The Sixth Annual

Federal Forecasters Conference - 1993

Papers and Proceedings

Crystal Gateway Marriott Hotel
Crystal City, Virginia
September 8, 1993

- Sponsoring Agencies -

Bureau of the Census ♦ Bureau of Economic Analysis ♦ Bureau of Mines
Bureau of Labor Statistics ♦ Central Intelligence Agency ♦ Department of Veterans Affairs
Economic Research Service ♦ Energy Information Administration ♦ Environmental Protection Agency
National Center for Education Statistics ♦ U.S. Geological Survey

THE SIXTH ANNUAL FEDERAL FORECASTERS CONFERENCE

PAPERS AND PROCEEDINGS

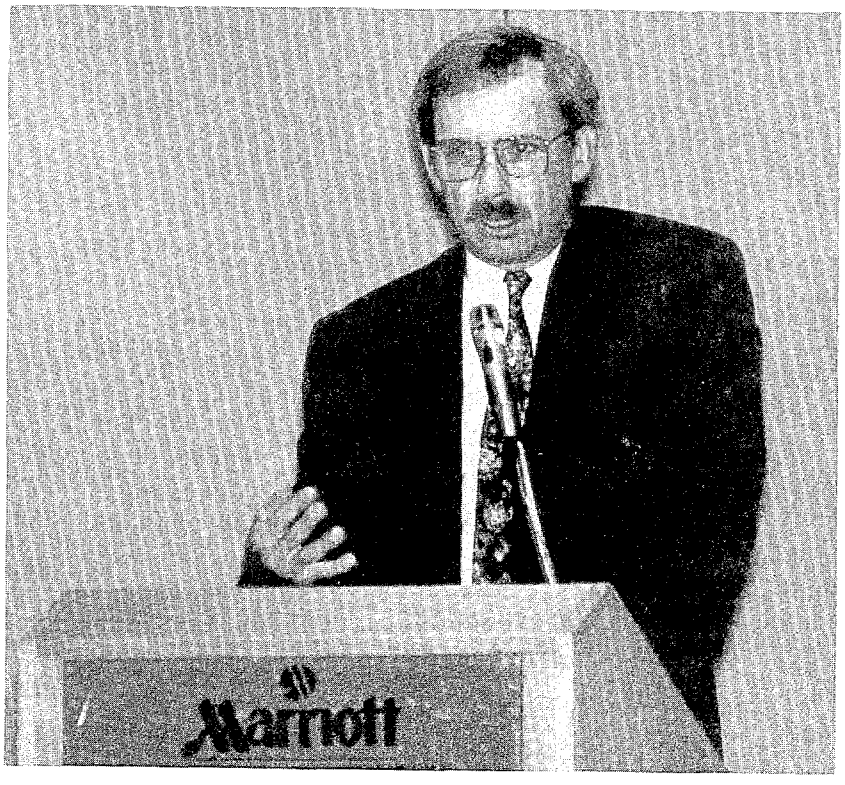
The Crystal Gateway Mariott Hotel
Crystal City, Virginia
September 8, 1993

Released for Printing: October 4, 1994

Printed with soy ink on recycled paper

SCENES FROM THE CONFERENCE

The 1993 Federal Forecasters Conference Organizing Committee



Larry Pettis, then Acting Administrator of the Energy Information Administration, delivers the opening remarks at the Conference.

TABLE OF CONTENTS

MORNING SESSION--FFC/93 1
Welcome and Opening 1
Award Presentations 2
Keynote Address 4
Panel Discussion 13
FFC-93 Survey Results 21

FORECASTING TECHNIQUES 23
John Parker, Economic Research Service 24
Alberto Jerardo, Economic Research Service 45
Andreas Muller, Department of Health Services Administration, University of Arkansas at Little Rock 66
Ramesh Dandekar and Nancy Kirkendall, Energy Information Administration 75
Harry Vroomen, Economic Research Service and Frederick Joutz, Energy Information Administration and the George Washington University 86
Kay Adams and Gus Mastrogianis, U.S. Patent and Trademark Office 98

LABOR AND PRODUCTIVITY 133
Howard N Fullerton, Jr., Office of Employment Projections, Bureau of Labor Statistics 134
Lea Swanson, Environmental Protection Agency 140
Lloyd D. Teigen, USDA Economic Research Service 146

FORECAST EVALUATION 155
Vernon Dale Jones and Stuart Bretschneider, The Maxwell School, Syracuse University 156
John Kitchen, Council of Economic Advisers and Economic Research Service 171
David Stallings and LaWanda Musgrove, World Agricultural Outlook Board, U.S. Department of Agriculture 181
Robin D. Reichenbach, Energy Information Administration 192

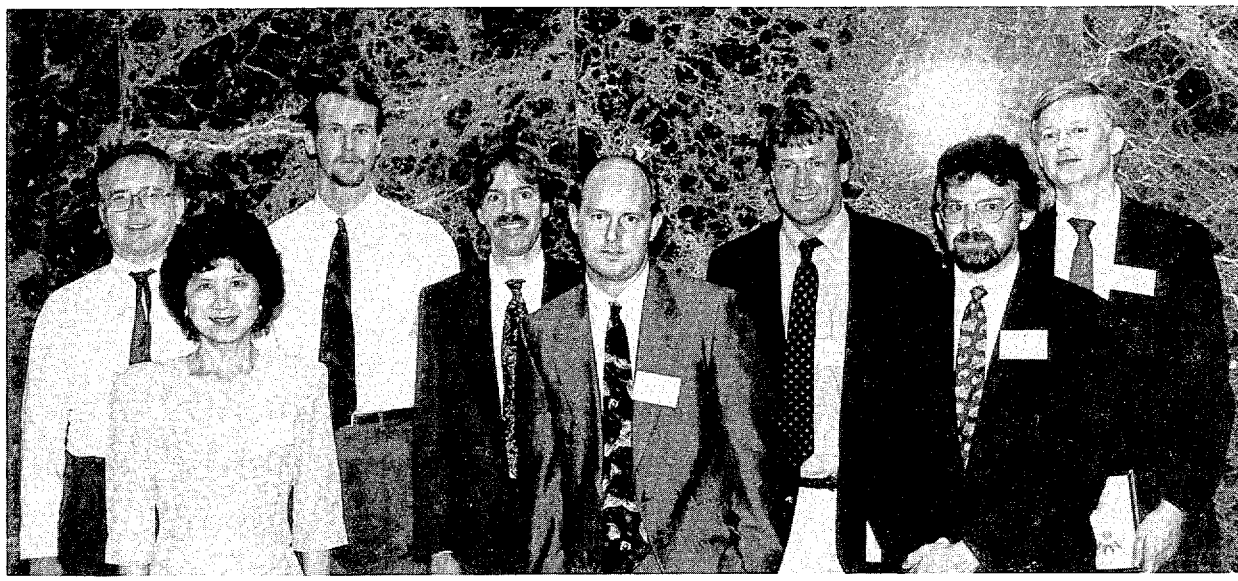
POLICY ANALYSIS 202
H.O. Stekler, Industrial College of the Armed Forces 203
Karen Fisher-Vanden, Jae Edmonds, and Hugh Pitcher, Pacific Northwest Laboratory 208
Jerome C. Glenn, United Nations Universtiy 245

EXPERIMENTAL ECONOMICS 260
Charissa Wellford, Federal Trade Commission 261
Steven Elliott, Oak Ridge National Laboratory 307
Jamie Kruse, University of Colorado 321
David J. Bjornstad, Oak Ridge National Laboratory 334

FORECASTING FUTURES MARKETS 340
Ronald A. Babula, Phil L. Colling, and Gregory Gajewski, Economic Research Service, U.S. Department of Agriculture 341
William Melick and Charles Thomas, Federal Reserve Board 355
David Torgerson, Economic Research Service 363

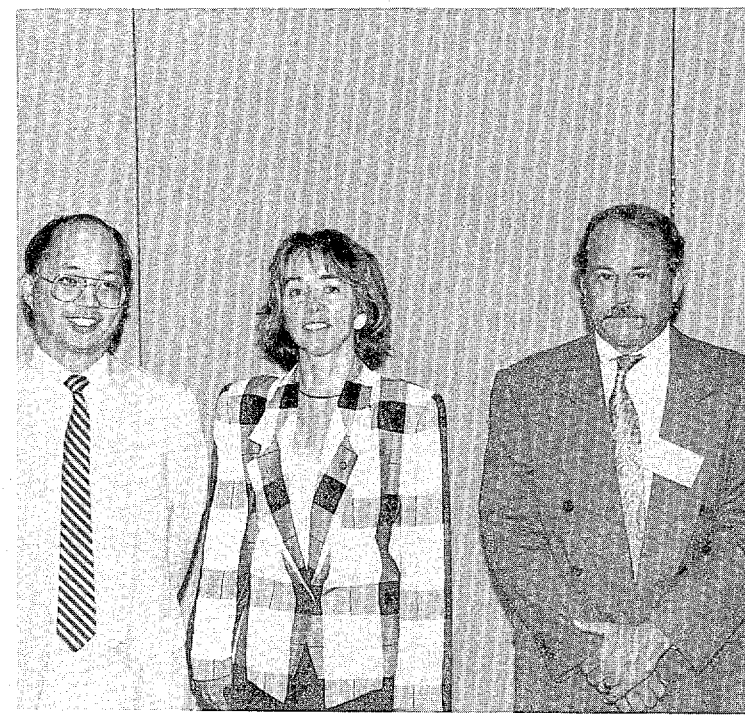
METAISSUES 366
George Wesley and Peg Young, Office of Inspector General, Department of Vetrans Affairs . . 367
Karen Hamrick, Economic Research Service, and Fred Mills, Federal Management Partners, Inc. 377

SCENES FROM THE CONFERENCE



Award winners from FFC-93.

Captain Dennis Egan of the National Performance Review delivers the keynote address.



The panel discussants on the "Role of Forecasting in the Anticipatory Government": Michael Woo, Professional Staff Member for the Committee on Energy and Commerce; Heather Ross, Special Assistant to the President for Economic Policy at the National Economic Council; and Mark Rodekohr, Director of the Energy Demand and Integration Division of the Energy Information Administration.

MORNING SESSION--FFC/93

Welcome and Opening

Mr. Earley: I want to welcome you to the Sixth Annual Federal Forecaster's Conference. I am Ron Earley and I am the Chair of this year's conference. This Conference is a collaborative effort between 11 different sponsoring agencies and over 50 attending agencies. This year we have approximately 300 registrants. We're holding true to form in terms of the growth and expansion of the conference. We have moved the site to a different facility this year. I would like to get some comments from you on this move.

I do want to say that it's a real pleasure to welcome you here today, and I think we have an interesting program. We've expanded participation this year to broaden it beyond the typical Federal base that we have. Represented here today are a number of private organizations, State Governments, as well as academia.

I particularly want to welcome a number of foreign visitors that we have today. A gentleman from BLS called me and said he had a training program and wanted to bring some of his trainees to the conference. So we have a large contingent of international representatives today. It was a little bit tough squeezing some of their names on their name tags. We had a little difficulty with spelling too, but I think we worked that one out.

I was talking to Debbie Gerald and she mentioned a story about a coworker who attended one of the town hall meetings at the Department of Education. At that town hall meeting, Vice President Gore asked for examples of things in Washington where Government employees developed and implemented programs. One employee spoke up and gave as an example the Federal Forecaster's Conference. I truly think that it is a good example of a grassroots activity built around 11 sponsoring agencies and all of the attending agencies.

I would like to pay particular thanks to those people who have served with me on the organizing committee this year. From the Bureau of Mines, Sandra Absalom and Ching Yu. From the Bureau of Economic Analysis, Tom Lienesch and Zoe Ambargis. From the Bureau of the Census, Paul Campbell. From the Bureau of Labor Statistics, Howard Fullerton. From the National Center for Education Statistics, Debra Gerald. From the Economic Research Service, Steven MacDonald and Karen Hamrick. From the Environmental Protection Agency, Joe Abe and Dave Rejeski. U.S. Geological Service, Ethan Smith. Department of Veterans Affairs, Peg Young, and finally, the Methodology Center of the Central Intelligence Agency as always, no named individuals. And of course, my colleagues at the Energy Information Administration, Arthur Andersen, Louise Bonadies, and Doug Hale. I also want to thank Kay, Brian, and Jason for all of their efforts while I've been preoccupied. A special "Thanks" is reserved for Sharon Shears, Shirley James, Linda Walsh, John Weiner, and Angela Renfrow for their diligent work on the logistics for this conference.

I would also like to pay particular thanks to the past chairs of this conference, Tom Lienesch from the Bureau of Economic Analysis, Norm Saunders and Howard Fullerton from Bureau of Labor Statistics, Karen Hamrick from the Economic Research Service, and Debra Gerald from the National Center for Education Statistics.

I want to thank you for your support and guidance during this process.

Last year's Chair, Tom Lienesch, operated under a little bit of a luxury in that he knew who the chairman for the next year's conference would be. I am not yet as honored as Tom was last year at this time. So what I will leave you with is the idea that the associations that I've had over the last six months in planning and working on this conference will be long remembered and pay benefits to me both personally and hopefully professionally. Thanks for the opportunity.

Now, to welcome us to this year's conference, I'd like to introduce you to Larry Pettis, who since January has served as the Acting Administrator of the Energy Information Administration and for many years now has been the Deputy Administrator of EIA. In these roles, he guides a very broad program of data collection, forecasting, and policy analysis. So to kick off the conference, I'd like you to please welcome Larry Pettis.

Mr. Pettis: Good morning. As I was thinking this weekend about meeting with a group of distinguished forecasters, I was reminded of an old saying which is "the most important thing is to know what you don't know."

Forecasting is a very humbling experience and we very quickly learn how much we don't know about the future, and I think it is in part that recognition that brings us all here today together. The Federal Forecasters Conference provides an excellent forum for us to develop professional relationships with other members of the forecasting community and to gain new ideas on how we can improve our forecasting activities.

The theme of this year's conference is the role of forecasting in Anticipatory Government. In their book, *Reinventing Government*, David Osbourne and Ted Gabler devoted a chapter to "Anticipatory Government, Prevention Rather Than Cure." The premise of this chapter is based on two points. One is that old adage that "an ounce of prevention is worth a pound of cure," and the other is that foresight must be built into the decisionmaking process.

With respect to forecasting, I'd like to read a quote from the book. It probably sounds familiar to some things you've heard before.

> Skeptics often argue that it is impossible to predict with accuracy ten years ahead. They are right. But the point is not to make projections that are 100 percent accurate. The point is simply to flag problems that loom ahead.

Most of the successes that are mentioned in this book are at the state and local level, and the authors point out a number of obstacles to Anticipatory Government at the Federal level. Included among those are the one-year budget cycle, accrual accounting which tends to focus on near term rather than long-term outlays, and the short-term political focus in Washington.

Despite these obstacles, as Federal forecasters, I think we can point to a number of successes where our work has allowed decision makers to make informed policy decisions that do have long-term importance to our Nation. Thinking about my own agency, there are a number of instances where our forecasting or analytical work related to those forecasts have been used in developing regulatory or legislative initiatives. Some that immediately come to mind are the tax reform back in 1986, the Clean Air Act amendments, and natural gas deregulation.

Last year at this conference, my agency, the Energy Information Administration, made a presentation in one of the concurrent sessions on an initiative that we've had underway in our agency for the past couple of years: the development of a new National Energy Modeling System, or NEMS, as we fondly call it. I'm pleased to report that at this point in time we're in the final stages of testing on this new modeling system and it will be used in our flagship forecast this year, the Annual Energy Outlook which will be developed later this Fall. We believe that NEMS is going to help us improve our forecasting and more importantly, understand some of the underlying trends behind that forecast.

But the goals of NEMS go beyond forecasting. NEMS has been designed as a policy analysis tool and we're optimistic that this will improve our ability to inform decision makers about how policy changes may change the future.

In closing, I would just like to recognize two groups that I think are in fact under-appreciated and under-recognized. First are forecasters in general. So I would like to acknowledge the fine work that you do as Federal Forecasters that helps both us as agencies and the policy makers in general. The second group which Ron has already alluded to are those people that get the pleasure of planning conferences. And I would like to acknowledge Ron and the coordinating committee for the excellent job they've done in putting together a timely and interesting conference for us today.

So with that, I welcome you to the conference and I hope that you enjoy it as much as I think you will. Thank you.

Mr. Earley: Thank you, Larry. I would like to turn the program back over to some of my colleagues. The contest that we embark upon every year is one of the fun aspects of the conference. I would like to have Karen Hamrick and Debbie Gerald come up and tell us who are the best among us.

Award Presentations

Ms. Gerald: Good morning. Welcome to FFC/93 and to one of its most awaited events. Believe it or not, we receive many calls regarding this contest and several of you take it rather seriously and we're glad to hear that. This is the third consecutive year that Karen Hamrick and I have done the contest. This year's contest has five items to forecast for various times during the month of August. They are: the U.S. civilian unemployment rate, the average bank prime rate, the cash price for an ounce of gold, the temperature, and the number of home runs by the home run leader in the American League.

I will be announcing the runner-ups. Karen Hamrick will announce the winner. Before we continue, we would like to give a special award for the entry with the most direct hits. This special award goes to Allan Eck, Bureau of Labor Statistics. Allan is not here but we'll make sure he receives it.

Next, I will announce the winners who will receive a Certificate of Honorable Mention. Please come up and receive your certificate as I call your name.

From the Bureau of Labor Statistics, Betty Su. This is the second time that Betty Su has received the certificate.

The next award goes to Thomas A. Amirault, Bureau of Labor Statistics.

The next award goes to Joel Green, Economic Research Service.

David Kass, Bureau of Economic Analysis.

The next award goes to Thomas Snyder, National Center for Education Statistics. This is the third year in a row that Tom Snyder has received the certificate.

The next award goes to Kenneth Beckman, Bureau of Mines.

Next is William Hussar, National Center for Education Statistics.

Next is Clifford Woodruff, III, Bureau of Economic Analysis. And this is also the third year in a row that Clifford Woodruff has received the certificate.

Next is Ken Wetzel, Environmental Protection Agency.

And last is Brian Unruh, Energy Information Administration.

Karen Hamrick will announce the winner.

Ms. Hamrick: I know this is the moment everyone's been waiting for.

The winner of our Federal Forecaster's Conference this year -- is John Burgan, from the Bureau of Labor Statistics.

Thanks to everyone who entered and congratulations to all our winners.

Mr. Earley: I'd like to introduce Tom Lienesch. Tom was the chair for last year's Federal Forecaster's Conference, and Tom has the distinct honor of giving some awards today for the papers that were presented at last year's conference.

Mr. Lienesch: This is the last few loose ends to tie up from last year's conference and before I get to the awards for the best technical papers presented last year, I'd like to first thank all the participants who either gave papers, participated in panel discussions, or otherwise contributed to last year's conference. You all materially contributed to what I thought was a successful conference and thank you, very much. I totally appreciate it.

Now to the awards. Borrowing fairly liberally from the criteria that Norm Saunders set out last year in setting up the award for the best technical paper, the general criteria of the selection was first, general relevance. The more people and programs that could use the knowledge, the better. And with that in mind, the paper that's on a generally applicable topic would probably fair a little better than one that was relevant only to a fairly narrow topic of, let's say, a particular program.

The second was the coherence of the paper, did it make sense? The third was the comprehensiveness of the paper, did it cover the topic well? And finally, the knowledge of the topic itself and related research. In addition to that, I thought it was fair to basically discount the entries of members of the organizing committee, so it didn't get like it was a little too incestuous. So under this scheme, two papers were deemed honorable mentions and one paper the award winner. There is no implied ranking of the honorable mentions. I think everything that was given last year was rather fine and it was a difficult task to come up with the winners. But again, I think the most important thing is everybody that participated, thank you very much.

In alphabetical order for the honorable mentions. If you're not here, they go to me by default. Let me get these certificates that Ron and Company drew up.

For Honorable Mention, Best Technical Paper, first honorable mention is for John Hisnanick from the Veteran's Administration for his paper in "Using ARIMA Models in an Unconventional Setting, Forecasting the Demand for Inpatient Hospital Services."

The second Honorable Mention is for Paul Sundell, Economic Research Service, for his paper, "Determinants of Short-Term Agricultural Loan Rates at Commercial Banks."

For the best paper of Federal Forecasters Conference 1992, I would like to award to Jeffrey Butler from Internal Revenue Service for his paper, "Unit Roots and Fractional Differencing of Time Series, Implications for Forecasting."

That wraps up the awards this year. Thank you.

Keynote Address

Mr. Earley: It seems as though, I don't use the term pejoratively, but old conference chairs, just never seem to fade away. They come back as contest winners and award presenters. It's somewhat remarkable to me. I keep on going back to the theme of the people who worked on contributing to this particular conference and you see many of those same names pop up. What was encouraging to me this year is the number of new people who joined organizing committee. I successfully succeeded in doubling its size, which is contrary, perhaps, to the initiatives announced yesterday for the entire federal government. Nonetheless, having more people serve improves the quality of the conference.

I do want to encourage other people to get involved in the Federal Forecaster Conference activities and if anybody is interested in participating in next year's session, let me know. It's fun to be with these people and talk to these people on the phone. And I very much appreciated Tom's efforts over the last year because the strongest communication that you have is the person who just ran the conference last year. Thank's, Tom, for your support and effort.

I want to tell you about the remainder of the day. When we were trying to design what the focus of this year's conference was going to be I had done some reading, basically in Osbourne's book and a few other books talking about reinventing government. When I read the section on the Anticipatory Government, I thought, you can't become an Anticipatory Government unless you are farsighted enough and you are forward thinking enough to anticipate some of the problems.

One of the things that struck me about the comments in the book is that much of what we do or what is perceived to be done in government at all levels, whether it be Federal, State, or local, is now dealing with problems after they've already occurred. In many ways it's much more cost effective and just smoother on the collective psyche if we can anticipate those problems before they occur and attack them before they get out of hand. Unfortunately, I think too many times we are now in a reactive mode and we should be attempting to try and resolve problems in advance.

So in order to accomplish that, I look at this audience of forecasters. In fact, one of the papers this afternoon asks should there be a new occupational category called forecasting? Well, when I saw that abstract, my first reaction was yes, definitely there should be, and they ought to get a one pay grade premium, mainly because of the inherent riskiness and uncertainty of their work. They're always guaranteed to be wrong but you want to try and be in the ballpark so that you can provide to decision makers good information about what likely will happen in the future. Lacking that information, and that information largely stems from people in this audience, lacking that information, bad decisions will be made. If they have good information, at least there's the chance that good decisions will be made. Whether those decisions will be made or not is sometimes more in the political realm and that may be a problem for us as forecasters, but our job is to provide as good a set of information as we can about what likely events are going to be out in the future.

We are involved in all aspects of this process. We are involved in doing the baseline forecast. We're involved in identifying alternatives to the baseline evaluating policy options. Another critical area is evaluating uncertainty.

One thing that has been the thrust of the conversation in this conference in the past and which I think in many agencies not enough effort is spent on, is evaluating our efforts. So we will talk about all of those things during the course of today, both in the keynote, in the panel, and then later on in the afternoon sessions. But the critical thing from my point of view is, in order to do our job well, we've got to make our findings constructive to the decision makers.

With that sort of preparatory comment, I would very much like to at this time welcome Dennis Egan as our keynote speaker for this year's conference. Dennis is a Special Assistant to the Deputy Director of Management and Integration for Vice President Al Gore's National Performance Review. In this role, he has been involved in the overall project coordination, management, production, and strategic communications networking. He is a captain in the U.S. Coast Guard and for the last two years has been a strategic planner on the staff of the Commandant of the Coast Guard specializing in areas of global economics and political and military affairs.

Captain Egan recently served as an economic consultant with the Defense Intelligence Agency in appraising economic forecasts for the People's Republic of China. He was the managing editor for a widely ranging future environmental scan, which was just recently published by the Department of Transportation entitled, "The Road to 2012."

In addition to his numerous degrees, Captain Egan is also a registered professional engineer in the State of Missouri. As you'll note in your program, he's also got a couple other lines on his vita that I would very much like to talk to him separately about. When he came out of the Coast Guard Academy he served with Jacques Cousteau in the Antarctic expedition. And then he also served as a maritime advisor to the Acquino Administration.

Captain Egan has certainly been around and seen a number of different aspects, both economic, political, military, both in the United States and around the world. Captain Egan is now representing his government by serving on the National Performance Review. And it seemed to me that as we look at the role of forecasting in the Anticipatory Government, that it was a natural to have somebody from the National Performance Review here to speak to us today.

So with that in mind, I am honored to present to you Captain Egan from the National Performance Review.

Captain Egan: Thank you. Ron forgot to tell you I've been around a long time. As a matter of fact, I was on Noah's Ark and I was a forecaster on Noah's Ark. It's true. You know, Noah had been going around ranting and raving for a long, long time. "It's going to rain, it's going to rain." Well, they hadn't had any rain in that area for years and he kept building that boat. He was sure of his forecast. Based on the projections at the time, all the king's forecasters said, "You're nuts, Noah. You're nuts. It hasn't rained for years and you say it's going to rain? Forget it." So he couldn't raise any federal Financing for his ark. And so he went out to the private sector and they weren't very interested either. So he kept building and building and he got two of every kind and finally, it started raining. After awhile there was enough rain and slowly the ark rose and we all took off.

Well, he needed two of every kind and so he had a forecaster there too. Normally they called them twocasters then because it was before glasses. But I wear glasses, so I was the first forecaster. Well, here we were floating along for forty days and forty nights. I said, "Noah, it's 8:00 in the morning, time for your daily forecast and here it is. Based on current projections, it's going to rain about another two feet today and tomorrow, more of the same, and I don't see much change out there frankly." Noah said, "Very well, keep it so." And the next day the rain stopped and soon we hit dry land. I got fired so here I am.

That's just to show you in a way the fact that as forecasters in an age of chaos oftentimes we are hit by things we didn't expect. And as things are changing so quickly today as professional forecasters, oftentimes we find ourselves more in the reactionary mode as a matter of circumstance. And to be able to cope with this type of environment, the best way for some of us to survive in that environment is to prepare our leaders and the people they lead for a variety of possible futures. And so they recognize the weak signals when they first start to appear and even though they may be reactionary, they've seen it before and they know how to handle it. That's about the best a lot of us can do right now. And some of us aren't even doing that.

But I had the privilege of serving on Al Gore's National Performance Review and as many of you have probably heard we presented the first of a series of products to the President yesterday. Fascinating, especially when you read the Washington Post. They're the newspapers that say, "Eh, nothing's new under the sun. We've tried it eleven times this century on the average of once every eight and a half years and if you figure that out, that's just a little longer than administrations, and we still haven't got it right."

Well, let me ask you, if you went to Las Vegas and rolled the dice and you finally rolled it eleven times, don't you think you'd come up with a seven? Well, we think we're going to benefit from the confluence of a number of interesting events, most of them not due to our own wisdom.

If any of you noticed since the walls came tumbling down in Eastern Germany and Berlin, we're in the midst of a global revolution. And it's a global revolution that's forced us to look back at the varied pillars of governments. Now what does it mean? A lot of things we took for granted in this country. We were beating a drum and shouting, "Democracy, democracy, whatever the forum may be." We're for democracy and we're against all that's not. And this really drove our foreign policy for years and years and years.

Well, an interesting thing happened. The walls came tumbling down. Eastern Europeans swarmed into the United States graduate schools looking to academia to figure out just what is this democracy? And they started raising questions that we haven't had to grapple with for years. And it raised a greater level of public awareness when Ted Turner's CNN was not only able to broadcast Madonna in little villages out in the jungles of Africa and the Philippines, but the counter effect occurred also. He also was able to bring these images from around the world

to show us what's what. And bring them into every house in America, in front of kids. Bring them into the Internet, a global spanning computer network where, even though many of us my age and beyond are computer frightened, we have a problem with computers and the Internet. We ask our teenage son or daughter to please hook us up and we have a wide, wide world around us.

The coup d'etat in the Soviet Union, as things were collapsing. How do you think the people in the Soviet Union found out what was going on? It happened in experiment with the Kosmodrome years before when a group of U.S. astronauts got together in partnership with Soviet Astronauts and they decided they were going to get a computer cable strung between Vienna into the Kosmodrome that would allow the scientists to use the Internet, to be able to share rocket launch information and be able to process it on U.S. super computers and European super computers.

Well, guess what? When the project finished, that cable stayed in. And the underground network of Internet conduit started branching out throughout the old Soviet Union and when things happened, you no longer had to rely on Pravda. Instead, you had packet message after packet message coming out on the Internet going out to the Western World, coming back, beamed by satellite, beamed by TV broadcast, beamed by radio, back to the very citizens that were real, live actors in a drama that unfolded before our own eyes.

So all of a sudden, governments are finding us much more difficult what they do and why, to hide the rationale from the people. And the people are getting smart. And we are seeing this in our own country. This last election was a prime example. These political trends where people all of a sudden developed an acute political awareness of such things as the budget deficit. The fact that we had this growing, growing monster that was increasingly limiting out flexibility to do with government what had to be done. They started limiting what we could do as far as making investments the things that are absolutely essential for our kids and for our own welfare. We just didn't have the financial flexibility to do it unless we fell back on old solutions or just taxed more. Well, the only problem was the economy wasn't healthy enough to just tax more. There had to be a new way of looking at things. There had to be a new way of approaching it.

So you look across the operation and you say, "Well, golly, we've taxed more. We've cut taxes. We've tried to cut government somewhat unsuccessfully by lopping off positions and trying to eliminate departments." We've got all sorts of approaches to this but the one approach we really haven't tried is something that organizations like General Motors, Xerox Corporation, many, many others learned the hard way in the latter '80s, and that's something called competitive awareness, a quality movement empowering employees and a new style of running business lean and mean, and down-sizing.

I don't think the Federal Government has been in this kind of a stark present danger situation before when down-sizing and the environment of down-sizing was so readily apparent to us. A quality movement and the types of management to be competitive and responsive to your customer have never been more clear. And this is an environment that is not just in the United States. It's an environment that's world-wide.

If you take a look from '89 to 1991 something happened that was absolutely amazing. If you go to France, Great Britain, Japan, Canada, or various leading industrialized nations around the world, there are huge vacuum in governance. Not government. There's plenty of that. But a huge vacuum in governance and people started getting angry. People got very upset and right now even you have some of the lowest popularity percentages of leadership across the world, not just in the United States, but across the world. People want government that works better and they want it to cost less.

Is this having your cake and eating it too? Not necessarily. I see some nodding heads in the room. And that is the type of conclusion that somehow or another we've got to grapple with because there are two different attitudes in America. One is very healthy. That's a healthy skepticism. The other is something that's grown and is very strong and apparent right now and that's cynicism. And when you have a healthy cynicism, you've got an oxymoron. You've got something that doesn't work. We call it a dog that doesn't hunt.

There's got to be hope out there and there is hope if you can get out of the way and you destroy all those hurdles that are out there and let Federal workers do what they've always wanted to do. The fact is you can trust Federal employees if you empower them; if you allow them to make the types of decisions they have to make; if you start dismembering this huge gorgon's knot and slice it. You need to remember your Greek mythology. You know that story very well. Alexander the Great came in and he saw this huge, huge knot that nobody could unsnarl before and he said, "I know how to do it" and he sliced right through it. No problem any more. But we're not talking about slicing right through it.

I'd like to rapidly go through some of the initiatives that were announced yesterday. I won't dwell too much on the types of things that you can see in the newspaper, and then I want to conclude by focusing on what does this

mean to planners and forecasters, because I think you're going to find that in a chaotic world despite the fact you may not be able to be proactive and really effect how things happen, you can influence both the people at work in Government and the leadership to be more sensitive to weak signals, to understand the type of environment they're functioning in, and give them a framework by which they can make this work. It's vitally important because we only have a few more years. This can't continue this way. We've got to make this work. The time has come.

The principles are pretty straightforward. Basically, a Government that puts people first is what we're seeking, cutting unnecessary spending, serving customers, empowering employees, helping communities to solve their own problems and fostering excellence. The products of the National Performance Review was a report called "From Red Tape to Results." And it meant that there is a way out, there is light at the end of the tunnel. This report contains recommendations to make Government more effective, efficient, and responsive.

At the end of this week you'll have a series of Executive Orders, and the report is one in which probably over half of the initiatives recommended can be initiated by Presidential Action. A series of Executive Orders 1) to mandate customer driven, strategic plans and organizations and 2) to create the President's Management Council, link the Federal Quality Institute into this organization and drive through the chief operating officers of each one of the departments and agencies across the board the initiatives and the spirit of reinvention of government.

There is a package of changes requiring Congressional action and we're working very closely on a bi-partisan basis, trying to get those through. It won't be easy but we've had an incredible ground swell of support in Congress in this measure. The problem was that we had basically industrial era bureaucracies in the information age. But I would like to suggest to you that the information age is over. And as forecasters, I would challenge you to come up with the idea of what's next. We've got enough information out there. And it's coming at us from every quarter.

Now, the question is what's next? And I would submit that it's probably an age of conversation and an age of communication, an age of consensus making solutions, an age of individual empowerment, individuals taking responsibility to make things work instead of coming up with excuses why they can't.

So we looked at a number of entrepreneurial organizations. We found a lot of successes. The Air Combat Command doubled productivity by basically empowering their workers and creating teams across the board, pushing decisions all the way down, removing the regulations that required 20 different signatures just to make routine decisions.

The Internal Revenue Service Center is competing against each other, using customer service performance standards as the marks of competition to determine their own success. Big success story over at IRS.

The Forest Service is streamlining itself again by squeezing middle management, going down to put the emphasis upon the workers in the field and giving them the ability and flexibility to make the decisions and to be able to excise their own budget.

Australia, Canada, New Zealand, Great Britain have all gone through the throes of agony of this type of transition. Canada recently issued their reinvention of government report about six months ago in the Mulroney Administration and it resulted in a 30 percent across the board cut in the operating cost of Canadian Government.

We're being a little less draconian in the United States, really. The winners had these common characteristics: they cut red tape, put employees first, they empowered employees to get results, and they cut back to basics. Are these just generalities? I'd like to get to some more specifics.

First of all, cutting back to basics. Basically requiring a report in 18 months on a base-closing initiative for Federal facilities, not for military facilities. Using more negotiated rule-making where you call it reg neg. And alternative dispute resolution techniques to bring the public, the stakeholders, to the bargaining table and become a good faith member at that bargaining table to hammer out regulations on a consensual basis, rather than a "King Solomon" basis.

Giving the President enhanced rescission authority is an old idea whose time has come. If we are really looking at running a brittle government, a government with the fat squeezed out, the President and his role of leadership has to have more flexibility.

Allowing all agencies greater freedom in setting fees for service and then also letting those agencies use those fees back in their own agency budget for reinvestment, to do their job better and at less cost. I don't know how this last one got thrown in there. That looks like that is probably too difficult.

Well, part of cutting back to basics and part of employee empowerment is also a matter of responsibility. You absolutely will not have the flexibility given the Federal employees for empowerment unless along with that goes flexibility. Right now, for instance, we found several situations where Federal employees had stolen from the Government and were not allowed to be released from their jobs by their supervisors until all avenues of a long,

excruciating process were fulfilled, resulting in 18, 24, 36 months after the fact. We need more flexibility. We need more flexibility on the part of management to handle these situations.

We need to improve the process for removing people who are no longer disabled from disability insurance rolls. The process takes incredibly long.

Organizations like Alaska Power that are prohibited by law from going into the commercial sector. Now we have a stabilized market in Alaska, to unload those types of initiatives back and make them competitive again.

Transferring law enforcement functions from the Drug Enforcement Agency to the FBI is just part of an overall initiative to try to reduce duplication and turf battles in the area of law enforcement. There's an awful lot of duplication and turf battles in the area of law enforcement.

Simplify the compliance and certification process, creating a National Spatial Data Infrastructure in cooperation with states and localities. Basically what that is, we've got an awful lot of Landsat information out there. We've got information that can give you images of everything down to your post office number out in front of your house. But what good is this type of information locked up in data banks out in Colorado? We're talking about releasing, getting this information out, getting it managed, getting it linked into state and local government entities where it can really be a tool for economic advancement.

Cutting red tape. I remember in the early '80s when the Federal Acquisition Regulations really came out in reams and reams and reams. And then I remember when the Department of Defense came out with their DAR, which was their own interpretation of that. Their own interpretation was about ten times the size of the original Federal Acquisition Regulations. I remember the Department of Transportation and the Department of Defense interpretation wasn't good enough for us so we came out with another half a dozen volumes called the TAR, and I think it was probably well named. What we didn't have was the feather, but right now we're going out with the feather. We're saying tar and feather the Federal Acquisition Regulation.

We need it simpler. A lot of other countries around the world have done just that. They've thrown out hundreds of pounds of procurement regulations. They made a very, very simple mission objective driven procurement processes with just a lot of authority delegated back to the front line management, front line workers.

Cutting in half the annual cost of headquarters staff and reducing federal government staff by 252,000 employees, which is about 12 percent. Instituting bi-annual budgets and appropriations, two year budgeting cycles rather than a one year budgeting cycle. And getting off of this micro-management by OMB and FTE ceilings that are placed in cost departments and really skew their operations in crazy sorts of ways ending up with lots of temporary employees that don't count against ceiling, all sorts of gyrations caused by FTE management, FTE floors. We're saying we need a process to get away from that and go directly to agencies that are driven by their overall operational budget themselves, and give the agency the flexibility to manage their own people, to hire, to fire, to recruit and these types of things. OMB indeed will stop using FTE ceilings and will instead use caps in operation costs.

We're talking about decentralling personnel authority to departments and agencies to conduct their own recruiting and examinations and simplifying the personnel classifications system, switching to pay bounds, just the personnel regulations. Yesterday it required two fork lifts bringing them out on the White House lawn. It was impressive. The piles were higher than this and that was just for a couple of agencies. I would have liked to seen if we had taken all federal agencies in the Washington, D.C. area and piled them all up. We would have had a mountain.

We want to reduce by half the time it takes to terminate employees. We want the agencies to roll over 50 percent of what they save in operation costs to the next year and establish a process by which agencies can more widely obtain waivers from regulations. And then we need to establish a cabinet level community empowering working group that drives deeply into the Federal relationship with state and local authorities on state local grant aid programs.

Putting the customer first. You're going to see an Executive Order requiring departments and agencies to create customer service programs equal to the best in the business. And this is going to require benchmarking, a skill that some forecasters are very good at. But it's going to be absolutely required to be able to survive as Federal agencies in this new environment.

We're talking about eliminating Government monopolies, such as the Government Printing Office and GSA, and having them have to compete against other alternatives given to the agencies. Create competitive one-stop career development centers. We've got in some states over twelve different employment programs with advisors, lots of field offices, and each one has its own select clientele and anybody else need not apply. And I'll tell you, I have some friends out in Oregon that told me some stories about coming into one and getting this walk around

where they finally visited eight different offices, finally to try to untangle what they had coming to and what they didn't. They finally gave up. You can cut a lot of cost by consolidating these types of field offices into one stop shopping.

Restructuring the Air Traffic Control system into a corporation, an initiative strongly supported by the airlines' industry and something that had been tried by several other nations successfully.

Issuing new accounting standards to identify the true unit cost of all government activity. This was an incredible realization as we went through agency to agency. We found we had no common cost accounting system. Nobody knew what the costs of operations really were. But they did know how much money they had to spend and that was what was driving things. No cost consciousness whatsoever. We've got to sort this out. We've got to come to a common cost accounting basis across Federal Government.

And finally, cross Government collaborative efforts to empower communities and strengthen families. You may have heard of worklife. Worklife, does that sound familiar? Putting emphasis in the work place of alternatives for child care? Putting emphasis in the work place for the family focussed type of benefits that are absolutely necessary to level out the working place with the stresses of today.

Enhanced technology, enhancement programs, like telecommuting, are being very heavily emphasized now as not only necessary because the environment and the congestion on the highways, but it's also family friendly. It also allows men or women to work from home and still be able to pull in a paycheck, still be close to their children. A lot of issues here and that's a good one.

Empowering employees to get results. I talked to you briefly about this President's Management Council. It's going to become a very important operative instrument of government. The President's Management Council will have a series of working groups on cross cut issues across agencies. Things like examining a new maritime policy for this nation. Things like rebuilding a new personnel system, a reinvented personnel system for this nation, rebuilding a new procurement system for this nation. These types of things will be driven by the President's Management Council which will be headed by Phil Lader who is currently the Deputy Chief of Management of the Office of Management and Budget.

Establishing performance agreements between the President and Cabinet Secretaries. There is an Act that started out as S-20 that got signed by the President in August, which is going to have a very strong impact upon forecasters and planners because it requires strategic plans to be developed in all agencies across the board. It requires at least every 3 years that these strategic plans be reinitiated, examined, shaken off and redirected. It requires that planning drive budgeting and planning drive budget execution. These are big things. And it requires that the budget execution be measured against the plans in annual reports submitted back to the President's Management Council.

So what we're seeing here is a shift from knowing how much money you have to spend to knowing what you have planned. Then measuring your results from your budgeting against the plan. And I think you can start to see the rationale for shifting to a two year budgetary process because it simply gives you a better horizon. Second of all, eliminating this business that you can't carry money across the fiscal year that results in this surge of spending at the end of the fiscal year. We're talking about eliminating that and allowing money to be carried across the fiscal year.

You saw that reduction in Federal workers. Part of that is due to an initiative to reduce the ratio of managers to employees from one to seven to one in fifteen. This is real cruel shock treatment across the board for middle managers. What it may mean is the whole concept of the middle manager is changed. The whole concept of middle manager is going to be pushed out to the field. A lot more hands on, a lot more closer to the customer contact. Corporation after corporation has found that this has been the necessity of operating in the type of environment that we are headed into in this 21st Century and I'll tell you, government simply has to do the same thing. We can't afford otherwise.

Initiating training at all levels starting with the top. We're talking about sensitizing our leadership with strategic management quality and information technology training because one of the realities we found was that you can't get there without very careful adaptation of information technology and I'm not simply talking about replacing typewriters with computers. I'm talking about conductivity linking up. I'm talking about the ability for employees to take personal empowerment to communicate with each other as Federal managers, Federal employees, to not have to go up and over hierarchial boundaries, to take the initiative themselves to enter the Internet and talk to each other.

Forming a labor management partnership. A labor management coalition work group will be reporting through the President's Management Council, will be taking a careful look at the many issues that affect workers in the Federal work place in this new environment.

Finally, creating a coherent financial management system. I don't want to get into the details on that. That will become more clear when that report comes out.

The schedule of events on what's next. The summary report called "From Red Tape to Results" is now available through the Government Printing Office and the Internet. Interesting, when the President announced acceptance of the report yesterday on the White House lawn, for the first time this report fired up on Internet and was distributed to a potential 6 million networks around the world within minutes of the President announcing it.

The number of copies that were printed by the Government Printing Office were 7,000. Those were gone in an hour and thirty minutes. People were calling like crazy trying to get copies of the report from around the country. And we said, "Talk to your librarian. Your librarian knows how to get on an Internet, if you don't." And we gave them dozens of different sites so they could download that document at no charge and have it in their hands immediately. We gave them the alternative. You can place your order with the Government Printing Office by calling this number and they will have it in your hands within four to six weeks after they have your money. This is truly an example of using technology to have government that works better at less cost.

It's time we had a new customer service contract with the American people. And we guarantee an effective, efficient, and responsive government. It's time we cut the red tape, trim the bureaucracy. It's time we took out of our bureaucracy the words "we've always done it this way." Those were the words of Vice President Al Gore and we're around to try to make this happen.

There's a very interesting follow-on organization called Net Results. That is suddenly appearing on the Internet and it's working its way from one forum, one conference to the next, to the next, just floating out there in cyberspace. And you can become a part of that simply by getting onto the Internet, getting your personal Internet account number from one of numerous ways and entering the discussion that's going to drive these issues deep into local government, state government, individuals all the way up to our leadership and our policy makers.

I want to close by focusing on what does this mean to planners and forecasters. I think you'll find that forecasting and planning has never been more vital to our success. Remember I was talking to you about forecasting for Noah and how I did a simple projection and based on my projection, I thought we were going to get a lot more rain. Well, that's kind of simplistic.

I think in an era of chaotic events, as forecasters to be able to survive and have our bosses survive, we've got to focus on something a little different than our own future. We've got to focus on our customers' future. And the biggest challenge of focusing on our customer's future is first of all knowing who that really is, and knowing that it's not just the people inside our organization as the quality movement in federal organizations have done very well over the last several years pointing out in developing management sensitivity to the fact that there are internal customers. But now we've got to start focusing on our external customers, that other 20 percent of our concerns.

We have to understand how that customer's future may change due to changes in the environment, changes in the economy, changes in the real environment around us. We have to understand how the very nature of a family has changed. We have to understand what this means as far as education and training. We have to think about such issues as health care, and how we're going to bear the burden of health care on this economy and in our organizations and what our responsibilities are, both as employers and also as servants of the public.

I think that focusing on external environmental trends is going to be an easy one for us because we're the best in the field in that but this business of doing GAP analysis, of determining where are we and were did we say we were going to be and being able to map that into the Federal Performance Reports that are submitted back up to the Office of Management and Budget depending on how your organization determines to do this, your services can really come into play in the area of GAP analysis and result measurements.

And then benchmarking, competitive awareness. The fact that you're going to have a lot of initiatives going on around Federal Government for cost servicing agreements. There are going to be a lot of other agencies out there to eat your lunch. You've got to make management aware of that. You've go to let your bosses know that. You've got to know what your competitors are doing. You've got to know who they are. And you have to focus on organizational survival by being the very best in the business.

Making best use of information technology. If you are not on Internet, if your agency is not in Internet, if you have your own internal E-mail communication system that can't talk to anybody, as a Federal employee, do not let that be an excuse for you personally not to enter that world that's out there because it's a tremendous world full of rich resources. You're going to find yourself expanding your capabilities expedientially as forecasters if you

learn how to use that Internet. You get out to the rest of the world, you start sharing, providing value added, you'll gain a lot more out of it than you'll put into it. I assure you.

There are organizations like Cap Access here in the Washington, D.C. area that can give you your Internet account free of charge. You can enter E-mail to anybody else that has a Cap Access account personally. We've got lots of details on all sorts of alternatives like this and I would implore you to get involved in this wave of reinvention by personal empowerment.

Finally, Government Performance & Results Act, Public Law 103-62 is the Act that I'm talking about. I don't know if everybody has looked at that or not. It started out as S-20 but it's very important to our collective future. If you haven't read it, take a careful look at it. It's implementation is going to be vital.

And finally, let me just say it's a pleasure to be here. I think you're in the catbird's seat in this new environment that we're entering into. I think it's clear and present danger. I think it's a world wide phenomena and I think we're extraordinarily lucky we live in these times where a rising tide raises all ships whether they intended to be raised or not. So thank you very much. It's a real pleasure.

Mr. Earley: At this time, if there are any questions from the audience for Captain Egan, we'd like to entertain a few.

Question: You've emphasized "customer" in the private sector as the person paying the bill. In the government, that's not the case. How do we deal with that problem?

Captain Egan: The question is that in the private sector, you deal with a customer because the customer is the one that pays the bill. In the government, you have a different situation. We all pay the bill.

Well, that's not entirely correct first of all. We're finding in an awful lot of cases that we have to do things in very specific way for a specific customer who was represented through the Congressional process and then micro-management is written right into our authorization. And whether we like to or not, whether it means good management or not, we've got to do it. And this is just one example of the type of entanglement we're trying to get cooperation from Congress. We're trying to do in an Administration that's got both Democrats and Republicans but in Congress for the first time, a Democrat in the White House plus a Democratic majority for the first time in twelve years. We think we have a rare consensus opportunity with some of the Perotites, in the feeling that Perot represented both sides of the political spectrum. We have strong Republican constituency that wants to have government that costs less and that does things better.

So what I'm trying to say is we've got the forces there that say, look, focus on the segment of the American population that your program was meant to serve. Congress, give as much flexibility as you can to those people who are performing that service and use the quality initiatives, use the professionalism that we all have if we're just allowed to function. And you're going to find a government that does things better, that can recognize a customer and this customer is the people we serve.

Why is it when we've got 2.1 million Federal employees that if I call most agencies in the Washington, D.C. area, I can't get a real person to answer the phone unless I know somebody in the agency. Then I can weed my way in, finally find out who's there. I mean, there's no excuse for that sort of thing. But that's what Americans are finding out about their Government. It's simply not responsive. The Americans are your customer but we're all smart people and I think we can focus pretty much upon who our target is as far as delivery of services go. We've got to know what their needs are. We've got to know what they need and don't need and we've got to have the flexibility to respond to that.

So this is not a matter of who pays. This is a matter of common sense. We've got common sense in our Federal, State, and local government. It's just the fact that we haven't been allowed to use it in the past. I think it's high time that we have that responsibility.

Next question?

Question: (Questioner not on microphone.)

Captain Egan: The question is one of time lines. First of all, over this next month you're going to have a series of agency specific and system specific monographs that get into some detail on what the initiatives are as far as procurement system, as far as finance from the government, management systems and these types of things.

You'll see the Executive Order for the President's Management Council stand up at the end of this week. And during these next two, three, four years, you're going to see the system start falling into place. You're going to see a lot of opportunities for employee involvement, for public involvement and basically this is going to take a process that has a horizon of about four or five years. And depending on the system you're talking about, depending on the flexibility of whether Executive Order or legislation or just simply bringing diverse opinions to a consensus, those dynamics have to have time. They can't be mandated top down. We've got to achieve consensus on it.

So figuratively speaking, I'd say a horizon of five years to get things in place but about 40, 50 percent is going to start the locomotion out of the station by Executive Order and by decision on the part of Cabinet Secretaries and administrators.

Question: (Questioner not on microphone.)

Captain Egan: The question as I understand it is how can we get more cross fertilization, I hesitate to use that word, cross fertilization from the corporate sector, from very good managers and consultants, I might say, into the Federal sector. It's a good question. This is really a question of training. And not only training, but I think probably employment opportunity on the part of those that are in that business.

How many of you have not heard of the Federal Quality Institute? Can I see a show of hands? Probably the majority of us haven't heard of the Federal Quality Institute. And I think that probably was because of the way it was handled by the previous administration, but it's alive and well and it was created a couple of years ago to bring the commercial sector, to bring business and business managers, those associated with the quality movement, these types of management techniques, to bring those types of educational resources to the table as far as agencies across the board that needed that type of assistance. With all the red tape cut, if you please, to be able to bring those types of people on board to help your particular situation out in your agency and be able to measure whether you're really accomplishing what you set out to do with your quality initiatives in your agency.

They're an important resource and their role is going to be considerably enhanced by linking their efforts to measurement initiatives with the President's Management Council and also increasing the involvement of public/private partnerships with the Federal Quality Institute in providing those services to agencies. So that's just one. There are many other initiatives I think that as Federal managers we will be able to bring on line, especially if some of the red tape can be cut to bring this stuff on line.

As far as employment opportunities, I really don't see a very bright picture out there in an environment of down-sizing for a lot of lateral motion in and out of the Federal work place. I see more out and less in but there's going to be some turnover and I think if you have these types of programs and you're emphasizing the importance of your management, at least as far as the training of those that are in government, the money is going to be made available as an investment to get those people trained in the latest management techniques and these types of things. So I see a hot and cold here, okay? The environment is not very conducive but the emphasis of getting to where we have to go requires it so I guess the answer is yes and no.

Question: These various suggestions that you want to reduce duplication and waste, and on the other hand you want to encourage bringing agencies together and doing the same thing. And you cite the law enforcement as one example. So on the one hand to get efficiency you create monopoly and on the other you say we don't want Government monopolies. You're going to have to have both. If you really want to encourage competition, you should encourage more Government agencies to do more things and compete with each other. And of course, if you really want pure and beautiful competition of the classical model, you should have so many of them that none can influence price.

Captain Egan: Thank you. That's great. Seriously, I think you're absolutely right, but it's situational. There are certain things in Government where you want to compete. Absolutely it's healthy. But you know the last thing I want is to have the police in Vienna competing against each other for who can write the most tickets.

The idea of having a police force compete against another police force to try to write more tickets, to enforce the law is not necessarily the competition you're looking for. You're looking for, I think, competition as far as alternatives of either cross servicing agreements, where on service provides a service internally for another agency because they can do it better and they can do it cheaper. And that allows the other agency flexibility,

reorganization to use their people better. This is just one example of the types of hybrid arrangements, I think, where competition can be very productive.

In other cases you may be looking at a mean and nasty environment out there where you have certain agencies trying to take over another agency's business as far as operations go. You've got them both doing it and after a period of evaluation or something like that, you say whoever wins will do it, whoever doesn't win has to get out of the business.

But I'm not talking about situations like we've had in the drug czar's office off and on for years where you've got turf battles just raging and unnecessary duplication of effort. I'm not talking about the border patrol and the customs and the types of turf battles that they've had that may not be too familiar to you but when we got into them, they were an incredible waste of money. So it's situational. You've got to take a careful look at it. Competition is great but competition can get out of hand too.

Mr. Earley: One last question.

Question: Over the years we have traded costs for the protection of certain groups. The minorities, women, disabled and so forth in our procurement policy and in our personnel policies. Are you willing now to give up all of those advances in order to conserve cost alone?

Captain Egan: No, absolutely not. There is a social cost of justice and there's an ethics of good government. And we're not going to turn our backs on that, absolutely not. Instead we've got to take a look at it from their point of view and their accesses to services in eliminating all the unnecessary steps they've got to do to get what their program should be offering them as a matter of law.

Unfortunately, a person that's in distress, a woman who may have children is out of a job, her house is going to be foreclosed on her, she has all these things coming down on her head in this environment and now she's got to go out and she's got to weed her way through 17 different programs just to get well baby money so she can keep her child healthy? I mean something's wrong there and there's just so much red tape. I think that if we can consolidate situations like that for one stop shopping on these types of programs along with the other service delivery programs we have, we're going to be doing a much, much better job.

Mr. Earley: I'd like to have us all thank Captain Egan for speaking to us today.

Panel Discussion

Mr. Earley: In thinking about the next part of the program, one way to view the discussion in terms of the role of forecasting in the Anticipatory Government is to take kind of a slice in time in the analytical time it takes to evaluate certain programs, to making a decision, and then eventually acting these decisions into law.

We all, as I mentioned earlier, have responsibilities for creating the initial baseline, evaluating various options, coordinating the analysis of these options across agency and across governmental lines and then working, of course, with Congress in terms of the ultimate resolution of these and the packaging these into programs that are put in Law.

So with that in mind, I would like to introduce to you Mark Rodekohr. Mark is presently the Director of the Energy Demand and Integration Division within the Energy Information Administration. He manages approximately 35 economists and other professionals. Dr. Rodekohr is responsible for mid-term and long-term modeling within the Office of Integrated Analysis and Forecasting.

He is also responsible for working on the EIA's Annual Energy Outlook and the International Energy Outlook. His major responsibility right now is the development of the demand and integration components of the National Energy Modeling System, an effort we've been working on within EIA for the better part of two years now. Prior to this current position, Mark was responsible for the International and the Contingency Information Division where he was responsible for much of the international data and analysis functions within EIA.

So to talk generally about the development of baseline and the analysis of policy options, I'd like you to welcome Dr. Rodekohr.

Dr. Rodekohr: Thank you. I'm afraid my talk is going to be a little bit more mundane than the previous one. I'm not going to attempt to reinvent government today. Although I probably will have a few things to say about

that because I just read part of Mr. Osbourne's book last night, and I found some interesting corollaries between what I was going to talk about and some things he said in his book.

I'm going to talk about two things this morning. One is creating a credible baseline projection, and that word credible is a dangerous word. And secondly, creating a policy analysis tool and I'll rely on our recent experience with reinventing our models. It's always a good idea to reinvent one's models. We seem to do it about every ten years and it's driven by necessity of changes in the policy environment as much as anything else.

So first I'll talk about establishing a baseline. What I'm going to do here is list a series of rules and talk about the rules and why we found that they appear to work. We've been issuing a projection of the energy economy since 1973 on an annual basis so we've got about 20 years of these projections under our belt at this point. Unfortunately I've been there for most of those 20 years so I also have a lot of experience there.

But nevertheless, a few things come to mind. One, unless you're dealing with short-term forecasts, just concentrate on trends. Trying to predict the timing and amplitude of cycles in the economy and energy prices is dangerous. It leads to a lot of controversy and you're seldom right. And it drags the discussion into an area where you'll find yourself ignoring the basic issues or policy issues that are of interest to people into talking about things of short-term nature that really policies can have little influence over.

So the moral of the story here is to concentrate on the trends and I'll show some graphs about that in a few minutes.

Secondly, explain differences from the conventional wisdom. Conventional wisdom is an amorphous thing. It changes every year but nevertheless, it's what many of the policy makers are thinking about, are being told is the way the future is going to be and right or wrong, if you're different from the conventional wisdom, you owe it to them, you owe it to the establishment of accredibility to explain why you're different and why the conventional wisdom is perhaps in error.

The third point is to always compare your forecast with those of others. I hate to say it, but quite frankly, people don't have a lot of confidence in federal forecasters. Their errors are noted more than their successes and therefore, it becomes imperative to understand and compare your forecast with those of others made by whatever party. But policy makers see others. They want to understand why you're different. They want to understand what the implications of those differences are for their policies. Does it make a policy a high risk option, a low risk option, whatever? And lastly, and this is sometimes over emphasized but uncertainty analysis can be very successfully used to incorporate other views.

A long term forecast of, let's say the economy, if you incorporate all the uncertainties, you explode your error bands on out, the policy maker will take one look at that and they'll say well, the economy can be between $5 and $15 trillion in the year 2010. Well, that doesn't tell them very much. That's not very helpful. It doesn't give them a sense of what they can do to enact a change or get to a position they want to be in. It just says nobody knows what's going on. So you have to be a little bit more specific than that.

Now, I want to talk about a phenomena. Being wrong is sometimes being right, or helping the situation. I haven't picked on our projections of world oil prices. I've used the International Energy Workshop poll, which is a group of people who project world oil prices. As you know, this is a very volatile market. The history has shown here the price shock of '73, of 1980, the resulting collapse in prices.

Nevertheless, projections made in the early '80s, the conventional wisdom was, well, oil prices have been going up and they're going to go up forever. And these projections were made in '81, '83, '85, '86. All that period of five years prices were going to go up, we had to do something about it. Well, the reality was, of course, the prices didn't go up and in fact, I've often said that if I walked into a forecasting conference in the early 1980s and said what actually was going to happen to prices did happen, I'd be out of a job. And I expect that's true.

Nevertheless, part of the reason prices didn't go up, is because these projections were saying, if something wasn't done, the status quo would be for this situation to happen. Policy makers saw that, said well, we can't live with that. That's too much of our GNP going to energy and governments in the U.S., outside of the U.S. Policies were enacted which helped to bring about the moderation in prices.

So it's an example of where being wrong is being right. You've pointed out a possible problem, policy makers reacted to that problem, and other events happened to make that not become a problem. And so it's important that you establish the reasons why you think things are going to go the way, if nothing else happens. Then the debate can take place. Then the debate can focus on, well, is this really going to happen if I don't do anything? And what do I need to do to effect it? And that's the important part of the policy maker.

Now, on to uncertainty. Everybody who's an oil market analyst has an opinion about what's going to happen to world oil prices. It's probably one of the most widely focused on variables. We've dealt with this

historically by using uncertainty to wrap a band around the projections of plausible prices and then talking about what happens if prices try to move outside of this range. And we've argued, and I think successfully, that market forces will tend to push them back in here. In reality, prices may bounce around in the future. I don't know. I can't predict the timing or the political events that will lead to supply disruptions or other things that will control the actual stability or instability of prices. But I can say with a fair amount of certainty it's unlikely the prices will lie outside of that band. That, to policy makers, I believe is useful.

The uncertainty range helps establish your credibility of your baseline. Even though people might not believe it, they're much more comfortable with the range that you see here and it helps add another dimension to your problem.

Now I'm going to turn to the area of policy analysis. And then again, what are the keys here to what we're doing. There's been a lot of talk this morning about customers and clients and so forth. I've used the term 'users' here, but you'll find that spending a lot of time identifying who uses your product and for what is very important. Then you've got to ask yourself as a manager, these views are sometimes competing and you can't make everybody happy. And so in a sense, part of your job is to decide where the emphasis should be, where the real policy focus is? At the federal level usually, because we are in the Federal Government. We constantly get comments, I need state forecasts or I want a forecast at the county level. Well, we can't do all of that.

So we have to make it clear who our users are and what is the goal of the analysis. And that's not so obvious and in fact, it takes a lot of work and a lot of thinking to define the goals of the analysis. For example, in the energy market our model probably projects thousands and thousands of different things. Only a few of those are really important to most policy makers. The level and price of oil imports is one of them that rightly or wrongly is a big focus. The cost of energy, the ability of the energy segment of the economy to damage the rest of the economy. Those are things that we've developed and started to look at with a lot of care and a lot of time.

That is, what should be measured is very important. Not everything should be measured or can be measured in your model.

Next, as an economist I have a bias and that is to focus on measuring the cost and benefits of proposed policies. Oftentimes this is not done. I was very dismayed last night when I was reading Mr. Osbourne's book on reinventing government. He noted several successes. One of them was a fire department in the west. They privatized the fire department. They made it mandatory that all homes have sprinkler systems and they said, you know it's a success. I mean, not as many buildings catch on fire. The fire department costs the city less. And that's exactly the problem with that kind of analysis. It's focused on the benefits. It hadn't focused on the cost at all. You don't put sprinkler systems in buildings for nothing. I mean, this is a cost. It may not be a cost to the government, but it's a cost to the consumer. No discussion of cost and benefits.

I find it extremely useful if in any analysis, people can step back and start to think about things in terms of costs and benefits. To whom is it a very important political issue and to society is it a very important economic issue.

Carefully explain the results in a simple, understandable manner. That can't be emphasized too much. Black box answers seldom make for a credible forecast or policy analysis tool. The killing words are "well the model says." You just can't use that as a descripter. I get in more trouble every time I say that and I keep trying not to.

The last thing I'm going to mention here is our reinvention of our model, the National Energy Modeling System and walk through an example of how we've been developing this and the kind of things we've done differently. And hopefully, this will be a little bit useful.

First of all, we took a lot of time in developing a simple statement of the objective of the model. What are we all doing this for? And this was to illustrate the energy, economic, environmental, and security consequences on the United States, not every country in the world, of various energy policies and assumptions by providing forecasts of alternative energy futures in the mid and long-term using a unified modeling system. That may seem rather simple and rather obvious but it does help to break the problem down and define it into something that's manageable. It's not designed to do everything for everybody. It simply can't. No model can do that.

Now, how did we go about doing this? Well, we started two years ago and we started with a small internal group that developed a series of working documents, sort of overall design documents. We also got the National Academy of Sciences involved. They gave us a lot of support and direction as to what the overall make up of the model should look like. What was good from our existing model? What should we keep, what should we throw away? How should we do it?

We then spent a tremendous amount of time and effort, and this cannot be over-emphasized, identifying your clients, to work with your client through your customers. And we started off with a broad-base. There is a tremendous industry in this country and around the world in energy. I think it's $400 to $500 billion industry in the United States. Something on the order of seven or eight percent of our GNP is devoted to energy related activities.

We then broke that industry down into the various components, the Governmental components, the state and local components, the industry itself, and those who study the industry. We attempted to go very systematically through each one of those components and to develop groups who review and coordinate wherever we could. In the department we had monthly working group meetings. We developed very detailed design documents before we actually built the model. And staff tells me we over designed it in a stack of reports about that high. It doesn't take a forklift to bring them in, but close to it. We developed very detailed design reports, which we distributed widely. I think the mailings were up to 8,000 or so the last time I looked, all over the place.

We developed an independent expert review program where we got outside experts in the various fields to come in and give us advice on the design and on the implementation of the models. We used a group of academic, business, and other associates, called the Energy Modeling Forum, to help us with part of the design and review effort. And we've had academic review as well, so we've tried to hit all parts of the community that have studied energy situations or energy issues and have an interest and an input into how they think it should be modeled.

We then developed outside users' groups. We went to trade associations. We went to, then again, back to academics. And we went to industry as well. And we've had a conference, of course, and we'll have another one, I'm sure. But anyway, as you might imagine, all of this took a tremendous amount of time, effort, but it was probably worth it in terms of we think we've got a model that has a fairly broad-base of consensus about what it should be.

Now we've made mistakes. We've probably gone off in a few wrong directions based upon some very strong comments from people that are very articulate but you're going to have some of those, no matter what you do. Anyway, the model is near completion, at least in its first version. It appears to be working as we think it should although it's certainly not without a lot of problems. But we do believe that it is a model that will be able to look at the issues of the '90s and not rehash the old wars of the '80s and the '70s.

But there has taken a tremendous amount of change in the energy industry in the last two decades. Some of that from a modeling sense, you can put behind you. And then you have to be forward looking. And we think that we've developed a way of doing that by carefully going out and talking to all of the constituents, people that use or have opinions about the forecasts and analysis and have incorporated their concerns into our effort.

I'd like to thank you for your time.

Mr. Earley: Next I'd like to introduce Heather Ross. Heather Ross serves as the Special Assistant to the President for Economic Policy at the newly formed National Economic Council. I somewhat pride myself in getting two of the new organizations in town on the dais today. I think it's good to hear from them. They represent some change here in Washington.

Dr. Ross has been involved in a lot of coordination efforts in her recent tenure here in the National Economic Council working on the Btu tax that was part of the original package submitted by the President. She is currently working on the initiative of reducing greenhouse gas emissions.

Dr. Ross performed public policy research at the Woodrow Wilson School and the Urban Institute. She joined the federal government working at the Senate Budget Committee, and then at the Department of Interior. During the '80s she moved to overseas assignments with British Petroleum and now she has returned to Government service since 1993 at the National Economic Council.

Please join me in welcoming Dr. Ross.

Dr. Ross: Thank you. My topic is coordination and from my arrival here over the last few months, I have to say that one of the things I've learned is that coordination is more than the mechanical coming together of people. We have a lot of working groups. There is a lot of convening and talking about where to go. There's not as much shared vision about where that is and that's what I'd like to talk about.

If we have a view of where we want to go, we can reinvent government to get there, but we really do need to start with a view of the future that people can rally around. And so when I saw the topic on Anticipatory

Government, I thought there really is in the country a new element seeping into our anticipation of the future. It's pessimism and it's fear.

This is not fear of an external enemy. We've coped with that. It's a concern about how our own economy is functioning, how society is functioning. And how if you have winners and losers in the new challenges to us, right among our own compatriots, how do we reach agreement on a way forward? The answer to that must be that we find some common ground in how to go ahead together. That is where people who look to the future come in. There are many people plying their trade. The importance now is to find some way to bring our views together into one that we have in common about what the future looks like and how we can prosper in it.

The opinion polls that we see show pessimism. Substantial majorities of people are concerned that we're on the wrong path. Confidence indices are down. And we have 20 year lows in insurance rates but not a surge of buying of houses or other major commitments. People are worried and there's a tremendous irony here because the world that we face is one that is better than we ever anticipated.

During the Cold War, we hoped to get people to develop along our model of democracy in markets and we've succeeded remarkably. There's major liberalizing trends throughout the world. Other nations are in fact becoming more like us and that is the challenge. They're coming our way and they're coming our way very rapidly.

The prospect for growth now in the Third World is six percent a year, maybe three times what the OECD is. And so we have an engine here of global growth. We have a center of gravity that is moving toward the developing countries. In 25 years they'll have perhaps 70 percent of the entire world GDP. This is an enormous change and it's enormous good news for us. It's a fact about the future that is central to our opportunities for success.

But we have to change in order to be part of that world. We have to move ahead as it catches up with us. It's hard to compete with a Mexican peasant coming down from the mountains and working in a plant. Maybe he can increase his productivity ten times over. We can't do those same things any more. We have to do new things.

We know that markets are really major drivers of those new things that we will do. But markets are in fact what are concerning people now with layoffs, with downsizing. We have to get through this period. We have to see into a future that looks beyond the transition that we are in now.

And so the idea that we will, in the work that we do, find a view of the future that we can share and that we can convince people is open to us and that we have the capacity to do things both in the private and the public sector to get there, is the most important decision that we face right now.

There is a wave out there that we have to ride, rather than trying to hold back those waters. And not only the actuality of the experience of that in the economy but the policy judgments to support that, are occurring not in some future time but right now.

The example I'd like to use is NAFTA, which is not this week but next week, as the harbinger. This growth in the Third World. Mexico is certainly in the forefront of that in terms of liberalization and in being right next door it is the instance before us. It's the event that we have before us right now. It's four percent of our economy, the Mexican economy, and pooling our future with theirs will be net gains for us in terms of growth of exports to their growing economy.

Those are net gains. That is, there will be some jobs that will not continue. Those have been estimated in the 100,000 to 150,000 level. This is almost nothing compared to a giant economy of our own and so it's not a make or break for us. This is an important juncture in the policy direction that we take, in the view that we have of the future. The symbolic importance of it is exceptional. We can choose a path that says we can handle what the world is becoming. We can be part of a virtual circle where we encourage that, where we are benefitted by the growth in the world generally and that we make the changes we need to in our own economy in order to bring everyone along so that they can be part of those gains and so we recognize a new world and we are going to stay in the lead in that new world.

The other side of that coin is a negative feedback loop where we say we can't handle the opportunity or the possibility of free trade with Mexico, which is an extraordinary statement for us to make, that we don't see how we can make adjustments in our own economy that will allow us to welcome the world that is coming toward us. We won't put in place the programs that we need to put in place in order to be able to stay ahead of the wave.

So we're looking at a pretty major struggle on NAFTA between these two visions of what the future holds and of our place in it. And this is not a conflict only of people's current interest. It's a conflict of world views. It's a conflict of how we see where the world is heading and where we are heading.

It's easy as we go into this to play on people's fears. In an economy where we've had real wage declines for a substantial part of the population for 20 years, alarming people is not difficult to do. The challenge is to see how to have this be an experience where people who have some constructive thoughts about the future and coordinate ways of facing the future can convince people that there are benefits for us here that are not worn off, that the path will be chosen, which is critical to a world that we're moving into, that there are gains and that the winners can compensate those who do not win initially. So that we don't have to fight it out over our current circumstances.

This may not seem to be the obvious thing to say to a group of forecasters. It's the essence of what looking at the future is, I think, for all of us, that we look out at the world and we see fundamentally how favorable it is to us. This is an enormously promising world. It's built in our image. It's trying to be like us. And so as we go about our work and try to look at what it is that we face and to develop a common view about it for all of us whose work is to try to anticipate, it's critical now that we see how we can make a new era of success for ourselves in this rather than a fight over the current concerns that we have about our future, that we still have a lot of opportunity to shape.

The only solutions to these things will be done collectively and so the set of people who are in the forefront of this have to be the set of people in the government who are working ahead within their own spheres to try to convince people about what the future is and how we can prosper in it. And so I am happy to be here and to try to put what appears to be these days a counter spin on what the future looks like and to wish all of you well in trying to convince people of an affirmative view of what we can do now together.

Thank you.

Mr. Earley: Dr. Ross brought up NAFTA. I think we've all heard the radio ad that plays on a lot of people's fears. I concur with her comments that it is our job to try to be analytically sound and to provide information to people who are making the decisions, to get rid of the hysteria of fear.

I would like to introduce Mike Woo who is our last speaker. Mike is a professional staff member for the Committee on Energy and Commerce, a position he has held since 1981. In this capacity he is primarily responsible for economic and regulatory analysis of a number of issues primarily focusing on energy, health and environment. In each of these efforts Mr. Woo is essentially involved in the development and analysis of policy options and probably as important if not more important, the development of consensus about what is a desirable policy option to put in place.

Mr. Woo served as a consultant to the Department of Energy through the firm of Sobottka and Company. He coordinated and assisted in the Interdepartmental Study of Contingency Planning Options to deal with severe oil shortfalls. He also served as a consultant to the Board of Education of the Chicago Public School System on the development of a desegregation plan there.

With all of the panelists today I've had some connection with them cutting through the issues and so it's with great pleasure that I introduce now, the legislative perspective on this Anticipatory Government. Please join me in welcoming Mike Woo.

Mr. Woo: Thank you. Let me first start out with a few comments about forecasting and then talk about the legislative view.

Earlier in my career, and I've been here almost as long as Mark, I focused more on economic analysis. I thought I was pretty good at what I was doing. However, I was put to the test with a group of assorted people who were interested in energy; Danny Boggs, Mike German, Dan Dreyfuss -- a whole bunch of us energy policy junkies. Each year during the early part of the Reagan Administration we would make a forecast of what the price of energy would be at the end of the year, how much imports would be, etc.

It was very sobering to contrast how good I thought I was when I actually had to make forecasts that were benchmarked against reality. I always ended up in the bottom half. However, even as bad as I was, others were worse. In the early '80s I also worked on the policy of whether or not we should legislate certain special provisions for the then potential Alaskan Gas Pipeline. An economic analysis I prepared concluded that if there was anything less than three percent real growth in the price of oil starting in 1980, as Mark showed was at the top of his chart, this pipeline was not going to be economic.

And as Mark showed, the forecasts of most people were way up there. I made the bald statement that it couldn't possibly be sustained at that level. The proponents of the pipeline got together a number of eminent energy forecasters and did a forecast of what the price was going to be. They showed "conclusively" that the price was

going to be $100 a barrel in 1990 and that the lowest it could possibly be would be $50 a barrel. That was in 1980 dollars and, of course, it wouldn't even make us close even in nominal dollars. So anyway, I've had some good and bad experiences in forecasting.

When I told a colleague of mine, who is a lawyer, I was going to speak at the Federal Forecasting Conference he said, "Wait a minute. Isn't that a contradiction in terms? Isn't that an oxymoron? Isn't that like military intelligence? Isn't there something wrong?" I said, "Wait, wait, wait.

Let's explore what he was talking about. One, maybe he was talking about Federal forecasting. I think it goes without saying that Federal forecasting is as good as private forecasting. Private forecasters use Government data and forecast. Government forecasters use private data and forecast. I don't buy the concept that the Federal Government is any worse than the private sector.

Maybe it's Federal forecasting. I think that may be where the problem is. Everyone remembers the forecast. No one remembers the caveats. The main caveat in all forecasts is there are no facts about the future. A very good Washington radio sports call-in-host is often asked to predict the outcome of games and he always says he can only evaluate strengths and weaknesses. And, then he makes a little forecast. However, he also says if he knew what the future was going to be, he would make one bet, retire, and become the biggest sports junkie in the world. Similarly, I think those who work in the Administration and Congress are often asked to forecast the future and most people remember the forecast but not the caveats.

However, saying that, the Congress depends on forecasts in much of its legislation -- from the Federal budget, to the environment, to health care. Let me give you some examples of what the dependency is and interaction between how good those forecasts are and what the policy turns out to be. The Federal budget is one of the more obvious examples. There is always a forecast of what the Federal deficit is going to be and the pressure always is to reduce the Federal deficit.

Now, if you can trust forecasts, that is, if you can trust what you think the baseline is going to be and you can trust what you think the policy changes are going to achieve, then often you depend on individual policies to achieve the overall goal that you're trying to achieve. In other words, you're trying to bring the deficit down so you enact a series of individual initiatives that you forecast will bring the deficit down. What has been shown, though, is that these forecasts have tended to be wrong on the low side, i.e., that the deficit keeps growing even though the forecasts would lead you to believe that if you achieve certain policies the deficit should start coming down.

What that leads to in terms of policy is a desire by a number of Members to put absolute limits on budgets. That was one of the battles that recently occurred in the House. Some of the conservative Democrats said what does CBO forecast health care entitlement spending to be? Okay. We won't cut that spending. All we'll do is limit spending to that level. The forecast acts as a spending cap. That in fact made it into the House bill and it was dropped eventually in conference because it violated the Byrd rule. But that type of analysis may come back again when we consider health care.

A different type of forecasting occurs in global warming. It's more physical, although, the underpinnings are still basically economic. It's essentially forecasting fuel use and efficiency of use. It often misses central questions like what are the emissions and a similarly difficult question about the future -- what are the environmental impacts of these emissions?

President Clinton committed to achieving a reduction in global warming emissions to 1990 levels by the year 2000. We face the same types of questions we faced in terms of trying to reduce the Federal deficit. A number of people will try to focus on individual policies. They say: here's what the baseline is and if you do these individual policies, you will reduce the emissions down to this level. Therefore, we should carry out these individual policies. There are those, however, who want to say let's put an absolute cap on emissions and reduce everybody's emissions to achieve 1990 levels in the year 2000.

I think the battle of forecasting has yet to be resolved. I believe the Administration will come up with a series of individual initiatives to try to achieve some reductions and the environmental community will be unsatisfied. Perhaps, the two potential options will battle it out next Congress.

A final area that I want talk about is how forecasts are used in the health care reform legislation. One of the main issues is the rapid escalation of private sector health costs. Continued growth in health care costs at the rate that has occurred over the last decade or two is unsustainable. If you do the graph, you go back 50 years, and project forward with a straight line, by the middle of next century, 100 percent of our money is spent on health care. While absurd, it nevertheless raises a series of questions about how sustainable this is and whether or not we need to change policies. My hope would be that we would be as lucky as we were in energy. We had that same

forecasts of unrelenting price increases. If it turns out that actual health care prices go down like actual energy prices went down, our policies will be a rousing success.

But anyway, the same questions are going to occur in health care that I talked about earlier. What's the baseline? What is that health care growth going to be? Is it going to sustain its double digit increase in cost? How do the policy changes that occur as a result of health care legislation change that baseline and again, the questions of relative effect and absolute effect. Should we try to achieve individual policies which will affect that result or put on absolute caps to try to keep that health care cost at a certain level.

It's also interesting to note that in terms of the overall policy, if health costs continue to go up at anywhere near the way they're going up now, the policies are probably not sustainable even with the changes that the President is talking about. If you talk about enacting an employer mandate to cover health care costs and health care costs keep going up as a percentage of payroll, it just won't work.

On the other hand, if the costs are contained and if anything like what happened to energy starting in 1980 happens to health care, then any policy will be thought of as a rousing success.

Let me summarize by saying everybody complains about Federal forecasting and, as I said, everybody remembers the forecasts and not the caveats. But having said that, it also is true that Federal forecasting is very integral to policy making in the Congress. Finally, I would like to say that I really liked Mark's presentation. I have been here all through that period of time. My chairman, Mr. Dingle, is considered to be the father of EIA and I have handled EIA issues for him. The way Mark described the NEMS model as being generated is the way that forecasting and policy models ought to be generated. The NEMS model will be used in the context of many policy discussions including global warming.

Thank you.

Mr. Earley: I want to thank the panel for their comments. The afternoon sessions will begin at 1:00 p.m.

FFC-93 Survey Results

Debra Gerald, U.S. Department of Education, National Center for Education Statistics and Karen S. Hamrick, U.S. Department of Agriculture, Economic Research Service

During FFC-93, we conducted a survey of conference participants. Our intent was to get basic demographic statistics on Federal forecasters. For FFC-93, 45 out of 315 conference registrants completed the survey form. Because there is no guarantee that this is a representative sample, we cannot apply the results to all Federal forecasters, or even to all FFC-93 registrants. We were pleased that so many people participated in the survey, and that most of the survey questions were answered by all of the respondents.

Over 60 percent of the respondents at FFC-93 had a degree in economics. Other fields represented at the conference were accounting, demography, geography, information science, mathematics, operations research, and sociology.

Nearly 30 percent of the respondents were female. Also 30 percent of the FFC-93 respondents were managers. On average, the FFC-93 respondents had nearly 12 years of federal service. This group of respondents had been forecasting for 9 years. Thirty-six percent had a Master's degree and 41 percent had a Ph.D. Of the FFC-93 respondents, 84 percent published their forecasts. Also 84 percent of the FFC-93 respondents did evaluations of their forecasts.

Most of the respondents include national forecasts in their scope of work. Regional/state forecasts concerned 36 percent of the FFC-93 respondents and local forecasts concerned nearly 20 percent of the respondents. Slightly more than 20 percent of the FFC-93 respondents were concerned with international forecasts.

In terms of primary forecasting techniques, respondents cited a variety of methods, including trend analysis, regression models, time series models, macroeconomic models, and judgement.

Among the issues facing Federal forecasters listed by respondents were data consistency and quality, forecast accuracy, budget resources, reliability of forecasts, and modelling uncertainty.

CHARACTERISTICS BY OCCUPATION

		TOTAL RESPONDENTS	MANAGERS	NON MANAGERS
YEARS OF GOV'T SERVICE	(AVERAGE)	11.9	15.9	10.1
DISTRIBUTION:	0-4 YEARS	27	8	35
(PERCENT OF TOTAL)	5-14 YEARS	34	31	35
	15-24 YEARS	32	54	23
	25+ YEARS	7	8	6
PERCENT MALE		71	77	69
AVERAGE GRADE (EXCL. EXECUTIVE SERVICE)		GS/GM 13.0	GS/GM 14.2	GS/GM 12.5
PERCENT GS/GM 13		22	15	30
ED HIGHEST DEGREE:	ASSOCIATE'S	2	0	3
PERCENT TOTAL	BACHELOR'S	20	15	23
	MASTER'S	36	38	35
	PhD	41	46	39
YEARS OF FORECASTING	(AVERAGE)	9.2	11.9	7.5
DISTRIBUTION:	0-4 YEARS	41	23	48
(PERCENT OF TOTAL)	5-14 YEARS	30	23	32
	15-24 YEARS	27	54	16
	25+ YEARS	2	0	3
PERCENT WHOSE FORECASTS PUBLISHED		84	92	81
PERCENT, FORECASTS EVALUATED		84	83	84
OF WHICH, PERCENT EVALUATION PUBLISHED		41	55	35

CHARACTERISTICS BY SEX

		TOTAL RESPONDENTS	MALE	FEMALE
YEARS OF GOV'T SERVICE	(AVERAGE)	11.9	12.4	10.4
DISTRIBUTION:	0-4 YEARS	27	15	58
(PERCENT OF TOTAL)	5-14 YEARS	36	42	17
	15-24 YEARS	31	39	8
	25+ YEARS	7	3	17
PERCENT MANAGER		30	31	25
AVERAGE GRADE (EXCL. EXECUTIVE SERVICE)		GS/GM 13.0	GS/GM 13.4	GS/GM 12.0
PERCENT GS/GM 13		25	30	9
ED HIGHEST DEGREE:	ASSOCIATE'S	2	0	8
PERCENT TOTAL	BACHELOR'S	20	12	42
	MASTER'S	36	39	25
	PhD	42	48	25
YEARS OF FORECASTING	(AVERAGE)	9.2	9.9	7.1
DISTRIBUTION:	0-4 YEARS	40	30	67
(PERCENT OF TOTAL)	5-14 YEARS	31	36	17
	15-24 YEARS	27	30	17
	25+ YEARS	2	3	0
PERCENT WHOSE FORECASTS PUBLISHED		84	82	92
PERCENT, FORECASTS EVALUATED		84	84	83
OF WHICH, PERCENT EVALUATION PUBLISHED		42	39	50

Note for tables: percent distribution figures may not add to 100 due to rounding.

FORECASTING TECHNIQUES

SESSION I

MODERATOR Peg Young, Department of Veterans Affairs

PRESENTATION 1	John Parker, Economic Research Service
PRESENTATION 2	Alberto Jerado, Economic Research Service
PRESENTATION 3	Andreas Muller, Department of Health Services Administration, University of Arkansas at Little Rock

SESSION II

MODERATOR Sandra Absalom, Bureau of Mines

PRESENTATION 1	Ramesh Dandekar and Nancy Kirkendall, Energy Information Administration
PRESENTATION 2	Frederick Joutz, Energy Information Administration and the George Washington University, and Harry Vroomen, Economic Research Service
PRESENTATION 3	Kay Adams and Gus Mastrogianis, U.S. Patent and Trademark Office

Forecasting Techniques

Use of a Trade Matrix System for Data on Saudi Arabia and Iraq

John Parker, Economic Research Service

Adequate trade statistics from the official sources in Iraq are difficult to obtain. Thus, a matrix of data from suppliers provides a tool to estimate Iraqi trade activity. Through July 1990, statistics for Iraqi trade in major agricultural commodities were available through the USDA Agricultural Trade Office reports from Baghdad. Since August 1990, economists and business people have not had the advantage of that source, and a computerized tabulation from trading partners provides the best estimate for Iraqi agricultural trade.

While trade books are published by Saudi Arabia, they are on a delayed basis, and 1991 was the latest official trade book published in late 1993. Saudi statistics for imports of many commodities are far below those reported by trading partners exporting agricultural commodities to Saudi Arabia. The matrix tables help economists to obtain more current and realistic estimates for Saudi Arabia which is the world's leading importer of barley and sheep.

It is possible to get an approximate idea of the quantity of food imported by a given country through compilation of export statistics from trading partners. This is a useful tool when the official statistics are either not available, or they report a number far below the expected range. The development of data which is in the appropriate range is a necessary step in building a history for production, trade, and consumption used in forecasting. FAO (Food and Agriculture Organization of the United Nations) in Rome, Italy, and USDA in Washington use the trade matrix system to obtain estimates for imports of specific agricultural commodities by certain countries. This paper reviews the use of trading partner data for obtaining estimates for imports of selected commodities by Saudi Arabia and Iraq, with a strong focus on cereals, vegetable oils and sugar.

The first step is to gather all the trade information which has been compiled from United Nations trade tapes and printed for the format compiled by the Economic Research Service of USDA. This provides trade information through 1990 for over 60 countries and for 1991 for over 33 countries. Second, manually add in statistics from available trade books of countries not yet compiled in the United Nations system, including FSU, Iran, and Yemen.

Third, add in data for countries not in either the UN system, or with proper official trade books, but with reliable unofficial reports on exports to Iraq, like Vietnam's rice. Fourth, make estimates for the transit trade through neighboring countries. This includes an evaluation of the transit trade through Dubai, UAE to Saudi Arabia. Many shipments destined for Aqaba, Jordan are known to be completely for Iraq. Fifth, review various publications using the trade matrix system, especially statistical bulletins of the International Wheat Council.

Saudi Arabia's Import Statistics and Matrix Comparison

Saudi Arabia was the leading importer of agricultural commodities among developing countries during 1990-92, with imports averaging nearly $5 billion annually (table 1) and (figure 1). Average prices paid by Saudi Arabia are near the world average, or slightly below because of the excellent Saudi banking system and modern inexpensive unloading facilities (table 2). Despite wheat exports which averaged about 2 million tons during 1990-92, Saudi Arabia's grain imports were over 7 million tons annually. Corn imports trended upward, with larger purchases from the United States, Argentina and China (figure 3). Feed imports allow local producers to provide about 40 percent of the poultry meat sold in stores. Imports of a wide range of products allow Saudi supermarkets to have the best selection of food available to customers in the Mideast at relatively low prices. Intense competition among supermarkets means low prices.

Statistics reported in the official Foreign Trade of Saudi Arabia through 1991 came within 25 percent of the range of the tabulation obtained from a matrix of suppliers for livestock products and horticultural items. However, the official statistics for imports of barley, corn, rice, and sugar were less than half the matrix total during 1988-91.

This means that the matrix estimates must be used to develop an appropriate data base for forecasting future Saudi Arabian grain imports. Saudi Arabia is expected to remain a growing importer of feed grains, and in the next

decade it is likely to be one of the largest importers of barley in some years, along with Japan and FSU. Some forecasts indicate that FSU feed grain imports are expected to decline as domestic production rebounds. This will cause Saudi Arabia's share to total world imports of feed grains, particularly barley, to rise.

During 1992, Saudi Arabia was the leading world importer of barley, purchasing nearly 4 million tons from the European Union (EU) alone (table 3). Imports of over 2 million tons from other suppliers pushed total Saudi barley imports to over 6 million tons in 1992 (figure 2). U.S. barley exports to Saudi Arabia declined from 1 million tons in 1991 to 669,000 tons in 1992. Canadian barley exports to Saudi Arabia averaged about 1 million tons annually during 1990-92. During 1988-91, the official statistics reported for barley imports were less than a fifth of the amount reported shipped to Saudi Arabia by trading partners.

The Foreign Trade of Saudi Arabia reports imports of corn and sorghum together, but the numbers are usually less than half the matrix number for exports of corn to Saudi Arabia. As a matter of fact, U.S. corn exports to Saudi Arabia are usually greater than the officially reported imports from all sources. In 1992, Argentina exported 296,000 tons of corn to Saudi Arabia, while U.S. exported 497,000 tons. Shipments by China, EU and other suppliers led to a combined matrix total of over 950,000 tons for shipments of corn to Saudi Arabia in 1992. Saudi Arabia's official statistics tend to record over 90 percent of the imports of basmati rice reportedly shipped by India and Pakistan, but the number for imports of rice from the United States and Thailand is less than half the expected level.

Saudi Arabia imported about 625,000 tons of rice in 1992, and during 1988-91, its imports exceeded 520,000 tons annually, about double the official import numbers. Saudi official statistics for imports of rice from Thailand and the United States are usually well below the amount shipped by these two suppliers.
In the last several years, India emerged as a leading supplier of Saudi rice imports (figure 4).

One reason that the statistics for Saudi Arabia's imports of specific commodities may be that they are recorded in a miscellaneous category, rather than that for the specific commodity. Efforts to use computers to classify where imports of a given commodity should be recorded may have may have aggravated the problem. Through 1986, official trade statistics for some commodities like barley were in an acceptable range, but then the import numbers plummeted in the late 1980's. The number of personnel to review the trade statistics for accuracy were apparently reduced, and errors in commodity classification of import invoices for the trade books increased.

Detail statistics for many commodities are provided in Foreign Trade of Saudi Arabia according to the Harmonized Statistical System. This includes statistics for imports of items like peanut butter and chocolates which are in an acceptable range. However, imports of frozen potatoes reported in the book are less than a third the shipments to Saudi Arabia reported by the Netherlands.

Matrix Data Used to Review the Rise and Fall of Iraqi Imports

The trading partner matrix data indicates that Iraqi agricultural imports increased sharply in the 1980's and peaked at about $2.8 billion in 1989 (table 5) and (figure 5). Imports were close to the same pace in the first half of 1990, but declined steeply after August 2. The sharp decline resulted in a setback in consumption of most food items, particularly livestock products. The trade matrix indicates that Iraqi agricultural imports declined about a third in 1990, and fell steeply to less than $800 million in 1991, before making a rebound in 1992, with larger imports of wheat, flour, rice, sugar, and potatoes.

The decline was virtually complete for Iraqi imports of mixed animal feed, corn, and soybean meal. This resulted in a crash for Iraqi broiler meat output (figure 6). An unusual surplus and intense competition among rice producers in Southeast Asia worked to Iraq's advantage, with very low rice prices and a willingness of exporters to provide credit. Iraqi rice imports declined in 1991, but rebounded in 1992 (figure 7), and reportedly reached a record in 1993 with Thailand and Vietnam as major suppliers (table 4). Wheat imports have remained far below the peak levels of the late 1990's. Although Australia has resumed shipments on a reduced level, no U.S. wheat shipments have been reported since the summer of 1990, causing total wheat and flour imports to lag (figure 8). This combined with only token feed grain imports caused total grain imports in recent years to remain far below the peak of 5.4 million tons in 1989 (figure 9).

Nearly a third of the agricultural imports came from the United States in the late 1980's, but other suppliers were available. Total Iraqi imports declined sharply during late 1990 and fell below $800 million in 1991 (table 5). no U.S. food exports to Iraq were reported in 1991, compared with $808 million in 1988.

Regardless of where the basic food commodities came from, Baghdad's planners were determined to reduce shortages and to provide a reasonably adequate diet for the people through 1990. Yet, imports luxury items like

processed foods were restricted soon after the war with Iran started in September 1980, and particularly after 1983, with an exception of limited imports from Turkey for foreign technicians and hotels.

Iraqi agricultural production could not keep up with the rising demand for food in expanding urban areas as imports lagged in 1991. Through 1990, an artificial boom evolved in the cities related to government spending for projects and for the military. Efforts were made to greatly improve housing, creating a lot of jobs in construction. In addition to the rising population, the average diet improved as incomes increased and income for the poorer urban population improved. Regulations and disputes over land ownership tended to disturb plans for expanding food production by Iraqi farmers. After 1984, programs focused on raising producer prices helped to boost production of poultry products and vegetables helped increase output in central Iraq.

Trade statistics from matrix tabulations are used to prepare food balances for Iraq, in combination with data on production of specified commodities. Statistics from the Iraq Statistical Abstract through 1991, provide information on grain production by province and production statistics for many specific crops provide an indication of what was happening in the 1980's. For 1992, official data is not yet available from Iraq Statistical Abstract, although some preliminary estimates for cereal crops harvested were made by the FAO office in Baghdad. The data indicates because of modest grain yields, and less than half the normal imports, that the average Iraqi diet in 1992 contained about a third fewer calories than during 1990.

The use of the official Iraqi agricultural production numbers with estimates for imports developed by FAO from trading partner information clearly show the growing reliance on imports in the late 1980's, and especially in 1989 when drought reduced Iraq cereal yields, which lagged behind neighboring countries.

Neighbors Maintain Significant Trade with Iraq

Exporters in neighboring countries have been more aggressive in obtaining United Nations waivers for exporting food to Iraq. Many small traders have found Iraq a convenient market. Since Turkey and Jordan depend completely on imports for petroleum, and the low prices and convenient location of Iraqi supplies is a major reason why traders are willing to cope with many problems encountered in trading with Iraq. Jordan has continued to buy most of its gasoline needs from Iraq, while Turkey has turned to other suppliers, particularly FSU, GCC, and Libya.

Turkey is the leading source of Iraqi food other than wheat during most of the summer months. The border trade with an exchange of seasonal fruits and vegetables from Turkey moving by truck through Silopi Pass to Iraq rises in the summer, and declines by January. The trucks return to Turkey with as many drums of gasoline as possible. The price for gasoline in Iraq is less than 30 cents per gallon, compared with $4 in Turkey. Kurds have tried to get a toll for trucks passing through their territory, since Saddam Hussein has his own boycott of trade with the Kurds, and forbids the sale of gasoline to Kurdish truck drivers.

Turkey made large exports of many items to Iraq during 1981-83, but reduced deliveries during 1984-89, prior to the sharp rebound in 1991. Prior to the Gulf Crisis, the barter arrangement which allowed Turkish traders to deliver many commodities to Iraq and then receive payment at Turkish banks handling accounts for Iraq's petroleum exports ran into problems. The Turkish currency steadily declined in value. Slowness in paying them through the barter arrangement sharply reduced their profits. The traders shifted heavily to exports to less troublesome markets, where prompt payments in foreign exchange were made. Turkey's agricultural exports to Iraq declined from $346 million in 1983 to only $61 million by 1989. Turkey's deliveries of food to Iraq increased sharply in 1991 and remained strong in 1992. Official Turkish data show large exports of wheat and flour to Iraq through Aqaba in 1991 and early 1992. Turkey's exports of sugar and tomato paste to Iraq were strong in 1991 and 1992.

Since Iraq delivers about $300 million worth of petroleum products annually to Jordan, and gets Jordanian currency as payment, food purchases in Jordan are easier to obtain than elsewhere. Jordan's agricultural exports to Iraq trended upward in the late 1980's, consisting mostly of wheat flour, eggs, and processed foods and beverages. Jordan's agricultural exports to Iraq reported in its official statistics rose to $64.2 million in 1991, compared with an average of $22 million annually during 1987-89. A striking increase in exports of vegetables from Jordan to Iraq occurred in 1991, when the quantity reached 100,000 tons, including 57,000 tons of tomatoes. While Kuwait banned vegetables from Jordan and 90 percent of the Saudi business was lost in 1991, Jordanian vegetable exports brought $275 per ton, compared with $119 per ton in 1989 when Kuwait and Saudi Arabia were major markets.

For political reasons, Iraq had little trade in food with Iran and Syria during 1980-89. However, imports of apples, pistachios, raisins, and dry apricots from Iran increased in 1991 and 1992. This occurred because Iraq suffered a severe shortage of luxury food imports at the same time that Iran was increasing exports of these commodities. Kurds found it advantageous to begin some trade with bordering Syria in 1991, since the ban on trade with them by Saddam Hussein for the rest of Iraq became a problem, and they no longer had to abide with his ban on trading with Syria. Kurdish farmers derived considerable benefit from imports of wheat and vegetable seed from Syria, allowing them to increase yields.

Conclusion

The use of a matrix for trading partners will allow economists and statisticians to make estimates for imports of specific commodities which are in the expected range. The estimates are subject to change, because it may be discovered that some significant suppliers were left out of the beginning matrix. Yet, the use of the matrix estimate contributes to the development of historical data to make reliable forecasts.

This helps to get a better measure of world imports demand for selected commodities, and contributes to a forecasting system which is more polished than data systems using official statistics only.

Since a number of countries are late in reporting trade data, the matrix is adjusted with manually added numbers used to provide numbers for imports in recent years not yet recorded in the electronic series, which would show up as a 0 until the country data is put into the system. For example, Vietnam provided good data on rice exports to Iraq through 1992, but its data is not yet properly recorded in the UN trade data system. Another major tool in filling the gap for Iraq trade during 1980-90 was the trade book from Moscow reporting trade between Iraq and FSU, which was never a part of the electronic data base.

Figure 1

SAUDI ARABIAN AGRICULTURAL IMPORTS

By Specified Commodities, 1987-92

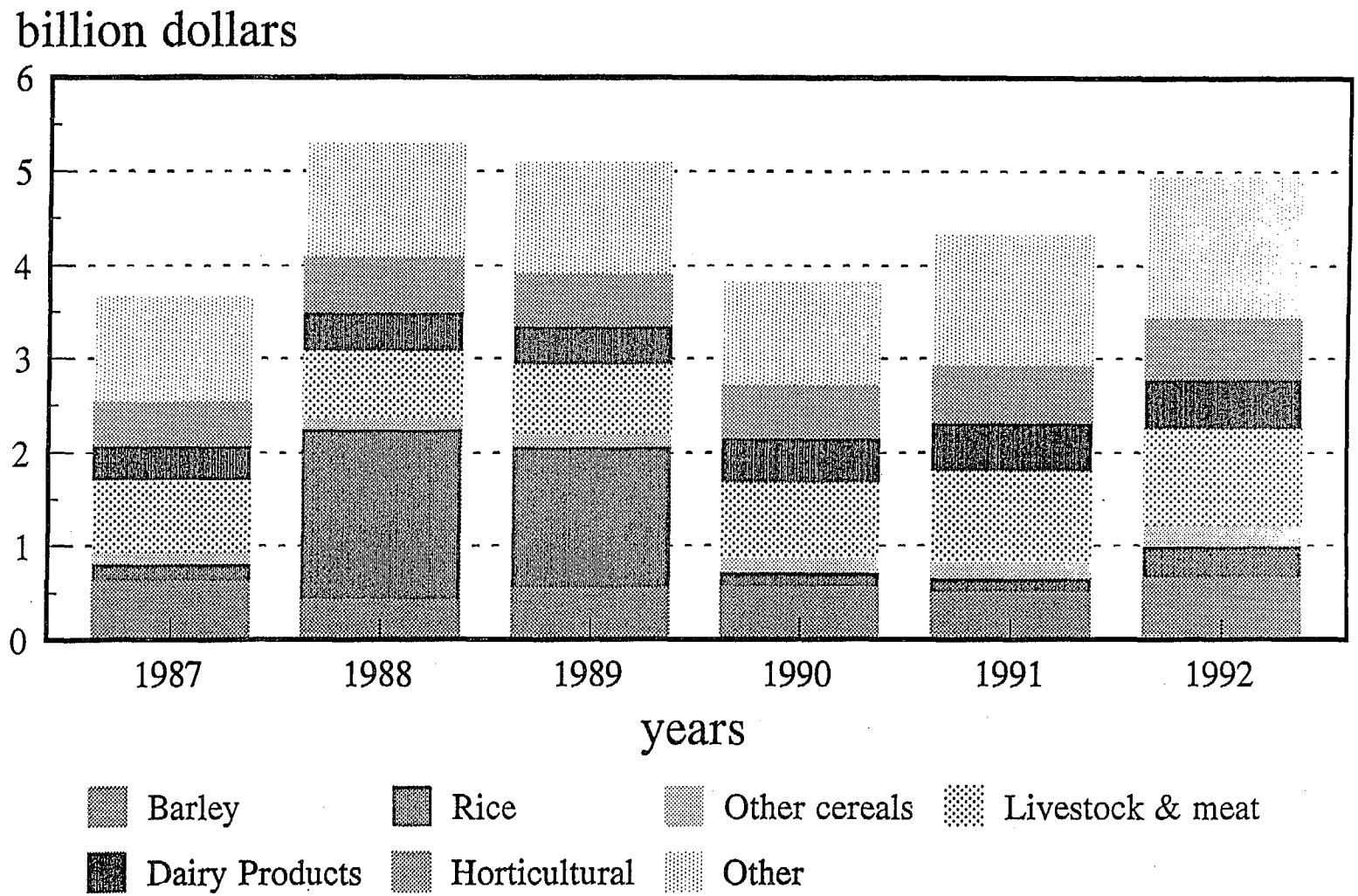


Figure 2

SAUDI ARABIAN BARLEY IMPORTS

From Specified Suppliers 1985-92

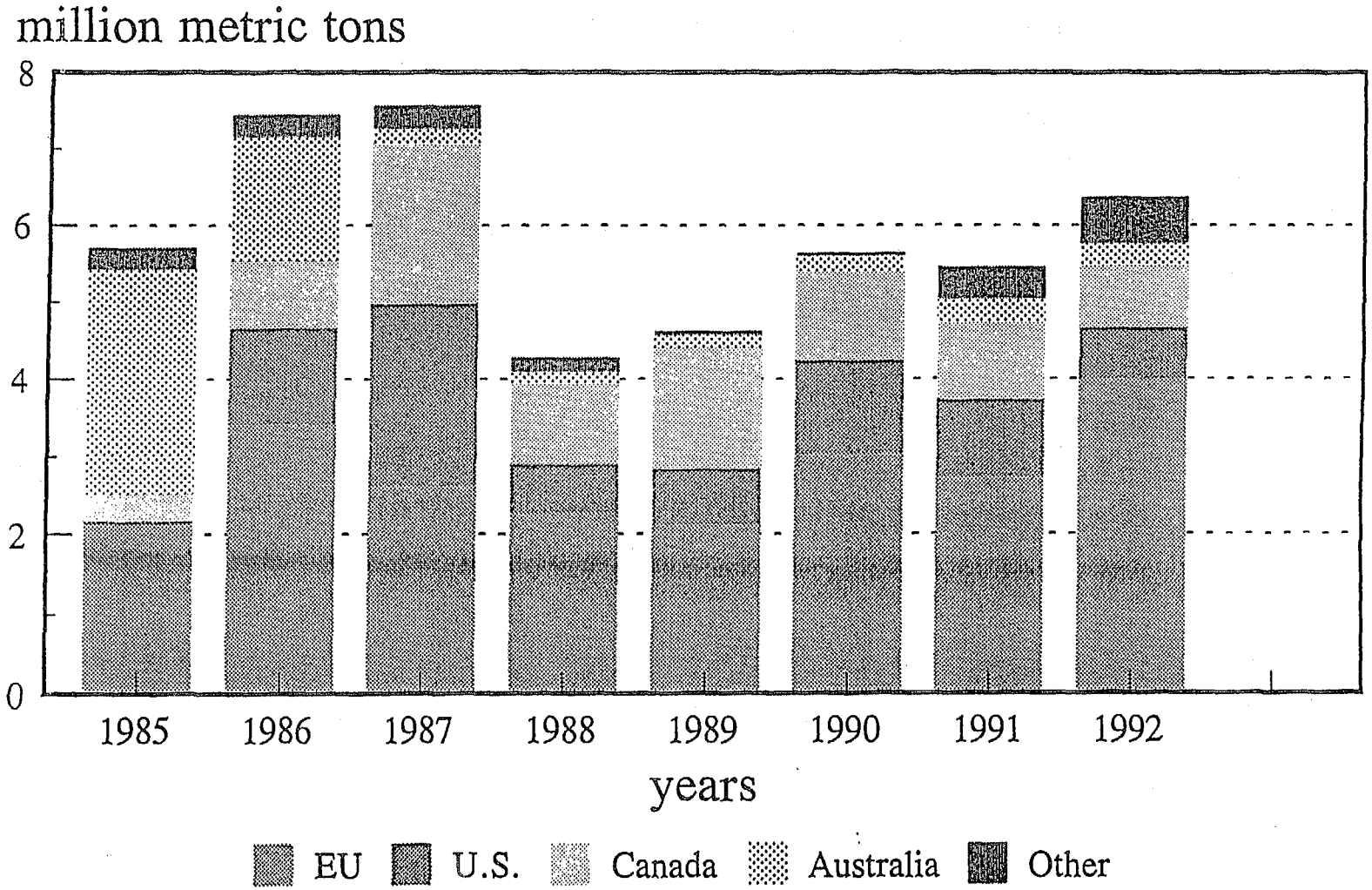


Figure 3

SAUDI ARABIAN CORN IMPORTS

From Specified Suppliers, Annual 1985-93

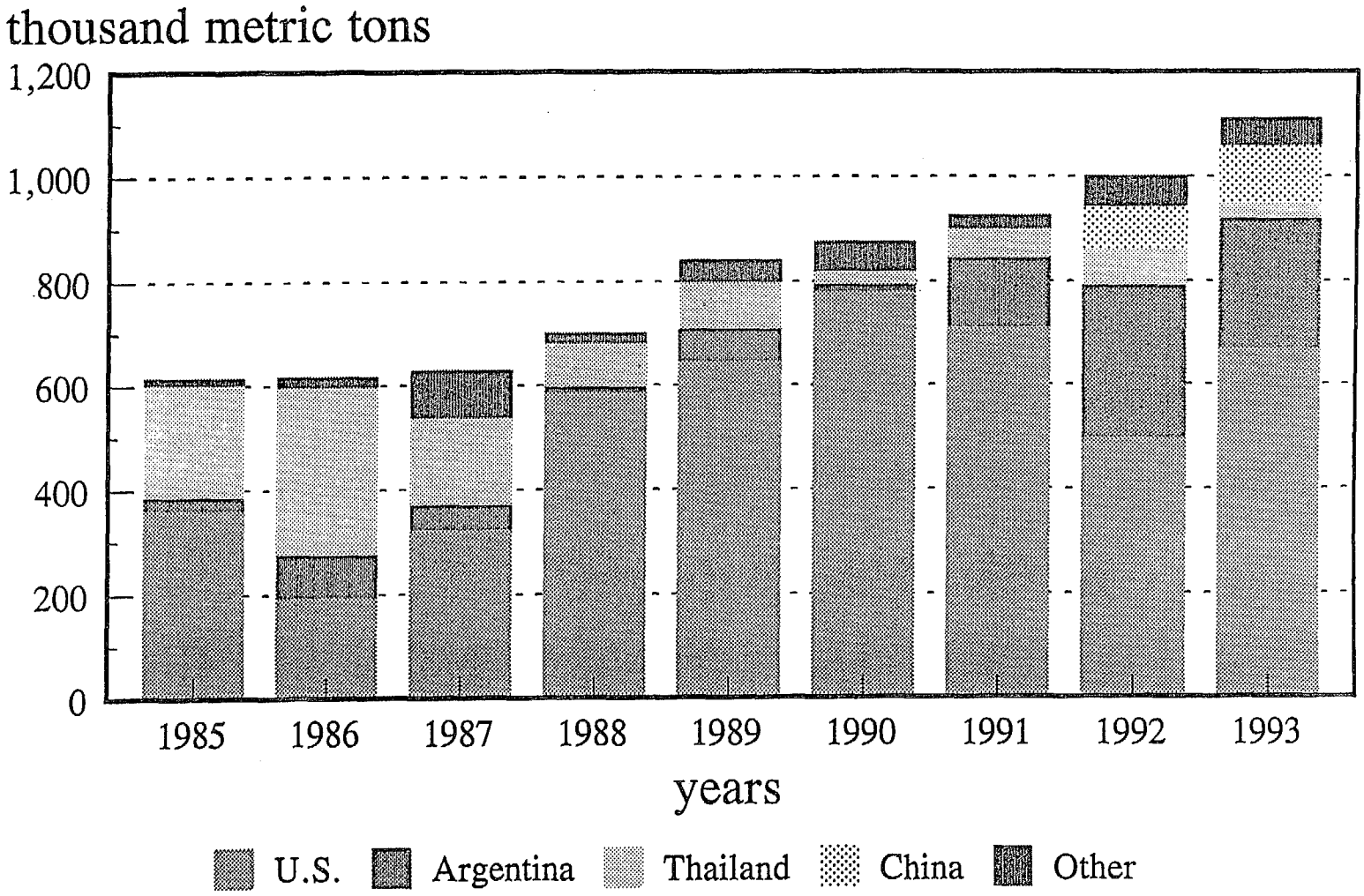


Figure 4

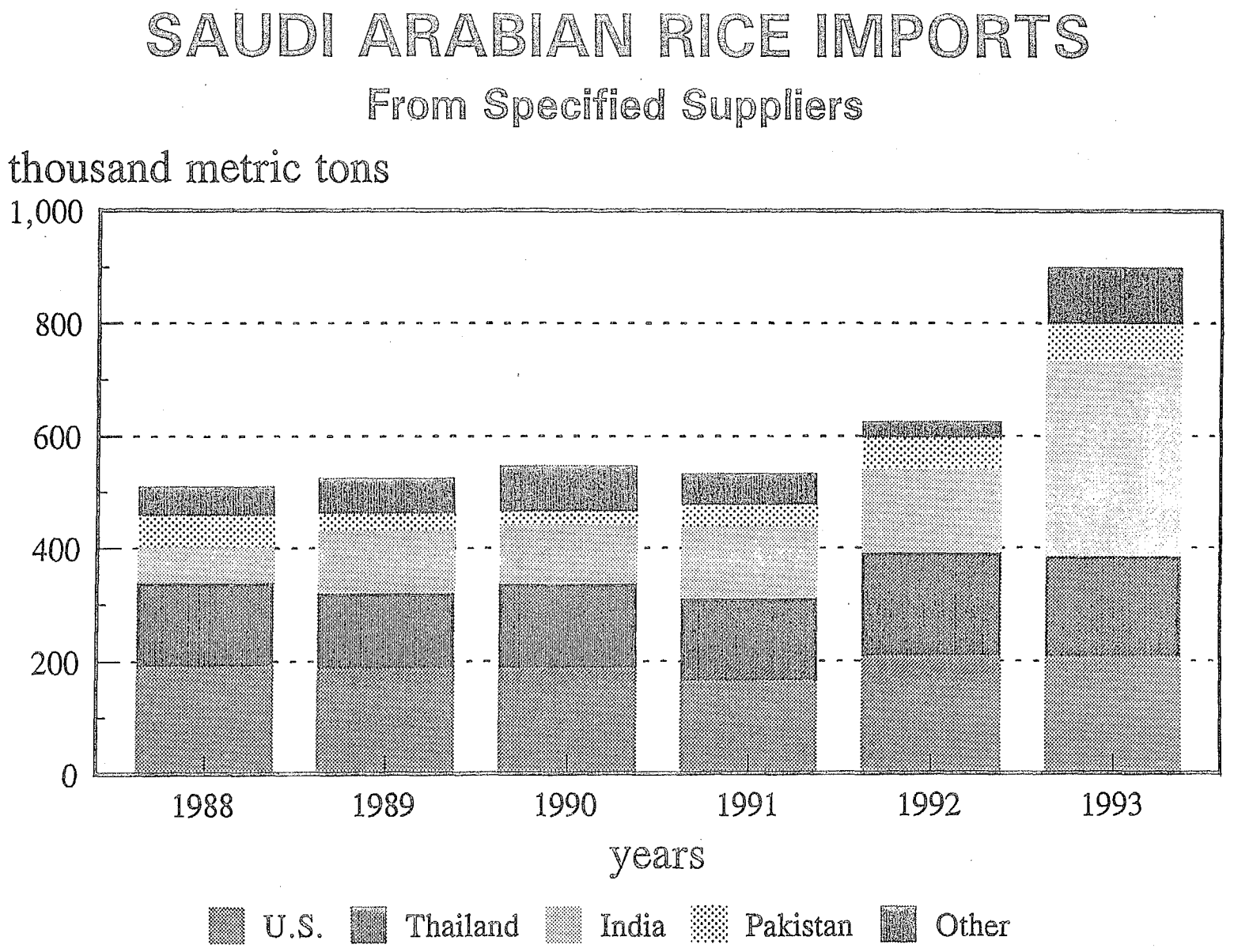

SAUDI ARABIAN RICE IMPORTS
From Specified Suppliers
thousand metric tons
1,000
800
600
400
200
0
1988
1989
1990
1991
1992
1993
years
U.S.
Thailand
India
Pakistan
Other

Figure 5

IRAQI AGRICULTURAL IMPORTS

BY COMMODITY AND VALUE, ANNUAL 1988-92

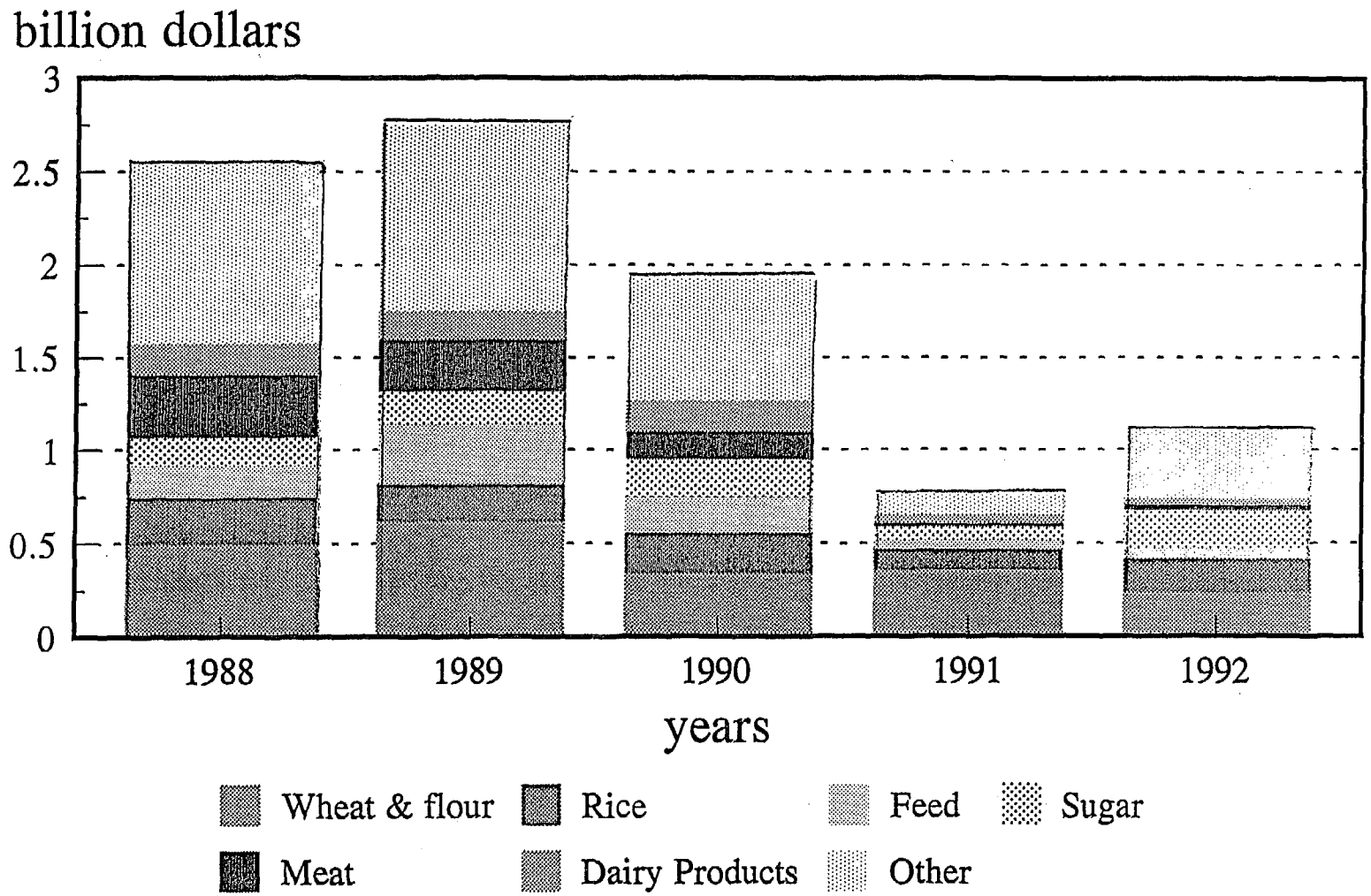


Figure 6

IRAQI POULTRY MEAT SUPPLY

OUTPUT AND IMPORTS, 1980-92

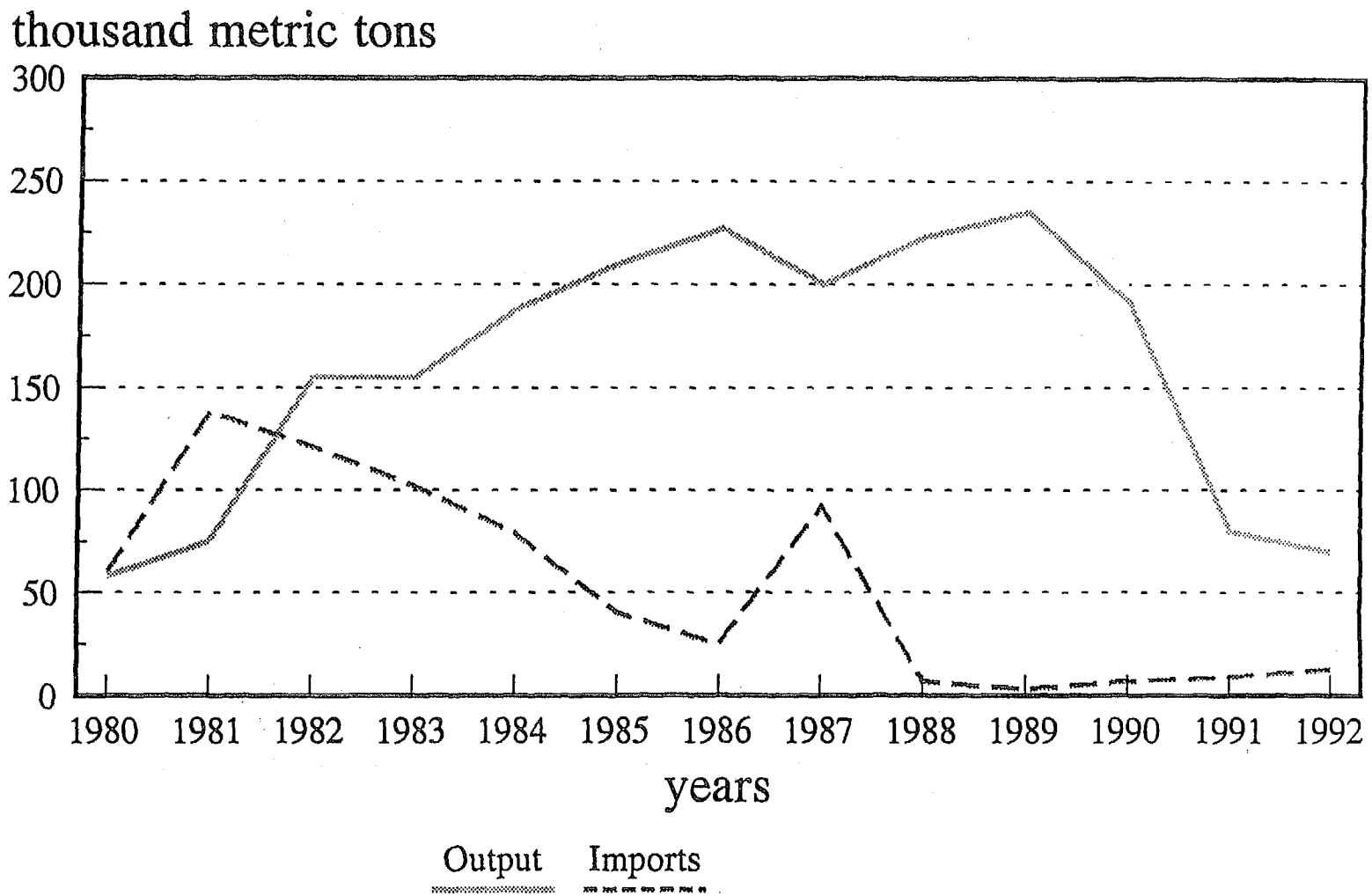


Figure 7

IRAQI RICE IMPORTS

From Specified Suppliers

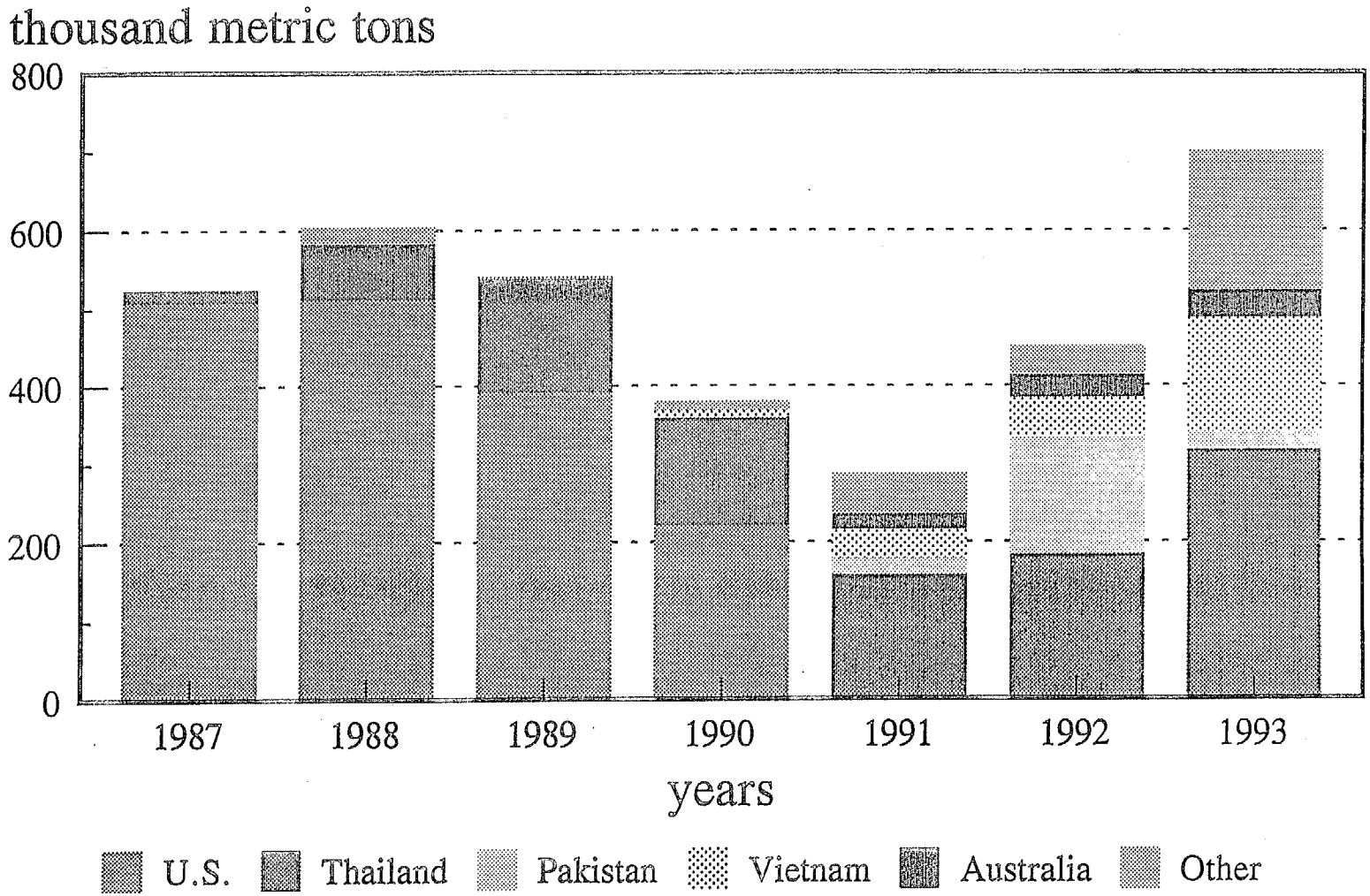


Figure 8

IRAQI IMPORTS OF WHEAT AND FLOUR

FROM SPECIFIED SUPPLIERS, 1988-92

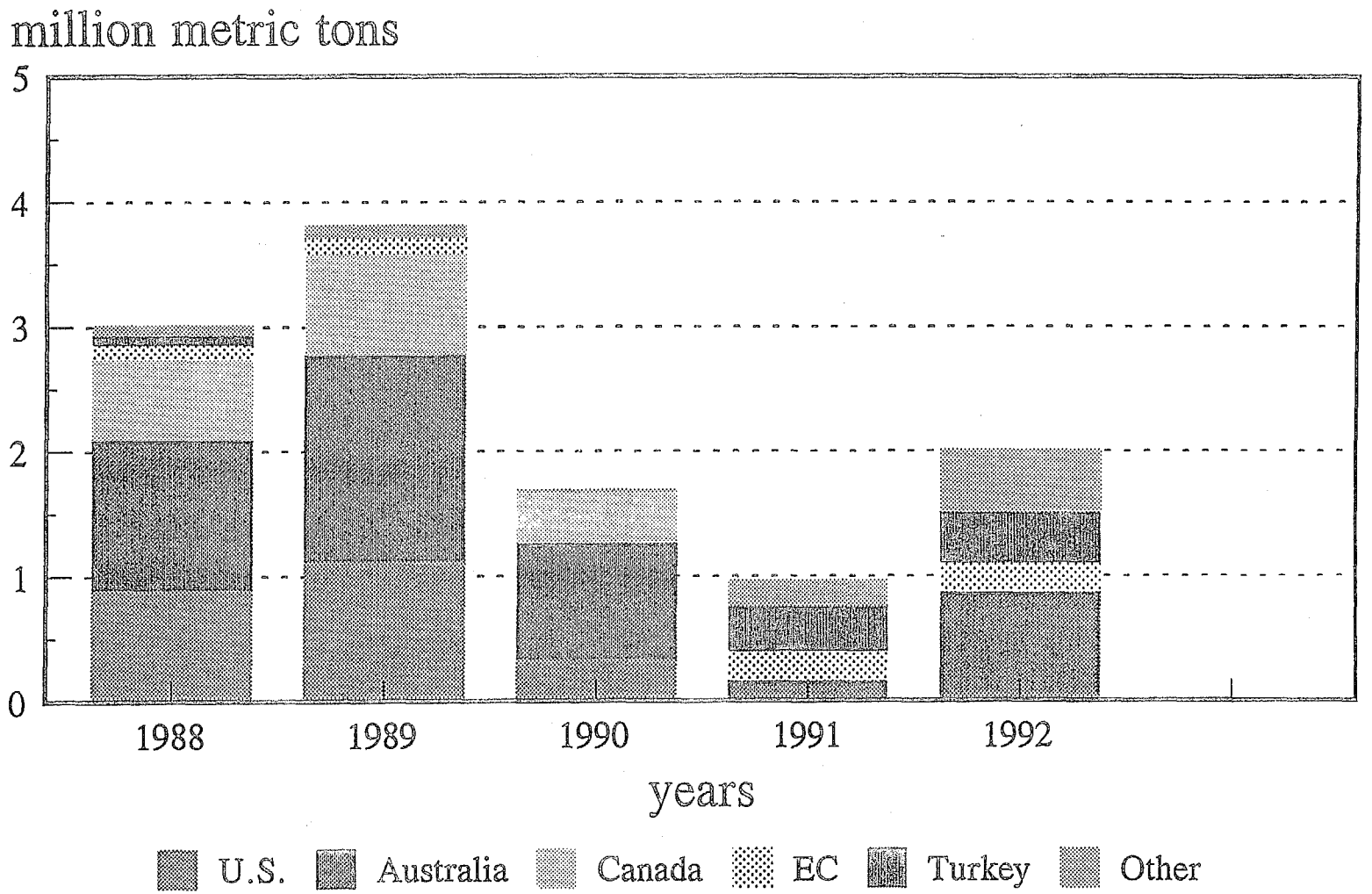


Figure 9

IRAQI GRAIN IMPORTS

BY TYPE, ANNUAL 1987-92

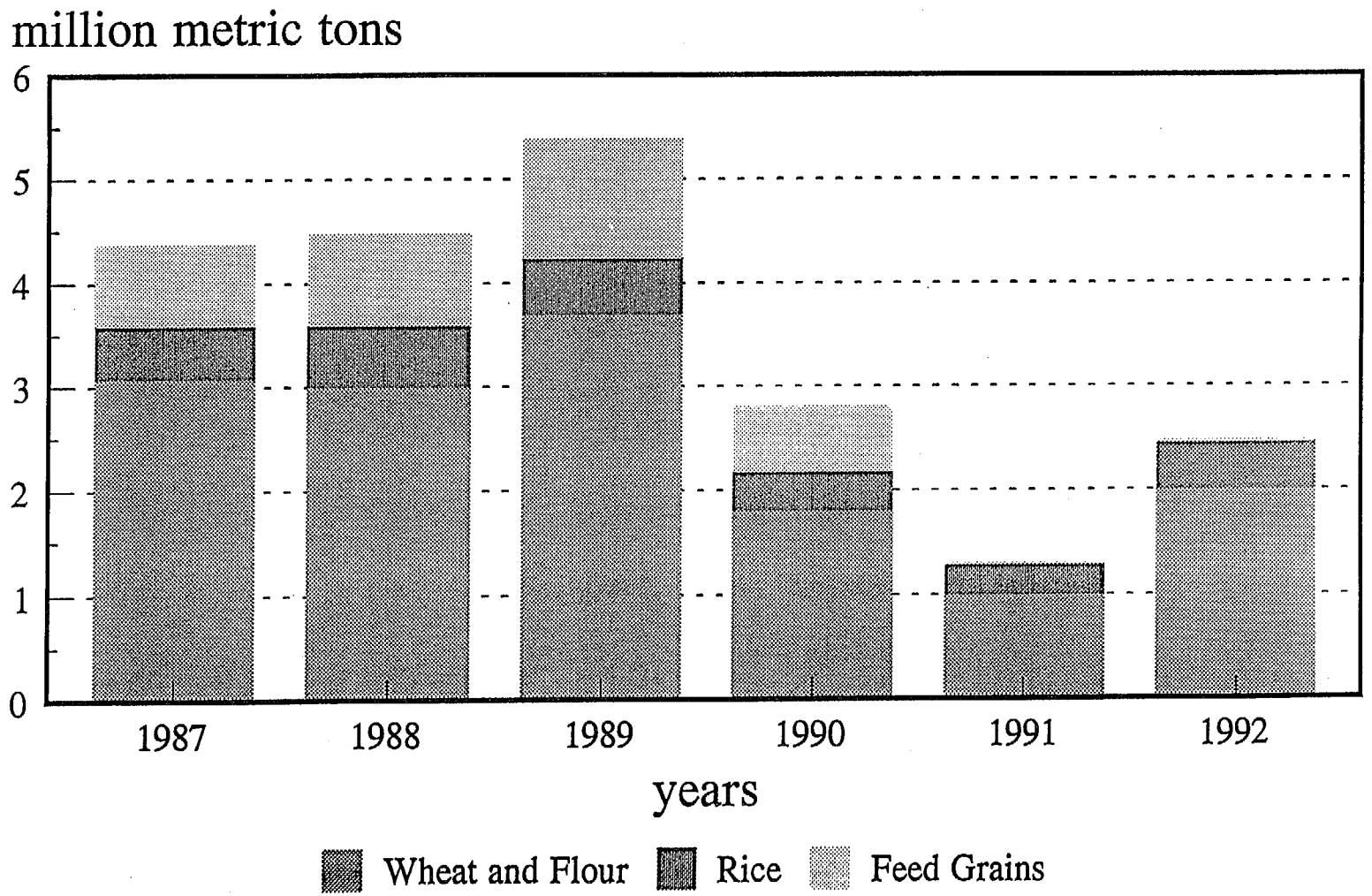


TABLE 1
SAUDI ARABIA: AGRICULTURAL IMPORTS BY COMMODITY, QUANTITY AND VALUE, ANNUAL 1987-92. 1989-92

COMMODITY	1989	1990	1991	1992	1989	1990	1991	1992
	metric tons				thousand dollars			
WHEAT	129737	127915	121906	173000	46917	45578	43644	90000
FLOUR OF WHEAT	7256	10276	22052	21000	1590	3170	7198	7400
MACARONI	22340	23777	21280	22300	19541	21973	21151	21151
PASTRY	41285	28854	29620	31200	73061	54977	60305	60305
MILLED PADDY RICE	524990	546770	532610	625000	146588	152166	147065	320000
BREAKFAST CEREALS	4976	4604	6427	6400	9818	10985	16595	16595
BARLEY	4611000	5622100	5243000	6356000	572000	576000	519000	680000
BEER OF BARLEY	28348	31431	42468	50000	16209	19846	28014	30000
MAIZE	839700	875000	925000	990000	89000	96000	117000	117000
OIL OF MAIZE	46000	60000	43495	56000	30000	44000	34101	38000
SORGHUM	9840	527	35000	20000	1951	210	7000	4500
POTATOES	135225	127074	129975	130000	24375	28717	33503	33500
FROZEN POTATOES	7500	9800	9500	8700	7700	9300	8400	8900
SUGAR REFINED	98560	101827	69935	100000	39349	46956	26587	37000
MOLASSES	2702	4615	8178	8178	1304	1173	2513	2513
SUGAR CONFECTIONERY	14435	12173	14936	12000	26063	23063	31018	28000
BEANS, DRY	2404	3138	2697	3000	1148	1789	1732	1600
BROAD BEANS, DRY	13395	16478	14700	14000	4141	4433	4700	6000
CHICK-PEAS	12503	10709	10720	12000	3778	4188	3815	5400
LENTILS	16249	16675	14037	16000	5068	7194	5677	8000
ALMONDS	1168	1057	1086	1086	2117	2073	2033	2033
PISTACHIOS	2614	3329	3382	3382	3811	4813	4642	4642
ALMONDS SHELLED	1541	1549	1494	1770	3757	3843	3004	4800
OIL OF SOYA BEANS	4011	4267	5134	5500	2549	2929	3947	4000
CAKE OF SOYA BEANS	248598	248500	262600	311700	79825	69497	60996	74000
GROUNDNUTS IN SHELL	1132	1015	1550	2000	506	677	1835	2000
PALM OIL	11925	53879	81598	159500	5522	17242	42908	90000
OLIVE OIL	5932	6641	6681	7000	10392	13483	13736	17000
OLIVES, PRESERVED	12310	10997	15811	14500	13101	10741	18859	17700
SESAME SEED	8203	16194	13105	15000	4626	16831	12021	13200
MELONSEED	7237	9294	8788	8788	1849	2303	3091	3091
CABBAGES	12849	11405	4673	6600	2330	2297	1586	2100
TOMATOES	143093	156165	136545	170000	23572	28899	32918	36000
TOMATO PASTE	25578	18902	10338	18000	21753	14727	8526	14500
ONIONS, DRY	154973	154665	146178	161000	26759	26939	27864	22000
GARLIC	8130	12681	10201	11900	3049	5696	5858	7000
VEGETABLES PREPARED NES	56420	59461	38579	43000	32286	35286	24994	27800
VEGETABLES FROZEN	7524	8364	6297	9100	6142	7810	6758	11100
BANANAS	114568	129332	124940	125800	31430	31412	47764	47000
ORANGES	197441	210803	192352	234000	58286	56732	54211	71000
TANG.MAND.CLEMENT.SATSMA	39884	42533	39652	39000	9286	10386	10694	10000
LEMONS AND LIMES	49742	54167	37356	46000	13615	14700	11664	18500

Saudi Arabian Agricultural imports continued

Table 1 continued

Saudi Arabian Agricultural imports continued

COMMODITY	1989	1990	1991	1992	1989	1990	1991	1992
	metric tons				thousand dollars			
(0515) APPLES	138989	134602	134588	110000	52026	44657	51787	48000
(0518) APPLEJUICE SINGLE STRENG	4917	3511	7728	7728	3384	2821	6406	6600
(0521) PEARS	13440	14382	13836	15000	5889	6001	5885	6000
(0534) PEACHES AND NECTARINES	25961	22925	9480	20000	8476	6345	3647	7600
(0560) GRAPES	31993	27812	27066	29900	16846	13038	14464	16000
(0561) RAISINS	5706	5170	4496	4400	5144	4995	3781	4100
(0567) WATERMELONS	15867	11086	9753	12000	4077	2980	3288	3500
(0568) CANTALOUPES OTH MELONS	10241	8790	13900	13900	2117	2292	6882	6882
(0571) MANGOES	3800	5557	7883	7883	2792	5450	8673	8673
(0575) PINEAPPLES, CANNED	9824	8204	6053	9600	8118	7951	4722	6100
(0577) DATES	8198	4346	300	100	4531	2588	180	60
(0583) MANGO JUICE	12356	10549	4956	4956	10219	8536	4661	4661
(0631) WATERS,ICE ETC	19874	15374	38084	38084	5964	4325	8585	8585
(0633) BEVERAGES NON-ALCOHOLIC	12839	8173	7916	7916	7402	5544	6343	6343
(0656) COFFEE, GREEN	6049	11761	16065	14000	18956	25905	37005	28000
(0666) CHOCOLATE PRODUCTS NES	16427	13821	21916	20000	45003	42533	69342	70000
(0667) TEA	9051	17890	19031	20000	30755	75865	74545	75000
(0702) NUTMEG, MACE, CARDAMONS	2755	6112	6678	7000	11753	32541	29991	32000
(0828) CIGARETTES	15949	18565	19362	19000	234345	261259	327434	343000
(0831) TOBACCO PRODUCTS NES	19190	15775	16369	16000	13594	12339	14488	14000
(0850) FEED SUPPLEMENTS	6775	12240	9547	16000	4262	7301	7955	12000
(0867) BEEF AND VEAL	15314	32534	47474	49500	38404	69638	88488	97000
(0874) SAUSAGES BEEF AND VEAL	1250	1381	5580	1500	2383	3224	14708	4500
(0882) COW MILK, WHOLE, FRESH	1669	1008	16108	14000	1421	1236	20392	23000
(0886) BUTTER OF COW MILK	16147	19097	17406	18000	33105	40439	36465	37000
(0894) WHOLE MILK, EVAPORATED	43522	44141	10662	8000	45580	46239	18395	14000
(0897) DRY WHOLE COW MILK	11958	13451	56218	61600	23570	27884	142546	165000
(0901) CHEESE (WHOLE COW MILK)	26733	30746	44483	43000	75015	91286	119523	115000
(0976) SHEEP	4930700	4775442	4618516	5100000	364711	341447	386435	410000
(0977) MUTTON AND LAMB	15640	22612	31159	36000	31072	45208	57088	70000
(1058) CHICKEN MEAT	194332	210076	244157	255000	229289	262482	333024	335000
(1182) HONEY	1450	838	2282	2700	3929	3301	6996	8000
OTHER	NA	NA	NA	NA	975959	1067751	1010383	1057666
(1882) AGRICULT PRODUCTS,TOTAL	NA	NA	NA	NA	3677496	3985395	4312343	4901000

Sources: Foreign Trade of Saudi Arabia, FAO and ERS matrix.

TABLE 2
SAUDI ARABIA: AVERAGE PRICE FOR SPECIFIED IMPORTS

	1989	1990	1991	1992
	dollars per metric ton			
(0015) WHEAT	361.63	356.31	358.01	520.23
(0016) FLOUR OF WHEAT	219.13	308.49	326.41	352.38
(0018) MACARONI	874.71	924.13	993.94	948.48
(0022) PASTRY	1769.67	1905.35	2035.96	1932.85
(0031) MILLED PADDY RICE	279.22	278.30	276.12	512.00
(0041) BREAKFAST CEREALS	1973.07	2385.97	2582.08	2592.97
(0044) BARLEY	124.05	102.45	98.99	106.99
(0051) BEER OF BARLEY	571.79	631.41	659.65	600.00
(0056) MAIZE	105.99	109.71	126.49	118.18
(0060) OIL OF MAIZE	652.17	733.33	784.02	678.57
(0083) SORGHUM	198.27	398.48	200.00	225.00
(0116) POTATOES	180.26	225.99	257.76	257.69
(0118) FROZEN POTATOES	1026.67	948.98	884.21	1022.99
(0164) SUGAR REFINED	399.24	461.14	380.17	370.00
(0165) MOLASSES	482.61	254.17	307.29	307.29
(0168) SUGAR CONFECTIONERY	1805.54	1894.60	2076.73	2333.33
(0176) BEANS, DRY	477.54	570.11	642.20	533.33
(0181) BROAD BEANS, DRY	309.15	269.03	319.73	428.57
(0191) CHICK-PEAS	302.17	391.07	355.88	450.00
(0201) LENTILS	311.90	431.42	404.43	500.00
(0221) ALMONDS	1812.50	1961.21	1872.01	1872.01
(0223) PISTACHIOS	1457.92	1445.78	1372.56	1372.56
(0231) ALMONDS SHELLED	2438.03	2480.96	2010.71	2711.86
(0237) OIL OF SOYA BEANS	635.50	686.43	768.80	727.27
(0238) CAKE OF SOYA BEANS	321.10	279.67	232.28	237.41
(0242) GROUNDNUTS IN SHELL	447.00	667.00	1183.87	1000.00
(0257) PALM OIL	463.06	320.01	525.85	564.26
(0261) OLIVE OIL	1751.85	2030.27	2055.98	2428.57
(0262) OLIVES, PRESERVED	1064.26	976.72	1192.78	1220.69
(0289) SESAME SEED	563.94	1039.34	917.28	880.00
(0299) MELONSEED	255.49	247.79	351.73	351.73
(0358) CABBAGES	181.34	201.40	339.40	318.18
(0388) TOMATOES	164.73	185.05	241.08	211.76
(0391) TOMATO PASTE	850.46	779.12	824.72	805.56
(0403) ONIONS, DRY	172.67	174.18	190.62	136.65
(0406) GARLIC	375.03	449.18	574.26	588.24
(0472) VEGETABLES PREPARED NES	572.24	593.43	647.87	646.51
(0473) VEGETABLES FROZEN	816.32	933.76	1073.21	1219.78
(0486) BANANAS	274.33	242.88	382.30	373.61
(0490) ORANGES	295.21	269.12	281.83	303.42
(0495) TANG.MAND.CLEMENT.SATSM	232.83	244.19	269.70	256.41
(0497) LEMONS AND LIMES	273.71	271.38	312.24	402.17

SAUDI ARABIA: AVERAGE PRICE

	1989	1990	1991	1992
	dollars per metric ton			
(0515) APPLES	374.32	331.77	384.78	436.36
(0518) APPLEJUICE SINGLE STREN	688.22	803.47	828.93	854.04
(0521) PEARS	438.17	417.26	425.34	400.00
(0534) PEACHES AND NECTARINES	326.49	276.77	384.70	380.00
(0560) GRAPES	526.55	468.79	534.40	535.12
(0561) RAISINS	901.51	966.15	840.97	931.82
(0567) WATERMELONS	256.95	268.81	337.13	291.67
(0568) CANTALOUPES OTH MELONS	206.72	260.75	495.11	495.11
(0571) MANGOES	734.74	980.75	1100.22	1100.22
(0575) PINEAPPLES, CANNED	826.34	969.16	780.11	635.42
(0577) DATES	552.70	595.49	600.00	600.00
(0583) MANGO JUICE	827.05	809.18	940.48	940.48
(0631) WATERS,ICE ETC	300.09	281.32	225.42	225.42
(0633) BEVERAGES NON-ALCOHOLIC	576.52	678.33	801.29	801.29
(0656) COFFEE, GREEN	3133.74	2202.62	2303.45	2000.00
(0666) CHOCOLATE PRODUCTS NES	2739.58	3077.42	3163.99	3500.00
(0667) TEA	3397.97	4240.64	3917.03	3750.00
(0702) NUTMEG, MACE, CARDAMONS	4266.06	5324.12	4491.02	4571.43
(0828) CIGARETTES	14693.40	14072.66	16911.17	18052.63
(0831) TOBACCO PRODUCTS NES	708.39	782.19	885.09	875.00
(0850) FEED SUPPLEMENTS	629.08	596.49	833.25	750.00
(0867) BEEF AND VEAL	2507.77	2140.47	1863.93	1959.60
(0874) SAUSAGES BEEF AND VEAL	1906.40	2334.54	2635.84	3000.00
(0882) COW MILK, WHOLE, FRESH	851.41	1226.19	1265.95	1642.86
(0886) BUTTER OF COW MILK	2050.23	2117.56	2094.97	2055.56
(0894) WHOLE MILK, EVAPORATED	1047.29	1047.53	1725.29	1750.00
(0897) DRY WHOLE COW MILK	1971.07	2073.01	2535.59	2678.57
(0901) CHEESE (WHOLE COW MILK)	2806.08	2969.04	2686.94	2674.42
(0976) SHEEP	73.97	71.50	83.67	80.39
(0977) MUTTON AND LAMB	1986.70	1999.29	1832.15	1944.44
(1058) CHICKEN MEAT	1179.88	1249.46	1363.97	1313.73
(1182) HONEY	2709.66	3939.14	3065.73	2962.96

TABLE 3. SAUDI ARABIA: GRAIN IMPORTS BY TYPE, ACCORDING TO SUPPLIELIERS

	WHEAT AND FLOUR (1,000 metric tons)						RICE (1,000 metric tons)							CORN				
Year	U.S.	EC	Australia	Argentina	Other	TOTAL	U.S.	Thailand	Pakistan	India	EC	Other	Total	U.S.	Argentina	Thailand	Other	TOTAL
1980	338	406	113	0	134	992	224	43	94	18	1	6	386	21	0	133	9	163
1981	518	71	126	46	11	773	236	143	111	10	1	7	507	27	0	201	17	245
1982	480	83	208	32	2	805	313	99	82	45	0	7	547	31	0	318	12	361
1983	310	114	58	60	57	599	281	71	112	38	9	60	572	22	0	448	33	503
1984	290	64	111	11	4	479	268	84	56	63	7	22	500	144	5	377	74	600
1985	65	53	34	3	0	155	195	145	60	67	10	15	492	362	26	215	12	615
1986	108	23	5	0	93	229	193	184	55	56	4	12	504	195	82	322	18	617
1987	137	10	7	0	0	154	201	150	50	75	10	24	510	323	49	167	90	629
1988	140	12	5	0	3	160	193	148	55	63	11	40	510	587	12	83	18	700
1989	133	3	2	0	12	150	189	133	27	113	12	51	525	646	63	90	41	840
1990	156	4	2	0	13	175	190	148	23	105	2	79	547	781	15	24	55	875
1991	161	3	2	0	6	172	167	145	39	127	3	52	533	713	132	55	25	925
1992	160	23	2	0	9	194	209	184	55	150	2	25	625	497	296	70	127	990
1993	36	35	3	0	26	100	208	179	65	348	10	90	900	670	270	30	135	1,105

	BARLEY (1,000 metric tons)						SORGHUM (1,000 metric tons)						OTHER GRAINS	GRAIN IMPORTS MATRIX
Year	U.S.	EC	Canada	Australia	Other	TOTAL	Thailand	Sudan	Argentina	U.S.	Other	TOTAL		
1980	16	773	4	145	291	1,229	137	301	0	0	70	508	2	3,259
1981	59	1,722	0	335	159	2,275	107	242	0	0	4	354	3	4,145
1982	444	1,497	0	695	1,224	3,860	223	349	0	0	6	578	5	6,150
1983	206	865	70	270	474	1,885	137	252	0	0	17	406	9	3,975
1984	318	2,475	110	953	2,020	5,876	207	3	4	0	0	214	7	7,670
1985	42	2,145	361	2,888	264	5,700	140	1	0	0	7	148	8	7,049
1986	1,236	3,434	879	1,615	282	7,446	80	288	0	0	3	371	9	9,182
1987	2,345	2,641	2,103	197	290	7,576	13	286	0	0	2	301	11	9,170
1988	634	2,284	1,033	162	170	4,283	0	310	0	0	0	310	14	5,963
1989	715	2,134	1,558	117	140	4,664	0	275	0	9	1	285	15	6,464
1990	1207	3,040	1,155	653	20	6,075	0	10	0	0	5	15	17	7,687
1991	985	2,753	997	243	610	5,588	0	5	0	10	2	17	18	7,235
1992	669	3,998	289	379	698	6,033	0	4	41	7	5	57	20	7,899
1993	335	3,370	450	90	455	4,700	0	3	40	5	2	50	22	6,855

Sources: Foreign trade of Saudi Arabia and matrix of suppliers trade run

TABLE 4. IRAQ: GRAIN IMPORTS BY TYPE AS REPORTED BY SUPPLIERS, ANNUUAL 1980-93.

WHEAT AND FLOUR (thousand metric tons)

Year	U.S.	EC	Australia	Canada	Argentina	Turkey	Other	TOTAL
1980	312	23	787	429	0	120	32	1,703
1981	95	90	208	453	134	97	.760	1,837
1982	178	128	860	259	280	200	252	2,156
1983	1,137	339	459	409	0	280	73	2,697
1984	1,064	206	1,290	492	0	311	70	3,432
1985	532	212	964	231	0	184	257	2,380
1986	816	70	980	552	0	0	240	2,658
1987	1,038	0	1,183	745	0	89	25	3,080
1988	895	122	1,200	640	0	78	68	3,003
1989	1,120	120	1,663	809	0	0	4	3,716
1990	335	0	930	423	0	0	112	1,800
1991	0	233	165	0	0	373	210	981
1992	0	221	879	0	0	412	512	2,024
1993	0	332	135	7	0	130	340	944

RICE (thousand metric tons)

U.S.	Thailand	Pakistan	EC	Australia	VN	Other	Total
269	53	26	0	0	0	0	348
74	54	76	0	0	0	146	350
235	129	0	0	0	0	56	420
282	134	0	1	0	0	23	440
448	38	0	1	0	0	9	495
							0
407	113	0	0	0	0	5	525
373	125	0	0	52	0	0	550
509	15	0	0	0	0	0	524
513	60	0	0	11	0	19	603
392	120	0	0	28	0	2	542
222	138	0	0	0	10	5	375
0	160	20	0	20	37	50	287
0	185	150	0	30	50	35	450
0	318	25	0	36	145	176	700

BARLEY (thousand metric tons)

Year	U.S.	EC	Canada	Turkey	Other	TOTAL
1980	148	63	16	6	233	466
1981	0	78	0	37	115	230
1982	0	0	30	135	0	165
1983	125	0	129	36	0	290
1984	276	0	254	17	10	557
1985	100	0	132	29	0	261
1986	0	0	10	0	10	20
1987	67	35	111	6	1	220
1988	153	25	20	24	0	222
1989	187	60	283	0	40	570
1990	104	40	56	0	19	219
1991	0	0	0	5	40	45
1992	0	0	0	10	20	30
1993	0	0	0	0	0	0

CORN (thousand metric tons)

U.S.	Thailand	TOTAL
0	83	83
0	6	6
0	70	70
57	25	82
339	0	339
240	30	270
323	7	330
542	4	546
647	3	650
565	10	575
413	2	415
0	5	5
0	4	4
0	2	2

OTHER GRAINS (thousand metric tons)	GRAIN IMPORTS	From U.S. Quantity	Share U.S. %
1	2,601	730	30.8
1	2,424	169	7.3
1	2,812	412	14.7
1	3,510	1600	45.6
1	4,825	2127	44.1
	0		
1	3,437	1279	37.2
2	3,560	1512	42.5
2	4,372	2156	49.3
3	4,481	2208	49.3
2	5,405	2264	41.9
3	2,803	1074	38.3
4	1,322	0	0.0
5	2,513	0	0.0
5	1,651	0	0.0

Sources: UN trade runs, U.S. Dept. of Commerce, FAO matrix
Sources: Grain and Feed Report of ATO Baghdad and matrix estimates

TABLE 5. IRAQ: AGRICULTURAL IMPORTS OF SPECIFIC AGRICULTURAL COMMODITIES, ACCORDING TO SUPPLIERS, 1989-92.

COMMODITY	1989	1990	1991	1992	1989	1990	1991	1992
	metric tons				thousand dollars			
(0015) WHEAT	3500000	1800000	500000	1100000	627000	342000	62000	165000
(0016) FLOUR OF WHEAT	101000	50000	237000	340000	25000	13000	49000	71000
RICE, MILLED	542000	375000	287000	450000	196000	220000	112000	181000
(0044) BARLEY	385362	210000	45000	30000	58254	35000	6500	3800
(1892) FEEDINGSTUFFS	257000	72000	58000	70000	117529	71000	44000	45200
(0049) MALT OF BARLEY	2000	9000	2500	6000	900	3600	850	2100
(0051) BEER OF BARLEY	14000	15000	5700	5000	14700	17000	4000	3500
(0056) MAIZE	570000	415000	800	1100	68500	52000	125	170
(0109) INFANT FOOD	6100	15000	3715	7800	18000	50000	10488	10488
(0116) POTATOES	31800	2500	140000	229000	10000	1600	28500	33000
(0164) SUGAR REFINED	656275	432400	166500	529000	283644	200000	73000	196000
(0191) CHICK-PEAS	28068	27000	18000	12000	14000	13000	6500	5000
(0201) LENTILS	20000	30000	19000	10000	16000	20000	11000	6000
(0211) PULSES NES	10000	10000	17000	28000	5341	6000	7900	9000
(0221) ALMONDS	1244	1244	400	900	4536	4536	1100	2300
(0238) CAKE OF SOYA BEANS	319500	101000	50	3300	77000	28000	20	1200
(0257) PALM OIL	257100	210000	21500	43100	133000	130000	13000	27000
(0258) OIL OF PALM KERNELS	10500	8100	300	450	7300	4500	220	350
(0268) OIL OF SUNFLOWER SEED	25400	22000	5700	3000	16300	14000	4500	2300
(0388) TOMATOES	25000	50000	57500	1400	7500	15300	15000	300
(0391) TOMATO PASTE	30000	25000	12257	14500	32000	24000	7624	7624
(0403) ONIONS, DRY	3500	400	65000	60000	800	200	12500	9400
(0472) VEGETABLES PREPARED NES	9000	6500	2000	2300	7800	5500	1400	1700
(0515) APPLES	10000	4500	88000	3300	5000	1800	23000	1200
(0561) RAISINS	2100	700	50	200	2900	1200	50	240
(0564) WINE	170	220	450	890	420	600	780	2300
(0653) FOOD WASTES	123000	60000	60000	67000	92000	44000	44000	44000
(0659) COFFEE EXTRACTS	300	200	200	230	2600	1800	1800	2100
(0666) CHOCOLATE PRODUCTS NES	450	1400	1340	1450	2000	4000	3000	3400
(0667) TEA	39973	32000	2600	1600	65185	70000	6500	1900
(0767) COTTON LINT	27259	50	50	11000	35605	58	64	25000
(0826) TOBACCO LEAVES	24108	13000	6000	5500	97552	55000	27000	24800
(0828) CIGARETTES	4000	5000	2500	3200	55000	75000	30000	35600
(0867) BEEF AND VEAL	78384	45000	18000	9800	232827	120000	21000	11600
(0886) BUTTER OF COW MILK	8942	3700	5600	12700	10492	5300	6000	13300
(1935) MILK COND EVAP	48700	39000	11000	3000	106918	115000	27000	6000
(0901) CHEESE (WHOLE COW MILK)	2313	10000	600	560	8353	32000	2100	1000
(0977) MUTTON AND LAMB	15581	10000	100	600	27477	14500	200	600
(1058) CHICKEN MEAT	5000	6000	500	2700	7000	8000	1000	1900
(1062) HEN EGGS	8080	6000	5000	6000	34797	21000	10000	11000
(1242) MARGARINE SHORTENING	28000	20000	57000	95000	26000	18000	43000	70000
OTHER	NA	NA	NA	NA	ERR	ERR	ERR	ERR
TOTAL AGRICULTURAL	NA	NA	NA	NA	2771000	1907900	772000	1096000

NA Not Applicable

Sources: FAO Agrostat, Trade Data of Suppliers, and ERS.

TABLE 6. IRAQ: ESTIMATED AVERAGE IMPORT PRICES, 1989-92

COMMODITY	1989	1990	1991	1992
	U.S. dollars per metric ton			
(0015) WHEAT	179.14	190.00	124.00	150.00
(0016) FLOUR OF WHEAT	247.52	260.00	206.75	208.82
RICE, MILLED	361.62	586.67	390.24	402.22
(0044) BARLEY	151.17	166.67	144.44	126.67
(1892) FEEDINGSTUFFS	457.31	986.11	758.62	645.71
(0049) MALT OF BARLEY	450.00	400.00	340.00	350.00
(0051) BEER OF BARLEY	1050.00	1133.33	701.75	700.00
(0056) MAIZE	120.18	125.30	156.25	154.55
(0109) INFANT FOOD	2950.82	3333.33	2823.15	1344.62
(0116) POTATOES	314.47	640.00	203.57	144.10
(0164) SUGAR REFINED	432.20	462.53	438.44	370.51
(0191) CHICK-PEAS	498.79	481.48	361.11	416.67
(0201) LENTILS	800.00	666.67	578.95	600.00
(0211) PULSES NES	534.10	600.00	464.71	321.43
(0221) ALMONDS	3646.30	3646.30	2750.00	2555.56
(0238) CAKE OF SOYA BEANS	241.00	277.23	400.00	363.64
(0257) PALM OIL	517.31	619.05	604.65	626.45
(0258) OIL OF PALM KERNELS	695.24	555.56	733.33	777.78
(0268) OIL OF SUNFLOWER SEED	641.73	636.36	789.47	766.67
(0388) TOMATOES	300.00	306.00	260.87	214.29
(0391) TOMATO PASTE	1066.67	960.00	622.01	525.79
(0403) ONIONS, DRY	228.57	500.00	192.31	156.67
(0472) VEGETABLES PREPARED NES	866.67	846.15	700.00	739.13
(0515) APPLES	500.00	400.00	261.36	363.64
(0561) RAISINS	1380.95	1714.29	1000.00	1200.00
(0564) WINE	2470.59	2727.27	1733.33	2584.27
(0653) FOOD WASTES	747.97	733.33	733.33	656.72
(0659) COFFEE EXTRACTS	8666.67	9000.00	9000.00	9130.43
(0666) CHOCOLATE PRODUCTS NES	4444.44	2857.14	2238.81	2344.83
(0667) TEA	1630.73	2187.50	2500.00	1187.50
(0767) COTTON LINT	1306.17	1160.00	1280.00	2272.73
(0826) TOBACCO LEAVES	4046.46	4230.77	4500.00	4509.09
(0828) CIGARETTES	13750.00	15000.00	12000.00	11125.00
(0867) BEEF AND VEAL	2970.34	2666.67	1166.67	1183.67
(0886) BUTTER OF COW MILK	1173.34	1432.43	1071.43	1047.24
(1936) MILK DRY	2195.44	2948.72	2454.55	2000.00
(0901) CHEESE (WHOLE COW MILK)	3611.33	3200.00	3500.00	1785.71
(0977) MUTTON AND LAMB	1763.49	1450.00	2000.00	1000.00
(1058) CHICKEN MEAT	1400.00	1333.33	2000.00	703.70
(1062) HEN EGGS	4306.56	3500.00	2000.00	1833.33
(1242) MARGARINE SHORTENING	928.57	900.00	754.39	736.84

Source: ERS calculations from trade matrix table.

References

"Country Data Sheets For Iraq "Food and Agricultural Organization, Rome, Italy, September 1993.

Foreign Trade of Saudi Arabia for 1991, Ministry of Finance and Economy, Central Statistics Department, Riyadh, November 1992.

"International Trade Matrix For Iraq", Economic Research Service, U.S. Dept. of Agriculture, Washington, DC. July 1993.

Exchange Rate Forecasts Using Eurocurrency Interest Rate Differentials

Alberto Jerardo, Economic Research Service

Abstract

The Eurocurrency market, in conjunction with the forward exchange market, represents the largest international pool of short-term funds tapped by arbitragers, hedgers, and speculators in financing and covering their currency positions. The forward covering of arbitrage positions in Eurocurrencies brings about covered interest rate parity. Empirical evidence of the correspondence of the forward premium and the interest rate differential between two currencies, especially when the U.S. dollar is involved, attests to the cogency of the interest rate parity theorem. The integration of the Eurocurrency market with the forward exchange market provides a framework in which interest rate parity can be exploited as a tool in generating exchange rate forecasts. In this paper, state space--a multivariate time series model, is employed in projecting interest rate differentials, which, by definition, estimate the forward premia. Taking the forward rate as the predictor of the future spot exchange rate, exchange rate forecasts are generated from the estimated forward premia. Notwithstanding satisfactory forecast results, the influence of interest differentials on exchange rate projections was not significant.

Keywords: Interest rate parity, exchange rate forecasts, Eurocurrency market, forward premium, state space, multivariate time series model.

Introduction

A country's export competitiveness depends to a significant extent on changes in currency exchange rates as they affect prices of traded goods and services. Prices, along with changes in income, in turn help determine the volume of exports and imports. While trade flows do influence exchange rates, their effect has increasingly been overshadowed by massive flows of funds between financial markets, which are driven principally by relative interest rates. Given current massive trade activity in short-term money assets, this paper attempts to forecast exchange rates from international interest rate differentials, specifically between Eurocurrency rates. The role of the forward exchange market in achieving covered interest rate parity (i.e., exchange risk is eliminated) is exploited in generating the forecasts.

Eurocurrency interest rate differentials largely determine the forward premium or discount that equates effective returns or costs--a broad characterization of interest parity. However, it is also the case that Eurocurrency rates adjust to changes in the forward premium (Herring and Marston, 1977). The joint determination of the forward exchange rate and the respective Eurocurrency rates is inherent in the way transactions are conducted--traders in one market base their quotations on rates established in the other market, and vise versa.[1] This interaction suggests the need for an integrated model, as opposed to independent models, of the Eurocurrency and forward exchange markets.

The linkage between forward exchange premia and international interest rate differentials brought about by covered interest arbitrage is the theoretical basis for the theory of international money market equilibrium. Interest rate parity--when the covered interest differential between convertible currencies is zero--provides the framework for simultaneous equilibrium in the forward and Eurocurrency markets. The linkage between these two markets was often not explicitly specified in past models of the forward market.[2] Interest parity provides a model for integrating these markets and for predicting exchange rates from interest rate differentials.

U.S. dollars, followed by Deutsche marks, and Japanese yen dominate the Eurocurrency and forward exchange markets. As such, the exchange rates for which forecasts are generated are the mark per dollar and yen per dollar. In general, a currency will achieve interest rate parity more quickly and with greater frequency the larger its external money and forward exchange markets.

[1] p. 80, R.J. Herring and R.C. Marston.

[2] pp. 56-57, 80, ibid.

Exchange Rate Forecasting Models

The three broad approaches to forecasting exchange rates use either structural models, arbitrage models, or pure prediction models. In structural models, spot or forward rates are derived from the relative size and direction of funds flowing between sources (supply) and users (demand). Alternatively, arbitrage models employ interest rate parity such that the (discounted) interest rate differential equals the anticipated percent change of the spot rate, as reflected in the forward premium if interest arbitrage is covered. That is, the forward premium depends on the size of the interest differential. Pure prediction models typically identify the underlying statistical structure of the spot rates after filtering out trend and seasonal patterns.

Whereas structural models show how underlying economic variables (including interest-rate differentials) jointly determine the market exchange rate, arbitrage models base forecasts on forward exchange rates, which is equivalent to adopting the interest parity relationship as the implicit reduced form of the true structural model (which is derived below). Arbitrage models, however, assume that interest rates are exogenously determined, at least independently of spot exchange rates, which is not consistent with Eurocurrency transactions where deposit or loan rates are jointly determined with forward rates.[3] A more realistic model for exchange rate determination can be constructed by making a Eurocurrency rate a function of the forward premium.

Time series models, which inherently cannot explain economic linkages, are nevertheless able to capture lag structures; and in the case of multivariate models, can correlate variables which are causally related. Estimating an arbitrage (interest parity) model by time series analysis will take care of the simultaneous bias from the linkage between Eurocurrency rates and the forward premium. Also, serial correlation in the residuals as well as multicollinearity problems are dealt with by the requirement of joint stationarity of the variables. This approach avoids potential econometric problems such as multicollinearity, simultaneity, and serial correlation in estimating a model that sets out to link forward rates and Eurocurrency rates--variables which are simultaneously determined and highly autocorrelated.

Interest Parity Theory of the Forward Exchange Rate

This theory postulates that equilibrium in the forward exchange market is reached when the covered return on a foreign asset is equal to the return on similar home assets--the forward premium is at interest parity with respect to home and foreign interest rates.[4] That is, if domestic and foreign assets, covered for exchange risk, are perfect substitutes, then interest arbitrage determines the forward premium. Risks such as capital controls, which are not generally covered in the forward market, are assumed to be zero. Exploiting interest rate differentials with the protection of forward cover (covered interest arbitrage), helps create, along with speculators, supply and demand for forward exchange that determines the forward premium.

The development and expansion of global financial markets under floating exchange rates have tied a country's monetary policy inextricably to its exchange rate policy. In the short run, the linkage between interest rates and exchange rates dominates the linkage between trade flows and exchange rates. When domestic interest rates rise, the home currency appreciates as capital flows in. However, the consequent loss in trade competitiveness usually takes some time to feed back into the exchange rate. In other words, changes in the real sector do not translate immediately into expectations about the currency's short-term value. As monetary policy instruments or targets, interest rates are more relevant for explaining short-term exchange rate movements than long-term trends such as trade flows or changes in income.

Floating exchange rates and highly mobile capital have also increased the interrelationship between international capital flows and interest rates and exchange rates. Decisions involved in holding financial assets, specifically those transacted on a covered basis, are based on both interest rate and forward exchange rate levels. Hence the direction of capital flows depends on relative interest rates as well as on the forward premium. Uncovered positions, on the other hand, may influence interest differentials when amounts are large enough, but will only affect the forward premium through any induced changes in the spot rate.

[3] op. cit.

[4] p. 60, ibid.

National money markets are subject to regulations that may inhibit interest arbitrage. In this case, forward exchange rates will diverge from interest parity with respect to national interest rates. In contrast, the Eurocurrency markets have been free of capital controls, and because of extensive interbank arbitrage, forward exchange rates quickly converge toward interest parity with respect to Eurocurrency rates. Tests have shown that interest parity holds with much greater accuracy and frequency between Eurocurrencies than between national money market deposits.[5] Figure 1 shows close adherence of the Deutsche mark's and the Japanese yen's forward premium (against the U.S. dollar) with the respective interest differential against the Eurodollar deposit rate.[6]

A model of the forward exchange market and the Eurocurrency market shall be constructed in the following sections with the objective that simultaneous market equilibrium results in the joint determination of equilibrium Eurocurrency interest rates and corresponding forward exchange rates. Covered interest arbitrage, in integrating the two markets, will equate the forward premium with the interest rate differential between two currencies. This relationship implies that once the interest differential is determined, the forward premium is determined as well, and conversely.

Interest Rate Parity in the Eurocurrency Market

Covered interest arbitrage is by nature transacted concurrently in the forward and Eurocurrency markets. An increase in demand for a particular Eurocurrency will affect the forward exchange rate, and an increase in demand for a particular forward currency will affect Eurocurrency rates. Equilibrium in both markets should be reached at the same time. But the interdependence of these markets has often been ignored in past studies, with conditions in one market generally assumed to be exogenously determined. To address this shortcoming, the structure of the exchange rate forecasting model will combine the two markets such that the forward premium and the Eurocurrency rates are simultaneously determined in the implicit reduced-form equation.

The Eurocurrency market can be described by a model in which all Eurocurrency assets (covered deposits and loans) are assumed to be perfect substitutes and perfectly mobile. It is further assumed that domestic interest rates are exogenously determined (by monetary authorities) so that arbitrage affects only Eurocurrency rates and the forward rate. Equilibrium in the Eurocurrency market must reflect the net (or excess) currency positions of all participants--Eurobanks as well as nonbank borrowers and investors.

The structure of the model is based on the equilibrium condition of equality between the net demand and net supply of Eurocurrency funds. The Euromark is used as the case currency, and Eurodollars as the foreign currency in formulating the model below, although a similar model for the yen can also be constructed. The demand for Euromark loans by nonbank borrowers is denoted by M^D, and M^S is the supply of Euromark deposits from nonbank investors. Eurobanks are not mere brokers, matching deposits and loans in each Eurocurrency. They typically take net loan or deposit positions in different currencies. That is, Euromark loans will equal the sum of Euromark deposits and the net Euromark position (NP) of the Eurobanks:

$$M^D = M^S + NP \tag{1}$$

These endogenous variables depend accordingly on the Euromark deposit rate (i_M), the Eurodollar rate ($i_\$$), and the forward premium (fp), in terms of marks per dollar:

$$\begin{aligned} M^D &= M^D(i_M^-, i_\$^+, fp^+) \\ M^S &= M^S(i_M^+, i_\$^-, fp^-) \\ NP &= NP(i_M^+, i_\$^-, fp^-) \end{aligned} \tag{2}$$

The demand for Euromark deposits is expected to increase in response to a rise in the forward premium since the cost of forward cover (in dollars) is lower. When banks cover their net Euromark positions in the forward

[5] p. 153, R.T. Baillie and P.C. McMahon (1989).

[6] The mean differential from 1980 to 1991 is not significantly different from zero at the 5 percent level (t test).

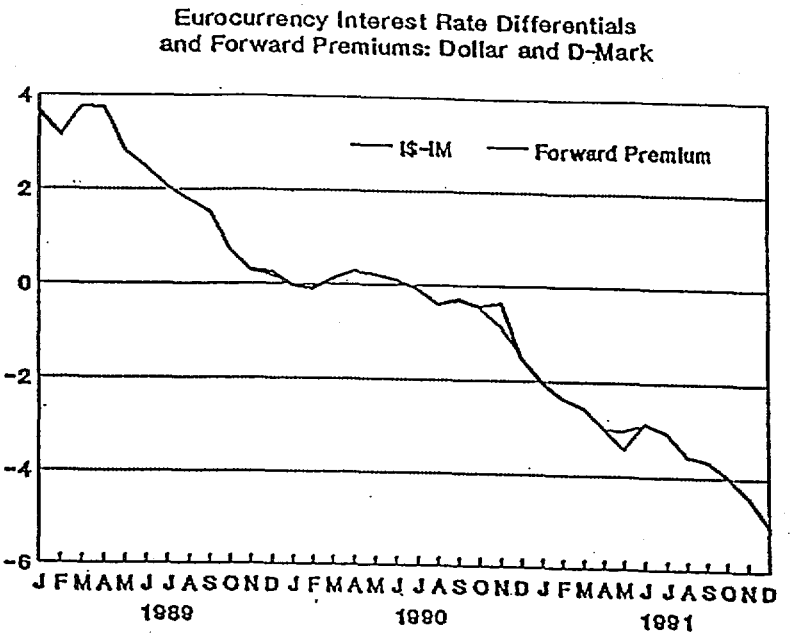


Figure 1a

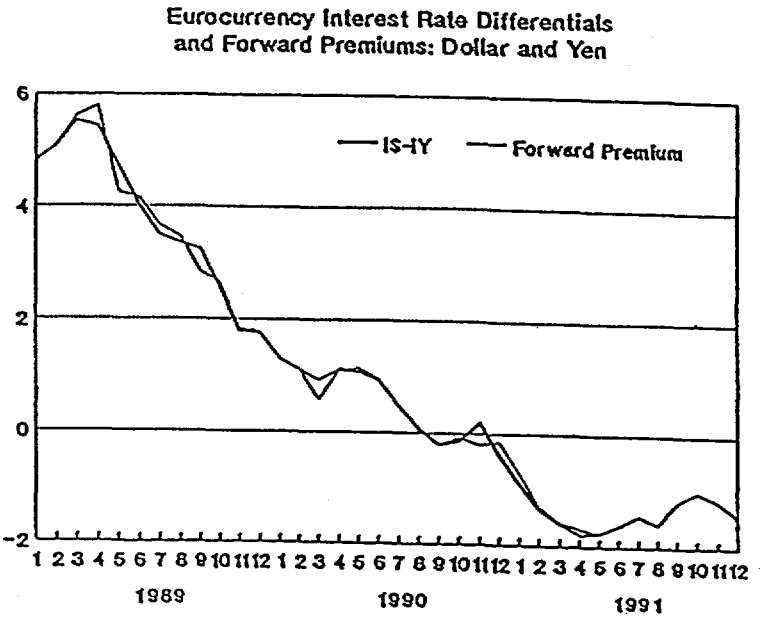


Figure 1b

market, they create a net demand (equal to NP) for forward dollars, the forward currency in this case. Competitive bidding leads to a point at which the net supply of forward exchange (FX) from speculators and hedgers must equal the Eurobanks' net demand for forward dollars. That is, at equilibrium,

$$FX = M^D - M^S \; (= NP) \tag{3}$$

Equation 3 can be expressed as an implicit function ($FX-M^D+M^S=0$), whose total differential equals zero:

$$(M_m^D - M_m^S)\, di_M + (M_\$^D - M_\$^S)\, di_\$ + (M_{fp}^D - M_{fp}^S)\, dfp = 0 \tag{4}$$

The subscripts indicate partial derivatives with respect to i_M, $i_\$$, and fp. Solving for the change in the forward premium, dfp, after dividing through by M_m^S:

$$dfp = \frac{\left(\frac{M_m^D}{M_m^S} - 1\right) di_M + \left(\frac{M_\$^D}{M_m^S} - \frac{M_\$^S}{M_m^S}\right) di_\$}{-\left(\frac{M_{fp}^D}{M_m^S} - \frac{M_{fp}^S}{M_m^S}\right)} \tag{5}$$

Under the condition of perfect capital mobility, where financial assets become perfect substitutes on a covered basis, these partial derivatives either approach infinity or negative infinity:

$$M_m^D = M_\$^S = M_{fp}^S = -\infty \quad and \quad M_\$^D = M_{fp}^D = M_m^S = \infty$$
$$Hence, \quad \frac{M_m^D}{M_m^S} = \frac{M_\$^S}{M_m^S} = \frac{M_{fp}^S}{M_m^S} = -1 \quad and \quad \frac{M_\$^D}{M_m^S} = \frac{M_{fp}^D}{M_m^S} = 1 \tag{6}$$

These conditions correspond to an infinitely elastic arbitrage schedule--a flat M^D curve in Figure 2 below, thus reducing the equilibrium equation to

$$dfp = di_M - di_\$ \tag{7}$$

which is the interest parity condition in differential form. This relationship represents all combinations of fp, i_M and $i_\$$ that will clear the forward market as well as the Eurocurrency market. Originally developed by Keynes, the interest-parity theory of forward exchange as described by Equation 7 holds that interest-rate gaps are always eventually matched by the forward premium or discount, making the covered interest differential zero.[7]

Market Equilibrium

The interest parity condition implies that changes in forward exchange rates have the same effect on capital flows as changes in the difference between Eurocurrency interest rates. Eurobanks offer loans and deposits at an interest rate of $i_M = i_\$ + fp$. Each combination of the Euromark and Eurodollar rates corresponds to a unique fp* in Figure 2 below, where M^S is infinitely elastic. Conversely, once fp* is determined, $i_M - i_\$$ is determined as well.

[7] John Maynard Keynes, A Tract on Monetary Reform (Macmillan, 1924), pp. 125-51. A more direct derivation of interest parity is described in Appendix A.

The degree of integration of the Eurocurrency and forward markets through covered interest arbitrage depends critically on the level of capital mobility. The speed of adjustment to new market interest and exchange rates is generally the measure of market efficiency, and capital mobility is required for these rates to clear the markets.[8] Any divergence of Eurocurrency rates or forward rates from parity should elicit instant realignment through new forward quotations, Eurocurrency quotations, or both.

Market equilibrium coincides with interest parity between Eurocurrencies as well as the equilibrium (or parity) level of the forward premium. In Figure 2, the equilibrium points in both the Eurocurrency and forward exchange markets are influenced by the level at which the net supply equals the net demand for forward currency, as well as by the degree of substitutability between domestic and foreign financial assets (which determines the slope of the net supply schedule for forward exchange, M^S).[9] Risks such as capital controls and other market imperfections that delay or limit the flow of funds cannot be covered in the forward market. They inhibit arbitrage by making covered forward assets less than perfect substitutes for domestic assets. In this situation, M^S is upward sloping.

Even in the absence of restrictions on capital flows, arbitrage does not bring domestic interest rates into relative parity. This is because national monetary policies bear heavily on domestic interest rates, and transaction costs between national money markets are contrastingly higher than in the Eurocurrency market. If there were perfect substitutability (M^S is flat), central banks will lose control over domestic interest rates and inevitably over capital flows. Since this is not the case, the forward premium adjusts to the differential between domestic and foreign interest rates as well as to the differential between the respective Eurocurrency rates. In Figure 2, the forward supply of D-marks simultaneously determines fp^* and i_M, given M^D and $i_\$$.

Forecasting Exchange Rates

Structural exchange rate models are usually specified by supply and demand equations, or their reduced form, representing the principal sources and users of funds of foreign exchange. Alternatively, exchange rate forecasts can be generated from observed prices of money market assets, which, if covered in the forward market, give the corresponding forward premium and interest rate. In efficient markets that reflect all available information, expectations about the future spot exchange rate are generally based on observed spot rates, forward rates, and interest rates. The Eurocurrency market, where the largest volume of (covered) money assets are intermediated internationally, represents an ideal structure for an exchange-rate forecasting model.

Because of the absence of capital controls, the Eurodollar and Euromark markets are tightly linked to domestic money markets in the U.S. and Germany. Consequently, deposit rates in those Eurocurrency markets correspond closely to domestic short-term rates for large deposits. Figure 3 below shows the close tracking of negotiable CD rates (for 3-month deposits) in the U.S. by the Eurodollar rate.

Domestic money markets in the United States and Germany are large in comparison to the size of the forward market for their currencies. An incremental change in supply or demand for forward dollars or marks will not ordinarily lead to a change in U.S.-German interest rate differentials. Rather, the change in forward supply or demand will usually be reflected in a corresponding change in the spot rate, leaving the forward premium at the unchanged interest-rate differential.

The interest parity relationship, supported by Figure 1, can be employed as the basis for forecasting exchange rates. Eurocurrency rates and the forward premium serve as the price variables that describe the market's structure. Models of the Eurodollar, Euromark, and Euroyen markets and their forward exchange markets, representing the bulk of global short-term (covered) financial flows, will provide the framework from which forecasts for each currency's exchange value can be generated.

State Space Model

Univariate time series estimation techniques--usually Box-Jenkins, do not distinguish between endogenous and exogenous variables. Multi-variable time series techniques, however, like econometric estimation, capture some

[8] Covered interest parity is said to be the only "unalloyed criterion" for capital mobility (p. 197, Frankel, J.A.).

[9] The Eurodollar rate is usually assumed as exogenously determined, since U.S. dollars make up about three-quarters of the Eurocurrency market. As such, the Eurodollar rate is influenced largely by U.S. monetary conditions.

Joint Equilibrium in the Forward and Eurocurrency Markets

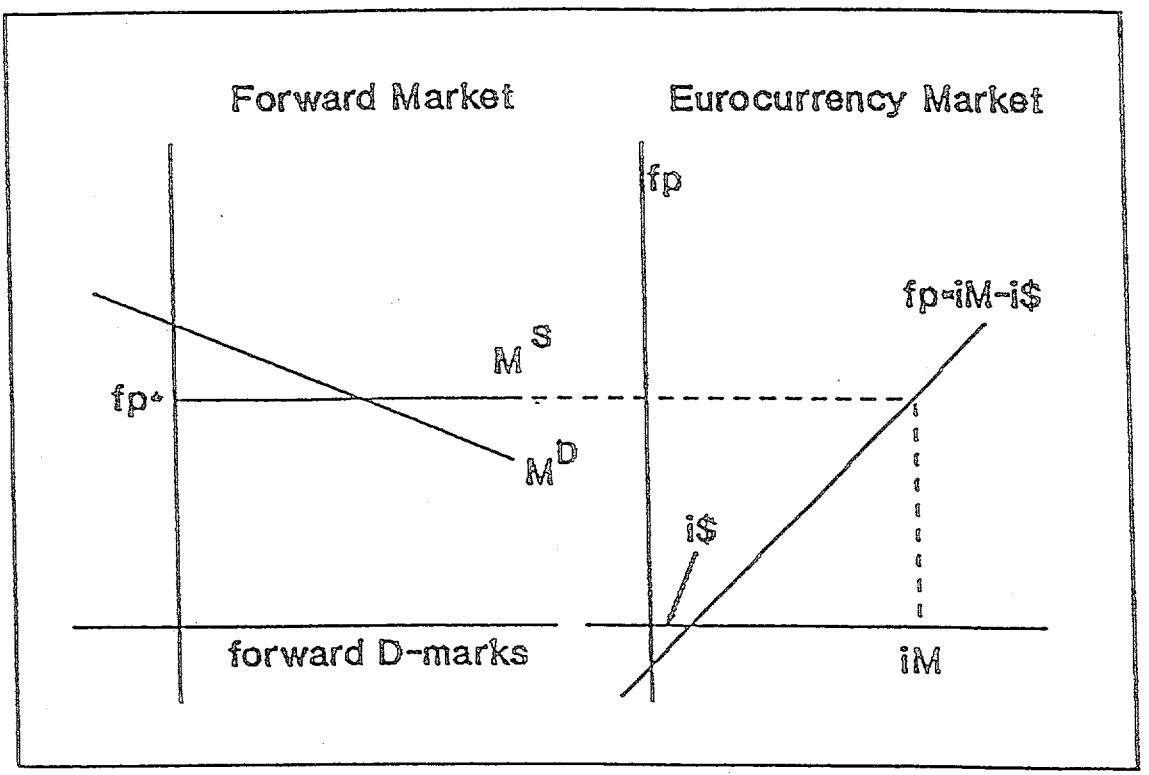


Figure 2

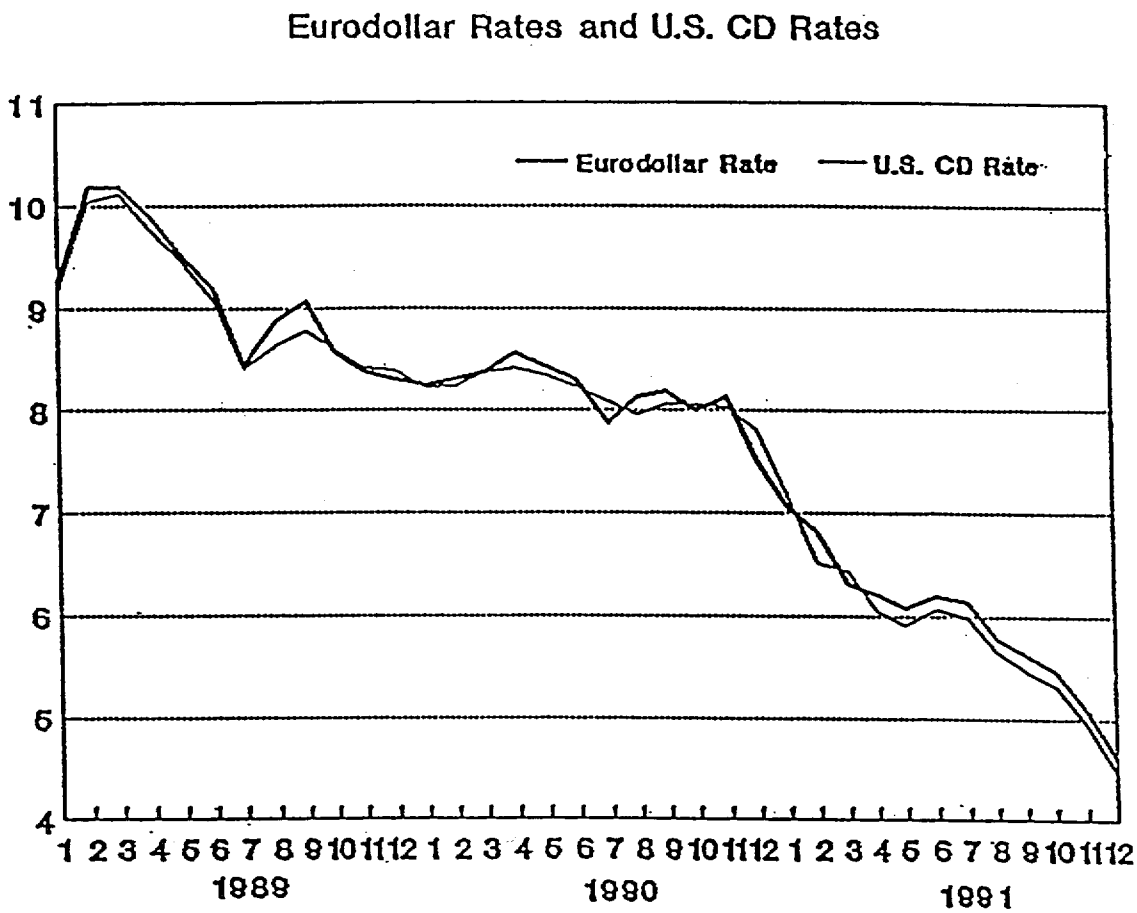

Eurodollar Rates and U.S. CD Rates
Eurodollar Rate
U.S. CD Rate
11
10
9
8
7
6
5
4
1 2 3 4 5 6 7 8 9 10 11 12 1 2 3 4 5 6 7 8 9 10 11 12 1 2 3 4 5 6 7 8 9 10 11 12
1989
1990
1991

Figure 3

causal relationship between the economic variables by differentiating between output and input variables. State space estimation, a generalized time series procedure, accounts for the effect of lead (input) variables on the output (dependent) variables.

A state space model is a multivariate time series forecasting model whose structure is made up of output and input equations in which values of the input variables are required in generating forecasts of the output variables. State space is a generalized form of Box-Jenkins and other statistical time series methods for estimating correlations and lag relationships in equations. As in Box-Jenkins, the output and input variables in the state space model must be stationary--i.e., have an unchanging mean, variance, and nonsignificant autocorrelations. Otherwise the residuals will not be white noise (when the autocorrelation function is not statistically significant for all lags).

A multivariate state space model is represented by two sets of linear equations:

Output equations: $\mathbf{y}_t = \mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{w}_t + \mathbf{v}_t$

Input equations: $\mathbf{x}_t = \mathbf{C}\mathbf{x}_{t-1} + \mathbf{D}\mathbf{z}_t + \mathbf{E}\mathbf{u}_t$

where: $\mathbf{y}_t$ = vector of output (endogenous) variables to be predicted
$\mathbf{x}_t$ = vector of input variables needed to predict $\mathbf{y}_t$
$\mathbf{w}_t$, $\mathbf{z}_t$ = vector of explanatory (exogenous) variables
$\mathbf{v}_t$, $\mathbf{u}_t$ = uncorrelated error terms
A,B,C,D,E are parameter matrices

As the model's structure suggests, the estimated values of the input variables are substituted into the output equations in order to generate forecasts of the output variables. The input variables are taken as exogenous when plugged into the output equations. Error terms of the input equations are assumed to have an autocorrelation structure (E) that accounts for the autoregressive effect of the lagged dependent variable $\mathbf{x}_{t-1}$.

Model Specification

Covered interest parity between two Eurocurrencies implies that when one Eurocurrency rate and the forward premium are determined, the other Eurocurrency rate is determined as well. Since Eurobank liabilities are largely denominated in U.S. dollars, the Eurodollar rate and the forward premia (in terms of the dollar) would be the endogenous variables in the output equation of the state space model. Once they are determined, the Euromark and Euroyen rates would then equal the sum of the Eurodollar rate and the respective forward premia. That is, given that arbitrage between Eurobanks ensures interest parity with the Eurodollar rate:

Euromark/yen rate = Eurodollar rate + mark/yen forward premium

Euromark and Euroyen rates differ from the Eurodollar rate only by the cost of forward cover.

In the state space format outlined below, the Eurodollar rate and the forward premia are the output variables (y_t). The Eurodollar rate is a function of the U.S. CD rate and the forward premia for the mark and yen. The forward premia in turn depend on the differential between respective deposit rates as well as the percentage change in relative export prices. The domestic deposit rates are the input variables (x_t), specified as functions of the treasury bill rate and CPI inflation rate. The explanatory variables in each equation were chosen because they normally lead or predict movements in the corresponding dependent variables. While in practice the forward premia for the mark and yen are determined simultaneously with their Eurocurrency rates, in the absence of capital controls, they can alternatively be estimated from interest differentials between national money market rates. They are also influenced to some extent by relative export prices in which exchange risk generally requires traders to cover exposed currency positions through the forward market. The following equations define the model's structure:

Output equations:

Eurodollar rate = f{U.S. CD rate, mark/dollar forward premium, yen/dollar forward premium}

Mark/dollar forward premium = f{U.S. CD-German deposit rate differential, percent change in relative export prices}

Yen/dollar forward premium = f{U.S. CD-Japanese time deposit rate differential, percent change in relative export prices}

Input equations:

U.S. CD rate = f{U.S. treasury bill rate, U.S. CPI inflation rate}

German deposit rate = f{German treasury bill rate, German CPI inflation rate}

Japanese deposit rate = f{Japanese time deposit rate, Japanese CPI inflation rate}

In forecasting exchange rates, the forward rate F is generated from the ratio of the Euromark/yen yield to the Eurodollar yield multiplied by the base spot rate S (the last actual value in the sample period), which follows directly from the assumption of interest parity (see Appendix A). That is, after equating the Eurodollar yield to the covered Euromark yield, the forecasting equation for F is as follows:

$$F = S*(1+i_M)/(1+i_\$)$$

$i_\$$ and i_M are the forecast values for the Eurodollar and Euromark rates. This relationship does not, however, imply any unidirectional causality. Rather, with the money and foreign exchange markets responding continuously to one another, the forward rate is determined simultaneously with the Eurocurrency rates (and the spot rate, which is assumed to be constant). Hence an exogenous disturbance causing a divergence from the parity condition induces simultaneous equilibrating adjustments in either the spot rate, the forward rate, or the two interest rates, or any combination thereof.

The model's specification depicts the linkage of national money markets to the international money market, with the presumption that cross-border capital movements are unrestricted. Domestic monetary conditions as represented in domestic interest rates are influenced by government monetary policy (through the treasury bill rate) and by price inflation. Since U.S. and German treasury bill rates and the Japanese time deposit rate normally lead their domestic deposit rates, they are used in place of the lagged domestic deposit rate to add explanatory power in the input equations. The model's recursive structure captures this progression of causality--from consumer and export prices to domestic interest rates and the forward premia, and then to Eurocurrency rates and the expected future spot rate, represented by F above.[10]

The order of estimation starts with the input equations for the national deposit rates, which are fed into the output equations to generate projections for the Eurodollar rate and the forward premia. The latter are then used to calculate the Euromark and Euroyen rates, whose covered returns when equated to the Eurodollar return provide forecasts for the forward exchange rate F in the forecasting equation. This process is equivalent to forecasting F, given the last actual spot rate, from projections of the interest differential, which equals the forward premium.

Projections of the exogenous variables in the input equations--Treasury bill rates and time deposit rate, inflation rates, and relative export prices--were generated by univariate Box-Jenkins. All short-term interest rates are monthly averages for 3-month deposits. The 10-year sample period covers January 1983 through December 1991, a total of 120 monthly observations; the forecast period is from January 1992 to June 1992.

[10] In contrast to the simultaneous estimation of parameters of the endogenous variables, a recursive structure sequentially estimates parameters.

Estimation Results

The equations were individually estimated and checked for autocorrelation in the residuals.[11] The requirement that variables be jointly stationary in each equation was relaxed as long as the Ljung-Box chi-squared statistic for error autocorrelation was statistically not significant. Large values of the Ljung-Box statistic indicate that the model does not fit the data. All dependent variables were identified as endogenous and explanatory variables as exogenous in the estimation procedure. The estimated parameters and test statistics of each equation are shown in Appendix B.

The Box-Jenkins models chosen for the inflation rate and relative export prices were poor fits, with large root mean square errors, but have nonsignificant Ljung-Box chi square statistics (the test for serial correlation). The forecasts generated by these univariate models generally exhibited constant patterns. This lack of variability in the forecast values of the exogenous variables should, however, produce more stable (or linear) behavior in projections for the endogenous variables. State space forecasts of the domestic deposit rates show linear trends with minuscule slopes. The Japanese deposit rate exhibited the least variability, which can be attributed to the relatively insulated money market in Japan. Japan's financial markets are not completely deregulated.

The root mean square forecast error for the Eurodollar rate was relatively small in comparison to the forward premia and the domestic deposit rates. This is because Eurodollars dominate the Eurocurrency market, and therefore are less vulnerable to external forces than other currencies. While the respective contributions of the forward premia with respect to the D-mark and the yen were statistically significant in explaining the variability of the Eurodollar rate, they paled in comparison to the U.S. CD rate's effect. The forward premium equations also show a significantly greater sensitivity to the domestic interest rate differentials than to relative export prices. Similarly, in the input equations, the domestic deposit rates responded more to the treasury bill rates than to consumer price inflation.

Evaluation of Forecasts

The out-of-sample projections for the exogenous (inflation rates and relative export prices) and input (domestic deposit rates and forward premia) variables are shown in Appendix C. The resulting Eurocurrency rate forecasts, when expressed as differentials, give the following exchange rate forecasts of the dollar's value in terms of the mark and the yen:

a. German marks per dollar:

1992	diff. of interest rate forecasts	diff. of actual interest rates	exchange rate forecast (DM/$)	actual exchange rate (DM/$)	absolute forecast error
January	5.6836	5.355	1.647	1.579	4.3%
February	5.7879	5.457	1.650	1.619	1.9%
March	5.8600	5.353	1.651	1.662	0.66%
April	5.9089	5.486	1.652	1.649	0.18%
May	5.9417	5.955	1.652	1.623	1.8%
June	5.9632	5.771	1.653	1.573	5.1%

Average forecast error: 2.32%

[11] The multivariate state space procedure of FORECAST MASTER PLUS (Version 1.01), a statistical forecasting package, was used in estimating the parameters of the equations.

b. Japanese yen per dollar:

1992	diff. of interest rate forecasts	diff. of actual interest rates	exchange rate forecasts (yen/$)	actual exchange rate (yen/$)	absolute forecast error
January	-1.089	1.074	126.63	125.46	0.93%
February	-1.100	1.077	126.69	127.70	0.79%
March	-1.104	0.628	126.68	132.86	4.65%
April	-1.106	0.458	126.74	133.54	5.09%
May	-1.107	0.779	126.68	130.77	3.13%
June	-1.107	0.635	126.68	126.84	0.13%

Average forecast error: 2.45%

The relatively low average forecast errors for both exchange rates are encouraging despite the narrow range of the exchange rate projections. The smaller interest differentials between Euroyen and Eurodollars accounted for the dominant influence of the base spot rate used in calculating the yen per dollar forward rate (the implicit future spot rate). This becomes more apparent by noting the large gap between the actual and projected interest rate differentials. To further test the accuracy of these forecasts, actual values of the explanatory variables in the forecast period were substituted instead into the output equations. Both the mark per dollar and yen per dollar ex post projections produced a 2.3 percent average forecast error (projections of variables are in Appendix C). Thus the out-of-sample forecasts in which predicted values for the explanatory variables were individually generated compare favorably with the ex post projections. These satisfactory results demonstrate the predictive ability of the interest parity theorem and the applicability of the state space model. The forward rate in this case provided an acceptably accurate predictor for the future spot exchange rate.

One advantage of a multivariate time series model over a univariate model derives from the ability to simulate various assumptions about domestic interest rates, inflation rates, or exchange rate policies and observe the impact on the output variables. When joint stationarity holds and residuals are white noise, econometric complications from multicollinearity, serial correlation, and simultaneity are avoided. For example, lagged dependent variables and serial correlation would have posed a serious problem had an econometric estimation method been applied instead. The state space model of relating Eurocurrency rates to lead variables lends itself uniquely to forecasting.

Summary and Conclusions

The role of forward exchange in covering short-term Eurocurrency deposits and loans provides a means for determining the expected future spot rate. Since the forward premium is the foreign exchange market's measure of a currency's expected appreciation, it can be used in predicting the future spot rate. By invoking the interest parity relationship, forecasts of Eurocurrency rates can be derived from estimates of the forward premia and the Eurodollar rate. Interest arbitrage leads to an equilibrium condition in which covered yields between Eurocurrencies converge, and where the forward rate represents the expected future spot rate. The Eurocurrency rates are determined as soon as the forward premia are determined. This way of forecasting exchange rates was suitably accommodated in the state space model for time series analysis. A recursive structure was constructed for the purpose of projecting Eurocurrency interest rates from forward premia. While satisfactory exchange rate forecasts were obtained from this model, the base spot rate was highly influential in determining the forecast values, overshadowing the effect of the interest differential.

This approach to forecasting exchange rates implicitly makes exchange rates a function of price variables--relative interest rates, inflation rates, and export prices, as opposed to underlying market demand and supply

variables. That is, domestic monetary and price conditions are reflected in the external market by their transmission through domestic interest rates to the forward premia. The assumptions that changes in domestic interest rates lead (by one month) changes in Eurocurrency rates and that price inflation affects domestic interest rates were built into the recursive structure of the model. Also, the forward premium is linked to changes in relative export prices and projected domestic interest differentials. Univariate Box-Jenkins forecasts of treasury bill rates, CPI inflation rates, and changes in the relative export price indexes were used to generate out-of-sample projections for the exogenous variables. The forecast results were relatively accurate because the actual exchange rates did not significantly diverge from the base exchange value. The marginal contribution to the forecast values of the interest differentials was beneficial in the case of the yen, where the discrepancy between the actual and projected interest differentials was particularly wide. These results, however, point out the model's vulnerability to exchange market volatility, especially when speculative activity is high. Had the exchange rate projections been extended to the second half of 1992 when the European Monetary System unraveled, the model's predictive ability would have suffered. But then, when projections based on one initial exchange value are made beyond a few months, the base exchange rate's influence always tends to recede.

The assumption that the forward rate represents the best guess of the future spot rate, as was made here, stems simply from the way the forward premium is computed--as the percentage difference over the spot rate. A forward rate bias is introduced by this assumption only if over time the average forecast error between the forward rate and the future spot rate is not zero. Similarly, if the forward premium and the interest differential consistently vary from each other, the forward rate would likely be a biased predictor of the future spot rate. While the simultaneous interaction of the forward premium and the interest differential due to the interdependence of the Eurocurrency and forward exchange markets may introduce a simultaneity, or endogeneity, bias in the forecast values of the Eurocurrency rates, taking differentials would minimize it. The unpredictable relationship between short-term interest rates and price inflation because of active government intervention in the money markets compounds some lead-lag assumptions in the model. Fortunately, if stationarity is achieved between these variables, state space estimation generates forecasts that are statistically sound. Last, not least, by purposely excluding exchange rates from the model, the problem of joint stationarity was avoided since exchange rates follow a random walk, thus nonstationary, process. Moreover, the inability by previous studies to find cointegration--a linear combination that is stationary--between nominal exchange rates and relative money supplies suggests that any structural linkage of exchange rates and interest rates, other than indirectly through the interest parity theorem, would not have been feasible in a time series model (see p. 548, Adams and Chadha).

From the premise that hedged and nonhedged holders of forward contracts on foreign-currency assets conduct their bidding by equating real rates of interest (adjusted for price inflation), exchange rates can also be predicted from inflation rate differentials (see p.891, Giddy). In practice, however, real interest rates tend to equalize only in the long run and observed exchange rates differ only partly because of price disparities. The forward premium, on the other hand, reflects the broadest range of market expectations, including inflation, and hence is a superior indicator of currency preference.

Appendix A: Covered Interest Parity Equation

The parity equation can also be derived by comparing expected returns between domestic and foreign investments. This simpler model depends directly on the price variables as opposed to structural variables. Under the same conditions of perfect competition and free currency convertibility, domestic and foreign yields (i_h and i_f) will tend to converge in equilibrium. Assuming continuous compounding of interest,

$$e^{i_h t} = e^{i_f t}\frac{r_t}{r_0}$$

$$\frac{r_t}{r_0} = (i_h - i_f)t$$

r_t is the forward rate of exchange for delivery at time t (in fractions of a year), and r_0 is the spot rate at time 0. Expressing the forward discount on domestic currency in percent per annum,

$$fp = \frac{r_t - r_0}{r_0 t}$$

$$\frac{r_t}{r_0} = 1 + fp * t$$

Therefore,

$$1 + fp * t = e^{(i_h - i_f)t}$$

After taking logs, the first-order approximation with respect to t gives

$$fp = i_h - i_f$$

which is the familiar equivalence between the interest differential and the cost of the forward contract. Market equilibrium is reached when the forward premium equalizes effective interest rates.

Hence in equating 1-year yields between Eurodollar and Euromark deposits, $i_\$$ and i_M,

$$1 + i_\$ = (1 + i_M) * \frac{r_t}{r_0}$$

which, if i_M is small, can be simplified to

$$i_\$ = i_M + fp$$

Appendix B: Estimation Statistics

A. State Space Diagnostics of Output Equations:

1. Eurodollar rate

Transformation for stationarity: differencing order=1; power=log
U.S. CD rate coefficient: 0.8789
D-mark/$ forward premium coefficient: 0.0052
Yen/$ forward premium coefficient: -0.0003
Adjusted R-square: 0.995
RMS (root mean square) error: 0.014
Ljung-Box chi square: 17.72 (0.526)
Schwarz criterion (BIC): 0.127
Error autocorrelation function: no significant autocorrelation (except at lag 1)

2. Mark/dollar forward premium

Transformation for stationarity: degree of differencing=1
Deposit rate difference coefficient: -.0989
(transformation for stationarity: degree of differencing=1; power=log)
Percent change in relative export prices coefficient: .007
Adjusted R square: 0.965
RMS error: 0.471
Ljung-Box chi square: 12.87 (0.201)
Schwarz criterion (BIC): 0.514
Error autocorrelation function: no significant autocorrelations

3. Yen/dollar forward premium

Transformation for stationarity: degree of differencing=1; power=inverse
Deposit rate difference coefficient: 0.19
(transformation for stationarity: degree of differencing=1; power=inverse)
Percent change in relative export prices coefficient: -.064
Adjusted R square: 0.00
RMS error: 4.757
Ljung-Box chi square: 1.98 (0.00)
Schwarz criterion (BIC): 24.946
Error autocorrelation function: no significant autocorrelations

B. State Space Diagnostics of Input Equations:

1. Domestic deposit rates

a) United States

transformation for stationarity: degree of differencing=1; power=log
Treasury bill rate coefficient: 7.102
CPI inflation rate coefficient: 0.0015
Adjusted R square: 0.980
RMS error: 0.027
Ljung-Box chi square: 26.78 (0.917)
Schwarz criterion (BIC): 0.2283
Error autocorrelation function: significant autocorrelations at lags 3, 11 (at 5% level)

b) Germany

Transformation for stationarity: degree of differencing=1; power=inverse
Treasury bill rate coefficient: 0.353

CPI inflation rate coefficient: 0.0018
Adjusted R square: 0.989
RMS error: 0.007
Ljung-Box chi square: 22.08 (0.771)
Schwarz criterion (BIC): 0.1724
Error autocorrelation function: no significant autocorrelation

c) Japan

Transformation for stationarity: degree of differencing=1
Treasury bill rate coefficient: 0.5806
CPI inflation rate coefficient: 0.00455
Adjusted R square: 0.992
RMS error: 0.123
Ljung-Box chi square: 24.95 (0.874)
Schwarz criterion (BIC): 0.134
Error autocorrelation function: significant autocorrelation at lag 3 (at 5% level)

C. Box-Jenkins Diagnostics of Exogenous Variables

1. Treasury bill rates

a) United States

Transformation for Stationarity: differencing order=1; power=inverse
first-order autoregressive coefficient: 0.493 (t statistic=5.743)
Adjusted R square: 0.967
RMS error: 0.0057
Ljung-Box chi square: 9.28 (0.047)
Schwarz criterion (BIC): 0.312
Error autocorrelation function: no significant autocorrelation

b) Germany

Transformation for stationarity: differencing order=1; power=inverse
first-order moving average coefficient: -0.2224 (t statistic=-2.489)
Adjusted R square: 0.961
RMS error: 0.0134
Ljung-Box chi square: 13.63 (0.247)
Schwarz criterion (BIC): 0.4076
Error autocorrelation function: no significant autocorrelation

c) Japan

Transformation for stationarity: differencing order=1
First-order moving average coefficient: -0.459 (t statistic=-5.607)
Adjusted R square: 0.971
RMS error: 0.2364
Ljung-Box chi square: 12.17 (0.162)
Schwarz criterion (BIC): 0.241
Error autocorrelation function: no significant autocorrelation

2. Consumer price index inflation rates

a) United States

Transformation for stationarity=none
Constant: 0.195
First-order moving average coefficient: -0.556

First-order seasonal autoregressive coefficient: 0.393 (t statistic=4.688)
Adjusted R square: 0.296
RMS error: 0.216
Ljung-Box chi square: 24.52 (0.861)
Schwarz criterion (BIC): 0.2247
Error autocorrelation function: no significant autocorrelation

b) Germany

Transformation for stationarity: seasonal differencing order=1
First-order autoregressive coefficient: 0.447 (t stat.=4.677)
First-order seasonal moving average coefficient: 0.844 (t statistic=21.173)
Adjusted R square: 0.299
RMS error: 0.203
Ljung-Box chi square: 2.548 (0.827)
Schwarz criterion (BIC): 0.2115
Error autocorrelation function: significant autocorrelation at lag 9 (at 5% level)

c) Japan

Transformation for stationarity: order of differencing=1; power=inverse
First-order moving average coefficient: 0.986 (t statistic=101.789)
Adjusted R square: 0.00
RMS error=2.50
Ljung-Box chi square: 15.42 (0.367)
Schwarz criterion (BIC): 0.9347
Error autocorrelation function: significant autocorrelation at lag 10 (at 5% level)

3. Change in relative export prices

a) Germany/United States

Transformation for stationarity: degree of differencing=1; power=log
First-order moving average coefficient: 0.6667 (t statistic=3.894)
Second-order moving average coefficient: 0.2698 (t statistic=3.118)
Adjusted R square: 0.00
RMS error: 0.808
Ljung-Box chi square: 11.90 (0.146)
Schwarz criterion (BIC): 1.2006
Error autocorrelation function: no significant autocorrelation

b) Japan/United States

Transformation for stationarity: degree of differencing=1; power=log
First-order moving average coefficient: 0.9218 (t statistic=25.424)
Adjusted R square: 0.00
RMS error: 0.805
Ljung-Box chi square: 10.19 (0.074)
Schwarz criterion (BIC): 1.0159
Error autocorrelation function: no significant autocorrelation

Appendix C: Projections

A. Out-of-Sample Forecasts

1. Input variables:

a) Treasury bill rates

		U.S.	Germany	Japan
1992	Jan	3.9185	9.1172	5.9903
	Feb	3.8263	9.1172	5.9903
	Mar	3.7824	9.1172	5.9903
	Apr	3.7612	9.1172	5.9903
	May	3.7508	9.1172	5.9903
	Jun	3.7457	9.1172	5.9903

b) Consumer price index inflation rates

		U.S.	Germany	Japan
1992	Jan	0.3196	0.5482	0.4505
	Feb	0.2577	0.2320	0.4505
	Mar	0.2577	-0.0016	0.4505
	Apr	0.2574	0.2798	0.4505
	May	0.2888	0.2175	0.4505
	Jun	0.3195	0.2746	0.4505

c) Domestic deposit rates

		U.S.	Germany	Japan
1992	Jan	4.1480	7.9632	5.9444
	Feb	4.0007	7.9588	5.9258
	Mar	3.9305	7.9688	5.9213
	Apr	3.8972	7.9490	5.9202
	May	3.8817	7.9315	5.9199
	Jun	3.8750	7.9005	5.9198

2. Output variables:

a) Percent change in relative export prices (relative to the U.S.)

		Germany	Japan
1992	Jan	2.1441	1.3922
	Feb	1.5792	1.3922
	Mar	1.5792	1.3922
	Apr	1.5792	1.3922
	May	1.5792	1.3922
	Jun	1.5792	1.3922

b) Forward premium (in terms of the U.S. dollar)

		Germany	Japan
1992	Jan	5.6836	-1.0891
	Feb	5.7879	-1.0996

Mar	5.8600	-1.1037
Apr	5.9089	-1.1057
May	5.9417	-1.1066
Jun	5.9632	-1.1070

c) Eurocurrency deposit rates

		U.S.	Germany	Japan
1992	Jan	4.3071	9.9907	3.218
	Feb	4.1597	9.9476	3.060
	Mar	4.0886	9.9486	2.985
	Apr	4.0545	9.9634	2.949
	May	4.0383	9.9800	2.932
	Jun	4.0311	9.9943	2.924

B. In-sample forecasts for the Eurocurrency rates (actual data provided for input variables--domestic deposit rates and forward premia):

		U.S.	Germany	Japan
1992	Jan	4.2210	9.6573	5.1677
	Feb	4.2279	9.6329	5.2616
	Mar	4.3918	9.6563	4.9978
	Apr	4.1613	9.6967	4.7789
	May	3.9846	9.7217	4.7999
	Jun	4.0143	9.7218	4.6705

Base spot exchange rate (last actual value of sample period):

Dec. 1991: D-mark/dollar = 1.563
Yen/dollar = 128.04

References

1. Abraham, Bovas and Johannes Ledolter. Statistical Methods for Forecasting. John Wiley, 1983.

2. Adams, Charles and Bankim Chadha, "Structural Models of the Dollar," IMF Staff Papers, vol. 38, no. 3, pp. 525-59.

3. Baillie, Richard T. and Patrick C. McMahon. The Foreign Exchange Market: Theory and Econometric Evidence. Cambridge University Press, 1989.

4. Clendenning, E. Wayne. The Euro-Dollar Market. Oxford University Press, 1970.

5. Cumby, Robert E. and Maurice Obstfeld, "A Note on Exchange-Rate Expectations and Nominal Interest Differentials: A Test of the Fisher Hypothesis," The Journal of Finance, June 1981, pp. 697-703.

6. de Grauwe, Paul. International Money. Oxford University Press, 1989.

7. Dufey, Gunter and Ian H. Giddy. The International Money Market. Prentice-Hall, Inc., 1978.

8. Frankel, Jeffrey A., "Measuring International Capital Mobility: A Review," Papers and Proceedings of the American Economic Association, January 1992, pp. 197-202.

9. Gibson, Heather D. The Eurocurrency Markets, Domestic Financial Policy and International Instability. St. Martin's Press, 1989.

10. Giddy, Ian H., "An Integrated Theory of Exchange Rate Equilibrium," Journal of Financial and Quantitative Analysis, Dec. 1976, pp. 883-92.

11. Goodrich, Robert L. Applied Statistical Forecasting. Business Forecast Systems, Inc., 1989.

12. Goodrich, Robert L. and Eric A. Stellwagen. Forecast Master Plus. Documentation Version 1.01, Business Forecast Systems, Inc., 1983.

13. Granger, C.W.J. and Paul Newbold. Forecasting Economic Time Series, 2nd ed. Academic Press, Inc., 1986.

14. Herring, Richard J., "The Interbank Market," Eurodollars and International Banking, Paolo Savona and George Sutija, eds. MacMillan, 1985.

15. Herring, Richard J. and Richard C. Marston, "The Forward Market and Interest Rates in the Eurocurrency and National Money Markets," Eurocurrencies and the International Monetary System, Carl H. Stem, John M. Makin and Dennis E. Logue, eds., American Enterprise Institute, 1976.

16. Herring, Richard J. and Richard R. Marston. National Monetary Policies and International Financial Markets. North-Holland Publishing Company, 1977.

17. International Monetary Fund. International Financial Statistics. Electronic database.

18. Kubarych, Roger M. Foreign Exchange Markets in the United States. Federal Reserve Bank of New York, August 1978.

19. Levi, Maurice D. International Finance, 2nd ed. McGraw-Hill Publishing Company, 1990.

20. Levich, Richard M. The International Money Market: An Assessment of Forecasting Techniques and Market Efficiency. Contemporary Studies in Economic and Financial Analysis, vol. 22, 1979.

21. Llewellyn, David T. International Financial Integration. John Wiley, 1980.

22. Marston, Richard C., "Interest Arbitrage in the Euro-Currency Markets," European Economic Review, January 1976, pp. 1-13.

23. McKinnon, Ronald I. Money in International Exchange: The Convertible Currency System. Oxford University Press, 1979.

24. Sohmen, Egon, "The Theory of Forward Exchange," Princeton Studies in International Finance No. 17, 1966.

25. Stein, J., "The Forward Rate and the Interest Parity," Review of Economic Studies, March 1965, pp. 113-26.

26. Stock, James and Mark W. Watson, "Variable Trends in Economic Time Series," Journal of Economic Perspectives, vol. 2 (Summer, 1988), pp. 147-74.

27. Wolff, Christian C.P., "Forward Foreign Exchange Rates, Expected Spot Rates, and Premia: A Signal-Extraction Approach," The Journal of Finance, vol. XLII, No. 2, June 1987, pp. 395-406.

Quantitative Policy Objectives: A Case for Moving Targets

Andreas Muller, Department of Health Services Administration, University of Arkansas at Little Rock

Introduction

Planners in government and large corporations are often called upon to propose quantitative objectives which are typically expressed in fixed rates pertaining to a future year. The corporate average fuel efficiency (CAFE) standard proposed to be 40 mpg by the year 2001 is one example[1]; another one is the objective to reduce the U.S. infant mortality rate to 7/1,000 live births by the year 2000.[2]

Stating policy objectives in quantitative terms has several advantages. By definition, quantitative targets allow numerical comparisons; that is, the difference between the base line period and a future period can be expressed as absolute or relative change. Quantitative objectives also help monitor target achievement and potentially make government agencies, or corporations more accountable. Quantitative targets also foster clearer communication between the policy designers and those who have to implement the policy. It is not surprising that in results-oriented circles of government and the corporate world there is growing support for expressing policy objectives in quantitative terms.

The Problem

Comparing quantitative targets with actual performance is a deceptively simple task. For example, a five-year objective is set 20% below the baseline level. A program is implemented which is expected to achieve the 20% reduction. Progress toward the objective is monitored and measurement in the last year of the planning period indicates a 25% reduction compared with the baseline period. A strong temptation exists to declare the program a success and to consider the objective "accomplished". On the other hand, if only a 2% reduction, or an unexpected 5% increase occurred, then the temptation exists to declare the program a failure. Indeed, none of the inferences is justified without assuming that policy relevant factors other than the program intervention did not change, or that those changes were mutually offsetting.

External factors, that is, factors over which government, or the corporation has little control, typically affect policy objectives and need to be taken into account. Without explicit consideration of major policy relevant factors, quantitative targets are likely to be misspecified and incorrect inferences will be drawn about target achievement.

The problem will be illustrated and an adjustment method suggested using the *National Health Promotion and Disease Prevention Objectives*, health status objective 9.3 which states:

> *"Reduce deaths caused by motor vehicle crashes to no more than 1.9/100 million vehicle miles traveled and 16.8 per 100,000 people. (Baseline: 2.4 per 100 million vehicle miles traveled (VMT) and 18.8 per 100,000 people (age adjusted)".*[2] (p.274)

The objective implies that the mileage-based death rate and the age adjusted death rate shall be reduced by 21% and 11%, respectively. The larger reduction target in mileage-based motor vehicle death rates reflects trend adjustment for the period 1980-87. The writers of the policy document acknowledge that the achievement of the objectives will depend upon factors such as the volume of mass transit and various transportation safety initiatives.[2]

---

[1] Insurance Institute for Highway Safety. 1992. *Status Report.* Arlington, VA; Insurance Institute for Highway Safety. 27(6): 5.

[2] U.S Department of Health and Human Services, Public Health Service. 1992. *Health People 2000: National Health Promotion and Disease Prevention Objectives.* Boston, MA; Jones and Bartlett Publishers.

Time Series Analysis of Motor Vehicle Death Rates

Figure 1 shows the population (left axis) and mileage- based motor vehicle death rates (right axis) from 1950-1990.[3] The horizontal lines from 1987 to 2000 indicate the respective year 2000 policy objectives. The population-based death rate are not age-adjusted, thus somewhat overstating the gap between actual and policy target.

Since 1950 the population based motor vehicle death rate stayed in the 20 to 30/100,000 range, with death rates approaching their highest levels in the mid-1960s. Motor vehicle death rates based on vehicle miles traveled show an overall downward trend; otherwise resembling the population based death rates. The vertical grid line indicates the onset of the 1973 oil crisis.

Figure 2 shows U.S. per capita alcohol consumption and the unemployment rate for the years 1950-1990. Per capita alcohol consumption represents the amount of pure drinking alcohol (ethanol) consumed by the U.S. population 14 years of age and older.[4] The unemployment rate is measured by the per cent of the U.S. civilian labor force 16 years of age and older out of work and looking for work in the last two weeks.[5] Prior research found alcohol consumption positively and unemployment negatively related to motor vehicle fatality rates, respectively.[6,7,8,9,10] The 1973-74 oil crisis reduced motor vehicle deaths[6]; its impact needs to be incorporated into the statistical model. A transfer function model[11] can be stated as follows:

$$\Delta Y_t = C + \omega_1 \Delta UR_t + \omega_2 \Delta PCA_t + \omega_3 OILP73_t + \omega_4 OILP74_t + \theta(B)/\phi(B) * a_t; \quad (1)$$

Y_t	= annual motor vehicle death rates;
C	= constant(deterministic trend);
UR_t	= annual unemployment rates;
PCA_t	= per capita alcohol consumption;
$OILP73_t$	= dummy variable pulse coded "1" for 1973, otherwise "0";
$OILP74_t$	= dummy variable pulse coded "1" for 1974, otherwise "0";
$\theta(B)$	= nonseasonal MA operator of order $(1-\theta_1 B-\theta_2 B^2 -\ldots-\theta_q B^q)$;
$\phi(B)$	= nonseasonal AR operator of order $(1-\phi_1 B-\phi_2 B^2 -\ldots-\phi_p B^p)$;
a_t	= residuals $(NI(0)\sigma)$;

[3] U.S. Department of Transportation, Federal Highway Administration. 1991. *Highway Statistics* (various years). Washington, DC; U.S. Government Printing Office.

[4] National Institute on Alcohol Abuse and Alcoholism, Division of Biometry and Epidemiology, Alcohol Epidemiologic Data System. 1991. *Apparent Per Capita Alcohol Consumption: National, State, and Regional Trends: 1977-1988.* Surveillance Report # 16. Washington, DC; CSR,l Inc. Table 1.

[5] Department of Labor, Bureau of Labor Statistics. 1992. *Employment and Earnings* 39(1):162.

[6] Partyka, S.C. 1984. Simple models of fatality trends using employment and population data. *Accident Analysis and Prevention* 16:211-222.

[7] Wagenaar, A.C., Streff F.M. 1989. Macroeconomic conditions and alcohol-impaired driving. *J. of Studies on Alcohol* 50(3):217-225.

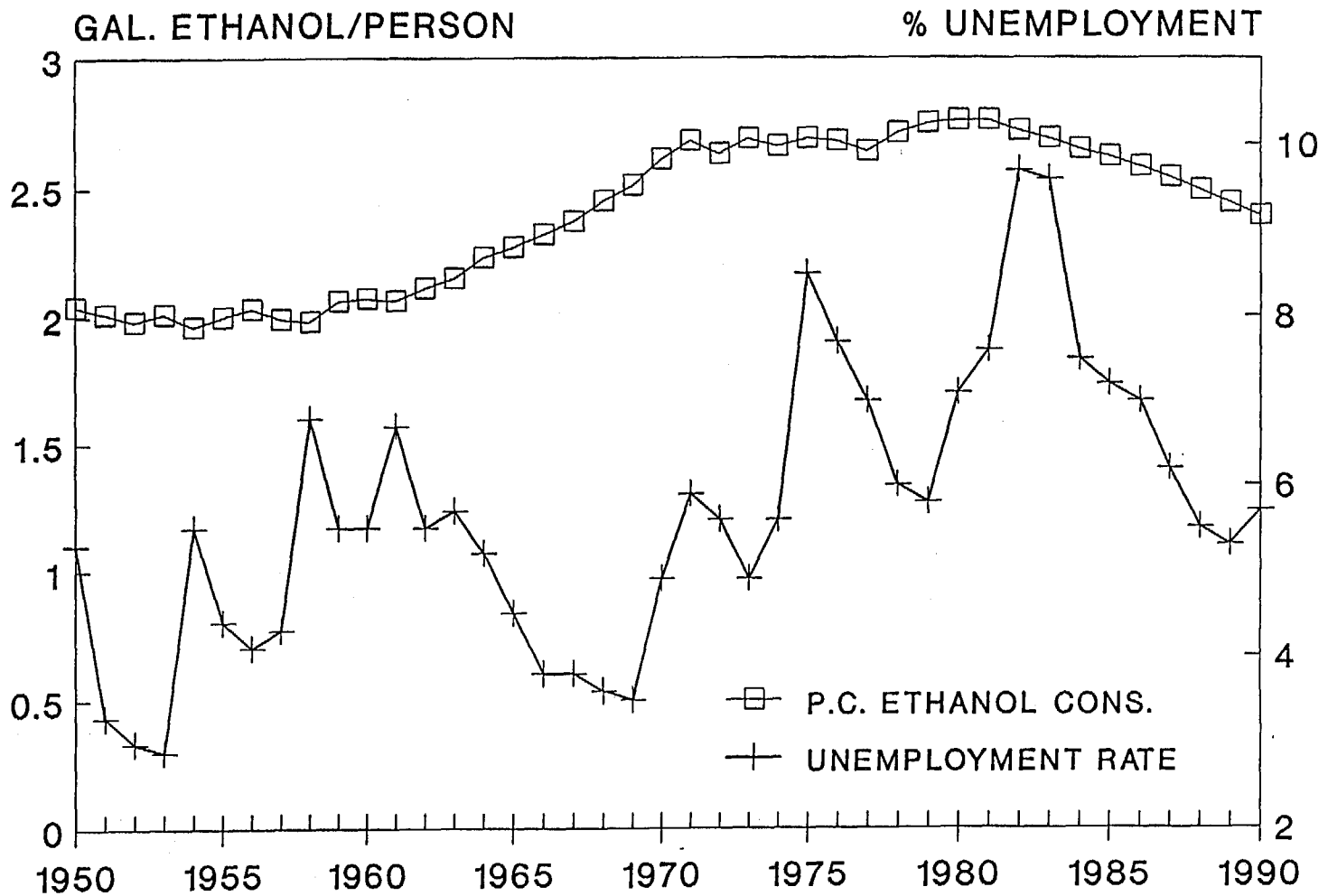
[8] Evans, W.N., Neville, D., Graham, J.D. 1991. General deterrence of drunk driving: evaluation of recent american policies. *Risk Analysis* 11(2):279-289.

[9] Muller, A. 1989. Business recession, alcohol consumption, drinking and driving laws: impact on Oklahoma motor vehicle fatalities and fatal crashes. *American J. of Public Health* 79(10):1366-1370.

[10] Leigh J.P., Waldon, H.M. 1991. Unemployment and highway fatalities. *J. of Health Politics, Policy and Law* 16(1):135-156.

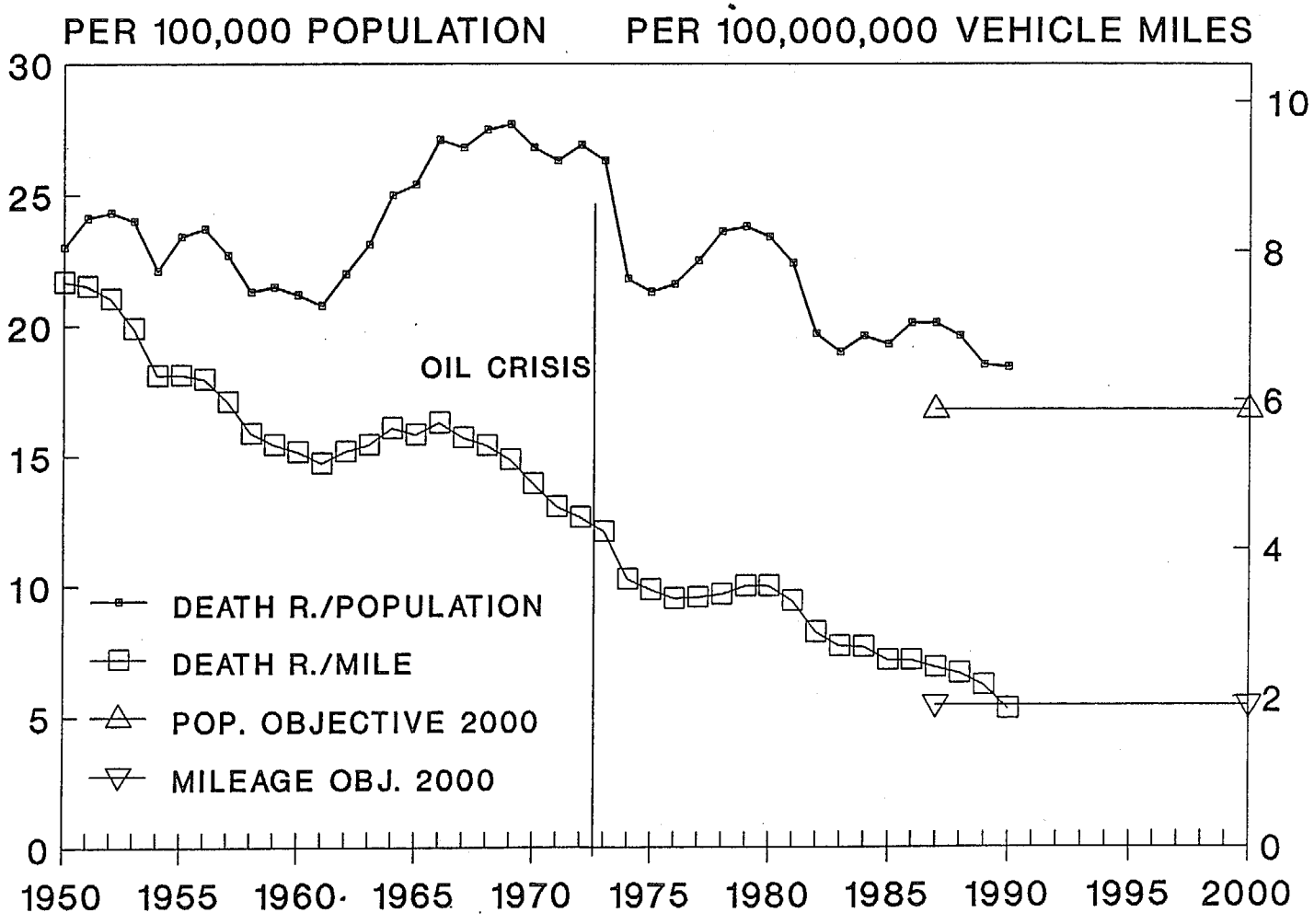
[11] Pindyck, R.S., Rubinfeld, D.L. 1991. *Econometric Models and Economic Forecasts* (3rd ed.). New York, NY; Macmillan.

U.S. PER CAPITA ALCOHOL CONSUMPTION AND UNEMPLOYMENT RATES, 1950-1990



NIAAA, DOL

U.S. MOTOR VEHICLE DEATH RATES (1950-90) AND RELATED SAFETY OBJECTIVES



HIGHWAY SAFETY STATISTICS

To stabilize the variance more, mileage-based death rates were ln-transformed. Since the autoregressive parameters of the output series approached the bounds of stationarity, all time series with the exception of the pulse inputs were once differenced (cf. Pankratz[12]). The model only stipulates contemporaneous unemployment and alcohol consumption effects. An examination of the assumption showed that neither lagged relationships, nor feedbacks were present.

The effect of the oil crises was measured by two dummy variable pulses for 1973 and 1974, respectively. In the final model specification a first-order decay effect associated with the 1974 pulse was found to be statistically significant (see Table 1). The nonseasonal *autoregressive-integrated-moving average* model $\{\theta(B)/\phi(B)a_t\}$ captures the lagged time dependencies of the output series. The model was estimated with RATS 4.0 software[13] using maximum likelihood estimation.

Table 1 shows that a 1% increase in the unemployment rate is estimated to reduce population based death rates by -.67 deaths/100,000, or 2.9%, and the mileage based death rate by approximately 2.0%. The coefficient pertaining to per capita ethanol consumption indicates that a 1 gallon increase in the consumption of alcohol increases the population based motor vehicle death rate by 6.95 deaths/100,000, or about 30% and the mileage based death rate by about 26%. These estimates are comparable with other studies explaining U.S. motor vehicle death rates.[8-10]

The impact of the 1973-74 oil crisis is estimated to have reduced population based death rates by 23% and decayed somewhat thereafter settling at a 19% below levels prevailing before the onset of the crisis. The impact of the oil crisis was less pronounced on mileage based death rates. There was 10.5% reduction in 1974 decaying to approximately 7% reduction over the next few years.

The significant constant term in the mileage-based model indicates a 3.5% annual reduction in the rates independent of other model inputs. It is noteworthy that neither the autoregressive term, nor the constant was statistically significant in the population based model indicating no longterm deterministic trend.

The results of the transfer function analysis illustrate that motor vehicle death rates are significantly affected by external factors such as the oil 1973-74 crisis, unemployment and alcohol consumption.[14] The findings further imply that any departure of policy related factors from their baseline values (1987) will effect the measure used to gauge policy achievement. That is, if policy related external factors depart significantly from baseline values, then the baseline, or the policy objective needs to be adjusted.

Adjusting Policy Baseline, or Objectives

Ideally, unemployment rates and per capita alcohol consumption would be forecasted with little error to the year 2000. Then, the forecasted values could be inserted in equation (1) and adjusted baseline values calculated for the year 2000. However, longterm forecasts of times series which show highly stochastic behavior, i.e. unemployment, may not be very useful since they tend to yield forecasts with large errors.[15]

Performing a sensitivity analysis is an alternate approach to dealing with uncertainty about the future. From past data reasonable ranges can be established for the input variables. For the period 1950-1990, unemployment varied between 2.9 and 9.7%, while per capita alcohol consumption ranged between 1.96 and 2.76 U.S. gallons ethanol. In 1987, the baseline year, unemployment rate was 6.2% and per capital alcohol consumption was 2.54 gallons. Likely future levels of unemployment and ethanol consumption can be expressed as differences from their 1987 baseline levels and multiplied by coefficients in Table 1, i.e. -.67 for unemployment and 6.95 for per capita alcohol consumption. The results are displayed in Table 2.

Table 2 estimates changes in motor vehicle death rates as a result of changes in unemployment and per capita alcohol consumption from their respective baseline levels. The italicized bold entries indicate conditions under

[12] Pankratz, A (1991) *Forecasting with Dynamic Regression Models.* New York, NY; John Wiley and Sons. 177-184.

[13] Doan, T.A. 1992. *Rats 4.0.* Evanston, IL; Estima.

[14] Alcohol consumption may be considered a partially external factor, since it is responsive to government interventions such as taxation, tougher drunk driving laws, and prohibition.

[15] McCleary, R., Hay, R.A. 1980. *Applied Time Series Analysis for the Social Sciences.* Beverly Hills, CA; Sage Publ. chpt. 4.

Table 1

TRANSFER FUNCTION MODELS OF U.S. MOTOR VEHICLE DEATH RATES*

INPUT	MV DEATHS/POPULATION ω_i	t	MV DEATHS/MILE(LN) ω_i	t
D-UNEMPLOYMENT RATE	-.67	-7.28	-.020	-4.61
D-P.C. ETHANOL CONS.	6.95	3.40	.256	2.01
OIL CRISIS PULSE '73	-1.48	2.39	-.040	-1.24
OIL CRISIS PULSE '74	-3.82	-6.28	-.105	-3.54
OIL CRISIS: δ_1	-.32	-2.24	-.493	-2.84
NOISE MOD. $\phi(1)$	-	-	.346	1.72
CONSTANT	-	-	-.035	-4.55
s.e.e.	.61		.032	
R^2 adj.	.948		.993	
DURBIN-WATSON	1.82		1.79	
Q(12)	8.31;	P=.60	Q(10) 7.28	p=.61
d.f.	36		34	
Mean	22.87		1.45	

*All time series are in first differences except for oil crisis which is measured by two pulses and a first-order decay effect.

Table 2

ESTIMATED CHANGE IN MOTOR VEHICLE DEATH RATES (100,000 POP.) DUE TO CHANGE IN UNEMPLOYMENT AND P.C. ETHANOL CONSUMPTION FROM 1987 BASELINE

UNEMPLOYMENT RATE	PER CAPITA ETHANOL CONSUMPTION (gal.) 2.00 (-.54)	2.25 (-.29)	2.50 (-.04)	2.75 (.21)
3% (-3.2)	-1.60	.13	1.87	3.60
4% (-2.2)	-2.28	-.54	1.21	2.93
5% (-1.2)	-2.95	-1.21	.53	2.26
6% (-.2)	-3.62	-1.88	-.13	1.59
7% (.8)	-4.29	-2.55	-.80	.92
8% (1.8)	-4.96	-3.22	-1.47	.25
9% (2.8)	-5.63	-3.89	-2.14	-.42
10% (3.8)	-6.30	-4.56	-2.81	-1.09

Numbers in parentheses indicate difference from 1987 baseline levels: per capita ethanol consumption=2.54 gal.; unemployment rate=6.2%.

which the year 2000 objective, a reduction in the population based death rate by 2/1000, is "achieved". For instance, if alcohol consumption were 2.5 gal. in the year 2000 and the unemployment rate 3%, the baseline, or the year 2000 objective would need to be adjusted by 1.87/100,000. In contrast, if the unemployment rate were 9% in 2000, then the baseline, or the policy objective would need to be adjusted downward by 2.14/100,000.

Table 3 shows the estimated per cent change in mileage based motor vehicle death rates by the year 2000 under varying assumptions of unemployment and per capita alcohol consumption. The per cent estimates incorporate an adjustment for the significant longterm deterministic trend[16], i.e. an annual reduction of 3.5% in mileage based death rates between 1987 and 2000.

Since all percentage estimates in Table 3 are less than the national objective (-21%), the year 2000 target is likely to be attained. This observation is not surprising, because the long run trend in mileage based death rates is expected to be about 37% lower, or 1.51/100,000,000 vehicle miles in the year 2000.[17] However, possible variation in per capita alcohol consumption and unemployment rates could result in mileage based death rates ranging from .99 (2.4*(1-.586)) to 1.8 (2.4*(1-.252)) per 100,000,000 vehicle miles by the year 2000.

Discussion

Planners interested in setting realistic objectives know that objectives need to be calibrated with regard to program, or policy effectiveness. What is less well known, or sometimes ignored, is that policies are carried out in a dynamic context in which factors not under government, or corporate control act independently upon policy outcomes. Under such circumstances, policy baselines are not well-described by fixed rates, or simple trends, but may be more like "moving targets".

Consequently, either baseline, or policy objective needs to be adjusted to compensate for departures from baseline values. *Barring such adjustments*, erroneous conclusion will be derived about policy targets, or their attainment. For instance, during boom times, public health and traffic safety programs may appear largely ineffective while during recessions they may appear spectacularly effective, when, indeed, neither conclusion is justified.

The statistical adjustment method proposed in this presentation can be extended to more input variables, although the resulting scenarios will get more complex. It is recommended that only the most important (best predicting) factors are selected for adjustment.

The adjustment weights (coefficients) can be derived from multivariate methods other than transfer function analysis (i.e. pooled, cross-sectional time series analysis, vector autoregressive models, state space models). A meta-analysis of the pertinent literature is recommended when many sources exist. Particular attention ought to be paid to trends when longterm objectives are of interest. Deterministic trends need to be carefully examined and, if present, incorporated in the adjustment.

A limitation of the simple sensitivity analysis proposed in this paper is that it is not helpful in answering questions such as: What is the probability of a reduction of 2 motor vehicle deaths per 100,000 population with or without government action by the year 2000? To answer such questions, more complex multivariate simulation methods ought to be employed.

Conclusion

Quantitative policy objectives ought to be adjusted for changes in major policy relevant factors. The common practice to assume static conditions, or to adjust only for "recent trends," appears inadequate and may result in misleading inferences about the assessment and attainment of policy objectives.

[16] Expressed in per cent, the trend adjustment for the period 1987-2000 was calculated by the formula $((1-.035)^{13}-1)*100=-37.1$. To simplify the presentation, forecast errors are ignored.

[17] The standard error associated with the year 2000 forecast is approximately $\pm .23$ in ln units; or ranging from 1.2 to 1.9/100,000,000 VMT.

Table 3

ESTIMATED PER CENT CHANGE IN MOTOR VEHICLE DEATH RATES (PER 100,000,000 VEHICLE MILES) DUE TO CHANGE IN UNEMPLOYMENT RATES, PER CAPITA ETHANOL CONSUMPTION AND LONGTERM TREND, 1987-2000

UNEMPLOYMENT RATE	PER CAPITA ETHANOL CONSUMPTION (gal.)			
	2.00 (-.54)	2.25 (-.29)	2.50 (-.04)	2.75 (.21)
3% (-3.2)	-44.4	-38.0	-31.6	-25.2
4% (-2.2)	-46.5	-40.1	-33.7	-27.3
5% (-1.2)	-48.5	-42.1	-35.7	-29.3
6% (-.2)	-50.5	-44.1	-37.7	-31.3
7% (.8)	-52.5	-46.1	-39.7	-33.4
8% (1.8)	-54.6	-48.2	-41.8	-35.4
9% (2.8)	-56.6	-50.2	-43.8	-37.4
10% (3.8)	-58.6	-52.2	-45.8	-39.4

Numbers in parentheses indicate difference from 1987 baseline levels: per capita ethanol consumption=2.54 gal.; unemployment rate=6.2%.

References

Insurance Institute for Highway Safety. 1992. *Status Report.* Arlington, VA; Insurance Institute for Highway Safety. 27(6):5.

U.S. Department of Health and Human Services, Public Health Service. 1992. *Health People 2000: National Health Promotion and Disease Prevention Objectives.* Boston, MA; Jones and Bartlett Publishers.

U.S. Department of Transportation, Federal Highway Administration. 1991. *Highway Statistics* (various years). Washington, DC; U.S. Government Printing Office. Table FI-1.

National Institute on Alcohol Abuse and Alcoholism, Division of Biometry and Epidemiology, Alcohol Epidemiologic Data System. 1991. *Apparent Per Capita Alcohol Consumption: National, State, and Regional Trends: 1977-1988.* Surveillance Report #16. Washington, DC; CSR, Inc. Table 1.

Department of Labor, Bureau of Labor Statistics. 1992. *Employment and Earnings* 39 (1) : 162.

Partyka, S.C. 1984. Simple models of fatality trends using employment and population data. *Accident Analysis and Prevention* 16:211-222.

Wagenaar, A.C., Streff F.M. 1989. Macroeconomic conditions and alcohol-impaired driving. *J. of Studies on Alcohol* 50(3):217-225.

Evans, W.N., Neville, D., Graham, J.D. 1991. General deterrence of drunk driving: evaluation of recent american policies. *Risk Analysis* 11(2):279-289.

Muller, A. 1989. Business recession, alcohol consumption, drinking and driving laws: impact on Oklahoma motor vehicle fatalities and fatal crashes. *American Journal of Public Health* 79(10):1366-1370.

Leigh J.P., Waldon, H.M. 1991. Unemployment and highway fatalities. *J. of Health Politics, Policy and Law* 16(1):135-156.

Pinkyck, R.S., Rubinfeld, D.L. 1991. *Econometric Models and Economic Forecasts* (3rd ed.). New York, NY; Macmillan.

Pankratz, A. (1991) *Forecasting with Dynamic Regression Models.* New York, NY; John Wiley and Sons. 177-184.

Doan, T.A. 1992. *Rats 4.0.* Evanston, IL; Estima.

McCleary, R., Hay, R.A. 1980. *Applied Time Series Analysis for the Social Sciences.* Beverly Hills, CA; Sage Publ. chpt. 4.

Latin Hypercube Sampling for Sensitivity and Uncertainty Analysis of Energy Forecasting Model

Ramesh Dandekar and Nancy Kirkendall, Energy Information Administration

Abstract

The Energy Information Administration (EIA) of the Department of Energy uses a wide collection of models for a mid-term forecast of U.S. energy production, supply, distribution, and consumption. The models are mostly deterministic and provide point forecasts for several times in the future. This paper addresses model forecast accuracy as a function of the accuracy of the model inputs. The use of Latin Hypercube sampling (LHS) technique for sensitivity and uncertainty analysis is demonstrated on a coal supply/demand model. The LHS technique provides the information needed for evaluating the impact of all model inputs and their interactions on the forecasts with a minimum number of computer runs. The results are compared with the commonly used "one-variable-at-a-time" sensitivity analysis.

Introduction

The Energy Information Administration (EIA) of the Department of Energy uses deterministic models for a mid-term energy forecast. These models are used to determine the "expected" response of the national (and international) energy system to "most likely" values of input parameters. As a result, the major focus of EIA's mid-term forecasting analysis is within a few percentiles of the middle of the distribution of input parameters. The accuracies of the point forecasts from these models depend on the accuracy of the model inputs. This paper addresses model forecast accuracy as a function of the accuracy of the model inputs. Only selected model inputs were varied. The results were examined, and observations were made concerning unusual features of the model. Some of these features may be well known to modelers, and easily explained. Others may illustrate potential problems with the model results.

The study also serves to demonstrate the applicability of Latin Hypercube sampling (LHS) technique to conduct sensitivity and uncertainty analysis of EIA's modeling system. LHS was originally proposed by McKay et. al. (1979). Correlations among the input variables was incorporated by using a method described in Iman and Conover (1982). For this first application the input variables were specified as being uncorrelated.

Assumptions

The LHS technique requires specifying the statistical distribution functions of all input variables. To implement a range around the "most likely" value for input variables, we chose to use a uniform distribution to characterizes all input variables. This should be conservative as it includes more observations in the tails of distribution.

In deciding on the range of the uniform distribution, our objective was to have a wide enough range to minimize the possibility of excluding a realistic combination of input variables from the analysis. A range of plus or minus 25 percent around the base values for all input variables was judged to be adequate for this purpose. The study assumes that the input variables are uncorrelated.

Candidate Model

This study used two of EIA's coal models. The first model, the Resource Allocation and Mine Costing (RAMC) model, generates coal supply curves using input geological and financial parameters. The second model, the Coal Supply and Transportation Model (CSTM), uses the coal supply curves from the RAMC model and coal demand forecasts from the Intermediate Future Forecasting System (IFFS) to produce coal production and price forecasts. The supply/demand linkage is done through a surface transportation network solution algorithm imbedded in the CSTM model.

The RAMC model consists of two sub-modules: the geological submodule and the financial submodule. In the geological submodule, the available coal reserves by coal type and coal supply region are used to create hypothetical mining operations with their own operating characteristics. In the financial submodule, the costing of these mines is done to determine their minimum acceptable coal prices. These two modules operate independently within the RAMC model. As a result, the decision was made to analyze the two modules separately using two independent LHS experiments. Due to the space limitations, the discussion in this paper is limited to the analysis of the financial sub-module.

The CSTM model connects demand from 47 coal demand regions (43 domestic regions and 4 export regions) from the IFFS modeling system with 32 coal supply regions in the RAMC model. The supply-demand linkage is done through a transportation network consisting of over 200 nodes and 700 links. These nodes and linkages include both rail and water transport modes. To satisfy input coal demands by region, the least cost coal supply quantities and prices are determined by use of a least cost optimization algorithm imbedded in the CSTM model.

As a part of this demonstration project, the coal demand coming from the IFFS modeling system was kept constant at its base level. Similarly, the transportation network from the base forecast was used without any changes throughout the project. Such a configuration allowed us to study the effects of changes in the coal supply side input variables on the changes in the regional coal outputs.

The effects of nine financial input parameters from the RAMC coal supply model were evaluated for the forecast year 2000.

Input Financial Variables

The financial input variables to the RAMC model belong to two broad categories. The variables belonging to the first category are related to coal extraction costs. These costs are strictly based on engineering considerations of a mining operation. Variables belonging to this category includes: capital costs (initial and deferred), labor costs, annual operation and maintenance (O&M) costs, utility costs (such as water, electricity and other fuels), and costs of other miscellaneous supplies. In addition, the engineering cost components include estimates for escalation rates, utility discount rates and rate of return on investment. In the RAMC model these costs are estimated either on a judgmental basis or are derived using a detailed engineering analysis of a hypothetical mining operation. This study focuses on the sensitivities of this category of financial variables.

The second category of financial input variables consists of social and welfare costs imposed by local and Federal governments. These costs are usually imposed as taxes, royalties, fees and insurance payments. In EIA forecasting models, including the RAMC model, these costs have been implemented at their current levels. This sensitivity analysis ignores the effects of changes in this category of financial variables.

Overview of Latin Hypercube Design

As a part of the experiment, fifty combined RAMC/CSTM computer model runs were made to estimate the effects of changes in the nine coal supply side financial parameters. The choice of making fifty computer runs was strictly arbitrary. All the nine financial parameters were changed simultaneously over the range +-25% of their base values in these fifty computer runs. LHS was used to identify the sample points in the multivariate input space. These financial parameters were assumed to be independent. A method developed by Iman and Conover was used to induce a desired rank correlation of zero among the 9 financial variables.

Output From the Experiment

The RAMC uses 32 coal supply regions. These 32 basic coal supply regions are further aggregated into eleven major coal supply regions. The demand component of the model consists of 43 domestic demand regions and four overseas demand regions. Coal production and price forecasts from the CSTM model provides information on the coal production for each coal supply region by coal mining method (surface, deep and total).

Each of the fifty model runs gave different values for the regional forecasts. The values from these fifty computer runs were used to calculate percent change from the corresponding base forecast values. The mean and standard deviation of the percent change values were computed for each variable in each region. Additionally, the empirical CDF was calculated along with 5, 50, and 95 percentile values. Although all input variables followed

uniform distributions, the output histograms for all output variables generally showed a mound shaped distribution with varying degrees of skewness. This indicates that the output variables are not dominated by changes in only one input variable. The summary statistics from the experiment for total of surface and deep mining activity is presented in Table 1. Similar tabular outputs, in the same format as table 1, were generated separately for surface and deep mining methods, but are not included here due to space limitations.

Summary Statistics

The first column in Table 1 shows the base forecast values in millions of tons. The remaining five columns provide the summary statistics obtained from the fifty computer runs. These statistics are measured as a percent deviation from the base value in the first column. Included are the mean, the standard deviation, the 5 percentile value, the median, and the 95 percentile value.

The mean and the median indicates the extent of possible shifts from the base case forecast. Positive values indicate that the forecast values tend to be larger than the base case, whereas negative values indicate that forecasts tend to be smaller.

The standard deviation (third column) is a measure of the dispersion of forecasts. The relatively large standard deviations in these tables indicate relatively high uncertainty in the forecasts.

The 5 and 95 percentile values provide an empirical 90 percent confidence interval for the forecast values. These percentiles can also be used to evaluate the symmetry of the distribution of forecasts.

The output from the fifty computer runs were also used to compute correlations between each of the model output variables. The variables showing very high correlation coefficients, both positive and negative (exceeding 0.80 in absolute value) are identified in Table 2.

A positive correlation indicates that the production in the two regions moves together, either up or down, showing that they react to changes in the input variables in the same way. A negative correlation indicates that production in the two regions moves in opposite directions in reaction to changes in the input variables. This could be the result of competitive markets for coal as implemented in the model.

Changes in Production Pattern

Confidence Intervals

As a result of changes in the financial variables, 90 percent confidence intervals for forecasts of surface production at the U.S. level were -4.75 percent to 3.59 percent; for deep production were -4.52 percent to 6.14 percent; and for total production were -.33 percent to .22 percent. The small impact on the U. S. Total production is expected as total demand was not changed. For finer levels of geographic detail, the 90 percent confidence intervals are much wider.

No Impact of Input Variables

There were four regions which had non-zero total production in the base case scenario, but which showed no response to changes in the financial variables. These were Shreveport, LA (deep mines), and Corsicana, TX, Winslow AZ and Healy AK (surface mines). It therefore appears that in the RAMC/CSTM model, coal produced from these four supply regions does not compete with coal from other regions. Based on the explanation by EIA coal industry specialist, the production for these four supply regions is dominantly lignite. Lignite is seldom transported more than a few miles from the mine.

High Variation

Regions highlighted below as having "high variation" are those which had base case production in excess of 10 million tons, and a standard deviation of greater than 10 percent. Those that had high variation in response to changes in financial variables include Cambridge OH, Clinton TN, Cordova AL and Carbondale CO. Based on the explanation by the coal industry specialist, the potential instabilities of coal production in Cambridge OH, Cordova AL and Carbondale CO are well known to coal supply modelers. They stem from model reliance on the assumption that consumer coal choices are dominated by minimization of short term delivered coal price.

TABLE 1:

SUMMARY STATISTICS FROM FIFTY LHS MODEL RUNS

SUPPLY REGION	(MM TONS) BASE VALUE	<- PERCENT DEVIATION FROM BASE VALUE -> MEAN	STD DEV	5 PERC	MEDIAN	95 PERC
PA JOHNSTOWN PA	61.27	0.51	6.24	-7.89	-0.55	11.18
OH CAMBRIDGE OH	20.45	0.23	16.47	-23.29	1.72	28.85
MD LONACONING MD	8.63	-0.36	8.11	-14.34	-0.60	12.85
NV CLARKSBURG WV	53.86	-1.21	2.92	-4.80	-0.68	1.06
SV CHARLESTON WV	149.82	-0.09	1.56	-2.20	-0.33	3.37
VA APPALACHIA VA	55.07	0.57	2.12	-2.20	0.30	4.26
EK HAZARD KY	143.41	-0.61	3.85	-7.44	-0.54	5.26
TN CLINTON TN	17.50	0.23	10.30	-18.05	1.70	16.65
AL CORDOVA AL	28.07	-1.63	16.11	-31.33	0.62	24.33
WK CENTRAL CITY KY	39.61	1.80	3.49	-4.26	1.20	8.35
IL CENTRALIA IL	56.90	-2.84	3.56	-9.17	-1.65	1.38
IN HUNTINGBURG IN	27.73	0.49	3.81	-4.83	0.05	5.94
IA OTTUMWA IA	3.50	0.27	1.53	-0.03	0.00	1.45
MO CLINTON MO	0.47	1.79	23.42	-28.06	-1.58	36.29
KS PITTSBURG KS	0.19	3.30	76.66	-100.00	9.79	106.70
AR RUSSELVILLE AR	0.84	6.83	17.29	-18.73	7.98	41.65
LA SHREVEPORT LA	4.11	0.00	0.00	0.00	0.00	0.00
OK TULSA OK	4.68	-6.29	17.32	-40.72	-3.41	18.64
TX CORSICANA TX	73.18	0.00	0.00	0.00	0.00	0.00
ND WILTON ND	30.05	-0.05	0.14	-0.44	0.00	0.00
SD LEMMON SD	0.00	0.02	0.04	0.00	0.00	0.13
EM SIDNEY MT	0.00	0.00	0.00	0.00	0.00	0.00
WM BILLINGS MT	51.42	0.22	1.41	-1.35	-0.30	2.71
NW GILLETTE WY	187.59	-0.29	0.30	-0.84	-0.23	0.10
SW ROCK SPRINGS WY	27.05	2.07	2.36	-1.26	1.98	6.25
CN HAYDEN CO	0.00	0.00	0.00	0.00	0.00	0.00
CS CARBONDALE CO	18.45	14.87	22.21	-15.60	4.03	52.54
UT SUNNYSIDE UT	30.12	-0.36	6.51	-9.49	-1.61	10.89
AZ WINSLOW AZ	10.65	0.00	0.00	0.00	0.00	0.00
NM GALLUP NM	23.23	-1.71	3.51	-7.12	-0.24	1.94
WA CENTRALIA WA	1.81	-2.70	9.82	-20.22	0.25	11.67
AK HEALY AK	2.01	0.00	0.00	0.00	0.00	0.00
U. S. Total	1131.65	-0.01	0.15	-0.33	0.00	0.22
NORTHERN APPALACHIA	144.21	-0.22	1.10	-1.65	-0.11	1.38
CENTRAL APPALACHIA	365.80	-0.18	2.34	-3.98	0.06	4.15
SOUTHERN APPALACHIA	28.07	-1.63	16.11	-31.33	0.62	24.33
MIDWEST	124.24	-0.62	1.54	-3.45	-0.36	1.52
WEST INTERIOR	86.98	-0.25	0.83	-1.70	-0.17	0.95
NORTH GREAT PLAINS	296.10	0.04	0.19	-0.15	-0.01	0.46
ROCKIES & SOUTHWEST	82.44	2.71	6.13	-6.33	0.67	13.31
NORTHWEST & ALASKA	3.82	-1.28	4.65	-9.57	0.10	5.51
APPALACHIA	538.07	-0.27	0.71	-1.58	-0.17	0.76
INTERIOR	211.21	-0.46	0.70	-1.72	-0.33	0.64
WEST	382.37	0.61	1.20	-1.06	0.14	2.85
EAST OF MISS. RIVER	662.31	-0.33	0.38	-1.06	-0.14	0.08
WEST OF MISS. RIVER	469.34	0.45	0.85	-0.68	0.08	2.01

TABLE 2

CORRELATION COEFFICIENTS AMONG SUPPLY REGIONS

SUPPLY REGION and MINE TYPE	Cor Coef	SUPPLY REGION and MINE TYPE
NORTHERN APPALACHIA SURFACE	-0.961	NORTHERN APPALACHIA DEEP
CENTRAL APPALACHIA DEEP	-0.930	NORTHERN APPALACHIA SURFACE
SOUTHERN APPALACHIA SURFACE	-0.871	NORTHERN APPALACHIA DEEP
SOUTHERN APPALACHIA SURFACE	-0.917	CENTRAL APPALACHIA DEEP
SOUTHERN APPALACHIA TOTAL	-0.943	CENTRAL APPALACHIA TOTAL
MIDWEST SURFACE	-0.800	NORTHERN APPALACHIA DEEP
WEST INTERIOR DEEP	-0.869	NORTHERN APPALACHIA SURFACE
WEST INTERIOR SURFACE	-0.892	NORTHERN APPALACHIA DEEP
WEST INTERIOR SURFACE	-0.887	CENTRAL APPALACHIA DEEP
WEST INTERIOR SURFACE	-0.832	WEST INTERIOR DEEP
WEST INTERIOR TOTAL	-0.811	SOUTHERN APPALACHIA TOTAL
NORTH GREAT PLAINS DEEP	-0.918	NORTHERN APPALACHIA SURFACE
NORTH GREAT PLAINS DEEP	-0.822	SOUTHERN APPALACHIA SURFACE
NORTH GREAT PLAINS DEEP	-0.846	WEST INTERIOR SURFACE
ROCKIES & SOUTHWEST TOTAL	-0.927	CENTRAL APPALACHIA TOTAL
ROCKIES & SOUTHWEST TOTAL	-0.887	WEST INTERIOR TOTAL
NORTHWEST & ALASKA DEEP	-0.864	SOUTHERN APPALACHIA DEEP
NORTHWEST & ALASKA DEEP	-0.890	ROCKIES & SOUTHWEST DEEP
NORTHWEST & ALASKA TOTAL	-0.902	SOUTHERN APPALACHIA TOTAL
NORTHWEST & ALASKA TOTAL	-0.929	ROCKIES & SOUTHWEST TOTAL
APPALACHIA SURFACE	-0.966	APPALACHIA DEEP
INTERIOR SURFACE	-0.853	APPALACHIA DEEP
INTERIOR SURFACE	-0.815	INTERIOR DEEP
WEST TOTAL	-0.959	APPALACHIA TOTAL
EAST OF MISS. RIVER SURFACE	-0.986	EAST OF MISS. RIVER DEEP
WEST OF MISS. RIVER TOTAL	-0.957	EAST OF MISS. RIVER TOTAL
WEST INTERIOR DEEP	0.816	CENTRAL APPALACHIA DEEP
WEST INTERIOR DEEP	0.818	NORTHERN APPALACHIA DEEP
ROCKIES & SOUTHWEST TOTAL	0.824	SOUTHERN APPALACHIA TOTAL
SOUTHERN APPALACHIA SURFACE	0.850	NORTHERN APPALACHIA SURFACE
WEST INTERIOR SURFACE	0.851	MIDWEST SURFACE
INTERIOR SURFACE	0.854	APPALACHIA SURFACE
NORTH GREAT PLAINS DEEP	0.864	WEST INTERIOR DEEP
WEST INTERIOR TOTAL	0.871	CENTRAL APPALACHIA TOTAL
WEST INTERIOR SURFACE	0.874	SOUTHERN APPALACHIA SURFACE
NORTH GREAT PLAINS DEEP	0.900	CENTRAL APPALACHIA DEEP
CENTRAL APPALACHIA DEEP	0.922	NORTHERN APPALACHIA DEEP
NORTH GREAT PLAINS DEE	0.925	NORTHERN APPALACHIA DEEP
WEST INTERIOR SURFACE	0.932	NORTHERN APPALACHIA SURFACE
NORTHWEST & ALASKA TOTAL	0.933	CENTRAL APPALACHIA TOTAL
NORTHWEST & ALASKA DEEP	0.935	WEST INTERIOR DEEP
NORTHWEST & ALASKA TOTAL	0.936	WEST INTERIOR TOTAL
ROCKIES & SOUTHWEST DEEP	0.948	SOUTHERN APPALACHIA DEEP

All but Clinton, TN also showed high variation in response to financial variables for deep mines. Cordova, AL showed high variation in response to financial variables for surface mines. Clarksberg WV, and the region Northern Appalachia also had high variation for surface mines in response to the financial variables.

The Southern Appalachia region is solely made up of Cordova, AL. As a result it too shows excessive variation in response to financial parameters. The Southern Appalachian region has a smaller base value and larger standard deviation than the seven other coal supply regions in the RAMC. Thus, the rationale for identifying Southern Appalachia as a separate supply region needs to be re-evaluated.

Change in Mean Forecast of Production

A simple statistical test of the mean percent change yields some indication of those regions where the symmetrical changes in the input variables tend to favor coal production in a region (a positive mean indicates increased production). The test indicates that the mean is significantly different from zero (at the 95 percent level) if it is larger in absolute value than the population standard deviation divided by 3.5, (roughly 2 divided by the square root of 50).

Table 1 indicates that the changes in the financial variables cause slight, but significant, increases in production West of the Mississippi, and corresponding decreases East of the Mississippi. Table 2 shows that the correlation between production values in these regions is -.96 for the financial variables.

The increase in the West occurs primarily in the Rockies and Southwest (Table 1). This comes from deep mines. The largest single contributor is Carbondale CO, which shows a large increase in total production of 15 percent and in deep production of 20 percent in response to changes in the financial variables.

Part of the decrease in production East of the Mississippi occurs in the Midwest and is due to a decrease in deep production. This occurs in response to changes in financial variables. Also in the Midwest there is a smaller, but still significant, increase in surface production due to changes in financial variables.

In response to financial variables there is a decrease in production from surface mines in Northern Appalachia and an increase in production from deep mines. The correlation between Northern Appalachian deep and surface mines was -.96 for the financial experiment.

Observations Based on Correlations

Table 2 shows correlations from the financial experiment. It shows that Southern Appalachian total coal production and production from the Rockies and Southwest (which are positively correlated with each other) are strongly negatively correlated with production in three different regions: Central Appalachia, West Interior, and Northwest and Alaska.

Based on the coal industry specialist, the inverse relationships with Central Appalachia and West Interior are to be expected, since these are the nearest competitive regions providing low sulfur and high sulfur steam coals respectively. Central Appalachia also competes with Southern Appalachia (Alabama) as a source of metallurgical coals in both the domestic and export markets.

The high negative correlation between Southern Appalachia and Northwest & Alaska is more difficult to explain, as these regions' products do not compete in any market. Production in these three regions also shows pairwise positive correlation. These correlation patterns could illustrate the effects of competition. The true potential for such shifts in the national coal market has been of interest to coal producers and consumers since mid-1970's. Competition between these sources has been sharpened by regulatory incentives for the use of low sulfur bituminous coal.

Southern Appalachia deep coal, which is solely made up of Alabama deep coal, shows strong positive correlation between deep production in the Rockies and Southwest. Both regions have strong negative correlation with deep production in the Northwest and Alaska. This indicates that deep coal produced from the Rockies and the Southwest in combination with Southern Appalachia deep coal could compete with deep coal produced from Northwest and Alaska. At the same time deep production in the Northwest and Alaska is positively correlated with deep production in West Interior, Central and Northern Appalachia and the Northern Great Planes.

Surface production in Southern and Northern Appalachia, West Interior and the Midwest are pairwise positively correlated. They are negatively correlated with deep production from West Interior, Central and Northern Appalachia.

Traditional Sensitivity Analysis Using Data From LHS

The analysis so far has been restricted to evaluation of summary statistics based on the model output. The next step was to quantify the average contribution of each of the input variables to the average model output uncertainty. Such an evaluation provides valuable insight into the relative importance of each input variable in determining model outcome.

To accomplish this objective, a linear regression model of the form

$$\frac{(Y-Y_b)}{Y_b} = \sum_{j=1}^{m} \alpha_j \frac{(X_j - X_{bj})}{X_{bj}}$$

was estimated. Where, Y = Model output variable, X_j = j^{th} Model input variable, Y_b = Base Forecast of the model output variable, X_{bj} = Base value of the jth model input variable and m = the total number of model input variables selected for the analysis. In this regression model all the variable values are expressed as a relative change from their base values. This regression model assumes that the cumulative effect of changes in all the input variables on the model output is linear. The regression coefficients and related R^2 values from these analyses are presented in Table 3. In the table the "*" next to the regression coefficient indicates that the related t statistic is less than 2 and therefore the regression coefficient is not significantly different from zero. Similar tables were generated for each coal mining methods and are used in the analysis that follows.

In table 3, a positive value for the regression coefficient indicates that the relative change in the output variable is in the same direction as the relative change in the input variable. The larger the value of the regression coefficient, the more influential is the input variable in determining the model outcome. The input variables with larger regression coefficients should be carefully evaluated. The R^2 value is located in the first column of the table. It is a measure of fraction of the output variation which is explained by the variation in the input variables. The R^2 value shown here is the "adjusted R^2" as this problem is stated as a regression through the origin.

Regression Fit

Tables 3 shows the results of regression analysis applied to total mines by supply region using data from the financial experiment. Similar tables were also generated for surface and deep mining methods but are not included in this paper.

For the U. S. total production (Table 3) all financial variables were statistically significant except for utility discount rate. However, only labor rate, labor productivity and escalation factors appear to be important in determining the relative changes in total coal production. For deep and surface mines at the National level the three important variables are labor rate, labor productivity and annual supplies & parts cost.

The regression coefficients for deep and surface mines at the National level are generally of opposite sign. Thus for example, an increase in labor rates tends to result in a decrease in deep production and an increase in surface production. Increases in labor productivity and annual supplies & parts cost, however, result in an increase in deep production and a decrease in surface production.

For most regions, it appears that there is little impact on production of changes in deferred capital costs, water and power costs, utility discount rates and industrial ROR. For these variables the estimated coefficients tend not to be statistically significant. Initial capital costs tend to be significant for only surface coal production in the West and West of the Mississippi river. The sign in those regions is negative, indicating that an increase in initial capital costs tends to result in a decrease in surface production.

The R^2 statistics shown in table 15 vary widely from one supply region to other. This shows that the linear model is good in some regions and poor in others.

Table 3 - REGRESSION COEFFICIENTS - Latin Hypercube Design - Total Mines

SUPPLY REGION	RSQUARE	INI CAP	DEF CAP	LABR RATE	LABR PROD	W&S COST	OTHR SUP	ESCA FACT	UT DIS RT	IND. ROR
PA JOHNSTOWN PA	0.912	0.030*	0.041	0.176	-0.259	0.014*	0.158	0.224	0.033*	0.039
OH CAMBRIDGE OH	0.904	-0.198	-0.040*	-0.501	0.631	-0.117	-0.381	-0.598	-0.065*	-0.136
MD LONACONING MD	0.978	-0.144	-0.011*	-0.271	0.338	-0.040	-0.204	-0.252	-0.006*	-0.068
NV CLARKSBURG WV	0.139	0.064	-0.004*	0.069	-0.047*	-0.050*	0.044*	0.061	-0.006*	-0.006*
SV CHARLESTON WV	0.000	0.014*	0.009*	0.003*	0.006*	0.014*	-0.019*	0.001*	-0.027*	-0.003*
VA APPALACHIA VA	0.816	-0.029	-0.014*	-0.093	0.080	0.012*	-0.028	-0.057	0.005*	-0.006*
EK HAZARD KY	0.940	-0.034	-0.019*	-0.153	0.179	-0.018*	-0.029	-0.114	0.006*	-0.015*
TN CLINTON TN	0.941	-0.013*	-0.069	-0.372	0.485	-0.017*	-0.192	-0.269	-0.052	-0.016*
AL CORDOVA AL	0.925	0.064*	0.042*	0.715	-0.712	0.117	-0.052*	0.413	0.036*	0.034*
WK CENTRAL CITY KY	0.000	0.011*	-0.007*	0.080	-0.114	-0.010*	0.053*	0.011*	0.024*	-0.028*
IL CENTRALIA IL	0.000	0.077	-0.005*	0.089	-0.093	0.009*	0.126	0.100	0.013*	0.016*
IN HUNTINGBURG IN	0.882	-0.088	-0.005*	-0.093	0.105	0.013*	-0.154	-0.117	0.002*	-0.024*
IA OTTUMWA IA	0.043	-0.023*	-0.013*	0.021*	0.001*	-0.008*	-0.033	-0.015*	-0.006*	-0.009*
MO CLINTON MO	0.915	-0.298	-0.041*	-0.364	0.466	-0.198	-1.253	-0.650	-0.019*	-0.204
KS PITTSBURG KS	0.936	0.995	0.157*	1.118	-1.406	0.522	4.170	2.229	0.025*	0.694
AR RUSSELVILLE AR	0.677	-0.224	0.184*	0.454	-0.660	0.076*	-0.759	0.219	-0.095*	0.089*
LA SHREVEPORT LA	0.000	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*
OK TULSA OK	0.704	-0.228	-0.079*	-0.540	0.792	-0.117*	-0.107*	-0.551	0.075*	-0.123*
TX CORSICANA TX	0.000	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*
ND WILTON ND	0.193	0.001*	-0.001*	0.005	-0.002*	0.000*	0.003	0.002*	-0.001*	0.001*
SD LEMMON SD	0.193	0.000*	0.000*	-0.002	0.001*	0.000*	-0.001	-0.001*	0.000*	0.000*
EM SIDNEY MT	0.000	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*
WM BILLINGS MT	0.648	-0.021	-0.008*	-0.031	0.036	0.000*	-0.039	-0.054	-0.006*	-0.007*
NW GILLETTE WY	0.000	0.007*	0.000*	0.003*	-0.003*	0.001*	0.004*	0.008	0.001*	-0.001*
SW ROCK SPRINGS WY	0.000	-0.054*	0.042*	0.037*	-0.036*	-0.010*	-0.052*	-0.014*	-0.015*	0.005*
CN HAYDEN CO	0.000	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*
CS CARBONDALE CO	0.339	0.161*	0.151*	0.754	-0.842	0.109*	0.445	0.786	0.092*	0.209*
UT SUNNYSIDE UT	0.864	0.079	-0.031*	0.094	-0.300	0.031*	0.180	0.192	0.012*	0.095
AZ WINSLOW AZ	0.000	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*
NM GALLUP NM	0.529	-0.002*	0.004*	-0.087	0.162	-0.001*	-0.060	-0.113	-0.024*	-0.022*
WA CENTRALIA WA	0.850	-0.124	-0.067*	-0.323	0.481	-0.094	-0.055*	-0.282	-0.015*	-0.094
AK HEALY AK	0.000	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*
U. S. Total	0.956	0.002	0.001	0.006	-0.006	0.001	0.002	0.006	0.000*	0.001
NORTHERN APPALACHI	0.221	0.000*	0.009*	0.013*	-0.018*	-0.032	0.017*	0.018*	0.002*	-0.009*
CENTRAL APPALACHIA	0.938	-0.013	-0.009*	-0.090	0.108	0.000*	-0.032	-0.066	-0.011*	-0.009*
SOUTHERN APPALACHI	0.925	0.064*	0.042*	0.715	-0.712	0.117	-0.052*	0.413	0.036*	0.034*
MIDWEST	0.449	0.019*	-0.006*	0.045	-0.056	0.004*	0.040	0.023	0.014*	-0.007*
WEST INTERIOR	0.750	-0.015	-0.003*	-0.023	0.036	-0.006*	-0.012	-0.027	0.003*	-0.006*
NORTH GREAT PLAINS	0.589	-0.004	0.002*	0.000*	0.001*	0.000*	-0.009	-0.005	-0.002*	-0.001*
ROCKIES & SOUTHWES	0.694	0.065*	0.024*	0.178	-0.252	0.036*	0.149	0.214	0.018*	0.075
NORTHWEST & ALASKA	0.849	-0.059	-0.032*	-0.153	0.228	-0.045	-0.026*	-0.134	-0.007*	-0.045
APPALACHIA	0.719	-0.005*	-0.001*	-0.021	0.031	-0.003*	-0.020	-0.018	-0.005*	-0.007*
INTERIOR	0.000	0.005*	-0.005*	0.017	-0.018	0.000*	0.019	0.003*	0.009*	-0.006*
WEST	0.608	0.010*	0.007*	0.037	-0.051	0.007*	0.025	0.041	0.002*	0.015*
EAST OF MISS. RIVE	0.000	-0.001*	-0.002*	-0.008*	0.015	-0.001*	-0.009	-0.010	-0.001*	-0.007*
WEST OF MISS. RIVE	0.543	0.006*	0.005*	0.026	-0.035	0.005*	0.018	0.028	0.003*	0.011*

Comparison of Results With One-at-a-Time Sample

To compare the results from LHS with those from the method of changing only one input variable at a time, an additional set of model runs were made. The set varied the nine financial input variables over the same range, -25 percent to +25 percent, one-at-a-time. As 50 observations of each variable in a one-at-a-time experiment would have required 900 additional computer runs, a reduced number of observations was used for each variable. A total of 10 observations were made for each variable at 5 percentage point intervals.

The same linear regression models used in the LHS experiments were also used with the one-at-a-time experiment. Results of the regression analysis applied to production by region are given in Table 4.

The findings from Table 4 for the one-at-a-time design are similar to those from Table 3 for the LHS design. In particular, conclusions concerning the important explanatory variables and the signs of the estimated coefficients tend to be the same. The actual values of the estimated coefficients are sometimes similar, and sometimes quite different. This may be due in part to interactions among the variables, and it may be due in part to the difference in sample sizes. The LHS design used a total of 50 runs, with 50 different observed values of each input variable. The one-at-a-time design used a total of 90 runs, with 10 different observed values of each input variable.

Conclusions

- The forecasts for coal supply regions of Alabama, Ohio and South Colorado appear to be highly variable in response to changes in both financial and geological input variables.

- Of the nine financial variables selected for the analysis; labor rates, labor productivity and escalation factors appear to be the three most important variables in determining the level of total coal production at the national level.

- Based on the financial input uncertainties, the national level 90 percent confidence interval for delivered coal prices is approximately plus and minus 5 percent. The demand regions east of the Mississippi have relatively wider 90 percent confidence intervals for price than the western demand regions.

- After labor rates and labor productivity, annual supplies & parts cost is the most important financial input variable to explain changes in surface and deep mine production for all supply regions.

- For the both financial and geological experiments, the LHS design provides results comparable to one-at-a-time sample design. By comparable, we mean that the two procedures identify the same input variables as being important and the signs of the estimated coefficients are the same.

- For both the financial and geological experiments, the values of the coefficients estimated by regression methods applied to the LHS data and to the one-at-a-time sample may be quite different. This is most likely due to interactions and other nonlinear effects, although for the financial variables the difference in the sample size may be a contributing factor.

Table 4 - REGRESSION COEFFICIENTS - Orthogonal Design - Total Mines

SUPPLY REGION	RSQUARE	INI CAP	DEF CAP	LABR RATE	LABR PROD	W&S COST	OTHR SUP	ESCA FACT	UT DIS RT	IND. ROR
PA JOHNSTOWN PA	0.774	0.069	0.017*	0.175	-0.255	0.020*	0.124	0.226	0.012*	0.029*
OH CAMBRIDGE OH	0.927	-0.195	-0.045*	-0.551	0.741	-0.096	-0.258	-0.630	-0.030*	-0.046*
MD LONACONING MD	0.975	-0.184	-0.020	-0.291	0.337	-0.027	-0.175	-0.248	-0.009*	-0.075
NV CLARKSBURG WV	0.000	0.036	0.002*	0.022*	-0.009*	0.005*	0.007*	0.030	0.004*	0.011*
SV CHARLESTON WV	0.000	0.014*	-0.014*	0.019*	0.004*	0.002*	-0.002*	-0.019*	-0.009*	-0.011*
VA APPALACHIA VA	0.794	-0.029	-0.020	-0.062	0.083	-0.006*	-0.049	-0.037	-0.005*	-0.013
EK HAZARD KY	0.951	-0.050	-0.011*	-0.153	0.171	-0.013	-0.024	-0.109	-0.006*	-0.027
TN CLINTON TN	0.935	-0.127	-0.123	-0.469	0.551	-0.060	-0.150	-0.313	-0.051	-0.081
AL CORDOVA AL	0.936	0.029*	0.098	0.769	-0.877	0.045*	-0.037*	0.494	0.037*	0.016*
WK CENTRAL CITY KY	0.000	-0.053*	0.016*	-0.015*	-0.046*	-0.005*	0.013*	-0.041*	0.000*	-0.018*
IL CENTRALIA IL	0.136	0.086	0.017*	0.120	-0.097	-0.006*	0.099	0.098	0.011*	0.038*
IN HUNTINGBURG IN	0.728	-0.069	-0.006*	-0.088	0.090	-0.006*	-0.202	-0.112	-0.008*	-0.031*
IA OTTUMWA IA	0.149	0.000*	0.000*	0.000*	0.000*	0.000*	-0.038	0.000*	0.000*	0.000*
MO CLINTON MO	0.883	-0.301	-0.079*	-0.366	0.474	-0.033*	-1.344	-0.899	-0.038*	-0.135
KS PITTSBURG KS	0.941	1.157	0.219*	1.264	-1.535	0.104*	4.431	2.916	0.190*	0.561
AR RUSSELVILLE AR	0.803	-0.339	-0.014*	0.291	-0.436	-0.002*	-0.991	0.005*	-0.009*	-0.171
LA SHREVEPORT LA	0.000	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*
OK TULSA OK	0.558	-0.210	0.001*	-0.339	0.623	-0.037*	-0.161	-0.166	-0.021*	-0.108*
TX CORSICANA TX	0.000	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*
ND WILTON ND	0.150	0.000*	0.000*	0.002	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*
SD LEMMON SD	0.150	0.000*	0.000*	-0.001	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*
EM SIDNEY MT	0.000	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*
WM BILLINGS MT	0.588	-0.004*	0.000*	-0.057	0.066	0.000*	-0.051	-0.080	-0.002*	-0.001*
NW GILLETTE WY	0.000	0.000*	0.000*	0.009	-0.009*	0.000*	0.009	0.012	0.001*	0.000*
SW ROCK SPRINGS WY	0.026	-0.031*	0.024*	0.032*	-0.006*	0.000*	-0.074	-0.004*	0.000*	-0.015*
CN HAYDEN CO	0.000	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*
CS CARBONDALE CO	0.375	0.594	0.330*	0.915	-1.001	0.236*	0.765	0.990	0.140*	0.407
UT SUNNYSIDE UT	0.723	0.034*	-0.096	0.091	-0.305	-0.006*	0.100	0.169	-0.012*	0.024*
AZ WINSLOW AZ	0.000	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*
NM GALLUP NM	0.760	-0.054	-0.001*	-0.164	0.184	-0.015*	-0.093	-0.177	-0.007*	-0.027*
WA CENTRALIA WA	0.831	-0.130	-0.011*	-0.284	0.456	-0.031*	-0.038*	-0.214	-0.013*	-0.064
AK HEALY AK	0.000	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*
	0.964	0.001	0.001	0.006	-0.006	0.000	0.002	0.005	0.000*	0.001
NORTHERN APPALACHI	0.000	0.004*	0.000*	-0.013*	0.013*	-0.005*	0.008*	0.003*	0.002*	0.005*
CENTRAL APPALACHIA	0.921	-0.025	-0.019	-0.084	0.107	-0.008*	-0.025	-0.071	-0.009*	-0.021
SOUTHERN APPALACHI	0.936	0.029*	0.098	0.769	-0.877	0.045*	-0.037*	0.494	0.037*	0.016*
MIDWEST	0.040	0.007*	0.012*	0.030	-0.039	-0.006*	0.005*	0.007*	0.003*	0.005*
WEST INTERIOR	0.596	-0.014	0.000*	-0.015	0.028	-0.002*	-0.017	-0.007	-0.001*	-0.007
NORTH GREAT PLAINS	0.852	-0.004	0.002	-0.002	0.005	0.000*	-0.010	-0.007	0.000*	-0.002
ROCKIES & SOUTHWES	0.605	0.130	0.039*	0.192	-0.283	0.046*	0.182	0.233	0.025*	0.092
NORTHWEST & ALASKA	0.830	-0.062	-0.005*	-0.134	0.216	-0.014*	-0.018*	-0.101	-0.007*	-0.030
APPALACHIA	0.560	-0.014	-0.008	-0.020	0.031	-0.004*	-0.017	-0.022	-0.004*	-0.012
INTERIOR	0.000	-0.001*	0.007*	0.012*	-0.011*	-0.004*	-0.004*	0.001*	0.002*	0.000*
WEST	0.532	0.025	0.010*	0.039	-0.055	0.010*	0.031	0.044	0.005*	0.018
EAST OF MISS. RIVE	0.117	-0.010	-0.004*	-0.011	0.018	-0.005*	-0.013	-0.016	-0.002*	-0.009
WEST OF MISS. RIVE	0.506	0.018	0.008*	0.029	-0.039	0.008*	0.022	0.034	0.004*	0.014

References

Doctor P.G. (1989) "Sensitivity and Uncertainty Analyses for Performance Assessment Modeling," Engineering Geology, 26(1989): pp. 411-429.

Iman R.L. and Conover W. J. (1982), "A Distribution-Free Approach to Inducing Rank Correlation Among Input Variables", Commun. Stat., B11(3): pp. 311-334.

Iman R. L. and Helton J. C. (1985), "A Comparison of Uncertainty and Sensitivity Analysis Techniques for Computer Models", Sandia National Laboratory.

Iman R. L. and Helton J. C. (1987), "An investigation of Uncertainty and Sensitivity Analysis Techniques for Computer Models," Risk Analysis, 8(1), pp. 71-90.

Iman R. L. and Helton J. C. (1991), "The Repeatability of Uncertainty and Sensitivity Analysis for Complex Probabilistic Risk Assessments," Risk Analysis, 11(4), pp. 591-606.

McKay M.D., Conover W. J. and Beckman, R. J. (1979), "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code", Technometrics 21(2): pp. 239-245.

Morgan M. G., and Henrion M (1991), Uncertainty, A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis, Cambridge University Press.

Short Run Forecasts and Long Run Dynamics: A Case Study Using Nitrogen Fertilizer Prices

Harry Vroomen, Economic Research Service and Frederick Joutz, Energy Information Administration and the George Washington University

The purpose of this paper is to compare alternative model specifications used in short run forecasts when serial correlation is present. The "textbook" correction for first-order serial correlation recommends the Cochrane-Orcutt procedure. However, this approach imposes restrictions on the long-run dynamics which can lead to a misspecified model and suboptimal forecasts. In addition, the textbook model produces biased confidence intervals. We demonstrate how to incorporate the long run dynamics into the short run model.

U.S. fertilizer prices have been highly variable since the mid-1970's and exhibited increased volatility since 1980. The prices of nitrogen materials, in particular, have varied significantly over this period. For example, the retail price of anhydrous ammonia fell 38 percent from May 1984 to October 1986, but rose nearly 29 percent by April 1989. Prices then changed direction again, falling 20 percent by October 1990, before climbing 18 percent by April 1993. Prices of other nitrogen fertilizer materials have followed a similar pattern (Figure 1).

This variability complicates the planning process for fertilizer suppliers and users. The USDA provides fertilizer price forecasts through the Situation and Outlook Program on agricultural inputs. Input manufacturers need to forecast fertilizer prices to plan production levels and decide on contract terms for future delivery. Similarly, accurate forecasts can assist farmers in making informed decisions with respect to crop mix and the timing of fertilizer purchases. Accurate fertilizer price forecasts can help reduce costs, maximize profits, and foster the efficient operation of the market.

This study compares five short-run price forecasting models for the retail prices of anhydrous ammonia (AA), urea (UREA), nitrogen solutions (NS), ammonium nitrate (AN), and ammonium sulfate (AS). These materials account for 97.3 percent of the single-nutrient nitrogen fertilizer used in the United States and 77.4 percent of the total nitrogen applied (Tennessee Valley Authority). Price forecasts are generated for the spring, the peak demand season for fertilizer.

The paper is organized into five sections. First, the model is explained. Then, issues relating to serial correlation and common factor restrictions are discussed. Third, we present the procedure for developing the forecasts. This is followed by the forecast comparisons. The conclusion is the fifth section.

The Model

We begin by assuming that retail prices are approximated by a linear relationship with the wholesale price for AA and transportation costs. AA is the source of nearly all nitrogen fertilizer used in the United States (Andrilenas and Vroomen, 1990). It may be applied directly to the soil or converted into other nitrogen fertilizers, such as UREA, NS, AN, and AS. The retail prices of these nitrogen materials would reflect price changes at the wholesale level if retailers followed a markup pricing scheme. However, retail prices should also be affected by changes in marketing costs not reflected in f.o.b. prices. Transportation costs represent a significant share of the final price a farmer pays for fertilizer, and rail is the predominant method of transporting fertilizer in the United States. Marketing costs are thus represented by the cost of rail transportation. Consequently, the retail price model for each nitrogen material is specified as:

$$RPN_{i,t} = \beta_0 + \beta_1 WHOLAA_t + \beta_2 RAIL_t + \epsilon_t \qquad (1)$$

where:

RPN =	retail price of nitrogen material i
WHOLAA =	wholesale price of AA
RAIL =	total rail freight rate index (Dec. 1984=100)
i =	AA, UREA, NS, AN, and AS, respectively
$\epsilon_{i,t}$ =	a stochastic disturbance term.

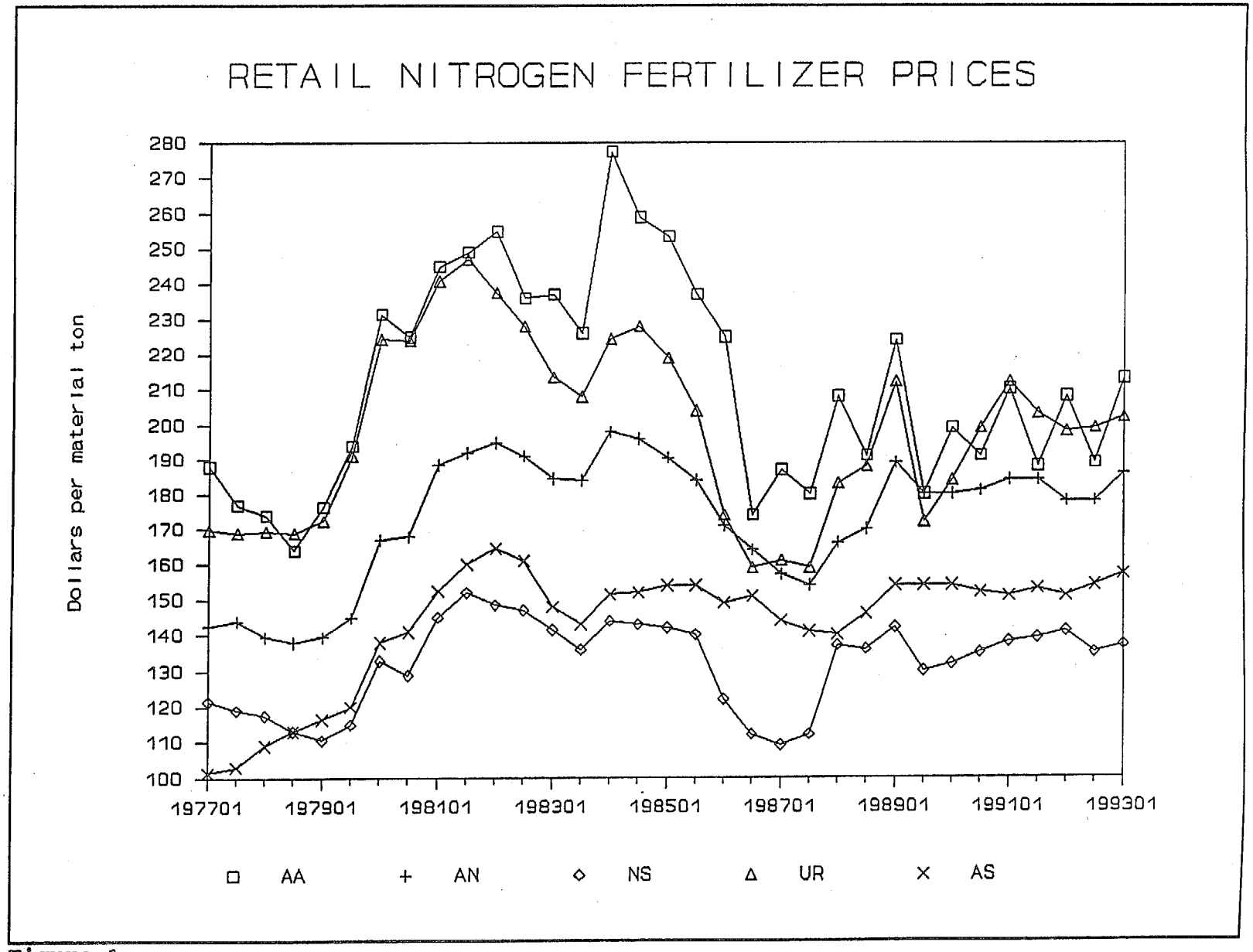

RETAIL NITROGEN FERTILIZER PRICES
Dollars per material ton
280
270
260
250
240
230
220
210
200
190
180
170
160
150
140
130
120
110
100
197701
197901
198101
198301
198501
198701
198901
199101
199301
AA
AN
NS
UR
AS

Figure 1

The data was collected from various issues of the U.S. Department of Agriculture publication, Agricultural Prices. From 1977 to 1985, retail fertilizer prices were reported for March, May, October, and December. Since 1986, however, retail prices have only been available for April and October. To form a continuous data set, March and May retail prices were averaged to construct an April price for years preceding 1986, while reported prices for April were used for subsequent years. Consequently, the retail price equations in (1) were estimated with biannual data (April and October) for 1977-92.

Table 1 contains the summary statistics from the OLS regressions for the retail price equations over the sample period 1977 through 1987.

All of the structural parameters have the hypothesized signs and all but two are significant at the 95-percent level. Four of the five autoregressive parameters are also statistically significant at 5-percent, the fifth is significant at 6-percent. In addition, R^2's indicate that the explanatory variables capture most of the variation in the retail prices of the nitrogen fertilizer materials. Preliminary results indicated that the Durbin-Watson statistic was significant in all the retail price regressions.

Liu (1987) argues that ignoring serial correlation of the residuals can greatly reduce the forecasting performance of a multi-variable model. All equations in (1) were thus estimated with a maximum-likelihood autoregressive technique. Autoregressive techniques use the time series part of a model as well as the systematic part in generating predicted values and so are useful forecasting tools.

Common Factor Restrictions And Serial Correlation Correction

This section discusses the implication of correcting for serial correlation in estimating models and making forecasts. We start by describing three maintained models: 1) a static model, 2) a model with AR(1) errors, and 3) a first order autoregressive distributed lag model, ADL(1,1). The appendix discusses the case where there are lagged endogenous variables in the original model and the AR and ADL models have longer lags. Hendry and Mizon (1978) addressed the issues related to autocorrelation and dynamic specification with a case study on the demand for money in the U.K.

The first two maintained models typify the conventional approach to econometric modeling. Gilbert (1990) refers to this as the average economic regression (AER) view of econometrics. The researcher uses a specification known to "be correct", our equation (1), because it is derived from theory. In the single equation case the static model is specified as:

$$y_t = \beta_0 + x_t\beta + \epsilon_t \tag{2}$$

where x_t is a row vector with k elements and ϵ_t is a random disturbance. The objective is to obtain estimates $b_0,\ldots,b_k$ of the coefficient vector $\beta_0,\ldots,\beta_k$. Several issues, among them, serial correlation, heteroscedasticity, multicollinearity, stationarity vs. differencing, simultaneity, and seasonality confront the applied researcher. These issues present themselves through low Durbin-Watson statistics, insignificant coefficients, wrong signs, extremely large t-values on the lagged dependent variable, etc. The researcher attempts to resolve these problems by respecifying the model and collecting more data.

In the process, variables are added and deleted from the model until the correct signs appear, variables are significant, first order autocorrelation disappears, and the explanatory power measured by $\bar{R}^2$ is relatively high. In the presence of a low Durbin-Watson statistic, the typical correction or respecification is:

$$\begin{aligned} y_t &= \beta_0 + x_t\beta + \epsilon_t \\ \epsilon_t &= \rho_1\epsilon_{t-1} + u_t \quad \textit{where } u_t \sim (0,\sigma_u^2) \textit{ and } |\rho_1|<1 \end{aligned} \tag{3}$$

The Cochrane-Orcutt procedure or maximum likelihood technique with grid search is often used to estimate the parameters. The autocorrelation coefficient is often referred to as a nuisance parameter necessary for efficient estimation of the β'_S. It is used to patch up the original theoretical model, because of "bad" test statistics.

There are three issues to consider in this case. First, the error terms may be autocorrelated of a higher order, AR(p). Further tests can be conducted using the Lagrange multiplier approach to detect the number of lags. Second, the model is not the true one. We shall assume that the relevant variables have not been omitted to avoid

Table 1
OLS Results for Retail Price Equations, 1977-1987

Retail Price	constant	WHOLAA	RAIL	RBARSQ	SSR	DW-statistic
AA	51.34 (16.47)	0.806 (0.098)	0.766 (0.171)	0.84	3486.0	1.54 (0.059)
UREA	82.4 (19.4)	0.794 (0.114)	0.213 (0.202)	0.72	4829.0	1.34 (0.019)
NS	70.14 (10.34)	0.342 (0.061)	0.198 (0.107)	0.66	1374.5	1.07 (0.002)
AN	60.53 (10.58)	0.371 (0.062)	0.742 (0.11)	0.84	1438.4	1.16 (0.005)
AS	43.83 (9.08)	0.139 (0.054)	0.925 (0.094)	0.85	1059.8	0.81 (0.1e-3)

Standard errors are given in parentheses. The exact Durbin-Watson statistic and probability were computed using SHAZAM 7.0.

this problem. Finally, the dynamic specification of the true model is incorrect. Certain variables are omitted which are themselves autocorrelated. This case, which is the focus here, suggests that the true model contains richer dynamics than that originally specified.

Model specification is sometimes referred to as an art form as much as a science. While it is unlikely that the researcher will finally discover the true dynamic specification, in cases where the order of autocorrelation is low, he can detect misspecification of the model. If the AR(1) model in (3) is maintained, it implies that:

$$y_t = \beta_0*(1-\rho_1) + x_t\beta - x_{t-1}\beta\rho_1 + \rho_1 y_{t-1} + \epsilon_t \tag{4}$$

or rewriting this using backshift notation:

$$(1-\rho_1 L)y_t = \beta_0(1-\rho_1 L) + (1-\rho_1 L)x_t\beta + \epsilon_t \tag{5}$$

A more general model is the autoregressive distributed lag ADL(1,1) where the first number in the parentheses is the number of lags of the dependent variable and the second is the maximum lag on the explanatory variables. This model is written as:

$$y_t = \pi_0 + x_t\beta + x_{t-1}\gamma + \rho y_{t-1} + u_t \tag{6}$$

The difference between the ADL(1,1) and the AR(1) models is that the former do not place (nonlinear) restrictions on the coefficients. The AR(1) expression in backshift notation reveals the common factor restrictions.

If the common factors are correct, then there is a direct relationship between the autocorrelation coefficient, the coefficient on the lagged dependent variable, and the coefficients on the lagged explanatory variables. The relationship imposes the restriction(s) $\gamma = -\beta*\rho$. This can be tested for using the procedure suggested by Sargan (1964) which is a Wald test. Unfortunately, there is no exact test like the t-test as in the linear restriction case. In finite samples, LaFontaine and White (1986) find that identical algebraic expressions for the (nonlinear) restrictions do not lead to unique statistical conclusions.

A general approach to testing for the common factor restrictions, or COMFAC analysis, is the asymptotic F-test or the Likelihood Ratio test. In finite samples, with lagged dependent variables and or nonlinearity in the parameters, the test is not exact. The statistic for testing the AR(1) model against the ADL(1,1) model is given by

$$\frac{(SSR_r - SSR_u)/r}{SSR_u/(n-k-r-2)} \sim F(r,n-k-r-2)$$

where the sum of squared residuals from the AR(1) model are SSR_r, and the sum of squared residuals from the ADL(1,1) are SSR_u. The degrees of freedom in the denominator are equal to the total number of observations, n, the number of variables in the static model, k, the number of additional unique parameters in the ADL model, r, one lagged dependent variable and the intercept term. The degrees of freedom for the test, r, refers to the number of identifiable parameters in the ADL(1,1) model. The number of restrictions, r, is not always equal to k. The Likelihood ratio test is

$$LR = \left[\frac{RSS_r}{RSS_u}\right]^{n/r}; \quad where \ -2\ln LR \sim \chi^2_{(r)} .$$

If the null hypothesis is not rejected from either the F-test or Likelihood ratio test, the imposition of the AR(1), common factors restriction, is a convenient restriction and not a nuisance. The model can be estimated more parsimoniously without loss of consistency.

In Table 2 we present the results for the tests of common factors using the LR test and asymptotic F-test. In all five retail price equations, the common factor restrictions from correcting for the first order serial correlation in (2) using the Cochrane-Orcutt procedure is rejected. The alternative model is an ADL(1,1).

We simplified the alternative model by deleting insignificant variables. The results are given in Table 3. The lagged dependent variable provides valuable explanatory power in every equation. The transportation cost variable, RAIL, appears to have been masking the dynamics of the lagged dependent variable. It is only significant in the UREA price equation. The lagged wholesale price, LWHOLAA, only appears in the AA equation. First order

Table 2
Test for Common Factors Restriction, 1977-87

Retail Price	AR(1) RSS	ADL(1,1) RSS	F-statistic	L.R. statistic
AA	3247.0	1401.3	18.54 (0.68e-4)	17.647 (0.15e-3)
UREA	2225.0	654.97	27.18 (0.72e-5)	25.681 (0.26e-5)
NS	805.58	286.48	22.50 (0.22e-4)	21.712 (0.19e-4)
AN	845.86	296.72	22.81 (0.21e-4)	21.999 (0.17e-4)
AS	583.34	366.33	12.74 (0.49e-3)	9.7701 (0.76e-2)

The p-value is given in parentheses.

Table 3
Autoregressive Distributed Lag Models for 1977-1987

Retail Prices	constant	WHOLAA	LWHOLAA	RAIL	LRAIL	LRP	RBARSQ	RSS	Durbin's h
AA_1	34.24 (13.70)	0.575 (0.079)	0.393 (0.087)			0.287 (0.066)	0.93	1240.2	-1.40
UREA	37.84 (11.49)	0.476 (0.067)		-0.276 (0.107)		0.633 (0.079)	0.95	855.9	0.21
NS	20.36 (8.92)	0.236 (0.035)				0.615 (0.078)	0.91	346.02	-0.98
AN	17.78 (7.76)	0.262 (0.0345)				0.707 (0.051)	0.96	349.56	-1.51
AS	17.05 (8.242)	0.119 (0.038)				0.785 (0.061)	0.92	472.17	0.34

1. The lagged retail price for AA enters with a "2-period" lag - one year instead of six months.

serial correlation is not present according to the Durbin's h statistic.

Since we will be constructing a set of rolling forecasts, we reestimated the model for each year. The models continued to reject the common factor restrictions based on an AR(1) specification for the error term. The ADL regression results for the sample period 1977-92 are presented in Table 4 as a simple test for model stability. The coefficients do not change much from the model estimated through 1977, however the model fit appears to decline slightly as all the $\bar{R}^{2\prime}s$ are lower.

Developing Model Forecasts

Retail price forecasts for the selected nitrogen materials are generated for the Spring, the peak demand season for fertilizer. Forecasts are made for each April from 1988-93. Out-of-sample forecasts are generated 6 months ahead at a time. For example, retail price forecasts for April 1988 were computed from model coefficients estimated with data through October 1987. This rolling horizon procedure was followed six times for each set of equations as the time period for each was sequentially updated until the final set of price forecasts for April 1993 was developed from equations estimated through October 1992. Sequentially updated forecasting incorporates new information in parameter estimates and is the efficient way to use this model.

In order to develop forecasts from the estimated retail price equations, we first need values for the explanatory variables appearing in these equations. The OLS, AR(1), and ADL models all require forecasts of $WPAA_t$ and $RAIL_t$. Forecasts of these variables are developed from autoregressive-integrated-moving-average (ARIMA) models.

The U.S. wholesale price of AA is determined at New Orleans, near most of the domestic ammonia capacity; thus, prices at this market are used in the study (Tennessee Valley Authority, July, 1989). F.o.b. prices, U.S. Gulf, are available for AA starting in February 1977 from Green Markets (McGraw-Hill, Inc., 1990). Following the iterative technique of identification, estimation, and diagnostic checking popularized by Box and Jenkins (1970), the series WHAA was identified as an SARIMA(2,1,0)(1,0,0):

$$(1-\phi_1 L-\phi_2 L^2)(1-\phi_{12}L^{12})(1-L)WHAA_t = e_t$$

where L is the backshift operator, ϕ are autoregressive parameters, and $\{e_t\}$ is a sequence of white noise. Maximum-likelihood estimation of this model over the period February 1977-October 1987 yields:

$$\phi_1 = 0.686(0.088) \qquad \phi_2 = -0.179(0.088) \qquad \phi_{12} = 0.240(0.090),$$

where the quantities shown in parentheses are standard errors. Parameter estimates are all statistically significant at the 95-percent level and lie within the bounds of invertibility. Diagnostic checks do not reveal any serious model inadequacies and the residuals of the model do not differ from white noise as indicated by the Q-statistics of 3.76 and 17.78, which are not significant when compared to a χ^2 statistic with 9 and 21 degrees of freedom, respectively (Box and Pierce, 1970). No intercept was used because it did not differ significantly from zero.

Similarly, using biannual data (April and October), the ARIMA model for the RAIL series was identified as an ARIMA(1,1,0):

$$(1-\phi_1 L)(1-L)RAIL_t = \delta + \epsilon_t$$

Data for the total rail freight rate index is available from the U.S. Department of Labor. Maximum-likelihood estimation of this model over the period April 1977-October 1987 yields:

$$\delta = 1.894(1.109) \qquad \phi_1 = 0.668(0.167)$$

Diagnostic checks reveal no serious model inadequacies. The Q-statistic of 5.43, which is not significant when compared to a χ^2 statistic with 11 degrees of freedom, indicates that the residuals of the model do not differ from

Table 4
Autoregressive Distributed Lag Models for 1977-1992

Retail Prices	constant	WHOLAA	LWHOLAA	RAIL	LRAIL	LRP	RBARSQ	RSS	Durbin's h
AA_1	53.73 (15.02)	0.513 (0.089)	0.306 (0.096)			0.299 (0.081)	0.86	3454.6	-0.22
UREA	45.07 (16.72)	0.456 (0.087)		0.101 (0.123)		0.462 (0.103)	0.82	3218.7	1.35
NS	28.71 (12.73)	0.158 (0.043)				0.646 (0.102)	0.72	1283.9	1.75
AN	20.46 (10.90)	0.178 (0.041)				0.768 (0.066)	0.87	1174.0	1.61
AS	16.72 (7.65)	0.08 (0.029)				0.83 (0.051)	0.90	653.14	1.24

1. The lagged retail price for AA enters with a "2-period" lag - one year instead of six months.

white noise.

The ARIMA models for WHAA and RAIL, estimated with data through October, are used to forecast these variables for the following April. Retail price forecasts from the OLS models are then developed by incorporating the forecasts for $WPAA_t$ and $RAIL_t$ into the respective equations for AA, UREA, NS, AN, and AS. Additional information is required to develop price forecasts from the AR(1) and ADL models. In addition to forecasts of $WPAA_t$ and $RAIL_t$, the AR(1) models require a lagged residual, e_{t-1}, which is available at time t. Similarly, the ADL models require lagged values of WPAA, RAIL and the dependent variable. These values are also available at time t.

Evaluation of Forecasting Models

Tables 5 and 6 provide summary statistics for the three models: ordinary least squares (OLS), Cochrane-Orcutt (AR1), and Autoregressive Distributed Lag (ADL). We report the mean percent errors (MPE), the root mean square error (RMSE), and the mean absolute percent error (MAPE). The statistics are for the six one step ahead forecast errors. The unconditional forecast results are given in Table 5 and the conditional forecast results are in Table 6.

The results are mixed as to which model produces the best forecast. However the ADL model appears to perform the worst of the three models. It produces the highest MPE in three of the five price equations using the conditional and unconditional technique. Similarly, the RMSE from the ADL model is also the highest in four of the five equations. The ADL model, while statistically superior, does not produce numerically better forecasts.

Conclusion

We test models used in predicting five nitrogen fertilizer product prices. The original model specified was a simple static model. It was found to have errors which exhibited first order autocorrelation. Then, we compared the "textbook" approach of correcting for this problem using the Cochrane-Orcutt procedure against the more general ADL model. The later was found to reject the common factor restrictions implied by the former. When the three models one step ahead forecasting power was compared there is no clear winner. We expected the ADL model to perform best, but it appeared to perform the worst. This may be due to several reasons: 1) the forecast horizon was restricted to one period, 2) the sample is too small, meaning the results are not significantly different, 3) this is a case where a well fitting model is not a good forecasting model.

Table 5
Comparison of Unconditional Forecasts 1988-1993

	Mean Percent Error			RMSE			MAPE		
Retail Prices	OLS	AR(1)	ADL	OLS	AR(1)	ADL	OLS	AR(1)	ADL
AA	1.62	1.73	8.51	1.94	2.52	4.23	2.01	2.33	8.51
UREA	8.56	1.49	6.50	4.62	6.53	7.72	5.86	5.08	8.18
NS	5.73	4.08	7.97	2.66	1.52	5.14	5.73	4.08	7.97
AN	1.50	0.01	5.77	3.89	3.60	3.55	3.44	2.59	5.81
AS	-2.34	-4.07	2.03	4.32	3.23	1.94	3.05	4.07	2.03

Table 6
Comparison of Conditional Forecasts 1988-1993

	Mean Percent Error			RMSE			MAPE		
Retail Prices	OLS	AR(1)	ADL	OLS	AR(1)	ADL	OLS	AR(1)	ADL
AA	-1.25	-1.30	6.61	5.45	5.78	8.12	4.44	4.70	7.78
UREA	3.39	0.15	0.50	5.41	3.78	6.59	4.97	2.76	5.33
NS	4.45	3.11	7.21	5.06	2.09	6.61	4.62	3.11	7.21
AN	0.38	0.53	5.17	4.35	2.67	5.31	3.09	2.07	6.11
AS	-2.84	-3.91	1.71	3.29	1.80	2.66	3.21	3.91	2.32

References

P. Andrilenas and H. Vroomen, "Fertilizer," in Seven Farm Input Industries, Ag. Econ. Report No. 635, US Dept. Agr., Econ. Res. Serv., September 1990, pp 14-29.

G.E.P. Box and G.M. Jenkins, "Time Series Analysis, Forecasting and Control," Holden-Day, Inc., San Francisco, CA, 1970.

G.E.P. Box and D.A. Pierce, "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models", Journal of the American Statistical Association, 65, 1509-1526(1970).

Gilbert, Christopher L.; 1990, "Professor Hendry's Econometric Methodology," Modeling Economic Series: Advanced Texts in Economic Series, ed. C.W.J. Granger, Oxford Press, chapter 13.

Hendry, David F. and Grayham E. Mizon; 1978, "Serial Correlation as a Convenient Simplification, not a Nuisance: A Comment on a Study of the Demand for Money by the Bank of England," The Economic Journal, 88, 549-563.

Lafontaine, F. and Ken White; 1986, "Obtaining any Wald Statistic You Want," Economics Letters, vol. 21, 35-40.

Lon-MU Liu, "Sales Forecasting Using Multi-equation Transfer Function Models," Journal of Forecasting, 6, 223-238(1987).

McGraw-Hill, Inc., Green Markets: Fertilizer Market Intelligence Weekly, New York, October 29, 1990, and earlier issues.

Sargan, J.D.; 1964, "Wages and Prices in the United Kingdom: A Study in Econometric Methodology," in Harts, Mills, and Whitaker (eds). Econometric Analysis for National Economic Planning. London: Butterworth, pp. 22-54.

Tennessee Valley Authority, Economics and Marketing Staff, "North American Fertilizer Capacity Data," Muscle Shoals, AL, July 1989.

Tennessee Valley Authority, Economics and Marketing Staff, "Commercial Fertilizers," Muscle Shoals, AL, December, 1989 and earlier issues.

U.S. Department of Labor, Bureau of Labor Statistics, Producer Price Indexes, monthly reports, 1977-92.

U.S. Department of Agriculture, National Agricultural Statistics Service, "Agricultural Prices," Washington, DC, various annual summaries, 1977-92, and April, 1993.

White, Ken; 1993, SHAZAM User's Reference Manual Version 7.0, McGraw-Hill.

Modeling and Forecasting U.S. Patent Application Filings

Kay Adams and Gus Mastrogianis, U.S. Patent and Trademark Office

Introduction

This paper summarizes the results from a research project to develop a methodology for predicting patent application filings at the U.S. Patent and Trademark Office (USPTO). Like many government agencies worldwide, the USPTO has switched to a fee-based funding environment; thus the need for reliable forecasts of revenue and resource allocation. Also, the USPTO is engaged in a long-term international cooperative effort with the European and the Japanese Patent Offices (EPO and JPO respectively) to track and project patent trends.

The research focus to date has been on aggregate patent application filings. We developed and compared naive, ARIMA, and econometric forecasting models at the one and two-year ahead horizon using annual and quarterly time series data. The Mean Error (ME), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) of the models from within-sample forecasting comparisons, suggest that the annual and quarterly econometric models are more reliable than the other two types of models. Both the annual and quarterly econometric models consistently outperform the naive and ARIMA models in predicting the actual patent application filing levels whether it is one-year ahead or two-year ahead within-sample forecasting.

The paper is organized into six sections. First, we present a review of the literature on patent application filings and related economic activities (e.g., the relationship between R&D and patenting). This motivates the construction of the econometric models. Description of the data used in the research is presented in the second section. The annual and quarterly empirical models are presented in sections three and four, respectively. Then, the within-sample unconditional forecast at the one and two-year horizons are compared. The final section concludes and discusses additional likely future research activities at the USPTO.

I. Patent and Macroeconomic Activity

Griliches (1990) stated that a patent represents an output of the inventive process having passed a two-part test of the economic threshold. First, the invention has been deemed worthy of an initial investment of resources by the inventor or the business organization. And once developed, an expectation of some level of utility or marketability is also tied to the invention. Thus, patent application filings represent an important measure of successful inventions meeting an expected economic value of the patent right.

In acknowledging the distinction between the process and output of innovation, Acs and Audretsch (Germany, 1989) provided empirical evidence which suggests that patents do provide a good, albeit imperfect, measure of innovative activity. Tempering this view somewhat, Griliches (1989) argued that patents (grants) are not a "constant yardstick indicator of either inventive input or output" (p. 291). Given that all inventions do not result in patents, that not all inventions are patentable, and that patents vary considerably in their economic impact, the problems associated with using patents as the sole measure of inventive output are easily recognizable. However, Griliches does contend that the aggregate count of patents may serve as a measure of shifts in technology.

Patents are intrinsically linked to the larger economic picture -- to the process of innovation and to technological and scientific change. Based on a review of domestic patent application filings from the Civil War to World War II, Schmookler (1954) concluded that industrial invention is economically caused. In his view, the level of inventive activity is represented by such factors as the number of technical trained workers, industrial (variable and not fixed) inputs, and the Gross National Product (p. 188). Schmookler also observed that "changes in the level of inventive activity cause changes in the number of patent applications filed" (p. 186).

Scherer (1983) tested the relationship between the R&D and patenting activity. He believed that the propensity to patent can be measured by examining the number of patents industrial corporations obtain per million dollars of company-financed research and development expenditures (p. 108). When expressed as a ratio, the measure becomes the number of patents to R&D expenditures.

Scherer found evidence which suggests that, within industries, patents rise in connection to R&D. Taken from 1976 and 1977, the survey covered 443 corporations and 15,112 U.S. patents originating from U.S. resident inventors and issued to the same corporations. This research supports the general view that R&D is a proxy measure of the inputs to the inventive system, while patents represent a qualified proxy measure of the output from the inventive process (Acs and Audretsch, 1989).

Based on the theory that R&D expenditures and patenting activities are positively related, Scherer's findings also showed:

1. patenting is closely associated with formally organized R&D activity; and
2. of the industries studied, approximately 75 percent had R&D expenditures with either constant or increasing rates of returns as demonstrated by patent grants (p. 111-117).

Pavitt (1982) supports the proposition that a strongly positive relationship exists between R&D and patenting activities. Scherer found evidence to support a lag of approximately nine months from the time of invention to the filing of a patent application (p. 108). In a related study, Bosworth and Westaway (U.K., 1984) recognized the existence of a lag between expenditures on R&D and the filing of a patent application. The model constructed for their study showed the time lag to be approximately one year. From this research, the authors proposed that a relatively high percentage of R&D programs yields fairly immediate results in the form of patents, typically on the order of one year (p. 144).

A study by Hall, Griliches, and Hausman (1986) focused on the analysis of the relationship between patenting and R&D activity at the firm level during the 1970s in the U.S. manufacturing sector. Based on previous research, recognition is given to the theory that successful research and development efforts generally lead to both patents and to the further commitment to invest in development. The authors saw annual R&D expenditures as investments which add to a firm's stock of knowledge. As they describe it, the stock of knowledge depreciates over time, rendering the contribution of older R&D less valuable. Using this assumption, they attempted to use patent application filings as a possible indicator of the value added to the knowledge base in any given year. Additionally, their study concluded there is a lag structure from R&D to patents, however, the timing of the lag remained uncertain.

Several other economic variables also have been shown to demonstrate evidence of relatedness to patents. For example, Griliches (1989) compared domestic patent applications with economic variables, such as real Gross National Product (GNP) and Gross Private Domestic Investment (GPDI). He concluded:

1. Fluctuations in R&D do affect the number of patent applications applied for, but less than proportionately, and to the extent testable, the causality runs from R&D to patents (p. 309-311).

2. An estimated 10 percent increase in defense expenditures causes a 5 percent decline in domestic patenting, implying that defense spending takes away resources from patentable research (p. 310).

3. The vast majority of the systematic short-run variability in aggregated domestic patenting is explained by the R&D and national defense R&D variables, and virtually none of the other economic variables (included in this study) make any significant contributions on their own (p. 311).

II. U.S. Patent and Macroeconomic Data

The USPTO patent applications filing data base was constructed by the Office of Long-Range Planning and Evaluation. The research goal was to forecast patent application filings by fiscal year. Most of the patent application filing data were readily available. However, some of the data formats varied. Patent application filing data differed by reporting periods or increments (e.g., weekly, biweekly, calendar, and fiscal year) and patent type/category (i.e., utility, plant, reissue, and design). In some cases, patent application filing data, most notably quarterly filings, were estimated using several past data reporting structures to derive the fiscal year utility, plant, and reissue (UPR) data format.

1. Annual USPTO Patent Application Filings

Annual patent application filings were available for total, domestic origin, and foreign origin. The amount and format of the patent application filing data varied among the three annual data sets. Total annual UPR patent application filings were available from fiscal year 1949 to fiscal year 1991 from the Commissioner of Patents and Trademarks Annual Reports. See Figure 1. Total annual patent application filings were consistently maintained as utility, plant, and reissue data since fiscal year 1949.

However, one inconsistency found in total annual UPR application filings was the switch to a new basis for a fiscal year. With the adoption of a new fiscal year in 1976, the year shifted forward by one quarter. Since quarterly data were not available prior to 1976, the switch was inconsequential for the quarterly data set. But the lack of quarterly data prior to that time made it impossible to adjust annual total UPR patent application filings to this change. Therefore, the years 1949 to 1976 are under the prior fiscal year basis (July to June) and 1977 forward are under the present fiscal year basis (October through September). The major reason for this concession was to have as many annual total UPR application filing observations as possible for the development of forecasting models.

In fiscal year 1949, patent application filings totaled nearly 66,000; by 1991 annual filings were approximately 168,000. Over the full sample period, the average annual growth rate was 2.06 percent. There is a clear upward trend to the data with a noticeable increase in the 1980s. During the 1980s, the average growth rate was in excess of 4 percent or about 2.5 percent higher than for the 1950s, 1960s, and 1970s. In fiscal year 1983, there was a decrease in filings of about 20,000 from 1982, because of an announced sharp increase in the patent application filing fees. The actual sample for the identification, estimation, and forecasting periods is 1956 through 1991. This was done because of data limitations for the other variables in the econometric models. Dropping the first seven observations did not significantly affect the ARIMA model's performance.

2. Quarterly USPTO Patent Application Filings

A quarterly data set for total UPR patent application filings was constructed from weekly, biweekly, and monthly total patent application filing data. A data set for the first quarter of 1976 through the third quarter of 1992 was transcribed from dozens of paper reports, patent application filing data varied by patent type and reporting periods. A process was developed at the USPTO to estimate total quarterly UPR patent application filings from the paper reports.

The first step of the process was to identify which quarters should be constructed since there are currently two time periods that can be used to construct quarterly UPR patent application filing data at the USPTO. One is based on the first and last day of the patent production process, and the second is based on estimating the first and last day of each fiscal quarter using monthly patent application filing data. There is a considerable degree of variability in the number of days in each quarter based on the patent production process. For instance, the cutoff dates for the end of each quarter in the patent production process may occur as many as fourteen days before the last day in a fiscal quarter or may start as early as fourteen days before the first day of a fiscal quarter.

Since the research includes quarterly economic data, which are based on the first and last day of a calendar quarter and easily converted to a fiscal quarter, total quarterly UPR application filing data were constructed based on estimated calendar monthly patent application filings. Total quarterly patent application filing data based on monthly application filings are consistent with calendar quarters. Total UPR, UPR&D (aggregate of UPR and design patents), and utility weekly and biweekly patent application filings were used to estimate monthly UPR application filings. Total monthly UPR application filings were then used to estimate the total quarterly UPR patent application filings.

Total quarterly UPR application filings are shown in Figure 2. There are several noteworthy characteristics of the data. First, the number of filings prior to 1982 was fairly constant. In 1982Q4, there was a big increase due to the dumping or "surge" of filings prior to the first and largest patent application filing fee increase since 1966. Starting in 1984, there is an upward trend to the data that appears to slow down in 1991-1992. The increase is from about 26,000 filings per quarter to 42,000 filings per quarter by 1992. Another notable feature of the data series is the saw-toothed pattern between consecutive observations around the average or trend.

3. USPTO Patent Application Filing Fees

In addition to patent application filing data, patent application filing fees prior to fiscal year 1965 were also needed for this study. Basic utility and reissue patent application filing fees were used to create a fee time series from fiscal years 1932 to 1965 and the first quarter of fiscal year 1976 to the third quarter of 1992. After fiscal year 1982, the filing fee data set reflects fees charged to large entities for utility and reissue patent application filings. Actual dates of filing fee changes (i.e., the dates when new filing fees became effective) were identified in U.S. Public Law documents and the Commissioner of Patents and Trademarks Annual Reports.

Patent application filing fee data were constructed for only utility and reissue application filings, which accounted for 93.95 percent of total UPR&D application filings in fiscal year 1991. From fiscal years 1949 to 1991, utility and reissue patent application filings have averaged 93.34 percent of total UPR&D application filings. It should be noted that filing fees for plant patents were excluded because plant fee data were not identified for fiscal years prior to 1965, and the application filing fee for plant patents is lower than fees for utility and reissue patents. Also, plant patent application filings only accounted for 0.23 percent of total USPTO patent application filings in fiscal year 1991 and averaged 0.16 percent of total filings since fiscal year 1965.

From fiscal year 1983 to the present, the USPTO has offered large and small entity filing fees. Only large entity filing fees were used to develop forecasting models. There were several reasons why only the large entity fee was selected. Large entities accounted for 66.61 percent of total patent application filings on average from fiscal year 1983 to fiscal year 1991, and both the large and the small entity fees have changed simultaneously during that period.

Although annual utility and reissue patent application filing fee data were available back to fiscal year 1932, it was only necessary to use the fee data from fiscal years 1949 to 1991 since these are the only fiscal years for which total annual UPR patent application filings are available. For quarterly utility and reissue fee data, a database for only the first quarter of fiscal year 1976 to the third quarter of fiscal year 1992 was created because of the similar constraint imposed by total quarterly UPR patent application filings. Once the fees were compiled, they were adjusted for inflation using the 1987 Gross Domestic Product implicit price deflator to derive real fee changes.

4. Economic and Demographic Data

Annual domestic economic and demographic data were available starting with calendar year 1949. Several of the economic indicators were only available in 1982 U.S. dollars, in nominal U.S. dollars, and in 1987 U.S. dollars. With the exception of Gross Private Domestic Investment (GPDI), the 1987 GDP implicit price deflator (GDP deflator) was selected for converting nominal economic data to real economic data. For GPDI, a separate 1987 GPDI price deflator was identified.

Quarterly domestic and foreign data were available for the majority of economic variables but not for any of the demographic variables (e.g., Scientists and Engineers engaged in R&D and graduates with a Bachelor's in Mathematics, Science, and Engineering). However, the data sets are more complete for the time period studied (1976-1992). One notable exception is the economic variable, R&D Expenditures (both domestic and foreign), that is not reported by quarter from the sources identified.

Table 1 lists all of the variables, their respective sample periods and coverage used in the study.

III. Annual Forecasting Models

1. Naive Model for Annual USPTO Patent Application Fillings

For the purposes of this project, the software package Forecast Pro was used in the expert system mode to select the best naive model. The Holt exponential smoothing model with a linear trend was found to minimize the root mean square error. This model's technical details are discussed below and followed by the empirical results.

In many real world applications, the data are not so simple in their patterns that simple exponential smoothing provides a reliable forecast. For example, the annual data on UPR patent application filings exhibit an upward trend over time. When there is a trend in the data, the simple smoothing forecasting model will have large errors that fluctuate from positive to negative or vice versa. A common method for adjusting the simple

smoothing forecasting model for trend is Holt's method, named after its originator C. C. Holt (1957). The exposition below follows Wilson and Keating (1990). Three equations and two smoothing constants are used in the model.

$$F_{t+1} = \alpha X_t + (1-\alpha)(F_t + T_t) \quad (3.1)$$

$$T_t = \beta(F_t - F_{t-1}) + (1-\beta)T_{t-1} \quad (3.2)$$

$$H_{t+h} = F_{t+1} + hT_{t+1} \quad (3.3)$$

where:

F_{t+1} =	Smoothed value for period $t+1$
X_t =	Actual value in period t
F_t =	Smoothed value for period t
T_t =	Trend estimate
α =	Smoothing constant $(0 < \alpha < 1)$
β =	Smoothing constant for trend estimate $(0 < \beta < 1)$
h =	Number of periods ahead to be forecast
H_{t+h} =	Holt's forecast value for period $t+h$
$(F_t + T_t)$ =	Actual forecast of X_t made in period t

Equation (3.1), the one period ahead forecast or smoothed value, adjusts F_{t+1} for a fraction of this period's level, X_t, and for growth of the previous period, T_t, by adding T_t to the smoothed value of the previous period, F_t. The forecast can also be expressed as an adaptive or error learning model.

$$F_{t+1} = (F_t + T_t) + a(X_t - F_t - T_t) \quad (3.4)$$

A one period ahead forecast is equal to the previous periods forecast plus the difference between the current actual value and its forecast.

The trend estimate is calculated in equation (3.2), where the difference of the last two smoothed values is calculated. Because these two values have already been smoothed, the difference between them is assumed to be an estimate of trend in the data. The logic behind β in equation (3.2) is the same logic used in the simple smoothing model for the constant α. The most recent trend, $(F_{t+1} - F_t)$, is weighted by β and the last previous smoothed trend, T_t, is weighted by (1 - β). The sum of the weighted values is the new smoothed trend value T_{t+1}.

Equation (3.3) is then used to forecast h periods into the future by adding the product of the trend component, T_{t+1}, and the number of periods ahead to forecast, h, to the current value of the smoothed data F_{t+1}. This method accurately accounts for any linear trend in the data.

Table 2 contains the results from the Holt exponential smoothing model. The model suggests that a one-year ahead forecast should be about 80 percent of this year's patent application filings plus 20 percent of the forecast error for this year. The Durbin-Watson statistic suggests that consecutive errors are uncorrelated, however, the statistic is not appropriate, because there is no intercept in the model [See Durbin and Watson(1971)]. The Ljung-Box-Pierce test for the correlation of errors through lag 18 is also insignificant.

2. Box-Jenkins (ARIMA) Model for Total Annual USPTO Patent Application Filings

Since patent application filings have a clear upward trend, the series was transformed to natural logarithms to reduce the relative increase from one year to the next. Table 3 provides the mean, variance, autocorrelation function, and partial autocorrelation function for the filings data in levels and first differences of natural logarithms.

The autocorrelation of filings in levels at lag 1 is 0.85, and it decays slowly or exponentially to 0.69 at lag 2, 0.56 at lag 3, and so on. The partial autocorrelation at lag 1 is the same by definition as the autocorrelation at that lag. It is significantly different from zero when divided by the asymptotic standard error 0.17. The magnitude of the autocorrelation at lag 1 and the slow rate of decay are indicative of a nonstationary series with a unit root. The Box-Jenkins procedure recommends transforming the data into first differences of natural logarithms. This has a convenient interpretation, because the data is in annual growth rates.

The second half of the Table 3 presents the results in first differences of natural logarithms. The average growth rate is approximately 2.3 percent with a standard deviation of 5.1 percent. The autocorrelation at lag 1 is -0.19, but the standard error of the autocorrelations is 0.17 indicating that last year's growth rate is weakly correlated (or a marginal predictor) of this year's growth rate. The implied t-ratio is greater than 1.0, but not greater than 1.65 if a 10 percent significance level were chosen. Similarly, autocorrelations at longer lags are not individually significantly different from zero.

The modified Ljung-Box-Pierce statistics are also tabulated in Table 3. The statistics test whether autocorrelations through a given number of lags are significantly correlated with the current value of the series. The statistics are distributed as Chi-square with the number of degrees of freedom, df, equal to the number of lags. The p-value(s) are the probabilities of rejecting the null hypothesis (of no correlation) when the alternative is true. If the values were less than 0.1, there would be some evidence of correlation between that number of lags and the current value. None of the Ljung-Box-Pierce statistics are significant suggesting that a possible ARIMA model would be in first differences without a constant (or drift) term for the growth rate. This specification, ARIMA(0,1,0), is referred to as a random walk in the time series literature. The standard error (deviation) is given below the annual growth rate. The estimate for the growth is insignificant from zero since the t-ratio is less than one.

Table 4 presents results of a test of the specification ARIMA(1,1,0) with a constant. The AR(1) parameter is as expected from the examination of the autocorrelation function. The t-ratio is slightly greater than one, suggesting that it provides marginal additional explanatory power. The constant term is about 2.7 percent, one half a percent higher in the simple random walk with drift model, but within a one standard error bound. The autocorrelation function and the Ljung-Box-Pierce statistic suggest that there is no further autocorrelation in the residuals of the estimated model. R-square is low, 0.0368, but this is not unusual in ARIMA models. At the bottom of the table is a F-test comparing the two models using the residual sum of squares. The F-statistic is 5.26 which suggests the random walk with drift model is rejected at the 5 percent significance level.

Thus, a candidate ARIMA model is the ARIMA(1,1,0) shown below.

$$\Delta F_t = \underset{(0.0095)}{0.02741} - \underset{(0.1684)}{0.1945} \cdot \Delta F_{t-1} + \varepsilon_t$$

The constant term can be interpreted as a year to year growth rate after the last year's effect on the growth of patent application filings. The negative coefficient on the lagged growth rate suggests that positive above average growth in one year will be followed by below average growth the next.

3. Econometric Model for Annual USPTO Patent Application Filings

The econometric model for patent application filings is based on a simple theory of innovation and economic growth. Firms and inventors invest resources to improve existing technologies and develop new technologies. One observable output of the process is a patent filing. This represents the firms' and inventors' (initial) claim to an invention. A filing is made to both protect the property rights of the claim and to earn rents or royalties from the invention.

It is assumed that the economic theory underlying the patent application filing process can be generalized to the aggregate level. A simple model of UPR patent application filings would assume that filings are a function of past inventions (or filings), the cost or price of a filing, and the resources invested toward inventions. The first component can be captured by including lagged values of filings in the model. The price of a filing is measured by the filing fee paid upon application; lagged or previous fees can be used to derive estimates of both short run and long run price elasticities. This measure is only a proxy, because it omits patent attorney fees. The resources invested for inventions can include both economic and demographic components. Economic variables could be represented by expenditures on research and development (R&D) and public and private investment expenditures. The number of scientists and engineers engaged in R&D can serve as a proxy for the human resources devoted to the process of invention.

The approach for finding an econometric model begins with a general model which is reduced to a parsimonious one via empirical tests of model fit and parameter stability. Below, the model building and testing process is described in two parts. First, an initial set of models are estimated and found to be deficient in terms of parameter stability.

The initial or most general model specified was a single equation autoregressive distributed lag regression (ADL) of filings F_t on filings for the last four years, the current and past filing fees, FEE_{t-i}, over the last four years, and the expenditures on research and development in the United States, ARD_{t-i} over the last four years. The ARD series is constructed as a three year moving average. This series attempts to smooth expenditure and reflects a commitment to research leading to patenting. The length of the moving average was chosen arbitrarily recognizing a potential degrees of freedom problem.

$$F_t = \alpha + \sum_{i=1}^{4} \phi_i \cdot F_{t-i} + \sum_{i=1}^{4} \beta_i \cdot ARD_{t-i} + \sum_{i=4}^{4} \gamma_i \cdot FEE_{t-i} + \varepsilon_t \qquad (3.5)$$

All variables are in natural logarithms. Both the fee and expenditure data were adjusted to 1987 dollars by the GDP deflator. The current fee is used, because the prices were known in advance by patent applicants and current prices can affect filing behavior. The sample period used is from 1956 to 1991.

The fee series can be treated as exogenous. R&D data enter the model through a lag, because current R&D expenditures are unlikely to result in a filing immediately. Also, the filing data are based on fiscal years whereas the R&D expenditure data are only available on a calendar year basis. Thus, using the one-year lag of ARD is effectively a three quarter lag. R&D expenditures are implicitly assumed to be exogenous in the model presented here. This implies that there is no feedback relationship between R&D and the output of patents, and the parameters of interest are invariant to the relationship. There was insufficient current data on scientists and engineers to be useful in the model. The most recent estimates are from 1989.

A more parsimonious model was searched for by reducing the number of lags across all variables and by variables individually. The lag length for the ADL is based on minimizing the Akaike Information Criterion (AIC) and Bayesian Schwartz Criterion (BSC) as recommended in Granger and Newbold (1986) and Judge et. al. (1988). The results are presented in Table 5. The unrestricted model with 4 lags yields an AIC and BSC of -6.85 and -6.19, respectively. It is minimized when there is one lag on filings, one lag on research and development expenditures, and the current and one lag of fees; the AIC is -7.02 and the BSC is -6.80. Other variations of the initial ADL restricting filings to one lag and both filings and fees to one lag were tried and yielded the same result.

Results from the final model are presented in Table 6. The econometric model chosen for forecasting purposes used past filings, a three year moving average of research and development expenditures, and a new variable for the price, LP. The price series was transformed to take account for the impact of the announced fee structure changes in 1983. Prior to 1983, the fee, lagged one and two years, is used and after that period the current and lagged fees were used in the new price series. Tests for autocorrelation resulted in a Durbin's H statistic of 0.84, suggesting the residuals are not autocorrelated at lag one. The Lagrange Multiplier statistics and Ljung-Box-Pierce statistics imply there is no further autocorrelation of the residuals. The tests for heteroscedasticity suggests that there is none present. The Jarque-Bera asymptotic LM normality test implies that the null hypothesis of no skewness and excess kurtosis cannot be rejected.

Sequential Chow Tests revealed a structural break in the model beginning in 1982. The Chow test statistic was calculated as:

$$Chow = \frac{(sse - sse1 - sse2)/k}{see/(n1 + n2 - 2k)}$$

where
sse = sum of squared errors for full sample
$sse1$ = sum of squared errors first part of sample
$sse2$ = sum of squared errors second part of sample
k = number of estimated coefficients
$n1$ = number of observations in first part
$n2$ = number of observations in second part

The statistic is distributed as an F, thus if it is less than the critical value from an $F(k, n1 + n2 - 2k)$, there is no evidence of a structural break. In the tests, the F statistic has 5 degrees of freedom in the numerator and 25 degrees of freedom in the denominator. The critical values are 2.6 at five percent and 3.86 at one percent. Despite this problem, the model will be used for forecasting purposes. Alternative models were tried, but could not encompass this model and in many cases had more severe problems.

The model chosen only includes a single lag of past filings, last year's three year moving average of R&D, and current and last year's real fees.

$$\underset{(0.4433)}{F_t = 0.21665} + \underset{(0.0725)}{0.81199} \cdot F_{t-1} + \underset{(0.0557)}{0.17632} \cdot ARD_{t-1} - \underset{(0.0191)}{0.1226} \cdot FEE_t + \underset{(0.0195)}{0.13268} \cdot FEE_{t-1}$$

The dynamics of the model are interesting. The short run effect of the three year moving average of R&D expenditures suggests that a one percent increase leads to 0.17 percent increase in patent application filings. However, the long run impact of a one percent increase is 0.9378 percent or almost one. This is consistent with research from the literature review. The immediate effect of a one percent increase in filing fees is -0.1226, while the effect from the previous year is a positive 0.1329. The net effect is zero. Patent application filing fees are set by statute. Historically, patent application filing fees have not been set as an impediment to the innovation and invention process. From an economic perspective, patent application filing fees can be viewed as user costs. The negative-positive sign pattern captures both the surge effect of announced fee increases and the declining real fee(s) between fee increases.

A simple adjustment was tried using a dummy variable, SURGE, which takes on the value 1 in 1982, -1 in 1983, and 0 in all other years. This was designed to capture the dumping or surge of filings just before the fee increases. In both cases, the fee increase was announced and passed by Congress as P.L. 97-247. The SURGE was used as a replacement for the FEE series and came up with the same lag specification for filings and R&D. Unfortunately, the model fit was not as "good" as with the FEE series.

IV. Quarterly Forecasting Models

The models of quarterly patent application filings are based on data from the first quarter of 1976 (1976Q1) through the third quarter of 1992 (1992Q3). Quarterly filings data are collected on a fiscal year basis, October 1st to September 30th.

1. Naive Model for Quarterly USPTO Patent Application Filings

The software package Forecast Pro, when run in expert mode, did not select any of its "canned" naive extrapolative models. Rather, it recommended a Box-Jenkins type (or ARIMA) type model. We decided to construct several naive models, because of the simplicity, ease of use, and forecasting potential. The first and most obvious was a simple random walk with trend starting in 1984. This was ruled out, because it would imply extremely high growth rates of 8.8 percent per annum in filings or a doubling of filing applications every of 7.9 years.

Another model which was ultimately chosen and performs quite well is a transfer function type model. It combines features of an autoregressive model and deterministic components. The functional form of the model is shown below:

$$F_t = \alpha + \beta_1 \cdot F_{t-1} + \beta_2 \cdot F_{t-2} + \beta_3 \cdot F_{t-4} + \beta_4 \cdot SURGE + \varepsilon_t \qquad (4.1)$$

Filings from three quarters ago did not contribute to the model fit so they were omitted from the specification. The SURGE variable is zero except 1982Q4 where it equals one and 1983Q1 where it equals minus one. The results from the model are presented in Table 7. All variables are significant except the constant which is kept in for testing purposes. The fit is extremely high, adjusted R-square is 0.95. There is no evidence of first order autocorrelation of the residuals from the Durbin's H statistic of -0.73. The residuals appear to be normal from the Jarque-Bera test, and the variance is constant from tests for heteroscedasticity.

2. Box-Jenkins (ARIMA) Model for Quarterly USPTO Patent Application Filings

The graph and autocorrelogram of the patent application filings showed that the series is nonstationary. As with the annual data, the quarterly data are transformed to first differences of natural logarithms. If only the post 1983 data were used, the ARIMA model would behave poorly. Furthermore, the annual growth rate of 8.8 percent for that period is not expected to continue as a long term trend. The ARIMA model for quarterly patent application filings will need to be used with caution, because of the drawbacks caused by the clear structural changes.

Table 8 contains the autocorrelation functions of the series in natural logarithms of levels and first differences. The nonstationary properties of the series are clearly presented by persistence (or slow decay) of the autocorrelations in levels. There is no clear pattern to either the autocorrelations or partial autocorrelations of the series in first differences. This suggests that a mixed model including both autoregressive (AR) and moving average (MA) components would be the best candidate model.

Since the maximum lag length for a significant partial autocorrelation is three, that was the longest lag chosen for either the AR or MA components. Various combinations of models were tried ARIMA(1,1,1), ARIMA(2,1,1), ARIMA(3,1,1), ARIMA(3,1,2), ARIMA(1,1,2), and ARIMA(1,1,3) were tried along with pure integrated AR and MA models. The model which fit best is an ARIMA(1,1,1); the results are given in Table 9. The negative sign on the AR coefficient is caused by the saw-toothed pattern in the data noted earlier. The residuals appear to be uncorrelated individually or as a group from the Ljung-Box-Pierce statistics. However, there is strong evidence of excess kurtosis.

3. Econometric Model for Quarterly USPTO Patent Application Filings

The quarterly model of patent application filings is based on the same economic logic as the annual model. Filings are an output of the inventive process.

Lagged filings from the same quarter in previous years are used to capture the dynamics and feedback from past inventive activity. Research and development expenditures are not available on a quarterly basis, so they were proxied for by using investment and gross domestic product. Because filings are the result of long run investment and economic trends, the explanatory variables are used or transformed to reflect this property. The real Gross Domestic Product, real Gross Private Domestic Investment and real Gross Private Non-Residential Investment series were converted to four quarter moving averages, referred to as AGDPQ, AGPDI, and AGIN respectively. This smoothes the series to reflect the trend in economic activity over the past year. The economic activity variables are on a calendar year basis. Filing fees were adjusted by the 1987 GDP deflator. An expected real fee series, ERFEE, was created to control for the announcement effect on fees. The series uses next period's fee as the fee in the current period. The SURGE variable is used to capture the impact of a change in the fee structure in 1983Q1 as in the naive model.

The technique for estimating the quarterly econometric model follows the general to specific approach used in the annual model. An ADL model with a fixed lag length is estimated. Fewer lags are used in an attempt to find a parsimonious model. The initial model is specified as follows:

$$F_t = \alpha + \sum_{i=1}^{4} \phi_i \cdot F_{t-4\cdot i} + \sum_{i=1}^{4} \beta_i \cdot ARDPQ_{t-4\cdot i} + \sum_{i=0}^{4} \gamma_i \cdot ERFEE_{t-i} + SURGE_t + \varepsilon_t \tag{4.2}$$

The two investment series were tried as replacements for AGDP, but did not appear to improve the predictive power. The AIC and BSC were used in selecting the appropriate lag length of the models. Tests for lag length were also performed using the two investment series, but they did not yield significantly better results. Also, lags two through four in the fee series provided no additional explanatory power, so the tests were conditioned upon the current quarter and one lagged quarter of the series. The results are summarized in Table 10. A model with filings lagged four and eight quarters, the four quarter moving average of real GDP lagged four and eight quarters, and the current and current expected real fee series and lagged one quarter appear to provide a parsimonious fit.

$$F_t = \alpha + \phi_1 \cdot F_{t-4} + \phi_2 \cdot F_{t-8} + \beta_1 \cdot AGDPQ_{t-4} + \beta_2 \cdot AGDPQ_{t-8} + \gamma_1 \cdot ERFEE_t + \gamma_2 \cdot ERFEE_{t-1} + SURGE_t + \varepsilon_t \tag{4.3}$$

Results from the model are provided in Table 11. The model fits the data quite well; adjusted R squared is 0.961, and the standard error of the estimate is 1259 with a mean of the dependent variable of 31329. The standard error is about four percent of the mean value. The Durbin's H statistic is -1.33 indicating first order autocorrelation is not a serious problem. There is no further autocorrelation from the residual correlogram. The null hypothesis of no heteroscedasticity cannot be rejected. The Jarque-Bera test for normality of residuals indicates there is no skewness and excess kurtosis.

The dynamics of the model are interesting. The short run effect of a billion dollar increase in the four quarter moving average of GDP lagged one year is associated with an increase in 10 patent application filings. A lag of one year is actually three quarters, because of the fiscal year calendar year distinction. On the other hand, the long run impact is about 17 filings. The dynamics are calculated as follows:

$$\text{Effect of \$1 Billion increase in GDP} = \frac{\hat{\beta}_1 + \hat{\beta}_2}{1 - \hat{\phi}_1 - \hat{\phi}_2} = \frac{8.8869}{1 - 0.47237} = 16.84$$

This result is interesting, because it implies for every billion dollars of real GDP there are 17 patent application filings. If real GDP is $5 trillion, that means there will be approximately 85,000 filings which is about equal to the number of domestic utility filings in 1991. The latter are about 49 percent of total UPR filings.

The net effect of the expected fees is zero, but they do appear to marginally contribute to the model. As previously mentioned, patent application filing fees are set by statute. Historically, patent application filing fees have not been set as an impediment to the innovation and invention process. From an economic perspective, patent application filing fees can be viewed as user costs. The negative-positive sign pattern captures both the surge effect of announced fee increases and the declining real fee(s) between fee increases.

V. Evaluation of the Forecasting Models

Tables 12 and 13 contain the results from the within-sample unconditional forecast evaluation of the annual and quarterly models respectively. The two-year ahead predictions in the annual model are dynamic in that they are based upon the one-year ahead forecasted value of patent application filings. The unconditional forecasts use the actual values for fees and R&D expenditures in the predicted period(s). Filing fees were known in advance, except for the inflation impact on their real values. The R&D measure enters into the two-year ahead prediction, and there it is through the last, most recent year in a three year moving average term.

The quarterly models forecast out eight quarters; the first four are summed for the one-year ahead prediction and the fifth through eight quarters are summed for the two-year ahead prediction. It is dynamic as well, because predicted the values of (lagged) patent application filings are used as the forecast horizon lengthens.

A rolling set of one to two-year ahead forecasts are constructed using naive, ARIMA, and econometric models estimated through 1984. Then, the models are re-estimated using data through 1985 to produce one and

two-year ahead forecasts. The procedure continues until 1990 when only a one-year ahead forecast is made. There are ultimately seven one-year ahead predictions and six two-year ahead forecasts.

In Tables 12 and 13, the actual filing level is given in the second column. Then the one and two-year ahead prediction errors from the naive, ARIMA, and econometric forecasts respectively are given in the next six columns. For example, when the model is estimated using data through 1984, the one-year ahead prediction error made in 1985 is 6,524 and the two-year ahead error is 10,629.

The annual econometric model outperforms the annual naive and ARIMA models. Its mean percent error is nearly half of the other two at 3.15 percent for the one-year ahead and 5.85 percent at two-year ahead. One note of caution, the forecasts from the annual econometric model appear to deteriorate starting in 1989. The RMSE from the econometric model is also the smallest among the three annual models. This may just be the result of using the exogenous variables, a criticism of forecast comparisons noted by Fair (1984). The naive model underpredicts in every case but the last one-year ahead forecast. The ARIMA model forecasts are less than actual levels in every case also. In the first half of the evaluation period, the econometric model overpredicts, thereafter it overpredicts.

In the within-sample forecast comparison, the quarterly forecasts using the econometric model generated smaller mean percent error and RMSE than the other two quarterly models for the one-year and two-year ahead forecast. The mean percent errors 2.18 at the one-year and 3.59 at the two-year ahead are smaller than the annual econometric model error. As with the annual model, the quarterly model overpredicts in the mid 1980s and then underpredicts in the late 1980s and early 1990s. The naive and ARIMA models also underpredict until 1990.

VI. Conclusions

A comparison of the one and two-year ahead *ex post* forecast suggests that the quarterly models consistently outperform the annual models in predicting the future levels of patent application filings. When the mean errors, the root mean squared errors, and the mean absolute errors from the *ex post* forecasts of the quarterly models are compared to the annual models, it is clear that the quarterly naive, ARIMA, and econometric models generate forecast errors much smaller than their respective annual models.

The error statistics also suggest that the quarterly econometric model outperforms the other two quarterly models. Judging from the *ex post* forecast errors, the quarterly ARIMA model appears to slightly outperform the quarterly naive model.

Among the annual forecasting models, it is the econometric model that outperforms the other two annual models in both one and two-year ahead *ex post* forecast comparison. However, unlike the case of quarterly models, the annual naive model appears to be much more reliable than the annual ARIMA model when the errors from *ex post* forecasts are considered. While the mean errors for the one-year ahead forecast from the annual econometric and naive models were 3,288 and 3,531 respectively, it was 5,565 for the ARIMA model. The annual ARIMA model performs particularly poorly relative to the other two models in the two-year ahead forecasts as indicated by its large mean error, RMSE, and mean absolute error.

In short, the quarterly models perform better than the annual models. This certainly is the case when their respective Mean Percent Errors, Root Mean Squared Errors, and Mean Absolute Percent Errors from the within-sample forecast tests are compared to those of the annual models. Based on the error statistics computed from the one and two-year ahead within-sample forecasting, the top three models in ranking appear to be the quarterly econometric, the quarterly naive and the annual econometric models. The USPTO will continue to examine the performance of the different models on a periodic basis with the most recent observations available.

FIGURE 1

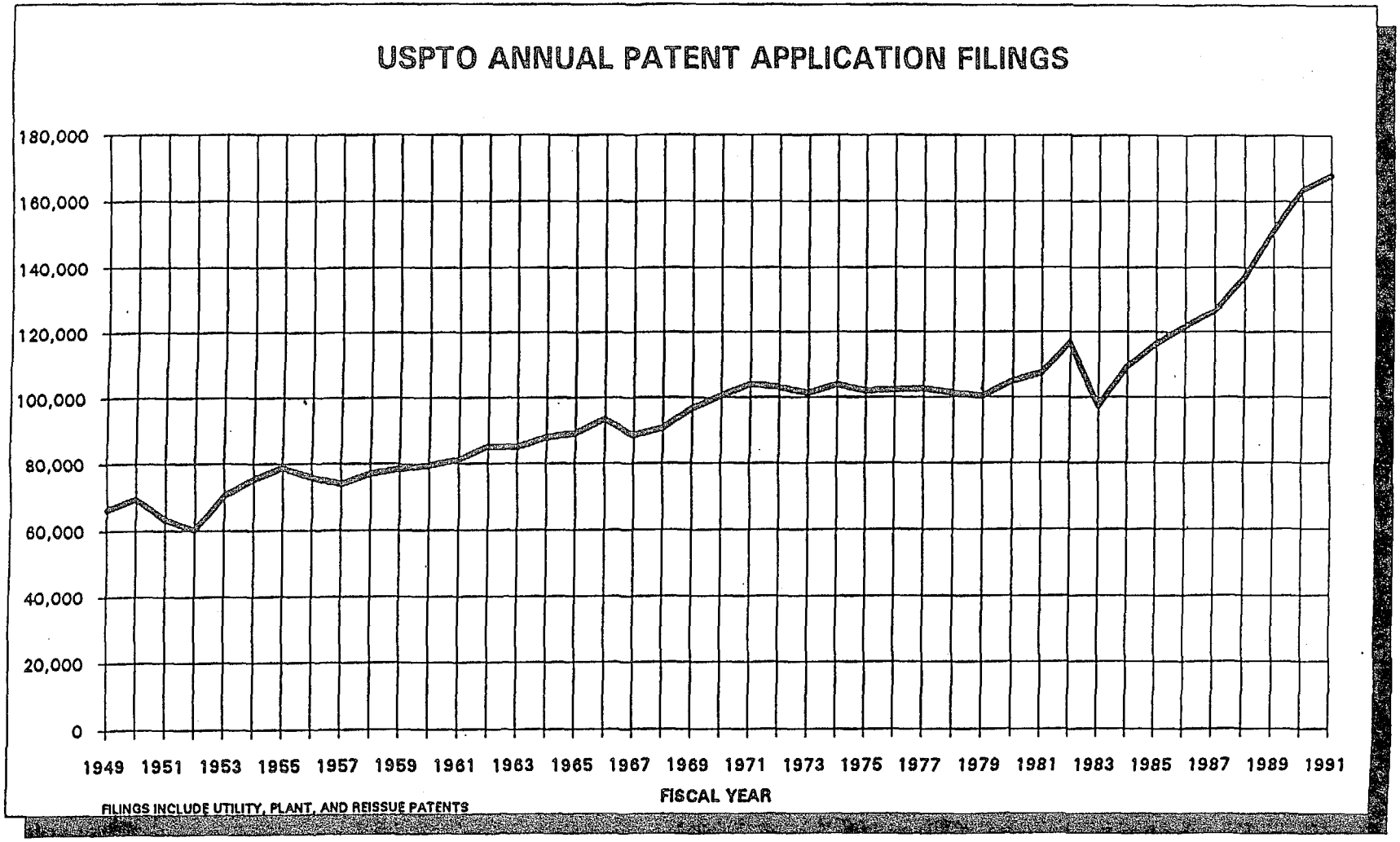

USPTO ANNUAL PATENT APPLICATION FILINGS
180,000
160,000
140,000
120,000
100,000
80,000
60,000
40,000
20,000
0
1949 1951 1953 1955 1957 1959 1961 1963 1965 1967 1969 1971 1973 1975 1977 1979 1981 1983 1985 1987 1989 1991
FISCAL YEAR
FILINGS INCLUDE UTILITY, PLANT, AND REISSUE PATENTS

FIGURE 2

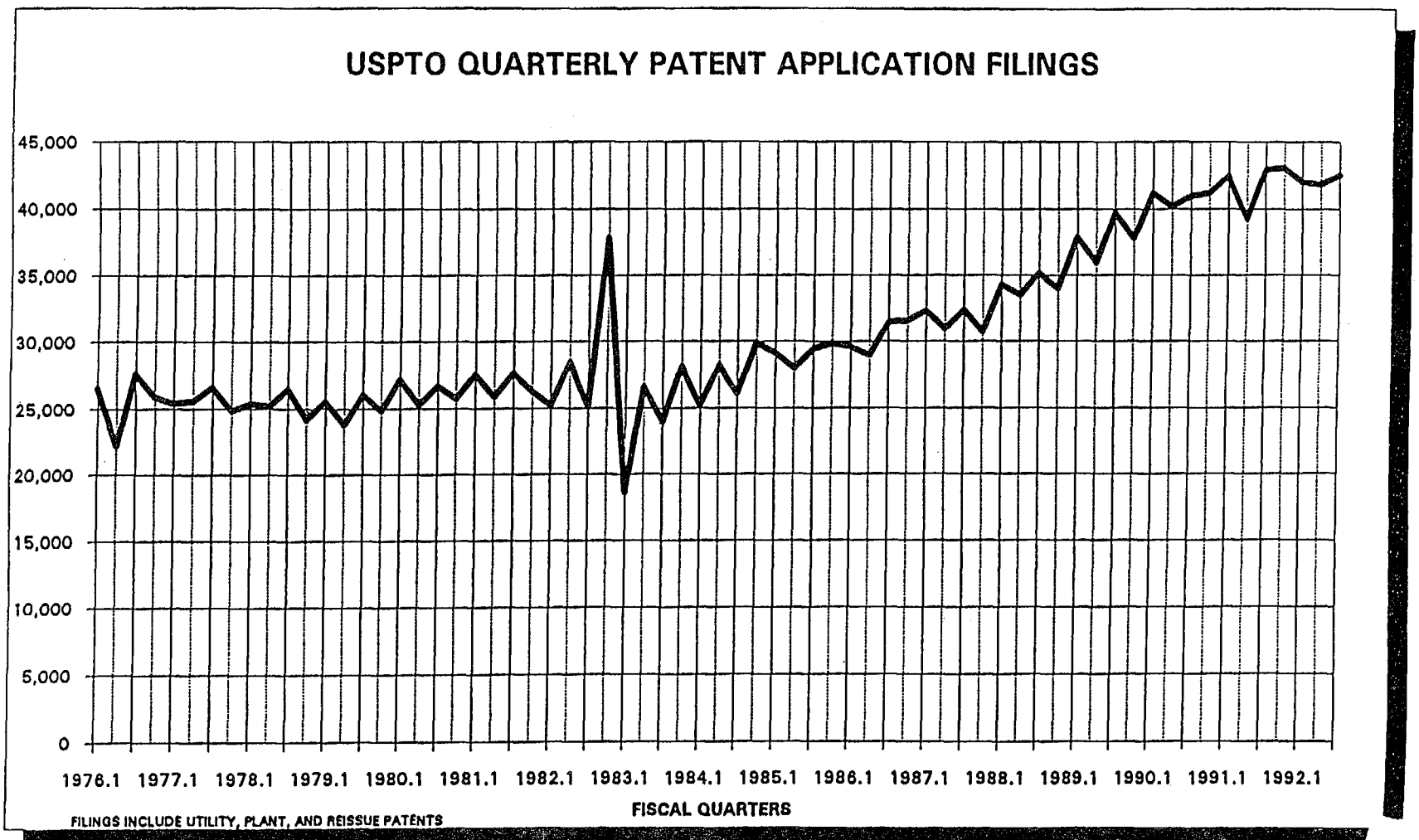

USPTO QUARTERLY PATENT APPLICATION FILINGS
45,000
40,000
35,000
30,000
25,000
20,000
15,000
10,000
5,000
0
1976.1
1977.1
1978.1
1979.1
1980.1
1981.1
1982.1
1983.1
1984.1
1985.1
1986.1
1987.1
1988.1
1989.1
1990.1
1991.1
1992.1
FISCAL QUARTERS
FILINGS INCLUDE UTILITY, PLANT, AND REISSUE PATENTS

FIGURE 3

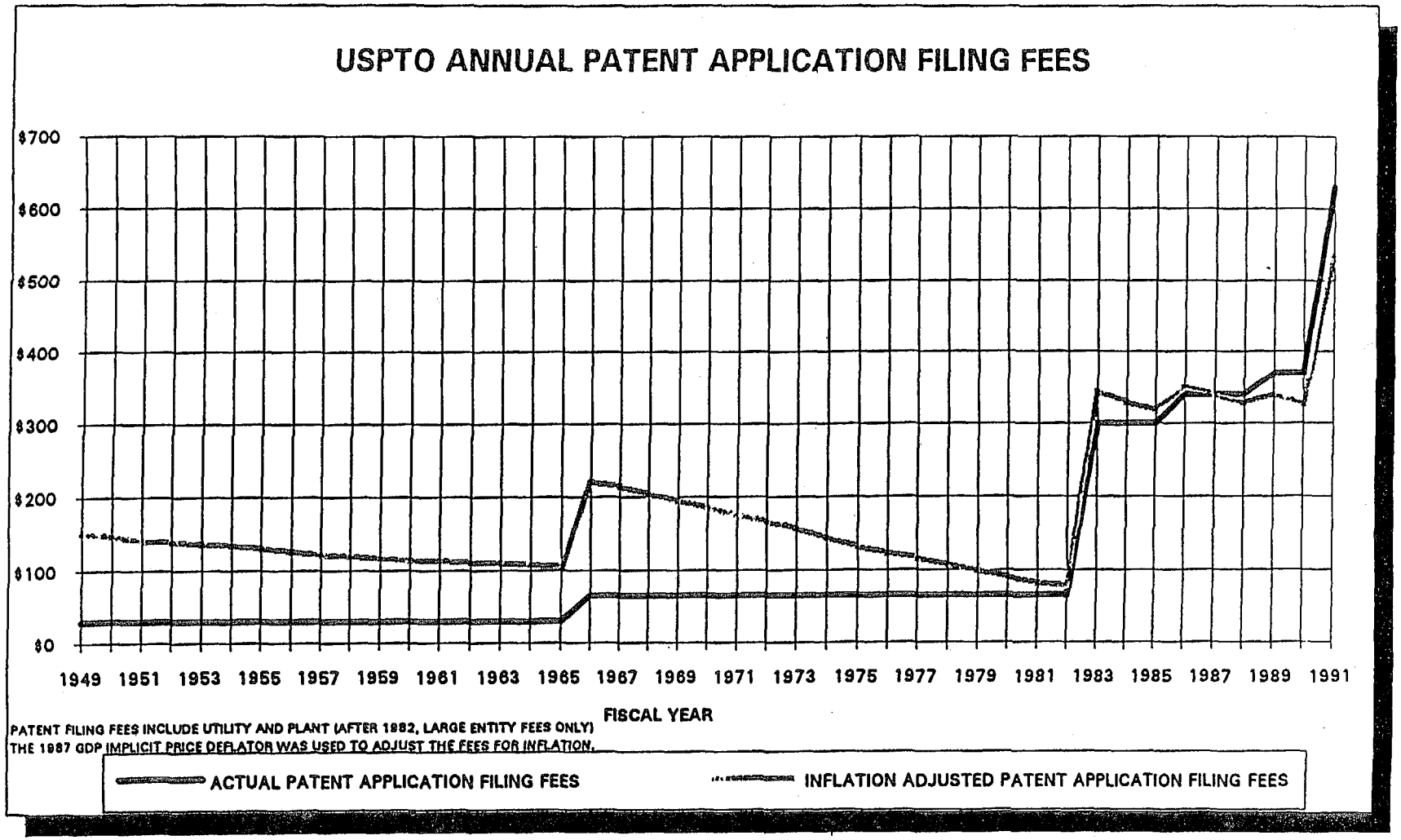

USPTO ANNUAL PATENT APPLICATION FILING FEES
$700
$600
$500
$400
$300
$200
$100
$0
1949 1951 1953 1955 1957 1959 1961 1963 1965 1967 1969 1971 1973 1975 1977 1979 1981 1983 1985 1987 1989 1991
FISCAL YEAR
PATENT FILING FEES INCLUDE UTILITY AND PLANT (AFTER 1982, LARGE ENTITY FEES ONLY)
THE 1987 GDP IMPLICIT PRICE DEFLATOR WAS USED TO ADJUST THE FEES FOR INFLATION.
ACTUAL PATENT APPLICATION FILING FEES
INFLATION ADJUSTED PATENT APPLICATION FILING FEES

FIGURE 4

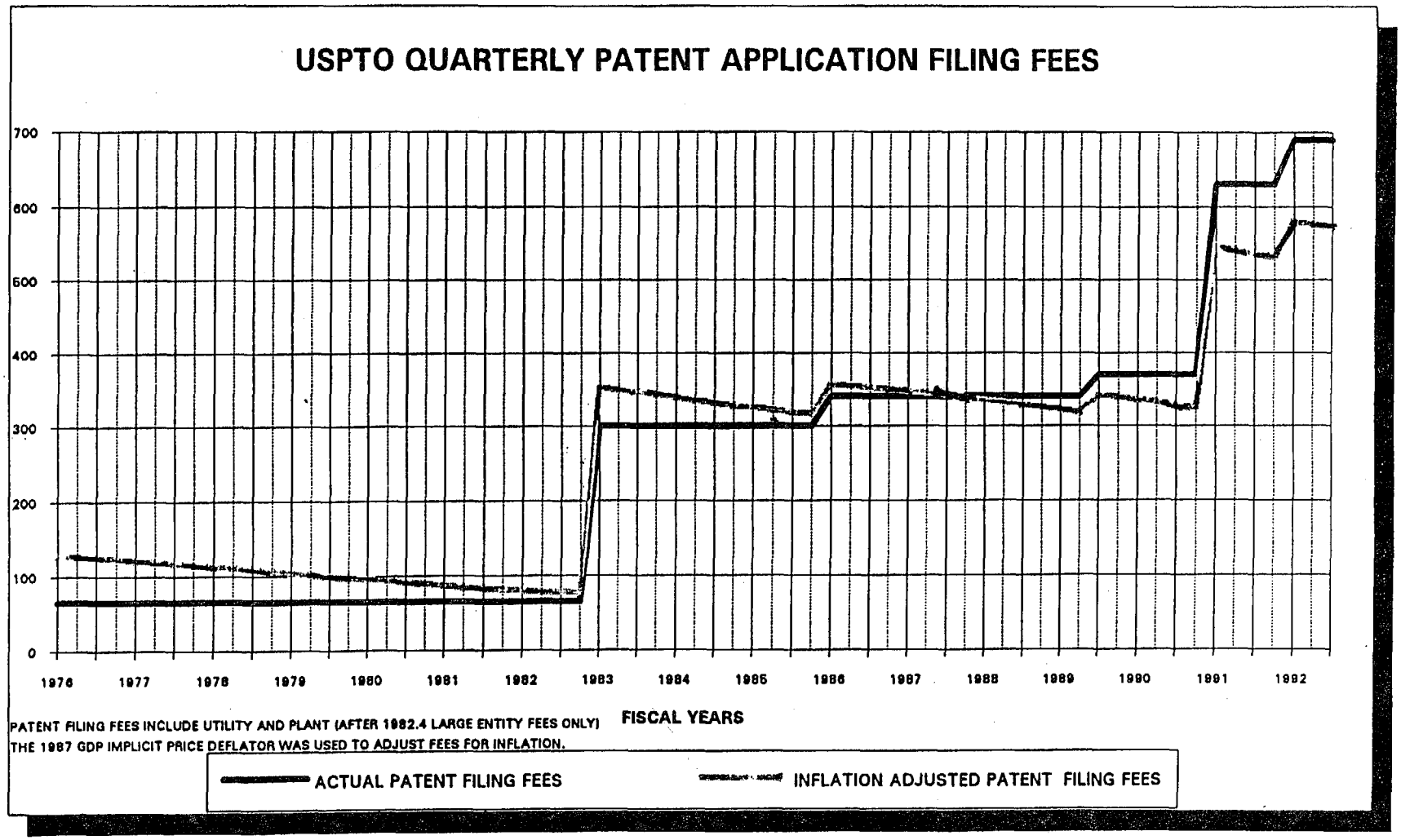

USPTO QUARTERLY PATENT APPLICATION FILING FEES
700
600
500
400
300
200
100
0
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
FISCAL YEARS
PATENT FILING FEES INCLUDE UTILITY AND PLANT (AFTER 1982.4 LARGE ENTITY FEES ONLY)
THE 1987 GDP IMPLICIT PRICE DEFLATOR WAS USED TO ADJUST FEES FOR INFLATION.
ACTUAL PATENT FILING FEES
INFLATION ADJUSTED PATENT FILING FEES

FIGURE 5

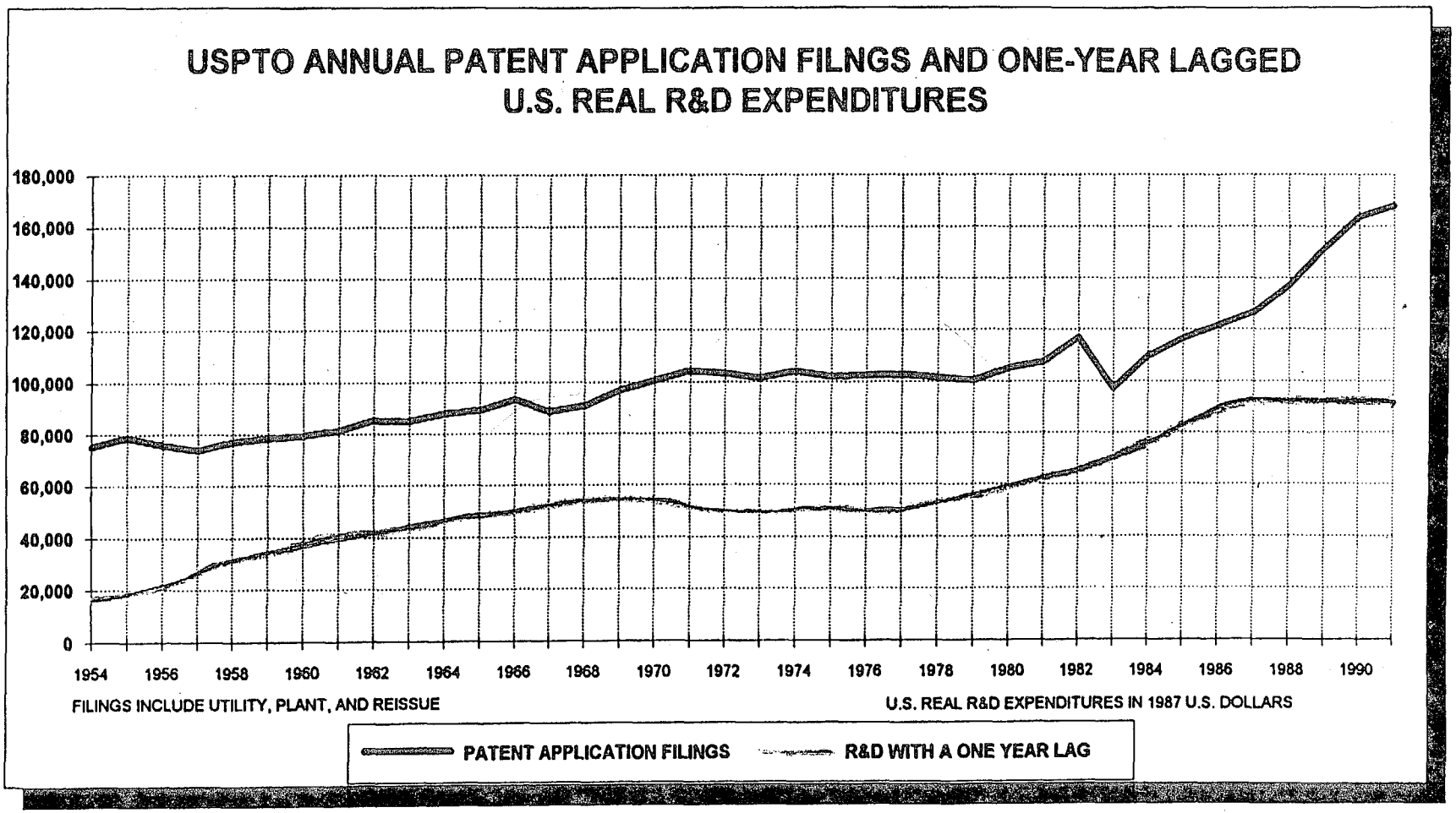

USPTO ANNUAL PATENT APPLICATION FILNGS AND ONE-YEAR LAGGED
U.S. REAL R&D EXPENDITURES
180,000
160,000
140,000
120,000
100,000
80,000
60,000
40,000
20,000
0
1954
1956
1958
1960
1962
1964
1966
1968
1970
1972
1974
1976
1978
1980
1982
1984
1986
1988
1990
FILINGS INCLUDE UTILITY, PLANT, AND REISSUE
U.S. REAL R&D EXPENDITURES IN 1987 U.S. DOLLARS
PATENT APPLICATION FILINGS
R&D WITH A ONE YEAR LAG

FIGURE 6

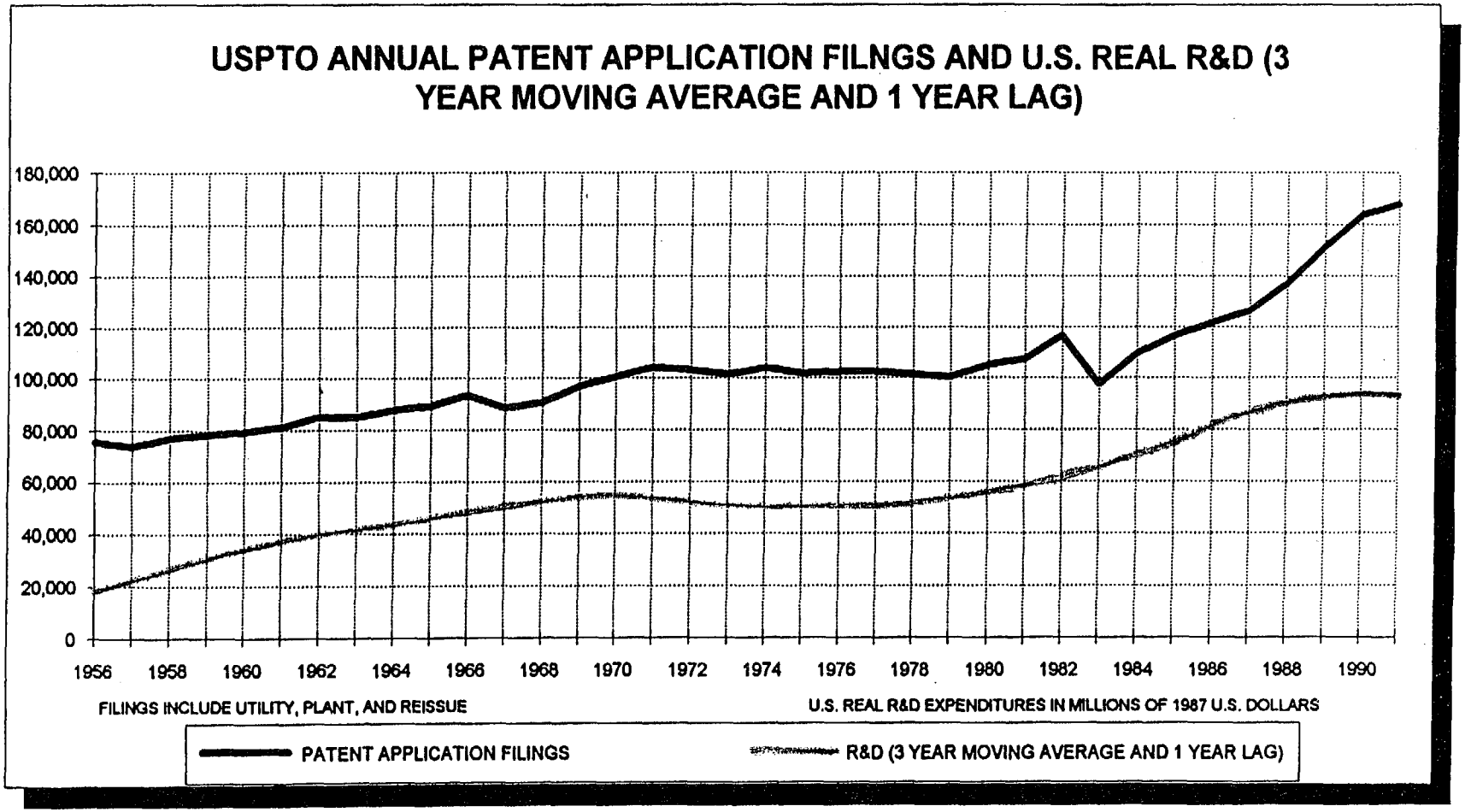

USPTO ANNUAL PATENT APPLICATION FILNGS AND U.S. REAL R&D (3 YEAR MOVING AVERAGE AND 1 YEAR LAG)
180,000
160,000
140,000
120,000
100,000
80,000
60,000
40,000
20,000
0
1956
1958
1960
1962
1964
1966
1968
1970
1972
1974
1976
1978
1980
1982
1984
1986
1988
1990
FILINGS INCLUDE UTILITY, PLANT, AND REISSUE
U.S. REAL R&D EXPENDITURES IN MILLIONS OF 1987 U.S. DOLLARS
PATENT APPLICATION FILINGS
R&D (3 YEAR MOVING AVERAGE AND 1 YEAR LAG)

FIGURE 7

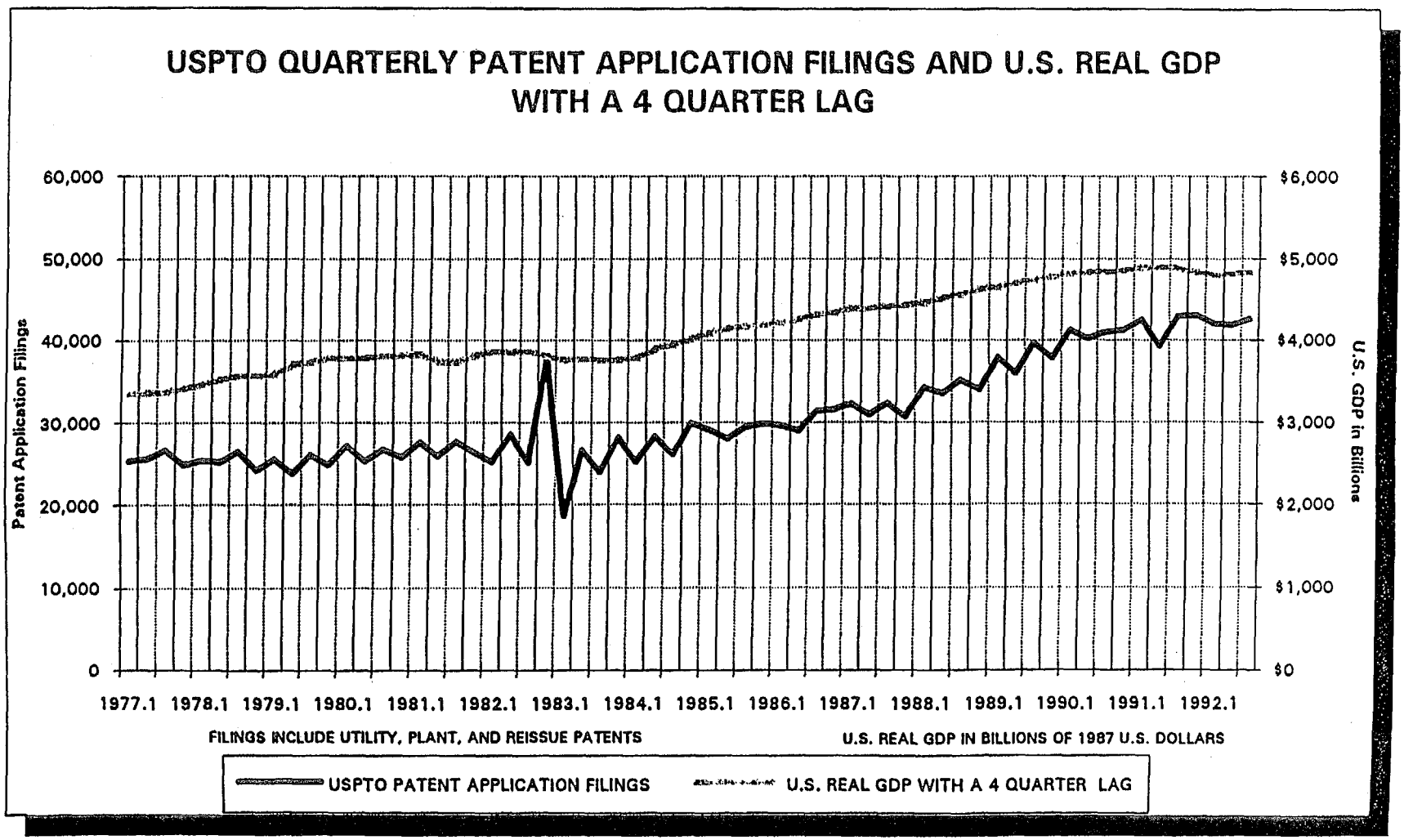

USPTO QUARTERLY PATENT APPLICATION FILINGS AND U.S. REAL GDP
WITH A 4 QUARTER LAG
Patent Application Filings
60,000
50,000
40,000
30,000
20,000
10,000
0
U.S. GDP in Billions
$6,000
$5,000
$4,000
$3,000
$2,000
$1,000
$0
1977.1 1978.1 1979.1 1980.1 1981.1 1982.1 1983.1 1984.1 1985.1 1986.1 1987.1 1988.1 1989.1 1990.1 1991.1 1992.1
FILINGS INCLUDE UTILITY, PLANT, AND REISSUE PATENTS
U.S. REAL GDP IN BILLIONS OF 1987 U.S. DOLLARS
USPTO PATENT APPLICATION FILINGS
U.S. REAL GDP WITH A 4 QUARTER LAG

FIGURE 8

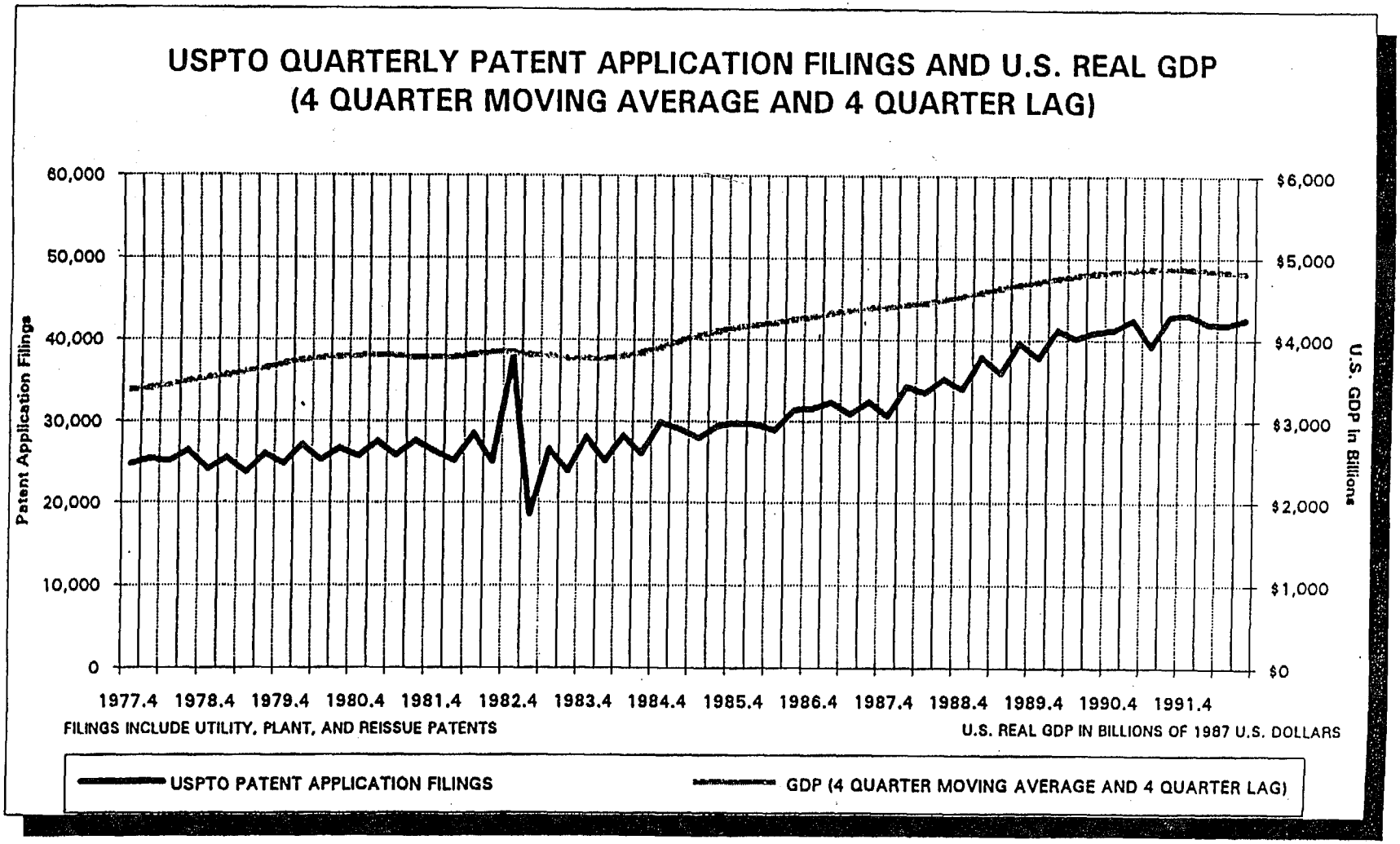

USPTO QUARTERLY PATENT APPLICATION FILINGS AND U.S. REAL GDP
(4 QUARTER MOVING AVERAGE AND 4 QUARTER LAG)
Patent Application Filings
60,000
50,000
40,000
30,000
20,000
10,000
0
U.S. GDP in Billions
$6,000
$5,000
$4,000
$3,000
$2,000
$1,000
$0
1977.4 1978.4 1979.4 1980.4 1981.4 1982.4 1983.4 1984.4 1985.4 1986.4 1987.4 1988.4 1989.4 1990.4 1991.4
FILINGS INCLUDE UTILITY, PLANT, AND REISSUE PATENTS
U.S. REAL GDP IN BILLIONS OF 1987 U.S. DOLLARS
USPTO PATENT APPLICATION FILINGS
GDP (4 QUARTER MOVING AVERAGE AND 4 QUARTER LAG)

TABLE 1

U.S. ANNUAL AND QUARTERLY ECONOMIC AND DEMOGRAPHIC DATA

ANNUAL ECONOMIC AND DEMOGRAPHIC INDICATORS	CALENDAR YEARS	BASE YEAR/CURRENCY	SOURCE
Real Gross Domestic Product in Billions	1959 - 1992	1987 U.S. Dollars	Citibase Database Services
Research and Development Expenditures in Billions	1953 - 1991	1987 U.S. Dollars	National Science Foundation
Scientists and Engineers Engaged in R&D in Thousands	1965 - 1988	-	National Science Foundation
Graduates in Science, Mathematics, and Engineering	1969 - 1985	-	U.S. Department of Education
Real Gross Private Domestic Investment in Billions	1959 - 1992	1987 U.S. Dollars	Citibase Database Services
10 Year Government Bond Interest Rates	1949 - 1992	-	Citibase Database Services
Business Formations	1963 - 1989	-	Dun and Bradstreet Corporation
1987 GDP Implicit Price Deflator	1959 - 1992	1987 U.S. Dollars	Citibase Database Services

QUARTERLY ECONOMIC AND DEMOGRAPHIC INDICATORS	QUARTERS	BASE YEAR/CURRENCY	SOURCE
Real Gross Domestic Product in Billions	1976.1-1992.3	1987 U.S. Dollars	Citibase Database Services
Research and Development Expenditures in Billions	NA	-	NA
Scientists and Engineers Engaged in R&D in Thousands	NA	-	NA
Graduates in Science, Mathematics, and Engineering	NA	-	NA
Real Gross Private Domestic Investment in Billions	1976.1 - 1992.3	1987 U.S. Dollars	Citibase Database Services
10 Year Government Bond Interest Rates	1976.1 - 1992.3	-	Citibase Database Services
Business Formations	NA	-	NA
GDP Implicit Price Deflator	1976.1 - 1992.3	1987 U.S. Dollars	Citibase Database Services

TABLE 2

NAIVE MODEL FOR ANNUAL PATENT APPLICATION FILINGS
HOLT EXPONENTIAL SMOOTHING: LINEAR TREND,
NO SEASONALITY
FORECAST PRO FOR WINDOWS VERSION 1.00B
SAMPLE PERIOD 1949 - 1992

Component	Smoothing Weight	Final Value
Level	0.837	1.7359e+005
Trend	0.193	6858.2

Standard Diagnostics

Sample size 44	Number of parameters 2
Mean 9.957e+004	Standard deviation 2.682e+004
R-square 0.9566	Adjusted R-square 0.9555
Durbin-Watson 1.974	Ljung-Box(18) = 12.69 P = 0.1902
Forecast error 5657	BIC 6023 (Best so far)
MAPE 0.0413	RMSE 5527
MAD 4007	

TABLE 3

ANNUAL PATENT APPLICATION FILINGS AUTOCORRELOGRAM
SAMPLE PERIOD 1956-1991

FILINGS IN NATURAL LOGARITHMS

MEAN = 11.526
VARIANCE = 0.41263E-01
STANDARD DEV. = 0.20313

LAGS	AUTOCORRELATIONS												STD ERR
1 -12	0.85	0.69	0.56	0.46	0.36	0.28	0.22	0.17	0.17	0.10	0.07	0.02	0.17
13 -24	0.00	-.01	-.02	-.02	-.04	-.06	-.08	-.11	-.16	-.20	-.24	-.26	0.38

LAGS	PARTIAL AUTOCORRELATIONS												STD ERR
1 -12	0.85	-.08	-.02	0.02	-.03	-.03	0.02	0.03	0.09	-.21	0.08	-.07	0.17

FILINGS IN FIRST DIFFERENCES OF NATURAL LOGARITHMS

MEAN = 0.22716E-01
VARIANCE = 0.26228E-02
STANDARD DEV. = 0.51214E-01

LAGS	AUTOCORRELATIONS												STD ERR
1 -12	-.19	0.09	-.02	0.18	0.02	-.06	-.11	0.13	-.14	0.05	-.01	-.11	0.17
13 -24	-.11	-.14	-.04	0.19	-.19	0.05	0.01	0.12	-.09	-.02	-.02	0.03	0.19

LAGS	PARTIAL AUTOCORRELATIONS												STD ERR
1 -12	-.19	0.05	0.01	0.19	0.09	-.07	-.16	0.05	-.11	0.03	0.09	-.14	0.17

MODIFIED BOX-PIERCE (LJUNG-BOX-PIERCE) STATISTICS (CHI-SQUARE)

LAG	Q	DF	P-VALUE	LAG	Q	DF	P-VALUE
1	1.36	1	.243	13	7.11	13	.896
2	1.65	2	.437	14	8.40	14	.867
3	1.67	3	.644	15	8.52	15	.901
4	3.09	4	.543	16	10.92	16	.814
5	3.11	5	.684	17	13.57	17	.697
6	3.25	6	.777	18	13.78	18	.743
7	3.86	7	.795	19	13.78	19	.796
8	4.62	8	.798	20	15.11	20	.770
9	5.55	9	.784	21	15.88	21	.776
10	5.69	10	.840	22	15.91	22	.821
11	5.70	11	.893	23	15.94	23	.858
12	6.43	12	.893	24	16.02	24	.887

TABLE 4

ANNUAL PATENT APPLICATION FILINGS ARIMA (1,1,0) MODEL

```
 R-SQUARE =  0.0368      R-SQUARE ADJUSTED =  0.0077
VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.26001E-02
STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.50991E-01
AKAIKE INFORMATION CRITERIA -AIC(K) =  -5.8379
SCHWARZ CRITERIA- SC(K) =  -5.7490
              PARAMETER ESTIMATES        STD ERROR     T-STAT
              AR( 1)  -0.19454          0.1684        -1.155
            CONSTANT   0.27412E-01      0.9482E-02     2.891

                              RESIDUALS
  LAGS                     AUTOCORRELATIONS                                STD ERR
  1 -12    0.02 0.05 0.04 0.20 0.05 -.08 -.11 0.08 -.11 0.03 -.02 -.15    0.17

 MODIFIED BOX-PIERCE (LJUNG-BOX-PIERCE) STATISTICS  (CHI-SQUARE)
    LAG    Q     DF  P-VALUE
     2    0.11   1  .741
     3    0.16   2  .923
     4    1.82   3  .611
     5    1.91   4  .751
     6    2.21   5  .819
     7    2.77   6  .837
     8    3.11   7  .874
     9    3.71   8  .882
    10    3.75   9  .927
    11    3.77  10  .957
    12    4.97  11  .933

              ANALYSIS OF RESIDUALS
              VALUES RANGE FROM      -0.1920 TO       0.0873
   SAMPLE MOMENTS OF RESIDUALS (USING THE DIVISOR  35) :
   MEAN =                -0.2351045E-03
   VARIANCE =             0.2451460E-02
   SKEWNESS =            -1.371450
   KURTOSIS =             7.257485
   STUDENTIZED RANGE =    5.640804
```

RSS(ARIMA(1,1,0)) = 0.08589 and RSS(ARIMA(0,1,0)) = 0.09879

$$F\text{-}test = \left[\frac{(0.09878-0.08589)}{0.08589}\right] * \frac{35}{1} = 5.25$$

critical values:
$F(0.05,1,30)=4.17$
$F(0.01,1,30)=7.56$

TABLE 5

SPECIFICATION TESTS FOR LAG LENGTH IN FINAL AUTOREGRESSIVE DISTRIBUTED LAG MODEL

Sample 1959-1991

FIL	RD	FEE	AIC	BSC	$\bar{R}^2$
UNRESTRICTED MODELS STARTING WITH FOUR LAGS					
1-4	1-4	0-4	-6.85	-6.19	0.9750
1-3	1-3	0-3	-6.90	-6.38	0.9759
1-2	1-2	0-2	-6.91	-6.54	0.9758
1-1	1-1	0-1	-7.02	-6.80	0.9778
RESTRICTING LAGS ON FILINGS TO ONE					
1-1	1-4	0-4	-6.93	-6.42	0.9764
1-1	1-3	0-3	-6.92	-6.49	0.9754
1-1	1-2	0-2	-6.64	-6.97	0.9767
RESTRICTING LAGS ON FILINGS AND FEES TO ONE					
1-1	1-4	0-1	-7.01	-6.63	0.9768
1-1	1-3	0-1	-6.96	-6.63	0.9756
1-1	1-2	0-1	-7.00	-6.72	0.9769

TABLE 6.
ANNUAL PATENT APPLICATION FILINGS ECONOMETRIC MODEL
SAMPLE 1959 - 1991

R-SQUARE = 0.9807 R-SQUARE ADJUSTED = 0.9778
VARIANCE OF THE ESTIMATE-SIGMA**2 = 0.77181E-03
STANDARD ERROR OF THE ESTIMATE-SIGMA = 0.27781E-01
SUM OF SQUARED ERRORS-SSE= 0.20839E-01
MEAN OF DEPENDENT VARIABLE = 11.561
LOG OF THE LIKELIHOOD FUNCTION(IF DEPVAR LOG) = -297.984
AKAIKE (1973) INFORMATION CRITERION- LOG AIC = -7.0242
SCHWARZ(1978) CRITERION-LOG SC = -6.7951

VARIABLE	SUM OF LAG COEFS	STD ERROR	T-RATIO	MEAN LAG
LFIL	0.81199	0.72516E-01	11.197	1.0000
ARD	0.17632	0.55689E-01	3.1662	1.0000
LP	0.10082E-01	0.18445E-01	0.54661	13.160

VARIABLE NAME	ESTIMATED COEFFICIENT	STANDARD ERROR	T-RATIO 27 DF	P-VALUE	PARTIAL CORR.	STANDARDIZED COEFFICIENT	ELASTICITY AT MEANS
LFIL	0.81199	0.7252E-01	11.20	1.000	0.907	0.7524	0.8120
ARD	0.17632	0.5569E-01	3.166	0.998	0.520	0.2606	0.1763
LP	-0.12260	0.1911E-01	-6.415	0.000	-0.777	-0.3384	-0.1226
LP	0.13268	0.1953E-01	6.793	1.000	0.794	0.3394	0.1327
CONSTANT	0.21665	0.4433	0.4887	0.686	0.094	0.0000	0.2166

DURBIN-WATSON = 1.6148 VON NEUMANN RATIO = 1.6669 RHO = 0.13653
RESIDUAL SUM = 0.12471E-12 RESIDUAL VARIANCE = 0.77181E-03
RUNS TEST: 19 RUNS, 15 POSITIVE, 17 NEGATIVE, NORMAL STATISTIC = 0.7443
DURBIN H STATISTIC (ASYMPTOTIC NORMAL) = 0.84684
COEFFICIENT OF SKEWNESS = 0.1109 WITH STANDARD DEVIATION OF 0.4145
COEFFICIENT OF EXCESS KURTOSIS = -1.0189 WITH STANDARD DEVIATION OF 0.8094

GOODNESS OF FIT TEST FOR NORMALITY OF RESIDUALS - 10 GROUPS

OBSERVED	0.0	0.0	3.0	7.0	7.0	5.0	6.0	4.0	0.0	0.0
EXPECTED	0.3	0.9	2.5	5.1	7.2	7.2	5.1	2.5	0.9	0.3

CHI-SQUARE = 4.7952 WITH 3 DEGREES OF FREEDOM

JARQUE-BERA ASYMPTOTIC LM NORMALITY TEST
CHI-SQUARE = 1.5250 WITH 2 DEGREES OF FREEDOM

HETEROSKEDASTICITY TESTS
E**2 ON YHAT: CHI-SQUARE = 4.268 WITH 1 D.F.
E**2 ON YHAT**2: CHI-SQUARE = 4.317 WITH 1 D.F.
E**2 ON LOG(YHAT**2): CHI-SQUARE = 4.218 WITH 1 D.F.
E**2 ON X (B-P-G) TEST: CHI-SQUARE = 5.685 WITH 4 D.F.
E**2 ON LAG(E**2) ARCH TEST: CHI-SQUARE = 1.356 WITH 1 D.F.
LOG(E**2) ON X (HARVEY) TEST: CHI-SQUARE = 1.519 WITH 4 D.F.
ABS(E) ON X (GLEJSER) TEST: CHI-SQUARE = 2.852 WITH 4 D.F.

TABLE 6 (continued)

RESIDUAL CORRELOGRAM
LM-TEST FOR HJ:RHO(J)=0, STATISTIC IS STANDARD NORMAL

LAG	RHO	STD ERR	T-STAT	LM-STAT	DW-TEST	BOX-PIERCE-LJUNG
1	0.1206	0.1768	0.6823	0.8316	1.6148	0.5105
2	0.3191	0.1768	1.8050	2.1196	1.1693	4.2028
3	-0.0970	0.1768	-0.5487	0.7566	1.8593	4.5557
4	-0.0010	0.1768	-0.0054	0.0077	1.6630	4.5558
5	0.0035	0.1768	0.0196	0.0245	1.5407	4.5562
6	-0.0574	0.1768	-0.3245	0.4091	1.6414	4.6940
7	0.1388	0.1768	0.7851	1.0700	1.2162	5.5322
8	0.0629	0.1768	0.3557	0.5152	1.3182	5.7114
9	0.0571	0.1768	0.3231	0.5412	1.2601	5.8657

LM CHI-SQUARE STATISTIC WITH 9 D.F. IS 4.977

SEQUENTIAL CHOW AND GOLDFELD-QUANDT TESTS

N1	N2	SSE1	SSE2	CHOW	G-Q	DF1	DF2
6	26	0.15528E-03	0.15612E-01	1.4152	0.20886	1	21
7	25	0.18658E-03	0.14608E-01	1.7975	0.12772	2	20
8	24	0.31967E-03	0.13292E-01	2.3364	0.15232	3	19
9	23	0.48792E-03	0.13277E-01	2.2610	0.16537	4	18
10	22	0.17435E-02	0.12001E-01	2.2711	0.49396	5	17
11	21	0.21412E-02	0.11672E-01	2.2378	0.48919	6	16
12	20	0.22545E-02	0.11033E-01	2.5007	0.43789	7	15
13	19	0.25666E-02	0.10965E-01	2.3759	0.40961	8	14
14	18	0.27004E-02	0.10842E-01	2.3707	0.35976	9	13
15	17	0.31733E-02	0.99995E-02	2.5607	0.38082	10	12
16	16	0.35858E-02	0.96380E-02	2.5338	0.37205	11	11
17	15	0.35861E-02	0.96378E-02	2.5338	0.31007	12	10
18	14	0.35901E-02	0.96316E-02	2.5349	0.25805	13	9
19	13	0.37930E-02	0.91410E-02	2.6892	0.23711	14	8
20	12	0.41664E-02	0.67908E-02	3.9682	0.28632	15	7
21	11	0.47074E-02	0.66937E-02	3.6423	0.26372	16	6
22	10	0.47109E-02	0.44279E-02	5.6332	0.31291	17	5
23	9	0.67747E-02	0.42261E-02	3.9349	0.35624	18	4
24	8	0.69278E-02	0.41767E-02	3.8572	0.26190	19	3
25	7	0.73916E-02	0.86829E-03	6.7008	0.85128	20	2
26	6	0.74613E-02	0.16656E-05	7.8861	213.32	21	1

CHOW TEST - F DISTRIBUTION WITH DF1= 5 AND DF2= 22

TABLE 7
QUARTERLY PATENT APPLICATION FILINGS NAIVE MODEL
SAMPLE 1976Q1 - 1992Q3

R-SQUARE = 0.9553 R-SQUARE ADJUSTED = 0.9522
VARIANCE OF THE ESTIMATE-SIGMA**2 = 0.18754E-02
STANDARD ERROR OF THE ESTIMATE-SIGMA = 0.43306E-01
SUM OF SQUARED ERRORS-SSE= 0.10877
MEAN OF DEPENDENT VARIABLE = 10.321
LOG OF THE LIKELIHOOD FUNCTION(IF DEPVAR LOG) = -539.228
AKAIKE (1973) INFORMATION CRITERION- LOG AIC = -6.2029
SCHWARZ(1978) CRITERION-LOG SC = -6.0328

VARIABLE	SUM OF LAG COEFS	STD ERROR	T-RATIO	MEAN LAG
LFIL	1.0324	0.31247E-01	33.040	2.1333

VARIABLE NAME	ESTIMATED COEFFICIENT	STANDARD ERROR	T-RATIO 58 DF	P-VALUE	PARTIAL CORR.	STANDARDIZED COEFFICIENT	ELASTICITY AT MEANS
LFIL	0.36832	0.5016E-01	7.343	1.000	0.694	0.3613	0.3683
LFIL	0.41114	0.6631E-01	6.200	1.000	0.631	0.3943	0.4111
LFIL	0.25296	0.6916E-01	3.657	1.000	0.433	0.2335	0.2530
SURGE	0.39772	0.3298E-01	12.06	1.000	0.846	0.3606	0.3977
CONSTANT	-0.31758	0.3219	-0.9866	0.164	-0.128	0.0000	-0.3176

DURBIN-WATSON = 2.1277 VON NEUMANN RATIO = 2.1620 RHO = -0.08533
RESIDUAL SUM = -0.76782E-13 RESIDUAL VARIANCE = 0.18754E-02
SUM OF ABSOLUTE ERRORS= 2.1372
R-SQUARE BETWEEN OBSERVED AND PREDICTED = 0.9553
R-SQUARE BETWEEN ANTILOGS OBSERVED AND PREDICTED = 0.9590
RUNS TEST: 39 RUNS, 32 POSITIVE, 31 NEGATIVE, NORMAL STATISTIC = 1.6537
DURBIN H STATISTIC (ASYMPTOTIC NORMAL) = -0.73832
COEFFICIENT OF SKEWNESS = -0.2209 WITH STANDARD DEVIATION OF 0.3016
COEFFICIENT OF EXCESS KURTOSIS = -0.6058 WITH STANDARD DEVIATION OF 0.5948
JARQUE-BERA ASYMPTOTIC LM NORMALITY TEST
CHI-SQUARE = 1.6056 WITH 2 DEGREES OF FREEDOM

HETEROSKEDASTICITY TESTS
E**2 ON YHAT: CHI-SQUARE = 0.823 WITH 1 D.F.
E**2 ON YHAT**2: CHI-SQUARE = 0.814 WITH 1 D.F.
E**2 ON LOG(YHAT**2): CHI-SQUARE = 0.833 WITH 1 D.F.
E**2 ON X (B-P-G) TEST: CHI-SQUARE = 0.924 WITH 4 D.F.
E**2 ON LAG(E**2) ARCH TEST: CHI-SQUARE = 2.636 WITH 1 D.F.
LOG(E**2) ON X (HARVEY) TEST: CHI-SQUARE = 1.555 WITH 4 D.F.
ABS(E) ON X (GLEJSER) TEST: CHI-SQUARE = 1.098 WITH 4 D.F.

TABLE 8

QUARTERLY PATENT APPLICATION FILINGS AUTOCORRELOGRAM
SAMPLE PERIOD 1976Q1-1992Q3

FILINGS IN NATURAL LOGARITHMS

MEAN = 10.311
VARIANCE = 0.39003E-01
STANDARD DEV. = 0.19749

LAGS	AUTOCORRELATIONS												STD ERR
1 -12	0.75	0.84	0.71	0.76	0.65	0.66	0.58	0.55	0.51	0.45	0.42	0.35	0.12
13 -24	0.33	0.25	0.25	0.19	0.20	0.11	0.11	0.05	0.06	-.01	0.00	-.07	0.39

LAGS	PARTIAL AUTOCORRELATIONS												STD ERR
1 -12	0.75	0.62	0.03	0.11	-.07	-.06	-.04	-.09	0.01	-.09	-.01	-.06	0.12

FILINGS IN FIRST DIFFERENCES OF NATURAL LOGARITHMS

MEAN = 0.71013E-02
VARIANCE = 0.17626E-01
STANDARD DEV. = 0.13276

LAGS	AUTOCORRELATIONS												STD ERR
1 -12	-.77	0.49	-.39	0.32	-.22	0.16	-.11	0.04	0.03	-.07	0.10	-.12	0.12
13 -24	0.16	-.18	0.16	-.17	0.21	-.19	0.17	-.19	0.19	-.18	0.17	-.19	0.23

LAGS	PARTIAL AUTOCORRELATIONS												STD ERR
1 -12	-.77	-.26	-.31	-.15	-.01	0.04	0.10	0.00	0.08	0.00	0.05	0.00	0.12

MODIFIED BOX-PIERCE (LJUNG-BOX-PIERCE) STATISTICS (CHI-SQUARE)

LAG	Q	DF	P-VALUE	LAG	Q	DF	P-VALUE
1	40.96	1	.000	13	86.43	13	.000
2	57.57	2	.000	14	89.17	14	.000
3	68.31	3	.000	15	91.32	15	.000
4	75.79	4	.000	16	93.79	16	.000
5	79.23	5	.000	17	97.81	17	.000
6	81.04	6	.000	18	101.12	18	.000
7	81.89	7	.000	19	103.92	19	.000
8	82.01	8	.000	20	107.26	20	.000
9	82.07	9	.000	21	110.91	21	.000
10	82.45	10	.000	22	114.11	22	.000
11	83.23	11	.000	23	117.14	23	.000
12	84.40	12	.000	24	121.04	24	.000

TABLE 9
QUARTERLY PATENT APPLICATION FILINGS ARIMA(1,1,1) MODEL

R-SQUARE = 0.6673 R-SQUARE ADJUSTED = 0.6568
VARIANCE OF THE ESTIMATE-SIGMA**2 = 0.59744E-02
STANDARD ERROR OF THE ESTIMATE-SIGMA = 0.77294E-01
AKAIKE INFORMATION CRITERIA -AIC(K) = -5.0294
SCHWARZ CRITERIA- SC(K) = -4.9298

PARAMETER	ESTIMATES	STD ERROR	T-STAT
AR(1)	-0.60018	0.1144	-5.248
MA(1)	0.54758	0.1210	4.525
CONSTANT	0.13219E-01	0.4533E-02	2.916

RESIDUALS

LAGS	AUTOCORRELATIONS												STD ERR
1 -12	0.04	-.08	-.08	0.17	0.08	0.04	-.05	0.02	0.06	-.01	0.04	0.00	0.12

MODIFIED BOX-PIERCE (LJUNG-BOX-PIERCE) STATISTICS (CHI-SQUARE)

LAG	Q	DF	P-VALUE
3	1.04	1	.307
4	3.18	2	.204
5	3.64	3	.303
6	3.78	4	.437
7	3.95	5	.557
8	3.97	6	.681
9	4.21	7	.755
10	4.23	8	.836
11	4.38	9	.885
12	4.38	10	.929

ANALYSIS OF RESIDUALS

VALUES RANGE FROM -0.3171 TO 0.2922

SAMPLE MOMENTS OF RESIDUALS (USING THE DIVISOR 66) :

MEAN = -0.5406873E-03
VARIANCE = 0.5702537E-02
SKEWNESS = -0.7533300
KURTOSIS = 10.25777
STUDENTIZED RANGE = 8.069505

TABLE 10

SPECIFICATION TESTS FOR LAG LENGTH IN AUTOREGRESSIVE DISTRIBUTED LAG MODEL

Sample 1976Q1-1992Q3

FIL	AGDP	FEE	AIC	BSC	$\bar{R}^2$
UNRESTRICTED MODELS STARTING WITH FOUR LAGS					
4,8,12,16	4,8,12,16	0-1	14.550	15.005	0.9575
4,8,12	4,8,12	0-1	14.525	14.890	0.9576
4,8	4,8	0-1	14.402	14.684	0.9610
4	4	0-1	15.376	15.580	0.8913
RESTRICTING LAGS ON FILINGS TO THREE					
4,8,12	4,8,12,16	0-1	14.523	14.939	0.9581
4,8,12	4,8,12	0-1	14.492	14.820	0.9584
4,8,12	4	0-1	14.456	14.748	0.9593
RESTRICTING LAGS ON FILINGS TO TWO					
4,8	4,8,12,16	0-1	14.484	14.863	0.9591
4,8	4,8,12	0-1	14.495	14.824	0.9583
4,8	4	0-1	14.495	14.741	0.9566
RESTRICTING LAGS ON FILINGS TO ONE					
4	4,8,12,16	0-1	14.483	14.824	0.9585
4	4,8,12	0-1	14.561	14.853	0.9548
4	4,8	0-1	14.504	14.750	0.9562

TABLE 11

QUARTERLY PATENT APPLICATION FILINGS ECONOMETRIC MODEL
SAMPLE 1976Q1 - 1992Q3

R-SQUARE = 0.9657 R-SQUARE ADJUSTED = 0.9610
VARIANCE OF THE ESTIMATE-SIGMA**2 = 0.15860E+07
STANDARD ERROR OF THE ESTIMATE-SIGMA = 1259.4
SUM OF SQUARED ERRORS-SSE= 0.80886E+08
MEAN OF DEPENDENT VARIABLE = 31329.
LOG OF THE LIKELIHOOD FUNCTION = -500.582
AKAIKE (1973) INFORMATION CRITERION- LOG AIC = 14.402
SCHWARZ(1978) CRITERION-LOG SC = 14.684

VARIABLE	SUM OF LAG COEFS	STD ERROR	T-RATIO	MEAN LAG
FIL	0.47237	0.70209E-01	6.7280	5.6077
AGDPQ	8.8869	0.87921	10.108	3.4866
ERFEE	-2.2129	1.9548	-1.1320	1.4223

VARIABLE NAME	ESTIMATED COEFFICIENT	STANDARD ERROR	T-RATIO 51 DF	P-VALUE	PARTIAL CORR.	STANDARDIZED COEFFICIENT	ELASTICITY AT MEANS
FIL	0.28252	0.7324E-01	3.858	1.000	0.475	0.2554	0.2723
FIL	0.18986	0.6984E-01	2.718	0.996	0.356	0.1485	0.1765
AGDPQ	10.028	1.075	9.331	1.000	0.794	0.7252	1.3281
AGDPQ	-1.1407	0.4352	-2.621	0.006	-0.345	-0.1257	-0.1443
ERFEE	0.93453	2.062	0.4532	0.674	0.063	0.0227	0.0083
ERFEE	-3.1474	2.620	-1.201	0.118	-0.166	-0.0749	-0.0281
SURGE	9029.1	963.2	9.374	1.000	0.795	0.2629	0.0000
CONSTANT	-19197.	2350.	-8.168	0.000	-0.753	0.0000	-0.6128

DURBIN-WATSON = 2.2698 VON NEUMANN RATIO = 2.3089 RHO = -0.14273
RESIDUAL SUM = -0.76056E-10 RESIDUAL VARIANCE = 0.15860E+07
RUNS TEST: 32 RUNS, 31 POSITIVE, 28 NEGATIVE, NORMAL STATISTIC = 0.4151
DURBIN H STATISTIC (ASYMPTOTIC NORMAL) = -1.3261
COEFFICIENT OF SKEWNESS = -0.3554 WITH STANDARD DEVIATION OF 0.3112
COEFFICIENT OF EXCESS KURTOSIS = 0.0721 WITH STANDARD DEVIATION OF 0.6133

GOODNESS OF FIT TEST FOR NORMALITY OF RESIDUALS - 12 GROUPS

OBSERVED	1.0	1.0	3.0	3.0	8.0	12.0	11.0	12.0	7.0	0.0	1.0	0.0
EXPECTED	0.4	1.0	2.6	5.4	8.8	11.3	11.3	8.8	5.4	2.6	1.0	0.4

CHI-SQUARE = 6.9283 WITH 2 DEGREES OF FREEDOM

JARQUE-BERA ASYMPTOTIC LM NORMALITY TEST
CHI-SQUARE = 1.1824 WITH 2 DEGREES OF FREEDOM

HETEROSKEDASTICITY TESTS
E**2 ON YHAT: CHI-SQUARE = 0.161 WITH 1 D.F.
E**2 ON YHAT**2: CHI-SQUARE = 0.214 WITH 1 D.F.
E**2 ON LOG(YHAT**2): CHI-SQUARE = 0.116 WITH 1 D.F.
E**2 ON X (B-P-G) TEST: CHI-SQUARE = 4.440 WITH 7 D.F.
E**2 ON LAG(E**2) ARCH TEST: CHI-SQUARE = 0.007 WITH 1 D.F.
LOG(E**2) ON X (HARVEY) TEST: CHI-SQUARE = 3.168 WITH 7 D.F.
ABS(E) ON X (GLEJSER) TEST: CHI-SQUARE = 3.436 WITH 7 D.F.

TABLE 11 (continued)

RESIDUAL CORRELOGRAM
LM-TEST FOR HJ:RHO(J)=0, STATISTIC IS STANDARD NORMAL

LAG	RHO	STD ERR	T-STAT	LM-STAT	DW-TEST	BOX-PIERCE-LJUNG
1	-0.1415	0.1302	-1.0867	1.1731	2.2698	1.2419
2	0.2013	0.1302	1.5465	1.7207	1.5748	3.8015
3	-0.0294	0.1302	-0.2260	0.2428	2.0216	3.8572
4	0.1806	0.1302	1.3872	1.6586	1.5848	5.9914
5	-0.0819	0.1302	-0.6294	0.7070	2.0867	6.4388
6	0.1676	0.1302	1.2877	1.5290	1.5301	8.3474
7	-0.0864	0.1302	-0.6635	0.7395	2.0290	8.8637
8	0.0211	0.1302	0.1621	0.1881	1.8071	8.8952
9	-0.1311	0.1302	-1.0068	1.1457	2.0985	10.1318
10	-0.0724	0.1302	-0.5560	0.6596	1.9699	10.5166
11	0.0371	0.1302	0.2849	0.3321	1.7245	10.6198
12	-0.2013	0.1302	-1.5464	1.8709	2.2001	13.7235
13	-0.0476	0.1302	-0.3654	0.4322	1.8525	13.9006
14	-0.1110	0.1302	-0.8525	1.0121	1.9691	14.8857
15	-0.1405	0.1302	-1.0793	1.3003	2.0041	16.5008
16	0.0565	0.1302	0.4336	0.5269	1.5906	16.7675
17	-0.0252	0.1302	-0.1936	0.2347	1.7073	16.8220
18	-0.0283	0.1302	-0.2174	0.2646	1.6683	16.8923

LM CHI-SQUARE STATISTIC WITH 18 D.F. IS 14.162

TABLE 12

COMPARISON OF THE FORECAST PERFORMANCE OF THE ANNUAL MODELS

		NAIVE		ARIMA		ECONOMETRIC	
Year	Actual Level	1 Year	2 Year	1 Year	2 Year	1 Year	2 Year
1985	116,427	6,524		11,902		-1,180	
1986	121,611	8,432	10,629	5,574	12,043	-380	-1,470
1987	126,407	9,766	12,034	4,020	7,643	-2,778	-3,127
1988	137,069	13,764	19,125	9,444	12,145	4,229	1,857
1989	151,331	15,980	26,550	14,627	21,094	12,085	15,455
1990	163,571	11,047	26,435	12,150	23,155	10,386	21,199
1991	167,715	-3,548	12,637	2,041	12,093	13,654	23,941
ME		8,852	17,902	8,537	14,696	5,145	9,643
MPE		6.50	12.24	6.17	10.02	3.15	5.85
RMSE		10,613	19,092	9,590	15,697	8,172	14,588
MAE		9,866	17,902	8,537	14,696	6,385	11,175
MAPE		7.10	12.24	6.17	10.02	4.16	7.08

TABLE 13

COMPARISON OF THE FORECAST PERFORMANCE OF THE QUARTERLY MODELS

		NAIVE		ARIMA		ECONOMETRIC	
Year	Actual Level	1 Year	2 Year	1 Year	2 Year	1 Year	2 Year
1985	116,427	7,905		-377		2,487	
1986	121,611	6,712	10,809	7,019	2,919	-1,112	100
1987	126,407	2,170	7,570	4,375	9,747	-1,181	-1,673
1988	137,069	8,759	9,352	8,147	12,378	5,795	3,288
1989	151,331	7,327	12,804	12,345	19,951	8,413	13,158
1990	163,571	-1,147	3,873	8,164	21,006	4,983	12,195
1991	167,715	-7,006	-8,679	-716	7,283	3,631	7,097
ME		3,531	5,955	5,565	12,214	3,288	5,694
MPE		2.91	4.56	3.94	8.25	2.18	3.59
RMSE		6,470	9,272	7,116	13,839	4,647	8,019
MAE		5,861	8,848	5,878	12,214	3,943	6,252
MAPE		4.31	6.28	4.15	8.25	2.71	4.03

References

Acs, Z. J. and D. B. Audretsch, "Patents as a Measure of Innovative Activity," *Kyklos*, vol. 42 (1989): p. 171-80.

Bosworth, D. and Westaway, T., "The Influence of demand and supply side pressure on the quantity and quality of inventive activity," *Applied Economics*, vol. 16 (1984): p. 131-146.

Box, George and Gwilym Jenkins, Time Series Analysis: Forecasting and Control, Holden-Day, Inc., San Francisco, 1970.

CITIBASE, 1992, *Citibank Economic Database* (1946-present), Citibank, New York, N.A.

Durbin, J. and Watson, G. S., "Testing for Serial Correlation in least-Squares Regression," *Biometrica,* vol. 38 (1971): p. 159-177.

Fair, Ray C., Specification, Estimation, and Analysis of Macroeconometric Models, 1984, Harvard University Press.

Granger, C.W.J. and Newbold, P., Forecasting Economic Time Series, Second Edition, Academic Press, Inc., 1986.

Griliches, Z., "Patents: Recent Trends and Puzzles," *Brookings Papers on Economic Activity: Microeconomics,* (1989): p. 291-330.

Griliches, Z., "Patent Statistics as Economic Indicators: A Survey," *Journal of Economic Literature*, vol. 27 (December, 1990): p. 1661-1707.

Hall, B., Griliches, Z., and Hausman, J., "Patents and R&D: Is There a Lag?" *International Economic Review,* vol. 27, no. 2 (1986): p. 265-83.

Holt, C. C., "Forecasting Seasonal and Trends by Exponentially Weighted Moving Averages." Office of Naval Research, Memorandum, no. 52, 1957.

Judge, G., Hill, Carter Hill R., Griffiths, W.E., Lutkepohl, H., and Lee, T.C., Introduction to the Theory and Practice of Econometrics, Second Edition, John Wiley and Son, 1988.

OECD "International Surveys of the Resources Devoted to R&D by OECD Member Countries." Directories for Science, Technology and Industry. Organization for Economic Cooperation and Development. Paris, 1987.

Pavitt, K., "R&D, Patenting and Innovative Activity: A Statistical Exploration," *Research Policy*, vol. 11, no. 1 (1982): p. 33-51.

Scherer, F.M., "The Propensity to Patent", International Journal of Industrial Organization (1983): p. 107-128.

Schmookler, J., "The Level of Inventive Activity," *Rev. Econ. Statist.,* vol. 36, no. 2 (1954):p. 183-190

Stellwagen, E. A. and Goodrich, R.L., *Forecast Pro for Windows* (Business Forecast Systems, Inc., MA), 1992.

White, K.J., S.A. Haun, N.G. Horsman and S.D. Wong, 1993, *SHAZAM: Econometrics Computer Program - User's Reference Manual, Version 7.0* (Mcgraw-Hill, New York).

Wilson, J. Holton and Barry Keating, Business Forecasting, Irwin, IL., 1990.

LABOR AND PRODUCTIVITY

MODERATOR Norm Saunders, Bureau of Labor Statistics

PRESENTATION 1 Howard Fullerton, Bureau of Labor Statistics

PRESENTATION 2 Lea Swanson, Environmental Protection Agency

PRESENTATION 3 Lloyd Teigen, Economic Research Service

Labor and Productivity

Measuring Rates of Labor Force Dynamics

Howard N Fullerton, Jr., Office of Employment Projections, Bureau of Labor Statistics

Keywords: Entrants, leavers, numerator analysis, projections

1. Introduction

In its most recent description of its labor force projection, BLS attributed the slowing labor force growth to a lag in population growth, noting that most labor force growth is explained by population growth (Fullerton, 1991). In a later section in the same document, the slower growth is attributed to an increase in the number of persons leaving the labor force while the number of entrants remained virtually the same. Although these are not competing explanations of slower labor force growth, it is not obvious that they are the same explanation. The analysis of labor force growth and population growth was done by dividing the growth rate of the labor force into components. The analysis of entrants and leavers was done by calculating the number of entrants and leavers for the 1975-90 and 1990-2005 periods and comparing them, a much lower level of sophistication.

In this paper, we will first explain how BLS measures entrants and leavers; then examine alternate measures of the rate of entry or separation.

BLS measures labor force *entrants* by comparing a specific sex-race birth cohort in 2005 with itself in the base year (1990). For most purposes, entrants is a better or more useful measure than net change. To be in the labor force, one has to be at least 16; thus, in 2005, none of those under 30 could have been in the labor force in 1990. All 16 to 29 year-olds, by definition, must have entered the labor force between 1990 and 2005. Older cohorts are examined by five-year of age groups and compared with their labor force size in the base year. This was done by sex and by race or Hispanic origin. Older cohorts had more members in the labor force in 1990 than in 2005, a difference termed labor force *leavers*. The difference between the aggregated entrants and leavers is *net change*, and should equal the net change calculated by comparing the total labor force numbers for the two years.

2. An example: calculating female entrants

To make the calculation of net entrants more concrete, we have included a summary table of entrants for women with no race or Hispanic origin detail. Those 16 to 19 in 2005 were aged 1 to 4 in 1990; all their 2005 labor force would enter between 2001 and 2005. Starting with ages 40 and over in 1990, the numbers in the labor force at each age decreased; these are the leavers. What is striking is that women are entering the labor force for an extended period of time; the cohort born 1951-55, which would be 55 to 59 in 2005, has more than 150,000 entrants over the 1990-2005 period. The baby boom generation is split between entering and leaving; the older groups, born 1946-50 are leaving; those born 1951-65 are still entering. The baby boom generation would contribute 3 million entrants to the 27 million entrants and 1.3 million leavers to the 12.7 million leavers.

3. Net change

The 2005 labor force of women is larger than in 1990 age group by age group. This reflects increased participation as well as population increases. However, between 1975 and 1990, the passage of the baby boom resulted in a decrease in some age groups. For men, especially, labor force participation rates did not increase enough to offset the population decline. Thus, the net change would have been negative.

Table 1. Entrants to and net change in the male labor force, 1975-90

	Labor force	Entrants	Share of entrants	Labor force	Net change	Share of net change
	1975	1975-90	1975-90	1990	1975-90	Percent
Men, 16 and over	56,299	28,135	100.0	68,232	11,935	100.0
16 to 19	4,805	3,866	13.7	3,866	-939	-7.9
20 to 24	7,565	7,291	25.9	7,291	-274	-2.3
25 to 29	7,734	9,583	34.1	9,583	1,849	15.5
30 to 34	6,457	5,425	19.3	10,230	3,773	31.6
35 and over	29,737	1,970	7.0	37,264	7,527	63.1

Using share of net change yields the surprising and perhaps not very meaningful result that workers over the age of 35 accounted for 63 percent of net change, though only about 7 percent of entrants. Men at the younger ages accounted for 10 percent of net change, considered as a share of "net" or "new" entrants.

Calculating shares when some components are negative and some positive leads to difficulties between 1985 and 2000, the younger portion of the labor force was projected to decrease in size. Analysis of net change is fraught with difficulty, even if the components reviewed do not have any apparent negative elements. Of course, the advantage of using net change is that the underlying age detail is not needed.

BLS has calculated that 55.9 million persons entered the labor force between 1975 and 1990. It has projected that 55.8 million will enter over the 1990-2005 period, virtually the same number. It also has estimated that 24.9 persons left the labor force over the earlier period and that 29.8 million will leave over the 1990-2005 period. The 1975 labor force was 93.8 million, the 1990 labor force 124.8 million. Adding 56 million persons to a 94 million labor force should have a larger impact than adding 56 million persons to a labor force of 125 million. Investigating, or quantifying this observation should help us understand the relation between population growth and entrants, between population aging and increasing numbers of leavers. The number that we seek is a rate. We have been conducting numerator analysis, what we must do is analyze rates; to do that, we need a denominator.

We of course know our numerators, which are available by sex and race or Hispanic origin and by age. However, what are the denominators—what do demographers consider a rate?

4. What is a rate

Shryock and Siegel (1973) discuss the difference between rates and ratios: "The term 'rate' most appropriately applies to the number of demographic events in a given period of time divided by the population at risk during that period. The population at risk is usually only approximated. It may be the population at the middle of the period, (which is roughly the average population), the population at the beginning of the period, or a more complex definition." The cost of obtaining the more complexly calculated population at risk, or a mid-year estimate may be a significant factor in determining the denominator.

The people at risk to entering the labor force (the events we are measuring) are those not in the labor force, the population less the labor force, and those emigrating to the United States. Those at risk to leaving the labor force are those in the labor force.

We may also distinguish between "crude" and more refined rates. For example, the crude birth rate is the births in a year divided by the mid-year population. Not all the population (babies and older women) are at risk to having a child at all and we are counting both parents. Further, risks vary significantly across population groups, specifically, marital status groups. For our measures, a quick approach would be the total population as the at-risk denominator for entrants and the labor force for leavers.

For more refined rates of entry, we would use the not in the labor force population group, with an allowance for immigration. We would also provide age-sex-race or Hispanic origin-specific rates. Hispanic origin data is not available before 1980 and the refinements for race and Hispanic origin will not be pursued in this paper.

Demographers follow actuarial science in distinguishing between probabilities and central rates. Probability-type rates are based on the population at risk at the beginning of the period; central rates are based on a mid-period group at risk. Shryock and Siegel point out that when the population is open, specifically, receiving immigrants, then average populations or person-years are used.

The most difficult problem in providing a denominator for entrants is supplying an estimate of immigrants. Generally, the Census Bureau provides estimates of net immigration with their estimates of the components of change {reference}. Although there is an assumption of immigration and emigration in these estimates, they are generally not available nor are they available by age. This suggests using central rates rather than probability-type rates. An intermediate value would have immigrants embodied in the estimates.

5. Using "probability-type" denominators

The easiest set of rates to compute uses the total population for both entrants and leavers. If the same base is used for both groups, the difference is meaningful:

Table 2. Crude rates of entry to and departure from the labor force, 1975-90 and 1990-2005, probability-type measures
[Numbers in thousands]

	1975-90		Population		1990-2005		Population
	Entrants	Leavers	1975		Entrants	Leavers	1990
Total	55,915	24,901	215,465		55,798	29,851	249,415
Men	28,135	16,199	104,876		28,197	17,090	121,599
Women	27,780	8,702	110,589		27,601	12,761	127,816
	rates per thousand		difference		rates per thousand		difference
Total	260	116	144		224	120	104
Men	268	154	114		232	141	91
Women	251	79	173		216	100	116

The entrant rates declined from 260 per thousand for the 1975-90 period to 224 per thousand 1990-2005. This represents a decline of more than ten percent. Given that the number of entrants remained essentially the same, while the labor force grew by 26 million, the rates would be expected to drop. Female and male rates dropped about the same amount. Notice that their rates of entry, like their levels, are fairly close.

The rate of leaving the labor force increased modestly from 116 per thousand over the 1975-90 period to 120 per thousand. For the labor force as a whole, the slowing of labor force growth was influenced more by the slowdown in entrants than the increase in leavers. The male rate of leaving dropped by 10 percent. That is, men are projected to be *less* likely to leave the labor force over the 1990-2005 period than they did over the 1975-90 period. The rate of departure for women increased by 27 percent or they are projected to be more likely to leave the labor force than in the earlier period The increase was more than the amount of the decrease in the rate of entry. The female labor force is projected to grow more slowly both because the rate of entering dropped and because the rate of leaving increased.

Since both rates are calculated with the same base, it is meaningful to subtract the rates. This rate of change is lower for the projected period; measuring the slower growth of the labor force. The male labor force has been growing more slowly than the female labor force; this is projected to continue.

6. Refining the rates—departures

Only those in the labor force may leave it; for leavers, a more meaningful measure would use the labor force as a denominator:

Table 3. Leavers, labor force, and rates of leaving the labor force, 1975 and projected 1990-2005
[In thousands]

Sex and age	Labor force	Leavers 1975-90		Labor force	Leavers 1990-2005	
		Number	per thousand		Number	per thousand
Age in 1975 or 1990	1975			1990		
Total, 1 year or older	93,775	24,901	266	237,893	29,851	125
Men, 1 year or older	56,298	16,199	288	181,339	17,090	94
1 to 14						
15 to 19	4,805			3,866		
20 to 24	7,565			7,291		
25 to 29	7,734			9,583		
30 to 34	6,457	220	34	10,230	346	34
35 to 39	5,288	348	66	9,261	738	80
40 to 44	5,112	1,098	215	8,007	1,564	195
45 to 49	5,299	2,528	477	6,237	2,988	479
50 to 54	5,101	3,909	766	4,939	3,644	738
55 to 59	4,205	3,671	873	4,012	3,447	859
60 and	4,732	4,425	935	4,805	4,363	908
Women, 1 year or older	37,477	8,702	232	56,554	12,761	226
1 to 14						
15 to 19	4,065			3,544		
20 to 24	6,185			6,552		
25 to 29	4,965			7,837		
30 to 34	3,708			8,152		
35 to 39	3,275			7,640		
40 to 44	3,231			6,936		
45 to 49	3,376			5,339		
50 to 54	3,307			3,977	54	14
55 to 59	2,662	172	65	3,059	1,310	428
60 and	2,703	8,530	3,156	3,518	11,396	3,239

This table also shows a decline in the rate of leaving the labor force between the 1975-90 and 1990-2005 periods. The drop is substantial, by more than half. However, only the male rate dropped, by almost two-thirds. Female departures remain at the same level. This measures leavers from the labor force, not asking their destination, retirement, death, or emigration. If we were concerned about a specific outcome, we would have to take mortality and emigration into account.

This more refined measure points in a different direction than the crude rate, which indicated no change in the rate of leaving the labor force. The crude rate did preserve the relationship between men and women's rates.

The age-specific data indicate that men begin leaving the labor force sooner than men (or possibly that men complete their entry to the labor force sooner). The rates rise until the oldest age break, 60 and over. Since the usual age of entitlement to retirement occurs after 60, that we do not have information on the pattern of withdrawal is unfortunate. It is an aspect of the wide interval 15 years over which we calculate the changes. By this measure, few women permanently leave the labor force before age 60. The women considered to depart for child-rearing are outweighed in both time intervals by the number of women entering. One difference between the historic period and that being projected is that the most rapid increase in the rate of leaving the male labor force occurs between ages 45 to 49 and 50 to 54 in 1975-90. According to the projections, that will be between ages 40 to 44 and 45 to 49.

For entrants, the concern is not about destination, but about sources. Although most entrants come from the United States, and there few deaths at the ages of entry, some entrants do arrive from outside the country.

Table 4. Entrants, and not in the labor force, 1975-90 and projected 1990-2005
In thousands

Age and sex	Not in the Labor Force 1975	1975-90 Entrants number	1975-90 rate per thousand	Not in the Labor Force 1990	1990-2005 Entrants number	1990-2005 rate per thousand
Total, 1 year or older	118,541	55,915	472	120,618	55,798	463
Men, 1 year and older	46,967	28,135	599	51,315	28,197	549
1 to 4	6,627	3,866	583	7,608	4,575	601
5 to 9	8,972	7,291	813	9,247	7,989	864
10 to 14	10,534	9,583	910	8,805	8,269	939
15 to 19	5,952	5,425	911	5,243	4,820	919
20 to 24	2,075	1,697	818	2,453	1,817	741
25 to 29	791	273	345	1,071	546	510
30 to 34	501			671	92	137
35 to 39	367			638	89	140
40 and over	11,148			15,579		
Women, 1 year and older	71,574	27,780	388	69,303	27,601	398
1 to 4	6,342	3,544	559	7,258	4,218	581
5 to 9	8,623	6,552	760	8,817	7,266	824
10 to 14	10,112	7,838	775	8,385	7,227	862
15 to 19	6,400	4,087	639	5,101	3,953	775
20 to 24	3,492	1,455	417	2,835	1,706	602
25 to 29	3,695	1,971	533	2,737	1,734	634
30 to 34	3,465	1,631	471	2,854	1,272	446
35 to 39	2,658	702	264	2,435	219	90
40 and over	26,787			28,881	5	0

According to this more refined measure, the rate of entry between 1990 to 2005 remained the same. Both sex groups remained about the same, with women's entry increasing slightly, men's decreasing slightly. The crude rate of entry indicated a more greater drop in the rate of entry, with even less difference in the change in rates of men and women. The not in the labor force population (NILF) increased for men and decreased for women. This decrease reflects women's labor force growing more rapidly than their population. The continued growth of the male NILF population simply reflects population growth.

Within the age data, men completed their labor force entry by age 29, more technically, the size of the labor force cohorts did not grow for any cohorts aged 30 and over in the first year of the period. For women, this age of entry extended to 35 to 39. Judging the entry rates or levels on the basis of labor force rates for those aged 16 to 24 yields an underestimate of the number of women entering the labor force, 4.3 million in the cohorts aged 25 and over entered the labor force between 1975 and 1990. Although entry is projected to extend over a longer period, the number of women entering the labor force after age 24 is projected to be smaller. The number of men projected to enter the labor force at these "older" ages is projected to more than double.

References

Fullerton, Howard N, Jr.,(1991), "Labor force projections: the baby boom moves on," *Monthly Labor Review*, November, pp. 31-44, reprinted in *Outlook: 1990 - 2005*, (BLS Bulletin 2402), April 1992.

Gendell, Murray, and Siegel, Jacob S., (1992), "Trends in retirement age by sex: 1950-2005," *Monthly Labor Review*, July, pp. 22-29.

Shyrock, Henry J., and Siegel, Jacob S., and Associates, (1973), *The Methods and Materials of Demography*, (Washington, U.S. Bureau of the Census)

U.S. Bureau of the Census (1990), "United States Population Estimates, by Age, Sex, Race, and Hispanic Origin: 1980 to 1988," *Current Population Reports*, Series P-25, No 1045, (Washington), pp. 24-29, and Table 5, "Estimates of the Components of Population Change for Calendar Years: 1940 to 1988."

Pollution Prevention and Unemployment Prevention are Parallel Arms of a Single Strategy for Sustainable Development

Lea Swanson*, Environmental Protection Agency

Summary

This paper argues that pollution prevention and unemployment prevention are parallel arms of a single strategy for national competitiveness -- defined to include the achievement of quantitative and qualitative societal goals. The level of national competitiveness achieved ultimately depends on productivity. And productivity, in turn, can be increased by reducing the inputs required to achieve the same level of production. However, this is only half of the story. If the technology/process applied to increase productivity results in spoiled production inputs, the spoiled inputs cannot be used for future production. Similarly, if the technology/process results in reduced labor inputs, and that labor cannot be employed in another industry, the reduced input costs will be offset by increased expenditures required to support (welfare) and retrain displaced labor. A productivity increase is not realized if the resources released for use elsewhere, cannot be used. Moreover, productivity increases must increase efficiency and effectiveness if they are to add value to national competitiveness.

Drivers of Economic Prosperity

Competition drives national economic development: that is, competition -- at both an international and local level -- for trade, skilled manpower, innovative technology, and strategic advantage. Indeed, when Michael Porter, a recognized authority on competitiveness,[1] refers to phases of economic development (factor driven, investment driven, innovation driven, wealth driven) he talks in terms of the "stages of national competitiveness". He also links vigorous local competition between firms to the creation and maintenance of competitive advantage in the global arena. This link is critical to understanding Porter's definition of competition as dynamic and evolving; a process that has no equilibrium. Rather, he says "competition is a constantly changing landscape in which new products, new ways of marketing, new production processes, and whole new market segments emerge. Static efficiency at a point in time is rapidly overcome by a faster rate of progress."[2] The essential character in competition is thus innovation and change. And innovation here refers both to improvements in technology and better processes of production, with productivity improvements the logical corollary.

Porter uses the terms economic prosperity and competitiveness interchangeably. Competition, then, not only drives economic prosperity, it is also an end in itself. The level of competitiveness achieved by a nation depends, in turn, on the productivity with which national resources are employed. Productivity is a function of innovation, and innovation "requires sustained investment in research, physical capital, and human resources" to develop skills and knowledge. Mapping back full circle to the vigor of local competition, Porter asserts that rivalry drives innovation and "upgrading of competitive advantage". Rivalry is especially important to Porter's hypothesis

*Lea Swanson serves as Special Assistant for Policy, Planning and Evaluation, Office of the Assistant Administrator, U.S. Environmental Protection Agency. Ms Swanson earned her Masters of Public Administration at Harvard University. She is currently finalizing her thesis on national competitiveness and sustainable development to complete requirements for her Masters of Business Administration at Monash University (Australia).

[1] Porter, Michael E., teaches at the Harvard Business School, and is an advisor to leading companies all over the world. He is the author of Competitive Strategy (1980), Competitive Advantage (1985), and The Competitive Advantage of Nations (1990), all published by The Free Press, New York.

[2] Porter, Michael, E. The Competitive Advantage of Nations, The Free Press, New York, 1990, p.20.

because it creates a fear of failure and thereby overcomes inertia.[3] For Porter, corporate rivalry is the critical driver of national competitiveness.[4]

Recent analyses of national competitiveness offer a range of perspectives that support Porter, disagree with Porter, or reinforce Porter's conclusions but through a different theoretical framework. For example, Stephan Schmidheiny and the Business Council for Sustainable Development[5] fully concur with Michael Porter: competitiveness depends on a nation's capacity to innovate, which in turn depends on the intensity of local competition. Efficiency is highlighted as an important determinant of competitiveness. This lends support to Porter's emphasis on resource productivity as an important measure of competitiveness. Another important determinant of competitiveness cited by the Business Council is national infrastructure. "Macroeconomic stability; free, open markets; clear property rights; and political stability" are important elements of an attractive environment for investment.

The World Bank,[6] on the other hand, characterizes economic prosperity as the driver of national development. The report states that "economic development is an essential means for enabling national development, but in itself is a highly imperfect proxy for progress." Rather, national development is about improving the well-being of people. As such, the essential elements of national development are higher standards of living; better education, health, and opportunity; improved environmental protection, and a more equitable distribution of the burdens and benefits of national policies. The World Bank effectively turns Porter's concept of economic prosperity/competitiveness on its head, by describing it as a means to achieving broad societal goals, rather than an end in itself. Yet, somewhat paradoxically, the World Bank supports Porter's framework for national competitiveness: "By promoting specialization and competition and encouraging technological progress, open trade and investment policies raise productivity and improve efficiency -- including efficient use of environmental resources."[7] The difference lies in the qualifications that the World Bank adds to Porter's concept of economic prosperity/competitiveness as an end in itself. For the World Bank "what matters is that the overall productivity of the accumulated capital -- including its impact on human health and aesthetic pleasure, as well as on incomes -- more than compensates for any loss from depletion of natural capital." It is to this end that the World Bank identifies macroeconomic policies such as open trade and foreign investment as the drivers of national development.

A view that is diametrically opposed to the World Bank's perspective comes from a study by the McKinsey Global Institute[8], together with Nobel Laureate Robert Solow. The report cuts a straight line from productivity to competitiveness. There is no intervening process of competition emphasized here. Rather, productivity is the means to an end. According to the report, productivity, including service sector productivity, is the ultimate measure of international competitiveness. There are no broader societal goals to be achieved here. Insofar as

---

[3] ibid. I have focussed on only one of the four preconditions for competitive advancement discussed by Porter. The four preconditions are:

1. factor creation mechanisms
2. motivation
3. domestic rivalry
4. demand upgrading

[4] In my forthcoming (December 1993) MBA thesis entitled "Resilience: A Three Dimensional Metric of National Development", I do not concur with Porter's assertion that corporate rivalry is the critical driver of national competitiveness. I argue instead for the paramount role of cooperation in guiding competitive forces toward win/win solutions. For example, when I explore the role of corporate rivalry/worker skill and experience in driving corporate performance, it is only when I introduce the concept of co-evolution that the dynamic interaction of both competitive and cooperative forces are invoked. In the resultant scenario, mutually interdependent firms are able to maintain their independence and reap the benefits of cooperative efforts. Likewise when I explore the role of innovation and experimentation/regulations and standards, co-evolution as opposed to competitiveness offers a dynamic that utilizes both competitive and cooperative forces and underscores the value of strategic alliances. And cooperation is not limited in these examples to the level of the corporation. At the core of the concept of co-evolution is the notion of cooperative partnerships in all facets of activity -- between employees and management; between competing firms and industries; between the regulated community and the regulators; between government, the private sector, and the public, and between nations.

[5] Schmidheiny, Stephan, with The Business Council for Sustainable Development, Changing Course, The MIT Press, Cambridge, Ma., 1992.

[6] World Development Report 1992, Oxford University Press, New York, 1992.

[7] ibid.

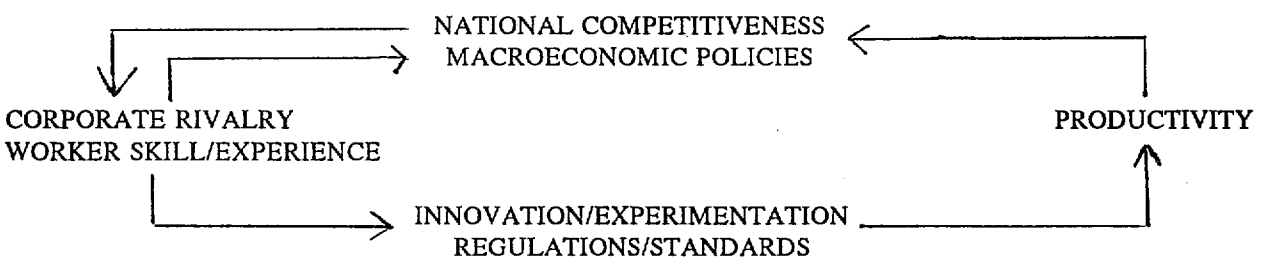
[8] Sylvia Nasar, "U.S. Rate of Output Called Best -- Study Affirms Lead Over Foreign Rivals," New York Times, October 13, 1992.

productivity describes, "not only a nation's standard of living, but also status in the world" -- productivity-cum-competitiveness is an end in itself.

On the topic of competitiveness, U.S. Secretary for Labor, Robert B. Reich, has quipped: "rarely has a term in public discourse gone so directly from obscurity to meaninglessness without an intervening period of coherence,"[9] That hasn't stopped Reich from offering an opinion on just what competitiveness means and how it should be measured. Reich firmly asserts that competitiveness is "the capacity of a nation's citizens to add value to the global economy without going into ever deeper debt."[10] It depends, he says, on the jobs that Americans perform which in turn determines standards of living and whether or not these standards are sustainable. National competitiveness, in this context, is synonymous with a nation's standard of living. Finally, the jobs that Americans perform depends on their respective skills and insights -- attributes that are a function of both training and experience. Of particular importance is the skill that finds the right fit between particular technologies and particular markets. "Experimentation", says Reich, "is the lifeblood of the high-value enterprise, because customization requires continuous trial and error." It is here that the corporation plays its vital role. The high-value corporation offers problem solvers, problem identifiers, and facilitators "membership in a small group engaged in a common task, sharing the risks of defeat and the potential rewards of victory." Thus Reich identifies the skill and experience of a nation's workforce (as distinct from Porter's corporate rivalry) as the critical driver of national prosperity and hence competitiveness.[11]

What's important here is that for Porter, national prosperity is synonymous with competitiveness, whereas for Reich, national competitiveness is synonymous with national prosperity. In each case the latter concept defines the former. Reich, not unlike the World Bank, defines national prosperity/competitiveness to include both the quantitative and qualitative terms imbued in standard of living and quality of life. Porter, Schmidheiny, and Solow, on the other hand, draw their definitions of national prosperity/ competitiveness from the narrower, and more immediate, quantitative perspective of economic competition.

Notwithstanding the differences in definitional perspectives, the components of the views discussed form a logical framework for exploring the process of national economic development. In summary, competitive forces, both global and local, drive a nation's economic development. However it is defined, national competitiveness, reflected in a country's macroeconomic policies, affects local corporate rivalry, and at the same time, is influenced by local competitive forces. Local rivalry, and worker skill and experience, drive innovation and experimentation. Acting within the parameters of a nation's regulatory environment, innovation, in turn, drives productivity improvements. The level of competitiveness achieved by a nation ultimately depends on national productivity.



Productivity

Productivity, then, reflects technology, innovation, invention, skills, and investment at both the macro- and micro-economic levels. It can be defined as the rate at which inputs (land, labor, capital, materials) are transformed into outputs, and it is usually measured by dividing GDP by total population. However, on recalling the World Bank's qualifications -- that the overall productivity of accumulated capital must more than compensate for any loss from the depletion of natural capital -- GDP per capita, as currently defined, is an inadequate measure of national

[9] Bob Davis, "Competitiveness is a big word in D.C., just ask the Vice-President," Wall Street Journal, July 1, 1992.

[10] Reich, R.B., "Who Is Us", Harvard Business Review, 1990.

[11] Reich, Robert B., The Work of Nations, Alfred A. Knopf Inc., U.S.A., 1991, p.119.

productivity.[12] Indeed the shortcomings of conventional measures of national income and productivity continue to be explored and documented by notable economists and environmentalists such as Herman Daly.[13]

The following is my own attempt, drawing on recent research and opinion[14], to state a transparent formula for measuring productivity and defining its relationship to national competitiveness. I defer here to the definition of national competitiveness discussed above that includes both quantitative and qualitative societal goals.

The Production Function:

National Production = Inputs (labor, capital, materials, natural resources) x productivity (the rate of transforming inputs into outputs)

In $ value terms:

Net Value of national production (Gross Value of Production - Value of Pollution) = Volume of production x marginal value to the nation of transforming inputs into outputs

Where:

National Margin = Gross value of national production - the total value of inputs (including labor and natural resources) - net benefits the nation must deliver to attract domestic and foreign customers and investors in the face of international competition[15] (such as investments in infrastructure, education, resource maintenance and restoration, as well as expenditure on welfare and environmental clean-up)

Now transparent, the production function reveals important characteristics of its various components:

1. The value of national production can be increased either by increasing the volume of production; decreasing pollution, or by increasing productivity.
2. Productivity increases can result from reducing the inputs required for production, and thereby the costs of inputs. This is usually accomplished through the application of innovative technologies.
3. If production inputs, for example labor, are reduced in one industry, but cannot be employed in another industry, the reduced input costs will be offset by increased net National benefits that must be delivered to support and retrain displaced labor.
4. If the technology or process applied to achieve increased productivity, results in spoiled production inputs (for example soil erosion, water pollution) then:
   a) pollution increases, leading to a decrease in the Net Value of Production

[12] Some of the Bank's current research in this area, carried out jointly with the United Nations Statistical Office (UNSO), is being conducted in two developing countries -- Mexico and Papua New Guinea. The UNSO Framework being used in these studies is a system for environmentally adjusted economic accounts. It tries to integrate environmental data with existing national accounts information, while maintaining national accounting concepts and principles as far as possible. Refer, Ernst Lutz and Mohan Munasinghe, "Accounting for the Environment", Finance & Development, March 1991.

[13] Daly, Herman, "Toward a Measure of Sustainable Social Net National Product", in Environmental Accounting for Sustainable Development, ed. Yusuf J. Ahmad, Salah El Serafy, Ernst Lutz, The World Bank, Washington D.C., 1989.

See also, by the same author, and John B. Cobb Jr., For the Common Good, Beacon Press, Boston Mass, 1989.

And also by the same author, and Robert Goodland, "Ten Reasons Why Northern Income Growth is not the Solution to Southern Poverty", The World Bank, February 1992.

Ernst Lutz, Salah El Serafy, and Roefie Hueting are among the recognized authorities on this subject. The following essays appear in Environmental Accounting for Sustainable Development, op.cit.

Serafy and Lutz, "Environmental and Resource Accounting: An Overview"

Serafy, "The Proper Calculation of Income from Depletable Natural Resources"

Hueting, "Correcting National Income for Environmental Losses: Toward Practical Solution"

[14] In addition to Herman Daly, op.cit, Ernst Lutz, Salah El Serafy, and Roefie Hueting, I refer to William P. Browne, et al, Sacred Cows and Hot Potatoes: Agrarian Myths in Agricultural Policy, Westview Press, Boulder Co, 1992. The authors articulate the important distinction between production and productivity; note the societal benefits of productivity increases, and emphasize that technological innovation is not to be blamed for unemployment.

[15] I have adapted a definition used by Pankaj Ghemawat, Professor of Business Administration, Harvard Business School, to describe an organization's margin. Refer his book entitled Commitment: The Dynamic of Strategy, The Free Press, New York, 1991.

b) the spoiled inputs, cannot be used for future production.[16] Again, the reduced input costs will be offset by increased net National benefits that must be delivered to restore, if possible, the spoiled resources, or to search for substitutes.

5. If labor retraining, or resource restoration, are required, there will be a time lag before productivity increases are realized. This lag may be indefinite depending on the nature and extent of retraining and/or restoration required.

The conditions that must hold for productivity increases to be realized are therefore more complex than is apparent from the simple equation that divides GDP by population. In particular, **a productivity increase is not realized if the resources released for use elsewhere, cannot be used.**

Even if a productivity increase is realized, however, this does not automatically signal enhanced national competitiveness. While the dollar value of productivity, national margin, serves as a measure of national efficiency, it may not serve as a measure of national effectiveness. National margin does not necessarily reflect the quality of production, or whether or not the investments are worthwhile.[17] Productivity increases must increase efficiency and effectiveness, and thereby enhance quality of life, if they are to add value to national competitiveness in its richer definition. This means they must:

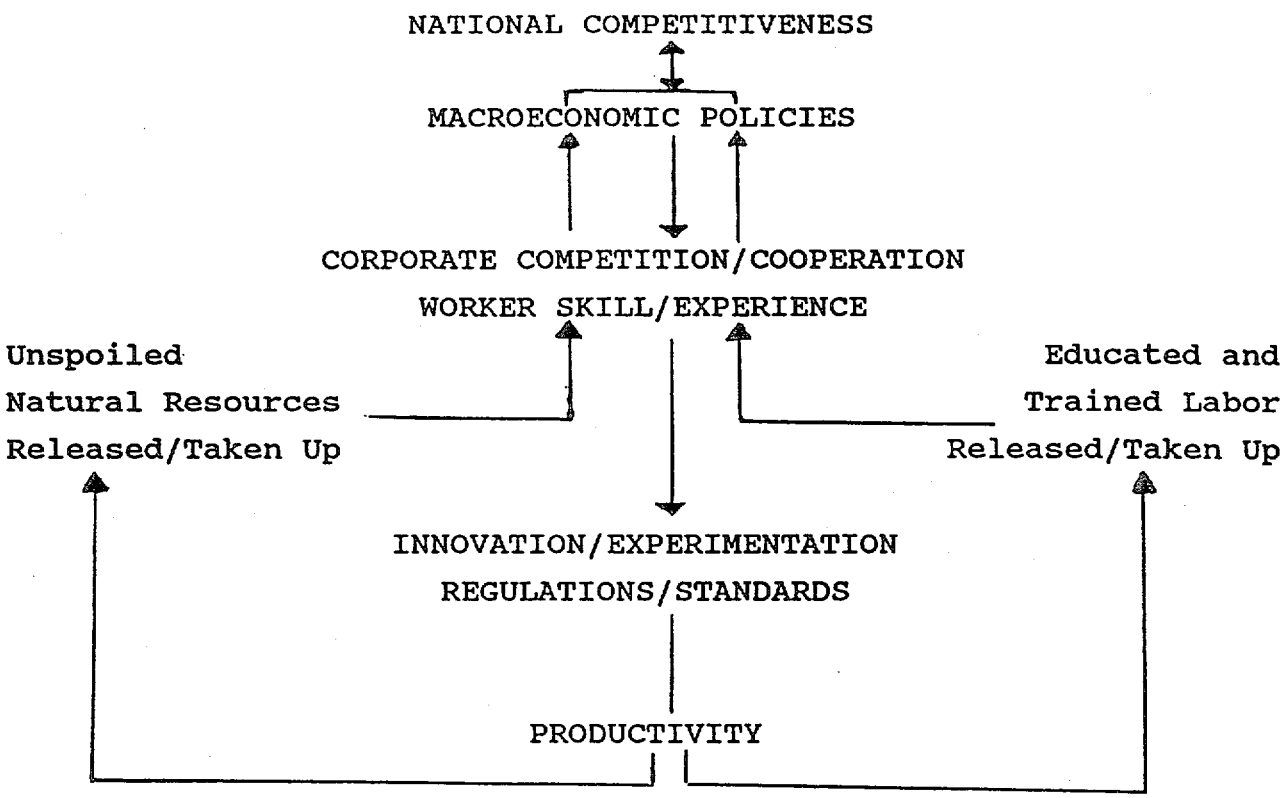
i) achieve worthwhile increases in the total volume of production in industries with unmet demand. An increase in the volume of production in industries where total production exceeds demand -- such as subsidized industries -- does not add value to national competitiveness. Instead, it results in overproduction which is a value-subtracting exercise, OR

ii) release useful inputs to production (unspoiled natural resources, educated labor) from a mature or declining industry to be used in emerging, or growing industries for the production of high quality goods and services

National competitiveness, which includes the achievement of quantitative and qualitative societal goals, ultimately depends on productivity -- as now defined. It is measured by expanding national accounts to include investments in environmental and human assets that contribute to economic activities. Similarly, imputed costs of resource depletion, costs of degradation, and costs of worker displacement that are directly related to productive activities and the generation of value added, are also included. Strategic thinking and acting to prevent both pollution and unemployment spurs innovation, and enables productivity gains to be fully realized. Hence pollution prevention and unemployment prevention emerge as parallel arms of a single, sustainable strategy to advance national competitiveness. Refining the diagram on page 142 emphasizes these critical linkages:

[16] This point is made in Sacred Cows and Hot Potatoes: Agrarian Myths in Agricultural Policy, William P. Browne, et al, Westview Press, Boulder Co., 1992.

[17] I refer here to the definitions of efficiency and effectiveness adopted by Osborne and Gaebler in their book Reinventing Government, A William Patrick Book, Addison-Wesley Publishing Company, Inc., Reading, Ma., 1992.


NATIONAL COMPETITIVENESS
MACROECONOMIC POLICIES
CORPORATE COMPETITION/COOPERATION
WORKER SKILL/EXPERIENCE
Unspoiled
Natural Resources
Released/Taken Up
Educated and
Trained Labor
Released/Taken Up
INNOVATION/EXPERIMENTATION
REGULATIONS/STANDARDS
PRODUCTIVITY

Employment and Income: Analysis and Prospects

Lloyd D. Teigen, USDA Economic Research Service

Abstract

Per capita income won't rise above 1990 levels, unless high-wage jobs are created faster than population projections. The reason for the peak in 1990 was record employment per capita. Creating new jobs in the high-wage (high-productivity) sectors of the economy will increase both real average wages and real per capita income. But in the absence of new high-wage jobs, the prospects are falling real wages, fewer jobs per capita, and falling per capita income, not prospects we'd like to see realized. This paper investigates sector specific trends in the relationship between employment per capita and relative wage rates.

Summary

Per capita income won't rise above 1990 levels, unless high-wage jobs are created faster than population growth. Real disposable personal income per capita peaked early in 1990, not to be surpassed until 1993-I and then by only a quarter of 1 percent. Real wage and salary disbursements per capita peaked in 1989 and have not recovered (1992 is less than 1987). The reason for the peaks in 1989 and 1990 is record employment per capita. From 1990 to 1992, employment per capita (EPC) has dropped 3.5 percent.

Relating EPC to real wages, a negative response is apparent. At lower real wages, more people are employed per capita. For the economy at large, two regimes are evident: when manufacturing was the dominant sector, the response was inelastic; when services became dominant, the response became elastic. At the 1990 EPC, the elasticity on the 1969-92 curve is one -- denoting the maximum attainable labor income per capita under that curve. However, the relationship between total wage payments and employment doesn't show enough curvature to fully corroborate the unit elasticity finding.

Table 1: Employment per 1,000 population, 1983-92.

	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992
Agriculture	14.4	14.1	13.3	13.1	13.2	12.9	12.9	12.7	12.8	12.6
Mining	4.1	4.1	3.9	3.2	3.0	2.9	2.8	2.8	2.7	2.5
Construction	16.8	18.5	19.6	20.0	20.5	20.9	21.0	20.5	18.5	18.0
Total Manufacturing	78.7	82.0	80.8	78.8	78.4	79.0	78.6	76.5	73.0	71.2
Durable Goods Manufacturing	45.7	48.6	48.1	46.6	46.0	46.4	46.2	44.5	42.0	40.5
Nondurable Goods Manufacturing	33.0	33.4	32.7	32.3	32.4	32.5	32.4	32.0	31.1	30.7
Transportation and Public Utilities	21.1	21.8	22.0	21.8	22.1	22.6	22.8	23.2	22.8	22.5
Wholesale Trade	22.6	23.6	24.1	24.0	24.2	24.7	25.2	24.8	24.0	23.4
Retail Trade	66.6	69.9	72.7	74.4	76.0	77.9	79.0	78.7	76.2	74.9
Finance, Insurance, and Real Estate	23.3	24.1	25.0	26.1	27.0	27.1	27.1	26.9	26.4	26.1
Services	84.1	88.0	92.3	95.8	99.8	104.8	109.6	112.4	112.1	113.2
Federal Government	11.8	11.9	12.1	12.0	12.1	12.1	12.1	12.3	11.7	11.6
State and Local Government	55.9	55.9	56.7	57.3	57.9	58.8	59.8	60.9	61.0	61.1
Total Armed Forces	9.1	9.0	9.0	9.0	9.0	8.7	8.6	8.1	7.8	7.0
Total Employment	408.5	422.9	431.3	435.7	443.1	452.4	459.5	460.1	449.2	444.1

Source: [1], Table B-41, Table B-30, and B-29.

In the upper chart, higher employment per capita generates more wage and salary disbursements per capita. In the lower chart, increasing jobs per capita --beyond 1990 levels-- by reducing real wages can't be counted on to increase real (wage-based) income. But, first, employment per capita has to rise to 1990 levels. That's why the 1990 level of per capita real income may be hard to surpass.

But, one equation doesn't tell the whole story. Each sector of the economy has its own story, and they are not marching in lock step. There are sectors where EPC is rising (services, wholesale & retail trade, finance insurance & real estate (FIRE), state & local government); where EPC has no trend (transportation & public utilities, construction, federal government); and where EPC is falling (agriculture, mining, manufacturing, and armed forces). These trends affect the prospect for per capita income, since incomes vary by sector: 2 construction jobs equal 3 service jobs, 4 manufacturing jobs equal 4 wholesale or 9 retail jobs, etc.

Sectoral evidence corroborates the aggregate relationship. About half of US jobs are in sectors with elastic-to-unitary demand for labor, about a fourth are in sectors with inelastic demand, and about a fourth in sectors without an acceptable relationship. Estimation revealed a highly elastic response of service EPC to relative wage rates. Employment in wholesale trade and FIRE also had an elastic response to relative wage rates. Retail EPC had an elasticity near one. Other sectors, notably agriculture, manufacturing, and transportation & public utilities were quite inelastic to wage rates. No acceptable relationships were found for mining, construction, or government.

To increase real per capita income, jobs have to be created faster than population growth. To do this, the relationships have to shift outward. Creating new jobs in the high-wage (high-productivity) sectors of the economy will increase both real average wages and real per capita income. But in the absence of new high-wage jobs, the prospects are falling real wages, fewer jobs per capita, and falling per capita income, not prospects we'd like to see realized.

To restore per capita employment to 1990 levels, the challenge is to create 3.6 million jobs now and 100,000 jobs a month hereafter. This means creating more than 10,000 new jobs a day beyond those lost to layoffs to recoup the loss in a year. That's a tall order, especially when individual companies announce plant closings and layoffs affecting tens of thousands in a single day.

This paper's major contribution has been to identify, and to some extent quantify, a problem in the US macroeconomy's performance. The empirical research I've done does not identify all the policy instruments needed to expand employment per capita. It does suggest that falling real wages expand per-capita employment. Real wages can decline from two factors: nominal wage rates may fall, or product prices in the employing sector are allowed to rise. The rise in product prices is a form of "beneficial inflation," if the labor demand elasticity is large enough to produce a significant growth of employment. ["Beneficial" in a partial equilibrium sense that abstracts from the effects on interest rates and, subsequently, on the capital market.

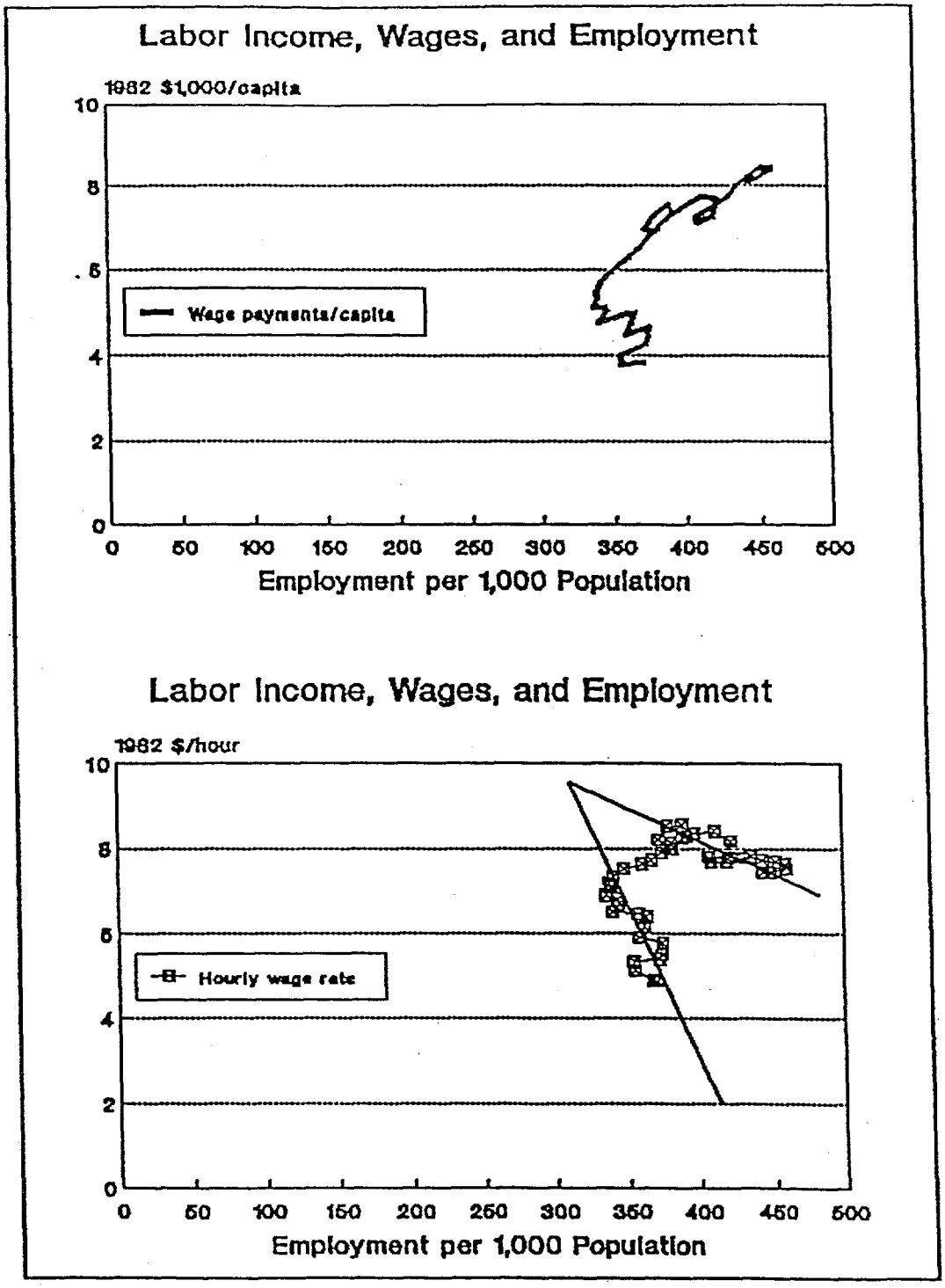
Empirical Evidence -- in Aggregate

My model of labor demand expresses employment as a function of wage rates and variable input prices, each deflated by an index of output prices for the respective sector. Because of aggregation, the coefficients change with the (unobserved) number of firms [in the sector, or using a technology] per 1,000 population. The lower half of the chart on page 118 illustrates this labor demand function. Two distinct trends are evident. The response is inelastic between 1947 and 1964, and elastic between 1969 and 1992.[1] Variations from the straight lines may be attributed to random disturbances, to cross price terms, and to intra-sample changes in the aggregate coefficients linked to individual technologies.

[1] What can we infer from the increasing elasticity of labor demand between the two epochs? In the early epoch, manufacturing dominated the employment picture. In the later period, service industries dominated. The more negative the second derivatives of the production function, the more inelastic the demand for labor. Interpret this as saying that both labor and capital are highly specialized, and that adding a little more of each greatly diminishes the productivity of both. That sounds to me like a manufacturing plant with a specific product line.

As workers shifted to service jobs, the nation's productive capacity became less specialized. Added capital or labor is more like existing capital or labor, and the marginal productivity does not diminish as rapidly. Labor demand consequently becomes more elastic. As the shift to the service sector becomes complete, specialties re-emerge in the work force, and the marginal productivity of labor and capital begins to fall off more rapidly. The elasticity of labor demand begins to fall -- out of the elastic range -- into the unitary range.

By including the average work-week, the unemployment rate, and the labor force participation rate for females, reasonably stable responses could be estimated over the entire period, 1948-92. Shorter work-weeks and high female participation correlate with greater employment per capita. Higher unemployment correlate with fewer workers per capita. Female participation rates explain the subsector trends moreso than the national aggregate. The elasticities with respect to relative wages in most of these cases were less than one.


Labor Income, Wages, and Employment
1982 $1,000/capita
10
8
6
4
2
0
Wage payments/capita
0
50
100
150
200
250
300
350
400
450
500
Employment per 1,000 Population
Labor Income, Wages, and Employment
1982 $/hour
10
8
6
4
2
0
Hourly wage rate
0
50
100
150
200
250
300
350
400
450
500
Employment per 1,000 Population

Representing the period 1947 to 1964, the elasticity is -.228 at ($6.13, 356), based on:

$$\frac{\text{Employment}}{\text{1,000 Popn}} = 436.6 \underset{(-5.64)}{-13.246} \frac{\text{Wage Rate}}{CPI_{82}} \quad R^2 = .62,\ S = 8.5$$

Representing the period 1969 to 1992, the elasticity is -1.217 at ($7.94, 415), based on:

$$\frac{\text{Employment}}{\text{1,000 Popn}} = 920.2 \underset{(-5.63)}{-63.597} \frac{\text{Wage Rate}}{CPI_{82}} \quad R^2 = .59,\ S = 18.2$$

The maximum labor income per capita occurs where the wage elasticity of labor demand is -1. This occurs at ($7.23, 460) under the 1969-92 relationship. By comparison, the 1990 data point was ($7.52, 460). Labor income per capita in 1990 was in fact larger than the maximum labor income possible under the smooth linear relationship.

To attempt estimation of the cross-price effects, two components of the Producer Price Index (PPI) were examined. These were the PPI for intermediate materials, supplies, and components; and the PPI for crude materials for further processing. The PPI for intermediate materials had a significant influence over 1947-64. But, neither had a significant influence during 1969-92.

During 1947-64, higher priced intermediate materials were associated with greater per capita employment. Qualitatively, intermediate materials substituted for labor or vice versa.

The estimated response remained inelastic during 1947-64. The wage elasticity was -0.20 and the cross-price elasticity was +0.41, at the point ($6.13, 1.04, 355). The estimated response is:

$$\frac{\text{Employment}}{\text{1,000 Popn}} = 279.3 \underset{(-4.54)}{-11.622} \frac{\text{Wage Rate}}{CPI_{82}} \underset{(1.88)}{+141.34} \frac{\text{PPI-Intermed.}}{CPI_{82}}, \quad \text{with } R^2 = .69 \text{ and } S = 7.9\ .$$

Under this specification, labor appears to be an inferior input. Higher output prices (viz. CPI_{82}) evoke lower employment per capita. A 10 percent rise in the CPI appears to reduce employment by 2.1 percent.

Finding no significant cross-price effect for 1969-92 was discouraging. Except for lower real wages, the only way to increase employment per capita is a random (or otherwise unidentified) event. Per capita employment can increase only at the expense of lower real wages. Given where the U.S. is on that relationship, labor income per capita cannot increase much with the additional per capita employment. Unless the relationship shifts dramatically, per capita labor income is not likely to exceed the levels of 1988-90.

Subsector Demand for Labor

Subsectors of the economy are more homogeneous than the economy as a whole. Thus, subsector estimates of labor demand are expected to have greater precision than estimates for the national aggregate. To help explain the shift in the national labor demand function, from an inelastic function before 1964 to a more elastic function after 1970, we expect sectors with growing per-capita employment to be more elastic than those with declining employment. That is, labor demand in the service, finance, and trade sectors is expected to be more elastic than labor demand in agriculture, mining, and manufacturing.

Because each sector represents a different aggregation of products or services, different price deflators and different wage concepts were used in each labor demand equation*.

The estimation revealed a highly elastic response of service employment per capita to relative wage rates. Employment in wholesale trade and in the financial sector (FIRE) also had an elastic response to wage rates. Per capita retail employment had an elasticity near one. All other sectors, notably manufacturing, agriculture, and transportation & public utilities were quite inelastic to wage rates. The elasticities are summarized in the following full-page box.

*All equations mentioned in this paper were estimated using Classical Least Squares. No attempt was made to correct for serial correlation or simultaneous equation bias. The Durbin-Watson statistic in virtually all equations was less than one. This is as much symptomatic of omitted variables or incorrect functional form as it is characteristic of the purely random part of the relationship.

Summary of Sectoral (per capita) Employment Elasticities (with respect to after-tax wages relative to sectoral product prices)

Sector (sample)	Elasticity	at (wage, employment)	Wage Concept
Agriculture			
1947-73	-1.34	($3.085, 31.1)	TPNAG-hr/PRFI
	-4.04	($4.313, 14.5)	
1974-92	-0.43	($4.542, 14.1)	TPNAG-hr/PRFI
	-0.40	($4.313, 14.5)	
	-1.00	($7.520, 10.1)	
Wholesale Trade			
1964-92	-1.59	($270.9, 21.7)	WHSL-wk/CPIU
	-1.00	($220.8, 28.0)	
Retail Trade			
1964-92	-0.98	($151.7, 63.3)	RETL-wk/CPIU
	-1.00	($153.2, 62.7)	
1969-92	-0.79	($146.2, 66.3)	RETL-wk/CPIU
	-1.00	($166.0, 59.3)	
Finance, Insurance, and Real Estate			
1964-92	-1.42	($236.2, 21.4)	FIRE-wk/CPservice
	-1.00	($201.6, 25.8)	
Services			
1964-92	-2.49	($208.4, 75.7)	SERV-wk/CPservice
	-1.00	($146.1, 132.0)	
1964-92	-2.34	($6.737, 75.7)	TPNAG-hr/CPservice
	-1.00	($4.810, 126.3)	
Transportation & Public Utilities			
1964-92	-0.09	($372.3, 21.9)	TPU-wk/CPpub.trans
Manufacturing (weekly wage rate not deflated by product price)			
Total Manufacturing			
1964-92	-0.18	($211.8, 87.0)	after-tax wage
Durable goods			
1964-92	-0.18	($211.8, 50.9)	after-tax wage
Nondurable goods			
1964-92	-0.19	($211.8, 36.1)	after-tax wage

No acceptable results for Mining, Construction, and Government.

The mnemonics are translated below:

PRFI	Prices Received by Farmers Index (1977 = 1.00, 1992 = 1.39), used in agriculture sector demand.
CPIU	Consumer Price Index Urban Wage-earners (1982-84 = 1.00, 1992 = 1.402), used in wholesale and retail trade sectors.
CPservice	Consumer Price Index for all services (1982-84 = 1.00, 1992 = 1.52), used in service sector and FIRE (finance, insurance, and real estate) sector.
CPpub.trans	Consumer Price Index for public transportation (1982-84 = 1.00, 1992 = 1.507), used in the transportation and public utility sector.
TPNAG	Total Private Nonagricultural Sector average wage [-hr = per hour], [-wk = per week] (1992 = $10.59/hour, or $364.30/week).
WHSL	Wholesale Trade average wage (1992 = $435.48/week).
RETL	Retail Trade average wage (1992 = $205.63/week).
FIRE	Finance, Insurance, and Real Estate average wage (1992 = $387.36/week).
SERV	Services average wage (1992 = $342.55/week).
TPU	Transportation and Public Utilities average wage (1992 = $523.41/week).
MFG	Manufacturing average earnings (1992 = $469.45/week).

Where the elasticity equals one, wage income in that sector (real income per 1,000 population) is at its maximum. The retail sector passed through that point when employment and wages reached 1976-77 levels. Subsequent employment increases came as a result of declining real wages that reduced total payments to retail workers (per 1,000 US population).

The finance, insurance, and real estate sector passed through that point when employment exceeded 1985 levels. The 1992 employment in FIRE (at 26.1 per 1,000 population) is just slightly higher than the point of maximum per capita wage income. Real per capita income originating in FIRE is just less than its maximum.

Neither wholesale trade nor services has yet passed the point where employment demand has unit elasticity, but both are close to that point. A 10-percent fall in real wages from 1992 levels in wholesaling would take us there. A 12 to 18 percent drop in the real wages in services would take that sector to the point of unit elasticity.

Agriculture, manufacturing, and transportation-public utilities are all in a region of inelastic demand for labor (per 1,000 population). That means that whenever real wages fall, total income in those industries also decreases, despite the small increase in employment there. Or, if real wages rise, total income rises, despite the loss of jobs. No acceptable relationships were found for the mining, construction, and government sectors of the economy. These 3 sectors account for slightly less than one-fourth of US employment, 100.2 workers per 1,000 population in 1992. Average wages in these sectors are substantially above the average in the rest of the economy.

In summary, slightly more than half the jobs are in sectors with elastic-to-unitary demand, a little less than one-quarter of the jobs are in sectors with inelastic demand for labor, and about a quarter of the jobs are in sectors without an acceptable relationship. The consequence is that the aggregate relationship is unlikely to be very elastic. Thus, increasing the number of jobs per capita can no longer be counted on to increase the (wage-based) income per capita. To increase per capita income, real wages have to increase or new jobs be created in the high wage (high productivity?) sectors of the economy.

Conclusion

I wish I could have identified the policy levers or program variables needed to solve the problem, and create the requisite number of high wage jobs. However, my econometrics did not focus on policy variables. Moreover, my Department is not the major actor on labor market policies. Thus, I must content myself with identifying the problem and bringing that to the attention of folks, like you, with the influence and information needed to solve the problem.

To repeat, the problem I've identified is: Per capita income is down, because the per capita wage payments are down. Per capita wage payments are down because employment per capita is down. Per capita wage payments can rise when lower wages increase employment only if labor demand is elastic. But, the evidence is that labor demand is becoming increasingly inelastic. Thus, to raise incomes, wages, and employment the entire demand schedule has to shift outward -- and that means creating more new jobs in the high-wage (high-productivity) economy.

Sectoral Labor Demand Equations, 1964-92, except as noted.

Sector[b/]	Intercept	Slope	Wage Concept[a/]	R^2	S.E.
Agric.[1/]	72.88	-13.547	TPNAG-hr/PRFI	.901	3.86
	(25.44)	(-15.11)			
		e =	-1.34 at ($3.085, 31.1)		
Agric.[2/]	20.26	-1.347	TPNAG-hr/PRFI	.918	0.34
	(44.94)	(-13.78)			
		e =	-0.43 at ($4.542, 14.1)		
Wholesale	56.08	-0.127	WHSL-wk/CPIU	.786	1.11
	(16.18)	(-9.95)			
		e =	-1.59 at ($270.9, 21.7)		
Retail	125.34	-0.409	RETL-wk/CPIU	.906	3.20
	(32.12)	(-16.09)			
		e =	-0.98 at ($151.7, 63.3)		
Retail[3/]	118.51	-0.357	RETL-wk/CPIU	.913	2.55
	(34.06)	(-15.17)			
		e =	-0.79 at ($146.2, 66.3)		
FIRE	51.60	-0.128	FIRE-wk/CPservice	.930	1.08
	(32.14)	(-18.98)			
		e =	-1.42 at ($236.2, 21.4)		
Services	264.08	-0.904	SERV-wk/CPservice	.881	7.69
	(19.74)	(-14.16)			
		e =	-2.49 at ($208.4, 75.7)		
Services	252.62	-26.262	TPNAG-hr/CPservice	.917	6.41
	(24.56)	(-17.32)			
		e =	-2.34 at ($6.737, 75.7)		
Transportation & Public Utilities	23.88	-0.005	TPU-wk/CPpub.trans	.235	0.60
	(34.06)	(-2.88)			
		e =	-0.09 at ($372.3, 21.9)		
Total Mfg	103.07	-0.076	MFG-wk[4/]	.837	3.34
	(68.62)	(-11.76)			
		e =	-0.18 at ($211.8, 87.0)		
Durable Mfg	59.99	-0.043	MFG-wk[4/]	.712	2.71
	(49.33)	(-8.18)			
		e =	-0.18 at ($211.8, 50.9)		
Nondur. Mfg	43.08	-0.033	MFG-wk[4/]	.946	0.79
	(121.59)	(-21.79)			
		e =	-0.19 at (211.8, 36.1)		

a/ All wage rates multiplied by (1-tx-ss) to obtain after-tax wages.
b/ Dependent variable is sectoral employment per 1,000 US population.
1/ 1947-73
2/ 1974-92
3/ 1969-92
4/ Manufacturing wages were not deflated by a product price index.

Sectoral Labor Demand Equations, 1947-64, except as noted.

Sector[b/]	Intercept	Slope	Wage Concept[a/]	R^2	S.E.
Nondur. Mfr	73.54	-5.653 (-20.10)	TPNAG-hr/CPIU	.962	0.71
	elasticity =		-0.68 at ($5.27, 43.8)		
Durable Mfr	74.48	-3.751 (-2.74)	TPNAG-hr/CPIU	.320	3.46
	elasticity =		-0.36 at ($5.27, 54.7)		
Total Mfr	148.02	-9.405 (-6.00)	TPNAG-hr/CPIU	.692	3.97
	elasticity =		-0.50 at ($5.27, 98.5)		
Mining	13.50	-1.623 (-23.60)	TPNAG-hr/CPIU	.972	0.20
	elasticity =		-1.96 at ($5.27, 4.9)		
Agriculture	116.85	-15.074 (-17.22)	TPNAG-hr/CPIU	.949	2.22
	elasticity =		-2.12 at ($5.27, 37.5)		
Nonag Self-employed & Household workers[1/]	63.83	-5.899 (-6.88)	TPNAG-hr/CPIU	.518	4.08
	elasticity =		-1.22 at ($5.94, 28.8)		
Transportation & Public Utilities[2/]	39.59	-2.797 (-13.64)	TPNAG-hr/CPIU	.809	0.97
	elasticity =		-0.72 at ($5.94, 23.0)		

a/ All wage rates multiplied by (1-tx-ss) to obtain after-tax wages.
b/ Dependent variable is sectoral employment per 1,000 US population.
1/ 1947-92. Includes unpaid family workers. This represents the difference between the employment concept in table B-30 and that in B-41 of the Economic Report of the President.
2/ 1947-92.

FORECAST EVALUATION

SESSION I

MODERATOR Thomas Lienesch, Bureau of Economic Analysis

PRESENTATION 1 Vernon Dale Jones and Stuart Bretschneider, The Maxwell School, Syracuse University

PRESENTATION 2 John Kitchen, Council of Economic Advisors and Economic Research Service

SESSION II

MODERATOR Steve MacDonald, Economic Research Service

PRESENTATION 1 David Stallings and LaWanda Musgrove, World Agricultural Outlook Board

PRESENTATION 2 Robin Reichenbach, Energy Information Administration

Forecast Evaluation

Organizational Pressures on Forecast Evaluation: Managerial, Political and Procedural Influences

Vernon Dale Jones and Stuart Bretschneider, The Maxwell School, Syracuse University

Abstract

This paper proposes a theory to explain why some forecasting organizations institutionalize forecast accuracy evaluation while others do not. The theory considers internal and external aspects of managerial, political and procedural factors as they effect forecasting organizations. The theory is then tested using data from a survey of the Federal Forecasters Group. Though some support for the theory is developed, multiple alternative explanations for results and the "public" nature of the sample organizations prevent wide-scale generalization. The results suggest that larger organizations are more likely to have some form of forecast evaluation compared to smaller units. The institutionalization of forecast accuracy evaluation is closely linked to internal managerial and procedural factors, while external political pressure tends to reduce the likelihood of institutionalization of evaluation of forecast accuracy.

Keywords: Institutional, organizational, managerial, political, and procedural pressures and influences on forecasting; implementation of forecasts; users of forecasts.

Introduction

Forecasting is applied to almost all decision-making and policy-making in organizations. Managers in government and business organizations, large and small organizations, and technically sophisticated and technically unsophisticated organizations, all rely on some kind of forecasting. Over the past few years, the scope of forecasting and its importance to organizations have grown to include planning, decision-making, as well as other management issues beyond the traditional technical concerns over which forecasting methods should be used.

As the scope and importance of forecasting in organizations grow, there are a variety of environmental pressures, both external and internal to organizations, which influence the forecasting process. Users of forecasts want forecasts which will enable them to succeed in an environment which is increasingly complex, interdependent, and uncertain. To produce accurate and credible forecasts, forecasters need an appreciation of factors which influence the forecasting process as conducted in their organizations. The purpose of this paper is to develop and test a theory to explain why some public organizations decide to implement formal processes of forecast evaluation. More specifically, since forecasting tends to be implemented by technical support units within larger organizations, this research focuses on why units responsible for generating forecasts useful to other elements of the larger organization are likely to formally evaluate themselves.

The first section of the paper reviews the literature associated with influences on forecasting. The next develops a formal model along with a series of empirical hypotheses that explain how external and internal organizational factors influence the likelihood a forecasting subunit will enact a formal forecast evaluation process. The third section of the paper describes the data collection process. Data to test aspects of the model come from the Federal Forecasting Practices Study Project. The empirical findings of the study are summarized in the fourth section. Finally, the paper concludes with a summary of our findings including recommendations for future research.

Forecasting Research

The Need to Focus on Broader Forecasting Issues

Research on forecasting emphasizes the development and testing of new forecasting methodologies. Furthermore, most of the empirical work focuses on the relative accuracy of techniques, ways of combining

forecasts and forecast methods to improve accuracy, methods for measuring accuracy, and techniques for achieving accuracy in different applied settings (Schultz, 1992). More attention needs to be focused on how managerial, political, and procedural factors influence the forecasting process in organizations; implementation of forecasts; uses of forecasts; and credibility of forecasts. A few forecasting researchers in recent years have taken note of this deficiency (Bretschneider and Gorr, 1989, 1991; Makridakis, 1991; and Schultz, 1992).

Influences on Forecast Accuracy

Bretschneider and Gorr (1987) argued that forecast accuracy of state government revenue forecasts was directly related to the level of political conflict/cooperation present in state government. They also argued that the organizational design of the forecasting process directly influenced the accuracy of the forecasts. Bretschneider, Gorr, Grizzle, and Klay (1989) tested and extended this theory. Their empirical results suggest that organizational, political, and forecast process variables influence accuracy as much as the forecasting methods. In particular, they found that forecast accuracy increases when independent forecasts from competing agencies are used. Furthermore, accuracy increases even more when formal procedures exist to combine competing forecasts generated by executive and legislative branch agencies. The use of outside expert advisors has a surprisingly negative effect on forecasting accuracy of state government revenues. Shkurti and Winefordner (1989) examined the relationship between forecasting and the political process in state revenue forecasting in Ohio. Their findings supported the earlier work and found that forecast accuracy is enhanced in an environment with strong institutional pressures and highly partisan politics.

Influences on Forecast Bias

Kamlet, Mowery, and Su (1987) found no evidence that partisan politics biased short-term economic forecasts by the executive branch or Congress.

On the other hand, Larkey and Smith (1984, 1989) and Feenberg, Gentry, Gilroy, and Rosen (1989) discovered a conservative bias or underestimation in state revenue forecasts explained by the need to generate a hedge or buffer against future revenue uncertainty. Cassidy, Kamlet, and Nagin (1989) found no systematic relationship between revenue forecast errors and state political and institutional factors but instead attributed forecast bias directly to the level of error (uncertainty) in economic projections used to forecast revenues. These researchers attributed a modest tendency towards conservative forecasting to economic uncertainty.

Furthermore, their research uncovered such a small amount of state revenue estimation bias that they cst doubt on previous findings of conservative bias. In a reexamination of bias in state government revenue forecasting, Bretschnieder and Gorr (1992) found that organizational design, political environment, and economic uncertainty factors all contributed to bias in forecasting state sales tax revenues. This study uncovered complex interaction effects between recessionary cycles (economic uncertainty) and the level of political conflict. The nature of bias and the impact of politics changed significantly depending upon whether a state government was at the beginning, middle, or end of the economic and political cycle.

Evaluations Within Public Organizations

A rich prescriptive literature exists on public program evaluation and evaluation research (Rossi and Freeman, 1985; and Dynes and Marvel, 1987). However, this literature is concerned with program conceptualization and design interventions, monitoring of program implementation, and assessment of program efficiency, effectiveness, and impact (Downs and Larkey, 1986). Additionally, the public strategic management and strategic planning literature addresses strategic management and planning influence on and contribution to public sector productivity (Nutt and Backoff, 1987; Bryson, 1988; and Halachmi, 1992). Little work has been done on why some organizations evaluate their activities while others do not. This study fills a gap in the literature by studying public organizations and the influences upon their decision to implement formal processes of forecast evaluation.

Research Theory and Hypotheses

Noting that little research has been conducted on organizational aspects of forecasting, Bromiley (1987) suggests areas for consideration. This study relates to three of his suggested areas by considering potential pressure factors from an organization's external and internal environment which influence the formality of forecast evaluation.

In this study, the unit of analysis is the forecasting unit of a large U.S. federal government agency. Therefore, "external" influences refer to those pressures emanating from outside the forecasting unit. This could include the large agency, Congress, the Office of the President, or even agency clients. "Internal" influences refer to those forces originating from within the forecasting unit.

Figure 1 organizes these ideas into a model, in which both internal and external pressures to evaluate forecasts may derive from managerial, political or procedural sources. Managerial pressures originate from the forecasting unit's organizational structure and role. For example, the level of resources committed to the forecasting unit, or its formal position within the agency are likely to influence both willingness and ability to carry out formal evaluation.

Political influences refer to both formal and informal efforts to manipulate the forecasting unit's activities and behaviors. For example, conscious policies regarding the level of importance placed on forecasting, the extent of usage of the forecasts by other organizations, and the degree of oversight of forecasting activities.

Procedural influences are task and technology characteristics of the forecasting unit's day-to-day activities. For example, characteristics of the forecasts being generated such as the level of aggregation in data, data quality, time horizon of forecasts (e.g. short vs. long), complexity of the forecast methods in use by the unit, and the stability of phenomena forecasted.

The dashed box in the model represents a formal forecasting group within an organization. The external and internal pressures influence the forecasting process by directly affecting the forecaster and the forecaster's development of forecasts. This occurs because the forecaster's valuation of criteria for forecast quality is influenced which in turn influences the forecasts produced. Environmental pressures alter the level of perceived importance placed on different criteria by the forecaster for evaluating the quality or usefulness of a forecast. Certain criteria can be assessed using formal evaluation while others mitigate the use of evaluation. For example, as external political pressure to manipulate forecasts increase, formal evaluation is likely to be less important since formal evaluation would tend to highlight the effects of political manipulation. On the other hand, as criteria related to technical accuracy or other scientific/professional norms become more important, then formal evaluation could provide valuable information towards enhancement of the forecasting process and the forecasts themselves.

Hypotheses

Six hypotheses have been formulated based upon our model. There is one hypothesis for each external and internal organizational factor influencing the forecasting process.

> *Hypothesis 1.* External managerial. If the forecasts are generated by the organization for its own use for external purposes, then the forecasts produced by the forecasting group will be subjected to greater scrutiny and the more formal forecast accuracy evaluation will be.

Forecasts are intended for external or internal managerial purposes by the organization producing them. Possible purposes of forecasts include financial management, management of direct service delivery of public goods, production planning, auditing the activities of others, and policy analysis and evaluation. Forecasts developed for external managerial purposes will generally be presented to more people, some outside in addition to those inside the organization. The visibility of forecasts produced for external managerial purposes is greater since the impacts go beyond the organization. Thus, the forecasting group's activities and forecasts are expected to come under greater scrutiny. Such closer examination is likely to result in greater pressure to make use of forecast evaluation in order to ensure higher quality forecasts.

> *Hypothesis 2.* Internal managerial. The greater the level of agency level resource support for forecasting, the more formal forecast accuracy evaluation will be.

An organization which places a high value on forecasting will provide more resource support for forecasting. Such support is best reflected in the relative amount of resources given to the forecasting group. The annual budget or the number of employees are two such measures of support. If an organization makes the decision to increase resources to the forecasting function at the expense of other suborganizational functions, then it is likely to require a return from its investment. Thus, the agency will produce external pressure on the forecasting unit to conduct formal evaluations of its activities to confirm that resources dedicated to forecasting are effective.

Hypothesis 3. External political. If there is external organizational interest in or use of the forecasts by or for political influence, then accuracy as a criteria is likely to be less important. Consequently, an increase in external political influence is likely to reduce the likelihood that a forecasting unit will formally evaluate its forecast accuracy.

External organizations sometimes are interested in or use the forecasts generated by a forecasting group. For example Senate members or their staffs, House members or their staffs, the Office of Management and Budget, judicial organizations, the press, and industrial groups are all potential consumers of forecasts generated by different executive agencies like USDA or the Department of Labor. For forecasting groups which routinely provide forecasts to external groups, it is expected they will experience great pressure to generate accurate forecasts. On the other hand, as competing political forces come into play, there is a growing potential that forecasts will be used to support preconceived policy positions without regard to the potential accuracy of the forecast. Thus, the direction of effect for external political influence depends on whether the external actors are motivated by overall accuracy or preconceived policy positions.

Hypothesis 4. Internal political. The greater the internal use of the forecasts and the greater the pressure for the forecasting group to deliver accurate forecasts, the more formal forecast accuracy evaluation will be.

An organization sometimes uses for internal purposes its own forecasts produced by the forecasting group. The organization's management or technical staffs use the forecasts for a variety of reasons. For forecasting groups which provide their forecasts to internal users and experience pressure from internal users for accurate forecasts, it is expected that forecast accuracy evaluation will be more formal.

Hypothesis 5. External procedural. If the forecasting group faces competition by external forecasting sources, then the forecasts produced by the forecasting group will be subjected to greater scrutiny and the more formal forecast accuracy evaluation will be.

External procedural factors such as laws, regulations, user preferences, and competition influence the forecasting process. Competition has a potentially strong influence on the forecasting process. An organization's forecasting group which competes with other external sources of forecasts will seek to generate accurate forecasts to remain a primary source of forecasts. Therefore, the forecasting group's procedures will come under greater scrutiny in order to determine forecast accuracy and consequent competitiveness. Thus, more formal forecast accuracy evaluation will occur.

Hypothesis 6. Internal procedural. The broader the scope and greater the complexity of the forecasting process, the more formal forecast accuracy evaluation will be.

Numerous methods can be used to forecast and include such categories as judgmental, time series extrapolation, single-equation econometrics regression, multi-equation/simultaneous equation econometrics regression, intentions survey, and combination. The more complex the methods used and the more frequent the more complex methods used, then the greater the procedural complexity for the forecasting group. Because of the greater procedural complexity, the organization seeks greater control over the forecasting process to ensure accuracy. Thus, it is expected that forecast accuracy evaluation will in turn be more formal.

Data Collection and Measurement

The Federal Forecasting Practices Study Project was sponsored by The International Institute of Forecasters, Syracuse University's Technology and Information Policy Program, and Carnegie Mellon University's School of Urban and Public Affairs.

Study Sample

Data were collected in 1990 from a mail survey of members of the Federal Forecasters professional organization. Members of the Federal Forecasters either have forecasting as part of their job responsibilities in the federal government or have an interest in forecasting. Membership is voluntary and the principle activity of the organization is an annual conference where ideas are exchanged and discussed through keynote speeches and presentation of papers. The sample frame consisted of 259 members listed in the organization's 1989 and 1990 directories. Thus, the sample for the study was not a random sample of all the forecasters working in the federal government. Nevertheless, the sample did represent a diversity of federal agencies by type and size and provides data for investigating managerial, political, and procedural influences on forecast evaluation. Respondents were from a broad representation of federal agencies including the Departments of Agriculture, Commerce, Defense, Energy, Health and Human Services, and Labor.

Study Survey

The overall response rate for the survey was 60%. Several types of responses were received other than completed surveys. These included individuals who returned uncompleted surveys because they no longer had job responsibilities which included forecasting or because others in their organization had already filled out a survey. Of the total number of respondents, over 75% were actively involved with developing or reviewing forecasts for their organizations. A total of 115 useable survey responses were obtained from 25 agencies for a 45% sample.

Additionally, the nature of the forecasting groups represented varies significantly. Not only were a wide variety of agencies represented in the sample, these organizations also varied significantly in size. Though the mean forecasting group employed nearly 13 full-time equivalent staff members and had a budget of over $800,000, the median group only employed five technical staff members with a budget of $300,000. It is clearly evident the sample contained a few very large forecasting groups, while most employ ten or fewer technical staff members.

Discussion of Analysis and Results

Variables

The dependent variable was the whether a forecasting unit had "no forecast accuracy evaluation process," " an informal forecast accuracy evaluation process," or "a formal forecast accuracy evaluation process."

Independent variables representing the six external and internal influences on the establishment of a formal evaluation process were developed. First, we considered managerial influences. To capture the effect of **external managerial** pressure we used a survey question designed to tap the extent to which the agency made use of forecasts for the purpose of policy analysis and evaluation. The variable is based upon combining two measures: the importance of the effect of proposed policy change and the importance of the effect of implemented policy change as purposes for forecasts. As the above hypothesis 1 suggests, a major external application of accurate forecasting is the use of forecasts for policy analysis by the agency and program evaluation. As the importance of forecasts generated by the unit for both policy analysis and policy implementation increases, the need for accurate prediction also increases. The increased need for accuracy results in more pressure on the forecasting unit to evaluate forecast accuracy. To capture the effects of **internal managerial** pressure we measured the number of full-time equivalent forecasting positions. This measure captured the impact of resource commitment by the agency to the forecasting process as well as the organizational slack resources the forecasting unit might have available to commit to additional activities. Increased resources enhance both potential and likelihood the forecasting unit can and will devote resources to forecast accuracy evaluation.

Political influences can derive from both the agency/forecasting unit or externally from Congress or other executive branch agencies. A variable measuring the extent of **external political** pressure and pressure to adjust

forecasts captured the effects of external political pressure of the forecasting unit. The variable comes from combining two external use factors (use by the Senate and use by the House) to create a measure of use and then interacting the result with a measure of political pressure to adjust forecasts. While the survey question focusing on political pressure to adjust forecasts tapped the aggregate effect of external political pressure, interacting it with use of forecasts by both houses of Congress focuses the effect on external political influence. As external political pressure mounts, there is an increased potential to make use of forecasts for political ends. This is likely to reduce the likelihood accuracy as a criteria will be important. Thus, this factor will tend to decrease the likelihood a forecasting unit will have a formal forecast accuracy evaluation process. Internal political pressure was captured by interacting the measure of the internal use of forecasts with the measure of political pressure to adjust forecasts. Internal political pressure for using forecasts by internal staff and pressure to adjust forecasts by other agencies of similar quantities (the presence of competition), and use of single-equation econometrics regression methods are the independent variables measured. Again, as this variable is designed to pick up the extent forecasts are being manipulated, the factor is political. The extent to which the origin of the activity is internal to the agency or forecast unit, the less likely this variable taps preconceived political positions. Thus, this factor is expected to be positively related to the existence of a formal evaluation process within the forecasting unit.

Procedural influence derive from institutional and process-based rules. External procedural influence is represented by whether other agencies produce forecasts of the same phenomenon. If institutional/procedural conditions have led to the existence of competition, then the forecasting unit is likely to be more motivated to evaluate their own forecasts. Finally, to tap the effects of internal procedural influence, an independent variable should focus on the technology of forecasting in use within the forecast unit. A major distinction in forecasting methodology occurs once the unit moves to the use of correlative/causal models for generating forecasting. Thus, a binary variable from the survey that indicates that the forecasting unit is using single-equation econometrics regression methods discriminates between units making use of causal models and those that do not. We anticipate that as more sophisticated methods are used that professional values and training will combine to enhance the likelihood that formal evaluation will take place.

Table 1 presents the means and standard deviations for the measures used for the forecast accuracy evaluation independent variables. Table 2 summarizes how the combined and/or interaction variables were derived.

The multinomial logistic regression procedure assumes that the dependent variable is a multi-level categorical measure, which is then modeled as a logarithmic function of an odds ratio. There are two odds ratios in this analysis with both having the probability of "formal forecast accuracy evaluation" in the denominator. The first estimates the conditional probability of "no forecast accuracy evaluation" relative to "formal forecast accuracy evaluation" (log y1/y3) and the second estimates the conditional probability of "informal forecast accuracy evaluation" relative to "formal forecast accuracy evaluation" (log y2/y3) . The parameters of the multinomial logistic regression model are estimated based on a maximum likelihood procedure.

A negative sign for an estimate indicates that an increase in the predictor value decreases the value of the odds ratio. This implies that the numerator probability would be expected to decrease relative to the denominator probability. In other words, a negative sign indicates that an increase in the independent variable decrease the likelihood of "no forecast accuracy evaluation" relative to "formal forecast accuracy evaluation." This also implies a decrease in "informal forecast accuracy evaluation" relative to "formal forecast accuracy evaluation".

On the other hand, a positive sign for an estimate indicates that an increase in the predictor value increases the value of the odds ratio. This implies that the numerator probability would be expected to increase relative to the denominator probability. Therefore, a positive sign indicates that as the independent variable increase the likelihood of "no forecast accuracy evaluation" or "informal forecast accuracy evaluation" will also increase relative to "formal forecast accuracy evaluation."

Results

Tables 3 and 4 present the results for estimating the multinomial logistic regression equations. Table 3 displays the results for the log odds of no evaluation versus formal evaluation, while Table 4 lists the results for informal evaluation versus formal evaluation. Given that only 8 forecasting units indicated no form of evaluation, it is not surprising that only one term in the model was statistically significant. We will therefore focus our attention on the results presented in Table 4.

The logarithm of full-time equivalent staff positions is essentially a measure of the size and resource endowments of the forecasting unit. This term is statistically significant in both models and suggests that as the

forecasting unit becomes larger and has greater resources it is more likely to have a formal forecast accuracy evaluation process. Clearly larger organizations have the resources and capacity to enact more specialized activities such as evaluation. This suggests capacity, not motivation.

The use of single-equation econometrics regression methods is also statistically significant in explaining the log odds for informal evaluation relative to formal evaluation. The direction of effect associated with a forecasting unit that makes use of more sophisticated techniques and causal modeling is to increase the likelihood that the unit will implement a formal forecast accuracy evaluation process. There are several possible explanations for this result. The more sophisticated methods and causal techniques are typically carried out by technically trained and professionally oriented staff. Thus, by background and training, individuals within the forecasting organization may be heavily influenced by cannons of professional behavior. For example, econometricians working on causal models have long been motivated to conduct ex-post sample evaluations of their models (McNees, 1992).

Table 1

Means and Standard Deviations for Measures Used for Forecast Accuracy Evaluation Independent Variables (n=97)

Variable	Mean	Standard Deviation
Number of full-time equivalent forecasting positions	12.70	17.50
Importance of effect of proposed policy change as purpose of forecasts	3.49[1]	1.70
Importance of effect of implemented policy change as purpose of forecasts	3.42[1]	1.67
Forecast use by Senate members or staff	2.42[2]	1.08
Forecast use by House members or staff	2.41[2]	1.09
Forecast use by agency technical staff	3.25[2]	1.14
Political pressure to adjust forecasts	1.63[3]	1.30
Use of aggregate single-equation econometrics regression methods	3.10[3]	1.56
Use of disaggregated single-equation econometrics regression methods	2.67[3]	1.76
Another agency forecasting similar quantities (competition)	0.58[4]	0.49

Means and Standard Deviations for Measures Used for Forecast Accuracy Evaluation Independent Variables (n=97)

Variable	Value	Frequency	Percentage
Level of forecast accuracy evaluation	None	8	8.5
	Informal	51	54.3
	Formal	35	37.2

1 1 = "insignificant" and 5 = "essential."
2 1 = "no use" and 5 = "heavy use."
3 1 = "very rarely" and 5 = "almost always."
4 Based on a nominal scale where 0 = "no" and 1 = "yes."

Table 2

Combined and Interaction
Forecast Accuracy Evaluation Independent Variables

Type Variable	Variable	Factors
Combined	Purpose of forecasts for policy analysis and evaluation	Importance of effect of proposed policy change as purpose of forecasts + Importance of effect of implemented policy change as purpose of forecasts
Combined and Interaction	External political pressure from internal use and pressure to adjust forecast	(Forecast use by Senate members or staff + Forecast use by House members or staff) * Political pressure to adjust forecasts
Interaction	Internal political pressure from internal use and pressure to adjust forecasts	Forecast use by agency technical staff * Political pressure to adjust forecasts
Combined	Use of single-equation econometrics (regression) methods	Use of aggregate single-equation econometrics methods + Use of disaggregated single-equation econometrics methods

Table 3.

Logistic Regression Results for the Probability of "No Forecast Accuracy Evaluation" versus "Formal Forecast Accuracy Evaluation" (n=94)

	Estimate	Standard Error	Chi-Square	Probability
Purpose of forecasts for policy analysis and evaluation	-0.24	0.20	1.40	0.24
Logarithm of full-time equivalent positions	-1.33	0.60	4.92	0.03*
External political pressure from external use and pressure to adjust forecasts	-0.46	0.32	2.09	0.15
Internal political pressure from internal use and pressure to adjust forecasts	0.005	0.25	0.00	0.98
Forecasting by other agencies (competition)	0.99	0.77	1.63	0.20
Use of single-equation econometrics models	-0.32	0.21	2.33	0.13

* Statistically significant at the 5 percent level.
** Statistically significant at the 10 percent level.

Table 4.

Logistic Regression Results for the Probability of "Informal Forecast Accuracy Evaluation" versus "Formal Forecast Accuracy Evaluation" (n=94)

	Estimate	Standard Error	Chi-Square	Probability
Purpose of forecasts for policy analysis and evaluation	-0.14	0.11	1.59	0.21
Logarithm of full-time equivalent positions	-0.98	0.29	11.18	0.0008*
External political pressure from external use and pressure to adjust forecasts	0.11	0.06	3.09	0.08**
Internal political pressure from internal use and pressure to adjust forecasts	-0.09	0.07	1.35	0.24
Forecasting by other agencies (competition)	-0.43	0.28	2.29	0.13
Use of single-equation econometrics models	-0.20	0.11	3.67	0.06**

* Statistically significant at the 5 percent level.
** Statistically significant at the 10 percent level.

One of the more interesting results presented in Table 4 are those associated with the effects of external political influence on the likelihood a forecasting unit will formally evaluate the forecast accuracy of their forecasts. In the original hypothesis, it was felt that to the extent external political influence was motivated by preconceived political positions, the effect would be to reduce the likelihood of formal evaluation. Ex-post accuracy would not be as meaningful to external political users as short-term political support for a pre-existing position. The empirical results provides some support for this view. The coefficient is positive suggesting that increasing external political influence increases the log odds that informal evaluation will occur relative to formal evaluation. The term is statistically significant but only at the 10% level.

The results in Table 4 suggest a mix of internal and external influences on the forecasting unit. Managerial and procedural factors tend to originate internally and have a positive influence while political influence is external and negative. Generally, internal factors have a greater impact on the establishment of a formal process than external factors. Size and resource endowments have the greatest impact on the likelihood that the forecasting unit will formally evaluate forecast accuracy.

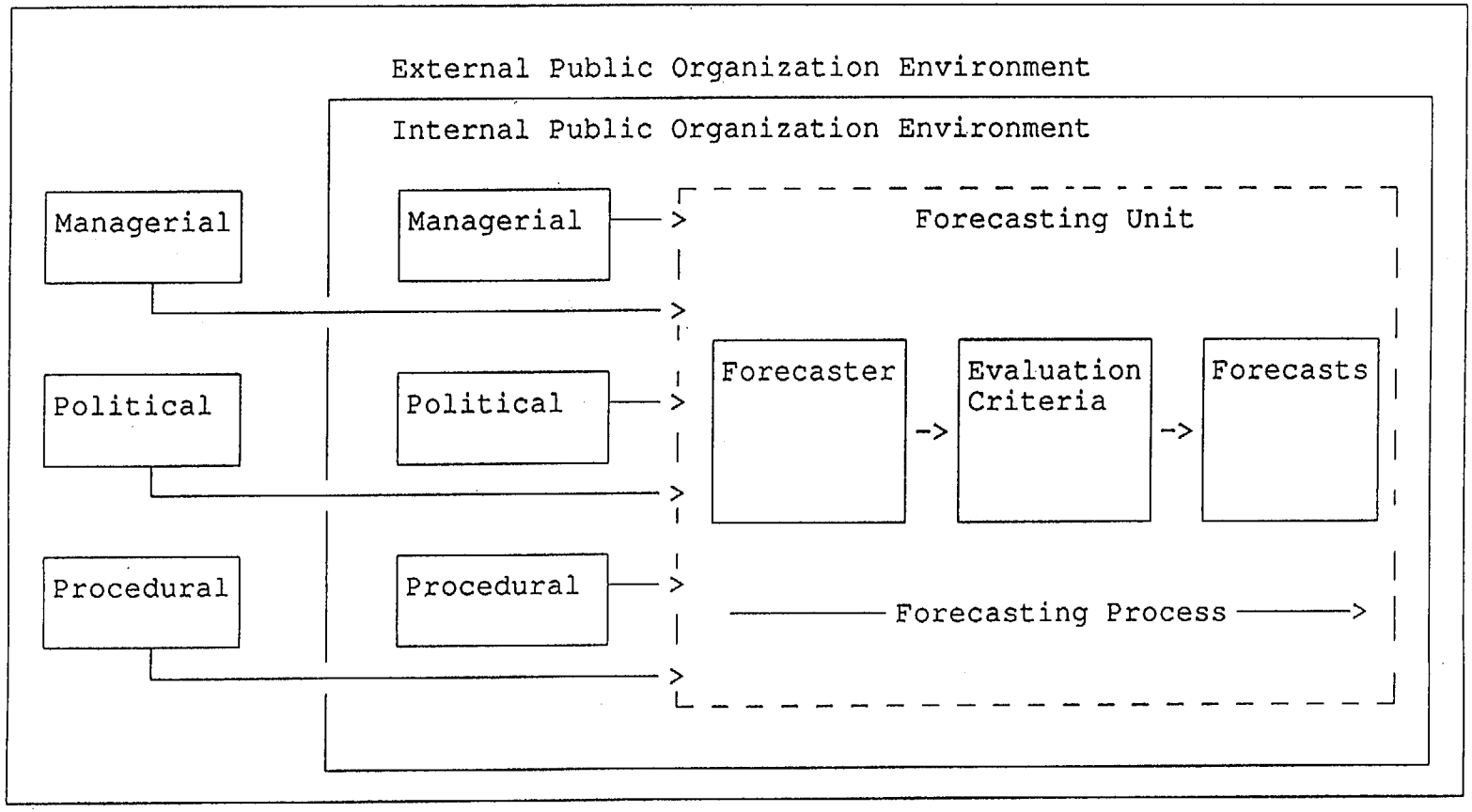
Conclusions

In considering the organizational impacts on implementation of formal process for the evaluation of forecast accuracy, several empirical results emerge. First, a mix of internal and external influences exist. The internal managerial and procedural influences tend to have a dominant and positive effect. "Good" management and process includes some form of evaluation. It is useful to note that over 90% of the sample had some form of accuracy evaluation in place. Larger organizations and those that make use of more sophisticated techniques of forecasting are more likely to establish formal forecasting accuracy evaluation processes than smaller units using less sophisticated techniques. This is not a surprise. Though part of the explanation is simple capacity to do formal evaluation, a part of the explanation is also related to the professionalization of the forecasting unit.

More sophisticated techniques can only be implemented by individuals with advanced training and high levels of professional training. As individuals obtain such training, usually in professional master's degree programs and doctoral degree programs, they become more heavily influenced by the norms of behavior associated with an academic discipline or professional group. In the context of forecasting, particularly with causal models, such norm invariably include the use of forecast accuracy evaluation.

More surprising, though, are results related to external influences on the use of formal forecast accuracy evaluation. Here we find that the principle external influence on the establishment of a formal evaluation is the extent of use by Congressional staff interacted with the level of pressure to adjust or change forecasts. This type of political influence suggests that the forecasts are being used within the context of the political process. Such forecasts are more likely to have short-term political value in the context of specific debates and have little long-term value. Thus, accuracy of a forecast, which can only be determined in a longer run context, has little value. Higher levels of external political influence reduce the likelihood a forecasting unit will formally evaluated accuracy. A similar conclusion was reached by Kraemer et. al. (1987), while studying the impact of various policy models used in the federal government.

Figure 1
External and Internal Organizational Factors Influencing Forecasting



References

Bretschneider, Stuart I. and Gorr, Wilpen L. "Economic, Organizational, and Political Influences on Biases in Forecasting State Sales Tax Receipts." International Journal of Forecasting, 7 (1992): 457-466.

Bretschneider, Stuart I. and Gorr, Wilpen L. "Forecasting Practices in the Federal Government." Federal Forecasters Conference 1991: Papers and Proceedings, (1991): 17-24.

Bretschneider, Stuart I. and Gorr, Wilpen L. "Editorial: Forecasting as a Science." International Journal of Forecasting, 5 (1989): 305-306.

Bretschneider, Stuart I. and Gorr, Wilpen L. "State and Local Government Revenue Forecasting." In The Handbook of Forecasting: A Manager's Guide, 2nd ed., 118-134. Edited by Spyros Makridakis and Steven C. Wheelwright. New York: John Wiley and Sons, 1987.

Bretschneider, Stuart I.; Gorr, Wilpen L.; Grizzle, Gloria; and Klay, Earle. "Political and Organizational Influences on the Accuracy of Forecasting State Government Revenues." International Journal of Forecasting, 5 (1989): 307-319.

Bromiley, Philip. "Do Forecasts Produced By Organizations Reflect Anchoring and Adjustment?" Journal of Forecasting, 6 (1987): 201-210.

Bryson, John M. Strategic Planning for Public and Nonprofit Organizations. San Francisco: Jossey-Bass Publishers, 1988.

Cassidy, Glenn; Kamlet, Mark S.; and Nagin, Daniel S. "An Empirical Examination of Bias in Revenue Forecasts by State Government." International Journal of Forecasting, 5 (3) (1989): 321-333.

Downs, George W. and Larkey, Patrick D. The Search for Government Efficiency: From Hubris to Helplessness. Philadelphia: Temple University Press, 1986.

Dynes, Patrick S. and Marvel, Mary K. Program Evaluation: An Annotated Bibliography. New York: Garland Publishing, Inc., 1987.

Feenberg, Daniel R.; Gentry, William; Gilroy, David; and Rosen, Harvey S. "Testing the Rationality of State Revenue Forecasts." Review of Economics and Statistics, 71 (2) (1989): 300-308.

Halachmi, Arie. "Strategic Management and Productivity." In Public Productivity Handbook, 551-563. Edited by Marc Holzer. New York: Marcel Dekker, Inc., 1992.

Kamlet, Mark S.; Mowery, David C.; and Su, Tsai-Tsu. "Whom Do You Trust? An Analysis of Executive and Congressional Economic Forecasts." Journal of Policy Analysis and Management, 6 (3) (1987): 365-384.

Kraemer, Kenneth L., Dickhoven, Siegfried, Tierney, Susan F., and King, John L., Datawars: The Politics of Modeling in Federal Policy Making, New York: Columbia University Press, 1987.

Larkey, Patrick D. and Smith, Richard A. "Bias in the Formulation of Local Government Budget Problems." Policy Sciences, 22 (1989): 123-166.

Larkey, Patrick D. and Smith, Richard A. "The Misrepresentation of Information in Government Budgeting." In Advances in Information Processing in Organizations, Vol. 1, 63-92. Edited by L. S. Sproull and P. D. Larkey. Greenwich, CT: JAI Press, 1984.

Makridakis, Spyros. "Editorial: Forecasting in the 21st Century." International Journal of Forecasting, 7 (1991): 123-126.

McNees, Stephen, How Large are Economic Forecast Errors?, New England Economic Review, (July/August 1992): 1-18.

Nutt, P. C. and Backoff, R. W. "A Strategic Management Process for Public and Third-Sector Organizations." Journal of the American Planning Association, 53 (1987): 44-57.

Rossi, Peter H. and Freeman, Howard E. Evaluation: A Systematic Approach. 3rd ed. Beverly Hills, CA: Sage Publications, 1985.

Schultz, Randall L. "Editorial: Fundamental Aspects of Forecasting in Organizations." International Journal of Forecasting, 7 (1992): 409-411.

Shkurti, William J. and Winefordner, Darrell. "The Politics of State Revenue Forecasting in Ohio, 1984-1987: A Case Study and Research Implications." International Journal of Forecasting, 5 (1989): 361-371.

Relative Information in Public and Private Sector Macroeconomic Forecasts

John Kitchen, Council of Economic Advisers and Economic Research Service

Introduction

Public sector macroeconomic forecasts--including those of the Administration and the Congressional Budget Office (CBO)--play an important information role for policy makers and the general public. Macroeconomic forecasts are a specific mechanism, and perhaps one of the more obvious ones, through which the Federal government plays the role of an anticipatory government. The forecasts provide information on the outlook for the economy as well as serving as a basis for calculating budget estimates--revenues, outlays, and deficits. Hence, forecasts of revenues, outlays, and deficits can only be as good as the underlying macroeconomic forecasts.

Macroeconomic forecasts can be used to evaluate how well the Federal government has functioned as an anticipatory government--at least in application to economic outlook and budget issues. One way to proceed in such an evaluation is to examine relative forecast accuracy. Alternative public sector forecasts--Administration and CBO forecasts--can be compared. Similarly, comparisons can be made between public sector and private sector macroeconomic forecasts. In fact, it is typically the case that the relative optimism or pessimism of government forecasts is evaluated in comparison to private sector "consensus" forecasts. In the early 1980s, Administration forecasts were criticized as being overly optimistic (through comparison with private forecasts) and the label of "rosy scenario" was born. For example, the 1981 Administration real GNP growth forecast had a cumulative 6-year forecast error that amounted to about 4 percent of GNP (although the forecast error in 1982 due to the recession more than accounted for the total). This issue helps to illustrate that proper functioning of the anticipatory role of government is necessary to maintain policy credibility.

Key questions emerge. Which of the forecasts was "better" at predicting the performance of the economy? Did one forecast carry more information than other forecasts? Did the consensus private forecast outperform public sector forecasts? The research reported in this paper is an attempt to answer such questions.

One might interpret the last question concerning the relative accuracy of public versus private forecasts as an attempt to determine whether the Federal government would have been better off over time to have used the private sector consensus forecast as a basis for its budget estimates, rather than having produced its own macroeconomic forecasts. However, as will be discussed later in the paper, such a view would be based on a misunderstanding of the manner in which Administration forecasts are produced. Administration forecasts are "policy" forecasts that incorporate the effects of the Administration's policy proposals. Hence, at times there may be a valid reason--a policy proposal that was not adopted--for a positive bias of Administration real growth forecasts. A recent example (not included in the sample to be analyzed in this paper) is the Clinton Administration forecast for real GDP growth from February 1993. That forecast was a policy forecast that included estimated short-run beneficial effects that would result from passage of a proposed stimulus package. When the stimulus package was not included as part of the final budget proposal, the Administration's expectation for real growth for 1993 and 1994 declined. In contrast, CBO forecasts are based on a current services baseline and do not include prospective policy changes, while private forecasts presumably incorporate prospective policy changes according to the (subjective) probability that they will pass into law.

Data and Methodology

Data for forecasts of key macroeconomic variables were collected for the 1980 to 1992 period for Administration, CBO, and the Blue Chip consensus forecasts. Data were available for 1-year-ahead forecasts and 2-year-ahead forecasts. This means there are relatively small sample sizes: 13 observations for the 1-year-ahead forecasts and 12 observations for the 2-year-ahead forecasts. For the Administration forecast, the source was the Budget of the U.S. Government; for the CBO, The Economic and Budget Outlook; and, for the Blue Chip consensus, Blue Chip Economic Indicators. While the Budget usually was not released until early February, Administration forecasts typically were made in December with adjustment possible into early January. A similar schedule typically applied to CBO forecasts. Comparable Blue Chip consensus forecasts were taken from January issues of Blue Chip Economic Indicators. Data for the actual values of the forecasted variables were the first-reported final estimates. For example, the first-reported final estimates of real GNP growth and GNP deflator

change for a given year were available at the end of March of the following year when the final estimate of GNP for the fourth quarter was released. There are advantages and disadvantages to this approach. Among the advantages is the fact that the approach reduces problems that might arise from methodological or base-year inconsistencies between the forecasts and actual values. A disadvantage is that the first final estimates do not always incorporate all relevant information, such as occurs when annual revisions are made to the NIPAs or when the accounts are rebenchmarked.[1] Given the alternatives, I chose to use the first-reported final values because they would yield the most legitimate comparison for the information set available when the forecasts were made. Of course, that choice also meant more work at assembling the data set.

Due to differences in published forecast presentations, only a subset of macroeconomic variables can be studied: real GNP growth, GNP deflator inflation, and the unemployment rate.[2] CPI inflation could not be used in the comparisons because, prior to 1992, Administration forecasts were CPI-W forecasts while CBO and Blue Chip forecasts were for the CPI-U (a simple mechanical adjustment between the two measures does not exist). Similarly, there was not a consistent set of interest rate data available for either the short-term or long-term interest rates for the sample period used--Administration and CBO forecasts reported Treasury security rates (and somewhat inconsistently for the CBO) while the Blue Chip survey results typically reported private rates such as commercial paper and corporate bond rates.

The empirical analysis used to compare the alternative forecasts is based on a procedure described by Fair and Shiller (1990) for comparing information in forecasts from alternative econometric models. To examine whether one model's forecast carried different information from an alternative model, Fair and Shiller regressed the actual change in GNP on the forecasted change from the two models. For example,

$$Y_t = a + b\,Y_{1t} + c\,Y_{2t} + e_t$$

where Y_t is the actual value of the variable in period t, Y_{1t} is the forecast of the variable in period t from model 1, and Y_{2t} is the forecast of the variable in period t from model 2. If neither model contained useful information for forecasting Y_t, then both coefficients b and c should be zero. If both models carried independent information, then both b and c will be nonzero. If one model carried all of the information of the other model plus additional information, then the coefficient on the model with more information would be nonzero and the other slope coefficient would be zero.

The application of the Fair and Shiller procedure for analyzing information carried by alternative public and private sector forecasts is straightforward. Regressions of the forecasted variable on the alternative public and private sector forecasts provide information on the relative information carried by the forecasts.

Results

Tables 1 and 2 present results testing for bias in the alternative forecasts as well as root-mean-square error (RMSE) statistics. To examine bias for the alternative forecasts, in each case the forecast error was regressed on a constant. The first column in each table shows the coefficient estimate for the constant term. The standard error and Durbin-Watson statistics from that regression are reported in the second and third columns. The RMSE is reported in the fourth column. If a forecast was unbiased, then the constant coefficient would be zero and the standard error of the regression would equal the RMSE.

For one-year-ahead real GNP growth forecasts, Administration forecasts exhibited the least bias but the greatest variation as revealed by the higher standard error of the regression and the higher RMSE; CBO forecasts had the smallest standard error and RMSE. For the unemployment rate, similar results were observed across the forecasts, with the CBO having a slightly smaller standard error and RMSE. For the GNP deflator change, all of

[1]Comparisons of results for the alternative actual values may be interesting. Estimations using the most-recent reported values have not yet been performed.

[2]Prior to 1992, Administration forecasts reported the total unemployment rate rather than the civilian unemployment rate--the difference attributable to military personnel. The forecasts for the total unemployment rate were raised by 0.1 percentage point to attain the equivalent civilian unemployment rate.

TABLE 1--BIAS AND RMSE FOR 1-YEAR-AHEAD FORECASTS

	CONSTANT	SE	DW	RMSE
GNP GROWTH				
ADMINISTRATION	0.008 (0.022)	1.241	2.31	1.241
BLUE CHIP	-0.200 (0.652)	1.105	2.30	1.125
CBO	-0.100 (0.347)	1.038	2.18	1.043
UNEMPLOYMENT RATE				
ADMINISTRATION	0.092 (0.665)	0.501	2.39	0.510
BLUE CHIP	0.092 (0.656)	0.507	2.06	0.516
CBO	0.092 (0.698)	0.477	2.28	0.486
GNP DEFLATOR INFLATION				
ADMINISTRATION	0.523 (2.800)	0.673	1.02	0.866
BLUE CHIP	0.600 (4.544)	0.476	1.40	0.785
CBO	0.462 (3.197)	0.520	1.69	0.708

NOTES: The first three columns show results from regressions of the forecast on a constant.
t-statistic in absolute value in parentheses.
SE is standard error of the regression.
DW is the Durbin-Watson statistic.
RMSE is the root mean square error for the forecast.

TABLE 2--BIAS AND RMSE FOR 2-YEAR-AHEAD FORECASTS

	CONSTANT	SE	DW	RMSE
GNP GROWTH				
ADMINISTRATION	1.242 (2.047)	2.101	2.19	2.469
BLUE CHIP	0.733 (1.345)	1.889	2.07	2.039
CBO	0.867 (1.685)	1.782	2.08	1.994
UNEMPLOYMENT RATE				
ADMINISTRATION	-0.142 (0.390)	1.259	1.64	1.267
BLUE CHIP	-0.092 (0.241)	1.317	1.25	1.321
CBO	-0.050 (0.146)	1.190	1.78	1.191
GNP DEFLATOR INFLATION				
ADMINISTRATION	1.008 (3.458)	1.051	1.44	1.508
BLUE CHIP	1.223 (3.304)	1.335	1.26	1.892
CBO	1.177 (4.003)	1.060	1.19	1.659

Notes: See Table 1.

the forecasts exhibited positive bias. Evidently, the extent of disinflation of the 1980s generally was unanticipated by public and private forecasters. The CBO again had the lowest RMSE, while the Administration had the highest.

For the two-year-ahead real GNP growth forecasts, each of the forecasts show evidence of positive bias, with the Administration forecast having the largest and most significant bias of about 1 1/4 percentage points. Administration real growth forecasts also had the highest RMSE. Blue Chip and CBO forecasts showed similar positive biases and RMSEs. For the unemployment rate, each of the forecasts had a small negative bias--consistent with the overestimate of real growth, although a typical Okun's law relationship would suggest absolutely larger negative biases given the size of the positive biases for real GNP growth.[3] CBO forecasts for the unemployment rate had the smallest RMSE. For the GNP deflator change, the forecasts had positive and significant biases in the range of 1 to 1 1/4 percentage points. Administration forecasts had the smallest bias and RMSE for the deflator change.[4]

The bias and RMSE estimates generally would indicate the CBO forecasts were the "best" of the three forecasts considered. However, the regression procedure for comparing alternative forecasts described above has some advantages over a simple comparison of bias and RMSE:

> First, if the RMSEs are close for two forecasts, little can be concluded about the relative merits of the two. Second, even if one RMSE is much smaller than the other, it may still be that the forecast with the higher RMSE contains information not in the other forecast. There is no way to test for this using the RMSE framework. (Fair and Shiller, p. 376)

Tables 3, 4 and 5 show forecast comparisons using the regression procedure described above to test for the relative information carried by alternative forecasts.[5] Table 3 shows the comparisons for real GNP growth forecasts. For the one-year-ahead forecast, the regression of real GNP growth on the Administration forecast and the Blue Chip forecast shows a significant coefficient[6] on the Blue Chip forecast and a coefficient on the Administration forecast that is not significantly different from zero. Hence, these results indicate that the Blue Chip forecast carried all of the information contained in the Administration forecast plus additional information. A similar result is observed for the comparison of Administration forecasts with CBO forecasts, with the CBO forecast carrying more information. The comparisons between CBO and Blue Chip forecasts indicate, at most, a marginal information advantage for CBO forecasts. The results for 2-year-ahead forecasts are similar to those for the 1-year-ahead forecasts, suggesting that the Blue Chip and CBO forecasts carried more information than Administration forecasts. However, one difference is that the coefficient on the Administration forecast is negative (significant at the 7 percent level) in the comparison with the CBO forecast. This result indicates that by including information from CBO forecasts, Administration forecast carried "counter-information," in the sense that the sign of the effect of the additional information provided by the Administration forecast varied inversely with the forecasted value. Inspection of the forecast errors showed that the Administration forecast error tended to have the same sign as the CBO forecast error, but the Administration forecast error was larger in absolute magnitude. That result is confirmed by the fact that the Administration forecast had the largest RMSE in Table 2 for 2-year-ahead forecasts.[7]

---

[3]An alternative explanation is that there was an overestimate of the rate of growth of potential GNP.

[4]The positive biases on real GNP growth and the GNP deflator change go a long way to explaining why we have had the Federal budget deficit problems we have had in recent years and over the past decade. Across the various forecasts, the total effect on nominal GNP growth for the 2-year-ahead forecasts is in the 2 to 2 1/4 percentage point range. As an example of the budget effect, current budget sensitivities indicate that a one-time lower nominal GNP growth shock of that magnitude would result in higher cumulative effect on the deficit over 6 years of around $130 billion, with the deficit in the sixth year alone being over $30 billion higher.

[5]Estimation was performed using ordinary least squares with the variance-covariance matrix adjusted for heteroskedasticity of the errors. The two-year-ahead forecasts may introduce a moving average process to the error term in the regression estimates. Only the errors for the regressions for the two-year-ahead GNP deflator change showed evidence of an MA(1) process. Estimation with correction for the moving average process in that case have not yet been performed.

[6]Significance determined at least the 5 percent level.

[7]In the cases that follow, negative coefficients of relatively high significance also appear to result from relatively larger (in absolute magnitude) forecast errors.

TABLE 3--FORECAST COMPARISONS: REAL GNP GROWTH

	CONSTANT	ADMIN	BLUE CHIP	CBO	SE	DW
1-YEAR-AHEAD FORECASTS						
ADMIN - BLUE CHIP	-0.071 (0.121)	-0.141 (0.288)	1.299 (2.657)		1.175	2.06
ADMIN - CBO	-0.126 (0.260)	-0.847 (1.303)		2.007 (2.939)	1.031	1.82
BLUE CHIP - CBO	-0.204 (0.417)		-0.182 (0.165)	1.326 (1.193)	1.101	1.98
2-YEAR-AHEAD FORECASTS						
ADMIN - BLUE CHIP	-1.803 (1.182)	-0.095 (0.169)	1.466 (1.302)		2.057	2.10
ADMIN - CBO	-1.078 (0.901)	-1.214 (2.080)		2.426 (3.054)	1.763	1.75
BLUE CHIP - CBO	-2.515 (1.277)		0.037 (0.021)	1.488 (0.941)	1.907	2.10

Notes: The results in this table are from regressions of the form $Y_t = a + b Y_{1t} + c Y_{2t} + e_t$ as described in the text.
t-statistics in absolute value in parentheses

TABLE 4--FORECAST COMPARISONS: UNEMPLOYMENT RATE

	CONSTANT	ADMIN	BLUE CHIP	CBO	SE	DW
1-YEAR-AHEAD FORECASTS						
ADMIN - BLUE CHIP	0.621 (0.742)	0.797 (1.440)	0.104 (0.198)		0.519	2.23
ADMIN - CBO	0.438 (0.326)	0.338 (0.225)		0.588 (0.360)	0.515	2.20
BLUE CHIP - CBO	0.250 (0.266)		0.099 (0.174)	0.853 (1.324)	0.516	2.17
2-YEAR-AHEAD FORECASTS						
ADMIN - BLUE CHIP	7.250 (2.393)	2.540 (2.915)	-2.543 (2.265)		1.104	1.68
ADMIN - CBO	1.952 (0.819)	-0.279 (0.270)		1.006 (0.912)	1.254	1.51
BLUE CHIP - CBO	5.143 (2.021)		-2.037 (2.060)	2.302 (2.786)	1.071	2.11

Notes: See table 3.

TABLE 5--FORECAST COMPARISONS: GNP DEFLATOR INFLATION

	CONSTANT	ADMIN	BLUE CHIP	CBO	SE	DW
1-YEAR-AHEAD FORECASTS						
ADMIN - BLUE CHIP	-0.519 (1.177)	-0.033 (0.110)	1.017 (2.847)		0.519	1.44
ADMIN - CBO	-0.074 (0.223)	-0.220 (0.792)		1.145 (3.911)	0.525	1.94
BLUE CHIP - CBO	-0.322 (0.645)		0.543 (1.197)	0.414 (1.154)	0.498	1.40
2-YEAR-AHEAD FORECASTS						
ADMIN - BLUE CHIP	0.278 (0.291)	1.708 (2.794)	-0.925 (1.897)		0.947	1.74
ADMIN - CBO	0.076 (0.076)	0.055 (0.066)		0.705 (0.872)	1.016	1.35
BLUE CHIP - CBO	0.770 (1.536)		-1.231 (2.764)	1.867 (4.114)	0.774	1.98

Notes: See table 3.

For the unemployment rate forecast comparisons, the regression results for the 1-year-ahead forecasts indicate that none of the forecasts carried distinct information relative to the other forecasts. For the 2-year-ahead forecasts, no information advantage existed between Administration and CBO forecasts. For comparisons with Blue Chip forecasts, both the Administration and CBO forecasts carried information while the negative signs (of relatively high significance) on the Blue Chip forecast indicates a similar relative information role as was described above for the Administration 2-year-ahead GNP growth forecast (note the higher RMSE for the Blue Chip forecasts in Table 2).

For the GNP deflator percent change forecast comparisons, Blue Chip and CBO forecasts each carried more information relative to Administration forecasts for the 1-year-ahead forecasts. There is very marginal evidence that CBO and Blue Chip forecasts each carried independent information. For the 2-year-ahead forecasts, Administration and CBO forecasts each carried more relevant information than the Blue Chip forecasts, while Administration and CBO forecasts did not carry independent information. The coefficients on the Blue Chip forecasts again had negative coefficients of relatively high significance (again note the higher RMSE for the Blue Chip forecasts in Table 2).

Alternative Explanations

Every effort was made to assure consistency across forecasts. Even so, additional explanations for differences in forecast performance exist, including (1) differences in timing of completion of forecasts, (2) differences in information sets, and (3) differences in policy assumptions on which the forecasts are based. Although the timing and information differences can be important, the difference in policy assumptions is likely the single most important factor for the performance of Administration forecasts relative to alternative forecasts.

The differences in policy assumptions could be considered a difference in information sets with forecasts conditional on the given information set. However, in the case of the Administration forecast, the information set used for the forecast was constrained to include proposed policy changes--whether they were likely to be adopted or not. Similarly, the CBO was constrained to use a current services approach in which policy proposals were not included. Hence, those information sets were not necessarily the information sets that those responsible for preparing Administration and CBO forecasts considered to be most likely.

The incorporation of policy assumptions in forecasts can be described in notation as:

$$F_t = B_t + \pi_t P_t$$

where F_t is the forecast of the variable of interest, B_t is a current services baseline forecast, π_t is the (subjective) probability of adoption of the Administration's proposed policy changes, and P_t is the effect on the forecasted variable attributable to the proposed policy changes. The subjective probability π_t takes on the following values across the different forecasts:

Administration:	$\pi_t = 1$
Blue Chip:	$0 \leq \pi_t \leq 1$
CBO:	$\pi_t = 0$

Let's consider again the results described above. The results indicate that the CBO generally tended (although there were some exceptions) to have forecasts that had a lower bias, a lower RMSE, and that were the most informative (based on the regression approach). One might interpret that result simply as a superior forecasting performance on the part of the CBO. An alternative explanation is that lack of action on the part of Congress, specifically, and the Federal government, generally, to adopt Administration policy proposals made status quo forecasts more reliable. For example, Administration forecasts had the largest bias (positive) and RMSE for real GNP growth in the 2-year-ahead forecasts. Those results should not be surprising *given the incorporation of Administration policy proposals with a probability of one* along with the fact that, over the years, *many of the Administration's proposals were not adopted.* Not only would the forecast be upwardly biased, but also the variation of the forecast would be higher given the inclusion of the additional forecasted component. Note that it is impossible to determine the relative role of the subjective probabilities for the Blue Chip forecasts, because the subjective probabilities are unknown.

Conclusions

The results and analysis presented in this paper indicate that there is no clear-cut answer to the question of which of the alternative public or private sector forecasts had superior performance over the 1980-92 period. A direct interpretation of the results would indicate that, *ex post*, the CBO forecasts of that period were more accurate and carried more information relative to the Administration and Blue Chip forecasts. However, such interpretations are clouded by the fact that Administration forecasts were "policy" forecasts that directly and completely incorporated the forecasted effects of proposed policy changes. The incorporation of the additional forecasted policy component in Administration forecasts (with probability 1) and Blue Chip forecasts (with varying probabilities) distort the statistical measures used for comparison of the forecasts. The fact that Blue Chip consensus forecasts did not outperform the CBO forecasts is notable, suggesting either that CBO forecasts were inherently more accurate and informative, or that the Blue Chip estimates of the probability of policy adoption were inaccurate.

References

Fair, Ray C. and Robert J. Shiller, "Comparing Information in Forecasts from Econometric Models," American Economic Review, June 1990, pp. 375-389.

Budget of the U.S. Government, Office of Management and Budget, 1980-92.

The Economic and Budget Outlook, Congressional Budget Office, 1980-92.

Blue Chip Economic Indicators, Capital Publications, January issues, 1980-92.

Evaluation of Long-Term Forecasts for Selected U.S. Agricultural Commodities

David Stallings and LaWanda Musgrove, World Agricultural Outlook Board, U.S. Department of Agriculture

The U.S. Department of Agriculture (USDA) issues estimates of total expenditures for the following fiscal year in January of each year, as part of the Government-wide budget estimates prepared by the Office of Management and Budget (OMB). The USDA has a wide variety of programs, ranging in responsibility from food stamps to farm programs to forestry. The most volatile spending component of the USDA budget in recent years has been outlays for farm commodity programs. Net outlays in this area over the past 10 years have varied from almost $26 billion in fiscal 1986 down to $6.5 billion in fiscal 1990.[1] Thus, an accurate picture of next year's Department expenditures depends largely on the accuracy of the forecasts of those variables that influence commodity program outlays.

There are three sets of variables that directly influence the amount spent on commodity programs. First are the program parameters themselves: the array of support prices, land setaside, and export programs that may require government checks to be written. Second is the supply and use of U.S. program commodities. The larger the difference between supply and use, other things equal, the greater the level of expenditures required to provide income support. Third is the state of the national economy. High real interest rates, for example, discourage private stockholding. Slow income growth often means a slow change in demand for livestock products, and hence the feedstuffs that are part of that process.

This report focuses on forecasts of the supply and use of major program commodities for the five-year budget "baseline" period. Program parameters are, for the most part, known well in advance of planting and harvesting for the first year. These can therefore almost always be incorporated into the one year out forecast. However, fiscal conditions changed markedly during the period covered by this study, 1970-92, resulting in sharp changes in commodity programs, such as support prices. The five-year baseline forecasts that are discussed here do not make any attempt to predict changes in the laws governing program payments. Only the effects of the current law, given forecast conditions, are considered.

The value of consensus

No one knows the future, even the near future. The simple reason is that all the information that is necessary to "know" the future is unavailable. Thus, people make assumptions about the most important determinants of a variable in question, and then estimate the magnitude of the effects of changes in these determinants on that variable.

There can, *a priori*, be wrong assumptions, incomplete empiricism, or faulty reasoning. Wrong assumptions mean that the wrong determinants are chosen, or that the magnitude of change in those determinants is poorly described.

Forecasts are often made with the help of empirical models, often developed by economists, but which increasingly include non economic information such as weather. Despite the increasing sophistication of such models, no econometric model (even one including weather) can include all variables that may be relevant.

A typical empirical model might express the relationship of wheat production (Q) to expected price (P^e) in the following way:

$$Q = \alpha + \beta P^e$$

A rise in P^e, other things equal, would have the effect of raising Q by an *average* of β. However, we know that not all farmers will react to a known P^e in the same way, and they will not react the same way from year to year. The judgment of the empirical model, Q, will be tempered by other variables that do not enter into the model: expected weather, expected costs, changes in taxes, farm value, export prospects, relative risk-aversion, policy

[1] *Agricultural Outlook*, various issues, Table 34.

changes, and technology. There is no way, for example, to systematically factor in the former Soviet Union as an export market.

The baseline forecast was designed in a way so as to minimize the problem of (1) poor assumptions, (2) empirical limitations, and (3) illogical reasoning. The consensus approach to forecasting is an effort that recognizes where limitations arise, and consensus is an attempt to minimize those limitations. It also creates an institutional process whereby learning may occur.

Consensus forecasts recognize that "models" of all types, formal and informal, abstract from reality. Thus, expertise from a variety of agencies is brought to bear in order to try to incorporate all relevant information. Simplifying, one may describe each agency's role as part of the puzzle:

- The Economic Research Service (ERS) will identify the most important economic effects, and implications for prices, quantity supplied, and quantity demanded. This may include information on such diverse areas as exchange rates, oil prices, the effects of domestic and foreign agricultural policy, and economic growth.

- The Agricultural Stabilization and Cooperatives Service (ASCS) will describe the current policy environment. How many farmers will join commodity programs? Econometric models are also important here, tempered by specialist expertise. There is considerable interaction between ERS and ASCS, and a good bit of crossover in function. The result is more complete economic analysis.

- The Foreign Agricultural Service (FAS) will provide an overview of foreign production and trade, with implications for U.S. exports. Will the Commonwealth of Independent States (CIS) buy grain? Where does the Export Enhancement Program (EEP) stand?

- The Agricultural Marketing Service (AMS) and the Federal Crop Insurance Corporation (FCIC) often provide relevant anecdotal detail.

- The Office of Budget and Policy Analysis assists in interpreting current farm legislation.

- The World Agricultural Outlook Board convenes the forums which issue the supply and use forecasts.

No one of these organizations, obviously, has all the answers to the question of what will happen. The only way that these pieces can be fit together is by cooperation. The question then becomes one of whether or not the cooperation, and consensus, of this very diverse group, produces a result that stands up to serious scrutiny. The question then may be restated by asking whether or not all the information available is used efficiently.

Evaluating forecasts

Whenever forecasts differ from actual outcomes, the use of the term "error" is ubiquitous. A forecast "error" is defined here as the forecast minus the correct answer. The question is really one of why the difference occurs. The forecasting exercise that has used reasonable assumptions, recognized (and accounted for) incomplete empiricism, and eliminated faulty reasoning may not be in "error." The drought in July-August 1988 was not the result of poor empiricism in January of 1988. A projection of drought before it actually occurs, especially one with the severity of that in 1988, cannot be considered "reasonable," given current knowledge. Thus, the term "error" must be used cautiously.

The baseline projections contain many variables. There are, unfortunately, also many indicators of forecast accuracy. Choosing all available indicators is cumbersome, and may, in some circumstances, lead to ambiguous results. A preferred indicator, for evaluating overall performance, would be useful across all variables. There are four criteria that will be used here to select a "single" indicator:

1. The measure must be unit-free, so that we can discuss stocks, prices, and production without changing our frame of reference.

2. An unbiased examination of forecasts minimizes outliers -- these would be unusual events (such as export embargoes and droughts) which may dominate particular years and variables. An outlier may distort an indicator, and thus imply that a forecast in a more normal year is less useful than is actually the case.

3. A useful indicator would be one that can be used over the entire forecasting period, to determine whether or not the estimates are converging or diverging from the "correct" result as the horizon comes closer.

4. Last, we want an indicator that compares the forecast with some alternative, such as a "naive" forecast. This is a very inexpensive way of determining if the baseline consensus process is better than the simplest forecast.

There are several popular measures of forecast evaluation that appear frequently in the literature. The question is whether or not they meet the criteria set out above.[2] Two that meet all the above criteria are the relative absolute error (RAE) and Theil's U2 statistic. The RAE is calculated as:

$$RAE = \frac{\sum_{i-1}^{n} \left| \frac{(F_{it} - A_{it})}{(A_{i(t-1)} - A_{it})} \right|}{n}$$

where:

F_{it} = the current projection or estimate,
A_{it} = the actual value, and
$A_{i(t-1)}$ = the actual value in the period immediately before the forecast, as it is known at the time of the forecast.

An RAE of zero means that the projection or estimate was exactly correct. A value of one tells the forecast user that the "naive" forecast is as close to the actual value as is the projection or estimate. A value greater than one suggests that the "naive" forecast is better. Any value between zero and one indicates that the forecast adds enough information to improve on the naive model. We should also expect that the RAE will approach zero as we come closer to the forecast horizon. The RAE suffers, however, if there is no change from the previous year to the current year (as often occurs with target prices, for example). Such a circumstance leads the value to become infinite, which seriously distorts the average.

An indicator that is equivalent to the RAE is Theil's inequality statistic (also known as Theil's U2):

$$U2 = \frac{\left(\frac{\sum_{i-1}^{n} (F_{it} - A_{it})}{n} \right)^{1/2}}{\left(\frac{\sum_{i-1}^{n} (A_{i(t-1)} - A_{it})}{n} \right)^{1/2}} = \frac{RMSE}{RMSE_{rw}}$$

The numerator is the same as the root mean squared error (RMSE) and the denominator is the equivalent of an RMSE, using a "naive" (or random walk, subscript *rw*) for the forecast. The interpretation is the same as the RAE. If the U2 is greater than one, then the "naive" (or random walk) forecast is superior. Values less than one tell the user that the current forecast is superior to a random walk. The influence of an outlier is minimized, as with the RAE; it is contained in both the numerator and denominator.

The efficiency of the forecasts may be tested by estimating the following via regression:

[2]An excellent review of single indicators of forecast error is contained in J. Scott Armstrong and Fred Collopy "Error measures for generalizing about forecasting methods: empirical comparisons," *International Journal of Forecasting* 8 (1992) pp. 69-80.

$$A_t = \alpha + \beta F_t + u_t,$$

where A_t is the actual outcome, F_t is the forecast of that outcome, and u_t is a random error. An efficient forecast would have the estimator of α equal zero, and that for β equaling one. Deviations from these "perfect" values are evidence of inefficiency; better forecasts would result from knowledge of these parameters' values. Because both α and β must equal their "ideal" for efficiency to be established, the test must jointly restrict the two coefficients (Holden and Peel).[3] One may either use a Wald test or an F-test to test the null hypothesis of $\alpha=0$ and $\beta=1$. The Wald test is a Chi-square distribution and will be the one discussed below.

However, Holden and Peel also show that even if α and β differ from their "ideal," this does not necessarily indicate that a projection is biased. Instead, regressing the forecast error on a constant term provides the necessary and sufficient condition for determining bias, with the t-statistic telling the result. This test is equivalent to restricting β to one. Finding bias means that, even if α and β do not statistically differ from their respective "ideals," that the results can be called into question. However, if both efficiency and unbiasedness can be demonstrated, then the results strongly suggest a good forecasting record; information is being used economically.

The results

Forecasts are issued in January of each year. The current crop year has, at that time, not been completed. Thus, many components of supply and use are still estimated, such as production, and some are still subject to significant revision, such as market prices, exports, and ending stocks. This current crop year is labeled "Current year" in the results detailed in Tables 1-4. The first year out has crop years that begin June 1 for wheat and barley; August 1 for rice, cotton, and peanuts; September 1 for corn, sorghum, and soybeans; and December 16 for honey. Forecasts issued in January of 1992 then had the current crop year as 1991/92; the year ahead forecast as 1992/93 (beginning June 1, 1992, for wheat and barley, for example); the two years ahead forecast would be for the 1993/94 crop year (June 1, 1993 for wheat and barley); and the three years ahead forecasts would be for the 1994/95 crop year (June 1, 1994 for wheat and barley).

This study covers the January forecasts from 1970 through 1992. Long-term (more than one year out) forecasts were first issued in 1975. There are then 23 observations for current year projections, 22 for one-year projections, 16 for two-year projections, and 15 for three-year projections. The "actual" value is assumed to be that reported as the immediate past year for each of the forecasts. Thus the "actual" 1991/92 outcome is assumed to be that reported in the budget baseline document of January 1993.[4]

Supply and use categories, and prices, were chosen to maintain as much consistency as possible between commodities, for comparison and easy summarization. Area planted and harvested, obviously, however, does not apply to honey; the appropriate variable is number of colonies. The announced loan rate is the applicable loan rate for each commodity.[5] There are no explicit market price forecasts made for cotton (forbidden by law), and no target prices for soybeans, honey, or peanuts. Imports are significant components of supply and use for only honey and peanuts, of the crops covered herein. Commodity Credit Corporation (CCC) stocks are those bought by the USDA, voluntarily or not, for purposes of price stabilization or export programs such as PL-480.[6]

---

[3] *The Manchester School of Economic and Social Studies* 58 (June 1990), pp. 120-7.

[4] The 1991/92 values would have been issued in the November 1992 issue of *World Agricultural Supply and Demand Estimates* for major program commodities. Thus, the estimates are two months old at the time the budget baseline is completed. This is the case for all the "actual" observations as well as the estimates.

[5] The quota loan rate for peanuts, for example.

[6] A basic discussion of price stabilization and stock functions appears in *The Basic Mechanisms of U.S. Farm Policy Parts One, Two, and Three*, from the U.S. Department of Agriculture, Economic Research Service, Miscellaneous Publications numbers 1470 (May 1989), 1476 (December 1989), and 1477 (December 1989). A discussion of USDA storage programs appears in *Storage Subsidy Programs* from the U.S. Department of Agriculture, Economic Research Service, Staff Report AGES 9075, December 1990, pp. 4-5. Export programs are described in *Agricultural Export Programs: Background for 1990 Farm Legislation*, by Karen Z. Ackerman and Mark E. Smith, from the U.S. Department of Agriculture, Economic Research Service, Staff Report AGES 9033, May 1990.

Ending stocks are the difference between total supply (beginning stocks, imports, and production) and total use (domestic use plus exports). Errors for ending stocks are therefore not independent of errors in supply and use categories. Forecast errors from ending stocks will be "carried over," so to speak, into total supply for the subsequent years. Other things equal, then, forecast errors will automatically increase for supply and use variables as the forecast horizon recedes. Furthermore, changes in stocks are also strongly associated with changes in market prices. Thus, errors in stock level forecasts will produce errors in price forecasts. No attempt is made to adjust for these cumulative processes.

Table 1 summarizes a basic indicator of forecast accuracy, absolute percentage error. This indicator should demonstrate progressively higher errors as the forecast horizon recedes. The largest percentage errors, other than for CCC stocks, are for exports and ending stocks. Exports have varied widely in the 1980's because of significant changes in exchange rates, foreign economic growth rates, and U.S. export programs. All are difficult to forecast. The large CCC stock forecast percentage errors occur because of great volatility in relatively small holdings. Government-held stocks serve a variety of public policy purposes (export promotion, food security, price stability), not all of which are (or can be) well-identified more than a year in the future. Percentage errors in almost all categories are systematically higher as the forecast horizon recedes.

Low percentage errors for ending stocks in the current year for barley and wheat correspond to low percentage errors in market price forecasts for those two grains. Higher percentage errors in ending stocks for corn, sorghum, rice, and soybeans are also associated with higher average percentage errors in prices. These lower errors for wheat and barley, of course, could be the result of more of the crop year being completed at the time that the forecast is made.[7] The relative size of these errors, and their correspondence, persist until the second out year.

Table 2 shows Theil's inequality statistic (U2) for the same set of variables. The random walk assumption has it here that the previous year's value is maintained throughout the forecast horizon. For example, 1990/91 is the random walk assumption for forecasts issued in January 1992 for the "current" year 1991/92 and subsequent "out" years. The values of U2 show that the baseline forecasts add value for the current and year-ahead projections, with the exceptions of bee colonies (area harvested for honey) and CCC stocks. Most of the error in bee colonies comes from the fact that no statistics were collected between 1982 and 1986. In 1986, the number of colonies was determined to have dropped by 25 percent (from 4.3 million to 3.2 million). Current data now reflect only beekeepers with at least five colonies.[8]

Market price forecasts for those commodities with the greatest budget exposure (all but soybeans; cotton still being excluded) are better than "naive" forecasts for the current year and first year out, according to the U2 statistic. Other things equal, these market price forecasts mean that the supply and use estimates for the following fiscal year are better than a "naive" forecast. Two years out, the only observations of prices that suggest using a naive forecast are soybeans and honey. The two and three year estimates for exports, in general, have the least acceptable U2 values, other than those for CCC stocks.

Efficiency statistics are summarized in Table 3. The null hypothesis being tested is the Wald test of $\alpha=0$ and $\beta=1$. The designation "not efficient" means that the null hypothesis could not be accepted even at the 1 percent level. A "10" is the strongest acceptance of efficiency. The "year ahead" forecasts begin the systematic process by which exports and ending stocks are the least efficient forecasts, in general. These lead to price forecasts also being largely inefficient. By the time that the third year ahead is reached, only yield projections are mostly efficient.

Bias statistics appear in Table 4. These results, are, generally, better than those for efficiency. There is almost uniform unbiasedness to the forecasts. The notable exception is a downward bias for area for peanuts. Peanuts operate under a supply control program based on poundage. Thus, there is no reason to economize on land resources; less productive land may be brought into use rather than applying expensive soil enrichment resources.

[7]See fn. 4.

[8]See *Honey: Background for 1990 Farm Legislation*, by Frederic L. Hoff and Jane K. Phillips, from the U.S. Department of Agriculture, Economic Research Service, Staff Report AGES 89-43, September 1989.

Table 1. Absolute percentage errors for program commodity supply and use, forecasts issued 1970-1992.

	Corn	Sorghum	Barley	Wheat	Rice	Soybeans	Upland cotton	Honey	Peanuts
Current year									
Area planted	0.3	1.4	0.4	0.3	1.3	0.6	2.4	--	1
Area harvested	0.9	2.2	0.8	0.4	1.4	0.7	1.8	4.9	0
Yield	0.8	2.5	1.3	0.3	0.6	1.5	1.8	6.8	1
Production	1.3	2.5	1.5	0.4	1.6	2.0	1.9	5.0	1
Domestic use	3.5	10.5	5.1	5.6	5.8	7.5	3.9	4.3	7
Exports	12.2	20.1	24.6	5.0	5.8	8.3	11.3	9.0	18
Ending stocks	24.5	34.9	11.1	11.2	25.1	29.9	15.0	13.6	23
CCC inventory	86.1	84.0	122.6	57.8	96.7	234.6	245.2	29.8	--
Announced loan		0.2	5.0	4.9	4.8	1.2	0.6	0.3	0.10
Market price	7.6	7.1	2.9	2.8	5.7	6.8	--	6.9	4
Target price	0.0	0.0	0.0	0.0	0.0	--	0.1		
Imports								4.4	35
Year ahead									
Area planted	3.5	6.8	10.8	3.7	8.7	4.1	6.7	--	5
Area harvested	3.7	7.0	11.8	4.3	8.1	4.4	7.7	7.1	5
Yield	11.6	12.2	7.8	7.5	3.7	7.2	9.8	15.6	12
Production	14.2	15.3	15.2	8.2	7.7	9.7	14.0	12.3	11
Domestic use	8.1	16.9	9.1	7.7	7.7	10.7	10.3	10.6	11
Exports	26.1	21.7	64.2	16.7	13.1	14.8	26.8	35.2	34
Ending stocks	59.7	63.1	30.9	35.2	49.7	60.3	55.2	33.8	20
CCC inventory	693.5	353.8	421.1	117.1	888.1	928.0	1,485.3	77.2	--
Announced loan		3.2	13.5	6.3	8.2	6.0	3.5	0.8	2.92
Market price	19.8	18.5	16.0	15.1	21.8	24.4	--	13.4	5
Target price	1.8	3.2	8.3	2.2	1.4	--	1.5		
Imports								39.7	55
Two years ahead									
Area planted	8.4	14.7	13.4	10.3	18.6	8.8	16.5	--	10
Area harvested	8.7	14.8	14.8	12.6	18.8	9.0	17.0	9.8	10
Yield	10.7	12.0	8.5	7.9	5.0	7.3	9.8	17.5	15
Production	15.6	19.3	18.7	16.7	15.0	13.2	18.3	14.3	15
Domestic use	7.0	24.3	11.7	14.3	10.1	7.6	13.1	13.4	11
Exports	32.9	18.3	43.4	20.9	22.1	21.8	27.8	32.7	43
Ending stocks	66.7	73.3	43.6	42.2	51.6	39.8	51.0	37.2	24
CCC inventory	205.4	192.4	224.0	138.7	2,802.4	155.2	196.5	98.9	--
Announced loan		13.5	13.5	18.3	14.4	12.8	11.3	4.4	6.35
Market price	20.0	22.7	17.9	22.0	33.2	19.9	--	12.3	9
Target price	5.5	7.7	14.2	5.8	3.5	--	4.9		
Imports								39.1	74
Three years ahead									
Area planted	8.8	15.4	14.1	12.7	23.8	10.6	20.4	--	10
Area harvested	9.0	14.6	16.0	17.9	23.9	10.8	20.9	13.6	10
Yield	11.0	11.2	10.9	9.2	6.7	7.5	10.5	18.8	18
Production	18.4	18.8	19.9	19.3	20.0	14.6	19.1	16.2	16
Domestic use	6.3	23.1	11.4	21.0	10.3	7.3	16.8	12.1	13
Exports	38.2	19.8	68.5	22.1	31.3	28.0	33.9	44.7	45
Ending stocks	65.9	80.5	63.9	54.7	57.4	38.3	34.2	40.2	24
CCC inventory	240.3	229.9	206.6	203.9	3,406.6	243.5	190.9	130.8	--
Announced loan		18.3	18.4	23.5	21.2	18.6	16.4	7.8	11.09
Market price	22.9	24.6	24.6	25.0	40.7	22.3	--	15.5	13
Target price	8.4	12.0	20.2	9.5	6.8	--	5.5		
Imports								50.0	79

-- No information available or not relevant.

Table 2. Theil's inequality statistic for program commodity supply and use, forecasts issued 1970-1992.

	Corn	Sorghum	Barley	Wheat	Rice	Soybeans	Upland cotton	Honey	Peanuts	Exceeding one?
Current year										
Area planted	0.05	0.15	0.07	0.04	0.11	0.10	0.18	--	0.29	0 of 8
Area harvested	0.11	0.18	0.09	0.04	0.13	0.13	0.13	1.44	0.20	1 of 9
Yield	0.06	0.24	0.19	0.06	0.16	0.17	0.16	0.53	0.17	0 of 9
Production	0.07	0.15	0.13	0.05	0.14	0.15	0.10	0.37	0.15	0 of 9
Domestic use	0.52	0.61	0.78	0.53	0.88	0.91	0.53	0.47	0.57	0 of 9
Exports	0.76	1.04	0.47	0.32	0.46	0.54	0.41	0.22	0.73	1 of 9
Ending stocks	0.39	0.45	0.43	0.38	0.53	0.79	0.35	0.52	0.87	0 of 9
CCC inventory	0.91	0.89	0.87	0.96	1.04	0.53	1.08	0.91	--	2 of 8
Announced loan	0.08	0.65	0.79	0.55	0.26	0.19	0.11	0.01	0.21	0 of 9
Market price	0.58	0.60	0.19	0.15	0.28	0.48	--	0.89	0.59	0 of 8
Target price	0.00	0.00	0.00	0.00	0.00	--	0.05			0 of 6
Imports								0.17	0.32	0 of 2
Year ahead										
Area planted	0.34	0.46	0.78	0.31	0.59	0.54	0.40	--	0.97	0 of 8
Area harvested	0.32	0.46	0.77	0.34	0.54	0.57	0.43	1.11	1.00	2 of 9
Yield	0.76	0.84	0.79	1.00	0.68	0.72	0.81	0.72	0.93	0 of 9
Production	0.58	0.63	0.80	0.57	0.51	0.71	0.52	0.65	0.81	0 of 9
Domestic use	0.83	0.86	0.83	0.74	0.82	0.93	1.00	0.91	0.93	0 of 9
Exports	1.02	0.89	0.93	0.73	0.92	0.84	0.74	0.88	0.99	1 of 9
Ending stocks	0.72	0.65	0.72	0.68	0.71	0.90	0.85	0.85	0.65	0 of 9
CCC inventory	1.19	0.90	1.09	1.00	0.96	0.98	1.35	0.95	--	3 of 8
Announced loan	0.38	0.73	0.77	0.59	0.63	0.36	0.16	0.19	0.41	0 of 9
Market price	0.81	0.80	0.73	0.67	0.84	1.25	--	0.74	0.66	1 of 8
Target price	0.20	0.51	1.21	0.27	0.34	--	0.36			1 of 6
Imports								0.74	0.48	0 of 2
Two years ahead										
Area planted	0.84	0.80	0.74	0.74	0.96	0.86	0.87	--	1.47	1 of 8
Area harvested	0.82	0.74	0.75	0.79	0.97	0.89	0.94	1.22	1.45	2 of 9
Yield	0.79	0.84	0.79	1.13	0.83	0.58	0.96	0.75	1.26	2 of 9
Production	0.75	0.71	0.80	0.91	0.86	0.79	0.95	0.81	1.10	1 of 9
Domestic use	0.65	0.99	0.82	0.92	0.65	0.70	0.98	1.28	0.93	1 of 9
Exports	1.19	0.81	1.00	0.83	1.17	0.93	0.78	0.70	1.05	4 of 9
Ending stocks	0.96	0.66	1.04	0.81	1.01	0.80	1.07	0.78	0.83	3 of 9
CCC inventory	1.50	0.80	1.24	1.23	1.45	0.85	0.62	1.12	--	5 of 8
Announced loan	0.67	0.67	0.84	0.66	1.05	0.69	0.31	0.39	0.63	1 of 9
Market price	0.87	0.89	0.97	0.89	0.96	1.18	--	1.23	0.87	2 of 8
Target price	0.36	0.67	1.07	0.43	0.51	--	0.46			1 of 6
Imports								0.86	0.44	0 of 2
Three years ahead										
Area planted	1.16	0.84	0.71	1.04	1.16	0.92	0.89	--	1.33	4 of 8
Area harvested	1.06	0.71	0.68	1.09	1.19	0.95	0.93	1.26	1.28	5 of 9
Yield	0.76	0.74	0.97	1.04	0.88	0.75	0.76	0.87	1.13	2 of 9
Production	0.87	0.66	0.75	0.99	1.02	0.90	0.83	1.16	0.88	2 of 9
Domestic use	0.59	0.82	0.75	0.96	0.54	0.67	1.01	1.30	1.16	3 of 9
Exports	1.29	1.09	0.97	0.88	1.32	1.21	0.78	0.90	1.14	5 of 9
Ending stocks	1.01	0.65	1.30	0.96	0.95	0.82	0.75	0.84	0.77	2 of 9
CCC inventory	1.95	0.79	1.24	1.50	1.27	0.93	0.61	1.13	--	5 of 8
Announced loan	0.70	0.70	0.88	0.70	1.17	0.80	0.47	0.49	0.81	1 of 9
Market price	0.94	0.99	1.28	1.13	1.19	1.68	--	1.34	1.12	6 of 8
Target price	0.50	0.84	0.94	0.52	0.63	--	0.34			0 of 6
Imports								1.00	0.49	1 of 2

--Not available or not relevant.

Table 3. Forecast efficiency for program commodity supply and use, forecasts issued 1970-1992.[1]

	Corn	Sorghum	Barley	Wheat	Rice	Soybeans	Upland cotton	Honey	Peanuts
Current year									
Area planted	10	10	10	10	10	5	10	10	10
Area harvested	NE	1	10	10	10	10	10	NE	1
Yield	10	10	10	10	10	10	10	1	10
Production	1	10	5	NE	10	10	10	10	10
Domestic use	10	10	10	5	10	NE	10	1	5
Exports	1	NE	10	10	10	10	10	5	10
Ending stocks	10	10	10	5	10	NE	10	10	N
CCC inventory	NE	1	1	NE	NE	NE	NE	10	--
Announced loan	NE	10	10	10	10	10	10	10	10
Market price	NE	NE	10	10	10	NE	--	NE	10
Target price	10	10	10	10	10	--	10		
Imports								10	N
Year ahead									
Area planted	10	10	10	10	5	10	10	10	10
Area harvested	10	10	10	10	10	10	10	NE	10
Yield	10	5	10	1	10	10	10	10	N
Production	10	10	10	10	10	10	10	NE	1
Domestic use	10	NE	10	5	10	NE	10	NE	N
Exports	NE	NE	NE	1	NE	5	10	NE	N
Ending stocks	10	10	NE	NE	10	NE	NE	10	10
CCC inventory	NE	NE	NE	NE	NE	NE	NE	1	--
Announced loan	NE	10	5	10	5	10	10	NE	10
Market price	NE	NE	NE	1	NE	NE	--	NE	10
Target price	10	5	NE	5	10	--	10		
Imports								5	10
Two years ahead									
Area planted	1	1	10	NE	NE	NE	5	10	N
Area harvested	1	10	10	1	NE	NE	5	NE	N
Yield	10	1	5	NE	10	10	10	10	N
Production	5	10	5	NE	NE	1	10	1	1
Domestic use	10	NE	10	10	10	10	10	NE	N
Exports	NE	1	1	1	NE	NE	5	NE	N
Ending stocks	NE	10	NE	NE	NE	5	1	10	5
CCC inventory	NE	NE	NE	NE	NE	NE	10	NE	--
Announced loan	NE	NE	NE	NE	NE	NE	NE	NE	1
Market price	NE	NE	NE	NE	NE	NE	--	NE	10
Target price	N	NE	NE	NE	1	--	5		
Imports								NE	10
Three years ahead									
Area planted	NE	NE	10	NE	NE	NE	NE	10	N
Area harvested	NE	10	10	NE	NE	NE	NE	NE	N
Yield	5	5	NE	NE	10	10	5	10	N
Production	1	10	5	NE	NE	NE	1	NE	1
Domestic use	10	NE	5	10	10	NE	10	NE	N
Exports	N	NE	NE	NE	NE	NE	NE	NE	N
Ending stocks	NE	10	NE	NE	NE	1	10	1	1
CCC inventory	NE	1	NE	NE	NE	1	10	NE	--
Announced loan	NE	NE	NE	NE	NE	NE	NE	NE	N
Market price	NE	NE	NE	NE	NE	NE	--	NE	1
Target price	NE	NE	NE	NE	NE	--	NE		
Imports								NE	10

NE = "Not efficient"

--Not available or not relevant.

[1]The test summarized is the Wald test, a joint test of the slope equaling one and the intercept zero. A "10" means this null hypothesis cannot be rejected even at the 10 percent level. "Not efficient" means that the null hypothesis is rejected. The 10 percent level is the most stringent criteria for acceptance of the null hypothesis.

Table 4. Forecast bias for program commodity supply and use, forecasts issued 1970-1992.[1]

	Corn	Sorghum	Barley	Wheat	Rice	Soybeans	Upland cotton	Honey	Peanuts
Current year									
Area planted	NB	-10	NB	NB	NB	10	NB	--	N
Area harvested	-1	-5	NB	NB	NB	NB	NB	10	N
Yield	NB	NB	-10	NB	NB	NB	-10	-1	N
Production	-1	NB	NB	-10	NB	NB	NB	NB	N
Domestic use	NB	NB	NB	NB	NB	NB	NB	NB	1
Exports	NB	-5	NB	NB	NB	NB	-10	NB	N
Ending stocks	NB	NB	NB	-10	NB	NB	10	NB	N
CCC inventory	NB	NB	NB	NB	NB	NB	NB	NB	--
Announced loan	NB	NB	NB	NB	NB	NB	NB	NB	N
Market price	NB	NB	NB	NB	NB	NB	--	NB	N
Target price	NB	NB	NB	NB	NB	--	NB		
Imports								NB	N
Year ahead									
Area planted	NB	NB	NB	NB	NB	NB	NB	--	-1
Area harvested	NB	NB	NB	NB	NB	NB	NB	5	-1
Yield	NB	NB	NB	NB	NB	NB	NB	NB	5
Production	NB	NB	NB	NB	NB	NB	NB	5	N
Domestic use	NB	10	NB	NB	NB	NB	NB	10	N
Exports	NB	-10	NB	NB	NB	NB	NB	NB	N
Ending stocks	NB	NB	NB	NB	NB	NB	NB	NB	N
CCC inventory	NB	NB	5	NB	NB	NB	10	NB	--
Announced loan	-10	NB	-5	NB	NB	NB	NB	-5	N
Market price	NB	NB	-10	NB	NB	NB	--	NB	-1
Target price	NB	NB	NB	NB	NB	--	NB		
Imports								NB	N
Two years ahead									
Area planted	NB	NB	NB	NB	NB	NB	NB	--	-5
Area harvested	NB	NB	NB	NB	NB	NB	NB	5	-5
Yield	NB	NB	NB	NB	NB	NB	-10	NB	1
Production	NB	NB	NB	NB	NB	NB	NB	5	N
Domestic use	NB	NB	NB	-10	NB	NB	NB	NB	N
Exports	NB	NB	-10	NB	NB	NB	NB	NB	N
Ending stocks	NB	NB	NB	NB	NB	NB	NB	NB	N
CCC inventory	NB	NB	NB	NB	NB	NB	NB	NB	--
Announced loan	NB	NB	-10	NB	1	NB	NB	-10	N
Market price	NB	NB	NB	NB	NB	NB	--	NB	N
Target price	NB	NB	NB	NB	10	--	NB		
Imports								NB	N
Three years ahead									
Area planted	NB	1	NB	NB	NB	NB	NB	--	-5
Area harvested	NB	10	NB	NB	NB	NB	NB	5	-1
Yield	NB	NB	NB	NB	NB	NB	-5	NB	1
Production	NB	NB	NB	NB	NB	NB	NB	10	N
Domestic use	NB	NB	NB	-10	NB	NB	NB	NB	N
Exports	NB	NB	NB	NB	NB	NB	NB	NB	N
Ending stocks	NB	NB	NB	NB	NB	NB	NB	NB	N
CCC inventory	NB	NB	NB	10	NB	NB	NB	NB	--
Announced loan	NB	NB	NB	NB	1	NB	NB	-10	N
Market price	NB	NB	NB	NB	NB	NB	--	NB	N
Target price	NB	NB	NB	NB	10	--	NB		
Imports								NB	N

NB = "Not Biased"

-- Not available or not relevant.

[1]The test summarized is based on a regression of the forecast error (forecast minus actual) on a constant. The t-test is used to determine whether or not the constant (average error) is significantly different from zero. A value of minus 10 in the table means that the average is significantly different from zero at the 10 percent level and that the forecast is biased downward. A positive 10 means that the forecast is biased upwards, and differs from zero at the 10 percent level.

The role of forecasts in policy and the role of policy in forecast error

The January baseline forecasts are, as mentioned earlier, based on current law; no attempt is made to forecast policy changes outside current legislation. However, these forecasts can serve as warnings to policy makers that changes are needed. One example, leading to the payment-in-kind (PIK) program, occurred in 1983.

Supply and use forecasts for corn in January of 1983 implied (Table 5) that the huge stock buildup in 1982/83, resulting from large production and carryover, would send market prices below the loan rate for most of 1982/83 and at least part of 1983/84. This would mean that the government would be forced to accept grain at the loan rate, make high program payments (the difference between the target price and loan rate per bushel), and carry large grain stocks into the near future. The purchase of an additional 1.1 billion bushels of corn at the loan rate would cost $2.6 billion dollars alone. Significant budgetary pressure demanded that something be done, both for the current year and the year ahead.

The decision was made to "pay" farmers not to produce in 1983/84 by giving them grain carried over from 1982/83.[9] The result would be to effectively lower current stock levels, raising current prices above the loan rate. At the same time, lower production in 1983/84 would reduce the supply pressure in that crop year.

The PIK program helped to reduce production and raise prices in both 1982/83 and 1983/84 (partially assisted by a drought in much of the Corn Belt).[10] One additional effect of PIK was to significantly increase the forecast error for 1983/84, as a result of the abrupt policy shift. The percentage error on ending stocks for 1983/84, for example, is almost 300 percent. For CCC stocks, the error is over 1000 percent. Whatever one believes about the efficacy of the PIK program, the forecast "error" introduced by the change in policy clearly affected the appearance of accuracy in Table 1.[11] The question is whether or not good-faith forecasts affect policy decisions. Clearly, they can. One could argue that it is, in fact, one of their functions.

Conclusions

The USDA prepares, for internal use, long-term forecasts of program commodity supply and use each January, as part of a government-wide effort to estimate obligations for the next fiscal year. These supply and use forecasts are the result of consensus by several different agencies. Absolute percentage errors tend to increase as the forecast horizon recedes. This is partly due to the propagation of past errors, assuming that ending stocks are a residual term after the current year. Errors are particularly large for exports, ending stocks, and government-held (CCC) stocks.

Theil's U2 statistic indicates that, up until exports and ending stocks in the second year out, USDA forecasts add value to what is known about the past. The forecasts are useful. However, the lack of efficiency, especially starting with exports and propagated from there to ending stocks in the first year ahead, indicates that crucial information is not being included. Despite being inefficient, there is very little sign of bias in the forecasts. This may imply that there is no systematic "fix" to the problem of inefficiency.

The PIK program was used as an example of how projections may have been used to revise existing policies or implement new ones. Thus, an accurate *ex ante* forecast can be put to use to offset "problems" before they become severe. However, a forecast that leads to such use implies, *ex post,* that the forecaster is very poor at the job!

---

[9]A description of the program choices faced by farmers subsequent to the announcement of PIK, but before its actual implementation, may be found in *Analyzing the 1983 Payment in Kind Program at the Farm Level* by Fred J. Benson and Paul Hasbagen, University of Minnesota, Department of Agricultural and Applied Economics, Staff Paper series number P83-7, January, 1983.

[10]One could argue, from Table 5, that the effect of PIK was to reduce area planted from 70 to 60 million acres, while the drought reduced yields by almost 30 percent. Thus, the drought effectively turned a three-year program into a one-year program. We are grateful to Paul Westcott of ERS for this observation.

[11]See "Agricultural Policy," by Richard Rizzi, in *Agricultural Outlook* (January-February 1983), pp. 20-1, for a further example of how projections for the 1982/83 and 1983/84 crop year were affected.

Table 5. Supply and use forecasts for corn, issued in January 1983, and actual outcome

	History[1]		Forecast			Actual		
	1980/81	1981/82	1982/83	1983/84	1984/85	1982/83[2]	1983/84[3]	1984/85[4]
Million acres								
Area planted	84.0	84.2	81.9	70.0	74.0	81.9	60.2	80.4
Area harvested (grain)	73.0	74.6	72.8	60.0	64.0	73.2	51.4	71.8
Bushels per acre								
Yield	91.0	109.9	114.4	114.6	114.8	114.8	81.0	106.7
Million bushels								
Beginning stocks	1,617	1,034	2,366	3,497	2,884	2,286	3,120	723
Production	6,645	8,201	8,330	6,876	7,347	8,397	4,166	7,656
Imports	1	1	1	1	1	1	2	3
Total supply	8,263	9,236	10,697	10,374	10,232	10,684	7,288	8,382
Domestic use	4,874	4,903	5,100	5,265	5,450	5,674	4,700	5,165
Exports	2,355	1,967	2,100	2,225	2,300	1,870	1,866	1,838
Total use	7,229	6,870	7,200	7,490	7,750	7,544	6,566	7,003
Ending stocks	1,034	2,366	3,497	2,884	2,482	3,140	722	1,379
CCC stocks	510	1,673	2,723	2,280	1,735	1,150	201	240
Dollars per bushel								
Loan rate	2.25	2.40	2.55	2.65	2.55	2.55	2.65	2.55
Market price	3.11	2.45	2.30	2.65	2.65	2.70	3.20	2.65
Target price	2.35	2.40	2.70	2.86	2.86	2.70	2.86	3.03

[1]As of January 1983. [2]As known in January 1984. [3]As known in January 1985. [4]As known in January 1986.

Evaluating the Accuracy of the Short-Term Hydroelectric Generation Model Forecast

Robin D. Reichenbach, Energy Information Administration

The Short-Term Hydroelectric Generation Model (STHGM) was developed by the Energy Information Administration (EIA) to provide up to a 27-month forecast of utility net hydroelectric generation for publication in the quarterly Short-Term Energy Outlook (STEO) reports.

Utility net generation is gross generation minus plant use. Utility net hydroelectric generation includes generation from conventional hydroelectric facilities as well as pumped storage plants. Conventional facilities include storage, run-of-river, and diversion facilities. Storage facilities feature reservoirs created by dams and typically have additional functions such as flood control and recreation. Run-of-river facilities use natural stream flow along with a small dam. Some run-of-river facilities also impound water behind the dam to store enough energy for use during peak electric demand hours. Diversion facilities involve a man-made channel or aqueduct with sufficient slope for the water flow to drive a turbine. Pumped storage facilities repeatedly recycle water by pumping water discharged from the turbines to a lower retaining pool back into an upper storage facility for peak power production. Pumped storage facilities have a negative net generation, since the electricity consumed to pump the water exceeds the amount produced.

Hydroelectric generation provided 9 percent of total U.S. generation in 1992, and historically has provided as much as 14 percent. In 1992, approximately 3,500 hydroelectric facilities existed in the United States, including 144 pumped storage facilities.

Average national net hydroelectric generation for 1970 to 1992 was 275 billion kilowatthours; however, annual hydroelectric generation fluctuated a great deal (Figure 1). Between 1970 and 1992, generation ranged from a low of 220 billion kilowatthours in 1977 to a high of 332 billion kilowatthours in 1983. This variation was caused by several factors, including the dependency of hydroelectric generation on precipitation, the relative share of conventional versus pumped storage capacity, and regulations governing reservoir levels for alternative water uses.

Seasonality appeared in the monthly data for 1990 and 1991 (Figure 2) and was also apparent for other historical years. Seasonality appeared in 1992 as well, although it was less obvious. In all three of these years, generation decreased in April, peaked in mid-summer, and then bottomed out in late summer. Many explanations can be found for this seasonality. One reason is the late summer decrease in generation due in part to storage facilities maintaining higher water levels, particularly in recreational areas.

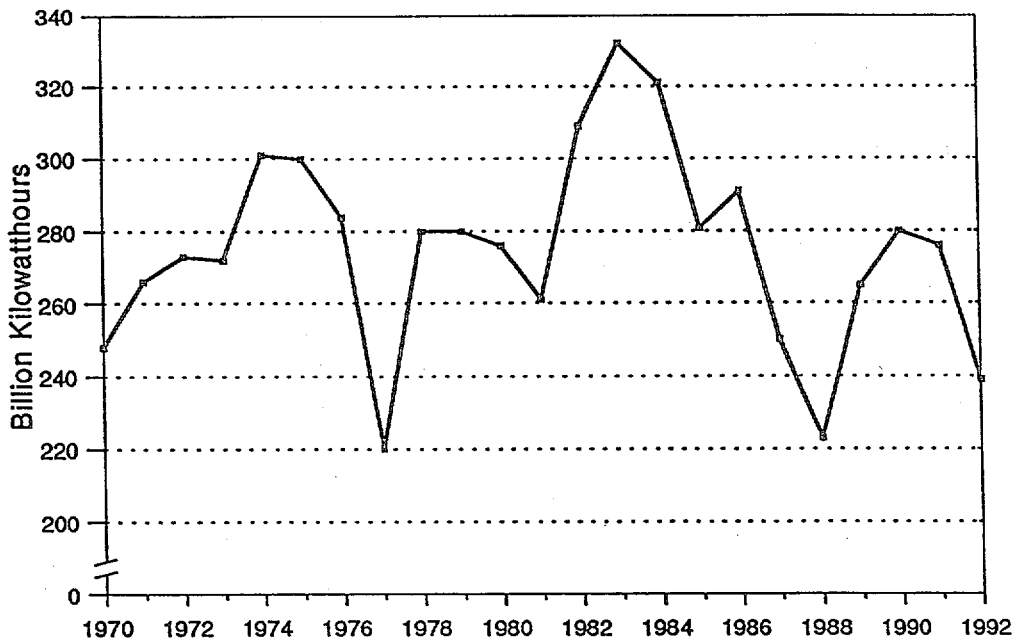
The hydroelectric generation data were also analyzed for trends in the data. From 1970 to 1992, conventional hydroelectric generating capacity increased by 18 gigawatts, from 53 gigawatts to 71 gigawatts. During the same time period, pumped storage hydroelectric generating capacity increased by 15 gigawatts, from 4 gigawatts to 18 gigawatts. However, even with these additions to capacity, generation from 1970 to 1992 remained stationary. No trend appeared in the generation data and, therefore, was not included in the model. It is believed that the increase in conventional generation was offset by the increase in net negative generation from the pumped storage facilities.

Census Division concentrations of hydroelectric generating stations are quite varied. Hydroelectric resources are not evenly distributed across the Nation, but are generally concentrated where precipitation and mountains combine to provide large water volumes. The Nation's most concentrated hydropower regions are the Pacific Division with 50 percent of total national hydroelectric generation in 1992, the Middle Atlantic Division with 11 percent, and the Mountain Division with 11 percent, although some major facilities are located in other areas (Figure 3). These three areas generated 172 billion kilowatthours of hydroelectric power in 1992 or 72 percent of the Nation's 239 billion kilowatthours of net hydroelectric generation. The most concentrated States (those generating more than 15 billion kilowatthours in 1992) were California, New York, Oregon, and Washington. All but New York are located in the Pacific Division.

With that background information, I will now describe the Short-Term Hydroelectric Generation Model. The Short-Term Hydroelectric Generation Model forecasts national net hydroelectric generation by utilities using an autoregressive integrated moving average (ARIMA) time series model with deterministic seasonality and precipitation as an explanatory variable, otherwise known as a transfer function model.[1] The model is based on 23 years of monthly historical data from 1970 to 1992.

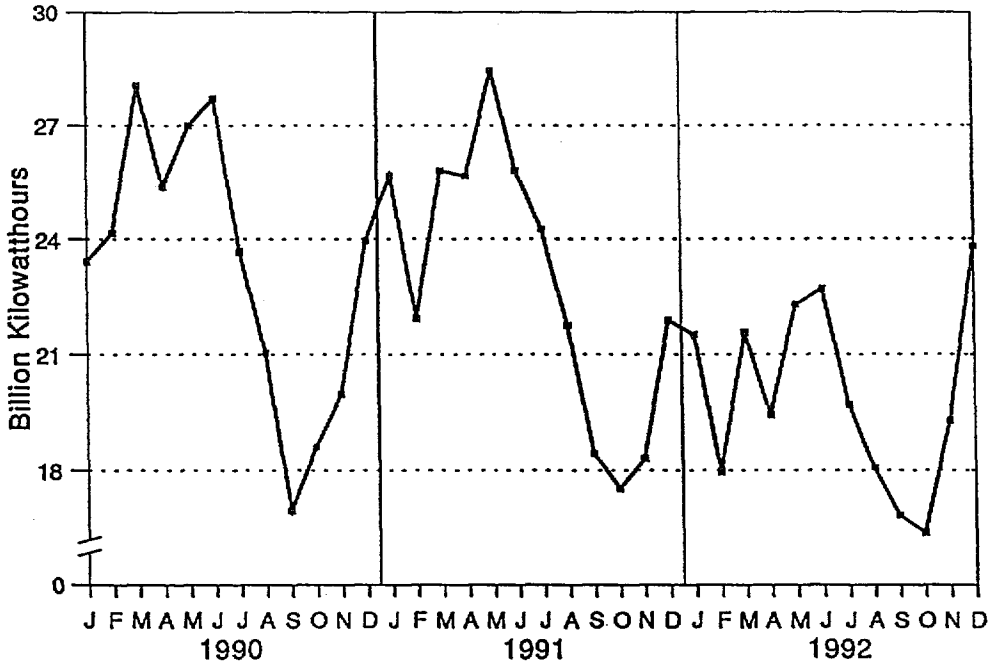
[1]George E. Box and Gwilym Jenkins, *Time Series Analysis, Forecasting and Control* (1976).

Figure 1
National Net Hydroelectric Generation by Year
1970 - 1992



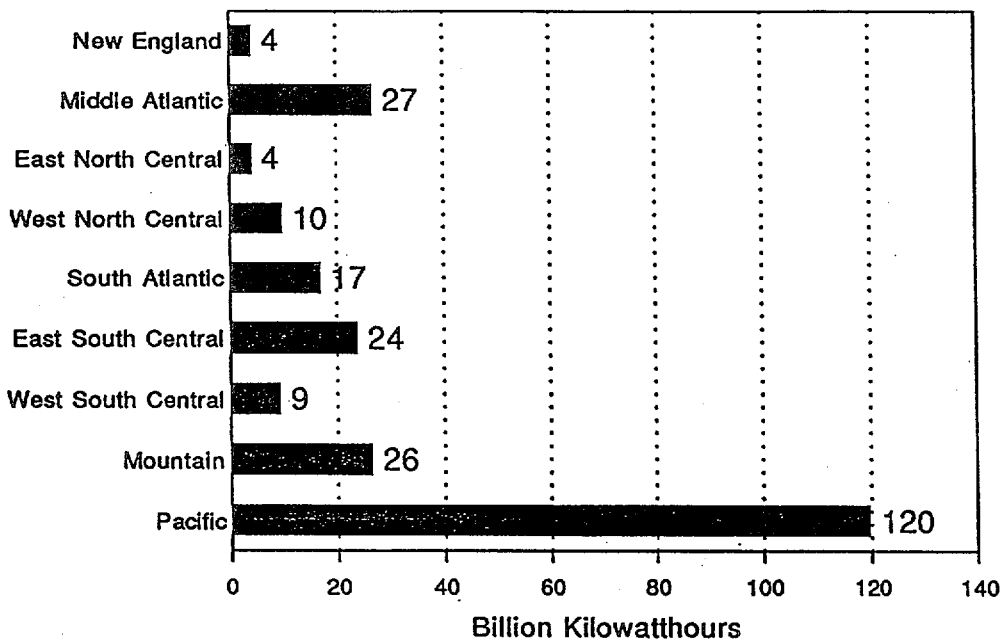
Source: Energy Information Administration, Form EIA-759, "Monthly Power Plant Report" and its predecessors.

Figure 2
National Net Monthly Generation
1990 - 1992



Source: Energy Information Administration, Form EIA-759, "Monthly Power Plant Report."

Figure 3
Net Hydroelectric Generation by Census Division
1992



Source: Energy Information Administration, Form EIA-759, "Monthly Power Plant Report."

Since historical monthly net hydroelectric generation data are correlated and the level of generation is affected by the availability of water, a time series model structure was selected. The model predicts future generation as a function of past generation, fixed seasonal components, and a lagged effect from precipitation levels. Precipitation levels determine the availability of water in reservoirs for electricity generation, as well as the effects of recent rain or snow on run-of-river hydroelectric facilities. It was found that lags of precipitation of 1, 2, 8, and 9 months were important in predicting generation.

The model is based on normal ARIMA modeling assumptions. That is, the data are assumed to follow a model which is linear and time invariant with constant coefficients. Fixed monthly values (dummy variables with estimated coefficients) are used to describe seasonality. For forecasting, precipitation is assumed to be "normal" from the last month of available data forward, where normal is defined to be the average monthly precipitation for a 57-year period.

Inputs to the model include two data series: (1) national totals of the plant-level net hydroelectric generation data reported monthly on the Form EIA-759, "Monthly Power Plant Report," and its predecessors; and (2) monthly precipitation data from the National Climatic Data Center.[2]

Specific model structure was selected from the family of ARIMA models based on standard diagnostics: the autocorrelation function, partial autocorrelation function, and inverse autocorrelation function of both the generation data and the model residuals, t-statistics for estimated parameters, the Portmanteau goodness of fit Chi square test for departures from model assumptions, the Akaike Information Criterion (AIC) to select the best from competing models, and comparisons of the model forecasts to actual data otherwise known as out-of-sample testing.[3]

The model was developed using the software SAS version 6. The initial step of the model reads in the historical net hydroelectric generation and precipitation data (Figure 4), and the data are divided by the number of days in each month in order to account for the different lengths of the months. Next, the first ARIMA model for generation is run to estimate coefficients for the monthly dummy variables and the appropriate lags of the precipitation data. The ARIMA procedure is a standard procedure in SAS and is based on the Box-Jenkins strategy for time series modeling. The procedure provides information for time series model identification, parameter estimation, and forecasting.[4]

As noted above, the model includes two functions that explain the time varying mean of the generation data, one due to seasonal variation and one due to the impact of precipitation. The seasonal effect on the generation is estimated using dummy variables each representing a month and estimated coefficients from the first ARIMA run. The impact of precipitation on the generation is estimated using the actual precipitation for past months and coefficients estimated in the first ARIMA run.

For purposes of forecasting, these two mean functions are subtracted from the historical generation data and the ARIMA procedure is run with this adjusted data, reestimating the autoregressive and moving average parts of the model. This ARIMA procedure provides the forecast for adjusted generation.

The seasonal effects and precipitation effects (using the assumed "normal precipitation") are added to the forecast of adjusted generation and the sum is multiplied by the number of days in each month. The output from this step is the final forecast.

After developing this model, procedures were established to evaluate the accuracy of the forecast and compare the results with our previous methodology. Until recently, the EIA had been using an informal methodology to produce the net hydroelectric generation forecasts. For each month during the current water year (which ends September 30) or a nine-month period from presently available data if longer, hydroelectric generation projections were based upon information obtained by phone from 10 utilities or organizations representing eight U.S. geographic regions. Total hydroelectric generation in each of the eight regions was projected to change by the same percentage as generation for the organizations contacted by phone.

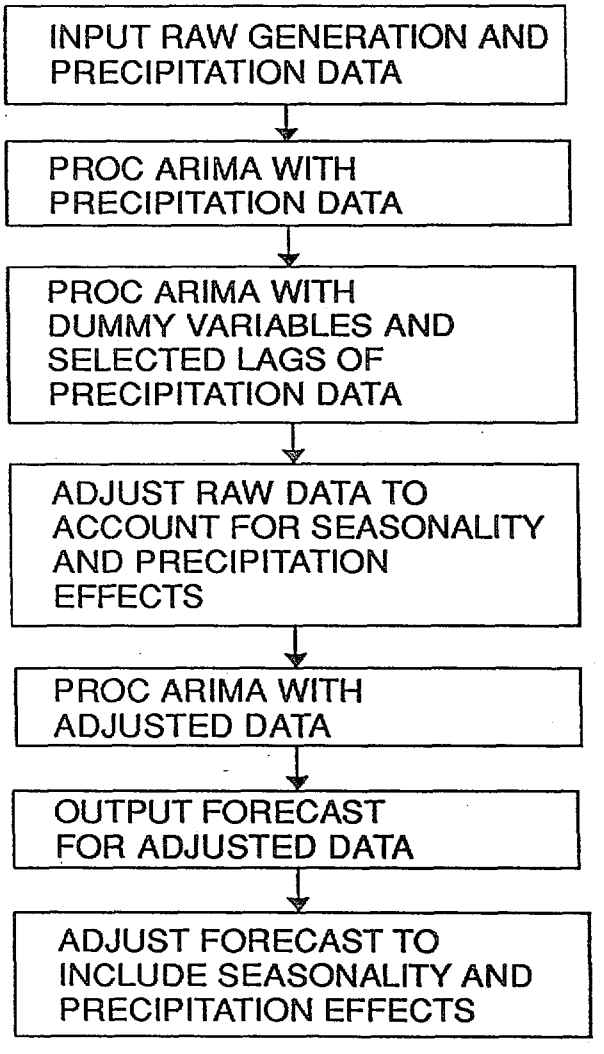
---

[2]**1970-1987:** National Climatic Data Center, *State, Regional, and National Monthly and Annual Precipitation Weighted by Area for the Contiguous United States January 1931 - December 1987* (Asheville, NC, August 1988), p. 66. **1988-1992:** National Climatic Data Center, *Monthly State, Regional and National Heating Degree Days Weighted by Population* (Asheville, NC, March, 1989 through 1993), Table 3.3, "Regional and National Average Precipitation."

[3]For more information, see: SAS Institute Inc., *SAS/ETS User's Guide*, Version 6 (Cary, NC, January 1989).

[4]For more information, see: SAS Institute Inc., *SAS/ETS User's Guide*, Version 6 (Cary, NC, January 1989), pp. 99-100.

Figure 4
Flow Diagram of the STHGM



Hydroelectric generation in the succeeding years was assumed to be normal. Normal generation was calculated by using hydroelectric capacity information from the Form EIA-860, "Annual Electric Generator Report," and historical monthly capacity factors averaged over 10 years.

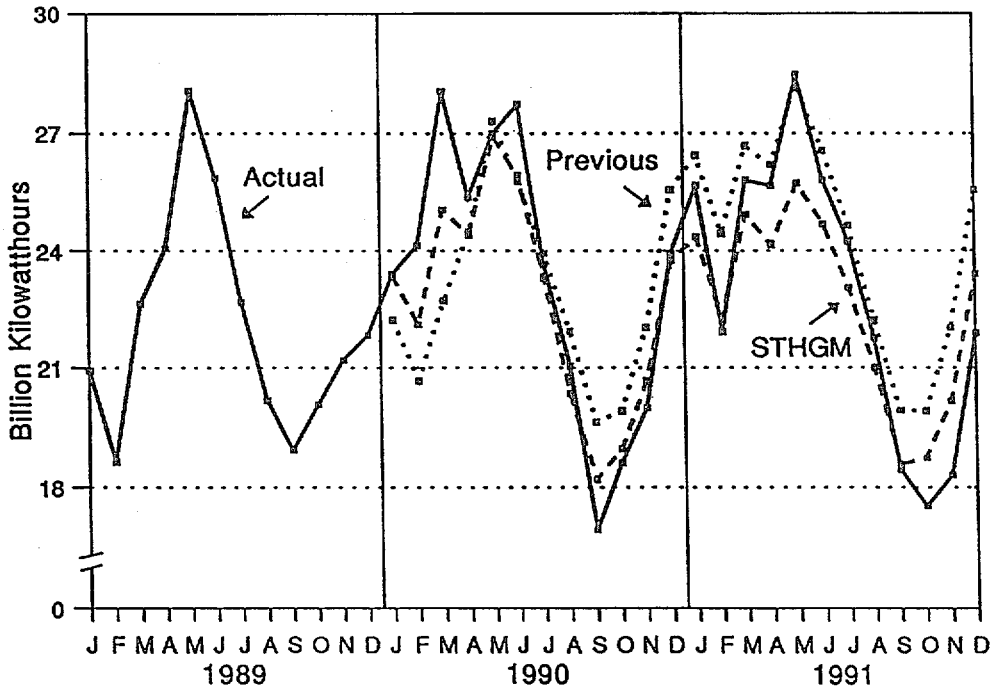
Evaluation of the model forecasts and forecasts from the previous methodology was performed using out-of-sample tests. That is, several different periods of actual data were deleted from the input data, the parameters in the model were reestimated, and forecasts made as described above for a 24-month period. These forecasts were compared to actual data and past forecasts using EIA's previous method obtained from memorandums produced when comparable historical data were available. Also, the Root Mean Square Errors (error) of the model estimates and the previous method were calculated and compared.

The first out-of-sample test was performed using historical data through 1989 (Figure 5). From the graph, the following were illustrated: (1) the model forecast was in general closer to the actual generation than the previous forecasts; and (2) the 1990 and 1991 actual data were close to the long run average hydroelectric generation and the model forecasts for these years was generally within 1 to 2 billion kilowatthours of the actual generation. The Root Mean Square Error (Table 1) of the forecasts using historical data through 1989 was 5 percent for 1990, the first year of the forecast, which was 45 percent less than the error for the previous method. For the 24 month forecast, the model had a 6 percent error, 38 percent less than the previous method.

The second out-of-sample test used historical data through 1991 (Figure 6). The graph illustrated the following points. First, for 1992, an unusually low year for hydroelectric generation, the forecast was not as good as for 1990 and 1991. This can be partially explained by the forecast assumption of normal precipitation while precipitation was below normal in 1992. Secondly, in 1992, with the below normal hydroelectric generation, the model still provided a closer forecast than the previous method. The Root Mean Square Error (Table 1) for 1992 for the model when using historical data through 1991 was 13 percent, 31 percent less than the previous method error. For both out-of-sample tests, the Root Mean Square Error in all cases was smaller for the model than for the forecasts from the previous methodology.

In conclusion, the Short-Term Hydroelectric Generation Model has been shown to improve upon EIA's prior procedure for projecting short-term hydroelectric generation. Of course, various uncertainties could affect the forecast, and, therefore, the validity of the forecasts rests on the legitimacy of the assumptions.

Figure 5
Comparison of the Forecast Methodologies with Actual Data, 1990 and 1991 (Forecasting with Actual Data through 1989)



Notes: • All forecasts are based on actual generation and precipitation data through the indicated year. • All forecasts assume normal precipitation during the forecast period.

Source: STHGM Forecasts: Energy Information Administration, STHGM run using HYDRO.TIME.SERIES.FINAL.D060193. Actual: Energy Information Administration, Form EIA-759, "Monthly Power Plant Report." Previous Forecasts: Energy Information Administration, Office of Coal, Nuclear, Electric, and Alternate Fuels memorandum to Office of Energy Markets and End Use dated March 13, 1990.

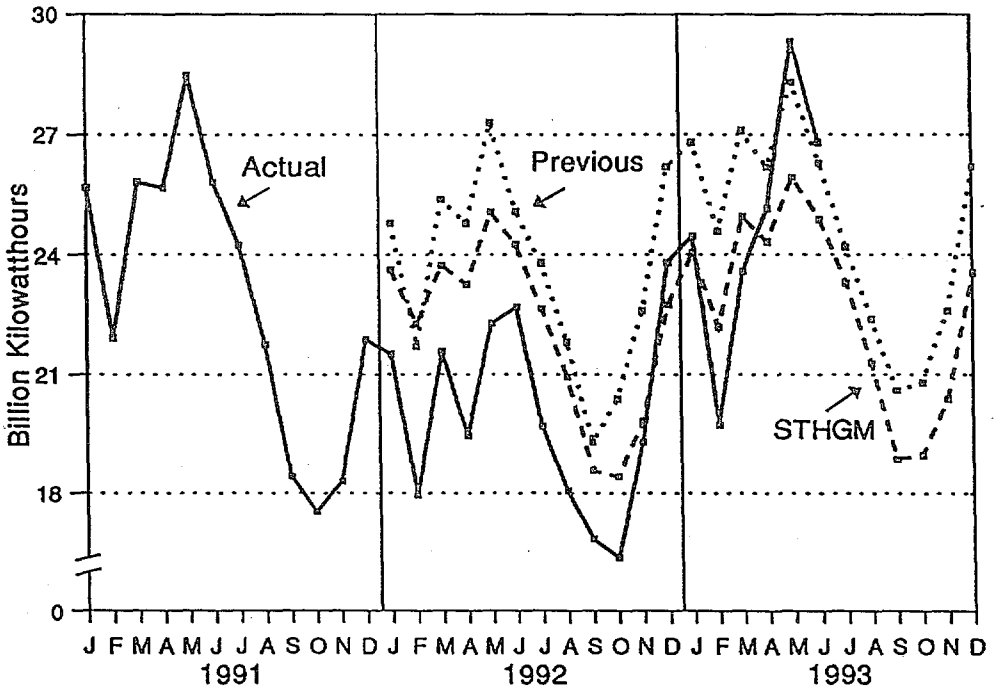
Table 1
Comparison of
Root Mean Square Errors

Using Historical Data through:	Year(s) of Forecast Included in the RMSE	Error (Million Kilowatthours)	
		STHGM	Previous
1989:	1990	1,267	2,289
		5.4%	9.8%
	1991	1,367	1,923
		6.0%	8.4%
	1990 and 1991	1,318	2,114
		5.7%	9.1%
1991:	1992	2,597	3,749
		13.0%	18.8%

Notes: • All forecasts are based on actual generation and precipitation data through the indicated year. • All forecasts assume normal precipitation during the forecast period. • Percents were calculated by dividing the error by the 12 (or 24) month average of the actual generation.

Source: STHGM Forecasts: Energy Information Administration, STHGM runs using HYDRO.TIME.SERIES.FINAL.D060193. Actual: Energy Information Administration, Form EIA-759, "Monthly Power Plant Report." Previous Forecasts: Energy Information Administration, Office of Coal, Nuclear, Electric, and Alternate Fuels memorandums to Office of Energy Markets and End Use dated March 13, 1990 and March 6, 1992.

Figure 6
Comparison of the Forecast Methodologies with Actual Data, 1992 and 1993 (Forecasting with Actual Data through 1991)



Notes: ◦All forecasts are based on actual generation and precipitation data through the indicated year. ◦All forecasts assume normal precipitation during the forecast period.

Source: STHGM Forecasts: Energy Information Administration, STHGM run using HYDRO.TIME.SERIES.FINAL.D060193. Actual: Energy Information Administration, Form EIA-759, "Monthly Power Plant Report." Previous Forecasts: Energy Information Administration, Office of Coal, Nuclear, Electric, and Alternate Fuels memorandum to Office of Energy Markets and End Use dated March 6, 1992.

POLICY ANALYSIS

MODERATOR	David Rejeski, Environmental Protection Agency
PRESENTATION 1	H.O. Stekler, Industrial College of the Armed Forces
PRESENTATION 2	Karen Fisher-Vanden, Jae Edmonds, and Hugh Pitcher, Pacific Northwest Laboratory
PRESENTATION 3	Jerome C. Glenn, United Nations University

Policy Analysis

Forecasting for Large Policy Changes

H.O. Stekler, Industrial College of the Armed Forces

Most forecasters working for the Federal government engage in the typical prediction exercises. We try to determine the effect either of small changes or of events about which we have some knowledge. In other words, we are able to make forecasts using models which replicate the past and which we assume will explain the future. For those incremental changes and for those events about which we have knowledge, this is the appropriate procedure. However, what should we do when there are fundamental structural changes occurring? In those situations the models of the past are probably not reliable for predicting the future. Moreover, we have no experience with the events that are taking place and have great difficulty in explaining the phenomena. We frequently then rely on theoretical analyses which may not yield accurate results in the real world because the assumptions of that theory do not hold in practice.

This paper will do two things. First, I look at the outcomes that were predicted to occur after a major policy change was enacted-airline deregulation. This involves a comparison of what was expected to happen when deregulation occurred with the outcomes we have observed and an explanation for the successes and failures. My thesis is that we must use one of two entirely different approaches to predict the effects of large policy changes. The approach that I will suggest is similar to the methods that the armed forces have used in determining the effects of alternative military strategies-simulated situations which replicate the conditions being investigated or experimentation.

I. Airline Deregulation

Airline deregulation occurred in 1978, but the expected effects of that policy change had been analyzed for a number of years prior to the event. These studies are too numerous to list here, but Levine (1987) has provided a comprehensive review of the major articles and books. Moreover there have been a substantial number of post-event explanations of the reasons that the actual outcomes differed from the predictions. Levine (1987) undertook the most comprehensive investigation of this question, while Kahn (1988) also provided his insights about the surprises that occurred as a result of deregulation. This paper will not replicate the analyses that were undertaken in these papers. Rather we shall only focus on the differences between the predictions and the outcomes, determine why they occurred and whether those discrepancies should have been surprises.

A. Predictions

The basic prediction was that a deregulated airline industry would display all the characteristics of a purely competitive industry and that prices would be substantially lower than they were with regulation. There is general agreement about those forecasts: the first is patently wrong (although there are individuals who argue that the industry is still in transition); the second is correct. Economists were correct in predicting what would happen to prices when the artificial constraints that regulation imposed on the industry were removed. They were unable to predict what the structure of the industry would be once deregulation occurred.

The predictions abut the post-regulation structure of the industry included statements that there would be as many as 200 airlines, each operating as few as six aircraft. Prices would be lower and uniform for all customers, with perhaps some discounts for off-peak flights and the airlines which were in existence at the time of deregulation would have a difficult time competing with the new entrants. Entry and exit from the industry would be easy since aircraft were mobile assets. (Kennedy Hearings, 1975.)

Finally, virtually all analysts focused on linear or point to point markets, for those were the markets which prevailed under regulation and were the only situations observed in the unregulated intrastate environments of Texas and California. In other words, the analysts based their forecasts on the experiences that they had and extrapolated into the future on the basis of this model and/or knowledge. They did not attempt to determine whether this would

be a good approach for predicting what would occur after the policy change took place and a new environment was created.

B. Outcomes

The outcomes differed markedly from these forecasts. The U.S. airline industry clearly does not resemble a purely competitive industry. There are only a small number of large firms, and, with perhaps one exception, all were in existence in 1978. In fact, the industry is now at least as concentrated as it was prior to deregulation, with three dominant firms. Although a number of the old line firms--such as Braniff, Eastern and Pan Am no longer exist, most of the new entrants also have failed. Prices are lower as predicted but they are not uniform with all types of restrictive fares in existence. The hub and spoke systems of the major carriers were clearly not foreseen. Why was there such a disconnect between the predictions and the outcomes? Moreover, why were these outcomes such surprises to those who made these predictions?

C. The Predictive Theory

Levine provided a partial answer to this question. He indicated (1987, p.403) that the proponents of deregulation had not considered the relationship between airline markets and the number of firms that would serve them. He, himself, had noted that there was an inconsistency between the assumptions of pure competition and the number of firms that were operating in the various markets, but left it at that. Prior to the actual deregulation of the industry the formal theory (based on the value of information and principal-agent analysis) that Levine used to explain the outcomes that deregulation produced was available in rudimentary form. However, it was not used to predict what might happen once the regulatory constraints were removed from the industry. The economists who analyzed the industry primarily focused on the costs that regulation imposed on consumers and the economy. They also were concerned about the absence or presence of economies of scale, i.e. the costs of producing a given number of seat miles and the ease of entry. Several years after regulation was abandoned, a new theory about contestable markets was proposed. This theory basically states that,if specific conditions prevail, merely the threat of competitive entry would be sufficient to produce the competitive market result. However, the structural changes that have occurred are inconsistent with this theory. These include hub domination, fare structures that are complex, computerized reservation systems that have been used for strategic advantage, frequent flier programs,etc.

Thus, we can conclude that prior to deregulation the theory which could have predicted what would happen once regulation was abandoned was still not fully developed and definitely not used for analysis while the formal theory which was proposed after the event was inappropriate. "It is difficult to be reassuring about the effects of phenomena that you did not predict and can not explain." (Levine, 1987 p.396.) Is this an appropriate way to undertake public policy?

Economic theory usually makes a number of simplifying assumptions in order to generate broad general principles or generalities. It can not generally predict exactly how the principals in a market will react. However, the economists who were analyzing the industry should have realized that the market participants who were present prior to deregulation would endeavor to protect their interests once the regulatory constraints were lifted. In other words, the institutional arrangements that existed or might develop and the possible strategic reactions of existing firms were not sufficiently well analyzed.

In fact, market participants and financial analysts were better able to predict what would happen in the years immediately after deregulation than were those economists who engaged in theoretical speculations. Moreover, many of the structural changes which actually occurred were foreshadowed in, an even cursory reading of, the trade literature. While it is dangerous to look for corroborating evidence after the fact, some of the evidence is quite clear and should have caught the attention of the analysts. A few examples will suffice to illustrate this point.

D. Some Industry Views

The evidence suggests that the advocates of deregulation were surprised by the creation of hub and spoke airline systems. Yet the testimony of United Airlines at the Kennedy Committee Hearings indicated that the firm considered the connecting flights as an important means of filling its longer distant flights. In other words, they were not merely concerned with point to point transportation. That airline indicated that deregulators were

conducting their analyses based on city pairs, for the CAB regulated on that basis but in reality the country's system was definitely not linear. (Aviation Week, April 7, 1975, p.21.) This lack of concern with actual institutional situations may have been the factor that prompted that airline to complain that there was too little management experience among the deregulators. (Aviation Week October 13, 1975, p.90.) United was not the only firm that was aware of the potential of hubs, for Aviation Week (February 17, 1975, pp.20, 22) indicated that both United and American were focusing attention on their hubs in Chicago and Dallas, respectively. Moreover, Boeing indicated that hubbing had been the norm with respect to airline operations and that point to point service was a relatively new phenomenon. (AW November 29, 1976, p.7.)

Similarly, the potential value of using computer systems for setting differential fares was general knowledge prior to deregulation. TWA noted that discounts did not have to be the same on every flight (AW February 17, 1975, p.22) while United indicated that it was possible to develop capacity controlled fares. (AW August 15, 1977, p.24.) Another surprise was the increased role that travel agents have played. Yet, there had been general discussions about permitting the agents to have direct access to the airlines' computerized reservation systems and of providing cash incentives to the agents for ticketing on specific (international) carriers. (AW September 22, 1975 p.28; December 1, 1975 p.27; February 14, 1977 p.24.) Finally, there were observations that customers were relying to an increasing extent on travel agents because fares had become more complex. (AW February 20, 1978 p.25.)

As for the structure of the industry, United envisioned the number of trunk lines being reduced to three by the year 2026 with an increased emphasis on international travel. There would be a large number of commuter airlines serving the domestic market. Thus that company foresaw some of the trends that have occurred since 1978. (However, it was wrong in one respect, for it predicted that there would be no discount fares.) It was also noted that exit is not costless for airline planning takes a long time and there are key connecting points. (AW April 7, 1975, p.21; February 16, 1976, p.7). According to the perceived theory prevailing prior to deregulation, predatory pricing should not occur, but it did and its potential was recognized by industry observers and financial analysts. (AW August 4, 1975; March 28, 1977, p.9.)

With hindsight it is always possible to find some analyst who correctly foresaw some of the trends that have occurred. That was not the objective of citing the aforementioned examples, for one could have also found a number of industry executives who failed to identify the observed trends. Rather, the idea was to note that there was a wide divergence of views with respect to the likely outcome of deregulation between the academic community and knowledgeable industry leaders. When such a wide discrepancy exists, it might be advisable to adopt another approach to analyze the impact of a proposed policy change.

Obviously it would not have been possible to engage in large scale experimentation either prior to or during deregulation, but there is another technique that can take account of institutional or industry knowledge, for predicting the likely outcomes of large policy changes. This is a technique which would use the technical expertise of the academic and the institutional knowledge of the industry analyst. The approach would be to adapt the methodology used by the military to analyze the possible outcomes of adopting a particular strategy or tactic, war games.

II. War Games

A war game is an analytical device which attempts to recreate the conditions of an actual historical battle or war. Alternatively, it might create the probable or possible situations which might prevail in a hypothetical future war. The players, who are knowledgeable about the institutional conditions, deploy and employ their forces and resources using alternative strategies to determine whether the outcomes would remain unchanged even if the combatants had made different decisions. It also permits an analysis of the conditions under which the outcomes would have been changed.

The individuals who take part in the game play the roles of all of the opponents who actually did (or possibly might) participate in a conflict. The rules of the game require each player or side to make a move, which might involve the movement of forces, the initiation of an attack, or the creation of a new strategic alliance. When all the players have made their moves the outcomes must be determined. Since the interactive moves involve uncertainties and an element of randomness, the outcomes are clearly not deterministic. The rules of the game determine the outcomes and in every case involve some probability generating function. Statistical tables are provided in the non-computer version of these games while the black box rules embodied in the computer determine those results. Upon the adjudication of that move, the process is repeated until the game is completed.

These games are not limited to single battles or theaters of war. In fact, one commercial game involves the recreation of the entire social, economic, political and military structure of the 14th century as the players become involved in the 100 Years War. This is an online interactive computer game involving 300 participants. It is obvious that these games can be very complicated and can embody many of the institutional factors that cannot be captured by abstract models.

The advantages of these games are obvious. By placing players in interactive competitive situations, they permit analysts to uncover and evaluate alternative strategies that actual competitors might employ in real life situations. In other words the games take over where theory ceases. In addition, the sensitivity of possible outcomes to alternative assumptions can be tested. This methodology might reduce the observed tendency of individuals to place unrealistically high confidence in their own theory, judgement or scenario. (Tversky and Kahneman, 1982; Ascher, 1993).

III. Application to Policy Changes

In my opinion, this methodology could have been applied to an analysis of the deregulated airline market. If, prior to deregulation, a game had been developed to analyze the new conditions and if the players had been individuals knowledgeable about the institutional factors,the results of the game might have been more realistic than the actual predictions. Then there would have been far fewer surprises about the strategies of the market participants, for the players would have assumed the roles of executives of new and old airlines, travel agents, various classes of consumers, the government, etc.

I would, therefore, recommend that such games be used whenever large policy changes might be implemented. These games can be developed relatively quickly by individuals versed in the art of devising war games. The participants should be people who are intimately connected with the policy under consideration and who might be affected by it. (One drawback of using executives of companies that might be impacted by such policy changes is that the firm's true strategies might not be revealed. An alternative would be to use former industry leaders, who have no vested interest in the outcomes). Repeated plays of the game would reveal the range of possible outcomes and would assist both the executive and legislative branches of the government in enacting an appropriate policy and to take account of the possible loopholes that were discovered before the policy was actually implemented.

This proposal has some elements in common with the Delphi method of forecasting, but it also deviates from that technique. The Delphi method asks knowledgeable sources what outcomes are likely to occur, and when those responses have been received,
they are all revealed to all these experts. The range of outcomes is then discussed and usually a consensus prediction is obtained. The similarity to a war game is that experts are used in both cases. The difference is that in the game the players are placed in an interactive competitive environment and must make strategic decisions in real time. The latter approach is likely to provide a more realistic prediction of what the real outcome will actually be.

The suggestion that policy changes be modelled through a war game-like analysis is not at all that radical. Business schools have frequently used computer games as decision making devices for analyzing the strategies of firms. What is new is the suggestion that such an analytic device be used in the public policy arena for predicting the outcome of a significant change in policy. The new health policy is the obvious candidate for such a mode of analysis. Moreover, in this situation it might be possible to combine this mode of forecasting with some experimentation, since alternative policies might be tried in different states.

IV. Conclusions

In this paper I have tried to suggest that the customary forecasting procedures might not be appropriate when large policy changes are being considered. The airline deregulation policy was used as an example to illustrate this point, for there were a large number of differences between what was predicted to occur and what actually happened after deregulation was implemented. Experimentation and simulations were suggested as alternative methods for predicting the outcomes of large policy changes.

References

Ascher, William (1993), "The Ambiguous Nature of Forecasts in Project Evaluation: Diagnosing the Over-Optimism of Rate of Return Analysis," International Journal of Forecasting, 9, 109-115.

Kahn, Alfred E. (1988), "Surprises of Airline Deregulation," American Economic Review, Paper and Proceedings, 78, 316-322.

Levine, Michael E. (1987), "Airline Competition in Deregulated Markets: Theory, Firm Strategy and Public Policy," Yale Journal on Regulation, 4, 393-494.

Tversky, A., and Kahneman, D. (1982), "Judgment Under Uncertainty: Heuristics and Biases," in D. Kahneman, P. Slovic and A. Tversky, Eds. Judgment Under Uncertainty: Heuristics and Biases (Cambridge University Press, Cambridge).

The Second Generation Model of Energy Use, The Economy, and Greenhouse Gas Emissions

Karen Fisher-Vanden, Jae Edmonds, Hugh Pitcher, Dave Barns, Richard Baron, Sonny Kim, Chris MacCracken, Elizabeth L. Malone, Ronald D. Sands, and Marshall Wise, Pacific Northwest Laboratory

1.0 Introduction

The global change problem has been a major issue on the public agenda since 1988, when a combination of events including a drought in the United States midwest, Congressional hearings, and an international meeting in Toronto Canada on the subject brought the issue to the forefront of public consciousness. The Toronto meeting set an arbitrary goal of reductions in fossil fuel carbon emissions of 20% for developed nations. The 1992 Framework Convention on Climate Change called on nations to stabilize greenhouse emissions at 1990 levels. And in 1993 the President of the United States pledged the United States to stabilize national greenhouse emissions at 1990 levels by the year 2000. In this paper, we explore policy options for achieving reductions of carbon emissions and their economic implications using the Second Generation Model (SGM) of human activities and global change, a new, computable general equilibrium (CGE) model designed to analyze global change problems. Further, we explore some of the greenhouse emissions consequences of policy proposals such as a BTU tax.

We begin our discussion with a brief description of the model, describe the derivation of reference year parameters, proceed to develop reference case assumptions, describe the resultant reference case scenario, and then explore the costs and benefits of selected policy options for stabilizing greenhouse emissions.

2.0 Model Description

In this section we provide a brief overview of the model and its numerical embodiment of the United States economy. The theoretical structure of SGM Version 0.0 is described in detail in Edmonds et al. (1993). The development of estimates for parameters is documented in Fisher-Vanden et al. (1993).

2.1 Origins

The SGM's intellectual roots can be traced to the modeling work of Edmonds, Reilly, and Barns, who developed and exercised the Edmonds-Reilly-Barns model of long-term global energy and greenhouse gas emissions (Edmonds and Reilly, 1985; Edmonds and Reilly, 1986; Edmonds and Barns, 1991). This model (member of the first generation of models) focused on emissions of energy related greenhouse gases. The model was simple and transparent, but had long time steps, did not consider non-energy related greenhouse gas emissions and sinks, and completely neglected impacts on human systems from global change. The SGM is structured to provide both increased scope (including both energy and non-energy related emissions activities and a framework designed to assess impacts of global change on human activities), and finer resolution of human activities (including a 5-year time step, enhanced technology descriptions, and an integrated demographic module).

The analysis reported here uses SGM Version 0.0. While 20 global regions are being developed for the global implementation of the SGM, only the United States module is used in this exercise.

2.2 Overview of SGM Version 0.0

In summary, the SGM is a member of the CGE class of models. The SGM Version 0.0 has nine producing sectors:

1. Agriculture
2. Crude Oil
3. Natural Gas Production
4. Coal Production
5. Uranium Production and Refinement
6. Electric Power Generation

7. Oil Refining
8. Natural Gas Transformation and Distribution
9. Other Production

We note that seven of the nine sectors of the economy in SGM Version 0.0 are energy sectors. The emphasis on the energy sector in SGM Version 0.0 reflects the central role energy plays in the determination of emissions of greenhouse gases to the atmosphere.

In addition to the producing sectors of the SGM, there are four final demand sectors:

1. Households
2. Government
3. Investment
4. Net Export

Producing sectors use goods and services produced by other producing sectors, itself, and primary factors of production to produce net output. The three primary factors of production are:

1. Land
2. Labor
3. Capital

There is a market, which is cleared by the price mechanism, for each producing sector of the economy. Similarly there is a market for each of these three primary factors of production.

The interactions of the main components of the SGM are shown in Figure 1, where the seven energy sectors are included in the energy box. In general energy, agricultural, and other production sectors supply net quantities of goods and services for each other and for final consumption by households, the government, investment, and net exports. Primary factors of production - land, labor and capital - are also used by the producing sectors. Labor and land are assumed to be owned by members of the household sector who supply them to the market. Capital is associated with producing sectors. Net profits are returned to the household sector, while taxes are collected from producers and households and provide the revenues for local, regional, and national governments.

Producing sectors are assumed to make decisions regarding production and investment with the objective of maximizing expected wealth. Governments are assumed to produce government services, including education, national defense, and general services, according to prescribed production functions. Households behave as if they are maximizing utility with regard to the allocation of resources to current consumption; however, the supply of savings and labor is modeled by exogenously specified rules.

All produced goods and services, and primary factors of production pass through markets which are assumed to clear in every period via a price mechanism. Estimates of gaseous emissions are computed in the model via the application of emissions coefficients to inputs and outputs of processes. Variation in these coefficients reflect variation in the emissions characteristics of alternative human activities.

In the remainder of this section, we will briefly highlight some of the more salient features of the SGM.

2.3 Subsectors and Technologies

Within the nine producing sectors of SGM Version 0.0, several sectors have subsectors and technologies defined. Each subsector, defined for a specific sector, is assumed to produce a homogeneous good which becomes part of the total production of the sector. For example, the Electric power generating sector has six subsectors defined: Oil, Gas, Coal, Biomass, Nuclear, and Solar/Hydro. Subsectors defined for the SGM Version 0.0 are shown in Table 1.

Table 1: Subsector in SGM Version 0.0

Producing Sector	Number of Subsectors	Description of Subsectors
1. Agriculture	1	Agriculture
2. Crude Oil Production	5	Crude oil is treated as a depletable resource with five grades modeled as subsectors.
3. Natural Gas Production	5	Natural gas is treated as a depletable resource with five grades modeled as subsectors.
4. Coal Production	5	Coal is treated as a depletable resource with five grades modeled as subsectors.
5. Uranium Production and Refinement	5	Uranium is treated as a depletable resource with five grades modeled as subsectors.
6. Electric Power Generation	6	Power generation by mode is defined for Oil, Gas, Coal, Biomass, Nuclear, and Solar/Hydro.
7. Oil Refining	3	Conventional Oil Refining, Coal Liquefaction, and Biomass Liquefaction.
8. Natural Gas Transformation	3	Natural gas transformation and distribution is defined for Natural Gas, Coal Gasification, and Biomass Gasification.
9. Other Production	1	Other Production

Just as sectors can have subsectors, subsectors can have technologies. Technologies are alternative modes for producing the product of the subsector. For example, there may be two technologies for producing electricity with natural gas (a subsector of Electric Power Generation): Conventional combustion, and Combined Cycle combustion. The framework of the model is sufficiently flexible to allow an arbitrary number of technologies to compete.

2.4 Natural Resources

Natural resources are treated explicitly in the SGM, which identifies two forms: depletable and renewable. Depletable resources are consumed in use, for example, fossil fuels. Renewable resources are not consumed in use, for example, land. This distinction is important to the treatment of energy production and transformation, as well as agricultural activities.

Fossil fuels and uranium are treated as depletable natural resources. They are divided into two categories, Resources and Reserves. Reserves are those energy sources whose location is known, which are producible using present technologies under present and anticipated economic conditions, and which investment for extraction purposes has been made. (Some crude oil, discovered conventional, and some Alaskan oil are examples of crude oil reserves.) Resources include all energy sources, including those which are known by location and those whose existence is inferred, those which are producible under present technologies and economic conditions, and those which may require greater incentives to exploit. Total resources include reserves. (In the SGM, reserves are created as economically recoverable resources are discovered). In the SGM, energy production from depletable natural resources occurs only from reserves. The rate of production from reserves depends on the amount of productive capacity put in place with discovery, and prices of inputs and outputs.

Biomass, solar power, and hydroelectric power are renewable resources. Biomass feedstocks are treated as a component of the agricultural production. Total domestic production is therefore constrained both by the total

available arable land, and the competition for the use of that land for other purposes. Biomass feedstocks are either transformed to liquids or gases, or consumed as a solid by electric utilities.

Electric utilities also consume solar and hydro power. Hydro power capacity is fixed for the purposes of this analysis. No resource constraint is placed on the solar power component, a source of electricity generation based on photovoltaic array technology.

Land is modeled as a separate factor of production. The identification of land reflects its potential importance as a constraint in emissions reduction strategies which utilize biomass or carbon sequestration by trees.

2.5 Expectations, Capital Formation, and Productivity

In the SGM the demand for capital by producing sectors depends on expected profits. Expectations of profits in turn depends on both the technology which describes the relationship between inputs and outputs, and expected prices over time, for inputs and outputs, including capital. The SGM provides a variety of options for describing the formation of price (and tax/subsidy) expectations, including assumptions that

1. Prices will remain at current levels over the expected life of the equipment (assumed in this exercise),
2. Prices will change at rates which reflect behavior over a prior period of experience, and
3. Price expectations are exogenously specified (which can be used to generate a rational expectations price path).

The last capability can be used to explore behavior when future prices are known. We note also that different combinations of expectation formation for market prices and taxes/subsidies can be constructed. Thus, the formulation of price expectations is not restricted to the assumption of perfect knowledge about future events.

Sectors and subsectors with the highest expected profits experience the greatest realized investments, although within a sector a logit function distributes investment resources across all subsectors which are economically attractive. Investment opportunities which have zero or negative expected contributions to wealth (after allowing for taxes and subsidies) receive no investment funds. The inclusion of subsidies in the expected contribution to wealth calculation is important.

The SGM uses a "putty-clay" specification of capital. That is, once an investment occurs, capital is permanently associated with that particular application. Thereafter, capital does not move from one sector to another. This is a particularly useful assumption for modeling energy applications where capital investments are highly specialized. It is also not particularly restrictive to the rest of the economy, which is modeled as a single aggregate sector.

Using a vintage approach to modeling capital also means that capital investments become a fixed cost of production. Thus, existing vintages continue to operate as long as they can cover their operating expenses, even if the rate of profit differs from that anticipated at the time the original investment was made. We note here that other factors of production are not fixed; they can be varied to either expand or contract output from an existing facility. Other factors of production also move freely among alternative applications.

Because a production function is associated with each capital investment by application, factor productivity can be manifest as both embodied and disembodied changes in input requirements. That is, combustion efficiencies may be embodied in the physical plant and vary depending upon the date of installation. On the other hand, agricultural productivities may be weakly associated with capital stocks, and more closely associated with changing management practices and seedstocks. The SGM can represent both types of productivity change.

2.6 Demographics

Estimates of population and its structure are developed within the SGM by a demographic module. The SGM is therefore capable of creating an array of internally consistent estimates of population and critical details such as the associated size of the work age population disaggregated by gender. The SGM demographic module builds population estimates from assumptions about age specific fertility rates, survival rates, and net immigration rates.

2.7 Emissions

The SGM was explicitly designed to provide estimates of gaseous emissions from all human activities, including those associated with energy, agriculture, and industrial processes. Not only does the model yield estimates for carbon emissions, but it also tracks emissions of CO, CH_4, VOCs, N_2O, NO_x, and SO_2. These emissions are associated with specific human activities, and where appropriate, with specific technologies.

2.8 Data Sources and Calibration

Parameters for the United States module of the SGM are derived so as to be consistent with behavior of the United States economy in 1985, investment behavior over the prior 26 years, and physical flows of energy resources in 1985, as well as the distribution of resources and reserves in that year, the demographic structure of the United States population, and emissions rates by human activity in 1985.

A variety of statistical information was used to develop calibration parameter estimates for the SGM. Table 2 contains a summary of these data, their sources, and an indication of their use.

Table 2: Data Sources for SGM Version 0.0 Calibration to 1985 United States Economy

Data Type	Data Source	Use
1985 Input-Output Table	DOC (1990)	Provide statistics for flows between different economic activities in 1985.
1985 Energy Flows	EIA (1991)	Provide benchmark flows of energy throughout the economy for 1985.
Investment Patterns 1959-1985	DOC (1987)	Provides pattern of capital accumulation by sector.
Resource and Reserve Estimates	EIA (1992) DOE (1991a)	Provide total resource constraints on fossil fuels and uranium, distinguish between reserves and resources, provide a basis for allocation of reserves and resources to economic grades.
Demographic Profile	DOC (1992)	Estimate fertility, survival rates, and net immigration in the demographic module.
Land	DOC (1992) USDA (1992)	Estimate payments to land by sector of the economy.
Emissions	DOE(1991a)	Provide estimates of emissions fluxes by human activity.

The methods used to transform data into model parameters are documented in detail in Fisher-Vanden, et al (1993).

3.0 Reference Case Assumptions: 1985-2030

We begin with the analysis of policy options for controlling United States greenhouse gas emissions by constructing a reference case against which to view policy derived impacts. We firmly believe that no reference case that we could construct today is capable of representing the future. An alternative to a single or multiple reference cases is the development of probabilistic scenarios. But while these map out a range of alternative futures systematically, they suffer from problems of specifying input parameter probability distributions as well as problems with covariance among exogenous variables. Furthermore, such scenarios lead to difficulties in analyzing and communicating results. We therefore chose the use of a reference case which is reproducible, and for which policy induced variations can be explained.

The SGM United States module requires input assumptions in six different areas: Demographics, Energy Resources, Productivity Change, Nuclear Power, International Trade, Fiscal Policy, and Emissions Coefficients. We discuss each in this section.

3.1 Demographic Assumptions

The demographic module in SGM Version 0.0 develops population estimates using exogenous fertility, survival, and migration rates. Population is differentiated by gender and 5-year age group.

Figure 2 shows the demographic profile of the U.S. for our benchmark year of 1985 (Department of Commerce, 1992). The bulge in the chart for the age groups of 20-40 years corresponds to the "baby boom" generation. Even by 1985, it is clear that the cohorts that follow are much smaller. By the first quarter of the next century, the population bulge will have reached retirement age, and the fraction of population that is available for the labor supply may be much smaller than it is currently.

In SGM version 0.0, fertility rates, survival rates, and migration rates are determined exogenously. We have specified initial values for these rates based on data for 1985. We also included a terminal value for each of these rates as a modeling parameter. Both the terminal values and the time to reach the terminal values can be specified by the user. Table 3 shows the initial values of the demographic rates which are assumed to persist throughout the modeled period.

Table 3: Demographic Rates (per thousand people)

Age	Fertility Rates	Male Birth Fraction	Female Death Rates	Male Death Rates	Migration Rates
0-4	0.0	.51	2.35	2.89	2.80
5-9	0.0	.51	0.20	0.3	2.30
10-14	0.0	.51	0.20	0.30	2.30
15-19	53.6	.51	0.50	1.10	2.30
20-24	111.5	.51	0.60	1.70	5.60
25-29	113.4	.51	0.60	1.70	3.70
30-34	73.7	.51	0.80	1.90	2.30
34-39	27.9	.51	1.10	2.30	2.30
40-44	4.8	.51	1.80	3.20	2.30
45-49	0.0	.51	2.90	5.20	1.40
50-54	0.0	.51	4.60	8.50	1.20
55-59	0.0	.51	7.30	13.50	0.50
60-64	0.0	.51	11.20	20.80	0.50
65-69	0.0	.51	16.70	31.00	0.50
70-74	0.0	.51	26.00	47.70	0.50
75+	0.0	.51	75.89	100.97	0.50

Source: Department of Commerce (1992).

3.2 Depletable Energy Resources and Reserves

By definition the resource base is the total quantity of a resource which could be produced over all time using any conceivable technology. It includes both discovered and undiscovered quantities. It is therefore the physical constraint on cumulative production. As indicated earlier, we distinguish energy resources for oil, gas, and uranium by grades reflecting variation in the ease of extraction which in turn drives costs. Five grades are defined for this exercise. The estimates of the energy content available in each of the five grades (excluding previously consumed quantities), as well as the relative cost of extracting each grade compared to Grade 1 are provided in Table 4.

Table 4: U.S. Energy Resources Remaining and Relative Costs by Grade

	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Total
Crude Oil (EJ)	80	170	208	253	379	1,090
Relative cost	1.0	1.4	1.8	2.4	3.4	
Natural Gas (EJ)	313	152	271	450	203	1,388
Relative cost	1.0	1.85	2.7	3.84	16	
Coal (EJ)	5,978	7,837	11,755	25,000	27,500	78,070
Relative cost	1.0	1.3	2.7	5.3	9.0	
Uranium (tons)	4,338	8,115	13,211	0	0	25,664
Relative cost	1.0	2.2	3.2	3.2	3.4	

Source: Based on estimates derived in Edmonds and Reilly (1985), and updated with information from DOE (1991a).

In the United States coal is by far the most abundant conventional fossil fuel resource. In fact, the domestic supply of coal in the cheapest grade--Grade 1--is enough to sustain current levels of production for nearly 300 years. Natural gas is the next most abundant domestic source of energy, although much of its supplies are in the more expensive grades. Gas resources in the first three grades are sufficient to sustain current production for approximately the next 35 years; therefore, the supply of relatively inexpensive natural gas becomes a binding constraint over the year 2030 study horizon of the model. Currently about half of the oil consumed in the U.S. is imported, so the implications of consumption on domestic resources are not as straightforward. Under current production levels, the supply of oil from the first three grades is clearly a constraint over the next 40 years. However, the future price of foreign oil plays a major role in determining domestic investment in the more expensive grades. Even so, domestic supplies of the more costly grades of oil are limited.

3.3 Productivity Change

One of the more important assumptions affecting the development of a reference description of the United States economy is the set of assumptions which govern productivity change in the variety of human activities in the SGM. Productivity changes in the SGM in either of two ways, through smooth changes in production function coefficients associated with new investment options, which we refer to as general productivity growth, or by the introduction of a new technology option at a specific point in time.[1]

Assumed rates of change for general, Hicks neutral, productivity in new investments are given in Table 5.

[1] Similarly the model can remove technologies from the available set; e.g., when new source performance standards are introduced.

Table 5: Assumptions of Exogenous Change in Total Factor Productivity for New Investments

Sector		Productivity Growth Assumption (%/yr)
1.	Agriculture	0.6
2.	Crude Oil Production	0.6
3.	Natural Gas Production	0.6
4.	Coal Production	0.6
5.	Uranium Production and Refinement	0.6
6.	Electric Power Generation	0.2
7.	Oil Refining	0.0
8.	Natural Gas Transformation	0.0
9.	Other Production	0.6

Since the Other Production Sector produces approximately 85 percent of total output, the productivity growth rate of this sector was adjusted to attain a growth forecast of the U.S. economy similar to what was assumed in the National Energy Strategy. Lack of detailed knowledge of variations in future productivity growth between sectors resulted in the assumption of no variation in productivity growth rates between the first five sectors and the Other Production Sector. Future versions of the SGM will incorporate current research in this area to better define variations in productivity growth between these sectors.

Productivity growth rates in two of the three energy transformation sectors are set to 0.0% per year to maintain energy balances between inputs and outputs. These three sectors--Electric Power Generation, Oil Refining, and Natural Gas Transformation--differ from the other producing sectors because they use large quantities of energy as inputs; most of which is passed through as output instead of being consumed. For example, the amount of energy leaving the gas distribution sector is nearly equal to the energy coming in, with almost no opportunity for improvement in the ratio of energy output to energy input. In these sectors changes in productivity are modeled as discrete changes in available production functions over time.

Although productivity improvements in the Electric Power Sector are handled mainly through the incorporation of new technologies in the SGM, small productivity improvements resulting from changes in management practices warrant the inclusion of a small neutral productivity growth assumption. This value (0.2) was chosen in order to attain results within the Electric Power Sector similar to those found in AEO93.

This analysis allows five new technologies to be introduced after 1990: liquids from biomass, gases from biomass, solar electric power, clean coal, and new gas turbine. Cost assumptions used to estimate production function parameters for these five technologies are given in Table 6.

Table 6: Annual Operating Costs for Hypothetical Plants

	Solar (1982 MM$)	Liquids from Biomass (1987 MM$)	Gases from Biomass (1986 MM$)	Clean Coal Technology (Atmospheric Fluidized Bed Combustion) (1989 MM$)	New Gas Turbine Technology (Natural Gas Combined Cycle) (1989 MM$)
Plant Capacity	1,000 MW	1000 dry tons per day	1000 dry tons per day	500 MW	300 MW
Materials	$16.50	$27.52	$15.67	$38.57	$15.39
Labor	$10.10	$6.92	$3.67	$18.15	$7.24
Capital (Production)	$187.10	$14.00	$10.88	$650.00	$180.00
Capital (Distribution)	$90.60	---	$4.17	$210.32	$126.22
Total Cost (without taxes)	$304.30	$48.44	$34.39	$917.05	$328.85
Indirect Business Taxes	$16.40	$1.85	$1.14	$9.83	$4.77
Total Cost (with taxes)	$320.70	$50.29	$35.53	$926.88	$333.62
Break-Even Price (1982$)	$0.11 per kwh	$15.71 per GJ	$8.71 per million BTU	$0.056 per kwh	$0.046 per kwh

Sources: Derived from Elliott, et al. (1990), Wan and Fraser (1990), DOE (1991a)

The two biomass technologies purchase their feedstock (wood) from the SGM's agricultural sector, thereby competing with other agricultural products for land. Future versions of the SGM will split the Agriculture sector into various crop sectors as well as a biomass feedstock sector.

The SGM allows for electricity sector investments in solar photovoltaics, a clean coal technology (Atmospheric Fluidized Bed Combustion), and a new gas turbine technology (Natural Gas Combined Cycle) beginning in the year 1990. New capital equipment is required for both generation and distribution. Operation and maintenance costs are very small relative to capital costs. Cost assumptions for solar are somewhat optimistic for the near term.

3.4 International Trade

Because world trade occurs, a single region model is not closed. Closing the SGM requires assumptions about the region's interactions with the rest of the world. These are given in Table 7. The crude oil market is assumed to be open, and the model can import or export as much as it wishes at an exogenously specified price. (This price path is taken from DOE (1991b)). The trade account is assumed to be balanced. This is accomplished by setting net exports of the Other Production sector, Sector 9, to the negative of the sum of net exports from other sectors.

Table 7: Closure Assumptions

Sector		Closure Assumptions
1.	Agriculture	1985 net exports fixed
2.	Crude Oil Production	World price path given exogenously
3.	Natural Gas Production	1985 net exports fixed
4.	Coal Production	1985 net exports fixed
5.	Uranium Production and Refinement	1985 net exports fixed
6.	Electric Power Generation	1985 net exports fixed
7.	Oil Refining	1985 net exports fixed
8.	Natural Gas Transformation	1985 net exports fixed
9.	Other Production	Set to balance trade account

3.5 Fiscal Policy

Budgets for all government activities taken together, including federal, state, and local, are assumed to be balanced. That is, federal deficits are assumed to be offset by state and local surpluses.

4.0 The Reference Case

The assumptions discussed in the preceding section generate a reference case trajectory for the United States economy, energy system, and greenhouse gas emissions. We discuss each in turn.

4.1 The Reference Case Economy

While the rate of population growth and level of economic activity are input assumptions for many models, they are results of the SGM. We begin by discussing trends in population and economic activity. Figure 3 shows the real gross national product (GNP) and its three principal components: consumption, investment, and government spending (recall that net exports are zero by assumption). GNP grows throughout the period of analysis, more than doubling between 1985 and 2030. We note, however, that the rate of growth slows distinctly after the year 2010. The rate of population growth slows to 0.29 percent/yr after 2010 compared with the post world war II period average of 1.01 percent/yr. By 2030, population is actually declining, even given relatively high immigration rates, and a decline in mortality rates.

In order to understand the economic forces at work within the SGM, it is useful to examine why GNP growth falls from 1995 to 2030. The decline is steady from 1995 to 2025 when growth rates appear to stabilize at about a half percent per year. This drop occurs primarily because of the fall in the size of working age population (Figure 4), from 182.5 million in 2010 to 167.8 million by 2030, an 8.0 percent decline. This leads to only a 2.9 percent reduction in the number of workers (149.2 million to 144.9 million) during the period 2010 to 2030 as labor force participation increases from 81.8 percent to 86.4 percent, and wage income per worker per year rises from $33,190 per year to $44,380 per year. As might be expected given the scarcity of labor, the capital/labor ratio increases over the same time frame from $136.9 to $165 million per 1000 workers, continuing the trend seen earlier (Table 8).

Table 8: Factors affecting Employment

Year	Work Age Population (millions)	At Work (millions)	Percent At Work	Wage Annual ($1000*)	Capital Labor Ratio ($1000* per worker)
1985	158.8	109.8	69.1	20.28	103.0
1990	163.0	116.3	71.3	21.88	108.7
1995	167.0	125.0	74.8	24.74	114.8
2000	172.5	134.0	77.6	27.56	122.8
2005	179.0	143.0	79.9	30.29	128.0
2010	182.5	149.3	81.8	33.19	136.9
2015	181.4	151.2	83.3	36.06	145.2
2020	177.3	149.8	84.5	38.78	154.0
2025	171.9	147.1	85.6	41.72	160.1
2030	167.8	144.9	86.4	44.38	165.0

* in 1982 baseyear (1985) dollars.

Since labor is scarce, the price of labor rises over time, leading to a substitution away from labor. This is seen in the much larger decline in the ratio of labor to output than in the ratio of capital to output. Figure 5 presents the path of prices for factors of production capital, labor, and land. Not only does the price of labor rise, but the fraction of the working age population at work rises as seen in Table 8.

The decline in the work age population has another route for affecting growth rates. Part of the determinant of the desired investment is the change in the rate of potential growth of the economy, namely the change in worker productivity and the change in working age population.

While productivity, measured as GNP/Worker, continues to grow rapidly (Table 9), working age population peaks in 2010 and then declines, thus depressing growth in investment. Examination of the components of GNP makes this clear, as investment becomes static after 2015.

This leads to an increase in the average age of capital, slowing down the rate at which technical change in realized in the economy. (By assumption, neutral technical change occurs at a constant pace in each industry).

Table 9: Factors Affecting the Capital Stock

Year	Capital Stock ($trillion)	Capital/Output Ratio	Interest Rate	Output/Worker ($1000)
1985	11.30	3.096	.0312	33.3
1990	12.63	3.066	.0313	35.4
1995	14.34	3.013	.0314	38.1
2000	16.45	3.024	.0312	40.6
2005	18.31	2.977	.0309	43.0
2010	20.44	3.006	.0306	45.6
2015	21.99	3.020	.0305	48.2
2020	23.00	3.029	.0304	50.7
2025	23.59	3.019	.0304	53.1
2030	23.87	2.968	.0303	55.5

4.2 The Reference Case Energy System

Total primary energy use increases steadily over the modeling period with an overall increase of approximately 55 percent between the years 1990 and 2030 (Figure 6). A period of rapid increase (averaging more than one percent annually) occurs during the period 1990 and 2005 mirroring the steady growth in the economy. A relative modest growth (less than half a percent average annual rate) in primary energy consumption occurs between 2005 and 2020 as a result of a slowdown in the economy, and an increase in oil and gas energy prices caused by the depletion of the less costly grades of domestic resources. This leads to a decrease in the ratio of energy use to GNP during this period. A period of accelerated growth in the last 10 years of the analysis (an average annual growth of slightly less than one percent per year) occurs due to a shift to electricity consumption increasing the consumption of coal.

Oil remains the dominant fuel over this period (despite the fact that its price is assumed to rise by more than 50 percent between the years 1990 and 2010) rising to almost 50 exajoules per year by the year 2015. Consumption is relatively stable throughout the remainder of the period of analysis. The escalation in oil consumption leads to a substantial increase in oil imports beginning in the year 1990.

The only primary energy carrier whose price remains relatively stable is coal. We note that the price of electricity rises by about 50 percent between the years 1990 and 2010, but the introduction of improved technologies and a decline in coal prices lead to a reduction of cost by more than one third between 2010 and 2030.

The relative stability of the price of coal compared to other fuels leads to an increase of domestic coal production and consumption which rises by more than 50 percent between 1990 and 2030. Natural gas consumption increases by 40 percent between 1990 and 2015, but depletion of the least expensive grades of domestic natural gas resources after 2015 causes a sharp increase in the price of natural gas and thus stabilizes consumption.

The assumption that no new nuclear powerplants are built leads to a stable pattern of nuclear electricity production in the near term, but after the year 2010 production declines steadily until all nuclear facilities are retired in the year 2025.

Electric power consumption as a whole expands steadily over the period of analysis. Larger increases in consumption of electric power occur after 2015, when a fall in the relative price of coal causes a shift towards electricity consumption.

4.3 Reference Case Greenhouse Gas Emissions

The focus of this analysis is to assess policy options for stabilizing United States greenhouse gas emissions by the year 2000 at 1990 levels. The reference emissions provide an important point of departure. We begin by examining anthropogenic carbon emissions. In the United States carbon is emitted predominantly by fossil fuel combustion, though total emissions include emissions from other human activities, notably cement manufacture and agriculture.[2] The time profile of emissions of carbon is shown in Figure 7. In the reference case emissions are anticipated to rise relatively rapidly between the years 1990 and 2000, at an average rate of 1.6 percent per year. The somewhat more rapid ramp-up in emissions than in total energy use is the consequence of an increased fraction of coal in the energy mix. While the rate of growth of carbon emissions declines somewhat over time, the continuing increase in the share of energy use in the form of coal results in a growth in carbon emissions that remains above that of total energy use.

Emissions from other greenhouse gases are displayed in Figure 8. Since the majority of carbon monoxide (CO) emissions are a result of oil combustion activities, we see the growth in these emissions mirroring that of oil consumption. Since we see a steady increase in oil consumption over the modeling period, we also see a steady rise in CO emissions. Other greenhouse gas emissions also experience a steady increase over the modeling period. SO_x emissions stabilize after 2000 but rise again after 2020, mirroring coal consumption during this period.

5.0 Stabilizing United States Greenhouse Gas Emissions

The issue which we consider in this paper is strategies for stabilizing United States emissions of greenhouse gases at 1990 levels by the year 2000. Due to the complexity of comparing the various gases' impact on the earth's radiation balance, we chose to focus our analysis in this paper on the stabilization of fossil fuel carbon emissions; the simplest concept to measure and regulate[3].

Two strategies for stabilizing carbon emissions were analyzed: (1) constant and variable taxes on fossil fuels based on each fuel's carbon content; and (2) variable energy taxes on fossil fuels based on each fuel's energy (BTU) content.

5.1 Carbon Taxes

The imposition of carbon taxes is the most direct price instrument to reduce carbon emissions. A policy of carbon taxes was modeled in the SGM as a tax on the production of fossil fuels applied as an additive tax based on the proportion of each fuel's carbon content (i.e., $ per ton of carbon (TC)). Carbon taxes affect carbon emissions and economic factors by raising the price of fossil fuel by the amount of the additive tax. Based on the known carbon content of various fuels, Table 10 shows how a $135/TC carbon tax (the lowest constant tax rate which would achieve carbon emissions stabilization) is translated into price increases for each of these fuels. The price increase is largest for coal since coal contains the largest concentration of carbon per 10^6 Btu.

[2] We note that the dominant non-energy related emissions of carbon are from ruminant livestock, and are given off in the form of CH_4.

[3] The concept of a global warming potential (GWP) was developed by the Intergovernmental Panel on Climate Change (IPCC) and elsewhere, to provide a metric by which to compare the various gases. But problems in defining these coefficients have emerged which have been most acute with regard to those gases with the greatest indirect effects on climate change; precisely those gases for which the GWP is most important.

Restricting the concept of greenhouse gas emissions narrowly implies that some potentially cost effective measures to reduce net carbon emissions to the atmosphere might be missed, implying higher than necessary costs for achieving any emissions reduction objective.

Table 10: Taxes Based on Carbon Content of Fuels

Fuel	$135/TC Carbon Tax
Crude Oil ($/barrel) Crude Oil ($/10^6 Btu)	$24.64 $2.84
Crude Natural Gas ($/1000 ft^3) Crude Natural Gas ($/$10^6$ Btu)	$2.03 $1.96
Coal ($/ton) Coal ($/10^6 Btu)	$75.60 $3.38
Gasoline ($/gal)	$0.35

Carbon Emissions

The SGM was used to determine the carbon tax rate required in each period to offset carbon emissions above the 1990 emissions level. These tax rates are shown in Figure 9. Tax rates begin at $43/TC in 1995 and rise steadily until 2015 when a peak tax rate of $140/TC is reached[4]. All revenue received from the carbon tax is recycled back to households as additions to personal income. Higher carbon tax rates after 2015 are not required due to a fall in the ratio of carbon (in TgC) to real GNP, implying a shift to more energy-efficient methods of production and less carbon-intensive fuel use.

Figure 10 compares the level of carbon emissions under a "business-as-usual" scenario with various "flat" carbon tax and "stabilization" carbon tax options. Assuming a carbon emissions reduction goal to maintain carbon emissions below the 1990 level, the SGM shows that a flat carbon tax of no less than $135/TC is required.

Impacts on GNP and Consumption

Figure 10 also shows that the $135/TC flat tax case results in an exceedance of the emissions reductions required in the years 1990 to 2030 to stabilize emissions. This will result in greater economic losses than the "stabilization" tax case. Figures 11 and 12 show the percent loss in GNP and consumption, respectively, resulting from a $135/TC "flat" carbon tax and "stabilization" carbon tax. As expected, the $135/TC tax case results in much lower growth rates than the "stabilization" case in the beginning years but converges with the "stabilization" case in later years.

It is interesting to note that the peak variable tax rate ($140/TC in the year 2015) is higher than the constant tax rate ($135/TC). This is attributable to the fact that the constant tax rate of $135/TC is built into earlier investment decisions (by way of expected prices) while the tax trajectory is not know in the case of the variable tax, thus not allowing producers to make the most optimal investment decisions. However, the difference between the two tax rates ($135/TC and $140/TC) is small, implying that the additional knowledge of future tax rates makes small differences in investment decisions.

The "flat" and "stabilization" carbon tax cases result in similar impacts, differing only in the size of the impacts; therefore, it would suffice to concentrate on the impacts of one. Since the "stabilization" tax case achieves stabilization of carbon emissions with the least economic cost, any future comparisons will concentrate on this carbon tax case.

[4] Using a 1985 price for crude oil, natural gas, and coal of $4.15, $2.26, and $1.15 per MBTU, respectively (Annual Energy Review, 1991), a $43/TC carbon tax would amount to an increase in price ranging from 22 percent for oil to 93 percent for coal while a $140/TC carbon tax would amount to an increase ranging from 71 percent for oil to 304 percent for coal.

Impacts on Primary Energy Consumption

Results from the SGM show a dual effect on primary energy consumption resulting from carbon taxes - a reduction in total primary energy consumption and a shift to lower carbon emitting fuels. Reductions in total primary energy consumption range from 8 percent in 1995 to 19 percent in 2030, while the share of coal consumed (out of all primary fuels) falls from 26 percent to 18 percent in 2030. This would indicate that the carbon taxes had a greater effect on reducing overall energy consumption than shifting energy consumption from higher carbon emitting fuels to lower ones. The Electricity Sector experiences similar impacts with the reduction in total fuel consumption outweighing the shift from more carbon-intensive fuels to less carbon-intensive ones. It is interesting to note, however, that although electricity fuel consumption drops by 28 percent in 2030, total electricity generated only drops by 9 percent. This can be attributed to a shift to less carbon intensive fuels, which are associated with more energy efficient methods of generation.

The variable carbon tax also has effect of making biomass fuels economically attractive in the later years. In the reference case, no biomass fuel production occurs throughout the modeling period 1990 to 2030. With the implementation of carbon taxes, we see both liquid and gaseous fuels from biomass being produced. By 2025, gaseous fuels from biomass make up 17 percent of total gas produced, while liquid fuels from biomass only make up 0.4% of total refined liquids produced.

Impacts on Energy Resources

Further evidence that carbon taxes have a greater impact on energy efficiency than on substitution of less carbon-intensive fuels can be seen in the depletion of energy resources. Instead of seeing a more rapid depletion of less carbon-intensive fuels (e.g., natural gas) than the reference case, production of these fuels falls with the imposition of carbon taxes, leading to less depletion of these resources. In addition, as would be expected, production of the more carbon intensive fuels is lower with the imposition of carbon taxes, leading to less depletion of these resources also.

Ratios

A shift to more energy efficient modes of production in reaction to carbon taxes is also apparent when we look at the ratio of energy use (in EJ) to GNP (in trillion $). Differences from the reference case range from 8.1 percent in 1995 to 18.3 percent in 2030. To a much lesser extent, the fossil fuel use (in EJ) to energy use (in EJ) ratio shows a shift to less carbon-intensive fossil fuels with a drop in this ratio of 2.1 percent in 2030 from the reference case, proving that carbon emissions reductions were obtained mainly through energy efficiency measures. Lastly, since population projections do not differ between cases, the fall in the ratio of GNP to population resulting from carbon taxes mirrors the loss in GNP. This loss ranges from 0.4 percent in 1995 to 0.7 percent in 2030.

5.2 Energy Tax

Another policy option which has the effect of reducing overall carbon emissions is a tax on energy use. The Clinton Administration, as part of its budget proposal, suggests a tax based on the BTU content of the various energy sources (i.e., all fossil fuels, and nuclear and hydro electric power generation). Points of tax collection would be at the minemouth for coal, refinery for crude oil, pipelines for natural gas, utilities for hydro and nuclear power, and points of importation for imported electricity and petroleum products. Nonconventional energy sources such as solar and wind would be excluded from the tax. Clinton's BTU energy tax calls for a basic rate of 25.7 cents per million BTUs with a supplemental oil tax of 34.2 cents per million BTUs (amounting to a 59.9 cents per million BTUs tax on oil). This energy tax is equivalent to $28.50, $17.70, and $10.10 per tonne of carbon for oil, gas and coal, respectively[5]. In comparison with the carbon taxes discussed above, the Clinton Administration's energy tax is relatively small. SGM results show small impacts by 2030 as a result of the proposed Clinton energy

[5]The conversion of an energy tax to a carbon tax was done by dividing the $/BTU energy tax by the amount of carbon per BTU in each fuel. The amount of carbon per BTU is highest for coal and lowest for natural gas; therefore, given a basic energy tax rate of $0.257/BTU across all fuels, coal has the lowest $/TC carbon tax rate. Oil has the highest carbon tax rate since it includes a $0.342/BTU supplemental tax.

tax with a 0.18 percent decline in GNP, a 3 percent decline in primary energy consumption, and a 3 percent fall in total carbon emissions.

In this modeling exercise, we impose the general structure of Clinton's energy tax on a policy to stabilize carbon emissions to the 1990 level. That is, the basic rate of 25.7 cents per million BTUs is increased until carbon emissions are stabilized in each period at the 1990 emissions level. (The supplemental oil tax is also increased in the same proportion as the basic rate). In the SGM, the energy tax is imposed on the refined oil, refined gas, coal and uranium sectors, and the hydro electric power subsector beginning in 1995. As in the carbon tax stabilization case, all tax revenue from the energy tax is recycled to the household sector as additions to personal income.

Carbon Emissions

As in the variable carbon tax case, the SGM was used to determine the energy tax rate required in each period to offset carbon emissions above the 1990 emissions level. These tax rates are shown in Figure 13. The energy tax rate trajectory follows a similar path to that of the carbon tax with rates rising steadily from 1995 to 2015 when the peak rate is reached. Energy tax rates begin at $1.03/MBTU in 1995 and rise steadily until 2015 when a peak tax rate of $3.53/MBTU is reached[6].

Figure 14 shows total tax revenue from both stabilization cases (carbon tax versus energy tax). As is expected, total tax required to stabilize carbon emissions is lower in each period with a carbon tax policy option. From a tax standpoint, carbon taxes are more effective at reducing carbon emissions since this tax targets the carbon content of fuels.

Figure 15 compares the stabilization carbon tax rates and energy tax rates (converted to $/TC) applied to the coal sector. We see a larger impact from a carbon tax on this sector in the years 1995 to 2015 because this tax has the effect of reducing use of the more carbon-intensive fuels first; notably coal. As coal is driven out in the beginning periods, carbon emission reductions in later years is obtained by reductions in other fuels. An energy tax, on the other hand, has the effect of reducing overall energy use and, due to the supplemental tax on oil, reducing refined oil consumption. Carbon emission reductions are achieved in each period by targeting reductions in the use of each energy source with a emphasis on refined oil. In Figure 15, carbon taxes have a greater impact on the coal sector in the early years since reductions in coal use is targeted with this policy option while an energy tax does not specifically target use of this fuel. In the later years, carbon emission reductions are still obtainable from the coal sector with an energy tax since this policy option did not drive out coal use in the beginning years like a carbon tax. Therefore, we see a shift where a energy tax has a larger impact than a carbon tax on the coal sector in the later years.

Impacts on GNP, Consumption, and Investment

Due to the higher total tax imposed on the economy with a stabilization energy tax, the impacts on GNP and consumption are slightly greater with this tax option than with a carbon tax as shown in Figure 16. Reductions in GNP from a stabilization energy tax range from 0.6 percent in 1995 to 1.7 percent in 2015 while reductions from a stabilization carbon tax range from 0.4 percent in 1995 to 1.2 percent in 2015. Reductions in consumption follow a similar track with reductions ranging from 0.1 percent in 1995 to 1.1 percent in 2015 with an energy tax while a carbon tax results in reductions of 0.02 percent in 1995 to 0.8 percent in 2015.

The energy tax has a greater impact on investment as compared to the carbon tax case especially in later years. Reductions in 1995 investment are 1.28 percent with an energy tax and 1.02 percent with a carbon tax while reductions in 2020 are 2.22 percent with an energy tax and 0.97 percent with a carbon tax. Since the model allows for shifts in capital inputs only in the long-term (the "putty-clay" assumption), it makes sense that investment under a carbon tax would shift to less carbon-intensive energy production. With the energy tax option, however, all energy sources are targeted (with emphasis on refined oil); therefore, the effects of this tax would be to decrease overall investment in energy production.

[6]Assuming a 1985 price paid by electric utilities for refined oil, gas, and coal of $4.35, $3.43, and $1.65 per MBTU, respectively (Annual Energy Review, 1991), an energy tax of $1.03/MBTU amounts to an increase in price ranging from 24 percent for oil to 62 percent for coal while an energy tax of $3.53/MBTU amounts to an increase in price ranging from 81 percent for oil to 214 percent for coal.

Impacts on Primary Energy Consumption

Figure 17 shows the impacts of both stabilization tax options on primary energy consumption. Both taxes result in similar impacts on energy consumption with the energy tax resulting in slightly larger impacts. Since an energy tax effects all sources of energy, we would expect this policy option to result in lower overall energy use when compared to a carbon tax policy option.

With both tax cases, we see a shift in electricity fuel consumption from oil and coal to natural gas and renewables. This is because the percent increase in price of oil and coal is greater than that of gas in both tax cases causing as shift to gas consumption[7]. The difference between the two cases is out of which fuel most of the shifting is taking place. Due to the supplemental tax on refined oil with the energy tax, we see most of the decline in electricity fuel consumption coming from refined oil; in fact, refined oil use in the electricity sector is completely phased out by 2000 under an energy tax whereas, under a carbon tax, oil is not phased out until after 2025. With a carbon tax, most of the decline in electricity fuel consumption comes from coal. In either case, the amount of electricity generated is essentially the same implying that each tax option has the same effect on electricity generation.

As with carbon taxes, we see the emergence of gaseous fuels from biomass with an energy tax with gas from biomass making up 14 percent (compared to 17 percent with the carbon tax) of gas production in 2025. No production of liquids from biomass occurs since the energy tax does not distinguish between feedstocks of the refined liquids in its supplemental tax on this sector. (We see a larger amount of the biomass fuels being produced with the carbon tax since these fuels are exempt from any carbon tax; unlike in the energy tax case where all sources of energy are affected).

Impacts on Energy Resources

In both tax cases, we see a faster depletion of natural gas than the reference case. This is due to a shift from oil and coal use in both cases to a more energy-efficient and less carbon-intensive use of energy; namely natural gas. As explained above, this shift is due to the larger percent increases in price of oil and coal than gas in both tax cases.

Ratios

As with a carbon tax, an energy tax results in more energy-efficient and less carbon-intensive energy use. As would be expected, since an energy tax targets all energy sources and does not target those which are more carbon-intensive, we see a slightly faster decline in the ratio of energy to GNP with an energy tax than with a carbon tax. In addition, we see a slightly slower decline in the ratio of carbon to energy with an energy tax than with a carbon tax for the same reason. Lastly, due to the larger reductions in GNP with an energy tax, we see a slightly smaller decrease in the ratio of GNP to population with a energy tax than with a carbon tax.

6.0 Conclusions

The purpose of this paper was to demonstrate some of the capabilities of the SGM and to specifically show its effectiveness as a long-term model of energy use, the economy, and greenhouse gas emissions. The model was designed to assess global emissions from all human activities and the major direct and indirect consequences of potential policies to reduce emissions, capabilities that no currently existing model possesses. As obvious from the model design, the SGM is capable of a wide array of modeling exercises which have not been presented in this paper. Examples include (1) various tax revenue recycling options; (2) reforestation as a policy option;

[7]Although the reasons why oil and coal experience a larger increase in price than gas under a carbon tax is straightforward (since the tax is based on carbon content, the more carbon-intensive fuels will face a larger tax), the reasons under an energy tax are more elusive. In the baseyear, coal is the cheaper fuel (in $/BTU) as compared to gas. (Conversely, more energy (BTU) per dollar is obtainable with coal). Imposing a constant energy tax ($/BTU) on both coal and gas, we see that the tax increases the price of coal more than gas since the baseyear price of coal is lower. (This can been seen by dividing the $/BTU energy tax by the $/BTU baseyear price of coal). The oil price increase is due to the same reason in addition to a supplemental oil tax.

(3) emissions trading as a policy option; (4) new investment in nuclear energy; (5) investment tax credits; (6) new source compliance; and (7) implementation of new energy technologies and retrofits, to name a few.

Future versions of the SGM will further expand on the capabilities of the model. Future expansions of the model include the provision of energy services and traditional biomass fuels, and the addition of a detailed description of the agriculture and health sectors to enable the model to consider the consequences of atmosphere and climate change in detail.

FIGURE 1

The Flow of Goods and Services in the SGM

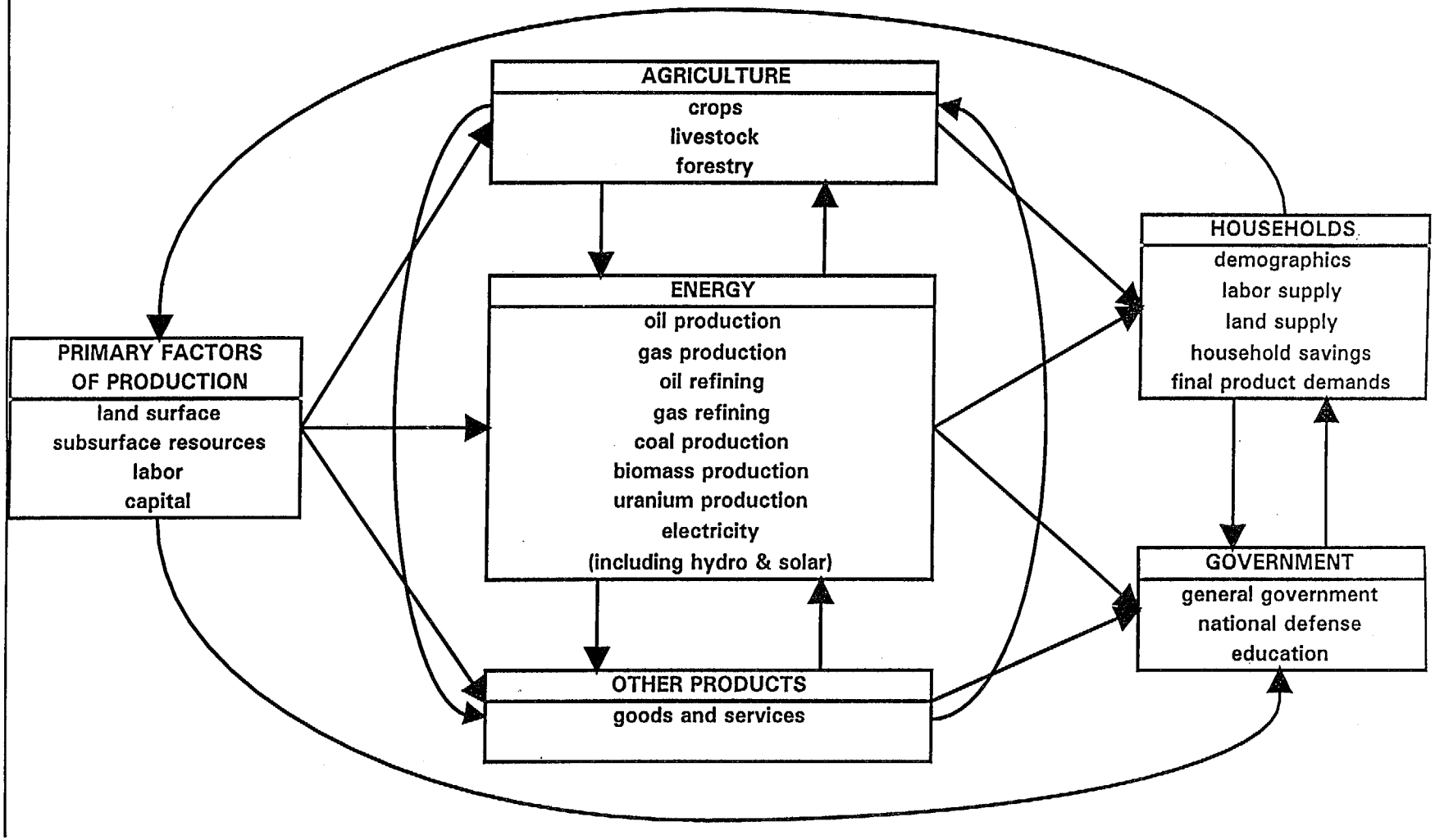


FIGURE 2: 1985 Population

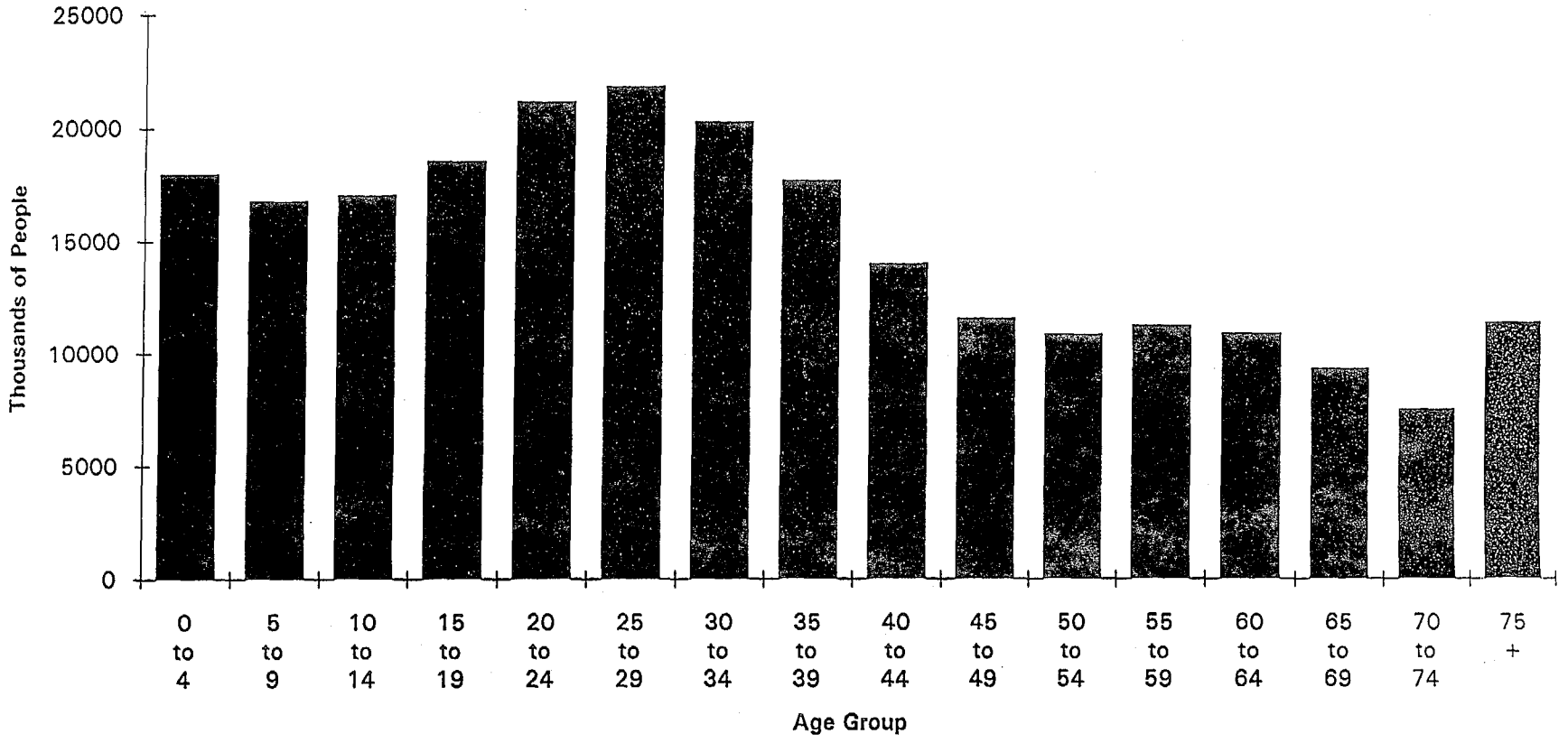


Figure 3: Components of Real GNP (1982$): Historical and Reference Case

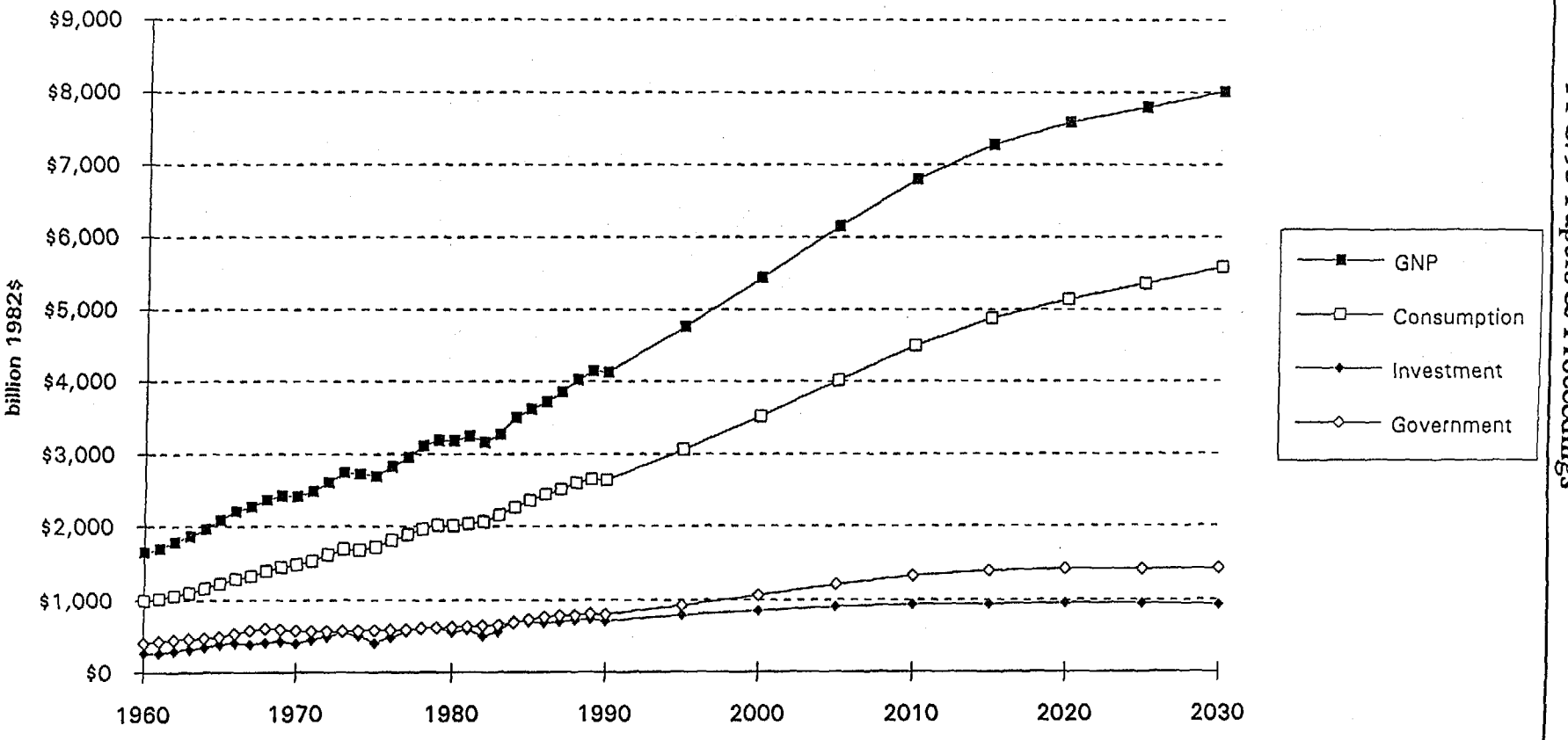


FIGURE 4: Working Age Population

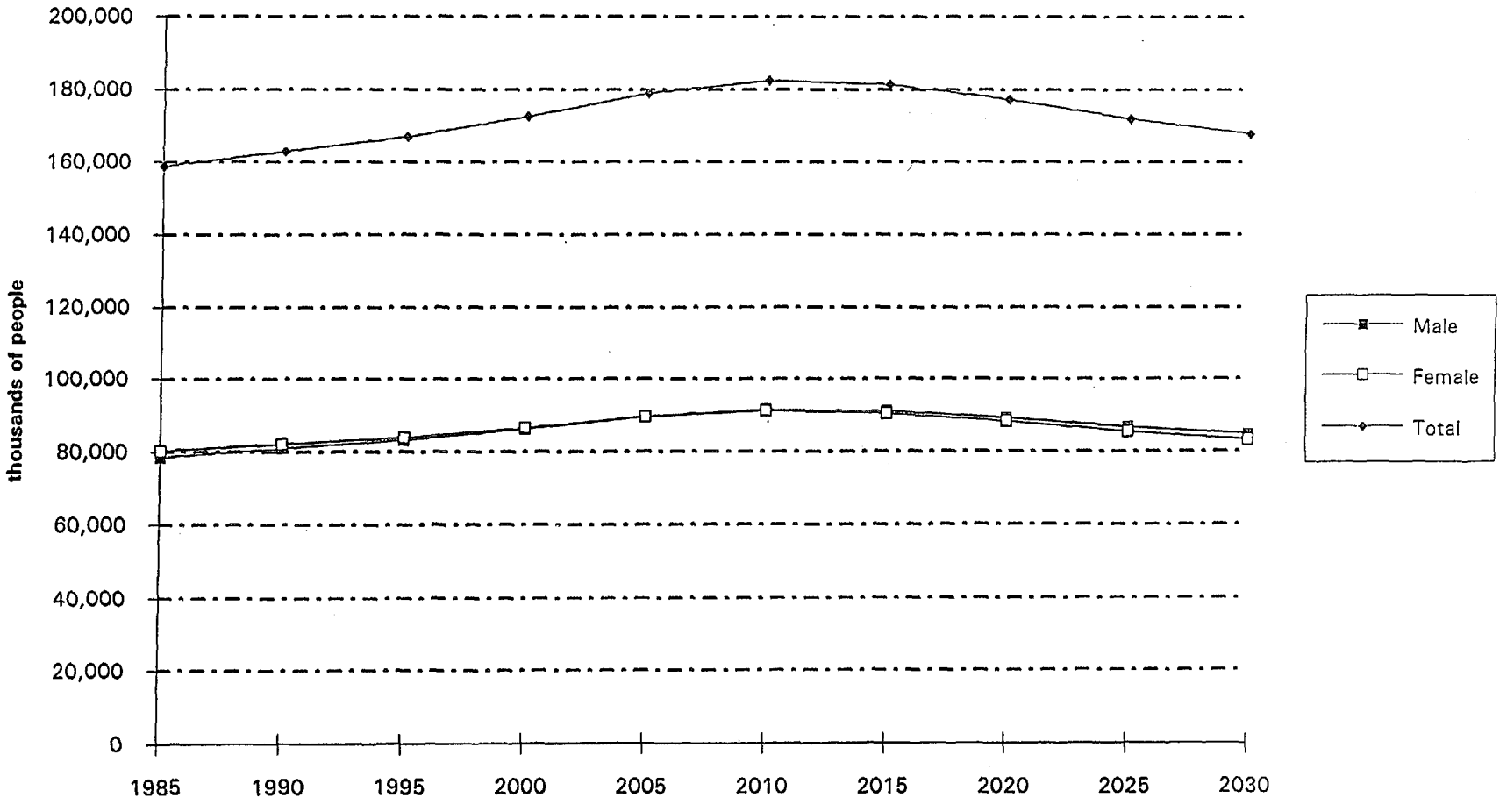


FIGURE 5: Prices of Primary Factors of Production (Base Year Prices)

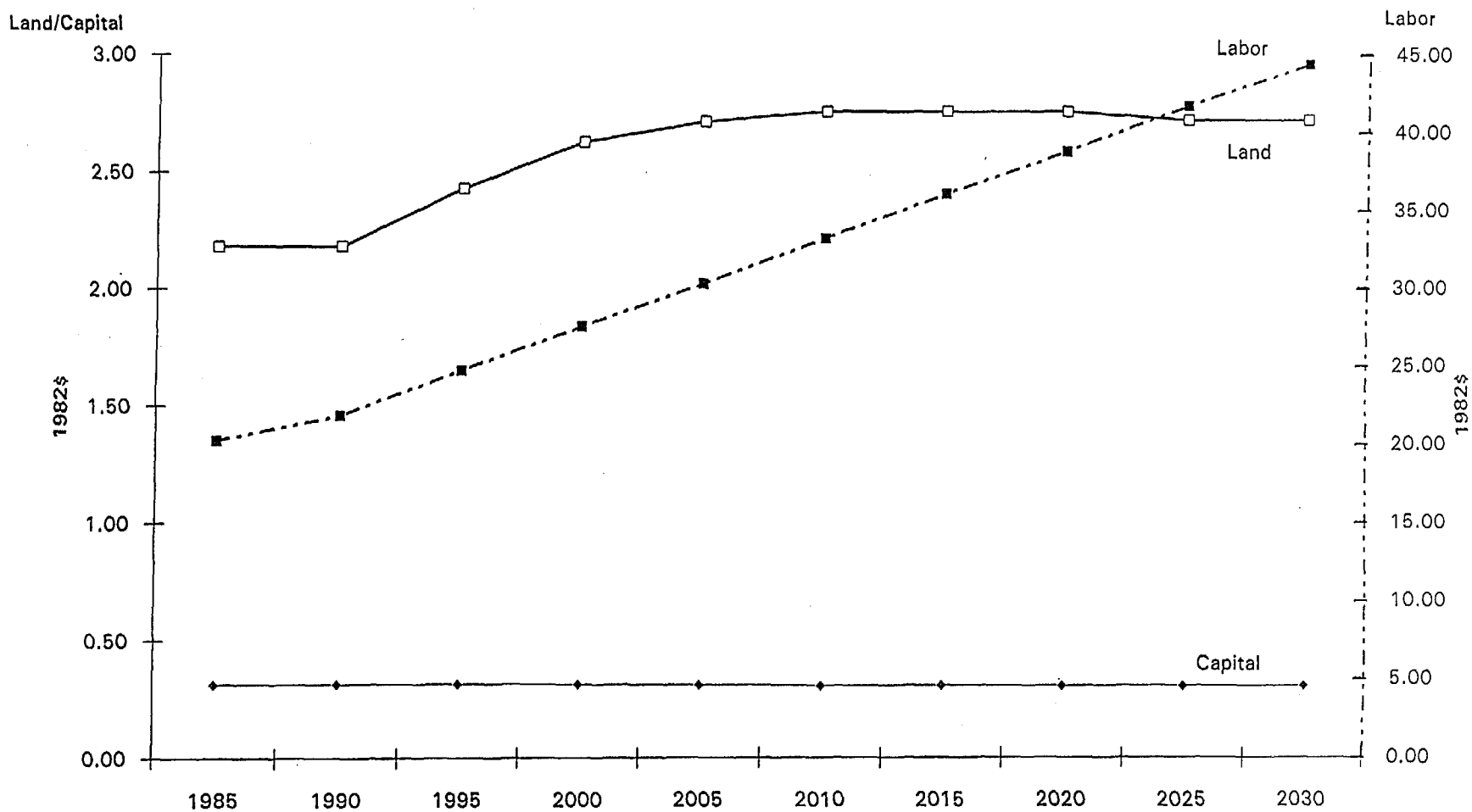


FIGURE 6: Primary Energy Consumption

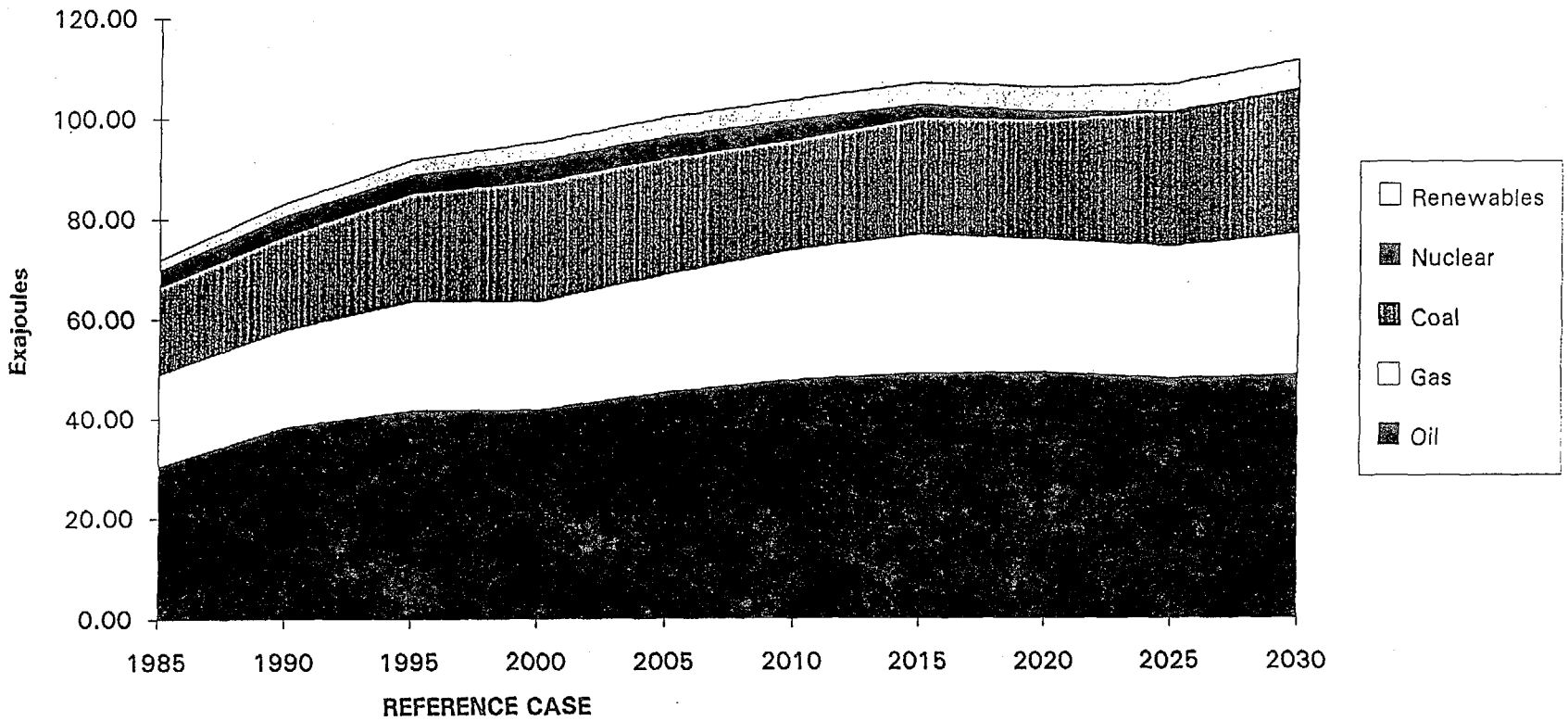


Figure 7: Co2 Emissions (TgC) : Historical and Reference Case

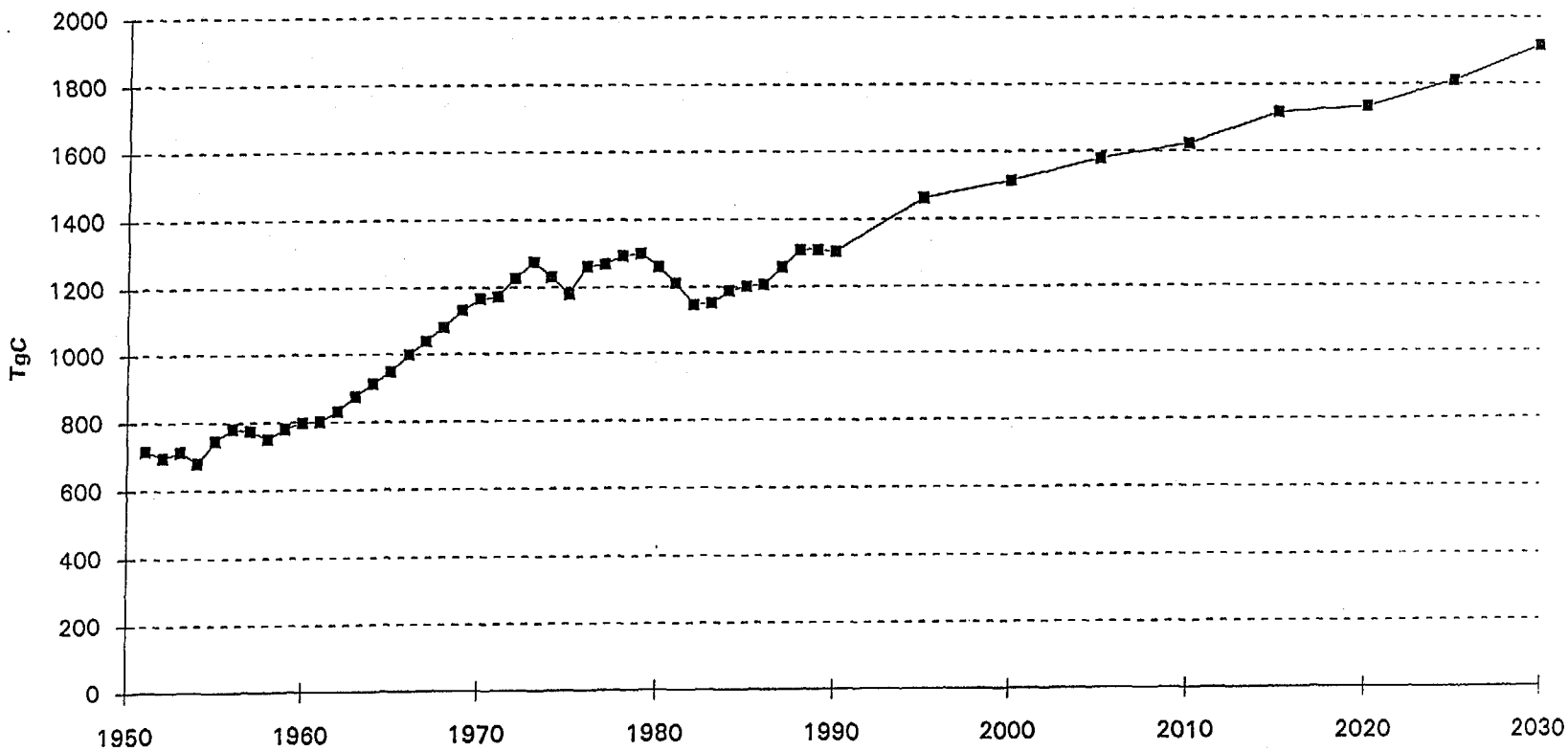


FIGURE 8: Other Greenhouse Gas Emissions

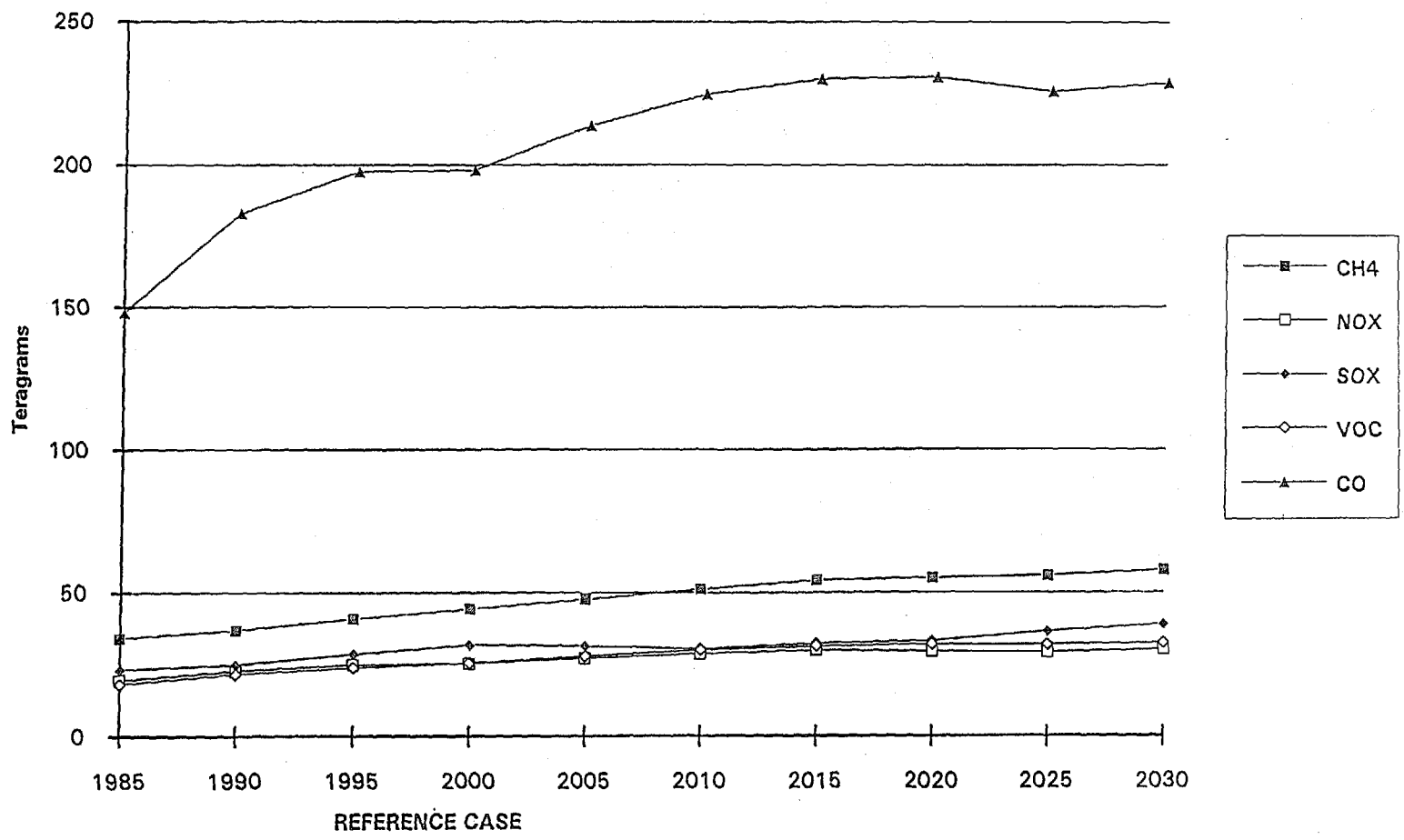


FIGURE 9: Carbon Tax for CO2 Stabilization to 1990 Level

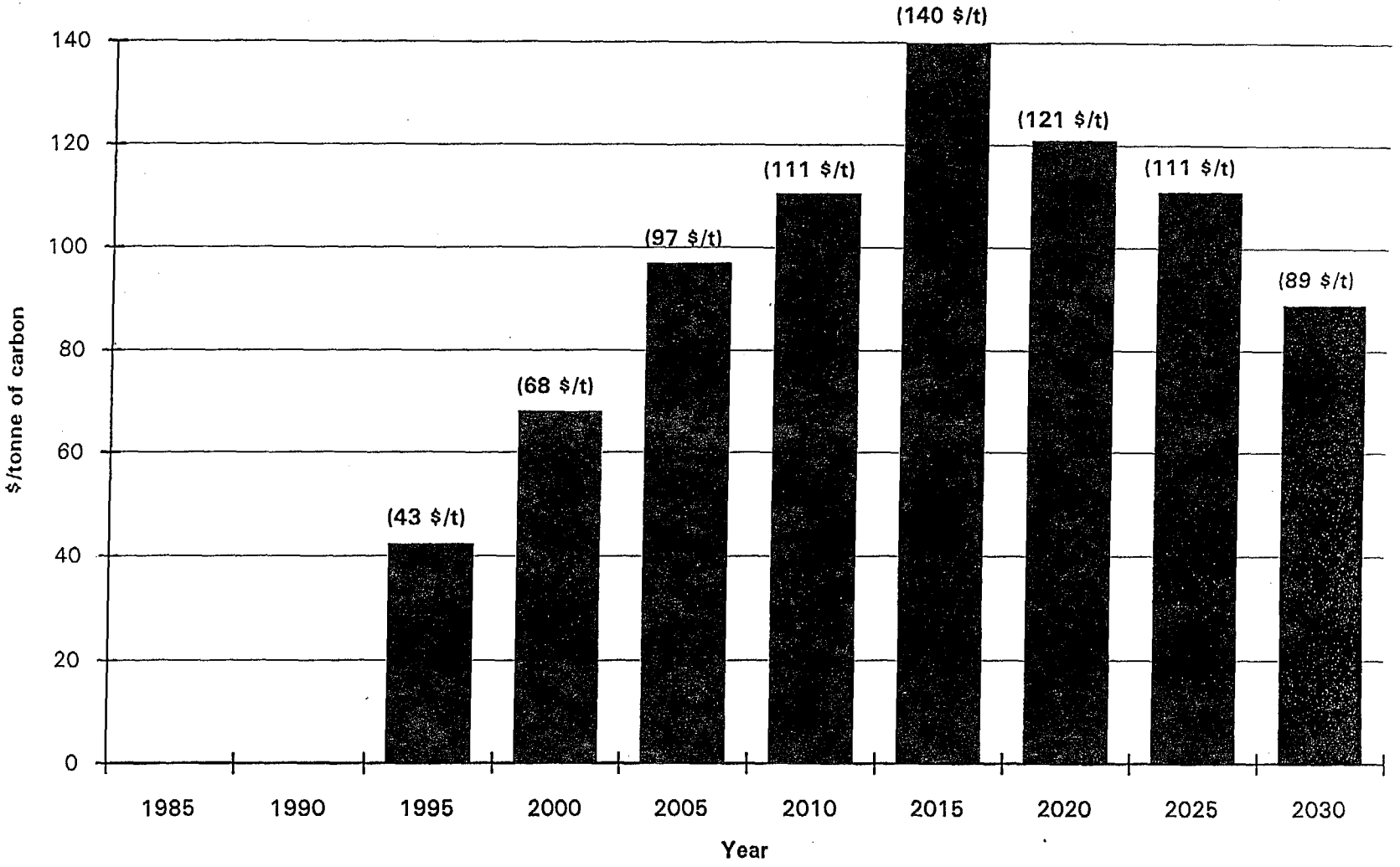


FIGURE 10: Total Carbon Emissions

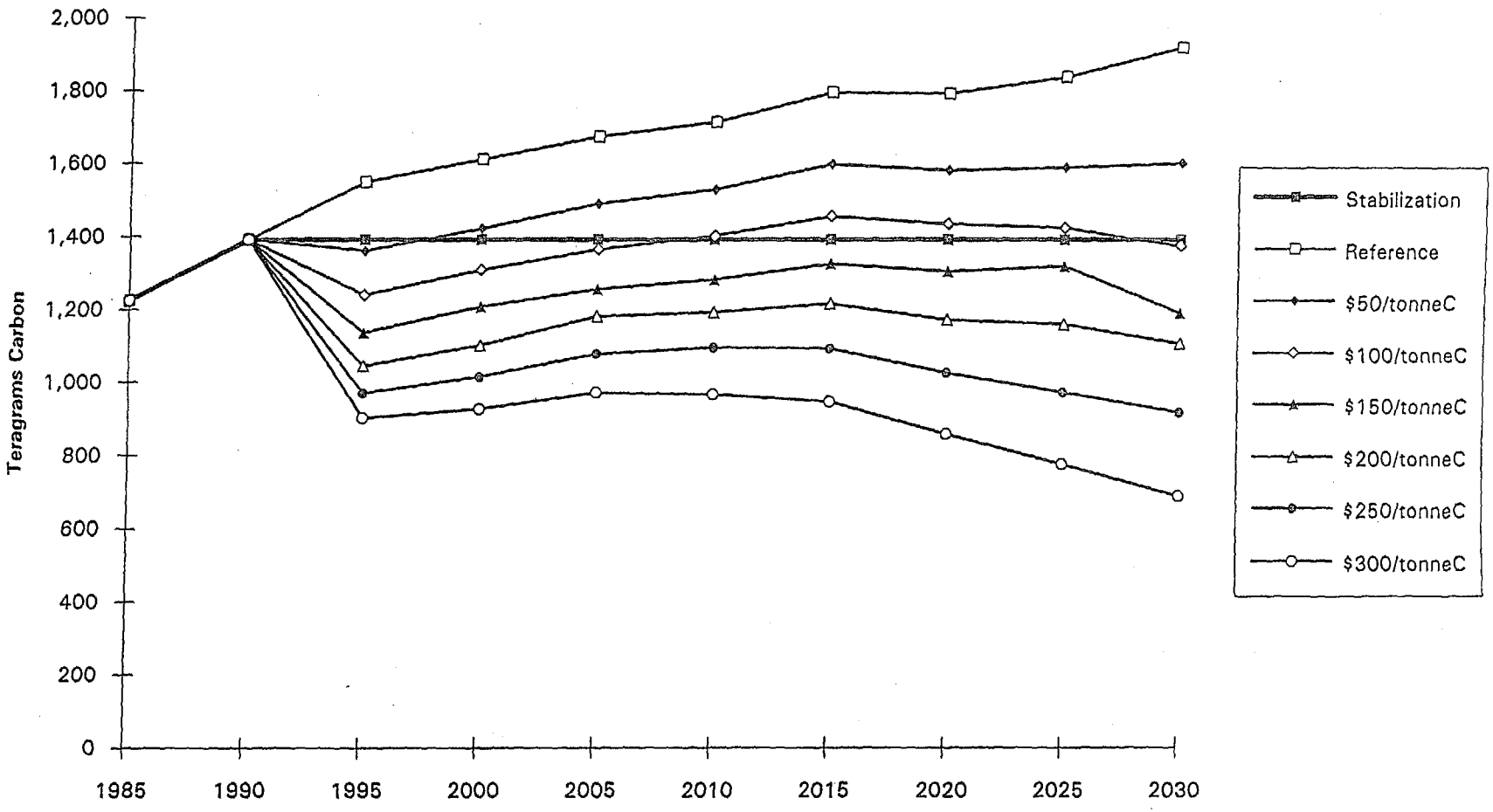


FIGURE 11: % Reduction in Real GNP from Reference Case

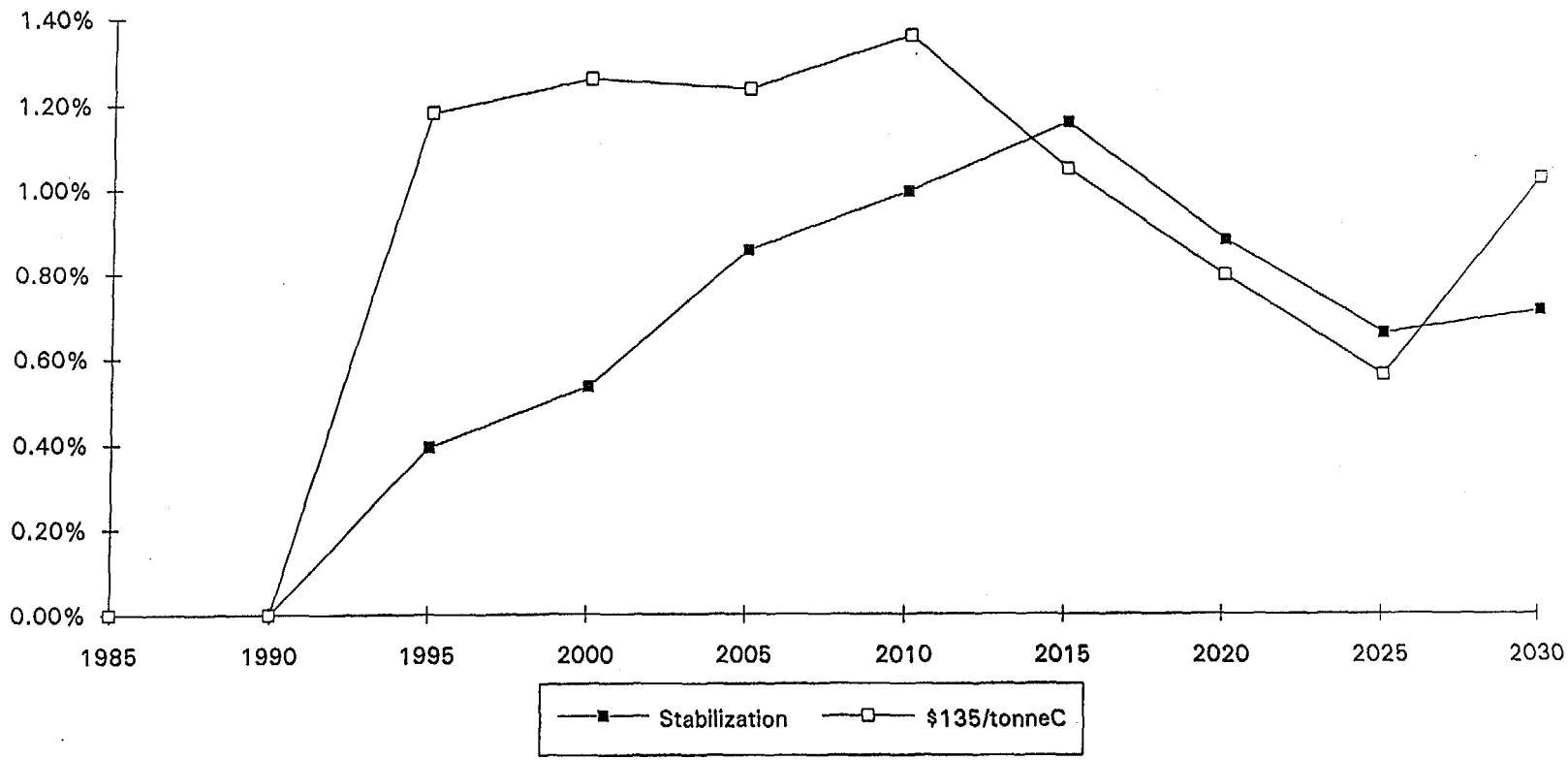


FIGURE 12: % Reduction in Real Consumption from Reference Case

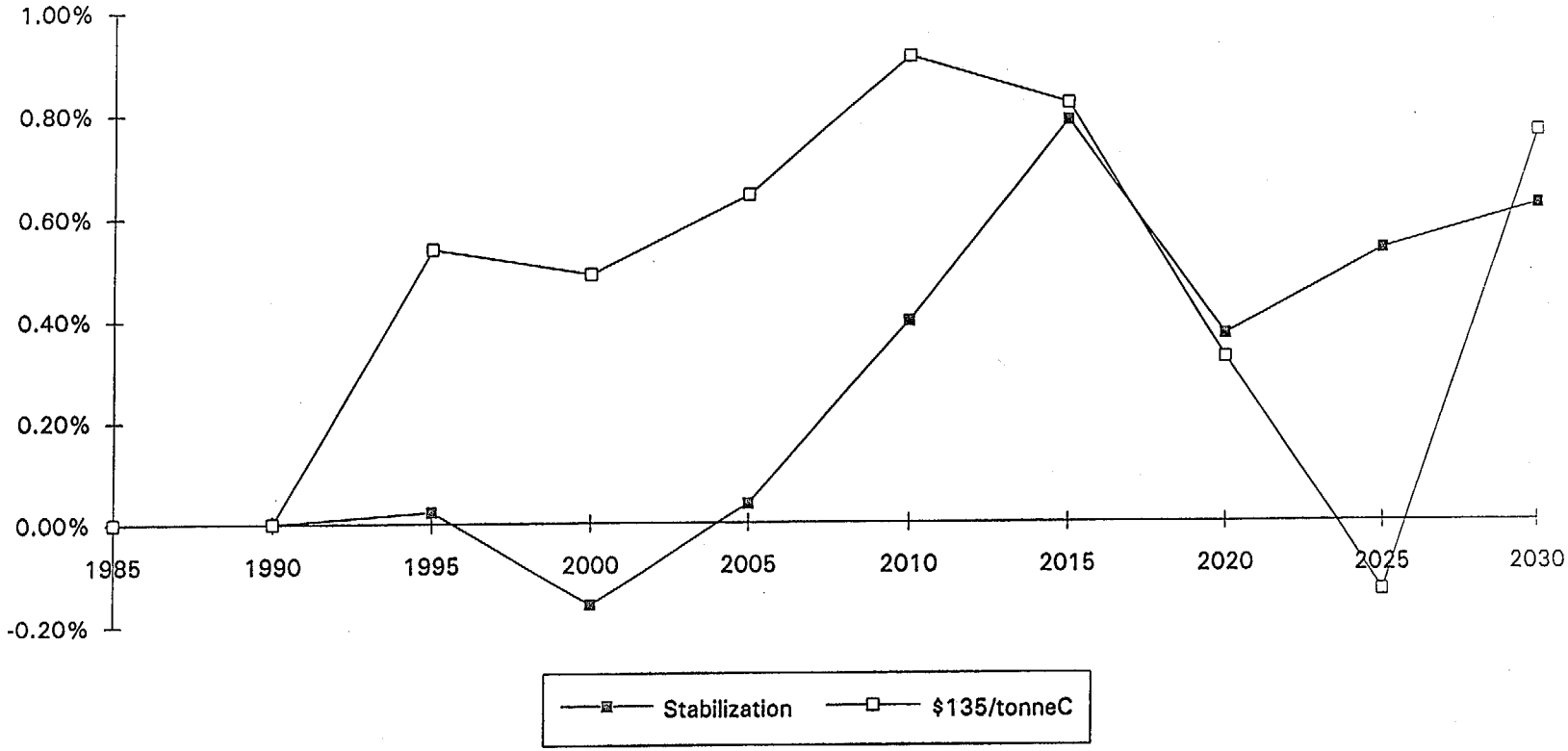


FIGURE 13: Energy Tax for CO2 Stabilization to 1990 Level

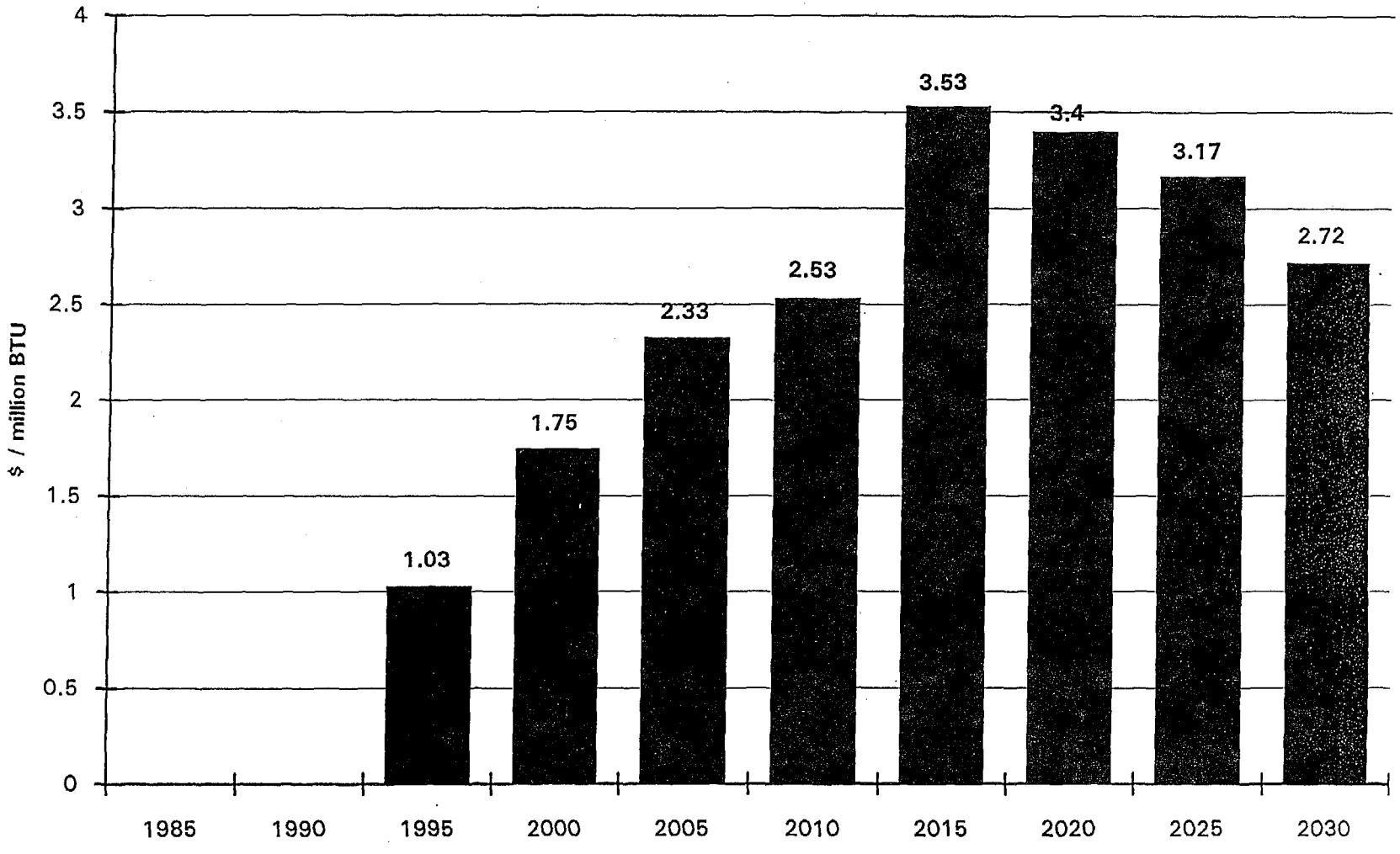


FIGURE 14: Tax Revenues

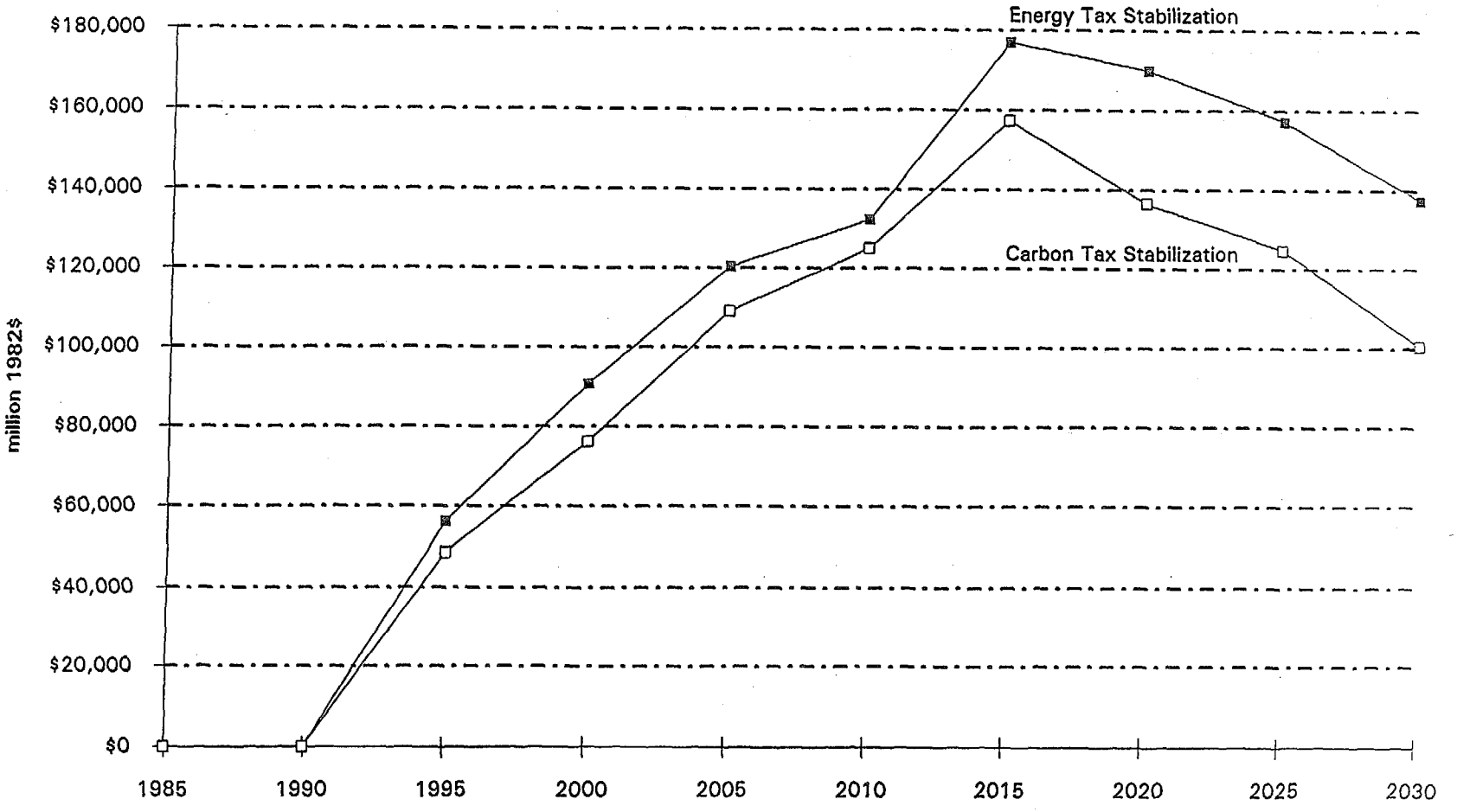


FIGURE 15: Taxes for CO2 Stabilization to 1990 Level (Energy tax converted to carbon tax based on carbon content of coal.)

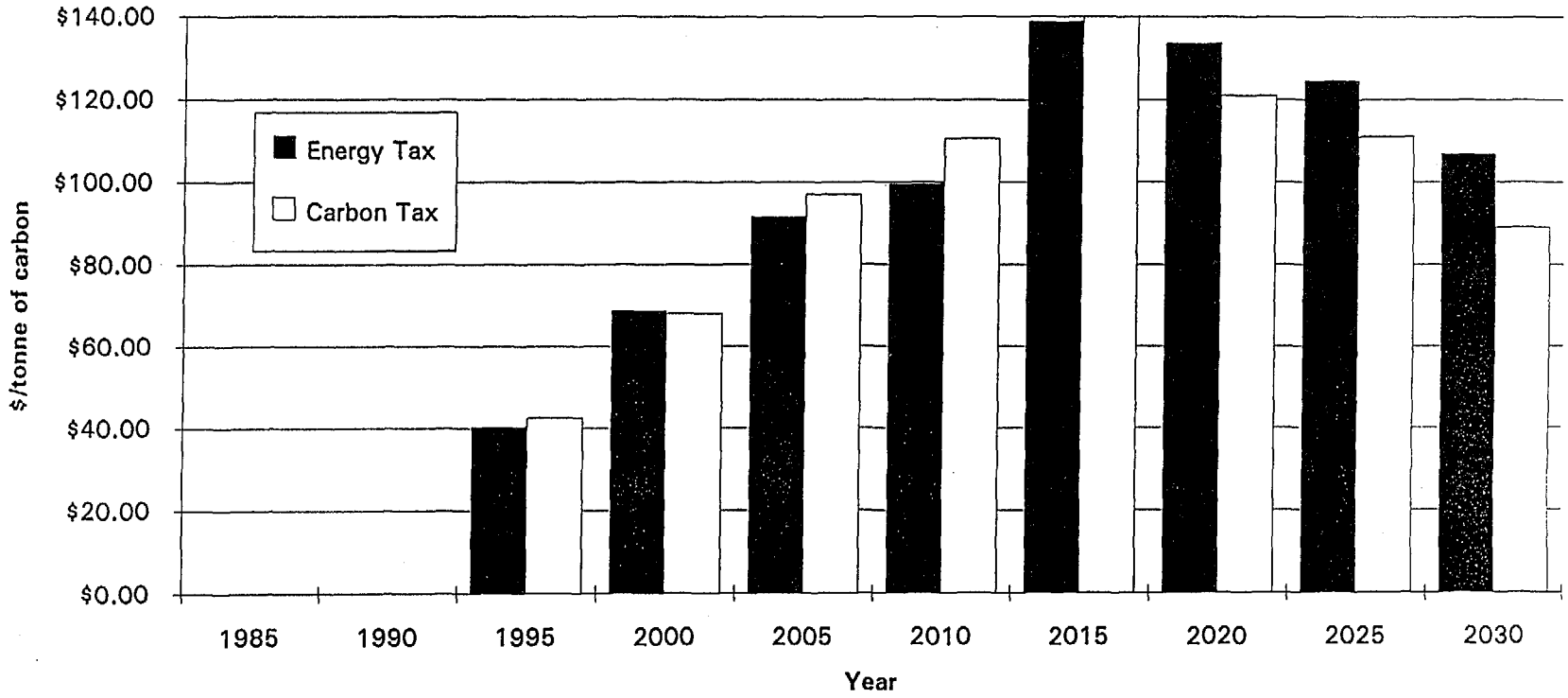


FIGURE 16: % Reduction in GNP from Reference Case (Base Year Prices)

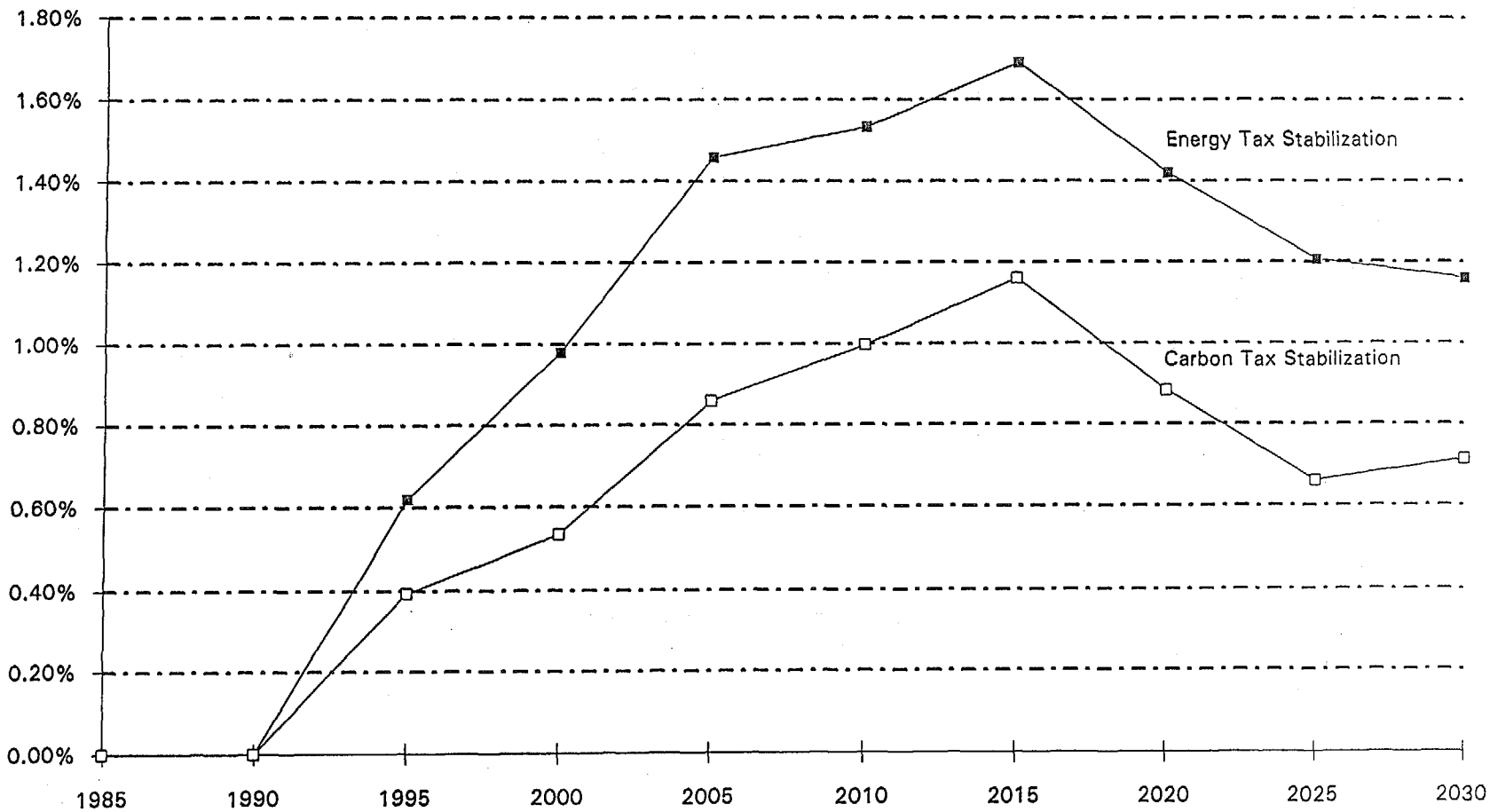
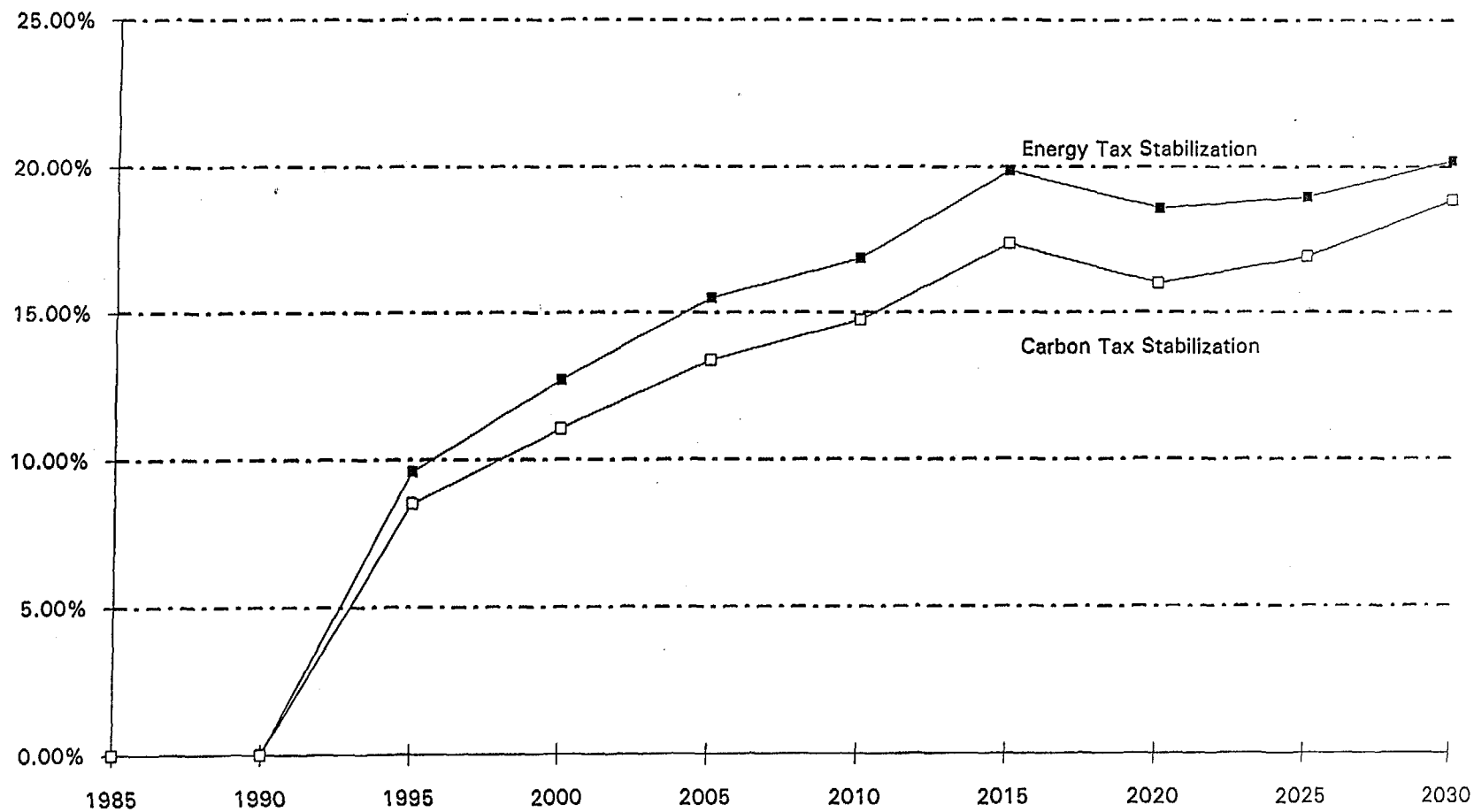


FIGURE 17: % Reduction in Primary Energy Consumption from Reference Case



References

Edmonds, J., and Barns, D.W., 1990. *Estimating the Marginal Cost of Reducing Global Fossil Fuel CO_2 Emissions*. PNL-SA-18361. Pacific Northwest Laboratory. Washington D.C.

Edmonds, J., and Barns, D.W., 1991. *Factors Affecting the Long-term Cost of Global Fossil Fuel CO_2 Emission Reductions*. PNL Global Studies Program Working Paper. Washington D.C.

Edmonds, J., and Reilly, J.M., 1985. *Global Energy: Assessing the Future*. Oxford University Press. New York.

Edmonds, J., Reilly, J.M., Gardner, R.H., and Brenkert, A., 1986. *Uncertainty in Future Global Energy Use and Fossil Fuel CO_2 Emissions 1975 to 2075*. TR036, DO3/NBB-0081Dist. Category UC-11. National Technical Information Service. U.S. Department of Commerce. Springfield, Virginia 22161.

Edmonds, J., H.M. Pitcher, D. Barns, R. Baron, and M.A. Wise. 1993. *Modeling Future Greenhouse Gas Emissions: The Second Generation Model Description*. Proceedings of the United Nations University Conference on "Global Change and Modeling." Tokyo, JAPAN. October, 1991.

Elliott, Douglas C. et al. 1990. "A Technical and Economic Analysis of Direct Biomass Liquefaction," in *Energy from Biomass and Wastes XIII*. Donald L. Klass ed. Chicago: Institute of Gas Technology.

Fisher-Vanden, K. et al. 1993. *Calibration of the Second Generation Model*. Pacific Northwest Laboratory. Washington, D.C.

IPCC (Intergovernmental Panel on Climate Change). 1990.

Jorgenson, Dale W. 1990. "Productivity and Economic Growth" in *Fifty Years of Economic Measurement*, edited by Ernst R. Berndt and Jack E. Triplett. National Bureau of Economic Research, published by the University of Chicago Press.

OECD (Organisation for Economic Co-operation and Development). 1992. International Energy Agency. *Electricity Supply in the OECD*. Paris, France.

U.S. Department of Agriculture, 1992. Conversations with E. Ball.

U.S. Department of Commerce, Bureau of Economic Analysis, 1987. *Fixed Reproducible Tangible Wealth in the United States, 1925-85*. Washington D.C., Government Printing Office.

U.S. Department of Commerce, Bureau of Economic Analysis, 1990. "Annual Input-Output Accounts for the U.S. Economy, 1985" in *Survey of Current Business*. Washington D.C.

U.S. Department of Commerce, Bureau of Economic Analysis, 1991. *Benchmark Input-Output Accounts for the U.S. Economy, 1982*. Washington D.C., Government Printing Office.

U.S. Department of Commerce, Bureau of Census, 1992. *Statistical Abstract of the United States*. Washington, D.C., Government Printing Office.

U.S. Department of Energy (DOE), 1991a. *Limiting Net Greenhouse Gas Emissions*. Report to the Congress. Bradley, R.A., E.C. Watts, and E.R. Williams (eds.). Volumes 1 & 2. DOE/PE-0101. Washington, D.C.

U.S. Department of Energy (DOE), 1991b. *Integrated Analysis Supporting the National Energy Strategy: Methodology, Assumptions and Results, Technical Annex 2*. DOE/S-0086P. National Technical Information Service. Springfield, Virginia.

U.S. Department of Energy, Energy Information Administration (EIA), June 1991. *Financial Statistics of Selected Investor-Owned Electric Utilities, 1989.* DOE/EIA-0437(89). Washington, D.C.

U.S. Department of Energy, Energy Information Administration (EIA), June 1992. *Annual Energy Review 1991.* DOE/EIA-0384(91). Washington, D.C.

Wan, Edward I. and Malcolm D. Fraser, 1990. "Economic Assessment of Advanced Biomass Gasification Systems," in *Energy from Biomass and Wastes XIII*. Donald L. Klass ed. Chicago: Institute of Gas Technology.

World Bank, 1990. *World Tables.* 1989-1990 Edition. Johns Hopkins University Press. Baltimore, New York.

Acknowledgements

The authors would like to express appreciation to the United States Department of Energy, Office of Energy Research and the Office of Policy, Planning, and Analysis, and to the United States Environmental Protection Agency, for support in the conduct of this research.

The Millennium Project and Global Issues of Population and Environment

Jerome C. Glenn, United Nations University

Abstract

This paper describes the United Nations University's Millennium Project, its feasibility study, and the results from a 2-round international Delphi study on population and environment.

Introduction

The pace, complexity, and globalization of change requires attention to the future. The change of millennia - the coming of the year 2001 - provides an opportunity, a psychological focus, for a global review of past achievements and problems and a unique chance to assess and reflect on future issues and opportunities.

Unfortunately, there is as yet no mechanism or information utility to support a worldwide effort to study past achievements and problems or future issues and opportunities. Although there are many individual, isolated, special purpose, and one-time study efforts underway, there is no international system that can provide coherence or continuity to these studies, including feedback and sharing of information, and, in particular, the systematic exploration of future possibilities and policy alternatives. With growing interest in the future, the spread of instantaneous and global communications, the advent of powerful new non-deterministic modeling techniques, the ability to evoke, capture and share information and perceptions with systematic questioning techniques and software, the proliferation of data bases, and knowledge visualization, it is now possible for futurists, scholars and others around the world to interact globally and take a fresh look at the future possibilities and policies in ways not previously possible.

As the World Bank provides an ongoing system for research and feedback to improve economic policy, so too the United Nations University could provide an on-going system for the improvement of futures research and its application to the policy process. According to a series of interviews, questionnaires, and meetings with leading futurists and scholars around the world, the proposed "Millennium Project" has the potential to become such a system.

The Concept of the Millennium Project

The Millennium Project was originally conceived as a major landmark in studies dealing with the future. Although its time horizon was set as the next century, the next two decades were to be given analytic emphasis and the next hundred years would be given more normative emphasis. The project was designed to synthesize judgments on a global basis about future developments and issues in a way that would enrich and illuminate policy discussions. Subsequent discussions have served to sharpen the central goal of the full-scale Millennium Project. In essence it is to:

> Organize a program of futures research and a facility for global research that will continuously update and improve humanity's thinking about the future, and to make that thinking available through feedback systems in a variety of media for consideration in public policy, advanced training, and public education.

The pre-feasibility study identified the following ten objectives for the Millennium Project:

1. Link futurists, scholars, creative thinkers, and institutions around the world

2. Create an international information system of forecasts, key questions, issues, lessons from history, and potential futures research agendas, that allows for feedback to improve the state of research and thought in these areas

3. Evaluate futures research methodology and potential for setting standards

4. Forecast important technological, social, and scientific achievements and their likely effects on the future human condition

5. Integrate forecasts to describe and evaluate, to the extent possible, the likely future conditions of people in the world over the next century

6. Explore possible strategies in coping with these issues - the strategies that relate to actions that might be taken by governments, other institutions, or individuals

7. Increase awareness of the future and its promise, while inculcating the view that the future could be shaped in the interests of society through thoughtful policies

8. Develop a set of educational materials useful at all levels, dealing with the history of achievement of the last century and social issues of the next century, critical choices, and policy options that relate to these issues, and other future-related topics

9. Conduct advanced training

10. Provide a legacy to the people of the next century: what was thought about the future and how we, at the turn of the century had hoped it would evolve.

The Millennium Project Feasibility Study has three phases; this report describes Phase I.

PHASE I, from which this paper is drawn, began in November 1992 with support from the U.S. EPA to identify and link futurists and scholars around the world. Two Delphi inquires were involved: a two round Delphi on the process of organizing the Millennium Project and a two round Delphi on population & environment (which composes most of this paper). The full 151 page report of this phase is available on Internet via Gopher futures.wic.epa.gov or on diskette from the Millennium Project, 4421 Garrison Street, N.W., Washington, DC 20016 (202)-686-5179.

PHASE II began in August 1993 with support from the UNDP/African Futures to produce a set of booklets and computer diskettes on futures methods and issues via an environmental scanning system.
Sixteen methods will be considered for review in Phase II and for presentation in individual booklets; these methods include:

1. Introduction & Overview
2. Participatory Methods
3. Delphi
4. Systems and Modeling
5. Simulation Modelling
6. Decision Modelling
7. Scenario Construction
8. Trend Impact Analysis
9. Technology Sequence Analysis
10. Cross Impact Analysis
11. Futures Wheels
12. Structural Analysis & Actors' Strategies
13. Normative Forecasting
14. Relevance Trees and Morphological Analysis
15. Simulation Gaming as Applied to Forecasting
16. Visioning, Genius Forecasting, Intuitive Approaches

17. Methods Frontiers, Integration, and Strategic Planning

Each methods booklet would have the following sections or chapters: 1) History of the Method; 2) What the technique IS; 3) How to DO IT; 4) Strengths and weaknesses of the method; and 5) Alternative uses of the method AND how to use it in combinations with other methods. An Appendix will also include several actual examples of the method and an annotated bibliography

Phase II will also include identification and analysis of futures issues of importance to Africa. These include:

1. Technological Capacity
2. International Economic Policy and International Trade
3. Agriculture and Food Security Trends
4. Global Life Support Systems and Sustainable Development
5. Population, Education and Human Welfare (special attention to AIDS)
6. Peace, Governance, and Culture

Each issue scan and analysis includes: 1) Historical forces; 2) How these forces are changing today and in the foreseeable future; 3) Alternative scenarios/impacts out to 2025 in Africa - futures chapter; 4) How to measure change on this issue; 5) How this scan was done and how to keep it going; 6) Template for data base ; 7) Key sources, individuals, institutions, computer network groups, etc. plus an appendix of relevant forecasts for The Futures Group scan in hard copy and on diskette

PHASE III will consist of a consultative meeting to focus on recommendations for the feasibility study report. The location and design of such a meeting will depend on the responses to this report and the results of Phase II. Use of advanced telecommunications systems could augment the proposed face-to-face consultative meeting. The primary product of this phase will be the final report on the feasibility of the Millennium Project.

As described in the full Phase I report, two 2-round Delphi inquiries were conducted to gather judgments from the international panel consisting of futurists, scholars, and interested participants. The first of these two-round Delphis was concerned with the process of organizing the full-scale Millennium Project. The second two-round Delphi focused on the leading issues identified by the international panel in the first Delphi: population and environmental change. The conclusions of this second Delphi study is below.

Results of Population and Environment Delphi Rounds I and II

In the first round, the panelists were asked to rate some forces that led to the reduction of the world population growth rate from 2.06% in the late 1960s to 1.7% currently and to assess how these forces might change over the next 25 years. They were invited to add forces, which were rated in the second round by the same Scale A below. Table 1 orders these forces by their historic influence; table 2 orders them by their future influence.

Scale A

Historic Influence	*Future Influence*
1 = Very Important	1 = Greatly increasing in importance
2 = Important	2 = Increasing in importance
3 = Marginally important	3 = Remaining the same in importance
4 = Unimportant	4 = Decreasing in importance
5 = Counter impact	5 = No longer a factor, or mixed

Table 1 - The importance of some HISTORIC factors on global population growth

2.0 Availability of inexpensive, simple effective contraceptives;
2.0 Family planning and public health programs
2.0 China's population policy
2.0 Legitimization of contraception

2.0 Decreases in infant and child mortality
2.1 Government policies in developing countries that encourage smaller families
2.1 Increasing number of years that women attend school
2.1 Rising incomes and the spread of middle class values
2.2 Move away from agricultural society/primary sector
2.3 Spread of new communications media (Television, etc)
2.3 War, famine, disease, and pestilence
2.4 Improved literacy by improved children's schooling
2.4 Decreasing Catholic Church's social influence
2.5 Rise of "woman's power."
2.9 Availability of male contraceptives
2.9 Environmental deterioration
3.1 Education about relation of environment and population
3.1 Increasing futuristic orientation
3.2 Demonstration by the North that fewer children can mean more wealth

Table 2 - The importance of some forces affecting population over the NEXT 25 years

1.8 Spread of new communications media (Television, etc)
2.0 Increasing number of years that women attend school
2.0 Environmental deterioration
2.1 Government policies in developing countries that encourage smaller families
2.1 Family planning and public health programs
2.1 Improved literacy by improved children's schooling
2.2 Availability of inexpensive, simple effective contraceptives
2.2 Legitimization of contraception
2.2 Education about relation of environment and population
2.2 Decreases in infant and child mortality
2.3 Rising incomes and the spread of middle class values
2.3 Availability of male contraceptives
2.4 Rise of "woman's power"
2.4 Increasing futuristic orientation
2.6 Move away from agricultural society/primary sector
2.6 Decreasing Catholic Church's social influence
2.7 War, famine, disease, and pestilence
2.8 China's population policy
2.8 Demonstration by the North that fewer children can mean more wealth

In the first round, the panelists were asked to rate some forces the could explain why populations rates have remained relatively high in the developing countries and to assess how these forces might change over the next 25 years. They were invited to add forces, which were rated in the second round by the same Scale A above. Table 3 orders these forces by their historic influence; table 4 orders them by their future influence.

Table 3 - Some reasons for high population growth in developing countries

1.6 Need of children for social security, to support parents in their old age.
1.7 Discrimination against women (little autonomy, education, and lack of social power)
1.8 Poverty
1.8 Low levels of literacy and lack of understanding ecological view of planet earth
2.0 Family-based, labor-intensive economies needing children's input
2.0 Low levels of literacy
2.1 Close adherence to religious tenets that lead to avoidance of contraceptives
2.1 Rural areas receive less attention from population programs than urban areas

2.2 Lack of information and access to contraceptives
2.2 High infant mortality
2.3 Masculinity associated with increasing numbers of children
2.4 Belief by governments that larger populations mean greater political strength
2.8 Government policies supporting large families
3.1 Liberal immigration policies in richer countries

Table 4 - How important will these forces be over the NEXT 25 years?
2.2 Low levels of literacy and lack of understanding ecological view of planet earth
2.4 Poverty
2.6 Rural areas receive less attention from population programs than urban areas
2.7 Low levels of literacy
2.8 Discrimination against women (little autonomy, education, and lack of social power)
2.8 Lack of information and access to contraceptives
3.0 Need of children for social security, to support parents in their old age.
3.1 Family-based, labor-intensive economies needing children's input
3.1 High infant mortality
3.2 Belief by governments that larger populations mean greater political strength
3.4 Close adherence to religious tenets that lead to avoidance of contraceptives
3.4 Government policies supporting large families
3.4 Liberal immigration policies in richer countries
3.5 Masculinity associated with increasing numbers of children

The panelists were asked to assess new forces and unprecedented events that might influence population growth in the future and suggest and assess additional such forces as to their likelihood of occurrence and impacts over the next 25 years using Scale B. **Table 5** shows the average of the panel's responses about their likelihood; **Table 6** shows their impacts over the next 25 years.

Scale B

Likelihood of occurrence	*Eventual impact on population within the next 25 years of growth*
1 = almost certain	1 = reduces growth rate by 30% or more
2 = likely	2 = reduces growth rate by 5-30%.
3 = even or 50/50 chance	3 = no impact.
4 = unlikely	4 = increases growth rate by 5-30%.
5 = almost impossible	5 = increases growth rate by 30% or more

Table 5 Likelihood of occurrence of new forces or unprecedented events that might influence future population growth

1.8 Simple test for identifying the sex of unborn children
1.9 Increasing survival in middle age and early old age due to curing or improved therapy for heart disease, cancer and stroke
1.9 Public health programs decrease mortality of infant and young adults by 5-10%
2.1 Simple, safe, effective male birth control pill
2.3 Changes in death rates due to spread of AIDS in developing countries
2.4 Massive starvation of the scale of Somalia at least once every three years
2.5 Long term (at least one year) contraceptives widely used and accepted as birth control pills are today
2.5 Important negative changes in the environment (e.g., accumulation of toxic wastes, failures of mono-agriculture crops, contamination of drinking water) resulting in increased mortality
2.6 New deadly viruses, including AIDS mutations

2.6 Wide spread use of relatively cheap and easy ways to affect the aging process, resulting in diminished mortality and extension of the life span by about 5 years
2.6 3% of births via new methods of impregnation and prenatal development ("Outside womb" fertility, artificial inseminating, surrogate motherhood, other such techniques)
2.8 Simple method for selecting sex at conception
2.9 Rising incomes in most developing countries
2.9 Increasing sterility, worldwide, by 10% due to environmental degradation, higher level of stress, and other factors
3.3 Change in the Vatican's position: use of contraceptives becomes available without limit
3.4 Two children per family becomes the social norm in the majority of the developing world
3.4 Increasing impotency by 10% due to environmental degradation, higher level of stress, and other factors
3.4 Doubling, worldwide, of today's level of male and female homosexuality
3.5 20% increase of people active in religions that encourage higher fertility
3.6 Successful new prototype habitats in oceans, cold regions, or in earth orbit gives stimulates popular frontier spirit and alternatives to previous urbanization patterns
3.9 Basic economic needs met for 90% of global population (minimum acceptable health care, food, and shelter)

Table 6 Eventual impact of these new forces and unprecedented events on population growth over the next 25 years.

2.0 Long term (at least one year) contraceptives widely used and accepted as birth control pills are today
2.1 Rising incomes in most developing countries
2.1 Two children per family becomes the social norm in the majority of the developing world
2.3 Simple, safe, effective male birth control pill
2.3 Change in the Vatican's position: use of contraceptives becomes available without limit
2.3 Increasing sterility, worldwide, by 10% due to environmental degradation, higher level of stress, and other factors
2.4 Changes in death rates due to spread of AIDS in developing countries
2.4 New deadly viruses, including AIDS mutations
2.5 Important negative changes in the environment (e.g., accumulation of toxic wastes, failures of mono-agriculture crops, contamination of drinking water) resulting in increased mortality
2.6 Massive starvation of the scale of Somalia at least once every three years
2.6 Simple method for selecting sex at conception
2.7 Simple test for identifying the sex of unborn children
2.7 Increasing impotency by 10% due to environmental degradation, higher level of stress, and other factors
2.7 Doubling, worldwide, of today's level of male and female homosexuality
2.9 3% of births via new methods of impregnation and prenatal development ("Outside womb" fertility, artificial inseminating, surrogate motherhood, other such techniques)
2.9 Basic economic needs met for 90% of global population (minimum acceptable health care, food, and shelter)
3.0 Successful new prototype habitats in oceans, cold regions, or in earth orbit gives stimulates popular frontier spirit and alternatives to previous urbanization patterns
3.2 Public health programs decrease mortality of infant and young adults by 5-10%
3.4 Increasing survival in middle age and early old age due to curing or improved therapy for heart disease, cancer and stroke
3.4 Wide spread use of relatively cheap and easy ways to affect the aging process, resulting in diminished mortality and extension of the life span by about 5 years
3.5 20% increase of people active in religions that encourage higher fertility

With the information provided by the panel in round I, we identified two possible future events that were judged to be unlikely, but nevertheless potentially effective in reducing population growth rates, if they were to occur. Both have been considered before and one is the focus of large international programs. Without repeating what has been said and tried many times, we asked for suggestions about novel policy approaches that

might be practical, and if implemented, improve the probability of these developments. A distillation of those suggestions follow:

1.4.1 Novel policies that could lead to a social norm of two children per family throughout the majority of the developing world:

- Contraceptives added to the water supply, to conceive potential parents obtain counter active agent.
- Family-size tax that increased substantially for every child over two.
- Prohibition of child labor.
- Universal and compulsory schooling, for all to age 15-16, with mandatory immunizations from birth onward, and free available family planning support. Clean water is also a necessity.
- Tie all government subsidies to number of children per family in an inverse relationship.
- Why not encourage acceptability of no children for some couples in all countries? Why must we all reproduce? New paradigm needed
- Policies should focus on women's access to work and education.
- With enough effort and resources, the present policies (national, bilateral, and multi-lateral) that try to make family planning services universally available and to promote information, education and communication about family planning, hence the two child norm should be virtually universal within 25 years.
- Study what would motivate those who have more than two to limit to two. Different programs for different groups, China's program seems to work; Indians would need to design one for India.
- Free circulation of contraceptives in public health programs supported by churches and other religious organizations.
- International information utilities in education, health, and training in information society.
- Adopt Chinese incentives
- Bi- and multi-lateral aid requires it.
- Without significant reductions in poverty and infant mortality, and increase in women's education and empowerment, these policies would need to be coercive in nature or at least provide strong economic Disincentives to having many children.

1.4.2 Novel policies that could lead to the Vatican's acceptance of contraception without limit:

- Remove Holy See from the United Nations on grounds that it is not really a country and give it the same status as the World Council of Churches.
- Promotion of contraceptive that is accompanied with strong dissemination of moral values.
- Separate personal religious beliefs from personal choices as to the use of contraceptives.
- Theological doctrine developed by U.S. Catholic Bishops in support of sustainable development (of Government Statements at UNCED 1992).
- New international convention of religious leaders - Vatican III - with supporting papal encyclical letter - 1995.
- Elect a non-European Pope.
- Schism
- It is not practical.
- Create philosophical shift among powerful within the Roman Catholic Church.
- Policies that focus on responsabilization of woman to make choices.
- Only decline of organized religion is likely to affect the situation.
- Allow priests and nuns to marry and pay for the raising of children.
- Reconsideration of the theology of St. Thomas Aquinas - especially the discarding of the Thomastic view of "natural law" as it applied to human sexuality. Christ said nothing whatsoever about human sexuality. Separate the notion of procreation as the only "natural end" from that of enjoyment.
- Try to influence Vatican that some kinds of contraceptions could possibly be acceptable-not "chemical" and for women contraception but "mechanical" and for men (condoms) contraception.

1.4.3. Other policy areas of your selection and how you would address it:

- Research for long-term implantable ovulation suppression device.
- Global televised debate on population policy, environmental protection, and social ethics.
- Sex education in schools
- All nations will have to eventually adopt policies that clearly state the freedom of individual choice.
- Policies that emphasizes people to take charge of their lives and reduce dependence on governments.

Environment Section

Panelists were presented several reasons why population growth may led to environmental degradation and were asked to suggest additional reasons. They assessed each using Scale C below. **Table 8** lists the average of the panel's rating.

Scale C

1 = of paramount importance
2 = of great importance
3 = of modest importance
4 = of little importance
5 = no effect

Table 8 Some reasons why population may lead to environmental degradation

1.7 The drive to improve "standard of living" through consumption
1.8 Sheer size
2.0 Lack of understanding about the environment
2.0 The generation of waste
2.0 Diminished carrying capacity of the environment
2.0 Increased energy consumption per capita
2.2 Increased material use per capita leading to increased global demand for local resources
2.9 Increased social stress, breakdown of community values

Panelists were asked to rate the forces that led to changes in environmental quality over the previous five decades and to asses how these forces might change over the next 25 years. There were also asked to suggest and rate new such forces. **Table 9** lists the averages of the panel's judgments on their historic influence over the past five decades; Table 10 lists the averages of the panel's judgement on their future influence using Scale D below.

Scale D

Historic Influence	*Future Influence*
1 = Very Important	1 = Greatly increasing in importance
2 = Important	2 = Increasing in importance
3 = Marginally important	3 = Remaining the same in importance
4 = Unimportant	4 = Decreasing in importance
5 = Counter impact	5 = No longer a factor, or mixed

Table 9 - Some historic influences on Environmental Quality in order of importance

1.4 World population growth
1.7 Lack of availability or use of clean energy generation systems compared to coal, oil, wood, or dung burning
1.7 Economic systems that treat the environment as a free good
1.8 Popularization of the automobile

1.8 Excessive use of industrial processes that have by-products of toxic wastes, without appropriate attention to the disposal of such by-products
1.9 Lack of political leadership, public support or will to address environmental problems
1.9 National accounting systems that do not reflect natural resources and environmental damage
1.9 Pricing of energy in a way that does not account for its full environmental costs
1.9 Lack of adequate waste management in most places in the world
1.9 Aggressive forestry
2.0 Short term, reductionist, anti-generalist, and reactive policy making with a lack of concern for future generations in most developed countries
2.0 Lack of economic incentives for corporation and individuals to be environmentally responsible.
2.0 Rise of a "throw-away society" in developing nations
2.0 Excessive consumption and contamination of water aquifers
2.1 Lack of understanding of environmental interrelationships by average person
2.1 Use of chemicals in agriculture
2.1 Lack of integrated action among land use planning, family planning, environmental protection, and sustainable development in most places of the world
2.1 Excessive farming on marginal lands
2.2 The desire by most people to be "modern," and by most nations to "industrialize," and value systems to be anthropocentric
2.2 Pricing of transportation, including public subsidies for transportation in a way that does not account for it full environmental costs
2.2 Military activity (also in the sense of diverting expenditures from other activities)
2.2 Mining and the search for and utilization of raw materials
2.3 Urbanization and suburban sprawl
2.4 Lack of funds for adequate environmental research, development and related environmental education and enforcement programs
2.5 Lack of effective communication among experts and with the public about environmental issues
2.6 Lack of academic attention, particularly from economists and social scientists
2.6 Fishing practices
2.9 Aspects of globalism and or nationalism that lead to loss of local culture and sense of individual responsibility
3.1 Lack of ecological ethic in religious values
3.1 Natural phenomena (e.g. volcano eruptions, etc.)
3.7 Failure to expand human community or habitation into oceans and space

Table 10 - Future influence of these forces over the NEXT 25 years in order of importance

2.0 World population growth
2.2 Lack of adequate waste management in most places in the world
2.2 Excessive consumption and contamination of water aquifers
2.3 Urbanization and suburban sprawl
2.3 Excessive farming on marginal lands
2.4 Lack of integrated action among land use planning, family planning, environmental protection, and sustainable development in most places of the world
2.4 National accounting systems that do not reflect natural resources and environmental damage
2.4 Lack of funds for adequate environmental research, development and related environmental education and enforcement programs
2.5 Popularization of the automobile
2.5 The desire by most people to be "modern," and by most nations to "industrialize," and value systems to be anthropocentric
2.5 Lack of availability or use of clean energy generation systems compared to coal, oil, wood, or dung burning
2.5 Lack of understanding of environmental interrelationships by average person
2.6 Lack of economic incentives for corporation and individuals to be environmentally responsible

2.6 Pricing of energy in a way that does not account for its full environmental costs
2.6 Excessive use of industrial processes that have by-products of toxic wastes, without appropriate attention to the disposal of such by-products
2.6 Aspects of globalism and or nationalism that lead to loss of local culture and sense of individual responsibility
2.6 Lack of political leadership, public support, or will to address environmental problems
2.6 Short term, reductionist, anti-generalist, and reactive policy making with a lack of concern for future generations in most developed countries
2.7 Aggressive forestry
2.7 Pricing of transportation, including public subsidies for transportation in a way that does not account for it full environmental costs
2.8 Economic systems that treat the environment as a free good.
2.8 Use of chemicals in agriculture
2.8 Lack of academic attention, particularly from economists and social scientists
2.8 Natural phenomena (e.g. volcano eruptions, etc.)
2.8 Lack of effective communication among experts and with the public about environmental issues
2.9 Rise of a "throw-away society" in developing nations
2.9 Mining and the search for and utilization of raw materials
3.0 Fishing practices
3.0 Lack of ecological ethic in religious values
3.1 Failure to expand human community or habitation into oceans and space
3.3 Military activity (also in the sense of diverting expenditures from other activities)

Panelists were asked to assess new forces and unprecedented events that might influence environmental quality in the future and to suggest additional such forces and events. They were asked for their judgments about the likelihood of occurrence and impacts of these forces and events over the next 25 years using Scale E below. Table 11 lists the averages of the panel's responses about the likelihood of occurrence of these events within the next 25 years; Table 12 lists these in terms of the future impact (only the responses on future impact are given from the first round, because the second round had a typographical error in Scale E asking impact on population instead of on environment).

Scale E

Likelihood of occurrence within the next 25 years.	*Eventual impact on environmental quality*
a = almost certain	1 = very positive impact.
b = likely	2 = positive impact.
c = even or 50/50 chance	3 = no impact.
d = unlikely	4 = negative impact.
e = almost impossible	5 = very negative impact.

Table 11 Likelihood of some new forces and unprecedented events that might influence environmental quality

1.9 Wide use of new clean energy generating technologies, producing 10% of total output
2.0 Economic competition for certain raw materials, jobs, and markets becomes much more intense; protectionism through tariffs and non-tariff barriers increases
2.1 Actual demonstration [irrefutable evidence] of greenhouse warming
2.2 Cessation in production/consumption of ozone depleting gases
2.3 World per capita GDP grows by a total of 25% from 1990 level
2.3 Doubling 1990 levels (as a percentage of GDP) of per capita financing for environmental work
2.3 10% reduction, through substitution or other means, in the use of the polluting sources of energy such as coal, wood, and dung in developing countries
2.4 Development and wide spread use of plant varieties that thrive in salt or brackish water

2.4 National CO2 emission goals established by most nations
2.4 Energy prices increase 100% by additional taxation in most developed countries
2.5 Observation of a rise in sea level as predicted by Intergovernmental Panel on Climate Change
2.5 Transportation prices (public and private) increased 200% by additional taxation in most developed countries
2.6 GNP/GDP accounting revised in most countries to account for environmental costs
2.6 "Sustainable agriculture" practiced by 50% of farms in western countries
2.6 Global depression including 20% unemployment in OECD countries for more than one year
2.7 Environmental education essentially everywhere
2.7 "Congestion tolls" established on most major highways in developed countries
2.7 Creation and implementation in essentially every country of economic incentives for corporations and individuals to act in an environmentally responsible way, through mechanisms such as pollution "rights," including prices that reflect environmental costs, etc
2.7 Detonation of one or more nuclear weapons in war, terrorist incident, or by accident
2.8 Global environmental fund established with payment based on CO2 emissions (Al Gore proposal)
2.8 Environmental ethics and sustainability thinking dominates policy of most governments
2.9 Reduction of annual greenhouse gases emissions by 25%
2.9 Sales of electric cars account for 10% of all new cars sold world-wide
2.9 Wide spread planting of genetically engineered trees that mature in 3-5 years
2.9 50% of humanity views waste as "immoral"
2.9 Primary raw materials prices increased 200% by additional taxation in most developed countries
2.9 Pesticide use reduced 50% worldwide
3.0 U.N. enforcement against a government, corporation, or institution for violating environment treaty
3.0 Most LDC governments institute helpful environmental policies
3.0 75% of all waste is recycled in OECD countries
3.0 General rise of conservative governments deny most environmental deterioration, exaggerate extent of progress in protection, and roll back many environmental initiatives
3.1 25% decline in per capita meat consumption worldwide
3.1 The rise of political power of "green" parties in essentially all countries
3.1 50% reduction in waste generation (e.g. Canadian green plan goal)
3.2 Commercial demonstration of production of electricity through fusion power process
3.2 Revival of spiritual values for most people in the world which includes treating plants and animals as partners
3.3 Widespread war occurs, at least as extensive as World War II
3.4 Tropical rain forests stop shrinking
3.5 Major volcanic explosion leads to widespread global cooling and crop failure
3.9 Eating beef considered immoral in western countries

Table 12 Some new forces and unprecedented events rated by Scale E for their eventual impact on the Environment (average responses from first round only)

1.7 Doubling 1990 levels (as a percentage of GDP) of per capita financing for environmental work
1.7 Most LDC governments institute helpful environmental policies
1.7 75% of all waste is recycled in OECD countries
1.7 Environmental education essentially everywhere
1.8 National CO2 emission goals established by most nations
1.8 GNP/GDP accounting revised in most countries to account for environmental costs
1.8 50% reduction in waste generation (e.g. Canadian green plan goal)
1.9 50% of humanity views waste as "immoral"
1.9 Tropical rain forests stop shrinking
2.0 Reduction of annual greenhouse gases emissions by 25%
2.1 U.N. enforcement against a government, corporation, or institution for violating environment treaty
2.1 The rise of political power of "green" parties in essentially all countries

2.2 "Congestion tolls" established on most major highways in developed countries
2.2 Wide spread planting of genetically engineered trees that mature in 3-5 years
2.3 Sales of electric cars account for 10% of all new cars sold world-wide
2.3 Development and wide spread use of plant varieties that thrive in salt or brackish water
2.6 Actual demonstration [irrefutable evidence] of greenhouse warming
2.9 Eating beef considered immoral in western countries
3.1 World per capita GDP grows by a total of 25% from 1990 level

With the information provided by the panel in round I, some possible future events were judged to be unlikely, but nevertheless potentially effective in reducing environmental degradation. All of these have been considered before and some have been subject to large scale international programs. Without repeating what has been said and tried many times, we asked for novel suggestions about policy approaches that might be practical, and if implemented, improve the probability of these developments. Below is a distillation of the panel's responses:

2.4.1 Novel policies that could lead to the end of tropical rain forests shrinkage:

- Total ban on tropical hardwood imports.
- Manufacture pharmaceuticals from cultivated "tropical" plants.
- Ocean farming
- National accounts including environmental costs.
- Restriction on use of unsustainably produced tropical woods.
- Strict laws for replanting for each tree cut.
- Provide technical support, funding, and environmental facilities to countries with tropical rain forests.
- International supervision and national trade of CO2 discharge right.
- Global fund to substitute governments for lost income from reduced logging.
- Rain Forest Guard Program that pays those who destroy rain forests to protect them from others who destroy
- Educate to value these forests as natural treasure house of medicines
- Proper valuation of living forests. This can be achieved in part by incorporating natural resources into the national accounting system. A key factor will be the inclusion of non-market values.
- Policies that adopt new incentives not based on the economic calculus.
- Adopt a forest incentive connect local communities to rain forests to sister communities around the world.
- Free plants and trees to those who recycle
- Credit for carbon dioxide sequestering. In Canada and many other countries, there are programs to reduce the emissions of "greenhouse gases". Certain industries that emit large quantities of these gases, such as fossil-fuel-burning utilities, are facing new restrictions on production and expansion. -- They have argued that they should receive credit for initiatives they undertake in other parts of the world. For example, a utility in North American could purchase or secure an area of tropical forest that will sequester carbon dioxide to balance new emissions from the utility.
- Debt for Nature swaps. I believe the World Bank and other lending organizations have recognized the value of living forests by retiring portions of a country's debt in exchange for preservation of tropical rain forest.
- Increase the commercial benefits of the living forest.
- Tropical rain forests produce unique biological materials for pharmaceuticals, chemicals, foods and other commercial products. The benefits of these materials accrue to the company that develops them commercially and establishes patents. Some mechanism should direct benefits to the originating country.
- Agricultural policies that favor harvest of endemic foods (nuts, berries, fruits, etc) over clearing and cultivation of introduced crops and livestock. This could include policies to support research into rain forest foods as well as marketing and consumer education.
- Support for farm co-operatives This may encourage smaller-scale farming and forestry and direct the benefits to the local people.

2.4.2 Novel policies that could lead to 75% of all waste being recycled in OECD countries:

- Fines for not recycling; tax relief for those who do.
- Don't collect/dispose of unrecycled garbage.
- Taxes on raw materials and subsidy on recycled materials could achieve this, although it would make no "economic" sense.
- International supervision and fine of pollution trade.
- Improvements in design of household items utilizing recycled material
- Copy Swedish example of deposit/reimbursement for cans.
- Policies in which waste recycling efforts are shared with both producers and consumers.
- Tax companies and households for waste generated/collected.
- Introduce simple ways for companies and households to separate their waste for recycling.
- ...the best policy for future is to eliminate waste rather than to recycle the accumulated waste. The recent experience in Germany shows that its recycling program has produced 400,000 tons of waste plastic for recycling instead of the expected 100,000 tons. Germany can recycle 80,000 tons (40%) of plastic a year at present, and will have to ship the rest abroad - much of it to Eastern Europe and the Third World. News reports tell of contractors filling false recycling reports, then tipping their loads in dumps. Nowadays recycling costs in Germany are as much as around $9 per pound. By 1997 the German government says it will be able to recycle 800,000 tons of plastic. Even this one example shows limited capacities of 'developed' countries to do recycling within their countries. Even they recycle 75% of their waste, because of its cumulative growth, if unchecked, the remaining 25% will severely affect the environment. Dumping waste, especially toxic, hazardous, nuclear, abroad following the not-in-my-backyard approach philosophy which is fueled by present day economics is a real future issue. I wonder how other respondents see it.
- Full-cost accounting. Policies that support full-cost accounting of waste disposal will generally increase the cost of disposal and encourage waste reduction and recycling.
- Ban the export of waste beyond national borders. This includes solid waste and sewage, but not materials destined for recycling.
- Tradable permits or quotas for waste disposal in international waters.
- Tradable permits are being implemented for companies emitting air contaminants. A similar system based on an international agreement could be applied to companies disposing of waste into oceans and international waterways. A company that recycles waste destined for ocean-dumping would be able to sell or trade its excess quota.
- more opportunities at community level to achieve higher levels of recycling

2.4.3 Novel policies that could lead to sustainable agriculture being practiced by 50% of farms in more economically advanced countries:

- Compulsory and clear labelling of all food products on whether they have been produced in a sustainable manner ("Eco-labelling") coupled with strong consumer education.
- Reduction in area of land set-aside under European Community common agricultural policy.
- Proportionate reduction of allowable agricultural imports.
- Heavy taxation on petro-chemical fertilizers and pesticides, even soil loss.
- Global sustainable agricultural awards for new breakthroughs
- Requirement that every farm should generate a 5-year conservation plan subject to regular review
- New scientific methods including bioresources would (in the long term) contribute to sustainable agricultural practices.
- subsidy for soil conservation, integrated pest management, polyculture, organic fertilizers.
- Popularization of the way of agricultural planning by macro-coordination science.
- Educate about value of healthier foods.
- Reduce subsidies and let them compete.
- Consumer education favoring produce from sustainable agriculture. This should include a credible system for consumers to identify produce that meets "sustainable agriculture standards".
- Increase monitoring of toxic chemicals in our environment. There is growing public concern about the threat of the environment to our health and a sensitivity to chemicals in our food, water and air.

- By monitoring and publishing data on toxic chemicals, pressure will grow to eliminate their sources, including agricultural chemicals.

2.4.5 Other policy areas:

- All environmental and population policy areas have to be re-invented. Policies that are currently in effect are completely inadequate for a world that is interdependent and has at its disposal safe and useful technologies.
- Encourage transition from competitive economy to synergetic economy.
- Establish the law of the Life Cycle Assessment in each country.
- Each country publish a forward looking statement every 5 or 10 years.
- Each country should hold an annual "search" conference to assess priorities and review public expenditure
- Massive increase in financial support in status of scientific research in tropical biology (NOT just agronomy or forestry).
- ecological economies (considering nature as a scare natural capital, properly valuing it, allocating natural commons similar to allocating electromagnetic spectrum through the ITU, and ecological taxes, make environment a global security issue, off-set investments that allow firms to remedy environmental damage in one country by cheaper countervailing measures in another, tradeable pollution permits hat fix global emission limits for countries or industrial sectors, reduce import tariffs on environmentally sound technologies, good and equipment, tax breaks for more environmentally sound practices),
- more flexible repatriation limits for income made from these technologies which could provide firms with necessary financial break to enable investments in more costly, green technologies in less developed countries,
- higher tariffs or taxes on polluting products or technologies, with the revenues collected to be used to subsidize the acquisition of environmentally safe technologies,
- bulk purchase agreements for a region,
- purchase guarantees by bilateral, multilateral or regional funding agencies which could underwrite less developed country purchases of sound technology,
- an international technology bank, funded by country pledges, could acquire the rights to innovative green technologies so as to make them easily available to environmentally less advantageous countries,
- an international center to settle investment disputes could curb restrictive business practices that block environmentally less advantageous country access to sound technologies, such as restrictive licensing arrangements and prohibitively high prices.
- debt-for-nature swaps.
- development assistance programs that could also provide additional impetus to green technology transfers. Some of these instruments have already been used. Their efficiency varies.

Demographics of the Population and Environment Delphi Panel (round I and or round II):

SECTOR/INSTITUTIONS	NUMBER RECEIVED	PERCENT OF TOTAL
Private Sector	13	26 %
Universities	12	24 %
International Organizations	10	20 %
Governments	8	16 %
NGOs	7	14 %
TOTALS	50	100 %

Number Sent	Percent of Total	Country/ Region	Percent responded of number sent by country/region
29	38 %	United States	38 %
19	25 %	Europe	53 %
13	17 %	Asia	77 %
7	9 %	Africa	71 %
5	7 %	Latin America	60 %
3	4 %	Canada	66 %
76	100 %	TOTALS	

EXPERIMENTAL ECONOMICS

SESSION I

MODERATOR Doug Hale, Energy Information Administration

PRESENTATION 1 Charissa Wellford, Federal Trade Commission

SESSION II

MODERATOR Andy Kydes, Energy Information Administration

PRESENTATION 1 Steven Elliott, Oak Ridge National Laboratory

PRESENTATION 2 Jamie Kruse, University of Colorado

PRESENTATION 3 David Bjornstad, Oak Ridge National Laboratory

Experimental Economics

Antitrust: Results from the Laboratory

Charissa Wellford, Federal Trade Commission

I. Introduction

Antitrust practitioners are well acquainted with the challenges of applying economic theory. Industrial organization issues have been studied in the laboratory for over three decades, and the results of this research offer useful insights for applied work in antitrust.[1] Experimental methods provide a means for enhancing our understanding of markets, broadening the base of knowledge that is appropriate to draw from as an expert witness. Some have suggested that laboratory findings could be admissible as evidence in court.[2] In this paper I present results from the experimental literature that address competition issues.[3]

In short, laboratory markets are created by defining the structure and rules of the market, and by paying volunteer market participants in cash according to how successful they are at achieving their economic goals within this setting. In contrast to theory (or computer simulations of theory), in which the nature of behavior is assumed, real human behavior directly impacts market performance.

Landes recalls that Ronald Coase's reason for growing tired of antitrust was

> because when the prices went up the judges said it was monopoly, when the prices went down, they said it was predatory pricing, and when they stayed the same, they said it was tacit collusion.[4]

In the laboratory, the experimenter knows the true competitive equilibrium and monopoly equilibrium, for the experimenter sets the costs, demand, and other market parameters that underlie these theoretical predictions. Thus, when a firm cuts price in a laboratory market, it can be unambiguously determined whether it is behaving in a competitive or a predatory manner.

The market performance resulting from behavior observed in the laboratory can be measured relative to the theoretical predictions. For instance, exploiting the experimenter's ability to control the underlying cost structure so as to create conditions that are theoretically amenable to predatory pricing, enables tests of the theory to be examined in a market setting. The behavioral deviations from the theoretical competitive, monopoly, and predatory predictions, which are known, can be measured directly. Similarly, the market performance in laboratory monopoly and oligopoly markets can be compared directly to monopoly and oligopoly theory. Measuring the

---

[1] Chamberlin (1948) typically receives credit for conducting the first laboratory market experiment, although his procedures do not satisfy current generally accepted methods of experimentation. Fascinated with the potential of Chamberlin's experiment in which he was a participant (*see* Smith (1991b)), Smith (1962) began running several series of experiments in 1955 that demonstrated the competitive equilibrium as predicted by the standard textbook intersection of supply and demand schedules. Smith implemented his test using the double-auction institution, which he developed based upon the basic trading rules of the New York Stock Exchange. Other important early research includes Hoggatt (1959) on oligopoly and Fouraker and Siegel (1963) on bargaining and oligopoly. The use of laboratory methods has expanded significantly in the past few decades, with over 40 universities worldwide installing and maintaining computerized laboratories. Many other sites conduct experiments without the use of sophisticated equipment.

[2] *See*, for instance, Kirkwood (1981).

[3] This brief survey of experimental results relating to antitrust is not a comprehensive compendium, but rather an introduction. Readers interested in further reading on industrial organization experiments with implications for antitrust are referred to Plott (1982, 1989) and Holt (1989, 1991). Isaac (1983) considers the role of experimental economics as a tool in public policy analysis and provides several examples. Smith (1981b) discusses the relationship of theory, experiments, and antitrust policy. The *Handbook of Experimental Economics*, edited by Kagel and Roth (1991), surveys a broad array of laboratory results. Smith (1990) summarizes the major findings from the first 30 years of experimental research, and Smith (1991a) contains reprints of many of the prominent papers in the field.

[4] *See* Kitch (1983), p. 191.

distance of the observed behavior from the competitive, Cournot, and the monopoly equilibria under different conditions is possible.

Control and the ability to replicate are the premier advantages of laboratory analysis. Each time that a design parameter (control variable) is altered, a different experimental *treatment* is created. Some examples of different laboratory treatments that relate to competition issues include: allowing sellers to openly discuss their choices, prohibiting direct communication except through direct market choices (*e.g.*, price, quantity), providing more information to some parties than others, creating market power, restricting or permitting entry, and changing the trading institution (rules of contracting). Changing one control variable at a time, the treatment effect can be measured directly. Replication of laboratory markets (repeating structurally identical markets with different groups of participants) at other laboratories by other scientists increases the confidence with which we can accept the results.

The ability to control many features of laboratory markets and to replicate experiments under the same conditions allows tests that discriminate among theories. This provides an important tool to antitrust practitioners, who are frequently faced with this very task of identifying the relevant theory. While laboratory results may not identify which theory holds for a specific antitrust case, they can influence how an antitrust practitioner interprets the relevance of a theory. Once a theory is shown to be robust in laboratory markets under the conditions that the theory itself suggests, it can be "stress tested" to determine under which other conditions the theory continues to predict behavior and those under which its predictive ability vanishes. This also provides insight as to how markets work and which theories are most useful in particular circumstances.

Those who advocate the policy application of a particular theory bear the burden of proving why the theory is appropriate as a basis for policy. One way to examine a theory is in the laboratory. When a theory fails to predict well in a simple laboratory setting under conditions the theory itself suggests, it is difficult to believe that it would predict better in a more complex market in the naturally-occurring economy.[5]

Experiments enhance what we learn from traditional empirical analysis of field markets. In general, laboratory analysis is well suited for some sorts of antitrust issues and not for others. For example, experiments are not practical for estimating some case specifics such as the true industry cost structure, demand elasticity, or proper definition of the market. Traditional econometric analysis is better suited for that task, as well as the analysis of marginal changes in specific field markets. In contrast, laboratory experiments are better suited for examining regularities under general industry structural characteristics, characteristics often influenced by both theory and empirical studies of naturally-occurring markets. As Smith (1992, p. 2) notes, laboratory analysis is less a substitute for traditional econometric/empirical methods than it is a supplement which seeks appropriate data for dealing with structural questions that are beyond the scope of aggregate observational data.

Both methods of empirical analysis share the same goal: coherent observation.[6] The laboratory enables the formation of data that address fundamental changes which often are difficult, if not impossible (or illegal), to address with field data. For instance, consider comparing the effects on competition of different information conditions, institutions, market concentration levels, or communication conditions (*e.g.*, permitting sellers to talk about price), while holding everything else constant. Moreover, as mentioned previously, the laboratory permits the observations to be evaluated against the true theoretical benchmarks.[7]

The prerequisite for effectively interpreting and applying experimental findings is understanding how a microeconomic system is created in the laboratory. The next section gives a succinct overview of methodology, followed by a discussion of laboratory results of relating competition issues.

[5] Plott (1987) and Smith (1992 and 1993) summarize several laboratory experiments designed to inform policy in field markets.

[6] *See* Smith (1985).

[7] The laboratory also facilitates the analysis of econometric estimators. *See* Cox and Oaxaca (1990).

II. Experimental Methodology

Every microeconomic system investigated in the laboratory consists of (1) an *environment*, (2) an *institution*, and (3) *behavior*.[8] An environment is made up of the utility functions and technology (knowledge and skill endowments) that each participant brings to the experiment, as well as the commodities initially endowed upon each individual by the experimenter. The utility functions and technology are inherently unobservable; however, their consequences are observable. In a market experiment, the reduced form of the environment is given by the supply and demand schedules, which summarize the incentives to exchange.

Elements of an institution include a language (comprised of messages sent by each participant), allocation rules, cost imputation rules, and adjustment process rules.[9] All components of an institution, which define each agent's property rights in communication and in exchange, are observable. An institution is given in the laboratory by the experimental instructions, which describe the rules of exchange that each participant receives. One example of a familiar institution is the English (or progressive) auction, commonly used to sell objects (*e.g.*, art or livestock) in art/antique auction houses or at county fairs. Bidders in an English auction must improve upon (*i.e.*, bid higher than) the standing bid in order for their bid (message) to be recognized by the seller. The last bidder, the one who bids the highest, is allocated the object and pays the seller the amount of the final bid.

Once an environment and institution are established in the laboratory, the third major element of a microeconomic system, behavior, can be observed. The resulting behavior of participants in an experiment is a function of both the environment and the institution. *Behavior is not induced*, the incentives that participants face are.[10]

A laboratory microeconomic system is created by *inducing value*, which involves assigning monetary value to units of experimental assets.[11] For instance, each seller in a laboratory market is assigned a cost for each unit of a commodity that he might choose to sell. The difference between the price at which he sells a unit and its cost, determines the seller's profit from the sale of that unit. Likewise, each buyer is given the redemption value, guaranteed by the experimenter, associated with each unit of the commodity that he might choose to purchase. The difference between the resale (redemption) value of a unit and the price he pays for the unit, determines the profit that the buyer earns on the unit. In this manner, the monetary value of a decision becomes well defined for each market participant.

Control of preferences in a laboratory microeconomic system requires that four precepts are satisfied:
(1) *Nonsatiation*. Utility increases in cash rewards (*i.e.*, each participant prefers more money to less).
(2) *Saliency*. Rewards properly reflect incentives, rewards increase (decrease) in the favorable (unfavorable) outcomes.
(3) *Dominance*. Rewards sufficiently compensate each market participant for any subjective costs associated with participation in the experiment.[12]

---

[8] For a more complete and rigorous presentation on the creation of microeconomic systems in the laboratory, *see* Smith (1982b). For a condensed version, *see* Smith (1987). My summary of methodology utilizes his terminology.

[9] Messages, for example, might include a bid or an offer, or acceptance of a bid or an offer. The allocation rule is a function of the messages of all participants, and specifies the allocation of commodities. The cost imputation rule gives the monetary payment each agent receives as a result of the messages. Adjustment process rules simply define the starting, transition, and stopping rules of the experiment.

[10] If behavior were induced, experimentation would be synonymous with simulation, in which behavior is assumed and some element of randomness is injected into the process. For instance, participants in a Cournot (quantity-setting market) institution are not constrained to choose the Cournot quantities. Rather, each participant selects a quantity from his choice set, which includes the Cournot quantity among several other production alternatives.

[11] *See* Smith (1976b) for the theory of induced value.

[12] Failing to satisfy dominance could lead to participants making their decisions in a purely random fashion, because the monetary compensation does not sufficiently compensate them for their decision costs. Many experiments have been run in which payoffs have been doubled, tripled, or even more. In general, these increases do not significantly affect the mean of observed participant behavior. As long as participants earn at least the opportunity cost of their time, the results are not significantly affected. Smith and Walker (1993) report that increasing the payouts usually does not change the mean of the data, although it generally reduces the variance of the observations about their mean.

(4) *Privacy*. Each participant receives information on his own reward schedule, but not the reward schedules of the other participants.

Precepts (1) through (4) permit tests of theory. However, an additional precept is necessary for addressing whether replicable laboratory results shed light on the behavior in and performance of naturally-occurring markets: (5) *Parallelism*. Hypotheses that are supported in one laboratory or field market experiment also hold in other laboratory or field markets when corresponding conditions are held constant.

Like theory, laboratory markets may or may not parallel the forces that drive naturally-occurring markets. Of course, if specific theories or laboratory markets are used to inform policy, it is desirable that they provide a credible parallel. The use of careful experimental design, sometimes enhanced by computerized laboratories, permits the study of fairly sophisticated markets.[13]

The above methodology is usually applied in university laboratories, and laboratory market participants are generally college undergraduates or graduate students, who likely are not as sophisticated as the agents in markets of interest. Hence, the methodological issue arises that laboratory market outcomes might vary with the pool of participants. Most economic theories are based on the concept of rationality, agents maximizing their utility given constraints. When market outcomes differ with the subject pool, this suggests either that rationality is not a generalizable trait (*e.g.*, students are not rational, but OPEC members are), or that experience with a parallel institution in the naturally-occurring economy is important. Participants usually receive training in a specific laboratory institution, especially for the more complex institutions. Laboratory markets of different types have been run with GE executives,[14] Chicago Mercantile Exchange over-the-counter stock traders,[15] Energy Administration officials, natural gas corporation executives,[16] Eastern European reformers,[17] and other professionals. In general, these groups do *not* exhibit behavior that is significantly different than that of student participants in laboratory markets.[18] Note that if subject pool selection becomes an issue, the laboratory accommodates its analysis.[19,20] Still, in the interest of satisfying payoff dominance while optimizing the research output subject to budget constraints (the opportunity cost of a student participant likely is lower than that of a Chicago Mercantile Exchange trader),

[13] *See*, for instance, the Rassenti, Reynolds, and Smith (1989) study of cotenancy in laboratory networks which are designed to parallel natural gas networks. *See also* Hong and Plott's (1982) telephone market study of rate filing policies for inland water transportation; and Smith (1993), which discusses the laboratory market foundation of the recently formed Arizona Stock Exchange (AZX).

[14] Fouraker and Siegel (1963).

[15] King, Smith, Williams, and Van Boening (*forthcoming*).

[16] Smith presented the results of running experiments with both regulators and then with some of the regulated parties when summarizing his research on natural gas networks in Rassenti, Reynolds, and Smith (1989) at the American Enterprise Institute Conference on Policy Approaches to the Deregulation of Network Industries, October 10-11, 1990.

[17] This data was collected by Michael Block and Vernon Smith, both of the University of Arizona, during a World Bank sponsored conference on market reform, Vienna, Austria (September 1992 and July 1993).

[18] *See* Ball and Cech (1993) for a review of subject pool selection in experimental economics. Many psychology experiments involve different subject pools, however, psychology experiments frequently violate the precepts of nonsatiation, saliency, and dominance (usually by not monetarily rewarding laboratory participants for their decisions). Hence, the results from such experiments should be interpreted with caution. *See* Cox and Isaac (1986) and Smith (1991) for a comparison of the laboratory methodologies utilized in economics and psychology. Wilcox (1989 and 1992) and Smith and Walker (1993) evaluate the use of monetary rewards in laboratory markets.

[19] *See*, for instance, Gerety (1987) which finds that undergraduates and prisoners do not exhibit significantly different propensities to conspire when presented with opportunities to do so in sealed offer auctions (markets in which each seller submits an offer to supply a good or service—similar to government procurement) without the threat of antitrust enforcement. However, as the Gerety hypothesized *a priori*, when the threat of enforcement was introduced, by means of a random monetary penalty for a violation, the subject pools viewed the threat differently. Namely, the students exhibited risk-aversion, concentrating excessively on the size of the penalty; while the prisoners were risk-seeking, focusing more on the low probability of detection.

[20] Dyer and Kagel (1992) test common value auction theory in the laboratory using students and professional construction contractors as participants. The use of the professionals highlighted the failure of current theory to parallel their industry. Several contractors requested information about the architect, suggesting that reputation, though not part of the theory, is an important factor in their decisionmaking.

and because most economics laboratories are located at universities, college students have been and likely will continue to be the preferred subject pool for most cases.[21]

In laboratory studies, as in econometric studies of naturally-occurring markets, it is possible to incorrectly implement the methodology and consequently draw inappropriate inferences. Beyond the basic precepts of good experimental design, discussed above, many other elements must be considered when designing an experiment.[22] A few other design matters will be discussed below in presenting laboratory findings that relate to competition issues.

III. Some Experimental Results

A. Competitive Behavior and Institutional Effects

Laboratories are well-suited to examining straight-forward (and falsifiable) theoretical predictions. For instance, it is often hypothesized that the competitive outcome, where total surplus is maximized at the intersection of the demand and supply schedules, requires many buyers and sellers, as well as perfect information. The competitive equilibrium prediction provides a relatively accurate description of market performance across many laboratory institutions and environments. However, the speed with which competitive equilibrium is reached, the proximity of observations to competitive predictions, and the propensity for cooperative behavior under various conditions varies with the institution. As Smith (1989) highlights, until the 1960s, economic theories remained largely institution free, with outcomes derived from the environment (market structure and agent knowledge) and *ad hoc* assumptions about demand (*e.g.*, perfect demand revelation or "price-taking" by buyers). For instance, the standard competitive equilibrium model says nothing about *how* equilibrium is achieved. Without specifying the mechanics of the underlying institution, such as the trading rules, the competitive model merely assumes that convergence will occur.[23] Important early developments in institution-specific theory include Shubik's introduction (1959) of the extensive form game to industrial organization theory; Hurwicz's theory (1960) of mechanisms; and Vickrey's work (1961) on (first-price, Dutch, second-price, and English) auctions. As Shubik notes (1959, p. 183):

> It is foolish to entertain the delusion that the Cournot, Edgeworth, Chamberlinian, or other 'reaction-curve' assumptions are good approximations of non-co-operative behavior in the market *and* that we can leave all aspects of asset and corporate structure safely assumed away in a *ceteris paribus* condition and still come out with a useful theory of oligopoly.

Before we are in a position to examine the detailed meaning of any type of behavioristic assumption or reaction-curve we must be able to describe with reasonable verisimilitude the salient observable features of a market in a dynamic setting. [*Emphasis in original*]

Institution-specific theories, and laboratory tests of institution-free theories across different institutions, suggest that institutions matter.[24]

---

[21] In most laboratory markets, each participant has the same degree of familiarity or *experience* with the institution. This, of course, can be controlled if it is important that the market participants have varying degrees of experience with the market.

[22] Methodological issues are discussed throughout the literature. *See*, for example, volumes of *Research in Experimental Economics*, Greenwich, CT: JAI Press, and Davis and Holt (1992).

[23] The concept of the Walrasian auctioneer comes to mind. The Walrasian model provides an institutional framework to describe how equilibrium is achieved. However, the abstract concept of employing a fictitious auctioneer raises the issue of whether the model parallels observable features of any institution in the naturally-occurring economy.

[24] The behavioral results from laboratory tests of auctions led to extensions of Vickrey's auction theories. For example, *see* Cox, Roberson, and Smith (1982) and Cox, Dinkins, and Smith (1993). Bargaining theory has also been altered to reflect laboratory results, *see* Bolton (1991). Other examples of theoretical work prompted by experimental results include Friedman (1984, 1991), Wilson (1987), and Easley and Ledyard (1988).

Several institutions have been studied in the laboratory. Two of the more important from an antitrust perspective are the *double-auction* and the *posted-offer* institutions.[25]

1. The Double-Auction Institution

The double-auction, formulated by Smith (1962, 1964), involves buyers who are free to make bids and accept offers, and sellers who are free to make offers and accept bids for units of a homogenous commodity. Trading takes place over several market periods of specified duration. As Smith (1976a) points out, these markets are similar to real estate and over-the-counter securities markets, as no unaccepted bid or offer is binding in a later period unless it is resubmitted. Another version of the double-auction institution imposes an improvement rule, permitting only those bids (offers) that are higher (lower) than the outstanding bid (offer). Under the improvement rule, the double-auction more closely parallels organized exchanges such as the American or New York Stock Exchange, as the traders cannot cancel a bid or offer that they have made that is currently standing as the lowest offer to sell or the highest bid to buy. With or without the improvement rule, the observed outcomes are roughly the same.

During the past decade, the majority of antitrust cases that were challenged or settled by the FTC involved nonretail markets.[26] The double-auction institution also parallels important features of wholesale or nonretail markets, which frequently involve negotiation of contracts.[27] Hong and Plott (1982) find that when an institution involves direct negotiations (telephone-based negotiations are a key feature of some of the markets that they study in the laboratory) the performance of the market appears similar to that of double-auction markets in the laboratory.

In the laboratory, double-auction markets have very strong *convergence* properties and exhibit particularly high *market efficiency*, as measured by the percentage of the total potential surplus captured by the traders (or alternatively, the absence of dead weight loss). The tendency for contract prices to converge from the side of the market that has the greater theoretical surplus is well documented.[28]

Figure 1 presents the cost and value assignments for a double-auction experiment and the resulting market performance with four buyers and four sellers inexperienced in laboratory markets.[29] Each seller (buyer) is assigned costs (redemption values) for three units of the experimental commodity. For instance, Buyer 1's redemption value (denoted B1 on Figure 1) for the first unit purchased is $5.65. The experimenter guarantees to redeem Buyer 1's first unit for $5.65. Thus, Buyer 1 earns the difference between the redemption value of the unit and the price he pays for it. As indicated in Figure 1, the redemption values for Buyer 1's second unit is $4.60, and the third $4.45. Similarly, Buyer 2 is assigned redemption values of $4.95, $4.75, and $4.50 for the first, second, and third units purchased. On the supply-side, units of the commodity are produced to order in this market, so that sellers do not bear the cost of production or inventory for unsold units.[30] Seller 1's cost (denoted by S1 in Figure 1) of producing the first unit is $3.75. Seller one earns the difference between the price at which he sells the unit and his cost. His production cost for the second unit is $4.80, and the third, $4.95. Redemption value and cost assignments did not change with trading period, each seller (buyer) faced the same costs (values) for three units each period. By the fourth round of trading, prices have stabilized at the competitive level. The competitive

[25] Thousands of double-auctions and posted-offer experiments have been conducted. *See*, for instance, Smith (1962, 1964, 1967, 1976b, 1981a); Plott and Smith (1978); Smith, Williams, Bratton, and Vannoni (1982); Friedman (1984); Isaac, Ramey, and Williams (1984); Ketchum, Smith, and Williams (1984); and Cox and Oaxaca (1990).

[26] I am indebted to Malcolm Coate, Andrew Kleit, and Rene Bustamante (1992) for use of their data set.

[27] I am indebted to John Morris, former FTC Assistant to the Director of Antitrust, for suggesting the parallel to me.

[28] For example, if the surplus is distributed asymmetrically such that the consumers surplus exceeds producers surplus at the static competitive equilibrium, prices will tend to converge to the competitive equilibrium from above. *See* Smith and Williams (1982).

[29] The market design and data presented in Figure 1 are reported in Ketcham, Smith, and Williams (1984).

[30] *See* Mestleman, Welland, and Welland (1987) and Mestleman and Welland (1988) for analysis of inventory effects in laboratory markets.

exchange quantity of seven units is a Nash equilibrium in this market.[31] Appendix 1 provides sample instructions for a seller in a hand-run double auction market.[32]

Other experiments have also shown that four buyers and four sellers can be sufficient to assure rapid and consistent achievement of the competitive price in double-auction markets. These results are robust with nonstationary demand and supply parameters. In other words, when the market demand and supply schedules change over time, the competitive prediction continues to describe the observed market performance.[33] Laboratory research in markets with nonstationary demand and supply schedules suggests the importance of *not* enforcing price discrimination laws. In a dynamic setting with nonstationary demand and supply schedules, the convergence to a new competitive equilibrium *requires* that different prices be charged to different customers. Thus, enforcement (or threat of enforcement) of price discrimination laws, such as the Robinson-Patman Act, in a dynamic setting likely would perversely encourage anticompetitive outcomes. Charging different prices to different customers in a dynamic market is consistent with competition.

In double-auction *duopoly* markets the outcomes remain competitive.[34] In fact, even with *one* seller, the hypothesis of monopoly pricing often is rejected in the double-auction institution.[35] Figure 2 shows data from a single seller, five buyer double-auction market.[36] The buyers in this market session signal one another by bidding low, strategically *withholding demand* so as to mitigate the power of the monopolist, who does not know the true demand curve. Withholding demand occurs when at least one buyer refuses to purchase units of the laboratory commodity, even when his marginal valuation for a unit exceeds the posted price. The seller(s) relies on the observed buyer behavior to estimate demand. Thus, if buyers under-reveal their true willingness-to-pay then the demand curve is altered in effect. The buyers withhold demand with the intention of influencing the future pricing decisions of the seller(s). The laboratory data on withholding show that the true competitive equilibrium is often preserved. In fact, the misrepresentations of the buyers encourages the competitive outcome. In the market presented in Figure 2, the buyers effectively exert downward pressure on price until the competitive outcome is achieved. Some double-auction monopolists are able to sustain prices above the competitive equilibrium; however, they usually are unable to sustain prices and profits at the level predicted by monopoly theory.

The data from double-auction experiments have profound implications for traditional competitive price theory. The laboratory results show that the competitive equilibrium is attainable with less stringent informational assumptions than thought necessary. The data indicate that each double-auction trader needs only know his own valuations; extensive knowledge of other traders' demand or cost valuations is not necessary to generate a stable competitive outcome. Moreover, the results show that the market need not involve a "large" number of buyers and sellers to achieve competitive results.[37] Trading experience is also not imperative to achieve competitive market performance, although experience may speed convergence. Moreover, the results do not require the standard assumption that agents must be "price takers" for the competitive equilibrium to be attained. In the double-auction

[31] This is to say that if all but one seller (buyer) traded each profitable unit at the competitive equilibrium price, the remaining participant could not do better than to offer (bid for) each profitable unit at the competitive equilibrium price.

[32] Double auction and other experiments either are hand-run, without the use of a computer network, or computer-run. Using the latter method involves assigning each market participant to a computer terminal, which is networked to the other participant's terminals, at which they can enter decisions and receive market feedback.

[33] *See* Miller, Plott, and Smith (1977); Williams (1979); Smith (1982a); Cox and Oaxaca (1990); Davis and Williams (1990a); and Davis, Harrison, and Williams (1993). Hoffman and Plott (1981) also supports this finding when speculators or "middlemen" are present.

[34] Smith and Williams (1989).

[35] *See*, for example, Smith (1981a); Smith and Williams (1981); and Smith and Williams (1989).

[36] The market design and data presented in Figure 2 are from Smith and Williams (1989).

[37] This is evident even in the early double-auction experiments (*see*, for instance Smith (1962, 1964). Even though the "large numbers" condition remains in many textbook explanations of competition (*see*, for example, Carlton and Perloff (1989) p. 67), theorists also recognize that the number of buyers and sellers required to achieve competitive outcomes in an industry is analytically unimportant. *See*, for instance, McGee (1971) or Fama and Laffer (1972). Demsetz (1973, p. 26) points out that "there has been an irresistible inclination among economists to identify real world monopoly power with the structure postulate of the monopoly model -- the one-firm industry. It is but a short step from this to the conviction that market concentration is an index of monopoly power."

institution, every trader is a price *maker*.[38] Thus, in the double-auction institution, the competitive prediction is more robust than theory indicates.

Thus, the results from double-auction experiments suggest that there should be limited, if any, antitrust concern in markets with institutional features that resemble those of the double-auction.[39] The strength of competitive tendencies in the double-auction institution arise from the ability of both sides of the market, buyers and sellers, to have a voice in the market. Buyers make bids to buy and sellers make offers to sell, both can accept or reject contracts. To the extent that an antitrust market exhibits such institutional features, experimental evidence offers little support for anticompetitive concerns.

2. The Posted-Offer Institution

Frank Williams (1973) was the first to examine experimentally the posted-offer institution, in which each seller sets production capacity (maximum output) and posts a "take-it-or-leave-it" unit price.[40] Sellers produce to order, so that no inventories are carried from one period to the next, and no production costs are incurred for unordered units. Each seller chooses his price before knowing his rivals' prices. Buyers are selected in random order each period and take turns completing their purchases. A trading period consists of a round of posting prices followed by shopping. When the lowest price producer exhausts his production capacity for the period, buyers move on to the next lowest price producer, and so on. All buyers may not be satisfied.

The posted-offer institution parallels several features of naturally-occurring retail markets. Sellers in retail markets set production capacities and post prices that cannot be easily changed for some interval. For instance, catalog marketers are unlikely to change their prices until the next issue. Airlines post special rates that are good for travel within certain dates. Stock-outs (hitting capacity constraints within a trading period) are also observed in both retail and posted-offer markets. Sometimes retailers post prices with disclaimers such as "while supplies last" or "subject to available seating." In the laboratory posted-offer institution sellers generally do not incur the cost of units that are not sold. Thus, subject to capacity constraints, production meets demand.

Plott and Smith (1978) and Ketcham, Smith, and Williams (1984) demonstrate that changing the institution from a double-auction to a posted-offer auction can affect outcomes. Prices tend to be higher on average over time in the posted-offer institution than in the double-auction institution. However, and more importantly from an antitrust perspective, prices in both market institutions converge to the competitive price. The reason that posted-offer prices tend to be higher on average than double-auction prices is that in posted-offer markets prices usually converge from above and take longer to adjust to the competitive level.[41] Market efficiency is higher in double-auction markets than in comparable posted-offer markets.[42]

---

[38] Each trader can actively participate in the pricing process: buyers make bids and sellers make offers. *See* Smith (1962, 1964, 1976a, and 1982a); Ketcham, Smith, and Williams (1984); and Plott and Smith (1978).

[39] Further evidence, presented below, enhances this conclusion.

[40] This analysis of the posted-offer institution was actually an attempt to compare the single unit per agent endowment feature of the Smith (1962 and 1964) double-auction experiments to a double-auction in which traders are permitted to buy or sell more than one unit per trading period. However, when altering the number of units per trader, Williams also unintentionally modified several of the trading rules. These modifications change the institution from a double-auction to what is commonly known today as a posted-offer auction. Williams reports behavior that is somewhat different than the behavior observed in the Smith single unit per trader double-auctions. Notice that the hypothesis Williams is testing is compound, contaminating the pure effect of moving from a single to a multiple unit assignment of costs and values in the same institution. Plott and Smith (1978) recognize this problem. They also classify and then test the difference between the double-auction and posted-offer institutions.

[41] Davis and Williams (1985) find that, unlike in double-auction markets, posted-offer contract prices converge to competitive equilibrium from above, irrespective of the asymmetry between producers' and consumers' surplus. The institutional effect also interacts with other design parameters. In posted-offer markets with no Nash equilibria, prices converge towards the competitive equilibrium; while in cases in which there exists a nonNash competitive equilibrium and a Nash equilibrium, prices lie closer to the Nash equilibrium than to the competitive or the limit price equilibrium. *See* Ketcham, Smith, and Williams (1984).

[42] This comparison of institutions is conducted holding all else equal. Thus, while the posted-offer markets exhibit lower levels of market efficiency than double-auction markets, in practice there remains the issue of institution efficiency. For example, consider grocery shopping under the two alternative institutions. The transactions costs of negotiating the price of each of many items under the double-auction trading

When sellers produce in advance, rather than to demand, and there is no carry over of inventory from one period to the next, the mean price across the two institutions no longer differs significantly. However, the double-auction remains more efficient, because more unsold units are produced in the posted-offer institution. Thus, the production characteristics of an institution may have an impact on market performance.[43]

In contrast to double-auction settings, buyers do not possess the ability to signal in posted-offer markets. While buyers can reject posted offers, these decisions are private, eliminating a means to signal to other buyers. Sellers, on the other hand, maintain the ability to signal to other sellers via pricing decisions, which are posted for all market participants to see. Figure 3 presents data from a posted-offer market with experienced sellers.[44] The underlying market structure is the same as that of the double-auction experiment presented in Figure 1.[45] However, for the posted-offer trial displayed in Figure 3, attempts are made to create conditions under which tacit collusion would have a good chance of being observed, by recruiting the sellers on the basis of their prior propensity to collude tacitly.[46] Although some signaling efforts (attempts to encourage supracompetitive pricing) arise in the early periods, they are eventually aborted. The trial can hardly be classified as a case of successful collusion. Prices stabilize at the competitive level. Ketcham, Smith, and Williams attribute the slight increase in price during the later periods to "end effect."[47]

Recall that in the double-auction institution, the monopoly model has little predictive power. The monopoly prediction does a better job of describing behavior in posted-offer markets with single sellers. Yet, while the prices lie above the competitive equilibrium in most of the posted-offer markets studied, the index of monopoly effectiveness (the percentage of theoretical monopoly profits actually captured by the single seller) is seldom 100 percent. For instance, across the posted-offer monopoly sessions reported in Isaac, Ramey, and Williams (1983), the average index of monopoly effectiveness is 35 percent.[48] The Isaac, Ramey, and Williams study uses *human* buyers. Brown-Kruse (1986) shows that when demand is fully revealed (generally the case when human buyers are replaced by *simulated buyers*) sellers extract more rents from posted-offer markets. This suggests that when sellers lack complete information on demand, human buyers can strategically influence the market outcome.[49]

B. Oligopoly

1. Institutions and Information

Theory presents several alternative models of oligopoly. The laboratory permits the examination different institutions suggested by the theories, as well as the impact of various information conditions upon market performance. Many oligopoly experiments use either double-auction or posted-offer institutions, but other institutions

rules could be substantial and avoided by trading according to posted-offer rules. In general, the gain from avoiding such transactions costs in retail settings by employing a posted-offer institution likely will dominate the loss in market efficiency from not adopting the double-auction trading rules.

[43] Mestelman, Welland and Welland (1987) and Mestelman and Welland (1988).

[44] The sellers have participated previously in posted-offer laboratory markets.

[45] The demand and cost parameters across the two experiments differ by a constant, $0.55, but the underlying structure is the same.

[46] *See* Ketcham, Smith, and Williams.

[47] Because subjects were experienced, they probably had a reasonable idea that the experiment would end around period 25. Generally, participants are not informed of the number of periods to be conducted in a given market session.

[48] The per-period measures range from -222% to 100%.

[49] Monopoly theory also encounters difficulties in a *posted-bid* institution. Smith (1982) shows that in markets with *one* seller who can accept or reject bids, and five buyers who post bids (a posted-*bid* as opposed to a posted-*offer* institution), monopoly power is not exercised. This results because the buyers successfully signal other buyers via their bids to withhold their true willingness-to-pay. Thus, through misrepresenting their willingness-to-pay, the buyers neutralize the seller's monopoly power, because the seller does not know the true demand curve. Market price and production converge to the competitive prediction in this institution and environment.

such as Cournot[50] and Bertrand[51] are also studied. In both Cournot and Bertrand settings, the market demand curve is provided to all sellers. In Cournot markets, sellers set quantities; in Bertrand markets, they set price.

As first shown by Carlson (1967) for large markets (more than 20 sellers), and later by Wellford (1989) for oligopolies (five firms), quantity-setting markets that are theoretically unstable in the cobweb or Cournot sense exhibit strong convergence properties in laboratory markets. Subjects behave "as if" the markets are stable. The hypothesis that subjects form expectations consistent with the cobweb model, *i.e.*, that price next period will be the same as price in the current period, is not supported by laboratory tests. Further, rational expectations also fails to predict well. Of the traditional models considered, adaptive and extrapolative expectations best describe the data.[52] However, the structure, not just the parameterization, of expectations seems to be adaptive. Market participants alter their expectation process when they realize that their myopic responses lead to relatively unfavorable outcomes.[53]

Information plays an important role in laboratory markets. Most oligopoly theories require *complete* and *common information* on preferences to achieve noncooperative equilibria.[54] This raises issues for applied work in antitrust and other areas, as it is difficult to believe that each market participant in the naturally-occurring economy actually knows the utility values for all other participants. In the laboratory competitive and noncooperative equilibrium outcomes are more likely to arise behaviorally however, under the conditions of *private information* than under complete information.[55] If *sellers* actually had complete information, they often would prefer to achieve a different equilibrium.[56]

Fouraker and Siegel (1963) document the effect of complete versus private information in their study of Cournot and Bertrand duopolies and triopolies. Tables 1 and 2 summarize Fouraker and Siegel's results, which have been confirmed in many other environments and institutions. The tables provide the percentage of sessions in which the market quantity (price) in a Cournot (Bertrand) market converges to (or near) a particular equilibrium concept after a specified number of periods.[57] Sellers in complete information treatments know the profit functions of their rivals. In particular, sellers know how each rival's quantity decision affects the rival's profit. In the private information treatments, each seller knows only his own profit function. Referring to Table 1, duopolies in the reported quantity choice markets under private information show no support for the shared monopoly outcome (*i.e.*,

[50] *See*, for instance, Fouraker and Siegel, *ibid*; Binger, Hoffman, Libecap, and Schachat (1990); and Wellford (1990).

[51] *See*, for example, Friedman (1967) and Friedman and Hoggatt (1980); and Hoggatt, Friedman, and Gill (1967).

[52] It is interesting to note that macro theorists are reverting from rational to adaptive expectations. *See*, for instance, Thomas Sargent's address in honor of Milton Friedman's eightieth birthday at the 1992 Western Economic Association meetings.

[53] Wellford, *ibid*. Schmalensee (1976) finds in a stark (nonmarket) setting that the speed of response in the adaptive structure retards during turning points. *See* Grether (1980 and 1992) and Camerer (1987) for tests of Bayes rule.

[54] Complete information holds when each market agent knows the preferences of all the other agents. Common information holds when all of the agents have information, and they know that all of the other agents possess this information and so forth.

[55] Complete information in the laboratory generally refers to information on the payout functions and previous decisions of other market participants. Under conditions of private information in the laboratory, market participants know only how their own decisions affect their payout, and can observe the market outcomes (*i.e.*, the market price and market production level). For other discussions on the effect of complete information in laboratory markets, *see* Smith (1989) and (1990), and McCabe, Rassenti, and Smith (1989). Holt (1985) conducts a quantity-choice duopoly game under complete information, and notes that the subjects want to earn more than the other firm, yet show no tendency to maximize the difference between their earnings. He conjectures that there "...probably would have been less variability in the data if the subjects in these experiments had not been given the complete information necessary to compute the other seller's profits," (p. 323).

[56] In general, a symmetric Cournot duopolist can earn more by choosing the joint profit maximization equilibrium rather than playing the noncooperative (Cournot) strategy.

[57] Fouraker and Siegel report data for period 21 in the Cournot markets, and period 14 in the Bertrand markets. A period is one round during which sellers set quantity (price) in each Cournot (Bertrand) market session.

the joint maximum), and the strongest support (87.5%) for the Cournot equilibrium.[58] By contrast, the joint maximum equilibrium concept received support in 31.25% of the Cournot markets conducted under conditions of complete information. While the Cournot prediction continues to describe behavior in 46.87% of the markets, it has diminished predictive power relative to the otherwise equivalent private information treatments. The ability of the noncooperative equilibria (*i.e.*, Cournot and Bertrand) to predict also improves under conditions of private information for the both Cournot triopoly and Bertrand duopoly markets. Table 2 suggests that the incentives in the Bertrand markets are sufficient to overcome the information effect as the number of sellers increases, so that the Bertrand equilibrium predicts well under both information treatments.

The ability of markets to economize information (*i.e.*, without intending to, market agents achieve efficient outcomes) was recognized by Adam Smith (1776) in his discussion of the invisible hand and described in more detail by Hayek (1945). However, until the advent of laboratory methods, it was difficult to test *directly* whether this hypothesis would hold. At least through the mid-fifties, the economics profession at large did not accept the relevance of the claim.[59] In fact, before he started conducting laboratory experiments, Vernon Smith (1993), who some call the "father" of experimental methods in economics, also doubted its applicability until his experimental subjects convinced him otherwise.

> In many experimental markets, poorly informed, error-prone, and uncomprehending human agents interact through the trading rules to produce social algorithms which demonstrably approximate the wealth maximizing outcomes traditionally thought to require complete information and cognitively rational actors.

While the outcomes of Smith's early experiments supported the notion that markets economize information,[60] Smith (1982a) provides a more stringent test of the Hayek Hypothesis, which he specifies as (p. 167):

> Strict privacy together with the trading rules of a market institution are sufficient to produce competitive market outcomes at or near 100% efficiency.

Using the double auction institution, three treatments are considered: (1) stationary, (2) cyclical, and (3) irregular shifts in demand and supply. Figure 4 presents the data from one laboratory session conducted under the most rigorous design reported. The data are representative of other sessions conducted under the same design. In this market, after the first week of trading (five periods), the demand and supply schedules shift from D_1 and S_1 to D_2 and S_2. The four sellers have identical unit costs of $5.70, with an 11 unit aggregate production capacity during the first week and a 16 unit one in the second. The four buyers have the same unit value of $6.80, with the aggregate capacity to demand no more than 16 units during the first week, and no more than 11 units the second.[61] The competitive equilibrium during the first (second) week occurs at the price of $6.80 ($5.70) with an output of 11 (11) units. This market design provides a rigorous "stress test" of the Hayek Hypothesis. The competitive equilibrium is not the unique price and production mix which satisfies 100% market efficiency. Complete efficiency occurs when 11 units are sold at *any* price ranging from $5.70 to $6.80. This provides a stringent test of the ability of price to converge to the competitive equilibrium in this design, because at the competitive equilibrium during week one (week two) all of the $12.10 per period exchange surplus accrues to the sellers (buyers). Thus, at the competitive equilibrium during the first week (second week), buyers (sellers) receive only a 10¢ trading commission per unit exchanged. The extreme asymmetry of the exchange surplus between buyers and sellers creates tremendous incentives for the market agents that are cut out of the exchange surplus at the

[58] This is not to say that the Cournot equilibrium was observed in 87.5% of the markets during period 21, rather that the observations lie *closer* to the Cournot prediction than to the alternative equilibria of interest.

[59] As Smith (1993) notes, the Chicago and Austrian schools of thought were the exceptions. *See also* Kitch (1983) for a remembrance of the days at Chicago prior to the general professional acceptance of these ideas.

[60] *See* Smith (1962 and 1964).

[61] Market participants are not informed in advance that market conditions might change. Smith (1982a, p. 172) notes that "from the point of view of the participants this change is rather subtle in that their individual value and cost assignments remain constant." Only the aggregate demand and supply capacities change.

competitive equilibrium to exert more resistance in the bargaining process. Such resistance could curtail, if not prevent, convergence to the competitive equilibrium.

The results presented in Figure 4 reveal an slow but steady convergence of price to the competitive solution by periods four and five in week one. It appears as though sellers' expectations of high prices and profits initially retard the rate of convergence after the shift in the demand and cost schedules. However, the new competitive equilibrium is achieved by periods nine and ten. These data provide impressive support for the Hayek Hypothesis in an environment with stationary demand and supply curves in the double-auction institution. Smith (1984b) also finds strong support for the hypothesis in experiments with dynamic environments. Van Boening and Wilcox (1992) have constructed markets with avoidable fixed costs in a double-auction institution. Their markets provide an even greater challenge to the ability of the market to achieve the competitive outcome, and have spawned research efforts to explore and design institutions that overcome more effectively the lumpiness inherent in such cost functions.[62]

Another interesting informational finding relates to situations of asymmetric information. Camerer, Lowenstein, and Weber (1989) find evidence that supports a popular theory in psychology, that better informed agents often fall victim to the "curse of knowledge." The curse occurs when agents fail to ignore information that they have that others do not, even when it is unnecessary information. For instance, investment bankers, who possess more information than the public about a stock to be offered publicly, must anticipate the demand for the securities by the relatively uninformed public when pricing the shares.[63]

The curse runs contrary to the conventional assumption in agency theory that agents with better information can correctly forecast the decisions of less informed agents.[64] More information may not always be an advantage. This is not to say that no information is necessarily better than some. In fact, the anticompetitive distortions of asymmetric information outcomes can be overcome in markets.[65] Market forces in the double-auction institution decrease the magnitude of the "curse" by about 50%, this decrease is largely due to the more rational traders disproportionate share of actions in the market. Thus, the disciplining forces of the market promote the paradox that individual irrationality can encourage collective rationality.

The curse is otherwise difficult to erode, even with very large incentives and extensive training, learning does not improve much. The curse remains. Camerer *et al.* note that the curse could affect strategic behavior in oligopoly settings, as firms erroneously decide to enter or exit. However, this remains to be tested. The curse has also yet to be analyzed in a systematic fashion in other laboratory institutions. Such study may enhance our knowledge of the breadth of the curse and the ability of markets to overcome it.

2. Market Power

The role of information, the form of institution, and other factors come into play in laboratory studies of market power. *Market power*, classified as a "unilateral competitive effect" by the Horizontal Merger Guidelines (1992, §2.2), is the ability of a firm to maintain price above the competitive level unilaterally and profitably. *Equilibrium Market Power*, a general game-theoretic concept, refers to a noncooperative equilibrium that is consistent with supracompetitive prices.

The laboratory facilitates the analysis of market power, in that it can be directly built into the structure of the market. Thus, the behavioral consequence of the existence of market power can be directly compared to markets that are similar except for the lack of theoretical market power.

Figure 5 presents a market design in which unilateral market power exists in a noncooperative game.[66] At the competitive equilibrium in this market with five sellers, 16 units are produced at the price of $2.60. In order

[62] *See*, for instance, Durham, Rassenti, Smith, Van Boening, and Wilcox (1993).

[63] Other examples include fashion, wine, and movies. Industry experts have informed opinions about the value of the product, but if they are subject to the curse of knowledge their prices will tend to biased upward (downward) for high (low) quality items relative to the uninformed general public's willingness to pay, and thus also relative to the profit maximizing level.

[64] *See* Conn, Young, and Bishop (1991) for tests of principal-agent theory.

[65] *See*, for instance, Isaac, Ramey, and Williams (1984).

[66] This design originates in Holt, Langan, and Villamil (1986) and is implemented in several subsequent laboratory studies on market power.

to encourage the purchase and sale of the marginal units, each market participant receives a 5¢ commission per unit exchanged.[67] Either of the low cost sellers, seller one (S1) or seller two (S2), could unilaterally and profitably increase their price from $2.60 to $2.85. At the competitive equilibrium, the two low-cost sellers would each earn $2.60 per unit for five units of production, less the $11.50 cost of producing the five units, plus $.25 in commissions, for earnings of $1.75.[68] However, assuming that the other sellers behave competitively, if either seller one or seller two withholds its two zero-profit units from sale, the *effective* supply curve shifts back so that the *effective* competitive equilibrium, where the demand and (effective) supply schedules cross, occurs at the price of $2.85. At this price, the seller who suppresses production by two units would earn $2.40, including commissions. Hence, it is profitable for either seller one or seller two to unilaterally increase price above the fundamental competitive equilibrium. In this example, the other low-cost seller, who withholds no units of production from the market would earn $3.00. Hence in this example, the two low-cost sellers increase their earnings by either $0.65 or $1.25. Any of the high-cost sellers (sellers three, four, and five) could also withhold a unit, and unilaterally and profitably increase price above the competitive level, but not as dramatically as the low-cost sellers. Unilaterally withholding one unit from the market would lead to a price between $2.60 and $2.85, thus a high-cost seller would increase his earnings by no more than 20¢ and could decrease it by less than 5¢.[69]

Table 3 presents a terse summary of several laboratory studies of market power.[70] I briefly discuss the overall findings of the research and some of the issues that they raises.

a). Supracompetitive Pricing, Efficiency, and the Hayek Hypothesis

While not consistently observed across all design treatments, supracompetitive pricing does occur in some of the laboratory markets. Probably the most striking result across the studies is that *even when market power is built into the structure of the market, the observed market efficiency remains relatively high.* Thus, the deadweight loss associated with the presence of theoretical market power seems to be small in laboratory markets.

What do the results of supracompetitive pricing and high market efficiency imply when evaluated in the context of the Hayek Hypothesis? One, interpretation of the Hayek Hypothesis, the *price interpretation*, implies that prices should converge to the competitive level. Another interpretation, the *efficiency interpretation*, suggests that the surplus captured should approach 100%. Davis and Williams (1991) find that when sellers recognize they possess market power, the price interpretation is often violated in double-auction and posted-offer markets. However, the efficiency interpretation of the hypothesis cannot be rejected, even when the price interpretation fails to hold. Across the studies reported in Table 3, the average market efficiencies consistently exceed 90%.[71]

Several of the studies note that sellers with theoretical market power rarely withhold units in the manner that static market power theory predicts. For instance, Davis and Williams (1991, p. 270) report that across their eight posted-offer markets the two low-cost sellers

> offered less than nine of the ten units collectively available to them in only eight out of one hundred and forty six periods. Five of these instances occurred in the first two periods of trading in four different experiments, while the remaining three instances were isolated instances in three different experiments.

The authors also report that one of four of their double-auction trials exhibits evidence of strategic withholding by sellers one and two. Yet, these low-cost sellers fail to achieve prices that are significantly higher than those in the other double-auction trials.

[67] Plott and Smith (1978) find that 5¢ commissions are sufficient to motivate marginal trades.

[68] As Figure 4 indicates, seller one (or two) incurs production costs of $1.60 for the first unit, $2.35 per unit for the second and third units, and $2.60 per unit for the fourth and fifth units.

[69] Withholding two units would unilaterally make the effective competitive equilibrium occur at the price $2.85, however any high-cost seller doing so would forego earnings altogether.

[70] Several of the studies (*see*, for instance, Holt and Davis (1990)) include a baseline treatment which are similar to the other treatments as possible except that no market power exists. These baseline market trials are not included in Table 3.

[71] A few studies do not report efficiencies.

These unexpected results could be due to the fact that sellers in the laboratory markets do not recognize that they have potential market power. Only in a few cases does experience with the institution seem to increase the sellers' awareness of their potential market influence to the point that they attempt to exercise their power.[72] The general absence of withholding in these markets suggests that the observed supracompetitive pricing is due to tacit collusion rather than the exercise of market power. When market power exists, the likelihood of tacit collusion increases; yet, the achievement of perfect collusion is infrequently observed.[73]

b). Institution Matters

When controlling for market power, larger deviations from the competitive equilibrium are observed in posted-offer markets than in double-auction markets.[74] The differential impact is most obvious when the *buyers*, rather than the sellers, have market power. When buyers have market power in double-auction markets, they are able to elicit subcompetitive prices,[75] however in posted-offer markets they are unable to overcome the institutional asymmetry (*i.e.*, that buyers cannot post bids, but sellers can post offers - - so that buyers have a more indirect influence on price) and prices converge to the competitive equilibrium.

Friedman and Ostroy (1991) argue that in most recent laboratory studies of market power, market agents have the opportunity to trade only a few units of an indivisible commodity. This indivisibility significantly reduces the profitable opportunities to raise price by withholding units. As the authors point out, in some markets the only way a trader can influence price is by exiting from the market and therefore giving up all profit for the trading period. When divisibility of sales is incorporated, it affects the ability of sellers to exercise market power in the double-auction institution. Most notably, when divisibility is allowed, supracompetitive pricing is not observed. Friedman and Ostroy describe the behavior of the market participants (p. 3),

> Through their misrepresentations, the subjects spontaneously imposed a perfectly competitive environment on each other...[the institution with divisibility] encourages individuals with potential monopoly power to exercise that power in such a way that it is neutralized.

Another institutional feature that affects the exercise of market power involves whether the message space of the buyers is limited to include only their true reservation values, as is often the case when buyers are simulated rather than human. As found in other environments under the posted-offer institution, prices are significantly lower when human buyers are present than when buyers are simulated to behave as the theory predicts.[76] Davis and Williams (1991) find that when human buyers are present and sellers have market power, price converges to the competitive equilibrium in posted-offer markets.

While the data from experiments reveal that institutions matter, the data also suggest that a feature of the environment, the nature of aggregate production capacity at the competitive equilibrium, might determine whether market power will in fact be exercised or whether tacit collusion is more effectively achieved. When aggregate

[72] Holt, Langan, and Villamil (1986) find one experienced seller who figures out the advantages of withholding. However, other studies (*e.g.*, *see* Davis and Williams (1991)) find no evidence of increased incidence of withholding of production by experienced participants. Another way to increase the sellers' familiarity with their potential market power would involve establishing the condition of complete information, so that each seller knows the cost and payoff functions of his rivals. It could also be important to require the sellers to "earn the right" to be in a position of potential power. Hoffman and Spitzer (1982) find that in laboratory tests of bargaining that if the preferential property rights are randomly assigned rather than "earned" (through a trivial game) that participants are less likely to fully exercise those rights. This would make it clear to all sellers and buyers (when human buyers are used) that there exist fundamental differences in cost structures across sellers. Barring these two design changes, complete information and "earned" assignment of property rights, it is unlikely that the opportunities to the sellers could be made clearer.

[73] *See*, for instance, Davis, Holt, and Villamil (1990).

[74] *See*, for instance, Holt (1989) and Davis and Williams (1991).

[75] This is found in one of two buyer-power market trials (not summarized in Table 3) from the Holt, Langan, and Villamil (1986) study. This trial, which has a symmetric distribution of exchange surplus between buyers and sellers, establishes that contracts occurring at prices below the competitive equilibrium are not due solely to the asymmetric distribution of exchange surplus (*see supra* note 24).

[76] *See* Brown-Kruse (1986 and 1991).

production capacity of the market is constrained at the competitive equilibrium, the likelihood of supracompetitive pricing might be different than when this aggregate capacity is nonbinding. The column in Table 3 marked "excess capacity at CE" indicates whether or not the production capacity is nonbinding. This issue is discussed in greater detail in the next section, as it also arises in research on the effect of communication upon collusion.

3. Tacit and Explicit Conspiracies in Restraint of Trade

Many laboratory studies of oligopoly focus on tacit or explicit conspiracy.[77] The laboratory provides a useful mechanism for studying conspiracies. By controlling directly the communication of market participants (*e.g.*, whether or what kind of communication is allowed), the impact of explicit versus tacit conspiracy can be compared to theoretical predictions. Collusion is studied in several laboratory institutions, including the Cournot, double-auction, and posted-offer auction.[78]

a). Opportunities for Tacit Collusion

Permitting communication only through the choice variables specified by the theory and institution allows tests of the ability of sellers to collude tacitly. In other words, sellers have no means of communicating directly with each other except through their market decisions.

In Cournot industries with *more than two sellers*, prices tend to lie between the static Cournot and competitive predictions, the joint profit maximum (*i.e.*, the monopoly outcome) is not achieved.[79] Theoretically, the Cournot solution approaches the competitive level as the number of sellers increases. Behaviorally, the fewer the number of sellers, the higher the observed variance in prices in laboratory quantity-setting markets.[80]

In laboratory Cournot *duopolies*, the behavioral results are less clear. Binger, Hoffman, Libecap, and Schachat (1990) observe prices converging to the collusive outcome from above the Cournot-Nash equilibrium under conditions of complete information, while Morrison and Kamarei (1990) find successful collusion in only one of their eight laboratory duopolies conducted under conditions of incomplete information. Under complete information in quantity-choice duopoly games, the Cournot-Nash equilibrium predicts better than the consistent conjectures and the collusive equilibria.[81]

Similar mixed results arise in price-choice institutions with two sellers. Dolbear, Lave, and Bowman (1966) mention more variability in outcomes with duopolies than with triopolies in price-setting markets. Davis, Holt, and Villamil (1990) also find that some duopolies exhibit rivalistic behavior in posted-offer markets. Benson and Faminow (1988) show that subject experience with the institution increases the incidence of tacit collusion in posted-offer duopolies, but they too observe experienced participants who compete fiercely. So, while tacit collusion between two sellers appears to be easier to coordinate than when there are more sellers in laboratory markets, head-on competition is also observed in duopolies.

Posner (1976) points out that the concentration level at which concerns of collusive pricing arise depends on the economist. For instance, Scherer (1970 p. 185) would be concerned when the four largest firms hold a forty percent industry share. Bork (1978 pp. 221-222), on the other hand, argues:

[77] R. M. Isaac (1990) provides a useful overview.

[78] *See* Isaac and Walker (1985) and Gerety (1987) for worthwhile laboratory studies of collusion in sealed-bid and sealed-offer auctions. Examples of these institutions in field markets include the application of sealed-bid auctions to assign off-shore oil leases, and sealed-offer auctions to select procurement contractors.

[79] This is found in many studies. *See*, for instance, Fouraker and Siegel, *ibid*; Binger, Hoffman, Libecap, and Schachat (1990); and Morrison and Kamarei (1990).

[80] *See*, for instance, Fouraker and Siegel (1963) and Wellford (1990).

[81] *See* Holt (1985). At the consistent conjectures equilibrium, each firm knows the other's output choice and, given the conjectured reaction of the other firm, any deviation from their own output level would be unprofitable.

> we are in the area of uncertainty when we ask whether mergers that would concentrate a market to only two firms of roughly equal size should be prohibited. My guess is that they should not and, therefore, that mergers up to 60 or 70 percent of the market should be permitted - a figure, curiously enough, that resembles the old Sherman Act..."

Bork also suggests that horizontal mergers that would create no fewer than three significant firms should be considered presumptively lawful. In general, applying these alternative rules of thumb for identifying presumptively collusive markets, the Bork rule outperforms the Scherer rule in laboratory markets.

Over time, average prices in laboratory posted-offer oligopolies are higher than in otherwise identical double-auctions. With more than two sellers however, prices in both the double-auction and posted-offer institutions eventually converge to competitive levels. Alger (1986, 1987) presents some evidence that longer experiments, which allow participants more time to coordinate a collusive strategy, can lead to more anticompetitive outcomes in posted-offer markets. At present it is not clear how to interpret his results in relation to other studies, as he made numerous design changes.[82]

As discussed in the previous subsection, the structural existence of market power in double-auction and posted-offer institutions sometimes leads to its exercise.[83] More often, it appears to increase the likelihood of tacit collusion. However, even when supracompetitive pricing occurs in the laboratory markets studied, the efficiency loss is generally small.[84] The posted-offer markets adjust more slowly than their double-auction counterparts to capture the gains from exchange. However, Smith (1982c) and Davis and Williams (1990b) find that the posted-offer institution mitigates market power and the ability to tacitly collude when the power is on the nonposting (*i.e.*, the buyers') side. Similarly, in *posted-bid* markets, in which buyers post bids and sellers accept or reject them, seller market power and tacit collusion is ineffective. Thus, the absence of a voice (*i.e.*, the ability to post either a bid or an offer) in the market appears to explain many of the behavioral differences observed in posted-offer and double-auction monopolies and oligopolies. Laboratory results suggest that the effect of the institution can dominate the effect of market structure, and thus, influence the likelihood of successful tacit collusion.

b) Opportunities for Explicit Collusion

Several laboratory studies research the impact of allowing open communication among sellers regarding price and other matters. Explicit communication among sellers about price, the "smoking gun" of antitrust, constitutes a *per se* violation of the Sherman Act in naturally-occurring markets.[85] The experiments with explicit opportunities to collude permit situations that would blatantly violate the antitrust laws in field markets. Thus, the laboratory markets provide benchmarks of behavior in a setting without the threat of antitrust enforcement.[86] Different types of permissible communication, such as communications akin to trade press announcements, can also be considered in laboratory markets.

When communication among sellers is permitted between trading periods in laboratory markets, participants quickly decide that price is an interesting topic to discuss.[87]In laboratory markets with open communication, buyers generally are (1) not informed of the sellers' deliberations, and (2) are physically isolated from the sellers. For

[82] Some of the changes include the number of sellers, the information available to sellers (*e.g.* quantity sold by rivals), simulated buyer behavior, the number of periods, the length of periods, the instructions, and the cost and demand configurations (*e.g.* a Walrasian excess supply at the competitive equilibrium). Taking one of the changes for instance, as shown by Brown-Kruse (1986 and 1991), altering the buyer rationing process is not a trivial modification. Thus, one cannot attribute the result of making numerous changes simultaneously to any individual change.

[83] *See* Davis and Williams (1990b, 1991); Holt, Langan, and Villamil (1986); Davis, Holt, and Villamil (1990); and Holt (1989).

[84] *See* Table 3.

[85] *See*, for instance, *U.S. v. Socony-Vacuum Oil Co. et al.*, 310 U.S. 150 (1940), which remains the precedent of *per se* illegality of explicit horizontal price fixing.

[86] The threat of antitrust enforcement can be incorporated in laboratory tests, although it is not commonly implemented. *See* Gerety (1987).

[87] Often the participants are prohibited from revealing their individual unit costs and from making side-payments. *See*, for instance, Isaac, Ramey, and Williams (1984).

example, Isaac, Ramey, and Williams (1984) geographically separate the buyer and sellers. The sellers, located in Arizona, transact with the buyers, who are situated in Indiana, via a computer network.

Even when open communication is permitted in double-auction markets in the laboratory with at least two sellers, prices still converge to the competitive equilibrium. By contrast, in posted-offer markets with at least two sellers, prices tend toward the higher Cournot-Nash equilibrium.[88] Binger *et al.* find that communication facilitates collusion in laboratory Cournot markets, in which sellers set quantity. When communication is permitted in these Cournot experiments, markets with two or five sellers are statistically indistinguishable from the monopoly outcome, *i.e.*, the joint profit maximum is maintained throughout. Brown-Kruse, Cronshaw, and Schenk (1993) show that communication can lead to collusive outcomes in otherwise competitive spatial markets.

Holt and Davis (1990) observe an absence of stable collusive outcomes in posted-offer markets under communications (nonbinding price announcements) of the type traditionally considered in the grey area of antitrust law. This result holds even when market power exists. In other words, in the Holt and Davis experiments, when (1) sellers are provided an opportunity to increase prices unilaterally and profitably by strategically withholding quantity and (2) nonbinding price announcements are made, the sellers do *not* exercise market power.[89]

Davis and Holt (1991) present four experiments showing that market power in posted-offer markets of three and five sellers leads to supracompetitive pricing. This result conflicts with the competitive results Holt and Davis (1990) find in similar markets of three sellers with nonbinding price communication (which would presumably only enhance the erosion of competition). The authors do not compare the results in their 1991 paper or the significant design change that seems to be driving the outcomes. The change has to do with the nature of aggregate production capacity. In the 1990 design, aggregate production capacity exceeded the competitive equilibrium level (as in Figure 5); while in the 1991 study, capacity is constrained at the competitive solution. Thus, it appears that the aggregate capacity constraint promotes the exercise of market power, when it exists. Further research is necessary to complete the analysis of the effect of capacity on unilateral market power, with and without nonbinding communication.

In markets not explicitly designed to investigate unilateral market power, Isaac and Reynolds (1990) find that the nature of aggregate capacity at the competitive equilibrium has a significant impact on market performance in posted-offer markets, even more so than a pure numbers effect of moving from four to two sellers. High excess aggregate capacity at the competitive equilibrium encourages more competitive behavior than the low excess aggregate capacity market designs. This supports the examples from the naturally-occurring economy presented in Scherer (1980, pages 209-212),[90] and contests the Davidson and Deneckere (1990) model of excess capacity as a device (a credible threat to punish defectors) to facilitate collusion.[91]

C. Mergers

Wellford (1990) examines horizontal mergers in laboratory Cournot markets. The study considers the ability of the critical levels of concentration demarcated in the *Horizontal Merger Guidelines* (1992) to screen for anticompetitive mergers in controlled markets.[92] Production efficiencies resulting from merger and methods of measuring concentration are considered in the study.

---

[88] Note that while prices may be higher in posted-offer duopolies when communication is permitted, the prices do not reach the perfectly collusive (monopoly) level. *See* Isaac, Ramey, and Williams (1984), Isaac and Plott (1981), and Smith (1981a).

[89] Before each period, one seller (selected in an announced sequence) is permitted to communicate via a computer network by completing the statement "___ is an appropriate price for the market this period." Other sellers can respond 'A' if they agree, and 'L' ('H') if they think it should be lower (higher). The announcements and responses are displayed to each seller every period.

[90] *See also* the experiments in Sherman (1971); and Brown-Kruse *et al.* (1993), which also shows that when sellers have capacity constraints, market behavior is best described by the Bertrand-Edgeworth "cycling" model. However, while the Bertrand-Edgeworth model predicts better than the mixed-strategy Nash equilibrium, the joint monopoly equilibrium, and the competitive equilibrium, each has some predictive power. None of the models is entirely consistent with the laboratory data.

[91] Moreover, the data from the field markets upon which the Isaac and Reynolds (1990) design is based, (*see* Isaac, Oaxaca, and Reynolds (1988)) are consistent with their laboratory results.

[92] The critical levels of concentration did not change in the revision of the 1984 Guidelines.

Table 4 presents the basic experimental design. The Cournot markets studied have either five symmetric firms (each with 20% of industry capital -- see column I) or eleven firms (three large, three medium, and five small firms, each with 15%, 10%, and 5% of industry capital, respectively -- see column II) before the merger. The initial market supply and demand conditions are identical for both of the industry structures. In the five-firm industries described in column I, merger creates a new entity with 40% of the industry's capital and increases the Herfindahl-Hirshman Index (HHI) of market concentration from 2000 to 2800. Based on the HHI levels, the Guidelines classify these industries as "highly concentrated" and mergers that "potentially raise significant competitive concern" (§1.51(c)). In the eleven-firm industries characterized in column II, merger occurs between two small firms, increasing the HHI from 1100 to 1150. This corresponds to a merger that the Guidelines state would be unlikely to have adverse competitive consequences and ordinarily require no further analysis" (§1.51(b)). The performance of the laboratory markets with eleven firms is indeed competitive. Yet, some competitive improvement is apparent when the merged firm enjoys economies of scale.

In light of the laboratory findings relating to complete versus private information conditions, the experiments are conducted under conditions of private information.[93] As in other laboratory investigations of Cournot markets, each seller chooses the quantity to produce and knows the true market demand curve. The experimental design prohibits entry into the laboratory industries.

Figure 6 presents the series of average prices by period from five Cournot laboratory markets with five symmetric firms premerger and four postmerger. No efficiency gains result due to merger in this experimental treatment. All market participants have experience in similar Cournot markets (without merger), so that they are familiar with choosing production quantities and calculating their profits. Figure 7 provides the average price by period from markets that have the same underlying conditions as those shown in Figure 6, except that the merged firm in each of five market sessions enjoys economies of scale in production.

As Fouraker and Siegel found in their study of Cournot private information triopolies,[94] the data in both Figures 6 and 7 tend to lie between the static competitive and Cournot predictions.[95] No tacitly collusive outcomes are observed in *any* individual market session. The market quickly disciplines sellers who try to cut output. Those who reduce production in one period generally increase it in the next.

When the merged firm experiences no economies from merging, the mean postmerger price across all market sessions in this treatment is not significantly greater than the premerger price. The data suggests that the critical HHI levels established by the Guidelines may be overly strict, especially considering that entry is precluded during, and the threat of antitrust enforcement is absent from, the experiments.

Economies of scale due to merger have a statistically significant impact on industry performance in the markets studied; efficiencies make the markets more competitive. Figures 6 and 7 illustrate the downward pressure on mean price by period when the merged firm experiences increasing returns to scale in production. Moreover, a statistically significant increase in mean total surplus and mean consumer surplus is observed. The efficiency result may have implications for mergers in declining industries, as it is generally believed that mergers in such industries exhibit increasing returns to scale cost structures.

The Guidelines state that concentration may be measured by either capacity or sales. Capacity is a control variable set by the experimenter, sales are observable *ex post*. All of the industries studied exhibit significantly higher levels of concentration as measured by sales than when measured based on capacity or on capital. In fact, when measured by sales, the eleven-firm industry HHI increases from 1900 premerger to 2000 postmerger (compared to 1100 to 1150 based on capacity). This measurement discrepancy is enough to move the behaviorally competitive eleven-firm industries into the range of concentration that the Guidelines classify as raising "significant competitive concern." For the mergers studied, relying on the HHI as measured by sales instead of capacity leads to inappropriately increasing the number of cases to be challenged. The laboratory data indicate that basing the HHI on capacity and altering the policy demarcation line would improve measurement of anticompetitive effects.

[93] *See* §III B(1), *supra*.

[94] *See* §III B(1), *supra*.

[95] It should be noted that these markets structurally differ from the Cournot markets reported in the previous section. Symmetric firm size is not maintained in these merger experiments, but is standard in the other market studies.

D. The Illusive Giant: Predation

Isaac and Smith (1985) set out to find predatory pricing under conditions that theory suggests are amenable to predation, noting that (p. 320), "if such behavior is a human trait we ought to be able to observe it in the laboratory." Their design involves two firms: one large firm (firm A) with "deep pockets" and a cost advantage over the other firm, a small firm (firm B). Both firms are given an up front capital endowment to help offset possible future losses. Firm A receives twice the capital endowment of firm B. Figure 8 gives the demand and supply conditions, specifying seller costs and buyer values. The market is structured so that firm A can exclude B and still experience a positive cash flow. The design sufficiently separates the alternative equilibria, including joint monopoly, dominant firm, Edgeworth cycling, predation, and competition.[96] Predation is not observed in this design with or without sunk entry costs to discourage entry. Satisfying the theoretical assumption of complete information on competitors' costs, in addition to sunk entry costs again proves fruitless, as does attempting to induce rivalistic incentives.[97] The dominant firm equilibrium best describes the data from all the treatments.[98] After unsuccessful attempts at creating conditions suitable for observing predation, Isaac and Smith plant a confederate instructed to consistently price at predatory levels, to verify that firm B can indeed be forced to exit in their design. Exit occurs.

Isaac and Smith also study the effect of imposing various antitrust remedies for predation aimed at markets where predation is theoretically likely.[99] In their experimental design, application of a remedy corresponds to a situation of type two regulatory error, as no predation is observed in these markets. Market performance is more anticompetitive (*i.e.*, higher prices and lower market efficiency) when regulatory sanctions are in place, compared to identical markets without the supposed remedies.

Harrison (1988) runs one experiment using the Isaac and Smith parameters, but makes several design changes. The first relaxes the deep pocket advantage of the incumbent. This likely biases the design against observing predation. The more substantive changes involve several levels of prior subject experience in laboratory monopoly and contestable markets, as well as a multi-market feature. The opportunity cost of participating in one market becomes the foregone profits of not participating in the some other market, rather than a simple constant cost factor as in the Isaac and Smith design. Harrison reports evidence of predation. Although, Rutstrom (1985) initiates a careful attempts to reconcile the design changes between the two studies, she fails to replicate the Harrison result in the several treatments considered.[100] The search for predation continues.

Jung, Kagel, and Levin (1991) explore for predatory pricing in an abstract game theoretic setting. They examine the reputation arguments of Kreps and Wilson (1982) and Milgrom and Roberts (1982). The experimental session each use three monopolists and four entrants in a round of decisions. Each monopolist faces a series of entry threats from the four entrants. The entrants do not know *a priori* whether a monopolist is 'strong' (will punish the entrant by predating) or 'weak.' However, over the course of the round, information on a monopolist's reputation can be collected. Weak monopolists may strategically choose to fight the entrant, to build a reputation

[96] This is to say that from each seller's perspective his profit in one equilibrium is sufficiently different than his profit in another (*i.e. dominance* in payoffs is satisfied not only over the entire experiment, but also across relevant alternatives at a given point in time). For instance, if each seller's earnings only differ by a cent in one equilibrium versus another, it would be unlikely to observe a clean behavioral separation in market predictions when an experiment is replicated.

[97] Inducing rivalistic incentives raises methodological issues, as it move away from the issue of whether predatory behavior is innately a human trait.

[98] The dominant firm equilibrium (*dfe*) entails firm B following firm A's lead in setting price. For the Isaac and Smith design, both firms set the same price of $2.84 in the *dfe* and firm A (firm B) earns $1.99 ($0.51) each period. Firm A earns $3.42 per period at the uncontested monopoly equilibrium, from $1.10 to $1.80 in the competitive range, and no more than $0.88 per period with a predation strategy.

[99] The "remedy" applied prohibits firm A from expanding production for beyond the maximum quantity ever set by the firm for two periods whenever firm B enters. Firm A also faces a semipermanent price reduction regulation, requiring that all its price reductions last for at least five periods.

[100] The Rutstrom paper is an unpublished manuscript which pre-dates the published version of the Harrison note. The paper does not test *all* of the design changes or the possible combinations of treatments implemented by Harrison. In some sense this is the case of cold fission in the economics laboratory. Until other researchers can replicate the data point, it remains shrouded in uncertainty.

of being a strong monopolist, in hopes that other potential entrants decide not to enter. Some predatory behavior is documented.

Unfortunately, it is altogether unclear how the abstract game theory results map into behavior in well defined market institutions. Further work building the bridge between analysis in such abstract settings and in well defined institutions is warranted.

Smith and Isaac make a point that is worth repeating. Current theories and experiments of predation fail to acknowledge the importance of capital. Assets of exiting firms will likely be sold, perhaps at a discount, to new or other potential entrants. This severely dilutes the benefits from a predatory strategy.

E. Entry: A Look at the Contestable Markets Hypothesis

Entry prevails as a key variable in antitrust analysis. The laboratory facilitates the study of entry. Most laboratory studies of entry involve tests of the contestable market hypothesis under various conditions in posted-offer markets.[101] The basic structure of the markets examined in these studies includes two firms with identical cost structures that exhibit declining average cost, with each firm owning capacity sufficient to satisfy the entire market. Thus, the conditions for natural monopoly are satisfied. Both strong and weak forms of the contestable markets hypothesis are considered. The former conjectures that the threat of entry is sufficient to assure that competitive equilibrium outcomes are achieved, while the weaker version predicts that market outcomes will converge toward competitive levels, and lie closer to the competitive solution than to the monopoly prediction.

Coursey, Isaac, and Smith (1984) find that when sunk costs associated with entry are zero, economies of scale in and of themselves are an *ineffective* barrier to entry. In fact, the contestable market hypothesis provides a reasonable prediction of behavior. Four of their six markets quickly converge to the sustainable competitive equilibrium, and the other two gradually approach, but do not reach the competitive range. Over all, the Coursey, Isaac, and Smith data support the weak version of the contestable market hypothesis, yet several of the markets also appear to be consistent with the strong interpretation.

The study includes a *baseline* treatment of markets with a single seller who has the same cost structure as either of the two duopolists in the contestability experimental treatments.[102] Even *without* the threat of entry in the baseline sessions, the unregulated monopolists have difficulty in achieving the monopoly prediction due to buyers strategically withholding demand. Holding back purchases in this design hits the seller at his most profitable units. However, the disciplining effect diminishes over time. Strategic withholding appears at the rate of roughly 9% in the Coursey, Isaac, and Smith monopoly markets, and 1% in the duopoly markets. Thus, the competitive results observed in the duopoly markets arise largely from the contesting discipline of sellers rather than the withholding behavior of buyers. Of course the mere existence of human rather than computerized buyers creates the a credible threat that withholding could occur. Brown-Kruse (1991) examines this using the Coursey, Isaac, and Smith (1984) design, and finds that markets with human, rather than robot, buyers converge faster.

When the sunk costs of entry are positive, Coursey, Isaac, Luke, and Smith (1984, p. 69) find that support for the strong version of the contestable markets hypothesis weakens, "yet the disciplining power of contestability remains impressive," as sustained monopoly pricing is not observed. Data from all twelve of the markets reported are consistent with the weak form of the contestable markets hypothesis. Although, in contrast to the zero entry cost treatment presented in Coursey, Isaac, and Smith (1984), the dynamics exhibit less of a monotone convergence path in seven of the twelve laboratory markets. No evidence of natural monopoly outcomes, market collapse, or sustainable tacit collusion surface.

Harrison and McKee take issue with (1) the use of human buyers in the Coursey *et al.* (1984) studies instead of computerized perfectly demand revealing buyers that are in closer line with what contestable market theory, as presented in Bamoul, Panzar, and Willig (1982), assumes and (2) the simultaneous nature of price choice when the theory involves a sequencing so that potential entrants can make their entry decision based on the current price set by the incumbent. With respect to the first criticism, Brown-Kruse (1991) points out, the fact that human buyers do not fully reveal their willingness to pay is less of a shortcoming of the experiments than in fact a

[101] *See also* section D. on predation above.

[102] A baseline set of market trails is a treatment against which all other treatments are compared. In the Coursey, Isaac, and Smith study, the benchmark against which contestability must be compared is the monopoly case. While the *theoretical* monopoly benchmark is known without conducting any laboratory markets, the *behavioral* benchmark is not.

weakness of the theory to adequately include strategic behavior on the demand side of the market. Taking both points (1) and (2) into account, the Coursey *et al.* experiments can be interpreted as boundary experiments of the contestable market hypothesis, and the theory is in fact fairly robust to the deviations in conditions reported.

When Harrison and McKee examine contestability under the behavioral demand conditions assumed in the theory, and with the Bertrand-Nash sequential choice condition in affect, they find support for the strong version of the contestability hypothesis. However, they fail to compare the two demand conditions *directly* (human versus computerized) or the Bertrand-Nash treatments in isolation, hence compounding the influence of the effects of the conjectures that they advance. In a later study, Harrison, McKee, and Rutstrom (1989) argue that simulation of buyers rather than using human buyers does not significantly affect the results. In fact, both Harrison *et al.* studies digress substantially from the Coursey, *et al.* design.[103] Brown-Kruse (1991) explicitly analyzes the demand conditions with data from experiments designed to address the buyer issue. Her data reveal that *prices are significantly lower when human buyers are present.* The sheer possibility that strategic withholding might occur sufficiently disciplines the market to assure competitive market performance. This supports the findings of Coursey, Isaac, Luke, and Smith (1984), which raised the possibility that buyers' strategic withholding behavior enhances the convergence to the competitive solution. Using the Coursey *et al.* market parameters, Brown-Kruse also tests the contestable markets hypothesis when each duopolist can opt to receive a certain rate of return in an alternate market instead of entering the contestable market. The data reveal no significant difference in adjusted price relative to markets in which the opportunity cost of entry is zero. The effect of sequencing remains to be isolated by holding all other treatment variables constant.

Millner, Pratt, and Reilly (1990) study contestability using a new institution, a real-time posted-offer flow market. In their computerized duopoly markets, sellers may change their price at any point in time, setting a price for a specified *flow* of output per second. The seller with the lowest price for the longest interval makes proportionately more sales the other seller, and costs depend on the rate of production per second. Sellers in this design have identical decreasing marginal cost schedules. Each firm can opt to accept a per second alternate rate of return instead of posting an offer in the flow market. The sunk costs of entry and exit are zero.

Buyers in the flow market are computer simulated to reveal demand fully, purchasing units of the laboratory commodity at their true willingness-to-pay for a given quantity per second. One experimental treatment allows only one seller, who is fully protected from potential entry, and provides a benchmark against which to measure the behavior observed in the theoretically contestable laboratory markets.

Under the real-time continuous setting, contestability ostensibly weakens. However, it is difficult to know to what degree the institution actually affects the outcome, for no direct comparison of institutions is made under the same industry parameters. Millner, Pratt, and Reilly find that their data in the theoretically contestable markets fail to converge to any stable equilibrium, including the sustainable Ramsey-efficient competitive equilibrium. Yet, the market performance is significantly more competitive than in the protected monopoly laboratory markets. Contestability increases welfare. Several of the markets exhibit behavior consistent with the unstable price hypothesis, as introduced in Coursey, Isaac, Luke, and Smith: Prices decline while two firms serve the market, until they fall sufficiently so that one firm exits, after which prices increase until entry occurs and the cycle repeats. In the Millner, Pratt, and Reilly experiments, for the periods in which units were traded, the market was contested by both sellers over 70% of the time. The impact of using human buyers in the flow market remains for later study.

All of the reported contestability experiments involve firms with decreasing marginal cost functions, a common condition of natural monopoly. While many antitrust markets that practitioners evaluate do not involve such specialized cost conditions, it is instructive to observe that the threat of entry indeed imposes competitive pressure on behavior, even in the most potentially problematic types of markets. The pressure increases when human buyers replace the perfectly demand-revealing robot buyers that theory assumes. Little laboratory work exists on entry under conditions of constant and increasing marginal cost, and remains an area for future study of interest to the antitrust community. However, preliminary results of (noncontestable) monopoly posted-offer markets under different cost scenarios and simulated buyers in Harrison, McKee, and Rutstrom (1990) suggest that monopolists exert monopoly power most effectively in constant cost industries, followed by decreasing cost and then

[103] Other treatment variables introduced include: different cost and buyer values, a new computerization of a posted-offer market, altered experience levels. The participants in Harrison and McKee are also prescreened for exhibiting risk neutral behavior using a variation of the method employed in Becker, DeGroot, and Marshak (1964).

increasing cost industries. The underlying behavioral difference may stem from the nature of equilibrium in the three price settings. This remains for later research, but may impact the manner in which theorists and antitrust practitioners think about monopoly, even when entry is perfectly suppressed.

F. Facilitating Practices: Elements of the Ethyl Case in the Laboratory

Grether and Plott (1984) designed experiments to parallel the industry structure and practices in the Ethyl case.[104] Both sides of the case agreed on the general structural form of the lead-based antiknock compound industry. Thus, the design of the laboratory market structure was relatively straight-forward. Features of the market included a homogeneous product, inelastic demand, no entry,[105] two large firms, two small firms, low transportation costs, and large buyers.

While the defense and complaint counsels agreed on industry structure, they argued about the conduct that resulted from the following seller practices: (1) most favored nation clauses (guarantees the buyer that no other customer will obtain a lower price for a similar quantity of the good); (2) advanced price announcements; and (3) delivered pricing (price is not contingent on location). Holding the economic parameters constant across experiments, the practices can be introduced, or omitted, so that the effect of the change can be studied. Delivered pricing is present in all of the Grether and Plott laboratory markets.

The FTC counsel contended that uncertainty about rivals firms' actions or reactions encourages competition, and therefore the most favored nation clauses and the advanced price announcements promote collusion. Grether and Plott find that without the practices, the observed prices lie closer to the competitive level than to any other prediction. Introducing both the most favored nation clauses and advanced price announcements leads to a decrease in competition, however the prices lie approximately halfway between the predictions of the competitive and Cournot models. The joint maximum, the perfectly collusive outcome, is rejected. Additionally, the most favored nation clause in isolation exerts a *negative* impact on price, the opposite result as argued. However, the price announcements do generate the hypothesized effect of increasing prices.

The conclusions of the Grether and Plott research are consistent with other experimental findings, several of which are discussed above. The application of laboratory analysis to the Ethyl case illustrates how experimental analysis can be useful for increasing our understanding of the prominent elements of complex antitrust situations.

IV. Concluding Remarks

The economics laboratory is an important tool which can enhance understanding of how markets work. The first three decades of laboratory research expands the general knowledge of theoretical industrial organization and applied field analysis relating to antitrust. From this literature, several significant themes suggest directions for policy or topics for further study:

- Institutional form matters more than theory and policy imply. In fact, institutional effects often dominate the effects of structural characteristics of the market. (For instance, the competitive outcome is more robust than economic theory suggests, markets with single sellers do not necessarily lead to monopoly outcomes in certain institutional settings.)

- Standard theories of noncooperative equilibria assume that all market participants have complete information on the payout functions of others, an assumption that is likely violated in most field markets. However, these theories actually predict better in laboratory markets under conditions of private rather than complete information.

- Communications of the form the antitrust laws prohibit sometimes lead to collusive outcomes. No support has been found to date that communications in the form of trade press announcements facilitate stable collusion in laboratory markets. Tacit collusion is difficult to coordinate in most

[104] *Ethyl Corp. v. F.T.C., 729 F.2d 128* (1984).

[105] Entry was unlikely, as the EPA was in the process of phasing out the use of lead in gasoline.

settings. When explicit collusion is permitted, the successful implementation of a cartel depends on the form of trading institution: (1) in double-auction markets, collusion usually falls apart; (2) in posted-offer markets, the sellers raise price to the Cournot-Nash level; and (3) in Cournot markets, the sellers sometimes achieve the perfectly collusive outcome.

- Predation is difficult to generate behaviorally in a market context, even under market conditions designed to give the theory its "best shot." Another lesson to be gleaned from research on this topic is that simple game-theoretic models may perform well when tested under the rigid assumptions they dictate however, when the models are imbedded in laboratory market institutions, the models often fail to describe behavior.

- Economies of scale resulting from merger in Cournot laboratory markets are passed on to consumers. The HHI critical values set in the Guidelines are overly restrictive for the markets studied, even when the merging parties do not enjoy economies.

- Market power, when it exists, may be exercised. The effect of institutional form, the nature of capacity, the divisibility of production, and the presence of human buyers are important in determining whether or not it will be. When market power exists and prices are supracompetitive, it is often the case that market power is not exercised fully (in the sense that sellers with power do not withhold production to the extent that theory predicts). Rather, the presence of market power seems to facilitate tacit collusion. Further study is warranted to better understand the relationship of the above variables on the effective exercise of market power, as well as on the incidence of tacit collusion.

Almost all of the results surveyed above involve industries with concentration levels sufficient to "raise significant anticompetitive concern" under the Guidelines (§1.51(b)), when using the Hirfindahl-Hirshman Index (HHI) of concentration to screen for potential horizontal merger cases. Most of the laboratory markets have HHIs of at least 2500 (*e.g.*, four symmetric firms) and many have values of at least 5000 (*e.g.*, duopoly or monopoly). In merger cases litigated by the Federal Trade Commission or the Department of Justice from 1982-1991, roughly 35% of the cases involved industries with HHIs of 2500 or less, 65% of 3333 or less, and 84% of 5000 or less.[106] Thus, the concentration of the markets studied in the laboratory is on average higher than the market concentration in litigated field cases.[107]

It is worthwhile to note that most of the laboratory analysis summarized above omits two features that are critical for antitrust applications: (1) the possibility of entry[108] and (2) the threat of antitrust enforcement.[109] The absence of these factors should bias the data in favor of observing collusive outcomes.[110] Thus, to the extent that collusion is not observed, or when it is, it is unstable, presents a strong result: Even *without* the threat of entry or antitrust penalty, many of the markets are still competitive. It is possible to incorporate both of these factors in

[106] *See* Coate (1992).

[107] This may not necessarily be representative of market concentration in cases that fold or settle prior to litigation.

[108] The most notable exceptions include the literature on predation and contestable natural monopolies in the laboratory, *see* §III (D) and (E) *supra*. In particular, little is known currently about entry in laboratory markets with constant or increasing marginal cost environments.

[109] Gerety (1987) incorporates the threat, *see* note 23 *supra*.

[110] By and large, existing laboratory studies and antitrust analysis assume that the structural parameters (*e.g.* costs and demand) of an industry are constant over time. In this regard, both methodologies likely overestimate the ease of tacit collusion if structural shocks are relevant, for the shocks would likely make coordination more difficult. It is possible to test directly whether the incidence of supracompetitive pricing decreases when the parameters are not constant.

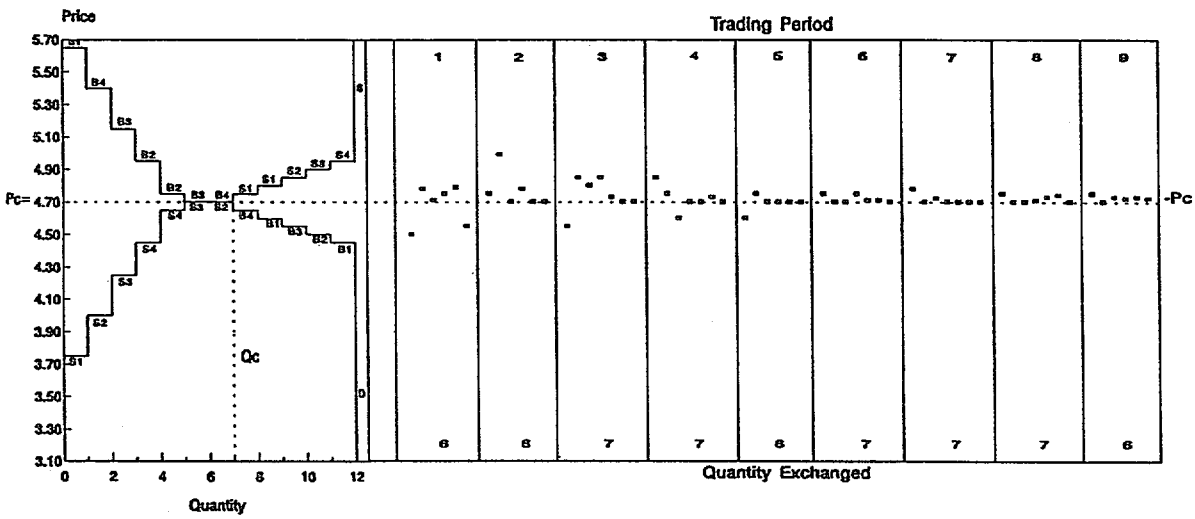
future laboratory studies.[111] Doing so would likely decrease the incidence of anticompetitive market outcomes, although, it remains to be tested.

Undoubtedly, future laboratory research will bring further progress and perhaps help to reach a consensus on how oligopoly markets work. As laboratory evidence accumulates, it will build an empirical foundation for our understanding of markets.

[111] Another feature that could be important is agency theory. Conn, Young, and Bishop (1991) consider principal-agent theory, but not imbedded in an oligopoly framework.

Figure 1*

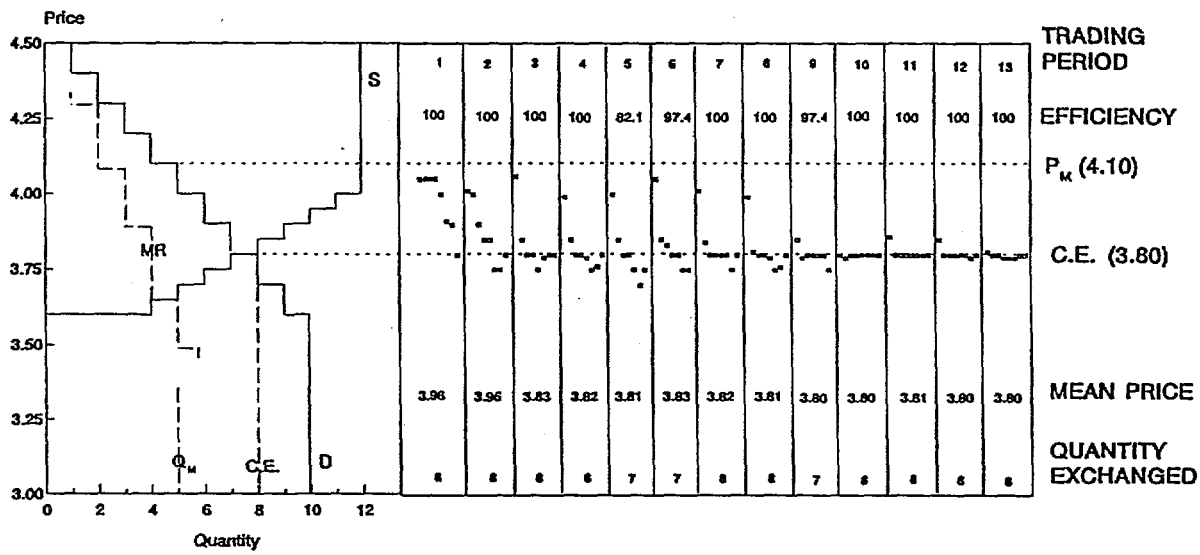
Double Auction 2DA 24



*Source: Ketcham, Smith, and Williams (1984), page 603.

Figure 2*

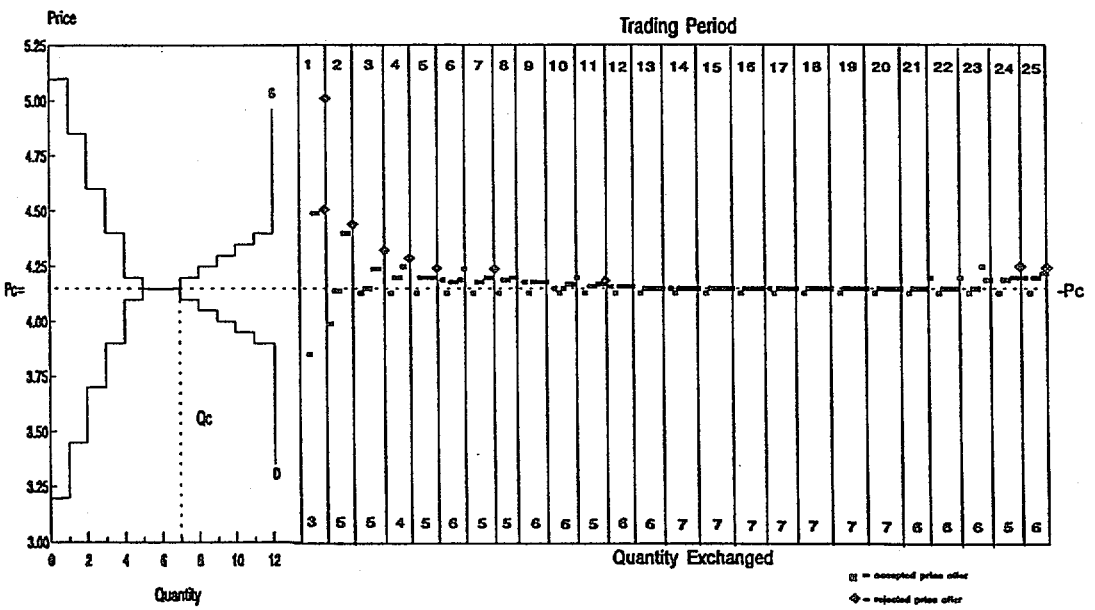
Double Auction Monopoly M1x



* Source: Smith and Williams (1989), page 38.

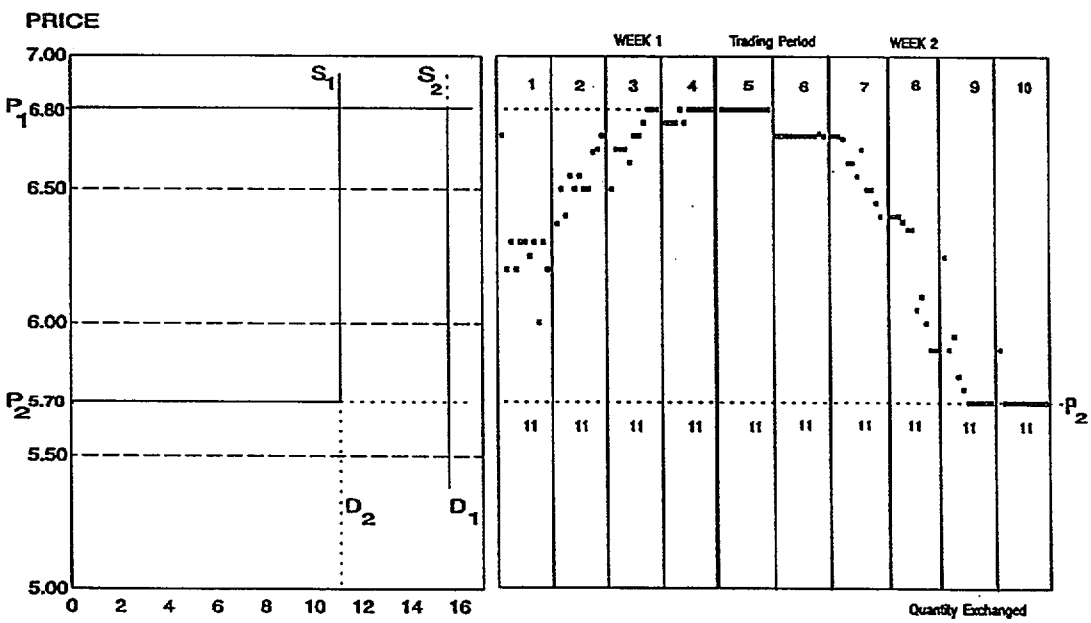
Figure 3*

Posted Offer PO17ixs



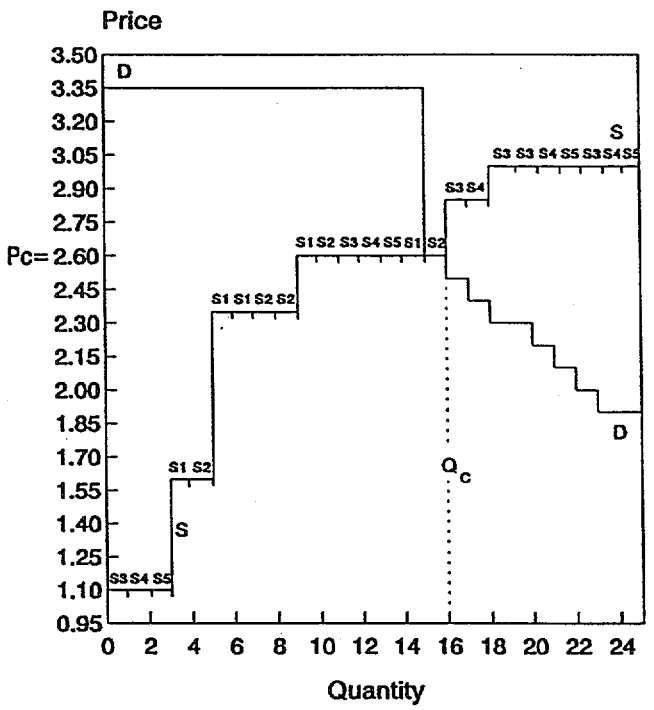
*Source: Ketcham, Smith, and Williams (1984), page 605.

Figure 4*
Double Auction llpda43



*Source: Smith (1982a)

Figure 5*
Market Power Design

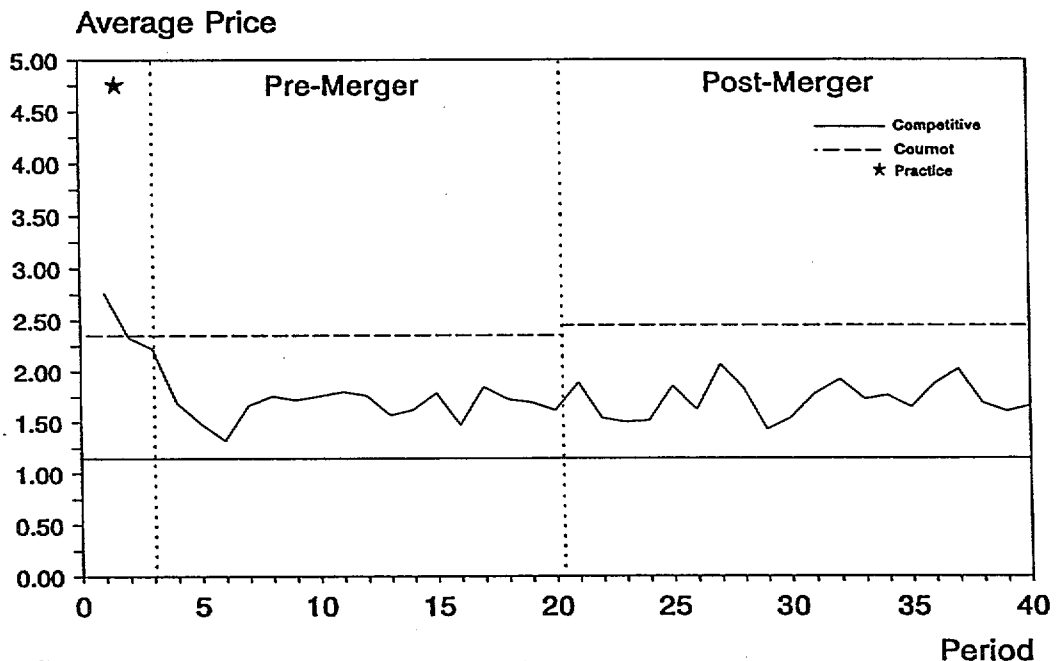


*Source: Holt, Langan, and Villamil (1986)

Figure 6*

Experiments 501 - 505

(No Economies)

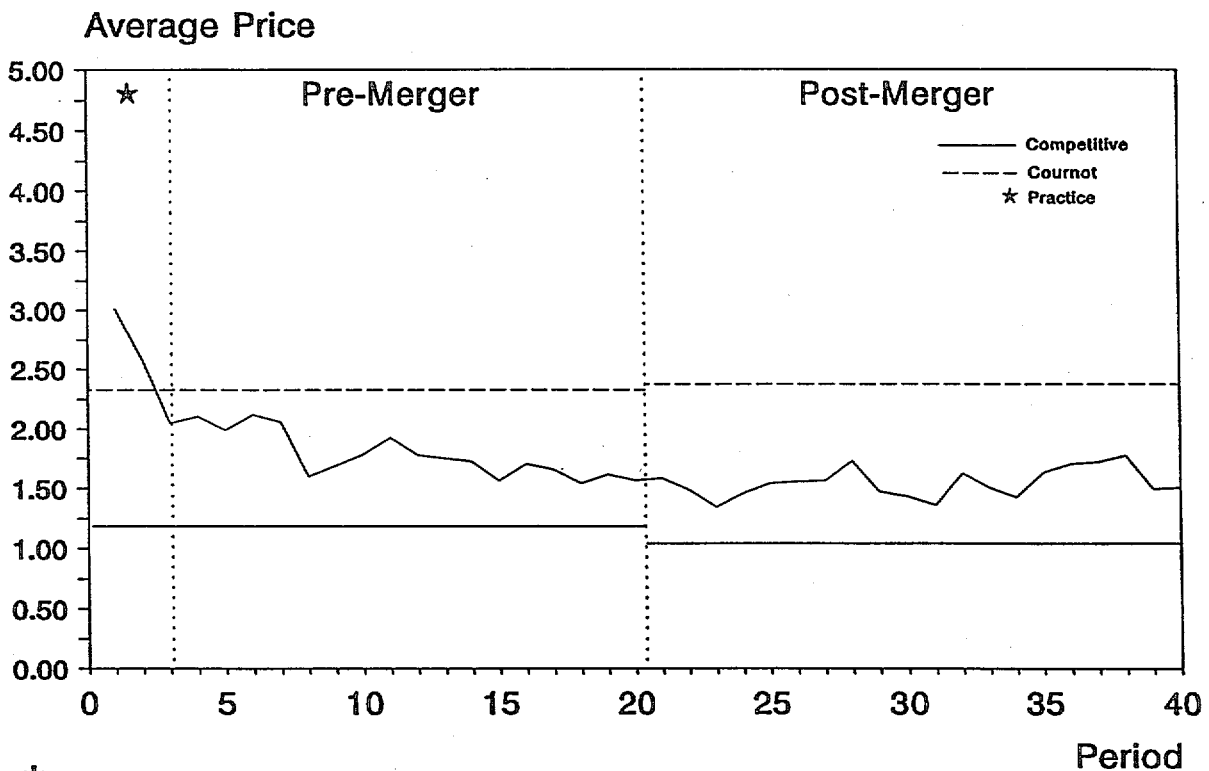


*Source: Wellford, 1992.

Figure 7*

Experiments 506 - 510

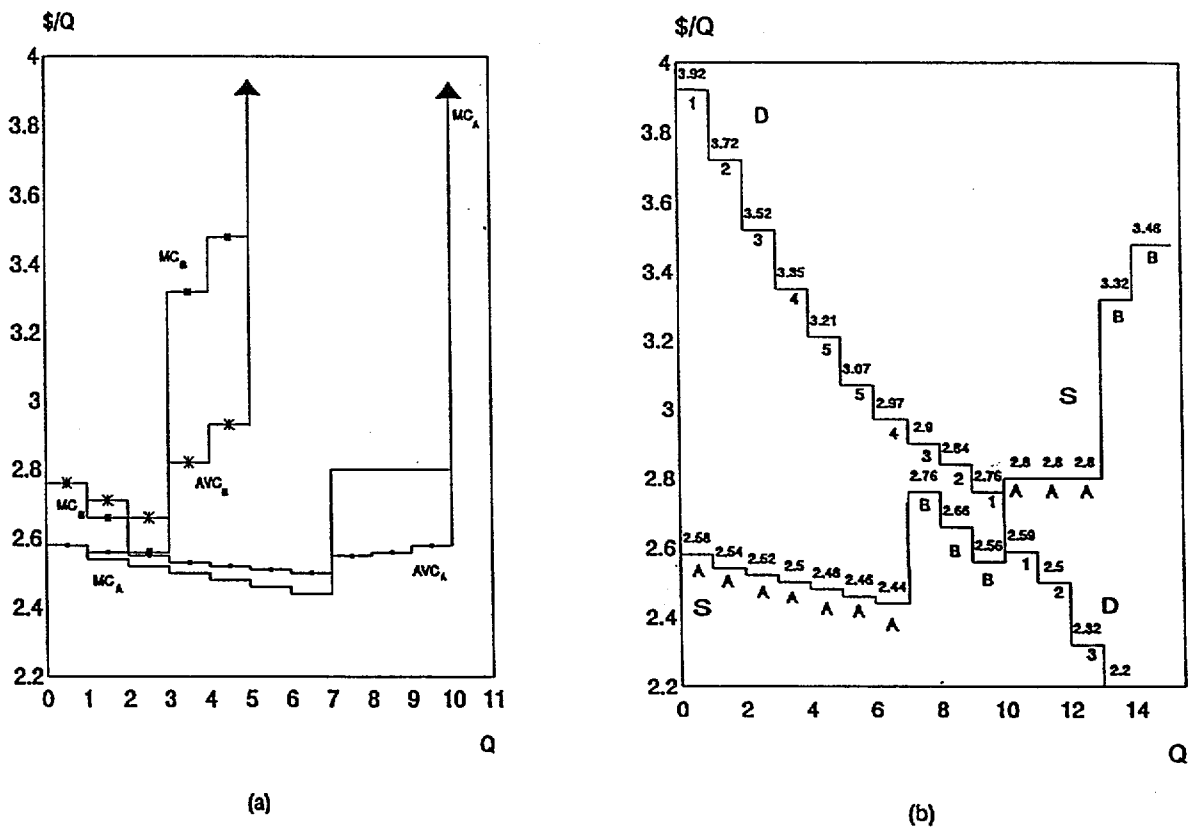
(With Economies)



*Source: Wellford, 1992.

Figure 8*

Cost and Demand Parameters Predatory Pricing Experiments



* Source: Isaac and Smith (1985), page 325.

Table 1

Percentage of Period 21 Market Quantity Observations
Supporting a Particular Equilibrium Concept
Under Private Information Conditions
in Fouraker and Siegel's Cournot Markets*
(*Complete Information Treatments in Parentheses*)

Equilibrium Concept	Duopoly	Triopoly
Joint Max	0.00	0.00
	(31.25)	(0.00)
Cournot	87.50	81.82
	(46.87)†	(45.45)
CE	12.50	18.18
	(21.88)†	(54.55)

* Compiled from pages 131, 134, 140, and 141.

† One observation is equidistant between the Cournot and the Competitive Equilibrium (CE). Each of the two equilibria is given equal weight of 0.5 for this observation in the calculation of the percentages.

Table 2

Percentage of Period 14 Market Price Observations
Supporting a Particular Equilibrium Concept
Under Private Information Conditions
in Fouraker and Siegel's Bertrand Markets*
(*Complete Information Treatments in Parentheses*)

Equilibrium Concept	Duopoly	Triopoly
Joint Max	0.00	0.00
	(35.29)	(0.00)
Bertrand	100.00	100.00
	(64.71)	(100.00)

* Compiled from pages 174, 177, 181, and 183.

Table 3

Laboratory Studies of Market Power

Study	Institution	Excess Capacity at CE	Sellers	Buyers	Trials	Market Performance	Average Market Efficiency
Holt, Langan, and Villamil (1986)	DA	Yes	5†	5	3, 2x	Two of the markets converge to the CE, the others converge to a price slightly above the CE price. On average, the deviations from the CE are observed with higher frequency when the sellers have previous experience in the PO institution.	100% (in period 8)
Davis and Williams (1991)	DA	Yes	5†	5	2, 2x	Supracompetitive prices are observed, but cannot be explained by the strategic withholding of quantity by sellers with theoretical market power.	98%
Friedman and Ostroy (1991)	DA with Divisible Units	In Some Markets	4	4	5x	Strategic withholding is negligible. By strategically misrepresenting their true valuations, market agents unintentionally achieve the competitive outcome.	> 95%
Davis, Holt, and Villamil (1990)	PO	Yes	either 2 or 3	*	2, 2x 2, 2x	Static market power leads to observed supracompetitive prices. Median price is less than (exceeds) the mixed strategy Nash equilibrium prediction in the duopoly (triopoly) markets. Some evidence of tacit collusion exists, but it is imperfect, with price lower than the shared monopoly outcome.	NR
Davis and Williams (1990b)	PO	Yes	5†	5	4x††	Markets with *buyer* market power are unable to overcome the institutional asymmetry, prices converge to the competitive levels.	99%
Holt and Davis (1990)	PO with Non-Binding Price Notices	Yes	3	*	2x	After early success at raising price above the pre-notice level (but not to the shared monopoly level), coordination breaks down and price converges to the pre-notice level.	NR
Davis and Williams (1991)	PO	Yes	5†	5(*)	4, 4x	Prices exceed CE price in all of the market trials with market power and simulated buyers. However, it is not clear whether market power is exercised, for the firms that structurally possess market power do not withhold units from the market. When human buyers are introduced into the analysis, price converges to the CE.	91%* 92%
Davis and Holt (1992)	PO	No	either 3 or 5	*	12x	Supracompetitive prices are sustained throughout all the market power trials. However, the elevated prices seem to be due to tacit collusion rather than the exercise of market power.	NR

CE Competitive Equilibrium; DA Double-Auction; PO Posted-Offer; NR Not Reported; * Simulated buyers, and sellers have full information on the demand curve; x Trials in which participants have prior experience with the institution; † This uses the Holt, Langan, and Villamil (1986) design in which two sellers have substantially greater power than the other three sellers; and †† Four market power trials are summarized with the Davis and Williams (1991) experiments.

Table 4

Experimental Design

Cournot Homogeneous Product Mergers

	I	II
Premerger	5 firms (each with 20% of industry capital) HHI = 2000	11 firms (3 with 15%, 3 with 10%, and 5 with 5% of industry capital) HHI = 1100
Postmerger	4 firms (1 with 40%, and 3 with 20% of industry capital) HHI = 2800	10 firms (3 with 15%, 4 with 10%, and 3 with 5% of industry capital) HHI = 1150

Appendix

SAMPLE DOUBLE-AUCTION INSTRUCTIONS

General

This is an experiment in the economics of market decisionmaking. The instructions are simple and if you follow them carefully and make good decisions you can earn money which will be paid to you in cash at the end of the experiment.

In this experiment we are going to conduct a market in which some of you will be buyers and some of you will be sellers in a sequence of market days or trading periods. Attached to the instructions you will find a sheet labeled Buyer or Seller, which describes to you the value of any decisions you might make. **You are not to reveal this information to anyone.** It is your own private information.

Specific Instructions to Buyers

During each market period you are free to purchase units from any seller or sellers. The first unit that you buy *during a trading period* you will receive the amount listed under unit 1 marked redemption value; if you buy a second unit you will receive the additional amount marked unit 2 redemption value. The profits from each purchase are computed by taking the difference between the redemption value and purchase price of the unit bought. That is,

your earnings = (redemption value) - (purchase price).

Suppose, for example, that you buy two units and that your redemption value for the first unit is 200 and for the second unit is 180. If you pay 150 for your first unit and 160 for your second unit, your earnings are:

earnings from first unit = 200 - 150 = 50

earnings from second unit = 180 - 160 = 20

total earnings = 50 + 20 = 70.

The blanks on the table will help you record your profits. The purchase price of the first unit that you buy during the first period should be recorded on your sheet *at the time of purchase*. You should then record your profit for that unit. At the end of the period record your total of profits for the period in the space provided. Subsequent periods should be recorded similarly.

Specific Instructions to Sellers

During each market period you are free to sell units to any buyer or buyers. The first unit that you sell *during a trading period* you obtain at a cost of the amount listed under unit 1 marked cost; if you sell a second unit you incur the cost listed for unit 2. Your profits from each sale are computed by taking the difference between the sales price of the unit and its cost. That is,

your earnings = (sales price of unit) - (cost of unit).

Suppose, for example, that you sell two units and that your cost of the first unit is 140 and of the second unit is 160. If you sell the first unit at 200 and the second unit at 190, your earnings are:

earnings from first unit = 200 - 140 = 60

earnings from second unit = 190 - 160 = 30

total earnings = 60 + 30 = 90.

The blanks on the table will help you record your profits. The sales price of the first unit you sell during the first period should be recorded on your sheet *at the time of purchase*. You should then record your profit for that unit. At the end of the period record the total of profits in the space provided. Subsequent periods should be recorded similarly.

Market Organization

The market for units is organized as follows. The trading period is open for <u>5 minutes</u>. Any buyer is free to bid and any seller is free to offer at any time that recognition is gained from the auctioneer. The bid (offer) is tendered by giving the sequence: name and bid price (offer price). For example, "buyer 7 bids 120" or "seller 6 offers 210." Recognized bids and offers are written on the board and will remain there until accepted, canceled or replaced by a higher (lower) bid (offer). Anyone is free to accept a standing bid (offer). For example, "seller 6 accepts buyer 7's bid of 120" or "buyer 7 accepts seller 6's offer of 210." Once a bid (offer) is accepted, a binding contract has been closed and both parties record the transaction. Completed transactions will be circled on the board.

Things to note

Always buy (sell) units in order of unit number.

You can only buy (sell) as many units as you have values (costs).

There is no carryover of untraded units across market periods.

SAMPLE SELLERS RECORD TABLE
SELLER # 2

PERIOD 1		UNIT 1	UNIT 2	UNIT 3	
	(A) Cost	4.00	4.70	4.75	
	(B) Sales Price				
	(C) Profit per unit [(B) - (A)]				
	(D) Profit this period				
	(E) Cumulative Profit				

PERIOD 2		UNIT 1	UNIT 2	UNIT 3	
	(A) Cost	4.00	4.70	4.85	
	(B) Sales Price				
	(C) Profit per unit [(B) - (A)]				
	(D) Profit this period				
	(E) Cumulative Profit				

PERIOD 3		UNIT 1	UNIT 2	UNIT 3	
	(A) Cost	4.00	4.70	4.85	
	(B) Sales Price				
	(C) Profit per unit [(B) - (A)]				
	(D) Profit this period				
	(E) Cumulative Profit				

PERIOD 4		UNIT 1	UNIT 2	UNIT 3	
	(A) Cost	4.00	4.70	4.85	
	(B) Sales Price				
	(C) Profit per unit [(B) - (A)]				
	(D) Profit this period				
	(E) Cumulative Profit				

References

Alger, D. (1986), *Investigating Oligopolies Within the Laboratory*, FTC Bureau of Economics Report.

Alger, D. (1987), "Laboratory Tests of Equilibrium Predictions with Disequilibrium Data," *Review of Economic Studies* 54, 105-145.

Ball, S. and P. Cech (1991), "Subject Pool Selection," Kellogg Graduate School of Management Working Paper 91-42, Northwestern University.

Bamoul, W. J., J. C. Panzar, and R. D. Willig (1982), *Contestable Markets and the Theory of Industry Structure*, San Diego: Harcourt-Brace-Jovanovich.

Benson, B. L. and M. D. Faminow (June 1988), "The Impact of Experience on Prices and Profits in Experimental Duopoly Markets," *Journal of Economic Behavior and Organization*, 9, 345-365.

Binger B. R., E. Hoffman, G. D. Libecap, and K. M. Schachat (1992), "An Experimetric Study of the Cournot Model," University of Arizona Working Paper 92-13.

Bolton, G. E. (December 1991), "A Comparative Model of Bargaining: Theory and Evidence," *American Economic Review*, 81(5), 1096-1136.

Bork, R. H. (1978), *The Antitrust Paradox*, New York: Free Press.

Brown-Kruse, J. L. (March 1986), "Laboratory Tests of Buyer Rationing Rules in Bertrand-Edgeworth Duopolies," Revised Manuscript, June 1990, University of Colorado.

Brown-Kruse, J. L. (Spring 1991), "Contestability in the Presence of an Alternative Market: An Experimental Examination," *RAND Journal of Economics*, 22(1), 136-147.

Brown-Kruse, J. L., M. B. Cronshaw, and D. J. Schenk (January 1993), "Theory and Experiments on Spatial Competition," *Economic Inquiry*, 31(1), 139-165.

Brown-Kruse, J., S. Rassenti, S. S. Reynolds, and V. L. Smith (1987), "Bertrand-Edgeworth Competition in Experimental Markets," University of Arizona Discussion Paper 87-12, forthcoming in *Econometrica*.

Carlson, J. A. (January 1967), "The Stability of An Experimental Market With a Supply-Response Lag," *Southern Economic Journal*, 33(3), 305-321.

Carlton, D. W. and J. M. Perloff (1989), *Modern Industrial Organization*, Glenview, IL: Scott, Foresman/Little, Brown Higher Education.

Chamberlin, E. (1948), "An Experimental Imperfect Market," *Journal of Political Economy*, 56(2), 95-108.

Camerer, C. F. (December 1987), "Do Biases in Probability Judgement Matter in Markets? Experimental Evidence," *American Economic Review*, 77(5), 981-997.

Camerer, C. F., G. Lowenstein, and M. Weber (1989), "The Curse of Knowledge in Economic Settings: An Experimental Analysis," *Journal of Political Economy*, 97(5), 1232-1254.

Coate, M. B. (1992), "Economics, the Guidelines and the Evolution of Merger Policy," *Antitrust Bulletin*, Winter, 997-1024.

Coate, M. B., A. N. Kleit, and R. Bustamante (1993), "Fight, Fold, or Settle?: Modeling the Reaction to FTC Merger Challenges," FTC Working Paper 200.

Conn, D., S. M. Young, and R. A. Bishop (1991), "Testing Principal-Agent Theory: A Framework for Analysis and Results From Three Experiments in a Non-Stochastic Setting," University of Southern California School of Business.

Coursey, D., R. M. Isaac, R. M. Luke and V. L. Smith (Spring 1984), "Market Contestability in the Presence of Sunk Costs," *RAND Journal of Economics*, 69-84.

Coursey, D., R. M. Isaac, and V. L. Smith (April 1984), "Natural Monopoly and Contested Markets," *Journal of Law and Economics*, 27, 91-113.

Cox, J. C., Dinkins, and V. L. Smith (1993), "Endogenous Entry and Exit in Common Value Auctions," University of Arizona Department of Economics Working Paper.

Cox, J. C. and R. M. Isaac (1986), "Experimental Economics and Experimental Psychology: Ever the Twain Shall Meet," *Economic Psychology: Intersections in Theory and Application*, A. J. and H. W. MacFadyen, eds., Elsevier Science Publishers B. V. (North-Holland).

Cox, J. C. and R. L. Oaxaca (1990), Using Laboratory Market Experiments to Evaluate Econometric Estimators of Structural Models," Manuscript, University of Arizona.

Cox, J. C., B. Roberson, and V. L. Smith (1982), "Theory and Behavior of Single Object Auctions," *Research in Experimental Economics, Volume* 2, Greenwich, CT: JAI Press, 1-43.

Davidson, C. and R. Deneckere (1990), "Excess Capacity and Collusion," *International Economic Review*, 31(3), 521-541.

Davis, D. D., G. W. Harrison, and A. W. Williams (1993), "Convergence to Nonstationary Competitive Equilibria: An Experimental Analysis," *Journal of Economic Behavior and Organization*, forthcoming.

Davis, D. D. and C. A. Holt (1991), "Capacity Asymmetries, Market Power, and Mergers in Laboratory Markets with Posted Prices," Manuscript, Virginia Commonwealth University.

Davis, D. D., C. A. Holt, and A. P. Villamil (1990), "Supra-Competitive Prices and Market Power in Posted-Offer Experiments," BEBR Working Paper 90-1648, University of Illinois Urbana-Champaign.

Davis, D. D. and A. Williams (1985), "The Effects of Rent Asymmetries in Posted-offer Markets," *Journal of Economic Behavior and Organization*, 7, 303-316.

Davis, D. D. and A. Williams (April 1990a), "Convergence to Nonstationary Competitive Equilibria: An Experimental Analysis," Manuscript, Indiana University.

Davis, D. D. and A. Williams (1990b), "Market Power and the Institutional Asymmetry of the Posted-offer Trading Institution," *Economic Letters*, 34(3), 211-214.

Davis, D. D. and A. Williams (1991), "The Hayek Hypothesis in Experimental Auctions: Institutional Effects and Market Power," *Economic Inquiry*, 29, 261-274.

Demsetz, H. (1973), *The Market Concentration Doctrine*, Washington, D.C.: American Enterprise Institute.

Deneckere, R. and D. Kovenock (1989) "Capacity-Constrained Price Competition when Unit Costs Differ," CMSEMS Discussion Paper 861, Northwestern University.

Dolbear, F. T., L. B. Lave, and G. Bowman (1966), "Collusion in Oligopoly: An Experiment of Numbers and Information," *Quarterly Journal of Economics*, 82(2), 240-259.

Durham, Y., S. Rassenti, V. L. Smith, M. Van Boening, and N. T. Wilcox (1993), "Can Core Allocations Be Achieved in Avoidable Fixed Cost Environments Using Two-Part Price Competition?" presented at the Western Economic Association Meetings, Lake Tahoe, NV, June 22.

Dyer, D. and J. H. Kagel (1992), "Experienced Bidders in Common Value Auctions: Behavior in the Laboratory versus the Natural Habitat," Manuscript, Department of Economics, University of Pittsburgh, presented in the FTC Bureau of Economics Seminar Series, Spring 1992.

Easley, D. and J. Ledyard (1988), "Theories of Price Formation and Exchange in Oral Auctions," Manuscript, California Institute of Technology.

Fama, E. F., and A. B. Laffer (1972), "The Number of Firms and Competition," *American Economic Review*, 62, 670-674.

Fouraker, L. and S. Siegel (1963), *Bargaining Behavior*, New York, McGraw-Hill.

Friedman, D. (March 1984), "On the Efficiency of Double-Auction Markets," *American Economic Review*, 74(1), 60-72.

Friedman, D. (1991), "A Simple Testable Model of Double-Auction Markets," *Journal of Economic Behavior and Organization*, 15, 47-70.

Friedman, D. and J. Ostroy (September 1991), "Competitivity in Auction Markets: An Experimental and Theoretical Investigation," Manuscript, University of California, Santa Cruz.

Friedman, J. W. (1967), "An Experimental Study of Cooperative Duopoly," *Econometrica*, 379-397.

Friedman, J. W. and A. C. Hoggatt (1980), "An Experiment in Noncooperative Oligopoly," *Research in Experimental Economics, Supplement 1*, Greenwich, CT: JAI Press.

Garvin, S. and J. Kagel (December 1990), "An Experimental Investigation of Limit Entry Pricing," Manuscript, University of Pittsburgh.

Gerety, V. E. (1987), *Formation and Control of Collusion in a Sealed Offer Market: An Experimental Examination*, University of Arizona Ph.D. Dissertation.

Grether, D. (1980), "Bayes Rule as a Descriptive Model: The Representativeness Heuristic," *Quarterly Journal of Economics*, 95, 537-557.

Grether, D. (1992), "Testing Bayes Rule and the Representativeness Heuristic: Some Experimental Evidence," *Journal of Behavior and Organization*, 17(1), 31-57.

Grether, D. and C. Plott (October 1984), "The Effects of Market Practices in Oligopolistic Markets: An Experimental Examination of the Ethyl Case," *Economic Inquiry*, 22, 479-507.

Harrison, G. (1987), "Experimental Evaluation of the Contestable Markets Hypothesis," *Public Regulation: New Perspectives on Institutions and Policies*, E. E. Bailey, ed., Cambridge: MIT Press, 191-225.

Harrison, G. (1988), "Predatory Pricing in a Multiple Market Experiment," *Journal of Economic Behavior and Organization*, 9, 405-417.

Harrison G. and M. McKee (Spring 1985), "Monopoly Behavior, Decentralized Regulation, and Contestable Markets: An Experimental Evaluation," *RAND Journal of Economics*, 16(1), 51-69.

Harrison, G., M. McKee, and E. E. Rutstrom (1990), "Experimental Evaluation of Institutions of Monopoly Restraint," *Advances in Behavioral Economics, Volume II*, Green and Kagel, eds., Norwood, N.J.: Ablex, 54-94.

Hayek, F. A. (September 1945), "The Use of Knowledge in Society," *American Economic Review*, 35(4), 519-530.

Hoffman, E. and C. R. Plott (1981), "The Effect of Intertemporal Speculation on the Outcomes in Seller Posted Offer Auction Markets," *Quarterly Journal of Economics*, 96, 223-241.

Hoffman, E. and M. Spitzer (April 1982), "The Coase Theorem: Some Experimental Tests," *Journal of Law and Economics*, 25, 73-98.

Hoggatt, A. C. (July 1959), "An Experimental Business Game," *Behavioral Science*, 4, 192-203.

Hoggatt, A. C., J. W. Friedman, and S. Gill (1967), "Price Signaling in Experimental Oligopoly," *American Economic Review*, 66(2), 261-266.

Holt, C. A. (June 1985), "An Experimental Test of the Consistent-Conjectures," *American Economic Review*, 75, 314-325.

Holt, C. A. (October 1989), "The Exercise of Market Power in Laboratory Experiments," *Journal of Law and Economics*, 32, S107-S130.

Holt, C. A. (1991), "Industrial Organization: A Survey of Laboratory Research," *Handbook of Experimental Economics*, Princeton University Press.

Holt, C. A. and D. D. Davis (December 1990), "The Effects of Non-Binding Price Announcements on Posted-Offer Markets," *Economic Letters*, 34, 307-310.

Holt, C. A., L. W. Langan and A. P. Villamil (1986), "Market Power in Oral Double-auctions," *Economic Inquiry*, 24, 107-123.

Hong, J. and Plott, C. (1982), "Rate Filing Policies for Inland Water Transportation," *Bell Journal of Economics*, 13(1), 1-19.

Hurwicz, L. (1972), "The Design of Mechanisms for Resource Allocation," *American Economic Review, Papers and Proceedings*, 63, 1-30.

Isaac, R. M. (July 1983), "Laboratory Experimental Economics as a Tool in Public Policy Analysis," *The Social Science Journal*, 20(3), 45-58.

Isaac, R. M. (1990), "Laboratory Studies of Monopoly and Conspiracy," Manuscript, University of Arizona.

Isaac, R. M. and C. Plott (1981), "The Opportunity for Conspiracy in Restraint of Trade: An Experimental Study," *Journal of Economic Behavior and Organization*, 2, 1-30.

Isaac, R. M., R. L. Oaxaca, and S. S. Reynolds (1988), "Competition and Pricing in the Arizona Gasoline Market," University of Arizona Discussion Paper 88-16.

Isaac, R. M., V. Ramey, and A. Williams (1984), "The Effects of Market Organization on Conspiracies in Restraint of Trade," *Journal of Economic Behavior and Organization*, 5, 191-222.

Isaac, R. M. and S. S. Reynolds (1989), "Appropriability and Market Structure in a Stochastic Invention Model," *Quarterly Journal of Economics*, 103, 647-672.

Isaac, R. M. and S. S. Reynolds (1990), "Two or Four Firms: Does it Matter?" Manuscript, University of Arizona.

Isaac, R. M. and V. L. Smith (1985), "In Search of Predatory Pricing," *Journal of Political Economy*, 93, 320-345.

Isaac, R. M. and J. Walker (1985), Information and Conspiracy in Sealed Bid Auctions," *Journal of Economic Behavior and Organization*, 6, 139-159.

Jung, Y. J., J. H. Kagel, and D. Levin (1991), "On the Existence of Predatory Pricing: An Experimental Study of Reputation and Entry Deterrence in the Chain-Store Game," Manuscript, University of Pittsburgh.

Kagel, J. and A. Roth, eds.(1991), *Handbook of Experimental Economics*, Princeton University Press.

Ketcham, J., V. L. Smith, and A. Williams (1984), "A Comparison of Posted-Offer and Double-Auction Pricing Institutions," *Review of Economic Studies*, 51, 595-614.

Kirkwood, J. (1981), "Antitrust Implications of Recent Experimental Literature on Collusion," in *Strategy, Predation, and Antitrust Analysis*, S. Salop, ed., FTC Bureau of Economics Report.

Kreps, D. and J. Scheinkman (1983), "Quantity Precommitment and Bertrand Competition Yield Cournot Outcomes," *Bell Journal of Economics*, 14,326-337.

Kreps, D. and R. Wilson (1982), "Reputation and Imperfect Information," *Journal of Economic Theory*, 27, 253-279.

Ketcham, J., V. L. Smith, and A. Williams (1984), "A Comparison of Posted-Offer and Double-auction Pricing Institutions," *Review of Economic Studies*, 51(4), 595-614.

King, R., V. L. Smith, A. Williams, and M. Van Boening (forthcoming), "The Robustness of Bubbles and Crashes in Experimental Stock Markets," *Nonlinear Dynamics and Evolutionary Economics*, I. Prigogine, R. H. Day, and P. Chen, eds., Oxford: Oxford University Press.

Kitch, E. W., ed. (1983), "The Fire of Truth: A Remembrance of Law and Economics at Chicago, 1932-1970," *Journal of Law and Economics* 26, 163-234.

Levin, D. (April 1988), "Stackelberg, Cournot, and Collusive Monopoly: Performance and Welfare Comparison," *Economic Inquiry*, 26, 317-330.

McCabe, K., S. Rassenti, and V. L. Smith (1989), "Lakatos and Experimental Economics," University of Arizona Discussion Paper 89-24.

McGee, J. S. (1971), *In Defense of Industrial Concentration*, New York: Praeger Publishers Incorporated.

Mestelman, S., D. Welland, and D. J. Welland (June 1987), "Advance Production in Posted-Offer Markets," *Review of Economic Studies*, 55(4), 641-654.

Mestelman, S. and D. J. Welland (October 1988), "Advance Production in Experimental Markets," *Journal of Economic Behavior and Organization*, 8, 249-264.

Milgrom, P. and J. Roberts (1982), "Predation, Reputation, and Entry Deterrence," *Journal of Economic Theory*, 27, 280-312.

Miller, R., C. Plott, and V. L. Smith (November 1977), "Intertemporal Competitive Equilibrium: An Empirical Study of Speculation," *Quarterly Journal of Economics*, 91, 599-624.

Morrison, C. C. and H. Kamarei (1990), "Some Experimental Testing of the Cournot-Nash Hypothesis in Small Group Rivalry Situations," *Journal of Economic Behavior and Organization*, 13(2), 213-231.

Plott, C. (1982), "Industrial Organization Theory and Experimental Economics," *Journal of Economic Literature*, 20(4), 1485-1527.

Plott, C. (1987) "Some Policy Applications of Experimental Methods," *Laboratory Experimentation in Economics*, A. E. Roth, ed. Cambridge: Cambridge University Press, 193-219.

Plott, C. (1989), "An Updated Review of Industrial Organization: Applications of Experimental Methods," *Handbook of Industrial Organization, Volume II*, eds. R. Schmalensee and R. D. Willig, Elsevier Science Publishers.

Plott, C. and V. L. Smith (1978), "Experimental Examination of Two Exchange Institutions," *Review of Economic Studies*, 45(1), 133-153.

Posner, R. A. (1976), *Antitrust Law*, Chicago: The University of Chicago Press.

Rassenti, S. J., S. S. Reynolds, and V. L. Smith (October 1989), "Cotenancy and Competition in an Experimental Auction Market for Natural Gas Pipeline Networks," Manuscript, University of Arizona.

Rutstrom, E. E. (December 1985), "In Search of a Reconciliation of Results in Predatory Pricing Experiments," Manuscript, University of South Carolina.

Sargent, T. J. (1992) "Milton, Money, and Mischief," presented at the Western Economic Association Meetings, San Fransisco, CA, Forthcoming, *Economic Inquiry*.

Scherer, F. M. (1980), *Industrial Market Structure and Economic Performance*, Second Edition, Boston: Houghton Mifflin Company.

Schmalensee, R. (January 1976), " An Experimental Study of Expectation Formation," *Econometrica*, 44, 17-41.

Sherman, R. (April 1976), "An Experiment on the Persistence of Price Collusion," *Southern Economic Journal*, 26, 489-495.

Smith, A. (1937), originally 1776, *The Wealth of Nations*, New York: Random House, Modern Library Edition.

Smith, V. L. (1962), "An Experimental Study of Competitive Market Behavior," *Journal of Political Economy*, 70(2), 111-137.

Smith, V. L. (1964), Effect of Market Organization on Competitive Equilibrium," *Quarterly Journal of Economics*, 78(2), 181-201.

Smith, V. L. (1967), Expe^Xxmental Studies of Discrimination Versus Competition in Sealed-Bid Auction Markets," *Journal of Business*, 40, 56-84.

Smith, V. L. (1976a), "Bidding and Auctioning Institutions: Experimental Results," *Bidding and Auctioning for Procurement and Allocation*, Y. Amihaud, ed., New York: New York University Press, 43-64.

Smith, V. L. (1976b), "Induced Value Theory," *American Economic Review*, 66(2), 274-279.

Smith, V. L. (1981a), "An Empirical Study of Decentralized Institutions of Monopoly Restraint," *Essays in Contemporary Fields of Economics in Honor of Emanuel T. Weiler (1914-1979)*, G. Horwich and J. Quirk, eds., Purdue University Press, 83-106.

Smith, V. L. (1981b), "Theory, Experiment, and Antitrust Policy," *Strategy, Predation, and Antitrust Analysis*, ed. S. Salop, FTC Bureau of Economics Report.

Smith, V. L. (April 1982a), "Markets as Economizers of Information: Experimental Examination of the 'Hayek Hypothesis'," *Economic Inquiry*, 20, 165-179.

Smith, V. L. (December 1982b), "Microeconomic Systems as an Experimental Science," *American Economic Review*, 72, 923-955.

Smith, V. L. (1982c), "Reflections on Some Experimental Market Mechanisms for Classical Environments," *Choice Models for Buyer Behavior: Research in Marketing*, Supplement 1, Greenwich, CT: JAI Press, 13-47.

Smith, V. L. (1987), "Experimental Methods in Economics," *The New Palgrave*, volume 1, J. Eatwell, M. Milgate, and P. Newman, eds., New York, NY: W. W. Norton and Company, 94-111.

Smith, V. L. (1988/1989), "New Directions for Economics," *Journal of Business Administration, Special Issue: Future Directions for Economics*, 18(1,2), 41-52 (an earlier version was presented on February 28, 1986 at the first official meeting of the Economic Science Association, Tucson, Arizona).

Smith, V. L. (Winter 1989), "Theory, Experiment and Economics," *Journal of Economic Perspectives*, 3(1), 151-169.

Smith, V. L. (1990), "Experimental Economics: Behavioral Lessons for Microeconomic Theory and Policy," Nancy L. Schwartz Memorial Lecture, Kellogg Graduate School of Management, Northwestern University.

Smith, V. L. (1991a), *Papers in Experimental Economics*, Cambridge: Cambridge University Press.

Smith, V. L. (1991b), "Game Theory and Experimental Economics: Beginning and Early Influences," University of Arizona Working Paper 91-5.

Smith, V. L. (1991c), "Rational Choice: The Contrast between Economics and Psychology," *Journal of Political Economy*, 99, 877-897.

Smith, V. L. (July 1992), "Experimental Economics: Evaluating Microeconomic Policies and Institutional Designs," Manuscript, presented in the Economic Analysis Group Seminar Series, Department of Justice Antitrust Division.

Smith, V. L. (November 1993), "Economics in the Laboratory," *Journal of Economic Perspectives*, forthcoming.

Smith, V. L. and J. Walker (April 1993), "Monetary Rewards and Decision Cost in Experimental Economics," *Economic Inquiry*, 31, 245-261.

Smith, V. L. and A. Williams (June 1981), "On Nonbinding Price Controls in a Competitive Market," *American Economic Review*, 71(3), 467-474.

Smith, V. L. and A. Williams (September 1982), "The Effects of Rent Asymmetries in Experimental Auction Markets," *Journal of Economic Behavior and Organization*.

Smith, V. L. and A. Williams (September 1989), "The Boundaries of Competitive Price Theory: Convergence, Expectations, and Transaction Costs," *Advances in Behavioral Economics, Volume I*, L. Green and J. Kagel, eds., Norwood, N.J.:Ablex, 31-53.

Smith, V. L., A. Williams, K. Bratton, and M. Vannoni (1982), "Competitive Market Institutions: Double-Auctions versus Sealed Bid-Offer Auctions," *American Economic Review*, 72(1), 58-77.

U.S. Department of Justice (1984), "Merger Guidelines," *Federal Register*, 49, 26284.

U.S. Department of Justice and Federal Trade Commission (April 2, 1992), "Horizontal Merger Guidelines," *Antitrust Trade and Regulation Report*, 1559.

Van Boening, M. V. and N. Wilcox (1992), "A Fixed Cost Exception to the Hayek Hypothesis," presented at the Economic Science Association Meetings, October 24.

Vickrey, W. (1961), "Counterspeculation, Auctions, and Competitive Sealed Tenders," *Journal of Finance*, 16(1), 3-37.

Wellford, C. P. (1989), "Price Dynamics and Expectations in the Cobweb Model: An Experimental Analysis," University of Arizona Discussion Paper 89-15, Revised Manuscript, 1991, Federal Trade Commission.

Wellford, C. P. (1990), "Horizontal Mergers: Concentration and Performance," Chapter 2 of University of Arizona Ph.D. Dissertation.

Wilcox, N. T. (1989), "Well-Defined Loss Metrics and the Situations that Demand Them," presented at the Economic Science Association Meetings, Tucson, AZ, October 28-29.

Wilcox, N. T. (1992), "Incentives, Complexity, and Time Allocation in a Decision-Making Environment," presented at the Public Choice/Economic Science Association Meetings, New Orleans, LA, March 27-29.

Williams, F. (1973), "The Effect of Market Organization on Competitive Equilibrium: The Multiunit Case," *Review of Economic Studies*, 40(1), 97-113.

Williams, A. (1979), "Intertemporal Competitive Equilibrium: On Further Experimental Results," *Research in Experimental Economics*, vol. 1, V. L. Smith, ed., Greenwich, CT: JAI Press.

Wilson, R. (1987), "On Equilibria of Bid-Ask Markets," *Arrow and the Ascent of Economic Theory: Essays in honor of Kenneth J. Arrow*, G. Feiwel, ed., London: McMillan Press, 375-414.

The Marketable Emissions Allowance

Steven R. Elliott, Oak Ridge National Laboratory

Abstract

This paper examines the results of experiments designed to study various aspects of tradable emissions permits as set forth by the Clean Air Act Amendments of 1991. In particular, this paper develops an understanding of some of the results that we attained in our experimental work in this area at the Universities of Arizona and Colorado. In these experiments subjects acted as managers of firms that must trade emissions permits. After an initial allocation of permits, subjects were allowed to trade them in both an "informal" double auction and a "formal" revenue neutral auction or in some sessions just the "formal" auction. Further, some sessions allowed subjects to hold permits across rounds for later use (intertemporal banking of permits), while others did not.

The results show, when faced with only one arena in which to trade permits, subjects are able to lock on the efficient trading price and come close to the efficient intertemporal allocation of permits. However, when we introduce the double auction mechanism to a setting with banking, results change dramatically. Now there are arbitrage possibilities across markets and permits become an asset. Price becomes very unstable across markets and periods, and it may show signs of bubbling and crashing as other assets have done in experimental markets. Further research in this area is outlined to understand more of the dynamics of these influences on the efficient solution predicted by theory.

I. Introduction

The 1991 clean Air Act Amendments (CAAA) created a system of tradeable emissions permits for the reduction of SO_2. The participants in this new system are large electricity generation facilities in the U.S.. By initiating such a system, proponents argue that reductions in emissions can be achieved at the least cost to society as a whole. The idea is to introduce market based solutions to what had traditionally been relegated to systems of command-and-control. This is the first test on such a scale of this type of incentive-based regulatory system. Its success or failure will have a profound influence on the adaptation of such systems in the future.

The trading system that has been mandated by the U.S. Environmental Protection Agency (USEPA) has three separate opportunities for permit transactions (see Federal Register 1991a and 1991b). The first mechanism is a formal auction process using a revenue-neutral auction mechanism. This system allows firms to submit both bids to buy and offers to sell permits. The bids are arrayed from highest price to lowest to form a demand curve for permits while the offers are arrayed from lowest price to highest to form a supply curve. In standard fashion, all those whose bids and offers are to the left of the intersection of the supply and demand curves trade. Those to the right are shut out of that particular trading opportunity. The auction is discriminative price- and revenue-neutral so that each purchaser of a permit pays exactly what he or she bids, and the auctioneer (in this case USEPA) does not receive any of the auction's proceeds. In the mechanism specified, the highest bidder purchase the permit from the agent making the lowest offer, and pays exactly what he or she bids. Therefore, the lowest offer receives the amount paid by the highest bidder and so on up and down the respective schedules.[1] This formal market is to occur once a year and will be the arena in which the permits withheld at the annual allocation are traded[2].

The second trading opportunity is an informal process. In this case, agents are allowed to negotiate individual deals for permits at any time. Such transactions have occurred already with the Tennessee Valley authority as principle purchaser of permits (Charlier, 1992).

The final opportunity for permit transactions is direct purchase of permits from USEPA. In this situation, agents contact USEPA to purchase permits at some posted price. The price is designed to be significantly higher

[1]As might be expected this type of incentive structure is not incentive compatible, and may create opportunity for profitable strategic bidding and offering. The reader is referred to Cason, 1992 for a full theoretical development of this auction's strategic characteristics.

[2]The units withheld from the allocation enter as permits with a zero offer price. Their revenue is distributed equal on the bases of the number of units actually withheld.

than the price established in the other trading markets. Agents will see this as an option of last resort that is used very infrequently.

Permits have a life of one year but can be banked. That is, once a permit is used to cover SO_2 emissions at a plant, it cannot be used again. However, permits need not be used in the year they are issued. Permits can be held or "banked" until such time in the future that the firm wishes to use or sell them. This introduces an intertemporal aspect to this emission right system.

While there has been much theoretical work in the past describing, and analyzing a system of marketable permits (see Montgomery, 1972 and Dales, 1968), very little literature has addressed the hybrid system that is currently in place. Instead this new program is being analyzed within the controlled laboratory environment of experimental economics.

In an initial study before the legislation was finalized, the discriminative price revenue-neutral auction mechanism was tested. That work focused on the size of the mandatory contribution to this market, and the effects of using a discriminative price versus a more well understood uniform price. The results of that work (Franciosi et al.,1992) added to our understanding of some of the fundamentals of this novel system.

This paper reports the results of two experimental economic investigation into the workings of the mandated pollution permit market. One study focuses on the formal revenue-neutral auction. While other contains both this formal auction and the opportunity for informal trades. As is explained in grater detail below, the informal market takes the form of a double auction. Each study allows for banking of permits.

The results show that one mechanism alone can be effective in realizing much of the gains from permit trade. Yet, by adding the second trading opportunity, the results begin to diverge from our equilibrium forecasts, and gains from trade fall substantially. Further, effective use of banking may be hindered by this complication.

It is important to note at this time that the comparison of these results is qualitative only. While both studies used the same experimental parameters, they used different procedures, so a rigorous statistical procedure is not possible. A study already underway will attempt to replicate these results using a standard procedure and parameter set. In this way we will be able to investigate formally the results reported in the present paper.

The remainder of the paper is organized in three sections. Section II briefly describes the design and procedures used in these experiments. The next section will turn to the results. Here we will focus on the issues of permit price fluctuation and system efficiency. The final section draws some tentative conclusions and outlines future analysis in this area to help understand the dynamics at work within this complicated institution.

II. Experimental Designs

These experiments were conducted in two different locations using two different procedures. The University of Colorado's Laboratory for Economics and Psychology (LEAP) undertook experiments that focused on the revenue-neutral auction. The University of Arizona's Economic science Laboratory (ESL) created an environment that contained the informal market and a simplified revenue-neutral auction. In the remainder of this section I focus on a short description of the procedures used by the specific experiments at each location. One important feature that will be discussed is the training procedure used at both locations, as these may have some influence on the results. For a detailed discussion on the exact procedures used see Franciosi et al. 1992 and Brown-Kruse and Cronshaw 1992.

Revenue-Neutral Auction Only

In these experiments run at LEAP, subjects acted as the managers of revenue-maximizing-rate-of-return regulated firms. That is, at the beginning of each round each firm was given a fixed revenue and its profit depended on its cost reduction behavior. Also at the beginning of each round, firms were given an allocation of emission permits. The allocation in these series of experiments was the same for all firms (6 permits to start) with the allocation being cut in half in round 5 to simulate the change that occurs in the late 1990's in the USEPA's regulation. Subjects knew in advance the length of the experiment and each period's allocation.

For these experiments firms would produce, as a matter of course, some emissions that had to be cleaned up. Firms could chose to use permits to cover the emissions produced or pay to clean them up. There were five firm types which varied in the amount of emissions and cost to clean them up. This gave incentive for firms with low emissions to sell permits to high emission firms. Table 1 contains the parameter set used in these experiments, including the value of permits to firms and their initial allocation.

Once the initial allocation of money and permits was made, subjects made decisions about use of permits. This involved deciding how many permits to buy or sell, and how many to use or bank. At the end of the round new balances were calculated and added to the subject's total. At the end of the experiments, subjects were paid in cash.

For this series of experiments subjects first participated in a "banking only" experiment. In this case, the initial allocation of money and permits was made, and the only choice subjects faced was how many permits to bank or use; there was no trading between subjects. The purpose was to introduce the subject to the somewhat complex concept of banking without including the further complexity of the auction. All subjects participating in this training session were paid as usual, but only those who made over a specified amount of the optimal profit were invited back for the full session.

The experienced subjects who were brought back faced the same experiment as their trainer expect it now featured both banking and the revenue-neutral auction. Then when the initial allocation of permits was made, some were withheld (see Table 1 for initial allocation and withholding quantity). Subject were then asked to entered their bids for permits and then were asked to enter their offers to sell. In this way, subjects could operate on both sides of the market and operate as true traders in permits.

Once all bids and offers had been entered, the computer ranked them as described above, and notified subjects as to the number of permits they bought and sold and of their net profit. Subject were given time to study this information before they had to decide on the number of permits to use or bank. Once all subjects had made and entered their banking decisions, a new allocation was made, and the auction was opened up. At the end of the experiment all permits had to be used or fell to a zero value. Subjects again were paid in cash at the conclusion.

Table 1
Experimental Parameters

	Firm No.				
Unit No.	1,6	2,7	3,8	4,9	5,10
1	3.80	15.20	7.60	19.00	22.00
2	3.40	13.60	6.80	17.00	20.40
3	3.00	12.00	6.00	15.00	18.00
4	2.60	10.40	5.20	13.00	15.60
5	2.20	8.80	4.40	11.00	13.20
6	1.80	7.20	3.60	9.00	10.80
7	1.40	5.60	2.80	7.00	8.40
8	1.00	4.00	2.00	5.00	6.00
9	.60	2.40	1.20	3.00	3.60
10	.20	.80	.40	1.00	1.20
Allotment (forced)					
Prd 1-4	6(2)	6(2)	6(2)	6(2)	6(2)
Prd 5-12	3(1)	3(1)	3(1)	3(1)	3(1)

Two Permit Markets

As noted, the procedure used at the ESL varied from that used at LEAP. This set of experiments utilized software designed by ESL for the Department of Energy for their continuing research effort in the area of marketable emissions permits. Subjects were told that the permits could be redeemed for cash. That is, instead of acting as firm managers, these subjects were simply utility-maximizing individuals. As in the other experiments, permits were allocated at the beginning of each round. Subjects knew the allocation for the entire experiment, and the redemption value for permits.

Training sessions for this series of experiments were also conducted independently of the sessions reported here. In these sessions, subjects went step by step through the computerized, on-screen instructions. At the conclusion of the instructions, subjects were allowed to participate in three rounds of the experiment to see how the procedures worked first hand. After these practice rounds had been completed a written examination about the procedures in the experiment was given to the subjects. Those subjects who scored greater then 60% were asked to be in the regular sessions. All subjects participating in this training session were paid a flat fee for their time.

The sessions using these experienced subjects again began with the computerized instructions. After subjects had reviewed the instructions the sessions started with an allocation of permits. Again some permits were withheld for mandatory participation in the auction. Once the allocation was made, an informal trading opportunity was opened. In these sessions this informal market was a double auction. This institution was chosen for its well-known properties of conversion to the competitive equilibrium. This market was open for 3 minutes, allowing all subjects to buy and sell permits as they liked.

At the close of the market, subjects were asked by the computer if they would like to voluntarily contribute more units to the auction that was coming up. In this set of experiments, there was no opportunity for subjects to put a reservation price on contributed units as in the LEAP experiments. After all voluntary contributions were registered by the computer the auction was opened. Here subjects were asked to indicate to the computer their bids to buy permits. Once all bids had been entered, the computer ranked bids from highest to lowest, and distributed them to those with winning bids. It also reduced their balance by the amount of their bid. Because the auction was still revenue-neutral, the proceeds were distributed on a per unit basis to all those who had units forced into, or voluntarily contributed to, the auction. As noted, this was a more simpler procedure than either the one used in LEAP or mandated by the USEPA rules. However, it was agreed that the complexity of the experiment necessitated this compromise.

After the results of the auction were posted, subject were asked to consult their individual redemption values to consider how many units they wished to redeem for cash and how many they wished to bank for the future; balances are adjusted accordingly. At the end of the experiment subjects are again paid in cash.

Thus we have results from one set of experiments that utilize only the revenue-neutral auction, and one that combines this mechanism with an opportunity to trade within a double market institution. By including this new mechanism we have introduced the opportunity to arbitrage across markets. However, as will become obvious this new feature may create problems for the efficient operation of this system of marketable permits.

III. Results

As noted in the introduction, the focus of this paper is on the prices of the traded permits and the overall efficiency of the system. There are of course other important points that should be considered such as the optimal banking pattern, the number of permits used (or redeemed) each period, and the bid and offer pattern. These topics can and will be addressed in later papers and are discussed in some detail in both the Franciosi, and Brown-Kruse and Cronshaw papers.

Figures 1-3 are representative of the results of the auction-only experiments. The lowest trading price is the last accepted bid in the revenue-neutral auction, and the mean trading price is the average price bid on, and paid for, permits. The adapted CE price is a measure of the expected competitive equilibrium given the banking pattern within the given experiment. That is, instead of comparing these price measures with a competitive equilibrium based on optimal banking patterns ($71), the adaptive CE is based on the previous banking patterns observed in the particular experiment and is a dynamic measure across the life of the experiment. It was hypothesized that subjects made price decisions based not on some expectation of optimal banking established in period one, but on their current behavior and that of the other subjects.

As can be observed from the figures, prices seem to converge to the adaptive CE price. Further, as the

FIGURE 1

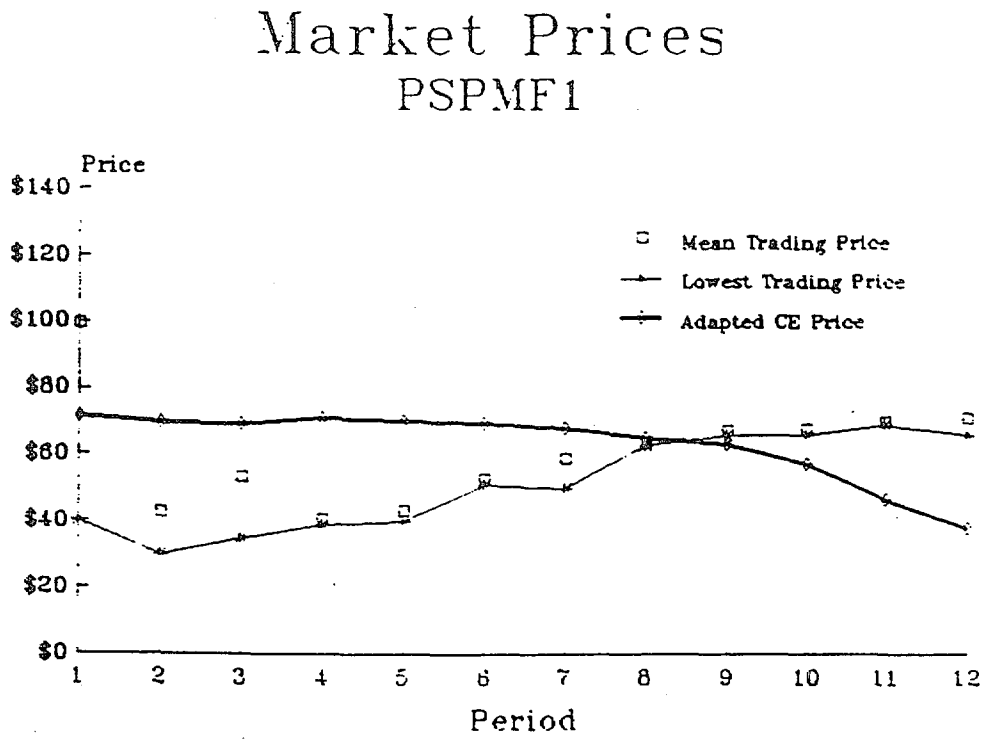

Market Prices
PSPMF1
Price
$140
$120
$100
$80
$60
$40
$20
$0
Mean Trading Price
Lowest Trading Price
Adapted CE Price
1
2
3
4
5
6
7
8
9
10
11
12
Period

FIGURE 2

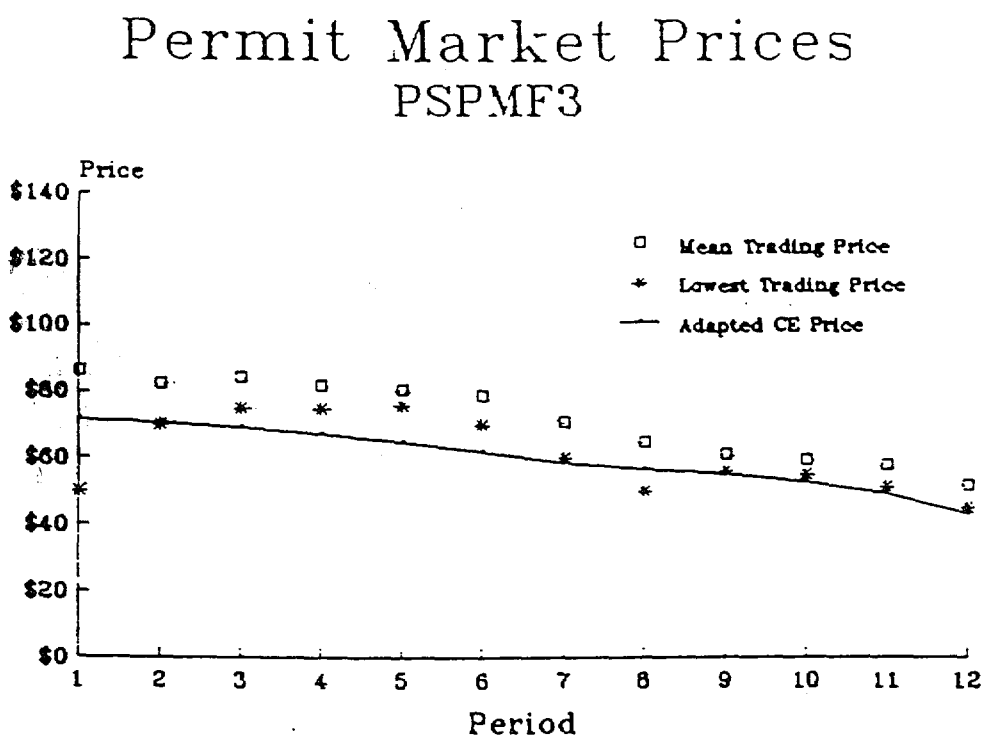
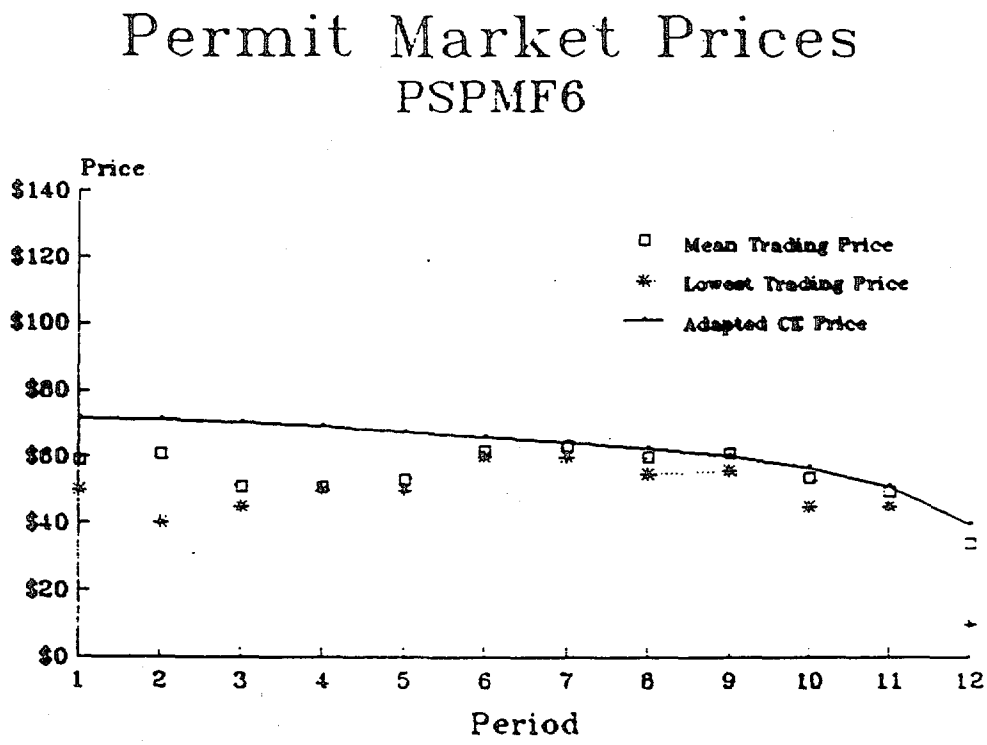

Permit Market Prices
PSPMF3
Price
$140
$120
$100
$80
$60
$40
$20
$0
Mean Trading Price
Lowest Trading Price
Adapted CE Price
1
2
3
4
5
6
7
8
9
10
11
12
Period

FIGURE 3


Permit Market Prices
PSPMF6
Price
$140
$120
$100
$80
$60
$40
$20
$0
Mean Trading Price
Lowest Trading Price
Adapted CE Price
1
2
3
4
5
6
7
8
9
10
11
12
Period

experiment progresses, the bids tighten. The variation across winning bids becomes very small; the mean and the lowest trading prices converge. The falling CE price is the result of the excessive number of permits banked (thereby making subsequent purchases less important). Even if we ignore this CE price, we can see that the price of traded permits does not move too far from the original expected price of $71. The system seems to operate effectively and efficiently.

Another important observation is that there is no price fluctuation when the allocation of permits is cut in half at the start of period five. This would indicate that banking is serving its purpose by helping to stabilize the prices in this market. This is consistent across all the sessions. Thus, prices in under this revenue-neutral auction seem stable across time and tend to converge to the adaptive CE price.

However, when we examine the results from the experiments with both the auction and the market we see a completely different story. Figures 4-6 show much of the same data as the previous figure. Again we can see the mean trading price, and in this case, the highest reject bid (the one right below the lowest accepted bid). Further, the solid connected points represent the contract price established in the double auctions. The dashed line at $7.10 represents the experiment's equilibrium permit price with banking, while the solid line is the equilibrium price with each round[3].

As can be observed, prices are unstable relative to the earlier results. We no longer see the convergence of the mean bid price and the lowest rejected bid. In fact there is quite a pronounced divergence. Also there is a definite effect when the allocation is reduced in period five. Here, at least in the Figures 4 and 5, we see a sharp increase in both the prices in the auction and in the market. It is possible that this pattern could be a "bubble and crash" phenomenon that has been reported in other experimental asset markets (Smith et al. 1988).

Further there is a significant divergence in the prices between the auction and the market. This would mean that it would be possible to buy a permit under one institution and sell in another at a profit. This is especially pronounced in Figure 6. Even though the market price seems to converge early to the banking equilibrium, it is well above the auction prices. It seems that the arbitrage is not working. By adding the second trading opportunity, the system seems to inherit a number of non-optimal characteristics.

This is reenforced when we examine the efficiencies within the experiments. Here we define efficiency as a percentage of the optimal profit realized across the entire experiment. As would be expected, the average efficiency in the auction-only experiments was quite high, at just over 80% across all the experimental sessions. This compares with efficiencies ranging from 7.5% to 47% in the auction and market experiments. Somehow, the introduction of this new market drastically affects the operation of the system in a negative way.

IV. Conclusions

It would seem that the introduction of a new trading institution dramatically effects the outcome in this permit market. Prices become wildly unstable, not only across time, but across markets in a period; we have no indication of an optimal social price for permits. Realizations of gains from trade are lost as banking becomes very erratic. In effect a system that operates near its expected optimum in the one-market case seems to fly apart when we introduce this double auction.

What is it about the introduction of this second trading opportunity that so adversely affects these results? The answer is not easy or obvious. It is important, before postulating too extensively to remember that these two sets of experiments used two different procedures -- not only to run the experiments but to create the experienced subject pool used. It will be important to study this in a more systematic approach by using one procedure both for data collection and training.

Also, it will be important to use a parameter set that is designed for this type of study. If we look at other results from the two-market experiments using a different data set, we see much higher efficiencies, similar to those seen in the one-market sessions. Of course these experiments still yielded prices that were similar to those in figures 4-6. But the radical and consistent fall in efficiencies seen with the introduction of the double auction may be caused by some quirk in the parameters.

Experiments are being designed to look at these issues. They will employ the software used in the two market experiments. A completely new parameter set also will be created. Further, subjects not from either location will be used to control the unlikely event there is some kind of subject pool effect. There experiments will

[3]The prices in the two experimental designs are stated in different units, but are comparable.

FIGURE 4

MARKET AND AUCTION PRICES

R10_12DC

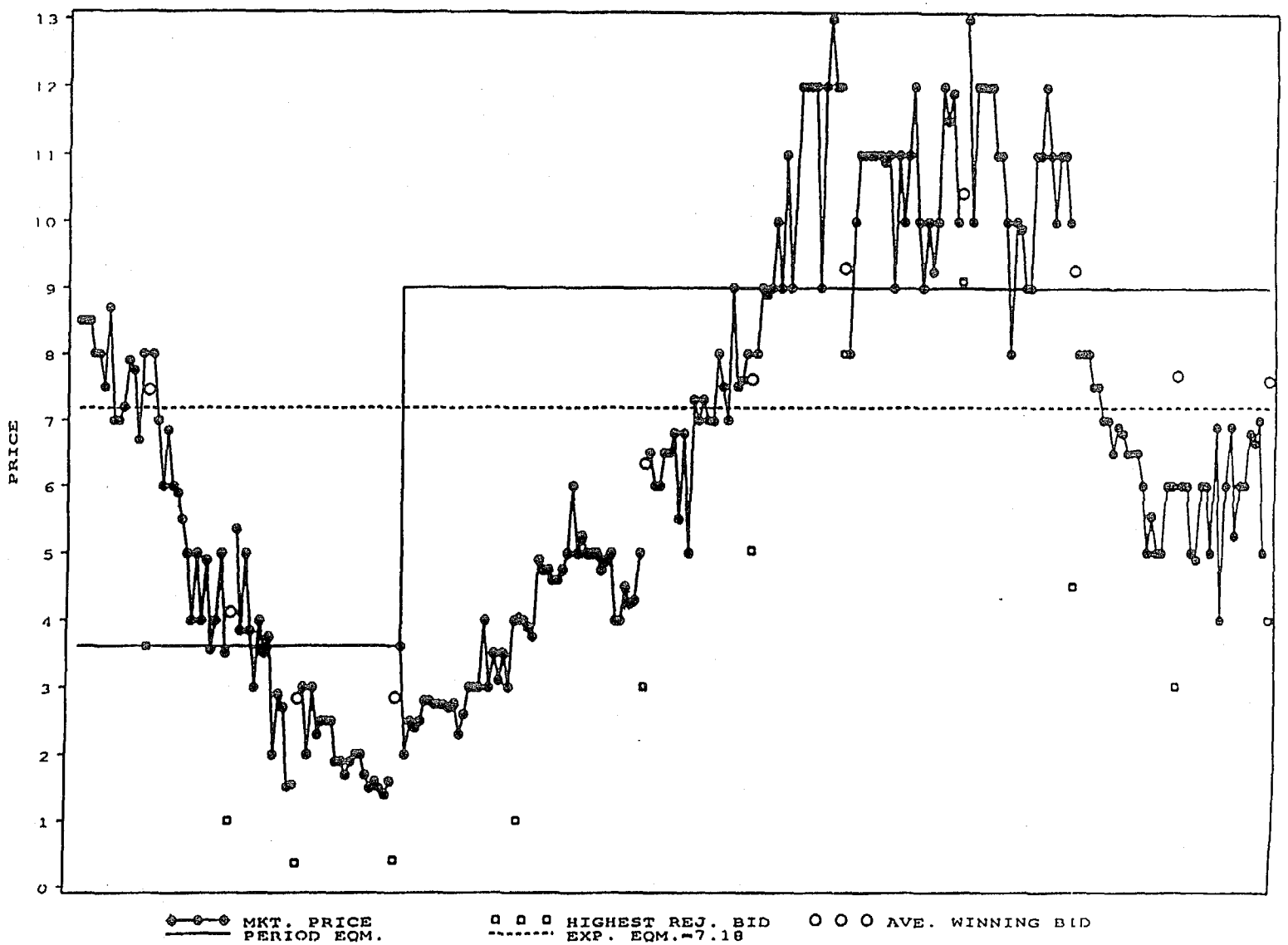


FIGURE 5

MARKET AND AUCTION PRICES

R11_12DC

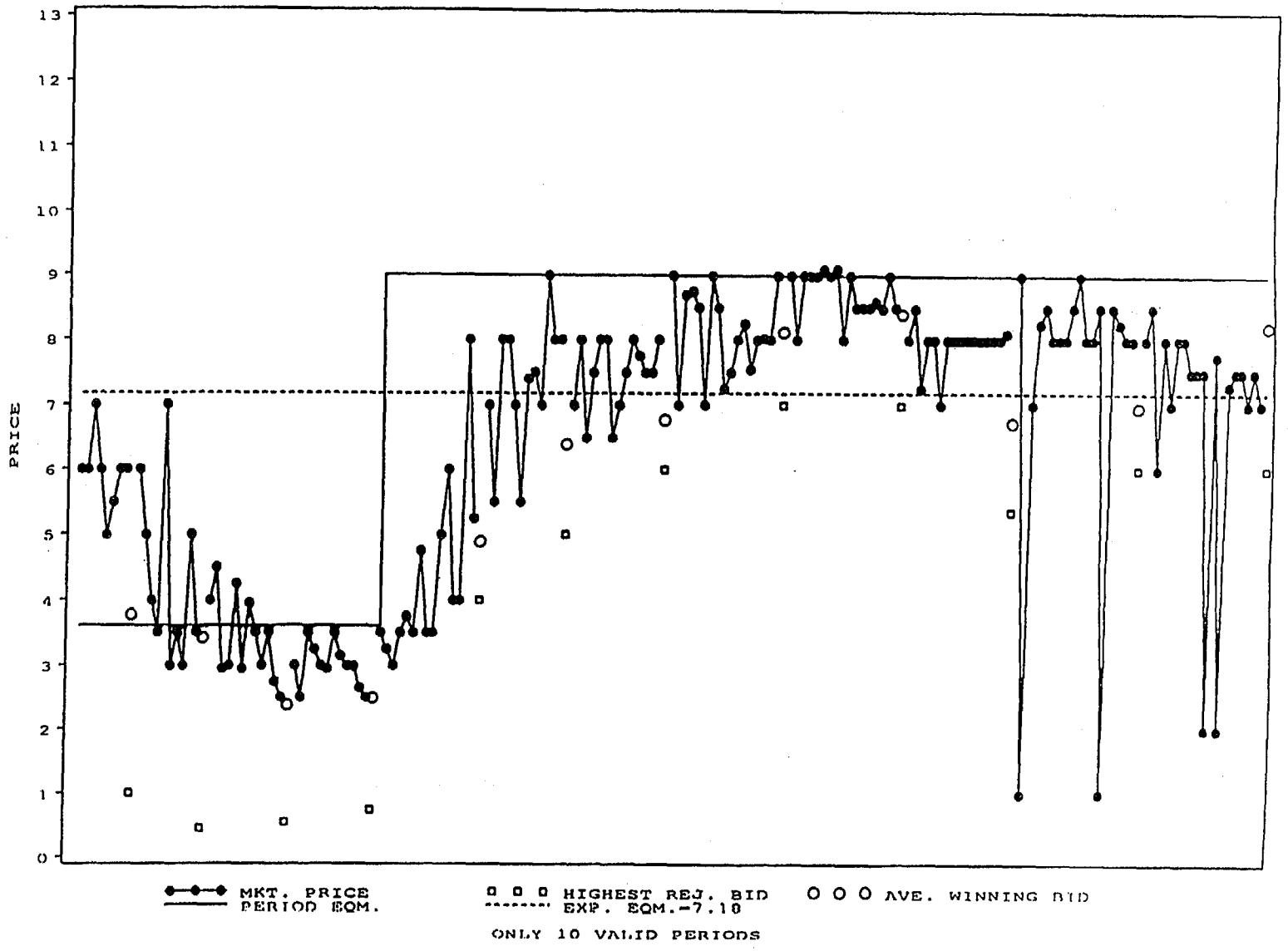
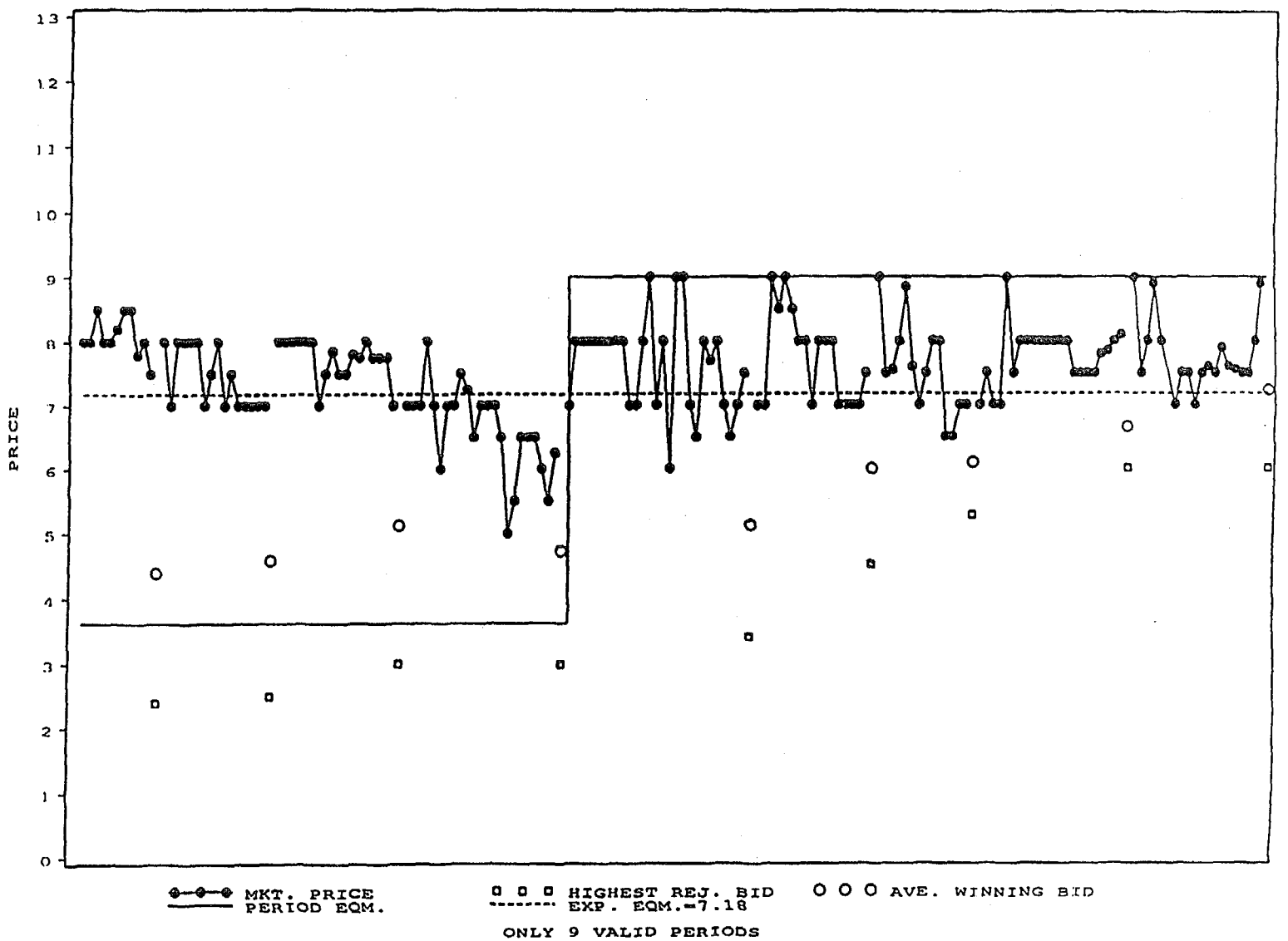


FIGURE 6
MARKET AND AUCTION PRICES
R12_10DC



be run both with and without the double auction. Also the training procedure will be varied systematically in order to control for training effects.

It is important that we understand what is going on in this market. With one experiment giving a thumbs up and another a thumbs down to this incentive-based system, we are sending radically different signals to the policy makers who are closely watching. If we can crate a system that controls emissions while reducing government "involvement," we may see widespread acceptance and use of incentive-based regulation. If we create a system that blows up (and we cannot explain why), this grand experiment may be for naught and we will continue with command-and-control.

Experimental economics offers a testing bed for this type of policy, but only if we can show its consistency, reliability and usefulness in identifying, explaining and mitigating problems such as those we have found here. We must carefully examine anomalies like those seen in this paper and methodically work to explain them. by doing so we will demonstrate the power of experimental economic methods as a policy analysis tool.

References and Selected Readings

Baumol, W.J.; Oates, Wallace E. (1988) The Theory of Environmental Policy. 2nd ed. Cambridge U.K: Cambridge University Press.

Brown-Kruse, Jamie; Cronshaw, Mark. (1991). Permit Market Experiments Preliminary Report. University of Colorado, Department of Economics. 92DOCS 016 - ET

Brown-Kruse, Jamie; Cronshaw, Mark; Elliott, Steven (1992-03). Experimental Investigation into Marketable Emissions Permits: Some Results and Questions. University of Colorado, Department of Economics.

Cason, Timothy N. (1992). Seller Incentive Properties of EPS's Emissions Trading Auction. University of Southern California, Los Angeles, CA. Dept of Economics. 92DOCS 028 - et

Cronshaw, Mark B.; Brown-Kruse, Jamie. (1991-3-8). Temporal Properties of a Market for Emission Permits. University of Colorado, Department of Economics. 92DOCS 018 ET

Cronshaw, Mark B.; Brown-Kruse, Jamie. (1992-05-26). An Experimental Analysis of Emission Permits with Banking and the Clean Air Act Amendments of 1990. University of Colorado, Department of Economics. 92DOCS 030

Cropper, Maureen L.; Oates, Wallace E. Envrionmental Economics: A Survey. Journal of Economic Literature 30(2):675-740.

Dales, John H. (1988) Pollution, Property and Prices. Toronto: University of Toronto Press.

Franciosi, R.; Isaac, R. Mark; Pingry, David E.; Reynolds, Stanley S. (1992-5). Extension of Research into Marketable Emissions Permits. 92DOCS 029

Hahn, Robert W. (1988). Promoting Efficiency and Equity through institutional design. Policy Sciences 21:41-66. 91DOCS 031

Hahn, Robert W. (1989) "Economic Prescriptions for Environmental Problems: How the Patient Followed the Doctor's Orders" Journal of Economic Perspecitives 3(2): 95-114.

Hahn, Robert; Noll, Roger G. (1982) "Designing a Market for Tradable Emission Permits" in Reform of Environmental Regulation. Wesley Magat (ed.) Cambridge, MA: Ballinger.

Hale, Douglas R.; Bjornstad, David J. (1991). Marketable Permits for Controlling Sulphur Dioxide Emissions. Oak Ridge National Laboratory, Energy Division. 92DOCS 019 - ET

Heggelund, Merete. (January, 1991). Emissions Permit Trading: A Policy Tool to Reduce the Atmospheric Concentration of Greenhouse Gases. Calgary: Canadian Energy Research Institute. 92DOCS 006

Montgomery, W. David. (1972) Markets in Licenses and Efficient Pollution Control Programs. Journal of Economic Theory 5(3): 395-418.

Plott, Charles R. (1983). Externalities and Corrective Policies in Experimental Markets. Economic Journal 93: 106-127. 91DOCS 043 ET Sheffield, Arthur (1992) Canada, Department of the Environment, Personal Communication.

Tietenberg, Tom (1988) Environmental and Natural Resource Economics. 2nd ed.

Smith, Vernon L., Svchanek, Gerry L., and Williams, Arlington W. (1988) "Bubbles, Crashes, and Endogenous Expectations in Experimental Spot Asset Markets," Econometrica, Vol 56, NO 5, September, 1119-1152.

Voluntary Selection of Real Time Pricing: Laboratory Evidence

Jamie Kruse, University of Colorado

Abstract

This paper reports an experimental study whose goal is to capture behavior in the laboratory that mimics observed field behavior. A laboratory model is created in which the access fee portion of a two part tariff can lead to acceptance or rejection of a real time pricing program. Further, changes in the laboratory conditions which generate derived demand for an input result in optimal reversal of the decision to participate in the program. This behavior mirrors that observed when such a program was implemented on a trial basis in the field. Further, two proposed adjustments are tested and both judged successful in increasing participation rate.

I. Introduction

In 1989, Niagara Mohawk Power Corporation (NMPC) offered on an experimental basis, real-time pricing service. The experimental service called Hourly Integrated Pricing Program (HIPP), is a two-part tariff. A customer specific access fee plus marginal cost-based variable charge comprise the HIPP tariff. The societal benefit of real-time marginal cost-based pricing is well documented (For example, see R. Willig, 1978; R. Schmalensee, 1982; H. Varian, 1987; and K. Roberts, 1979). Under HIPP the goal of customer specific access fee is to preserve revenue neutrality--the customer is at least as well off under the two part-tariff as he would have been under conventional pricing. In order to preserve incentive compatibility, the current or future access fee cannot be affected by the customer's current decisions. The access fee is based on the customer baseline load (CBL) which was determined prior to creation of the program. Specifically the access fee equals the CBL multiplied by the nonnegative difference between the conventional and real-time price. Under stable conditions, this access fee plus real-time price yields exactly the same revenue to the utility and profit to the customer as conventional pricing <u>if the customer maintains the same load pattern</u>.[1] Figure 1 presents a simple example. In case a, P_c is the conventional price and x the quantity the customer would use based on derived demand D. Suppose x becomes the customer's CBL and the real time price P_R is realized. The customer can shift load to x_R to take advantage of the low input price. The customer's gain from the transaction at P_R is equal to the area of acP_R minus the access fee (area $P\underline{c}bdP_R$). However in its experimental offering of HIPP, NPPC found that customers would enroll in HIPP and then retire from it later when economic conditions changed.

As long as the customer's economic conditions do not worsen, the benefits to the customer and society are both positive. A problem arises with the fixed access fee when economic conditions change. When conditions change so that customers' consumption under conventional pricing would differ from their established CBL, HIPP is no longer revenue neutral. Figure 1 case b illustrates an example that would prompt the customer to retire from the real time pricing program. If economic conditions deteriorate for the firm, then the relevant derived demand is D' as depicted in case b. In this case, with the favorable real time price, P_{Rt}, the customer will use $x_{R'}$. The customer's gain from the transaction is the area $a'eP_{Rt}$ minus $P\underline{c}bdP_{Rt}$--which makes him worse off than with the conventional price. At such a time, this customer would retire from the HIPP program only to reenter when an upturn in the business cycle makes HIPP preferable again. When customers leave the HIPP program, the societal benefit of marginal cost-based real-time pricing is lost for the time being.

The primary goal of this study is to determine whether laboratory conditions similar to those described above prompt a behavioral response similar to the customer reaction observed in the field. Our secondary goal is to compare two adjustments to the original program which changes the incentives to remain in the real time pricing program.

The laboratory environment was created so that subjects would self-select into two groups: New Program customers and "conventional" customers. From this base case, we then introduce a series of changes in the laboratory environment. Subjects experience a "recession" of known length. A subset of the New Program

[1] By stable, we mean that the economic conditions that determine customer profitability do not change. Product prices may be stochastic, but the underlying distribution is constant.

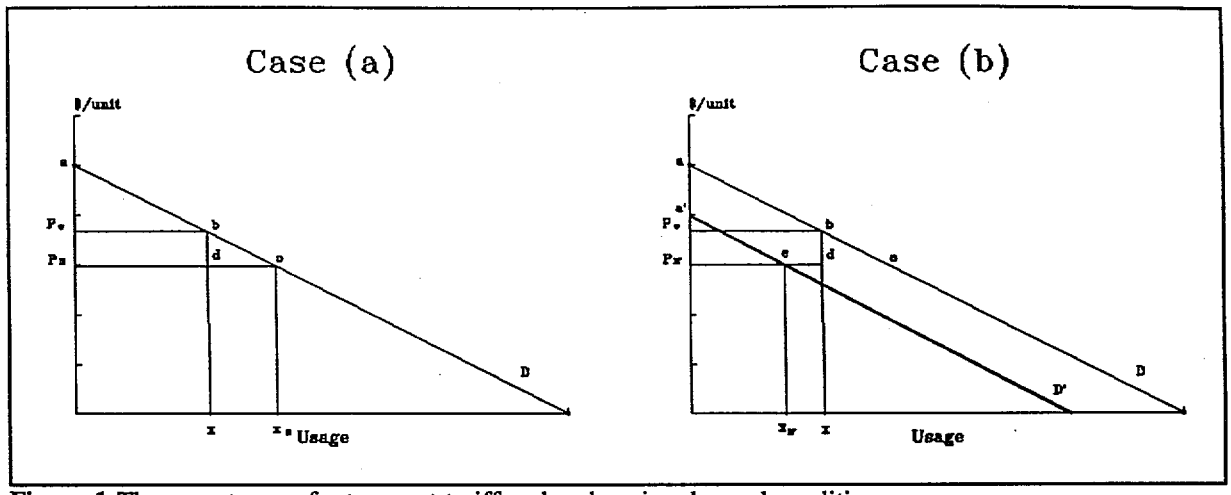


Figure 1 The acceptance of a two part tariff under changing demand conditions.

participants find their CBL-based access fee so high that the conventional program is more profitable than the New Program during the recession. A second recession is initiated with either the additional stipulation that retiring from the New Program means that a mandatory waiting period must pass before reentry is allowed or a reduced access fee for the New Program accompanies the recession. In the next section we will elaborate on the experimental design and discuss expected profit maximization based incentives. This is followed by our experimental results and discussion in the final section.

II. Experimental Design and Theoretical Predictions

The experiment is divided into **Parts, Cycles,** and **Rounds.** Each of the four **parts** of the experiment places the subject under a different set of rules which govern his/her decision. Figure 2 demonstrates the relationship between parts, cycles and rounds. In each **part** there are several **cycles.** At the beginning of a **cycle,** the subject chooses either Conventional pricing or the New Program. In general, the subject does not know how many **cycles** are in each **part.** Each **cycle** contains four **rounds** (a, b, c, & d) in which the subject makes production decisions.[2]

In **Part I,** subjects are told that they are making production decisions for a firm which uses a single input to produce a fictitious product. All firms face the "conventional" input price in Part I. The product price for each firm is determined by the toss of a six sided die. The numbers on the die map into a set of six product prices. There are six different mapping. After the die toss is entered, the computer calculates the subject-specific profit opportunities for all possible input decisions. Figure 3 shows a typical screen display that would be presented to a subject. The profit maximizing production decision within a **round** is transparent--costs, revenue, and profit for every possible input choice are calculated for the subject. The subject can simply choose the number of input units that yield the highest profit for the round.

Part I familiarizes subjects with the general layout and timing of each **cycle** and **round.** There are one or two practice **cycles** containing four **rounds** each. One additional cycle in Part I establishes the subject's historical baseline usage which is used to calculate the access fee for the New Program.

Part II introduces the "new" real time pricing program. Subjects are informed that at the beginning of each **cycle,** they can choose either the constant (conventional) input price or a "new program" in which they face a lower average input price plus an access fee. The access fee is calculated as follows:

Access Fee = (old input price - new program input price) x Historic Usage[3]

[2] A complete set of instructions, protocol and profit calculations is available from the author on request.

[3] The minimum access fee is zero. If the HIPP input price exceeds the old input price, then the access fee is set to zero.

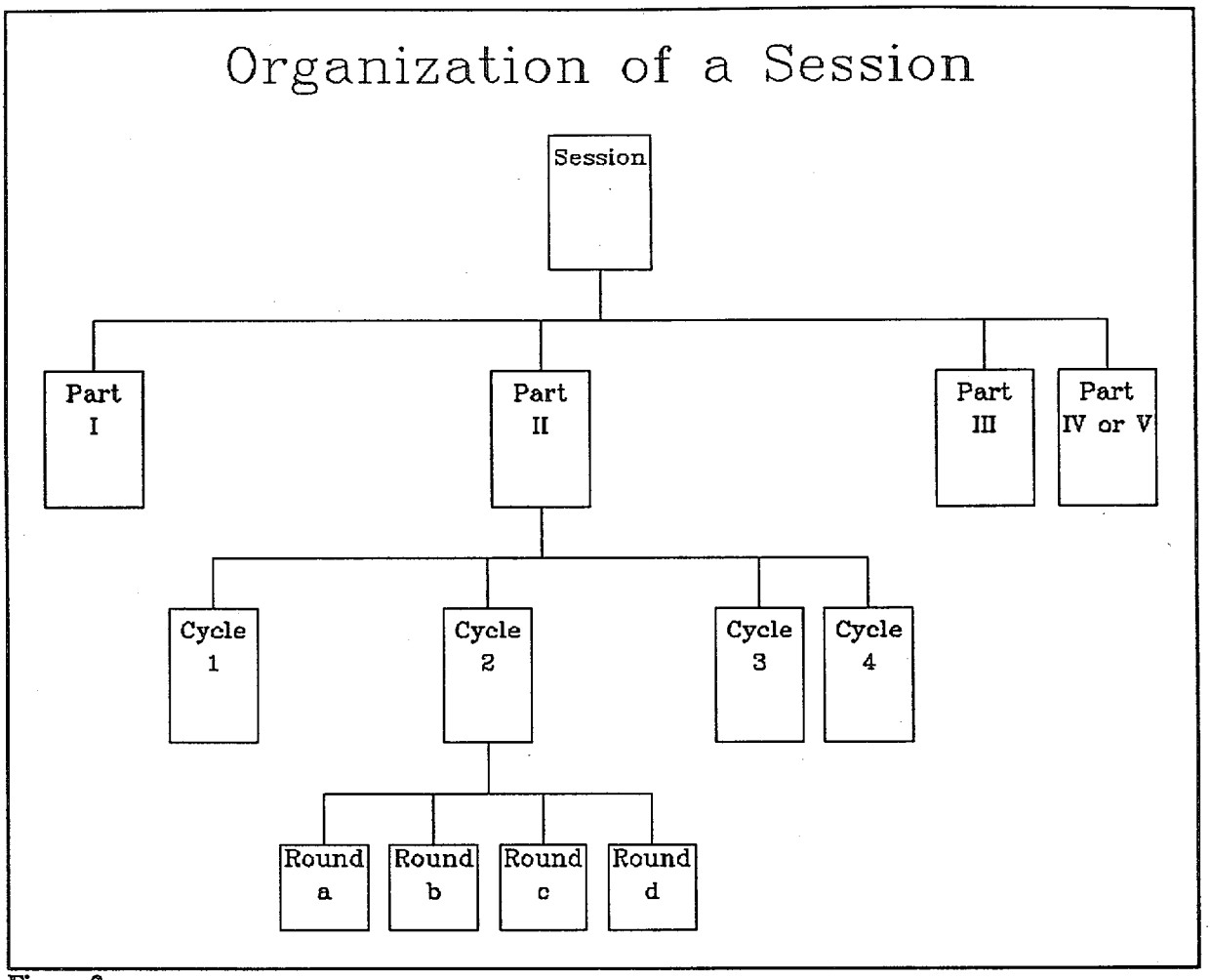

Organization of a Session
Session
Part I
Part II
Part III
Part IV or V
Cycle 1
Cycle 2
Cycle 3
Cycle 4
Round a
Round b
Round c
Round d

Figure 2

Figure 3.

PRODUCTION TABLE

Input Units	Input Cost	Product Units	Product Revenue	Profit
1	2.40	3.95	3.16	0.76
2	4.80	7.80	6.24	1.44
3	7.20	11.55	9.24	2.04
4	9.60	15.20	12.16	2.56
5	12.00	18.75	15.00	3.00
6	14.40	22.20	17.76	3.36
7	16.80	25.55	20.44	3.64
8	19.20	28.80	23.04	3.84
9	21.60	31.95	25.56	3.96
10	24.00	35.00	28.00	4.00
11	26.40	37.95	30.36	3.96
12	28.80	40.80	32.64	3.84
13	31.20	43.55	34.84	3.64
14	33.60	46.20	36.96	3.36
15	36.00	48.75	39.00	3.00
16	38.40	51.20	40.96	2.56
17	40.80	53.55	42.84	2.04
18	43.20	55.80	44.64	1.44
19	45.60	57.95	46.36	0.76
20	48.00	60.00	48.00	0.00

INFORMATION
Part 1 Cycle 1 Round B

Input Price	2.40
Product Price	0.80

Input Units	Product Revenue	Profit
----	----	----
Balanc	4.00	

INPUT REQUEST
Enter Input Units
----->

Historic usage was established in Part I. The New Program input price is determined by the draw of a poker chip from a bag each round. There are twenty chips of each color in the bag and one hundred chips in all. A poker chip is drawn each round to determine the New Program input price and the die is tossed at the beginning of each cycle to determine the product price.

The first two cycles of Part III represent an economic downturn which hits all firms. The lower Product price Schedule B is used for the first two cycles. The two "recession" cycles are followed by three "normal" (Price Schedule A) cycles for a total of five cycles in Part III.

Part III is followed by either Part IV or V. In Part IV, if a subject does not choose the New Program during a recession cycle, then he/she must wait four cycles before being given the New Program input price option. In Part V, the access fee is scaled down during recession cycles. Both Parts IV and V last a total of six cycles.

All sessions were conducted in the Laboratory for Economics and Psychology (LEAP) at the University of Colorado. Subjects were volunteers from undergraduate classes at the University of Colorado at Boulder. Subjects were allowed to participate once. Typically, the duration of the experiment was two hours with subjects earning from $10 to $46. Table 1. summarizes the number of subjects and treatments used in each session.

Table 1. Summary of Sessions and Treatments

ID#	# of Subjects	I	II	III	IV	V
NMUCB1	11	X	X	X	X	
NMUCB2	11	X	X	X		X
NMUCB3	8	X	X	X	X	
NMUCB4	11	X	X	X	X	
NMUCB5	6	X	X	X		X
NMUCB6	9	X	X	X	X	
NMUCB7	6	X	X	X		X
NMUCB8	8	X	X	X		X
NMUCB11	9	X	X	X	X	
NMUCB12	6	X	X	X		X
NMUCB13	7	X	X	X	X	

III. Parameters

The experiments reported here were designed and parameters chosen so that under the assumption of risk neutral profit maximization subject/firms self select into three groups. Roughly one third of the subjects would maximize expected profit by always choosing the New program when it is offered. Another one third of the subjects maximize expected profit if they choose the old program. The remaining one third of the subjects would maximize expected profit if they choose the New Program during nonrecession cycles and choose the old program during the recession cycles.

All subjects have a production function of the form, $Q=ax^2 + bx$. Q is the quantity of output which is sold on the product market. x is the quantity of the input used. In general, $a<0$ and $b>0$. Specifically, the production function all firms use is $Q = -0.05x^2 + 4x$. Firms face a product price that is drawn from a known distribution. The input price and product price are known by the subject before he/she must make a production decision. The old program input price is $2.40 per unit. The set of product prices and their associated probabilities (in parentheses) are shown below. In addition, the optimal quantity of input is listed below the price and probability.

Product					
Price	$0.65	$0.70	$0.80	$0.90	$0.95
(probability)	(1/6)	(1/6)	(1/3)	(1/6)	(1/6)
Optimal usage	3	6	10	13	15

The experiment is designed so that all subjects face the same distribution of product prices. However, the random draw signifies different prices for different subjects. In other words, when one subject sees the draw and receives a product price of $0.90, another subject receives $0.80, another receives $0.65 and so on. In part I, the historic usage is recorded and used to calculate the New Program access fee for subsequent parts of the experiment. Our design generates subject decisions which should yield historic usage which ranges from 3 to 15 units of input and thus divides the subject pool. The proportion of the subject pool with a specific historic usage is roughly equal to the probability of the price associated with that optimal usage. So about one sixth of the subjects should have historic usage equal to three, one third of the subjects will have a ten unit history, etc.

The "New Program" is offered for the first time in Part II. Subjects can choose the old program with fixed input price or the new program with access fee and input prices that are drawn from a known distribution. The access fee equals the historic usage times the difference between $2.40 and the new program input price and is nonnegative. If the new program input price exceeds the old program price of $2.40, the access fee is zero. The subject must decide whether to subscribe to the new program before he/she knows the current cycle's product price. The New Program input prices, their associated probabilities and the Access Fee are shown below:

New Program					
Price	$1.90	$2.00	$2.05	$2.30	$2.50
probability	(1/5)	(1/5)	(1/5)	(1/5)	(1/5)
Access Fee	H*0.50	H*0.40	H*0.35	H*0.10	0.00
H=historic usage					

Restricting our attention to the historic usage predicted by simple profit maximization in Part I, we can look at the difference in *expected* profit between the Old Program and the New Program. Recall that profit maximizing firms in Part I will establish historic usage of 3, 6, 10, 13, or 15 units. Table 2. below compares their expected profit given their participation in the Old or New Program.

Using expected profit maximization as the decision criteria, subjects with historic usage of 3, 6, or 10 should choose the New Program and subjects with histories of 13 or 15 units should choose the Old Program.

Parts III, IV, and V have a two-cycle recession. The product prices during a recession are shown below.

RECESSION					
Product					
Price	$0.65	$0.67	$0.70	$0.73	$0.75
(probability)	(1/6)	(1/6)	(1/3)	(1/6)	(1/6)

Table 2. Profit Expectations from the Old (conventional) program and New Program.

Historic Usage	Old Program	New Program	difference in expected profit/round
3	$4.63	$6.89	$2.26
6	$4.63	$6.08	$1.45
10	$4.63	$5.00	$0.37
13	$4.63	$4.19	-$0.44
15	$4.63	$3.65	-$0.98

Table 3 compares the expected profits during a recession. Expected profit maximizing firms with histories of 3 or 6 should would choose the New program. Firms with a 10 unit history would choose the Old Program as would firms with historic usage of 13 or 15 if the decision criteria is expected profit maximization. Contrast the

risk neutral prediction for recession v. nonrecession cycles for the firms with a 10 unit history. During a "normal" cycle they would choose the New Program and during a "recession" cycle, they would choose the Old Program.

Parts IV and V change the incentive to participate in the New Program during recession cycles. Comparison of the last two parts is analogous to comparing a "carrot or stick" approach. Part IV penalizes subjects if they leave HIPP during a recession by not allowing re-entry for four cycles. Under the penalty system, firms with histories of 3, 6, and 10 should choose the New Program, whereas the Old Program is the expected profit maximizing choice of firms with historic usage of 13 or 15 units.

In Part V we reduce the access fee so that the New Program dominates the Old Program in terms of expected profit for the firms with a ten unit history. The magnitude of the difference is about the same as for the nonrecession case.

The experimental design and the parameters set the stage so that the choices of Old Program always, New Program always, and New Program during normal cycles and Old Program during recession cycles could be observed for some subset of the subjects. Using risk neutral profit maximization as the decision criteria, we can partition the subjects precisely as shown in Table 4. However, this partition is not necessarily a prediction. Risk attitudes can shift the classifications. The risk neutral partition provides a baseline for comparison with observed choices.

IV. Results and Discussion

One major concern in conducting this study was whether or not we could create conditions in the laboratory that would give rise to behavior that mimics field behavior. We wanted to create a set of conditions that would induce some subjects to participate in the New program during "normal economic periods" but retire from the program during economic downturns. All subjects had identical production functions, faced the same product price distribution and potentially had the same per unit input costs. The only differences induced by the lab environment were differences in historical usage which affected subjects' access fees. The parameters were chosen so that the array of subject decisions would cover all possibilities. The next step is to access whether the variety in subject decisions were due to purely random behavior or to the laboratory stimuli we used.

Figure 4. summarizes the participation rate in the New Program for all subjects. Normal cycles were unrestricted cycles in which schedule A product prices were used. Recession cycles were unrestricted cycles in which the lower schedule B product prices were used. Penalty cycles were also recession cycles with reentry allowed into the New Program after four cycles had passed since retirement. Finally, Lower Fee cycles were recession cycles with a reduced access fee. The New Program participation rate went from 59.2% in normal cycles to 49.8% in unrestricted recession cycles. In addition, both the penalty and lower fee resulted in higher participation than in unrestricted recession cycles.

The incentives for different decisions by otherwise identical firms were created by setting the individual decision conditions so that subjects would have different historical usage and thus different access fees. Firms with a low CBL would find the New Program very attractive because of a relatively low access fee. On the other hand, high CBL firms would find the access fee prohibitively high. If we can separate decisions on the basis of their CBL, then we have evidence that subjects are responding to the stimuli we created. Figure 5 provides graphical evidence that subject CBL can be used to classify subject decisions. High CBL firms had the lowest New Program participation rate on average for all cycle types. Low CBL firms had the highest New Program participation rates for all cycle types. A 2x3 ANOVA was used to test for differences in participation rate in unrestricted normal and recession cycles. We can reject the hypothesis that subjects with different CBL's responded similarly. In addition for alpha greater that .061, we can reject the hypothesis that participation in the New Program was the same for recession and normal cycles. These results are summarized in Table 5.

We compared subject participation in the New Program during the three types of recession cycles: unrestricted recession cycles, penalty recession cycles and lower fee recession cycles. The difference in participation rate according to CBL class and recession cycle type is highly significant in both cases. Table 6 provides a summary of the results of the ANOVA described above.

Analysis of Covariance was used to determine whether subjects responded differently to the penalty v. the lower fee. Subject participation during unrestricted recession cycles was the covariate. There is no significantdifference in the rate of subject participation in the two adjustments to the New Program. Apparently, subjects reacted to the two incentive programs in a similar fashion. Table 7 presents the results of the ANCOVA.

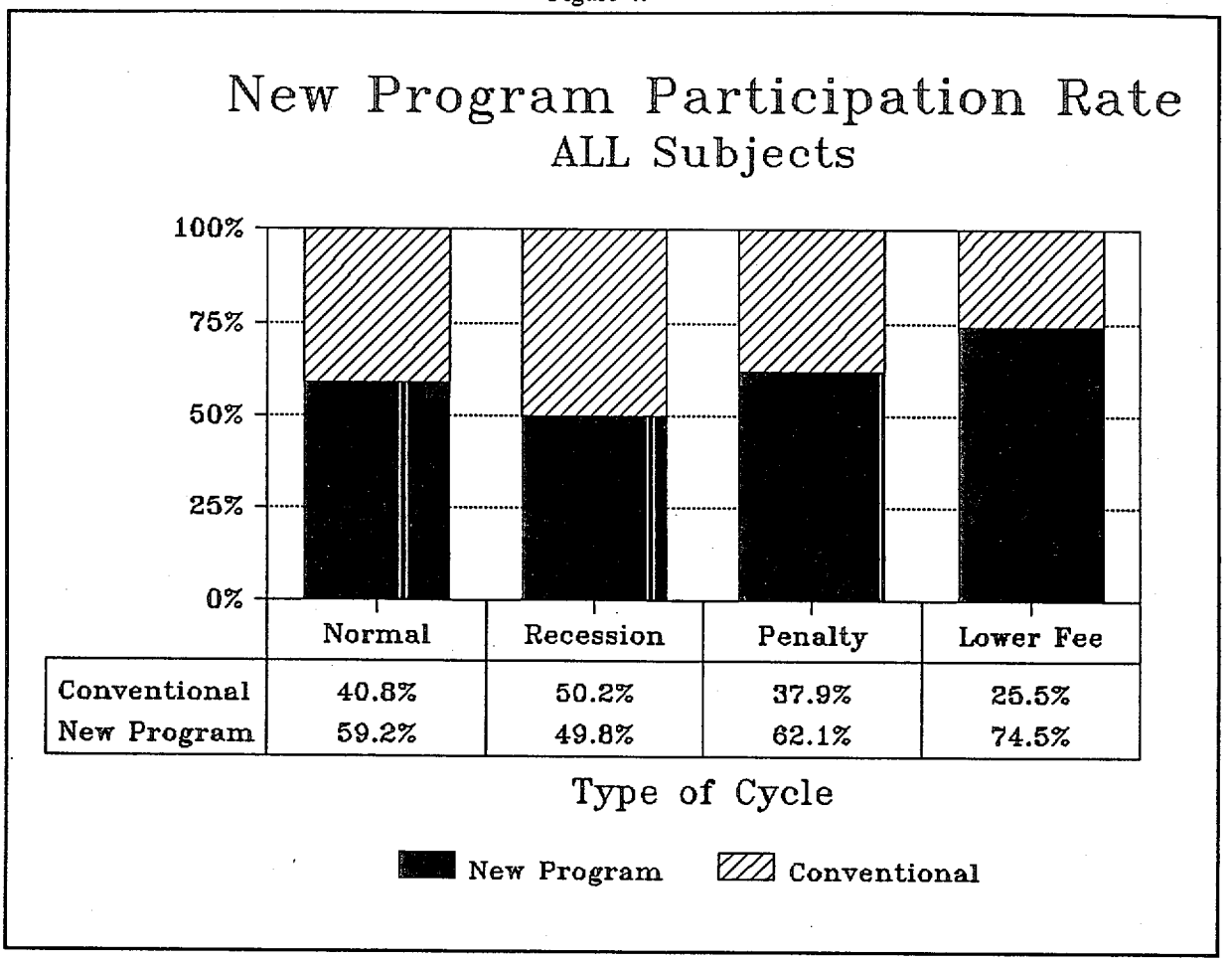
Table 3. Profit Expectations from the Old (conventional) and New Program during recession cycles.

Historic Usage	RECESSION Old Program	New Program	difference in expected profit/round
3	$1.24	$2.62	$1.39
6	$1.24	$1.81	$0.58
10	$1.24	$0.73	-$0.50
13	$1.24	-$0.08	-$1.31
15	$1.24	-$0.62	-$1.85

Table 4. Partition of sample based on expected profit maximization.

Historic Usage Part I	Part II	Part III cycle 1-2/3-5	Part IV cycle 1-2/3-6	Part V cycle 1-2/3-6	Proportion of Sample
3 or 6	New	N/N	N/N	N/N	1/3
10	New	O/N	N/N	N/N	1/3
13 or 15	Old	O/O	O/O	O/O	1/3

Figure 4.



	Normal	Recession	Penalty	Lower Fee
Conventional	40.8%	50.2%	37.9%	25.5%
New Program	59.2%	49.8%	62.1%	74.5%

Figure 5.

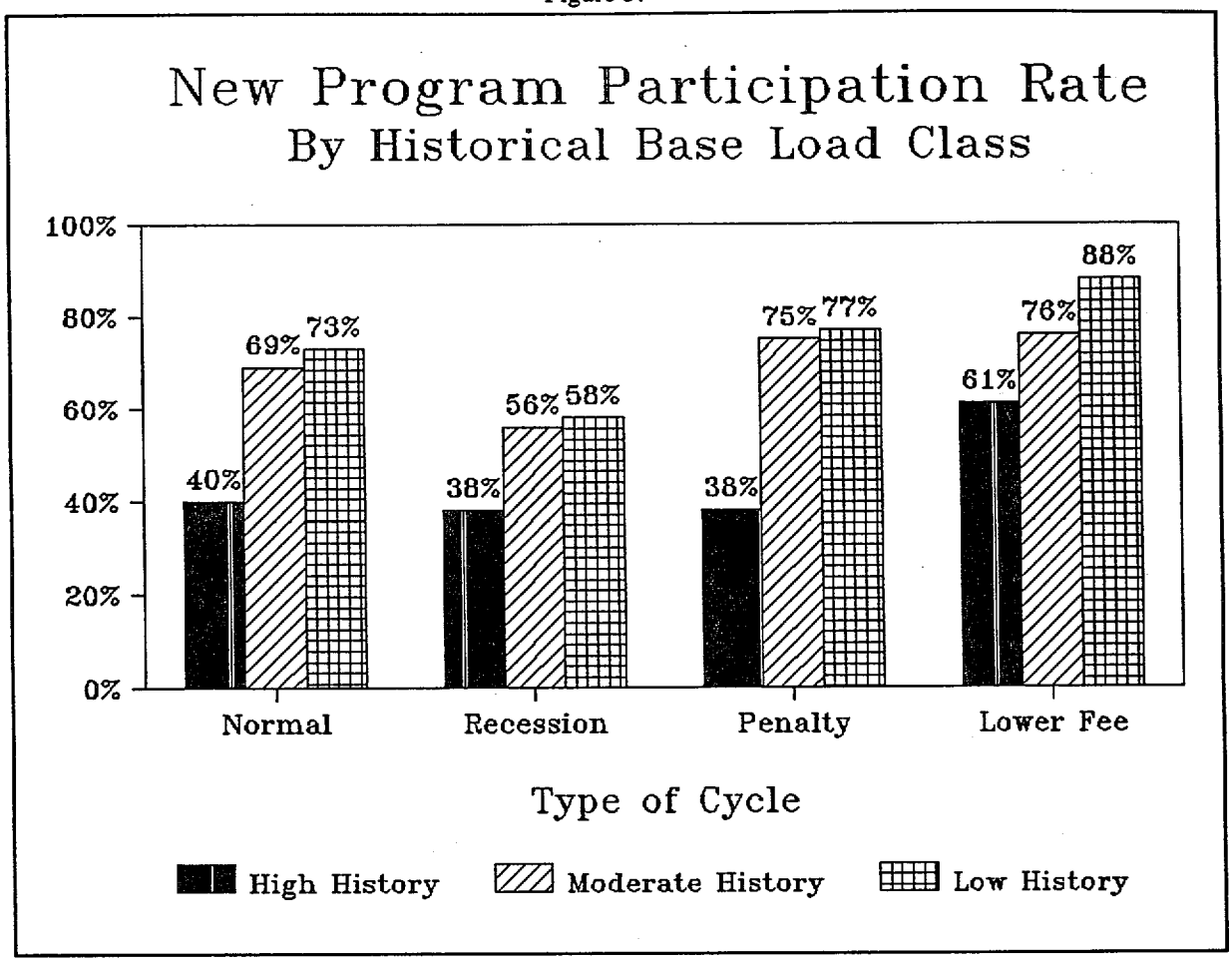

New Program Participation Rate
By Historical Base Load Class
100%
80%
60%
40%
20%
0%
40%
69%
73%
38%
56%
58%
38%
75%
77%
61%
76%
88%
Normal
Recession
Penalty
Lower Fee
Type of Cycle
High History
Moderate History
Low History

Table 5. ANOVA Summary Table for Unrestricted Cycles and CBL Class
2 (normal and recession cycles) x 3 (high, moderate, low)

Source of Variation	sum of squares	d.f.	mean-square	F-ratio	P[4]
CBL class	2.813	2	1.407	13.362	0.000
Unrestricted cycle type	0.376	1	0.376	3.569	0.061
CBL x cycle type interaction	0.223	2	0.111	1.059	0.349
error	18.634	177	0.105		

Table 6. ANOVA Summary Table for Recession Cycles and CBL Class
3 (recession, penalty, lower fee) x 3 (high, moderate, low)

Source of Variation	sum of squares	d.f.	mean-square	F-ratio	P[5]
CBL class	4.169	2	2.084	14.049	0.000
Recession cycle type	2.504	2	1.252	8.437	0.000
CBL x cycle type interaction	0.529	4	0.132	0.891	0.470
error	35.758	241	0.148		

[4] P is the calculated probability of a type I error. If P<alpha then we can reject the null hypothesis that the means are equal.

[5] P is the calculated probability of a type I error. If P<alpha then we can reject the null hypothesis that the means are equal.

V. Conclusions

The impetus for this study came from observed field behavior. Customers of a real time pricing program instituted through a two-part tariff would "bail out" during recessions and return when business improved. A model based on risk neutral profit maximization generates implications that give rise to the same behavior. Theoretically and behaviorally, the magnitude of the access fee is a determinant of the decision to subscribe to the New Program. Subjects did respond to adjustments in the New program to retain customers. Both of these adjustments work equally well from a statistical viewpoint although customers probably prefer the reduced access fee to the penalty.

The point of departure for this study is what makes it unique. Smith(1986) contrasts nomotheoretical experiments that compare theory and observation with nomoempirical that compare the effect of different institutions and/or environments as a means of documenting replicable empirical "laws". A subset of the nomoempirical are heuristic experiments that provide empirical probes into new topics. This study was not theory driven--although the theory plays an important part in the design and parameterization process. We do test different institutions-the alternative designs of the new program, but with a theoretical basis for the hypotheses we test. The experiment should not be classified as heuristic. The theoretical basis, once established, plays an important role. This study represents an attempt to capture the motives which give rise to a pattern of behavior that has been observed in the field--to generate similar behavior in the lab--and to test changes in the lab environment intended to modify the behavior. One remarkable consequence of this effort is that we had some degree of success.

Table 7. ANCOVA Summary Table for Penalty and Reduced Fee Recession Cycles using the Unrestricted Recession Rate as a Covariate

Source of Variation	sum of squares	d.f.	mean-square	F-ratio	P[6]
Recession cycle penalty or reduced fee	0.015	1	0.015	0.0849	0.772
Covariate: Unrestricted recession cycle participation rate	1.899	1	1.899	10.626	0.002
error	15.910	89	0.179		

[6] P is the calculated probability of a type I error. If P<alpha then we can reject the null hypothesis that the means are equal.

References

Katz, M. 1983. "Nonuniform Pricing Output and Welfare Under a Monopoly." *Review of Economic Studies*, 37-56.

Oren, S., S. Smith and R. Wilson. 1983. "Competitive Nonlinear Tariffs." *Journal of Economic Theory*, 49-71.

Roberts, K. 1979. "Welfare Considerations of Nonlinear Pricing." *Economic Journal*, 66-83.

Smith, V. 1986. "Experimental Methods in Economics." in The New Palgrave: A Dictionary of Economic Theory and Doctrine. Macmillan Press, Ltd. New York.

Spence, M. 1977. "Nonlinear Prices and Welfare." *Journal of Public Economics*, 1-18.

Willig, R. 1978. "Pareto-superior nonlinear outlay schedules." *Bell Journal of Economics*, 9(1): 167-207.

Investment, Uncertainty, and Risk

David J. Bjornstad and Steven Elliott, Oak Ridge National Laboratory

Abstract

Traditional economic theory suggests that investment will occur when the expected return on the investment is at least equal to the cost of capital (the interest rate). However, observations of investment behavior suggest that the predictions from this theory overstate the penetration of new technologies into markets. A new literature has developed to explain why investment seems to require a rate of return greater than the traditional cost of capital. We propose to use controlled laboratory experimental methods test the implications of this new literature, and thereby develop an understand of what drives investments in such things as energy saving technologies.

Introduction

The level of a nation's technology base is closely related to the nature of its capital stock. Changes in that base are therefore affected by decisions to make changes in the capital stock. Stated differently, technological change is determined by investment behavior, whether by firms in the production of goods for the marketplace or by households in the production of household services. In the aggregate, the summation of individual decisions regarding durable goods purchases determines the introduction of technology into the economy. Many of the nation's perceived problems rely to a great extent on technological change as a solution. Environmental protection, reduced CFCs, and reduction of greenhouse gas emissions have all been linked to new, innovative technologies. Hence understanding the underlying basis of investment is key to developing efficient policy responses to these issues.

The traditional theory of investment is based in fundamental Marshallian analysis, in which factor productivities are valued in product terms and related to opportunity costs. For investment, the opportunity cost is viewed as the cost of capital obtained through retained earnings or through capital markets. The firm is expected to invest when presented with opportunities to earn a rate of return that exceeds its opportunity costs. Technical innovations that lower costs or otherwise present profit opportunities are often forecasted to penetrate markets promptly. Conversely, factors whose value of marginal product falls below opportunity costs are forecasted to be discarded promptly.

This theory has proven a sound base for approaching the fundamental behavior of the firm as it trades off the purchases of various factor inputs. Yet it has been a poor device for forecasting the timing or the scope of individual investment decisions. In an effort to sharpen the forecasting properties of the theory, a new body of literature has developed that takes into account the irreversibility (sunk cost) attributes of the decision to invest and the role played by uncertainty. In simple terms, this literature postulates a theory of "optimal inertia" by creating a cost component that represents the option value placed by the firm on the opportunity to wait until more information becomes available and uncertainty is reduced.

This theory is linked to technological change by the assertion that the larger the difference between the attributes of an innovative durable good and the traditional durable good it seeks to replace, the larger is the uncertainty surrounding the decision. One theory (Dixit [1992]) argues that the value of the option to wait is directly and positively related to uncertainty, from which it follows that firms will be more reluctant to adopt technologically innovative products or processes then more traditional ones. Stated differently, the "hurdle rate" the firm applies to investment decision making (opportunity cost of capital adjusted for uncertainty) is systematically higher for technologically innovative goods than for traditional ones. Moreover, specific hurdle rates will be conditional on the market structure, the cost of information to the decision making unit, and the ability of the decision making unit to diversify, among other factors.

In this paper we outline a plan to investigate the relationship between investment behavior and choices among durable good opportunities characterized by different degrees of performance uncertainty. The approach to be taken will apply the methodology of experimental economics which offers the ability to control such institutional concerns as market structure, ability to diversify, etc., while systematically varying parameters of interest, such as

risk. It therefore offers an intermediate option for gathering information between pure theory and expensive and relatively complicated gathering of field data.

This work is of particular interest because of its potential ability to shed light on behavior that must be understood clearly in order to develop policies that influence market acceptance of new technologies. Past research has indicated that new technologies typically do not achieve performance levels anticipated by the bench-testing. Past research has also established that technologies penetrate at far slower rates then would be predicted by the simple Marshallian model. Despite this, Marshallian postulates remain at the heart of much of investment analysis particularly at the federal level. This proposed research would provide a foundation for analysis that could lead to an improved understanding of investment behavior controlling for confounding influences.

Background

The theory of investment has been codified in a rich and diversified literature, which for present purposes can be best understood through two seminal pieces. Pindyck [1991] summarizes this literature and introduces the notion that because decisions to purchase capital goods are characterized by irreversibility, uncertainty, and the ability to delay the decision until a future time period, investment opportunities can be arrayed and analyzed in a manner similar to financial options. In another paper Dixit [1992] retraces many of the steps taken by Pindyck while emphasizing to a greater extent the notion of "optimal inertia." Taken together, these two pieces essentially formulate the theory of investment as a problem of optimal timing rather than the more traditional approach of choosing an optimal level.

Pindyck takes as a point of departure the notion that a firm's value lies in part in the option value of the alternatives open to it. Once exercised, this option value is translated from a potential activity to an actual one and evaluated in real terms rather than financial ones. However, the firm cannot, in general, retrace its steps, exchanging the real asset for a financial one at full value. Hence, the firm waits to invest while it gathers information to reduce uncertainty. As we note below, there is an implied value of information problem embedded in the approach.

Dixit emphasizes the symmetry between the decision to invest and the decision to abandon (cease operations). Casting his work (upon which Pindyck draws heavily) as a theory of optimal inertia, he relates uncertainty to the size of the hurdle rate needed to undertake investment. He then demonstrates that firms have a symmetric incentive to maintain unprofitable operations to degrees that exceed the point dictated by the payment of variable costs. Thus, the firm both invests and disinvests more slowly than under the traditional model.

Taken together, this literature yields three primary conclusions:

1. When there is uncertainty, irreversibility and opportunity to wait, firms delay decisions to modify durable good holdings longer then traditional analysis would forecast.

2. The scope of delay is directly related to the uncertainty associated with the performance of the durable good. By inference, the greater the departure from current practices, i.e., the more innovative the technology, the greater will be the inertia.

3. From the perspective of the firm, the cost of this inertia is less than is evident from the cash flow implications of a simple investment analysis, because the market attaches values to the options available to the firm.

Lind [1992] seizes upon this literature as a means to place a value on R&D related to global warming issues. He argues that the opportunity to wait, coupled with uncertainty and irreversibility, is analogous to segmenting a larger decision such as immediate implementation of a carbon tax into a sequence of smaller steps, and permits comparison of the costs and benefits of gathering additional information to the costs and benefits of some irreversible action.

Thus there exists an alternative theoretic foundation for examining investment decisions. However, the data that exist make direct testing of such theories difficult. For this reason we turn to the arena of experimental economic methods to conduct empirical tests of this complex investment process.

The use of direct experimental testing of economic principles and theories began in the 1950's and grew the 1960's and 1970's as economists such as Vernon Smith and Charles Plott began to demonstrate the benefits of

studying economic phenomenon under controlled conditions. The field has taken off in the 1980's and 1990's and is now recognized through numerous publications in prestigious economic journals.

A summary of this approach is contained in Smith [1992] that was prepared for the Secretary of Energy Advisory Board Workshop on Innovative Tools for Economic Analysis. This paper reviews the influence of experimental methods in the analysis of many economic issues such as energy policy. In this paper, Smith points out the difficulty of obtaining a complete picture of economic transactions in market settings. In most cases only aggregate data are available, with many compounding and confounding components that make analysis difficult. Regarding standard econometric data he points out that

> . . . [data] limitations are spotlighted in standard econometric textbooks which begin with an ideal world in which the observations for econometric analysis have been derived from an experiment in which the independent variables are controlled treatments with randomization used where appropriate. Necessarily the econometric exercise has been largely defined in terms of what to do when your data do not come from such an experiment. Experimental economists have sought to open up the constraints on traditional empirical methodology in economics by means of a direct investment effort in experimental methods (Smith [1992] p.1).

Economic experiments that study the operation of markets have a core of three elements: (1) the value/cost **environment**; (2) an **institution** that defines the communication of information within the market; and (3) **behavior** of the market participants. The **environment** defines the conditions and the motivations for exchange. Examples include the preference structure imposed on subjects, their initial endowments, and the technology or costs that they must face in the experiment. The **institution** defines the rules under which subjects make their decisions. One easily recognizable institution is an auction. As an example, under the English Auction, such those for paintings or antiques, participants raise the bid on an object until only one bidder remains. This person must then purchase the item up for auction, and pay the price she or he last bid. Institutions may be quite formal or casual depending on the issue under examination. The final element is **behavior**; the actions that the subjects reveal in the experiment. The experimentalist controls the environment and the institutions and then observes the behavior to draw inferences about the market, theory or policy being tested.

It should be noted that experimental economic methods are one of the strongest links between the "science" of economics and other "hard" sciences such as physics and biology (Smith [1992]). The rigorous, controlled settings employed by the experimental economist have appeal to other disciplines that demand this kind of controlled method to investigate the validity of their hypotheses. Thus, experimental methods offer a strong tool for gaining more widespread acceptance of many of the theories within the field of economics. Therefore, we will investigate issues of investment decisions as expounded by Dixit and Pindyck in the controlled environment of the laboratory as described by Smith.

Experimental Methods and the Study of Investment Decisions.

This section will paint, in broad strokes, our protocol or the investigation of investment decisions by various economic agents using experimental economic methodology. We begin by restating that in its simplest form experimental economics places real people in a controlled economic environment to test various economic theories. For example, in a basic situation, human subjects are placed in a room with instructions to buy and sell a "good" to test predictions implied by the Law of Supply and Demand. In more complicated experiments, subjects may act as managers of firms who decide whether it is best to enter a particular market and compete for customers, or to stay on the sidelines. Subjects are provided positive incentives to behave rationally through cash payments based on the profits earned through their decisions.

The tasks described below outline a multi-year program of research in the area of investment behavior. Each focuses on an aspect of the investment process described by the theory above. The first task outlines baseline experiments that strip the problem to its most salient elements. The results of these experiments will be used to judge the performance of the more complicated ones in the following tasks. In Task 2, the information about investment payoffs is manipulated. Portfolio effects and trade offs are examined in Task 3 by introducing multiple investment opportunities. In Task 4, we change the basic structure of the baseline experiments to understand better the effects of investment decisions across time. Finally in the last task, we examine various market structures and the influences that they exert on investment decisions.

Task 1: Baseline Experiments.

In these experiments, as in all the other experiments, subjects will act as managers of firms. In the baseline experiments, these firms have the potential to invest in a new technology for their production process. They will be given an endowment of cash, and will be asked if they would like to put some or all of this in a fund (the investment) that has some type of uncertain return. This distribution of possible returns on their investment represents the present value of the future stream of the revenues. For these experiments, the distribution will be known to all subjects. That is, subjects will know all the possible values and respective frequencies for the return on their investments. Any and all money not put in the investment is placed in a safe haven. These funds will earn some of rate of return that is less then the expected return from the investment (though it is possible that the realized return from the investment will be below the safe haven).

When all subjects have made their investment decisions, each will draw their return from the distribution (with replacement). Subjects will be paid their return, and the experiment will move to another investment decision. In subsequent rounds, the experimental procedure works much the same with the distribution of investment returns varying in order to discover evidence of the "hurdle" rate and other phenomenon described above.

In this experiment, Marshallian investment theory predicts that as long as the new investment's certainty equivalent is greater then the return from the safe haven, subjects should invest. However, for the models of Dixit and Pindyck to be theoretical improvements, this rate of return must be systematically and predictably greater then the certainty equivalent rate. For example, if marshallian theory best describes investment behavior, then if the expected rate of return in the investment is even marginally higher then the safe haven funds should be placed in the investment. However, if subjects do systematically put some funds into the safe haven, we can hypothesize that investment behavior follows another paradigm such as those discussed above. If the traditional Marshallian approach is shown to predict behavior in the laboratory, we will begin to manipulate parameter of this environment to test its limits. If the data support another theory, we will again begin to manipulate the experimental environment in order to understand more about this behavior. The following tasks are predicated on the assumption that Marshallian theory will fail in these baseline experiments; however, the tasks can be easily modified to investigate the bounds of this traditional theory.

Task 2: Investments Under Uncertainty.

An important variation will be making the distribution of payoffs more uncertain. This can be done by changing the way the payoff distribution is presented to the subjects. In this way, we will be able to capture many important aspects of the uncertainty of making investments in new technology. In a case of less certainty, subjects will be given the mean, the variance, and the bounds of the distribution of returns. In decreasing information from this point, subjects will receive only the mean and variance. Removing more certainty, subjects will receive only the mean of the distribution with the variance and the bounds left unknown. In a final case, it is possible to imagine a case where the distribution of returns is completely unknown. Here, subjects can pay to receive informative draws from the distribution.

Task 3: Multiple Investment Opportunities.

Another line of investigation is the area of multiple investment opportunities. In this case, instead of subjects deciding between a single investment or the safe haven fund, there are two or more investment opportunities. In some cases, these options will be mutually exclusive (i.e., subjects can invest all or nothing in one or another), or may have some divisibility (i.e., some funds may be invested across a number of different options). In the first case we examine competing technologies, and in the second, we look at ways that firms can diversify their portfolio of technologies and mitigate the risk they face.

An important aspect to this line of inquiry is the manipulation of information, as was described above. That is, while at first it may seem obvious that a subject would invest all her/his funds in the investment with the highest expected return, this is predicated on the assumption that all the distributions are the same. Therefore, we need to vary the distributions around the various investment options. This means that while two distributions may have the same expected payoff, we may find that one distribution that does not have a negative payoff but has no large high payoffs may be preferred to one with high potential gains matched by high potential losses.

This is only one example of how the information variables above can be combined with the multiplicity of investment opportunities. Each environment captures a different aspect of the investment process in the field, and each gives us a deeper insight into how investments are made and what affects their outcome.

Task 4: The Time Element in Investment Decision.

To this point our investigation intentionally has ignored time. Investment decisions have been made in a single time periods and the returns have been paid out in the same period. This task will lift this restriction and make time a part of the decision process. The theory cited above suggests that time allows firms to postpone an investment today to wait for more or better information in the future. As noted, this is a rational strategy if the payoff to waiting is greater then the income foregone by not investing (and therefore earning some return) today.

To introduce time into our experiments, we will have payoffs for an investment drawn every period throughout the experiment (or life of the investment). That is, if a subject chooses to invest in the first period, he or she will draw an income from some distribution in subsequent rounds. By adding this dimension, it is possible for subjects to postpone investing at present, and wait until more information is available or until a more favorable distribution is presented. It may be explained that more information will be made available as the experiment progress, or that different distributions will be attached to investments in different rounds to point out that there is some difference between investing right away and waiting and postponing the investment.

It should be noted that the information or change of distribution need not necessarily be a positive change. In other words, the future information may be that the investment will only pay off for two more rounds and then it will return nothing. In this case the subject has already foregone some number of rounds of income. Likewise, a change in information may indicate the introduction of a distribution with a higher probability of a negative income from the investment. Because all of these possibilities exist, it is not always true that waiting is an optimal strategy. Another important issue brought out in the theory is that it might be optimal to spend a little today to insure the option to invest more at a later date. This can be used to examine some issues of sunk cost.

In terms of our experiments we can give subjects the chance to place a little of their cash in a pot that will allow them to take advantage of investment opportunity in the future should they so desire. They are told that whatever they spend now cannot be returned to them if they choose to invest, but some part may be applicable to the price of the investment. In this way we can begin to examine directly the issues of option value and the sequencing of investment.

Another facet of investment behavior characterized by moving to this more realistic representation of time is the option of abandoning or mothballing projects if they become unprofitable. In some treatments, subjects will be given the option of setting the investment aside (abandoning it) and not drawing a return on it. In other treatments, subjects can set aside the investment but reserve the right to pay some price in the future to bring it back and draw income from it again.

All of these treatments can be combined with salient variables that are identified in Task 2 or 3. Interaction of time and various information conditions as noted will be quite important in developing our understanding of investment decisions in the field. This is also true of the interactions with various uncertainty conditions. Multiple investments will open the door to how subjects react to shifting portfolios of investment opportunities.

Task 5: The Effect of Market Structure on Investment.

The experiments described thus far have assumed that each subject is making investment decisions in a vacuum; the decisions of one subject have no impact on the investment opportunities or payoffs of the other subjects. In market terms, the subjects are monopolists in their individual markets with no threat of entry of a competitor. Yet with a few notable exceptions, this is not the structure of most markets.

In a purely competitive market, if one firm is able to make greater returns then the competition, others will enter (or adapt that firm's technology), which will bring profits back into equilibrium. While this is also an extreme case, it points out that interactions between firms are important to consider in the analysis. If one firm invests in a new technology, others may be forced to take action just to keep up.

There are many ways to introduce this type of interaction into the laboratory setting. It is possible that the investment opportunity will be limited (there is only a discrete number of funds that can be divided among a number of subjects). In this case, an auction may be necessary to achieve allocative efficiency amongst the subjects. In another treatment, subjects have to invest in order to keep unwanted competition from entering the market and

running the established firms out of the experiment. Pay-offs may also be affected by the number of subjects investing in a given opportunity or fund. In a simple example, the draws of return from the distribution may be made without replacement, thus complicating the investment decisions and their impacts.

As with the experiments in the other tasks, the market structure experiments can be combined with aspects of the other scenarios that we more fully understand. This will also allow for the relaxation of some of the controls of the laboratory and bring the actions of subjects closer to that which is observed in the field. While this task list is not meant to be an exhaustive description of the possible impacts and interactions that affect investment decisions, it seems to capture many of the important ones.

Conclusion

The tasks outlined above form the basis of a multi-year research program in the area of investment decisions. Because of the time involved in the careful planning and execution of experiments, and the attention that must be paid to the analysis of the results, we will undertake no more than Task 1 and 2 in the first year of this project. Once we have this framework in place, we can more easily address the other tasks and the issues they embody.

It is also important to note that this entire research program is but a first step in itself. It does not address the important area of consumer investment. As pointed out above, a large part of the problem with technological innovation lies in acceptance by consumers. Yet it can be argued that individuals do not necessarily make investment decisions the same way firms do. That is, we do not view the purchase of a refrigerator as having a return on our investment. Instead, we see many of these as things that we need to exist in the modern world, or live the life we would like.

But even at the end of a research program on consumer investment behavior, we would only have an understanding of two investment sectors in isolation. It will be important to put both parts together and observe how they interact. We can create a situation where the return to a firm's investment decision is not determined by a random draw from a distribution, but by the investment decision of a consumer also in the experiment. Thus, as we begin to understand and to draw all the pieces together, we can more clearly understand the entire process.

References and Selected Bibliography

Davis D.D. and Holt, C.A. [1993], *Experimental Economics*, Princeton, Princeton University Press.

Dixit, A. [1992], "Investment and Hysteresis," *Journal of Economic Perspectives*, Vol 6, No 1, p 107-132.

Lind R.C. [1992], "Global Warming Policy: The Need to Consider Real Options and Take a Sequential Approach," Final Report submitted to the United States Environmental Protection Agency, May.

Lind, R.C. [1992a], "The Analysis of Intertemporal Decisions Under Uncertainty in Energy Policy," Prepared for Workshop by the Task Force on Economic Analysis and Modeling Research Agenda, Secretary of Energy Advisory Board.

Pindyck, R.S. [1991], "Irreversibility, Uncertainty, and Investment," *Journal of Economic Literature*, Vol 29, No 3, p 1110-1148.

Smith V.L. [1992], "Experimental Economics: Evaluating Microeconomic Policies and Institutional Designs," Prepared for Workshop by the Task Force on Economic Analysis and Modeling Research Agenda, Secretary of Energy Advisory Board.

Thaler, R.H. [1992], *The Winner's Curse: Paradoxes and Anomalies of Economic Life*, New York, Free Press.

FORECASTING FUTURES MARKETS

MODERATOR Ethan Smith, U.S. Geological Service

PRESENTATION 1 Ronald Babula, Phil Colling, Gregory Gajewski, Economic Research Service

PRESENTATION 2 William Melick, Charles Thomas, Federal Reserve Board

PRESENTATION 3 David Torgerson, Economic Research Service

Forecasting Futures Markets

Employing Historical Price Dynamics to Aid in Forecasting Price Effects: The Case of Lumber Price Increases on Housing-Related Prices

Ronald A. Babula, Phil L. Colling, and Gregory R. Gajewski, Economic Research Service, U.S. Department of Agriculture

The increased volatility of the U.S. lumber market in late-1992 and early-1993 was evident from the large swings in the prices of lumber and lumber futures (hereafter, futures price) chronicled by Babula, Gajewski, and Colling (1993) and Gorte (1993). Lumber futures prices doubled between October 1992 and March 1993, and then fell 25 percent in April. Prices of lumber have also been moving widely, though by less than futures prices. Lumber prices were nearly 34 percent higher over the year ending April 1993, and then fell 4.7 percent in May. Since May 1993, lumber and futures prices have remained volatile, while levels have receded somewhat from early-1993 highs. Such swings are the largest since November 1978-August 1979 when lumber futures price rose 34 percent, and since November 1982-July 1983 when lumber price rose 22 percent with heightened housing activity during the economic recovery of the early 1980's (Gorte 1993; Babula, Gajewski, and Colling 1993).

Reasons for the increased volatility in the prices of lumber and lumber futures include: a recurring lumber demand increase during recovery from such recessions as that of the early-1990's; a reduction in timber harvests from the Pacific Northwest because of legal and environmental constraints; and perhaps because of some increased Japanese purchases, as that nation emerges from its recession (Gorte 1993). Whatever the cause, the movements (increases and decreases) over the last year should have impacts on lumber-related prices in the construction and housing sectors of the economy. Our aim is to use historical dynamic patterns of interactions among lumber and lumber-related housing and construction prices to discern the future effects and effect patterns elicited by the recent wide swings in the prices of lumber and lumber futures.

We use data-oriented statistical methods to ascertain how much and with what dynamic patterns related construction and housing prices should rise (fall) due to rises (declines) in lumber and lumber futures prices. For reasons stated below, we use vector autoregression (VAR) methods, in two separate experiments, to map the historical dynamic effects of:

(1) a one-time 10 percent rise in lumber price on futures price, the price of construction materials, the consumer price of housing, and the consumer price of shelter (hereafter called the lumber price experiment).

(2) a one-time 10-percent rise in lumber futures price on lumber price, the price of construction materials, the consumer price of housing, and the consumer price of shelter (hereafter, the futures price experiment).

Specifically, we set out to answer six questions concerning how shocks in lumber and lumber futures prices dynamically affect the remaining four respondent prices in each experiment: (a) What are the reaction times required for the prices to begin responding to each shock?; (b) What dynamic patterns do the monthly responses take?; (c) To what degree do the prices ultimately respond to each shock?; (e) What are the strengths of relationships among the five prices?; and (f) What are the differences in (a) through (e) elicited by a shock in lumber prices as opposed to a shock in lumber futures price (hereafter futures price)? Question (e) includes whether futures price responds to a greater (lesser) degree to lumber price movements than lumber price responds to futures price movements. Question (e) also includes whether lumber price movements elicit more (less) pronounced materials and housing-related price effects than effects elicited by futures price movements.

While not forecasts, such dynamics reveal how, on average, construction- and housing-related prices have historically responded to changes in the prices of lumber and lumber futures. These dynamics therefore provide insight concerning patterns, timing, and size for specific monthly 1993-1994 forecasts in the wake of the recent volatility in lumber and futures prices. As with all models, actual event-specific 1993-94 levels and movement patterns of lumber-related prices may differ from what the model's captured long run dynamic patterns would

predict and what the model's impulse response function would simulate. Yet the average historical dynamic patterns reflected by the model constitute an evidential point of reference for determining which 1993-94 point-forecasts (and forecasted price patterns) from competing models are "reasonable" by historical standards to characterize what may happen in the wake of the recent movements in lumber and lumber futures prices.

Methods, Model, and Data

Common sense, observed history, and economic theory suggest that large increases in the price of such a major construction/housing input as lumber should elicit increases in the prices of related lumber, construction materials, and housing/shelter prices. And our results presented below confirm as much. Answering questions (a) through (f) implies focusing not so much on if respondent prices react, as on how they react to shocks in lumber and futures prices. Conventional econometric models that intensively use static economic theory are equipped to handle questions concerning what happens at the static equilibria before and after the shock (Bessler 1984a, b). Such "structural" models often have little or nothing to say about what occurs dynamically between equilibria to the observed choice variables -- that is in answering the above six dynamic questions about each experiment's price responses (Bessler 1984a, b). Such time series models as Sims' (1980) vector autoregression (VAR) and Johansen and Juselius' (1990, 1992) maximum likelihood methods of vector error correction (VEC) models for cointegrated systems are equipped to address dynamic issues (a) through (f).

We chose to model the five-price system with a vector autoregression (VAR) model in logged levels over a vector error correction (VEC) model. For a VEC to be appropriate, the variables in levels must be individually nonstationary (integrated of order d, $d>0$), and exhibit stationary behavior as a system. That is, for a VEC to be appropriate, the five prices must be individually nonstationary, but form at least one stationary long run (cointegrating) relationship such that the prices move tandemly through time (Granger 1986; Hendry 1986; and Johansen and Juselius 1990, 1992). Yet results from Dickey-Fuller (1979) tests performed on the logged series suggest that evidence at the five-percent significance level is sufficient to reject the null hypotheses that each levels-variable is nonstationary.[1] The variables are stationary, thereby permitting us to model the system of lumber, futures, materials, housing, and shelter prices as a system (Johansen and Juselius 1990, 1992).

VAR econometric methods are equipped to address the dynamic inter-equilibria issues (a) through (f). The technique is data-oriented and imposes as few a priori theoretical restrictions as possible, so as to permit the dynamic regularities in the time-ordered data to reveal themselves (Bessler 1984a, b).

The literature is replete with detailed summaries and derivations of VAR methods. Those interested should consult Sims (1980), Bessler (1984a, b), and VanTassell and Bessler (1988). Our five-equation VAR model takes the following form:

$$
\begin{aligned}
x_t = a_{0,x} &+ a_{x,T}*\text{TREND} \\
&+ a_{x,1}*\text{LUMBER}_{t-1} + \ldots + a_{x,10}*\text{LUMBER}_{t-10} \\
&+ a_{x,11}*\text{FUTURES}_{t-1} + \ldots + a_{x,20}*\text{FUTURES}_{t-10} \\
&+ a_{x,21}*\text{MATLS}_{t-1} + \ldots + a_{x,30}*\text{MATLS}_{t-10} \\
&+ a_{x,31}*\text{HSG}_{t-1} + \ldots\ldots + a_{x,40}*\text{HSG}_{t-10} \\
&+ a_{x,41}*\text{SHELT}_{t-1} + \ldots\ldots + a_{x,50}*\text{SHELT}_{t-10} + R_{x,t} \qquad (1)
\end{aligned}
$$

The subscript t denotes the current value, while subscript (t-i) refers to the ith lag from the period-t value. The upper-cased subscript T represents the coefficient on time trend or TREND. On the left hand side, x = LUMBER, FUTURES, MATLS, HSG, and SHELT. The latter variable labels reflect, respectively, the prices of lumber, lumber futures, construction materials, consumer housing and consumer shelter. The coefficient with a nought subscript represents the intercept. $R_{x,t}$ represents white noise residuals.

Monthly producer and consumer price indices (PPIs, CPIs) obtained from the U.S. Bureau of Labor Statistics (BLS) served for all but the lumber futures price. The BLS PPI for lumber serves as lumber price

[1]The procedures for Dickey and Fuller's unit root test are described in Fuller (1976) and Dickey and Fuller (1979). The "pseudo-t" values for the five prices ranged from -3.35 to -6.92. Since all five pseudo-t values were negative, and of an absolute value exceeding that of the τ_μ critical value of -2.89, then evidence at the five percent significance level was sufficient to reject the null hypotheses that each variable in levels was nonstationary. The variables in levels are integrated of order zero (I(0)). Hence were the five prices modeled as a VEC in Johansen and Juselius' equation 1.2 (1990, p. 170), then π would be of full rank.

(LUMBER). The wholesale price of construction materials (MATLS) is represented by the PPI for construction materials. The CPI of all urban consumers for housing services represents the consumer price of residential housing services (HSG). What was desired was a broader price than that of newly-built home-owned residential units. The CPI for housing includes prices of owned and rented shelter services, as well as prices of upkeep, household expenses, and furnishings that are to some extent, lumber-dependent. The CPI of all urban consumers for shelter (SHELT) is a more narrowly defined price than the housing CPI. The CPI for shelter reflects the consumer price of rented and owned residential shelter.

The lumber futures price used is the closing price of the nearby futures contract (that contract nearest to expiration). To avoid potential problems associated with contract delivery, we did not use the nearby contract during its delivery month. If the nearby contract was into its last month, which is the delivery month, the next nearby contract was used. To stay consistent with the timing of the BLS index data, we used the closing price on the Tuesday of the week containing the 13th of that month.

Following VanTassell and Bessler (1988) and Bessler (1984a, b), the VAR model's lag structure was chosen using Tiao and Box's (1981) likelihood ratio test procedure (also see Lutkepohl 1981). Results (not reported here) suggest a 10-order lag. Each equation includes a constant, a time trend to account for time-dependent influences not of direct interest to this study, and a series of 11 indicator variables to account for seasonal influences.

The five VAR equations may have contemporaneously correlated innovations. Failure to correct for contemporaneously correlated current errors will produce impulse responses not representative of historical patterns (Sims 1980). A Choleski decomposition was imposed on the VAR for each experiment to orthogonalize the current innovation matrix, such that the variance/covariance matrix was identity in each experiment. Choleski decompositions resolve the problem of contemporaneous feedback.

Each decomposition requires an arbitrary imposition of a Wold causal ordering among the current values of the dependent variables (see Bessler 1984a, b). In the lumber price experiment, the chosen ordering was LUMBER, FUTURES, MATLS, HSG, and SHELT. The chosen ordering in the futures price experiment was the same except that the ordering of the lumber and lumber futures prices was reversed. The choices of these orderings were based on a number of considerations. First, common sense, observed history, and economic theory all suggest that there is a valid line of causality from lumber and lumber futures price movements to the lumber-dependent prices of construction materials, housing, and shelter. Second, futures and lumber prices move tandemly. Third, Sims (1989) and Bessler (1984a, b) suggest that when there is a valid line of causality, as from lumber and lumber futures prices to the other prices, then the variable shocked is placed atop the ordering. Fourth, the bottom three prices in each experiment were ordered as MATLS, HSG, and SHELT. Lumber is a prime residential housing input, so a shock in lumber or futures price could be expected to influence materials price before housing or shelter prices, suggesting MATLS' third place in each ordering.

VAR Model Diagnostics

Data for all five prices was available from January 1974 through December 1992 or 1974:1-92:12 (hereafter, months are denoted numerically with 1 reflecting January and 12 reflecting December). A 24-month period, 1974:1-75:12, was chosen as a period over which we applied the Tiao-Box lag search procedures. The 24-month period, 1991:1-92:12, was reserved as the validation period, over which the model's out-of-sample forecasts were evaluated. This left the period 1976:1-90:12 as an estimation period over which we initially estimated the model. We provide in-sample and out-of sample diagnostics for the initially estimated model. Because this model met the accepted diagnostic standards of in-sample fit and out-of-sample predictive accuracy, we then re-estimated the model over the "simulation" period of 1976:1-92:12, which includes the estimation and validation periods. The model estimated over the simulation period generated the two experiments' impulse response results and the forecast error variance (FEV) decompositions that we analyzed. The re-estimation was necessary to incorporate the maximum information possible for simulation of the near-future (1993-94) conditions (see Babula and Bessler 1990). All estimations and simulations were generated by Doan's (1990) package, Regression Analysis of Time Series (RATS).

The diagnostic test results for the model are reported. Ljung-Box portmanteau or Q tests and Dickey-Fuller tests on VAR equation residuals are two in-sample tests concerning model adequacy. Theil-U values are calculated for the 24 step-ahead forecasts permitted by the validation period, and provide insight on how reliably the estimated VAR model predicts events yet "unseen" beyond its own sample (Bessler and Kling 1984; Doan).

The Ljung-Box Q values, calculated for an equation's residuals, tests the null hypothesis that the equation has been adequately specified (Harvey 1990; Granger and Newbold 1986, p. 100). Evidence at the five percent significance level was not sufficient to reject the null hypotheses that each VAR equation was adequately specified. The Q values ranged from 32.4 to 46.4, and were less than the 58.1 critical chi-square value.

Stationarity (white noise residuals) of the estimated equations is required. We therefore tested for the stationarity of the innovations or residuals of each equation using the Dickey-Fuller (DF) tests. The pseudo- or t-like values on the nondifferenced lagged regressors in the τ_μ and τ_τ tests ranged from -14.04 to -14.46. Not only were these values negative, they also had absolute values which exceeded the critical values of -2.89 for the τ_μ test and -3.45 for the τ_τ test (Fuller 1976, p. 373). Evidence at the five percent significance level was sufficient to reject the null hypotheses that each VAR equation's residuals were nonstationary. Combined with the Ljung-Box evidence of model adequacy, these ADF tests on VAR equation residuals point to model adequacy.

Each equation (initially estimated over the 1976:1-90:12 period) generated as many "step-ahead" forecasts as the validation period would allow. The forecasts were run through a Kalman filter. Thus the 24-month validation period permitted 24 one-step-ahead forecasts; 23 two-step-ahead forecasts, etc. Theil-U values were provided for each of the 24 forecast horizons (i.e., 24 Theil-U's for each equation). A Theil-U value of less than unity suggests a superior (more accurate) performance at a horizon than the "naive" or random walk model forecasts. A naive model forecast equals the previous period's actual value. Further, a sub-unity Theil-U suggests that there were gains, in terms of forecast accuracy, from modeling the VAR equations as opposed to expending relatively no modeling effort through naively forecasting (Babula and Bessler 1990). Gains to modeling were apparent in the following three equations with the following proportions of Theil-U values below unity: 15/24 for materials price and 24/24 for housing and shelter prices.

Lumber and futures price forecasts confirm the FEV decomposition results (below) suggesting that these two prices are highly exogenous in each experiment. Typically, highly exogenous variables do not predict well beyond the sample, although they can (and do) add much to the prediction and simulation of relevant endogenous variables (MATLS, HSG, and SHELT). No more than five of the lumber and futures price Theil-U values were sub-unity. Lumber price is affected by other influences beyond the price influences of housing-related markets. Lumber price does not forecast well, as expected, and two reasons are cited. First, only one lumber futures price contract is traded on the Chicago Mercantile Exchange: the contract for Spruce-Pine-Fir, 2x4, Standard and Better Grade (Babula, Gajewski, and Colling 1993). This lumber is harvested primarily in Canada. Hence the lumber futures price, while perhaps a good "surface signal" of lumber price movements and anticipations, is too narrowly defined to serve as an instrumental variable for, and directly reflect, the supply and demand forces in all of the different U.S. lumber markets (Babula, Gajewski, and Colling 1993). And second, the anticipatory nature of the futures price renders its prediction hampered by expectations, and sometimes even emotions, which are difficult to capture in a model.

We conclude that, on balance, all five equations are adequately specified. Most (three: MATLS, HSG, and SHELT) of the five equations "passed" all three diagnostic checks, while all five equations passed most (at least two) of the three sets of diagnostic checks. While lumber and futures prices failed to generate Theil-U's that suggest gains to modeling, these equations did generate evidence of adequate model specification by accepted standards.

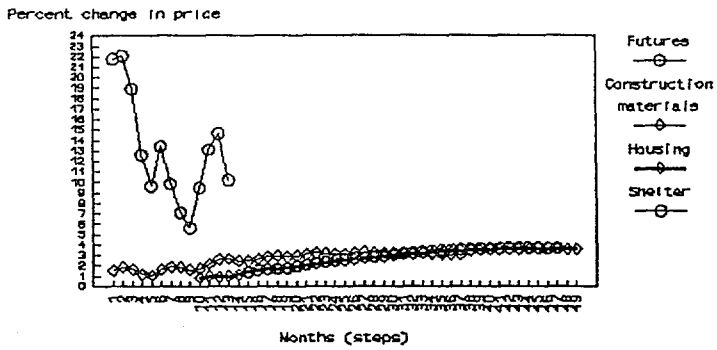
Impulse Responses in Related Prices to Shocks in Lumber and Futures Prices

The impulse response function simulates, over time, the effect of a one-time shock in one of a VAR's series on itself and on other series in the system (Bessler 1984a, b). Increases of 10 percent were chosen because one does not currently know at this writing to what point lumber and lumber futures prices will ultimately climb or recede. A 10-percent shock is conveniently sized because of the VAR model's linearity. The shapes of both experiments' impulse patterns in figures 1 and 2 would remain the same, with only the scales of the vertical axes varying with differently sized shocks.[2]

The impulse responses are reported for the lumber price experiment in figure 1 and the futures price experiment in figure 2. Dynamic aspects obtained from the impulse response results are summarized in table 1.

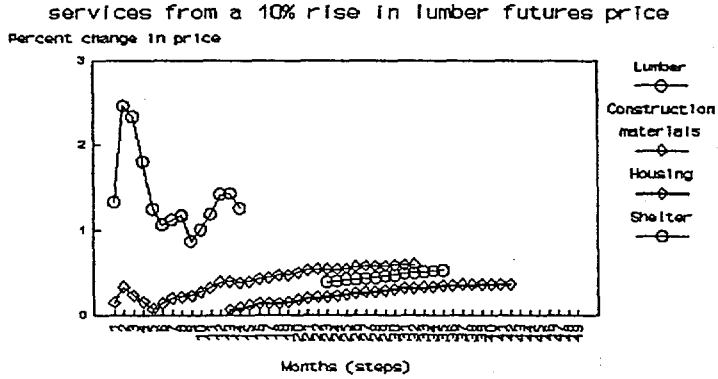
[2]For example, one can, by the model's linearity, characterize the impulse response simulations to 20-percent shocks by simply multiplying the impulses from the 10-percent shock experiments by a scaler of 2.0.

Figure 1. Impulse responses in the prices of lumber futures, construction materials, and housing services from a 10% rise in lumber price



Impulse responses are statistically non-zero at the 5-percent significance level. Impulses in the lumber futures price are included.

Figure 2. Impulse responses in the prices of lumber, construction materials, and housing services from a 10% rise in lumber futures price



Impulse responses are statistically non-zero at the 5-percent significance level.

Table 1 -- Dynamic aspects of price response patterns of the lumber and futures price experiments.

Dynamic Aspect:	Lumber Price	Futures Price	Materials Price	Housing Price	Shelter Price
Reaction times (months):					
Lumber price exp.	----	0	0	9	14
Futures price exp.	0	----	0	12	22
Response directions:					
Lumber price exp.	----	rise	rise	rise	rise
Futures price exp.	rise	----	rise	rise	rise
Response patterns:					
Lumber price exp.	---	sharp, then decay	shallow, then accelerate	shallow, then accelerate	shallow, then accelerate
Futures price exp.	sharp, then decay	----	shallow, then accelerate	shallow, then accelerate	shallow, then accelerate
Response durations (months):					
Lumber price exp.	----	13	37	40	33
Futures price exp.	14	----	32	31	13
Multipliers:					
Lumber price exp.	----	1.6	0.48	0.54	0.50
Futures price exp.	0.38	----	0.22	0.26	0.10

These results reflect patterns averaged over all of the sample's interactions. The results reflecting such average historical dynamics are valid in characterizing the dynamic effects of the recent rises in lumber and lumber futures prices insofar as recent/current conditions are similar to history's average conditions captured by the model. These long run average dynamics can provide standards by which to judge the "reasonableness" of monthly 1993-94 point-forecasts of competing models, as well as provide some indication on how forecasted price levels will be dynamically achieved.

Impulse responses are approximate changes in the non-logged prices, and are not price levels. Kloek and Van Dijk's (1981) Monte Carlo methods generated t-values for each impulse response. These values test the null hypothesis that each impulse is zero-valued. We focus on those impulses that were statistically nonzero at the five percent significance level.

Impulse responses: the lumber price experiment

Reaction times required for futures, materials, housing, and shelter prices to respond to shocks in the lumber prices have varied historically. As perhaps expected because of its speculative and anticipatory nature, futures price has begun reacting during the same month as (within 29 days of) the lumber price movement. And perhaps because of the short times required for construction materials to be manufactured, construction materials prices have also begun reacting to the lumber price shocks during the same month as the shock. On average, reaction times for the remaining prices have been longer: nine months for housing price and 13 months for shelter price. Reaction times of nine to 13 months may reflect the lags inherent in planning, constructing, and marketing (selling or renting) residential units. In addition to shelter price, housing price includes such other housing-related prices as furnishing and upkeep service prices that may be lumber-dependent, and that may respond to lumber shocks sooner than shelter prices. Hence, the housing price's reaction time is less than that required of the shelter price's responses.

Generally, history has had rises and falls in lumber price elicit similarly-directioned movements in futures, materials, housing, and shelter prices. While there may be event-specific examples in the past where this was not true, the model captures patterns that are averages over all interactions. Hence, these average patterns suggest that lumber price and the other prices generally moved up and down in a generally tandem manner.

The impulse patterns of the materials, housing, and shelter prices have been more enduring than those of futures price. Dynamic history has had futures price respond rapidly, sharply, and then take-on a gradually decaying pattern, while lasting just over a year, 13 months. This is consistent with recent media reports of sharply escalating and volatile lumber futures prices (see Taylor 1993). The impulse response patterns of the materials, housing, and shelter prices have differed from those of futures prices in being generally more muted in magnitude; in lasting for longer periods of time; and in taking-on patterns where responses gradually accelerate, rather than sharply decay, over the response period. These initially shallow patterns begin at low magnitudes and then accelerate to higher magnitudes that are still muted when compared in magnitude to the futures price impulses. The three patterns have endured from month-one through month-37 or about three years for materials price; from month-10 through 49 or just over three years for housing price; and from month-15 through month-47, or almost three years, for shelter price.

Multipliers of price response to lumber price shocks are calculated, and suggest the degrees to which prices ultimately respond to the shocks.[3] The multipliers provide history's average percentage point increase per point rise in lumber price. The futures price's multiplier for the lumber price experiment is 1.6 suggesting that, on average, each 10-point rise (fall) in lumber price has historically elicited a larger 16 percent rise (fall) in futures

[3]By a VAR model's very definition, each variable is posited as a function of a specified number, here 10, lags of each endogenously modeled variable in the system. Hence a one-time shock to the system places all five variables into cycles of monthly pulsation, including the shock variable. Insofar as the data levels are modeled in natural logarithms, then shocks to and impulse responses in, the logged variables constitute proportional changes in the non-logged variables, and percent changes in the non-logged variables when multiplied by 100. As an example, consider the materials price's multiplier. One first adds up the 37 statistically nonzero materials price impulses to obtain a cumulative percent change in MATLS response. One secondly sums the corresponding shock variable impulses into a cumulative change in lumber price. Finally, one then divides the percent change of the shock variable into the percent change in the response variable to obtain, here, the materials price's multiplier of percentage point response to a point change in lumber price. These are history's average responses. One calculates such an elasticity for the four respondent prices in each of the two experiments. A positive multiplier suggests that rises/falls in the shock variable have generally elicited the same in the respondent variable for which the multiplier is calculated.

price over a 13 month period.[4] Likewise, each 10-percent increase in lumber price has elicited, over periods ranging from 33 to 40 months, increases of 4.8 percent in materials price; 5.4 percent in housing price; and 5.0 percent in shelter price. The effects of the lumber price shock have been less than one-for-one on all prices except futures.

Impulse responses: the futures price experiment

The dynamic aspects with which lumber, materials, housing, and shelter prices have historically responded to futures price shocks are reported in table 1 and figure 2. Lumber and materials prices have historically reacted during the same month as the futures price shock. The housing and shelter prices have historically taken from 12 to 22 months to start reacting to the futures price shock. Generally, history has had lumber futures price and the four other prices move in similar directions.

Futures price increases have elicited lumber price increases that were pronounced early in the response cycle, that have decayed through the cycle's remainder, and that have endured, on average, for just over a year (14 months). The impulse responses of the materials, housing, and shelter prices have differed from those of lumber price in being generally more muted in magnitude; in lasting for longer periods of time; and in taking-on patterns that gradually accelerate, rather than sharply decay, over the response cycle. Construction materials price responses began in month-one and lasted for 32 months. Housing price impulses endured for a 31-month period beginning in month-13. Shelter price responses have lasted for 13 months after requiring nearly two years (22 months) to begin.

Multipliers for the futures price experiment are reported in table 1. These multipliers suggest that responses to futures price movements have been less than one-for-one. On average historically, a 10-percent increase has elicited an increase of 3.8 percent in lumber price over a 14 month period; an increase of 2.2 percent in materials price over a 32-month period; an increase of 2.6 percent in housing price over a 31-month period; and an increase of one percent in shelter price over a 13-month period.

Strength of Relationships: Analyses of Forecast Error Variance Decompositions

Analysis of decompositions of forecast error variance (FEV) is another tool of VAR econometrics for discerning relationships among the modeled system's time series. FEV is, at alternative horizons or steps, attributed to shocks in each of the dynamic system's series, such that a measurement of relative "strength" of relationships emerges (Bessler 1984a, b). Error decompositions attribute within-sample variance to alternative series and thus give measures which are useful in applied work. In table 2, the top portion contains the FEV decompositions of the lumber price experiment, and the bottom portion contains the FEV decompositions of the futures price experiment. Recall that the lumber price experiment was conducted under the following Choleski causal ordering:LUMBER, FUTURES, MATLS, HOUSING, and SHELT. The ordering for the futures experiment was the latter one with the reversal of the first two prices such that futures price is first and lumber price second.

Ordering influences the results, but perhaps not to as great an extent as is sometimes thought. Certainly, the materials, housing, and shelter prices, being lumber-related, are plausibly situated at the bottom of the orderings of the lumber and futures price experiments. Given this, FEV decompositions for these three variables, the sub-ordering of which is constant across experiments, take on the same patterns of FEV decompositions in both experiments. One can verify this by simply comparing FEV decompositions down a particular column across experiments. Further, the <u>combined</u> FEV decompositions take on the same patterns across experiments. For example, the combined explanation of materials price FEV from futures and lumber prices at month-24 is about 71 percent in both experiments. What differs across experiments is the breakdown of the combined FEV decompositions for these two prices. Hence, the degree to which lumber or futures price influences or explains uncertainty in the other prices depends on the ordering, that is on which of the two prices is placed causally first in the ordering.

[4]The PPI for lumber is an index of a number of lumber product prices at many pricing points, whereas the futures price is for a certain grade of lumber at one point. Yet the futures price can serve, and has often served, as a representative price for a wider array of lumber products. Therefore, we can not conclude that the futures price response is 1.6 times the "lumber" price response. However, the figures and table 1's results do indicate that futures price tends to overreact in the short-term. This is particularly evident from the futures price impulses of figure 1.

Table 2 -- Decompositions of forecast error variance (FEV) for the lumber and futures price experiments.

		Percent Explanation of Forecast Error Variance from:				
Variable:	Step	Lumber Price	Futures Price	Materials Price	Housing Price	Shelter Price
FEV results from the lumber price experiment:						
Lumber price:	1	95.36	3.96	0.03	0.23	0.42
	3	84.88	6.06	4.64	4.10	0.31
	6	70.37	4.49	10.19	12.26	2.69
	12	65.08	3.62	11.32	16.09	3.88
	18	58.79	3.35	20.53	13.17	4.15
	24	53.30	3.39	28.58	10.67	4.26
	30	47.20	3.63	35.21	9.47	4.48
	36	42.78	3.92	39.04	9.42	4.83
Futures price:	1	35.35	64.17	0.12	0.03	0.33
	3	35.22	54.26	5.88	4.33	0.31
	6	34.49	47.51	10.66	6.77	0.57
	12	37.04	38.83	13.61	9.77	0.76
	18	36.13	36.49	15.94	10.64	0.80
	24	35.67	35.09	18.21	10.25	0.78
	30	34.51	33.93	20.59	10.09	0.88
	36	33.59	32.91	22.05	10.26	1.19
Materials price:	1	42.65	1.35	55.26	0.71	0.03
	3	39.24	0.73	52.25	7.52	0.26
	6	42.45	1.20	42.24	13.01	1.09
	12	53.67	0.85	24.59	20.22	0.67
	18	65.05	1.04	16.07	17.20	0.64
	24	69.00	1.59	16.36	12.17	0.88
	30	67.41	2.16	20.99	8.45	1.00
	36	63.59	2.64	26.57	6.29	0.91
Housing price:	1	0.10	0.18	0.09	99.16	0.47
	3	0.29	0.21	0.09	97.95	1.47
	6	0.72	1.39	0.44	95.52	1.93
	12	9.68	2.82	0.55	82.63	4.32
	18	27.67	5.50	1.10	58.59	7.13
	24	46.15	5.71	0.92	37.81	9.41
	30	58.23	4.82	0.91	25.73	10.30
	36	64.84	3.97	2.78	18.42	9.99
Shelter price:	1	0.23	0.24	0.19	74.61	24.73
	3	0.51	0.16	0.37	73.79	25.17
	6	0.66	1.88	1.42	69.45	26.59
	12	5.30	2.10	1.60	57.03	33.97
	18	16.04	2.64	2.76	41.75	36.81
	24	30.04	2.35	2.79	30.28	34.54
	30	42.78	1.94	2.12	23.33	29.83
	36	52.14	1.66	3.44	18.21	24.55

FEV results from the futures price experiment:

Lumber price:	1	54.95	44.37	0.03	0.23	0.42
	3	44.01	46.93	4.64	4.10	0.31
	6	37.01	37.84	10.49	12.26	2.69
	12	34.83	33.87	11.32	16.09	3.88
	18	31.13	31.01	20.53	13.17	4.15
	24	27.63	29.01	28.38	10.67	4.26
	30	23.97	26.86	35.21	9.47	4.48
	36	21.46	25.24	39.04	9.42	4.83
Futures price:	1	1.21	98.31	0.12	0.03	0.33
	3	1.89	87.58	5.88	4.33	0.31
	6	2.71	79.28	10.66	6.77	0.57
	12	6.48	69.39	13.61	9.77	0.76
	18	6.78	65.84	15.94	10.64	0.80
	24	7.12	63.64	18.21	10.25	0.78
	30	6.90	61.54	20.59	10.09	0.88
	36	6.83	59.67	22.05	10.26	1.19
Materials price:	1	29.37	14.63	55.26	0.71	0.03
	3	27.12	12.85	52.25	7.52	0.26
	6	33.11	10.54	42.24	13.01	1.09
	12	37.71	16.82	24.59	20.22	0.67
	18	42.46	23.62	16.07	17.20	0.64
	24	42.37	28.23	16.36	12.17	0.88
	30	39.49	30.08	20.99	8.45	1.00
	36	35.86	30.37	26.57	6.29	0.91
Housing price:	1	0.00	0.28	0.09	99.16	0.47
	3	0.06	0.44	0.09	97.95	1.47
	6	0.08	2.03	0.44	95.52	1.93
	12	3.50	8.99	0.55	82.63	4.31
	18	10.57	22.60	1.10	58.59	7.13
	24	20.51	31.35	0.92	37.81	9.41
	30	28.65	34.41	0.91	25.73	10.30
	36	33.83	34.98	2.78	18.42	9.99
Shelter price:	1	0.07	0.40	0.19	76.61	24.73
	3	0.31	0.36	0.37	73.79	25.17
	6	0.29	2.25	1.42	69.45	26.59
	12	2.09	5.31	1.60	57.03	33.97
	18	7.21	11.46	2.77	41.75	36.81
	24	15.82	16.57	2.79	30.28	34.54
	30	24.42	20.29	2.12	23.33	29.83
	36	30.72	23.08	3.44	18.21	24.55

Lumber price is highly exogenous in the lumber price experiment, especially at horizons of 18 months or less, with 59 to 95 percent of lumber price's FEV self-attributed. After 18 months, this percentage drops below half and suggests lumber price's increasing exogeneity at the longer horizons. After 18 months, materials price takes on increasing importance in explaining lumber price FEV -- steadily up to 39 percent by the 36th month horizon. These FEV results coincide with the same experiment's impulse response results: lumber price shocks have required from two to three years before related price effects have played themselves out.

In the futures price experiment, lumber price has been more endogenous than in the lumber price experiment. The lumber price FEV has been from 21.5 to 55 percent self-attributed, and from 25 to 47 percent attributed to futures price variation.

Futures price becomes more endogenous in the lumber price experiment beyond the month-3 horizon, when no more than about 47 percent of its FEV attributed to own-variation. From 34 to 37 percent of futures price FEV is attributed to the shock variable, lumber price. Futures price is more exogenous in the futures price experiment, with from 60 to 98 percent of its FEV having been self-attributed.

Whether lumber price affects futures price or futures price affects lumber price depends on which of the two prices is placed causally first in the ordering as the shock variable, that is on which variable which moves first. Each shock variable exhibits high degrees of exogeneity, and explains comparable FEV proportions of each other, in the two experiments. So across experiments, the patterns with which lumber price accounts for futures price variability is similar to the patterns with which futures price accounts for lumber price variability. But considering these FEV results in combination with the two experiments' impulse response results adds further insight. While lumber and futures explain similar percentages of each other's FEV across experiments, the impulses suggest that the sizes of effects vary noticeably across experiments. Compared with the futures price experiment, lumber price accounts for similar percentages of larger price effects.

Materials price's FEV becomes increasingly dependent on the combined variation of lumber and futures price after the one-year horizon. After a year, no less than 66 percent of materials price FEV is attributed to variation in lumber and futures prices. This coincides roughly with the materials price impulses (figures 1 and 2) which require 12 to 18 months to achieve or approach peak strength levels in response to the shocks.

Housing price is highly exogenous early on, with no less than 59 percent of its FEV being self-attributed through the 18-month horizon. Thereafter, the combined variation of lumber and futures prices accounts for most (52-69 percent) of housing price's FEV. These results reinforce the housing price impulse responses which take from nine to 12 months to activate, and from 2.5 to three years after the shock to peak in strength (figures 1 and 2).

Shelter price remains highly endogenous in both experiments, with from 25 to 37 percent of its FEV being self-attributed at all reported horizons. The price's exogeneity peaks at month-18, after which contributions of own-variation to FEV steadily and moderately decline. These results are not inconsistent with the shelter price impulses of figures 1 and 2, where from 14 to 22 months are required for responses to begin, where responses remain within a narrow band of impulse magnitude over the response cycle, and where responses require from 2.5 to three years to approach peak strength.

Comparative Dynamics of the Lumber and Futures Price Experiments

Compared with a futures price increase, a 10-percent rise in lumber price has elicited price increases that form similarly shaped patterns, but which have engaged more rapidly; have endured for longer periods; and have achieved generally greater magnitudes. This is evident from table 1. When reaction times for a price across experiments have differed, those of the lumber price experiment have been shorter. A lumber price shock's effects were generally more enduring than price effects elicited by a futures price shock (see table 1, response durations). Figures 1 and 2 reveal, however, that the impulses of the two experiments take on similar monthly patterns -- reaction times and durations notwithstanding.

The most evident difference in the price effects from shocks in lumber and futures prices lies in the effect magnitudes. Differences are apparent from comparing the scales of figures 1 and 2 which contain plotted impulses, and from table 1's response multipliers. Lumber and futures price movements elicit statistically significant responses in each other, as well as in materials, shelter, and housing prices. Yet a lumber price shock has elicited far greater impulse magnitudes than a similarly sized shock in futures price.

Another interesting comparison involves futures price responses to lumber price shocks as opposed to lumber price responses to futures price shocks. Figures 1 and 2 reveal that lumber and futures prices

respond to each other with similar shapes and for comparable periods. Yet figures 1 and 2, as well as table 1's multipliers, suggest that futures price "over-reacts" relative to lumber price. The multipliers suggest that historically, percentage futures price response to lumber price movement has been four times the percentage lumber price response to futures price movement.

There were similarities and dissimilarities in FEV decomposition patterns across experimental orderings. The percentages of FEV of all prices attributed to variations in the three prices not influenced by the experimental orderings -- materials, housing, and shelter prices -- were the same across experiments. The combined influence of lumber and futures price uncertainty on the FEV's of the system is the same across experiments, although the distribution of this combined influence on the prices' FEV's varies according to which of the two prices is placed first in the ordering and engages the system's shock. The lumber price percentage of this combined influence is greatest in the lumber price experiment, while the futures price's percentage is greatest in the futures price experiment. But taken together with the impulse response results, another result emerges: the FEV decompositions are similar across experiments, but these similar percentage patterns hold for differently sized effects.

Summary and Conclusions

The average monthly dynamic patterns captured by the model can characterize what may happen to the five prices from the recent volatile swings in lumber and lumber futures prices.

Lumber price increases should elicit materials price increases that begin within a month, that gradually accelerate over about a year before peaking in magnitude, and that endure for about three years. Response should be less than one-for-one, with materials price rising by about half of the percent increase in lumber price over this period. A futures price increase should similarly influence materials price, but by less and for shorter time periods. Some time (up to 14 months) would elapse before lumber price increases noticeably influence housing and shelter prices. Patterns of housing and shelter prices increases would begin at low levels and then gradually gain strength, while enduring at least 2.5 years. Responses would be less than one for one, with housing and shelter price rising by about half of the percent increase in lumber price. Housing and shelter price impulses from futures price shocks would be more delayed, weaker, and shorter-lived than those elicited by lumber price movements.

Futures price should respond to a far greater degree to lumber price than lumber price would respond to futures price. Each percent rise in lumber price should elicit a greater 1.6 percent increase in futures price over about a year, while lumber price should respond by about a quarter of that percentage to similar increases in futures price.

Therefore, each increase, say of 10 percent, in futures price should elicit responses in construction and housing related prices that are often more delayed, that are weaker in strength, and that are less enduring than similar movements in lumber price. Compared with lumber price responses to futures price movements, futures prices should over-react to changes in lumber price. So to elicit given changes in materials, housing, and shelter prices, futures price would have to swing more widely than futures price.

References

Babula, R.A., G. Gajewski, and P. Colling. "Lumber Prices Linked to Housing Markets." Agricultural Outlook AO-197(June, 1993):27-29.

Babula, R.A. and D.A. Bessler. "The Corn-Egg Price Transmission Mechanism." Southern Journal of Agricultural Economics 22(no. 2, Dec. 1990):79-86.

Bessler, D.A. "An Analysis of Dynamic Economic Relationships: An Application to the U.S. Hog Market." Canadian Journal of Agricultural Economics 32(1984a):109-24.

Bessler, D.A. "Relative Prices and Money: A Vector Autoregression on Brazilian Data." American Journal of Agricultural Economics 66(1984b):25-30.

Bessler, D.A. and J.L. Kling. "A Note on Tests of Granger Causality." Applied Economics 16(1984):335-42.

Dickey, D. and W. Fuller. "Distribution of the Estimates for Autoregressive Time Series with a Unit Root." Journal of the American Statistical Association 74(1979):427-31.

Doan, T. Regression Analysis of Time Series, Users' Manual, Version 3.0. Evanston, Illinois:VAR Econometrics, 1990.

Ericsson, N. "Cointegration, Exogeneity, and Policy Analysis: An Overview." Journal of Policy Modeling 14(no. 3, June 1992):251-80.

Fuller, W. Introduction to Statistical Time Series. New York:John Wiley and Sons, 1976.

Gorte, R.W. "Lumber Prices." Memorandum, Environment and Natural Resources Policy Division, Congressional Research Service, Library of Congress, Washington DC, March 10, 1993.

Granger, C.W.J. "Developments in the Study of Cointegrated Economic Variables." Oxford Bulletin of Economics and Statistics 48(1986):213-28.

Granger, C.W.J. and P. Newbold. Forecasting Economic Time Series. New York:Academic Press, 1986.

Harvey, A. An Econometric Analysis of Time Series. Cambridge, MA: The MIT Press, 1990.

Hendry, D. "Econometric Modelling with Cointegrated Variables: An Overview." Oxford Bulletin of Economics and Statistics 48(1986):201-12.

Johansen, S. and K. Juselius. "Testing Structural Hypotheses in a Multivariate Cointegration Analysis of the PPP and UIP for the UK." Journal of Econometrics 53(1992):211-44.

Johansen, S. and K. Juselius. "Maximum Likelihood and Inferences on Cointegration: With Applications to the Demand for Money." Oxford Bulletin of Economics and Statistics 52(1990, no. 2):169-210.

Kloek, T. and H.K. VanDijk. "Bayesian Estimates of Equation System Parameters: An Application of Monte Carlo." Econometrica 46(1978):1-20.

Larue, B. "Farm Input, Farm Output, and Retail Food Prices: A Cointegration Analysis." Canadian Journal of Agricultural Economics 39(1991):335-53.

Lutkepohl, H. "Comparison of Criteria for Estimating the Order of a Vector Autoregression Process." Journal of Time Series Analysis 6(1985):35-52.

Sims, C. "Macroeconomics and Reality." Econometrica. 48(1980):1-48.

Sims, C. "Models and Their Uses." American Journal of Agricultural Economics 71(1989):489-94.

Taylor, J. "Bull Market in Lumber May be Over." Wall Street Journal
Section C, April 2, 1993, pp. C1 & C12.

Tiao, G. and G.E.P. Box. "Modeling Multiple Time Series: With Applications." Journal of the American Statistical Association 76(1981):802-16.

VanTassell, L.W. and D.A. Bessler. "Dynamic Price Adjustments Between Commercial and Purebred Cattle Markets." Southern Journal of Agricultural Economics 20(no. 2, Dec. 1988):137-43.

Reading the Market's Mind: Using Crude Oil Options to Recover the PDF

William Melick and Charles Thomas, Federal Reserve Board*

Energy economics is often concerned with decision making under uncertainty. Derivative markets, used to offset risk, can help quantify this uncertainty. Option markets in particular, given their contingent nature, have been used by economists in order to better understand the uncertainty in a give market. (One such example is Overdahl and Matthews (1988), who used options to construct confidence intervals for the future spot price of oil.)

Naturally, the standard tools developed in finance have been brought to bear on these questions. However, these tools are not well designed to handle some of the situations encountered in the energy markets. During unsettled times (that is, most of the time in energy markets) the assumption that prices are drawn from a lognormal distribution, which lies behind the Black-Scholes (1873) option pricing model, does not seem appropriate. Yet the usual alternative from the finance literature, that prices follow a jump diffusion process (e.g. Bates (1990), seems overly restrictive.

An alternative method to quantify uncertainty has been developed by Melick and Thomas (1993) (hereafter MT). They use option prices, imposing as little structure as possible, to recover the market's implied distribution for the underlying commodity. As an application, they consider options on crude oil futures during the Persian Gulf crisis. This paper presents an overview of their technique, using data from the time of the Persian Gulf crisis and the summer of 1993 for illustrative purposes. The paper is organized as follows. The first section presents the intuition behind the MT procedure, using the straightforward case of European-style options as a starting point. The second section examines several episodes from the Persian Gulf crisis, while the third section used illustrations from the most recent period. Conclusions are presented in the final section.

I. Option Prices and PDFs

A call option allows its holder to buy the underlying commodity at the agreed upon strike price. Therefore, the call option has value at expiration (is in the money) whenever the price of the commodity is above the strike price. Unlike an American option, European options can be exercised at the time the contract expires. Therefor, the value of a European call option, with a strike price denoted by X, can be expressed as the product of two terms

$$C[X] = \int_X^{\infty} f(p)\, dp \cdot \left(\left(\int_X^{\infty} p \cdot f(p) dp \Big/ \int_X^{\infty} f(p)\, dp \right) - X \right)$$

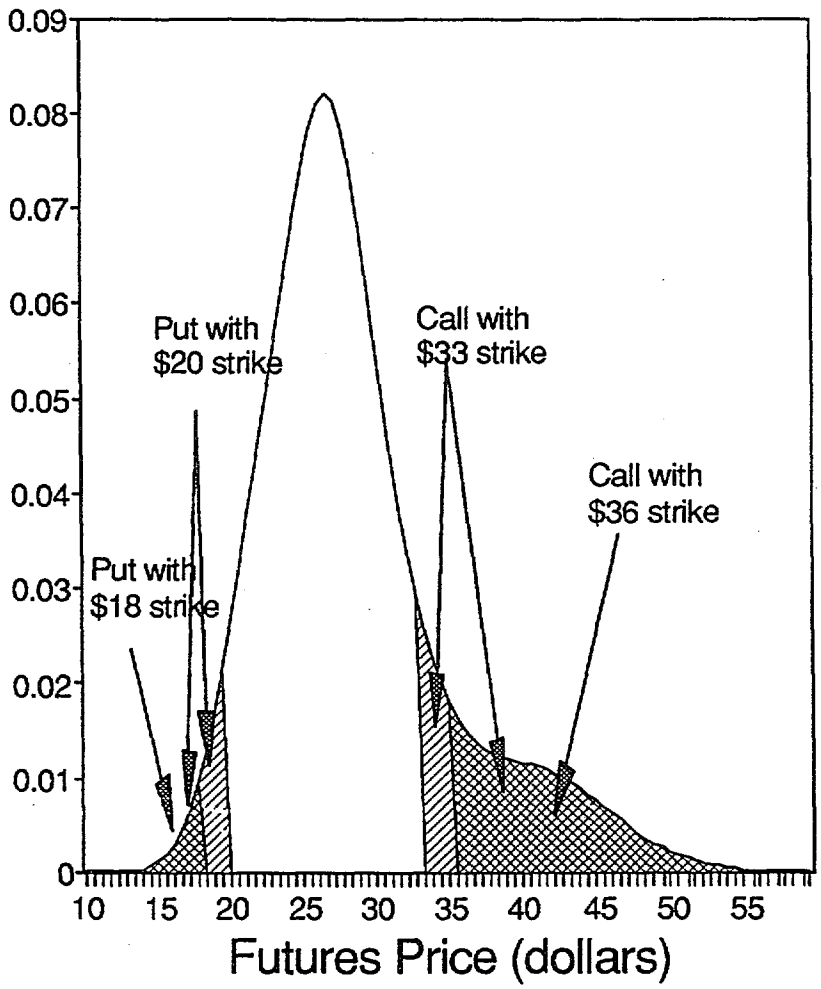
where p is the price of the commodity, and f(p) is its PDF. The first term represents the probability that the price of the commodity, at expiration, is above the strike price. The second term represents the difference between the expectation for the price of the commodity, given that it is above the strike price, and the strike price. Both of these terms are integral of the probability density function (PDF) for the underlying commodity's price. Give prices for European options with different strike prices, a functional form for the PDF, and a nonlinear solver, it is relatively straightforward, to "invert" a set of equations like (1) and recover the parameters of the PDF. The technique involves searching for the parameter values that minimizes the pricing errors generated by the set of equations.

These concepts are illustrated in Chart 1 which considers what portions of a PDF could be recovered with four options: two puts with strikes of $18 and $20, and two calls with strikes of $33 and $36. The $36 call would only have value if the price of the underlying commodity were above $36, therefore, it only offers information about the portion of the PDF above $36. The $33 call also provides information about the PDF above $36, as well as filling in the portion of the PDF between $33 and $36. The story for the puts is similar. The $18 put helps to

* This paper represents the views of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or other members of its staff. The authors may be reached by writing to William Melick, Mail Stop #42, Federal Reserve Board, Washington, D.C. 20551, or by calling (202) 452-2296.

Chart 1

Options Values and the Density Function



recover the tail of the PDF below $18. The $20 put also helps to recover this lower tail, as well as filling in the part of the PDF between $18 and $20.

It is useful to note how data limitations and the assumed functional form for the distribution interact. The fact that strikes are at discrete intervals and, more importantly, that they do not span the entire possible range of the price of the commodity, places an important limitation on what the option prices can reveal about the distribution. The recorded option prices only contain information about the conditional expectation and probability mass in the following segments of the range of the price: 1) the segment below the lowest strike, 2) the segments between each strike, and 3) the segment above the highest strike. Any number of distributions could generate the same results for these conditional expectations and probabilities. For example, for any given distribution we can construct a second distribution out of a series of non-overlapping uniform densities which will be observational equivalent to the given distribution. Any estimated distribution requires careful interpretation, especially in the regions below the lowest strike and above the highest strike. For crude oil, strikes are almost always $1.00 apart (in a few instances $5.00), allowing a fine demarkation of the distribution within the range of strikes. However, the shape of the distribution in the tails will depend importantly on the functional form assumed for the distribution.

The recovery of the PDF is more complicated for American options. These options can be exercised at any time prior to expiration, adding to their value relative to European options. Thus, the value of an American call option, in terms of the PDF, cannot be written as compactly as equation (1). The idea of MT is to bound the value of the American option using two equations (similar in spirit to equation (1)). As before, parameter values are found that minimize the pricing errors generated by a weighing of the pairs of bounds in the set of equations.

The method of MT requires one important condition to derive the bounds; namely, that the price of the underlying commodity martingales. This is not unreasonable for options on futures, as there is no cost in holding a futures contract. In brief, the martingale assumption means that today's price for the commodity equals the expectation for the price of the commodity at expiration. (For evidence that crude oil features martingale see Dominguez (1989), Kumar (1992), and Deaves and Krinsky (1992)).

The intuition behind the bounds is fairly straightforward. Given that we know the value of a European option, the upper bound will maximize the premium associated with early exercise, subject to the martingale assumption. The early exercise premium will be maximized if prices moved tomorrow and stayed fixed thereafter through expiration. Conversely, the early exercise premium would be minimized if the price did not move until just before expiration, so that the option holder would only exercise today or at expiration. In this situation there would be little value associated with the right of early exercise. Both the upper and lower bound, given the martingale assumption, can be written in terms of the PDF for the price of the underlying commodity. A detailed derivation and further intuition is presented in MT.

II. Application to Crude Oil: Persian Gulf Crisis

Section I showed that the MT method allows a researcher to recover the PDF for the price of the commodity underlying an American option given: 1) observed options prices, 2) an assumed functional form for the PDF, and 3) a nonlinear optimization routine. Throughout the Persian Gulf crisis, market commentary focused on three distinct outcomes: 1) a return to pre-Crisis conditions (e.g. Iraq would peacefully withdraw from Kuwait), 2) a severe disruption to Persian Gulf oil supplies (e.g. damage to Saudi Arabian facilities during a war), and 3) a continuation of unsettled conditions over the relevant horizon (e.g. a prolonged stalemate in which outcome 1 or 2 might eventually occur). Given theses three possibilities, we chose a mixture of three lognormals (MLN) as the form of the distribution to be estimated, written as

$$g = \pi_1 \cdot g_1[u_1,\sigma_1] + \pi_2 \cdot g_2[u_2,\sigma_2] + \pi_3 \cdot g_3[u_3,\sigma_3].$$

If in fact market participants felt that prices were likely to be drawn from a trimodal distribution this could be easily captured by the mixture. Ex ante, we expected that as news hit the market, the relative weighing of the three lognormals might change, as well as the parameters of each of the three lognormals. For example, news of an Iraqi rocket attack on a Saudi Arabian oil field might increase the weighing on the lognormal distribution with the highest mode, as well as increase the revellent range encompassed by this lognormal distribution.

As an alternative, we also followed Overdahl and Matthews and recovered a single lognormal density (roughly the Black-Scholes model) using the Barone-Adesi and Whaley (1987) approximation to account for the American style options. These estimations correspond to a single lognormal (SNL), and would be roughly equal to those from MLN in which $\pi_1 = \pi_2 = 0$.

Estimated distributions for selected events during the Persian Gulf crisis (both SLN and MLN) are found in Charts 2 and 3. Comparing the estimated PDFs from the two models immediately before and after the receipt of "news" allows us to infer how the market interpreted the news and highlights the differences between the NLN and SLN models.

On Thursday, October 25, 1990, the London Financial Times carried a report that Iraqi forces had attached explosives to 300 of Kuwait's 1000 oil wells, quoting a senior Kuwaiti engineer who had left Kuwait one week earlier. This revelation pushed oil prices up sharply, with the futures contract nearest to expiration (December) rising $3.17 per barrel. Chart 2 plots the PDFs from MLN and SLN for October 22 (top panel) and October 25 (bottom panel)using the January contract. On October 22, market expectations for future prices were centered quite tightly around $24 per barrel. The news of the mining widened each model's distribution significantly, with the MLN allowing for a sizeable probability mass between $60 and $70 per barrel.

The largest one-day change in oil prices in NYMEX history occurred in Thursday January 17, 1991 when 1) several governments announced a coordinated release of oil from their emergency inventories and 2) it became clear that the coalition forces had total air supremacy. On that day the settle price for the March contract fell $9.66 while the settle price for the April contract fell $7.82. The six panels of Chart 3 trace the evolution of expected PDFs on the days surrounding January 17. Prior to the first air strike (as can be seen in the first two panels), the market was still expecting a fairly significant chance of a major oil market disruption (perhaps Iraqi damage os Saudi Arabian oil facilities) that could push prices to the $40-$60 per barrel range. On January 17th these PDFs tightened dramatically, and on ensuing days the PDF generated from MLN moved closer and closer to that from SLN. By January 23, there was little difference between the two PDFs, as the market returned to almost a pre-crisis distribution.

III. Current Events: The Return of Iraq

Oil prices fell from June through mid-August of 1993 (the time this paper went to press), after recovering earlier in the year. OPEC production restraint in the first part of the year had buoyed prices, but Kuwait's decision not to participate in the cartel's third-quarter quota allocations resulted in price declines. Most recently, signs of apparent progress in talks between Iraq and the United Nations, coupled with frictions within OPEC on the appropriate response to this progress, have pushed prices back down to 1992 year-end levels.

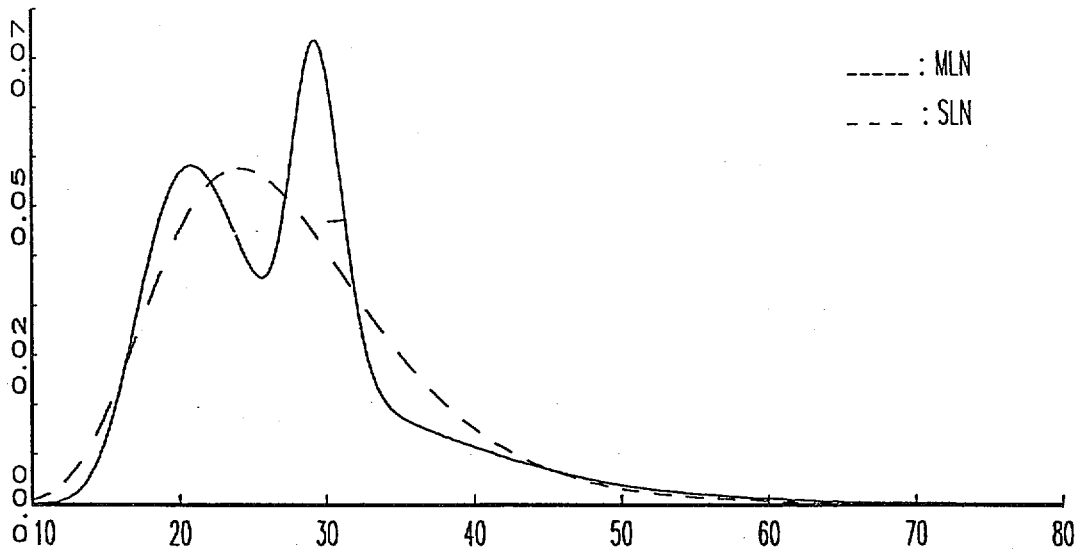
The one-sided nature of Iraq's return to the oil market argues against a standard application a la Overdahl and Matthews to options on crude oil futures. However, the m mixture of three lognormals used by MT for the Persian Gulf crisis seems excessive. Therefore, a mixture of two lognormals was used to recover the MLN PDFs drawn in Chart 4. Ex ante it seemed reasonable that market participants might envision a bimodal distribution for oil prices during this period, with a significant chance that Iraq could begin exporting as will as a significant chance of a continued embargo.

As can be seen in the top panels of the chart, on June 30, 1993, before reports of progress in talks with Iraq, expected oil prices at the expiration of both the November 1993 contract (the left panel) and the March 1994 contract (the right panel) were distributed in a fairly smooth bell shape typical of a lognormal distribution. However, by mid July, when indications of progress in the talks were greatest, market perception had changed. Oil prices had softened, and as shown in the middle panels, for both delivery months (but especially March 1994) the market now entertained the possibility of sharply lower prices. Comparing the two middle panels, market participants felt that the chance of an Iraqi sale was increasing with the passage of time, as seen by the much larger left-hand tail for the March 1994 contract than for the November 1993 contract. By March of 1994, the market felt there was a seven percent chance that oil might be selling for less than 15 dollars per barrel.

Data from August 9th were used for the two bottom panels. As prospects for an Iraqi sale have diminished so has the mass in the left-hand tail, with prices once again following a bell-shaped distribution -- albeit with a lower mean price than at the end of June. As of August 9th, the market only saw a 3 percent chance of oil selling for less than 15 dollars per barrel.

Chart 2

JAN Contract on 22-Oct Futures = 28.95



JAN Contract on 25-Oct Futures = 32.19

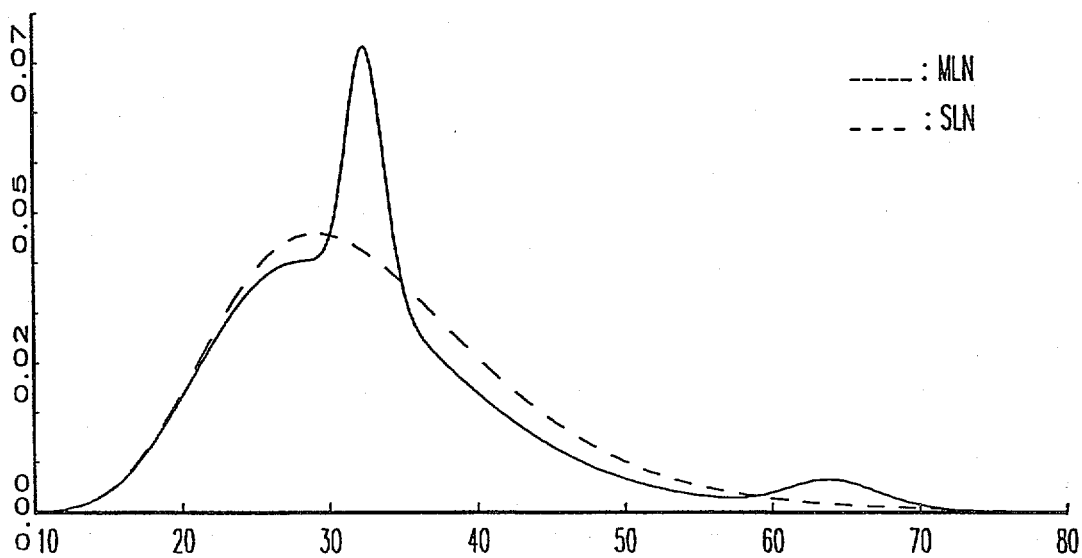
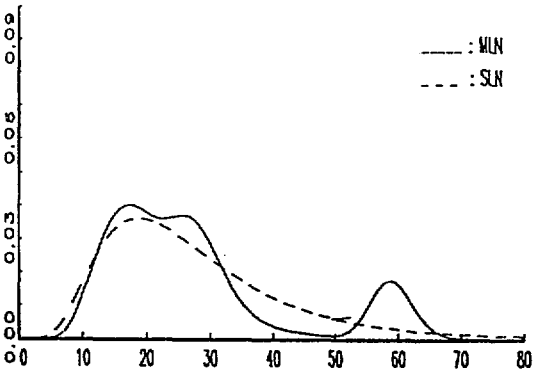
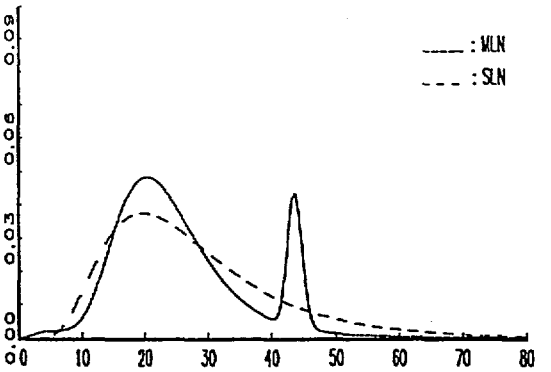


Chart 3

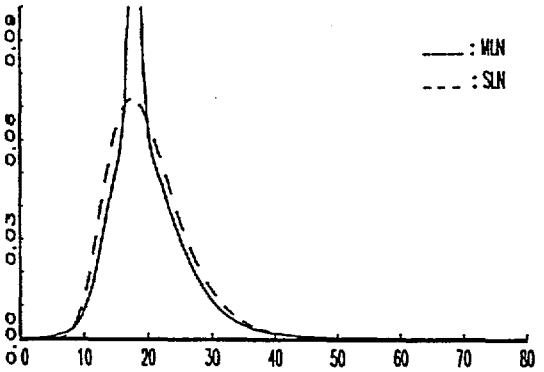
APR Contract on 14-Jan Futures = 26.74



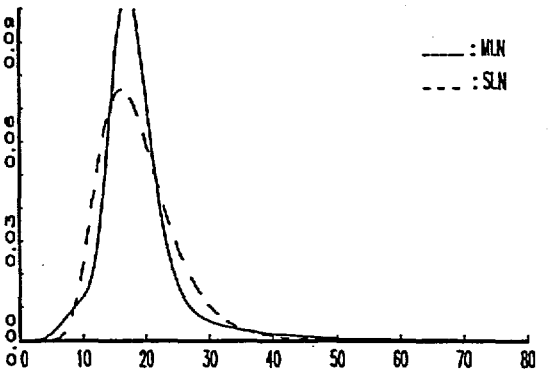
APR Contract on 16-Jan Futures = 27.95



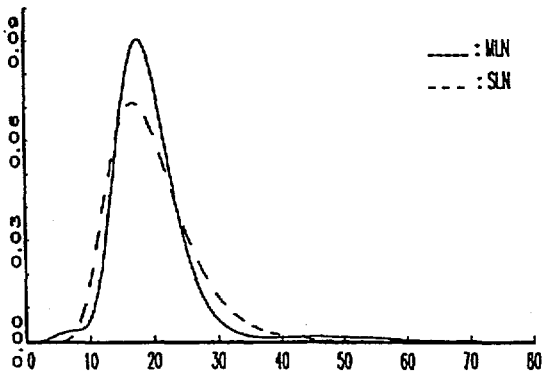
APR Contract on 17-Jan Futures = 20.13



APR Contract on 18-Jan Futures = 18.78



APR Contract on 21-Jan Futures = 19.77



APR Contract on 23-Jan Futures = 21.02

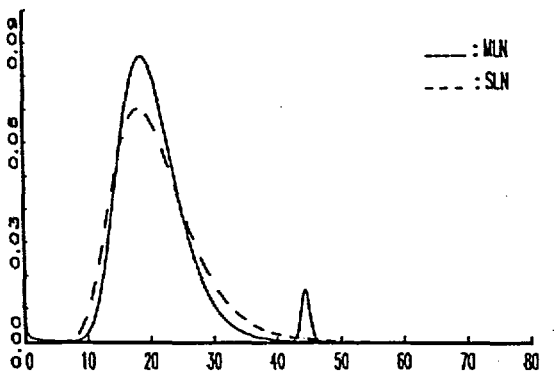
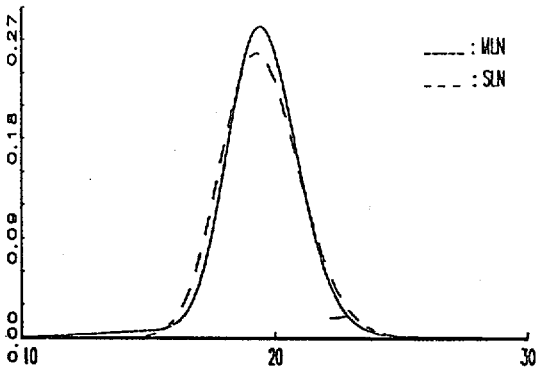
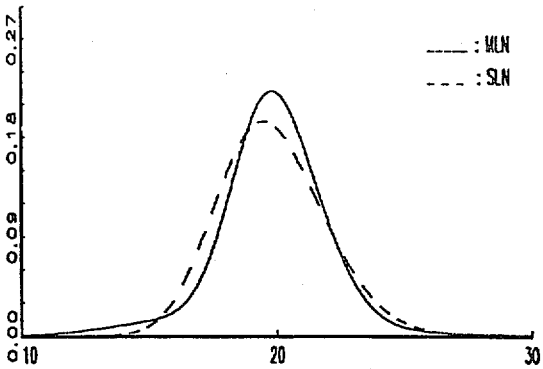


Chart 4

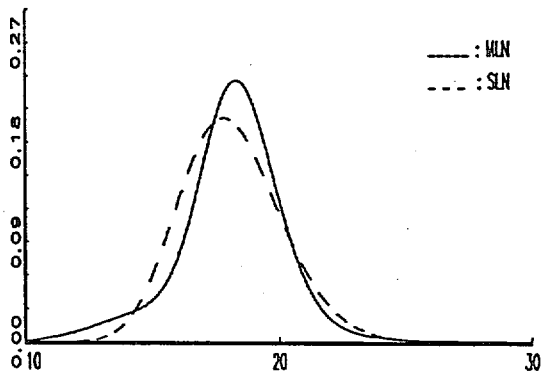
NOV Contract on 30-Jun Futures = 19.46



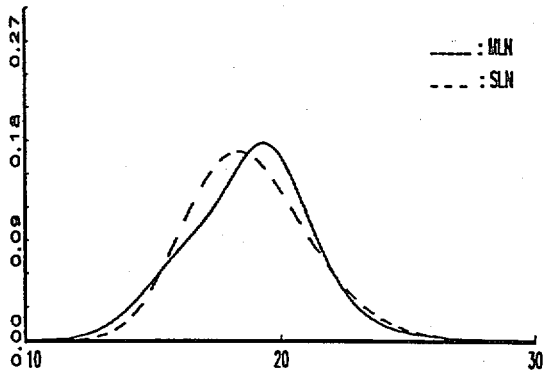
MAR Contract on 30-Jun Futures = 19.80



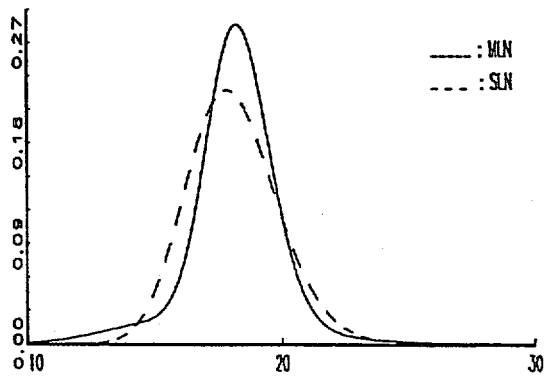
NOV Contract on 14-Jul Futures = 18.13



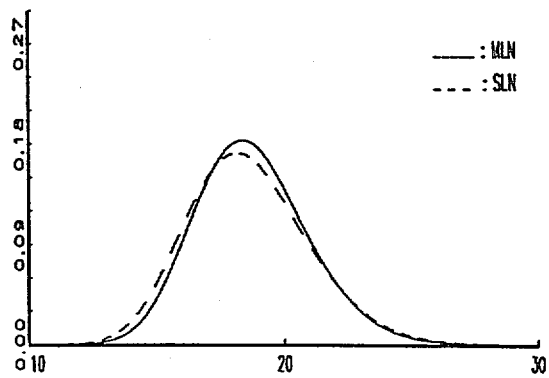
MAR Contract on 14-Jul Futures = 18.79



NOV Contract on 09-Aug Futures = 18.17



MAR Contract on 09-Aug Futures = 18.73



IV. Conclusion

This paper briefly describes the MT method for extracting the market's probability assessment of future events using American option prices, in particular the PDF for the option's underlying asset. The method had many applications as there is a broad spectrum of indices, commodities, and financial instruments on which option are traded, including interest rates and exchange rates. The major limitation of the method is that it is only applicable to those markets where the underlying asset or index martingales. In the application to the oil market, examination of particular days confirmed the large shirt in market expectations that occurred when significant news reached the oil market.

References

Barone-Adesi, Giovanni and Robert Whaley, "Efficient Analytic Approximation of American Option Values", Journal of Finance 42(2):301-320 (1987).

Bates, David S., "The Crash of '87: Was It Expected? The Evidence From Option Markets", Journal of Finance 46(3):1009-1044 (July 1990).

Black, Fisher and Myron Scholes, "The Pricing of Options and Corporate Liabilities", Journal of Political Economy 81:637-659) (1973).

Deaves, Richard and Itzahak Krinsky, "Risk Premiums and Efficiency in the Market for Crude Oil Futures", The Energy Journal 13(2):93-118 (1992).

Dominguez, Kathryn, "The Volatility and Efficiency of Crude Oil Futures Contracts," in Oil and Money: Coping With Price Risk Through Financial Markets, Harvard International Energy Studies 5:48-87 (1989).

Kumar, Manmohan, "The Forecasting Accuracy of Crude Oil Futures Prices", IMF Staff Papers 39(2) (June 1992).

Melick, William and Charles Thomas, "Recovering an Asset's Implied PDF From Option Prices: An Application to Crude Oil During the Gulf Crisis", mimeo, Federal Reserve Board, (1993).

Overdahl, James A. and H. Lee Matthews, " The Use of NYMEX Options to Forecast Crude Oil Prices", The Energy Journal 9(4):135-147 (1988).

Does the Commodity Research Bureau Futures Index Predict Inflation?

David Torgerson, Economic Research Service

Abstract

Is the Commodity Research Bureau futures index (CRB) related to inflation? Employing a procedure analogous to that done by Brown and Yucel (1993) I used pairwise Engle-Granger cointegration to search for relationships between the CRB and the Producer Price Index (PPI), the Producer Price Index for food (PPIfood), Consumer Price Index (CPI), and Consumer Price Index for food (CPIfood), respectively. First, all these variables, as is typical for macroeconomic variables, are nonstationery under the usual transformations (levels, logs, logged differences). Second, under the maintained hypothesis of cointegration between the CRB and the measures of inflation except for the PPIfood. The results put in a regression framework suggest that a 1 percent increase in the CRB is associated with an immediate .1 percent rise in the PPIfood and a one-year later increase of .05 percent. I note that this result is consistent with Babula (1993). While this link appeared to be reliable, the size would not suggest that the CRB alone would do a good job as a potential forecaster of PPI food inflation as a futures price is considered as a good predictor of the cash price of a commodity.

PROBLEM The Commodity Research Board and Future Food or General Inflation?
NO

The Commodity Research Board futures index (CRB) has been suggested as a useful indicator of future inflation, either general or food. Some analysts suggested that the recent rise in the CRB, which took place after the extent of the midwest flood became known, would imply sharply higher inflation either general or food. What is the relationship of these variables and the current and lagged values of the CRB?.
Common inflation indices are the producer price index (PPI), the consumer price index (CPI), and their respective subindices for food PPIfood and CPIfood, respectively.

I tested all four variables for stationarity using single-variable unit root tests--they all failed. They all failed to be cointegrated. jointly cointegrated in levels, in first differences and logs. All but the PPIfood failed to be cointegrated in logged differences (percent change in the levels) and the logged differences of PPI-food and the CRB were cointegrated at the 10 per cent significance level and at a lag of 12 months the error terms of both variables are significantly correlated. A regression of the percent change PPI-food on the percent change CRB contemporaneously and lagged 12 periods indicated that a 1 percent permanent change in the CRB is associated with an immediate .15 percent change in the PPI-food and .05 percent change 12 months ahead. The time trend in this relationship was essentially zero so the regression was justifiable.

The bottom line is that the CRB does not show great promise as an indicator of food or general inflation. Since only the PPIfood is cointegrated at the 10 percent confidence level and the direct and lagged effects add up "explaining" only 20 percent of the change we are not talking about a fine point such as whether a futures price provides an unbiased estimator of a future cash price and an implied elasticity of .9 is estimated.
If analytically one wants to forecast inflation food or general one needs a richer model than is provided by use of CRB and its lags.

Data

The CRB is a daily index of selected futures prices of twenty commodities has roughly 20 observations per month. The various measures of inflation one could consider such as CPI, PPI, PPI-food and CPI-food which are monthly. Of course, one uses a monthly average for the CRB--which may lose some of the information coming from the data. The data are monthly from January 1974 and May 1993.

Methodology and Results

The unit root testing of the individual variables is routine and can be found in the manuals for most econometric packages. For the time-series aficionado the two-variate case of Engle-Granger (1987) approach has given way to the maximum likelihood several variable methods of Johansen and Juselius (1990). We will sketch how we used the extended Dickey-Fuller test to derive the results outlined above.

We note that in testing the relationship between CRB and the inflation indices we are bending over backwards to make it come out. The only "successful" error correction equation (ECC) equation is based on equilibrium being:

(1) $PCPPIfood_t - a*PCCRB_t - b*TREND_t = e_t$, where

$PCPPIfood_t$ is the logged difference of the PPIfood,

$PCCRB_t$ is the logged difference of CRB,

and $TREND_t$ is the time trend,

and e_t is white noise, Normally distributed with mean zero.

All other formulations were statistically inconsistent with the analogous equilibrium condition. Even the above relationship with a constant term is statistically inconsistent with cointegration between the percent changes in CRB and PPIfood.

In (1) if we set et equal to zero, then the long-run relationship between variables is evident. In doing the cointegration test we allowed lags of up to 24 months. The estimate values of a and b are tested using the residuals or errors from equilibrium. In allowing a lag of up to 24 months one allows for error terms potentially influencing future error terms up to two years later. The calculated t-statistic associated with the estimated b is calculated using the Dickey-Fuller tau statistic since the distribution differs from the ordinary t. See Hall, et al (1990) for general discussion of this technique.
The estimated b from this procedure was .22 with significant autocorrelation of the error term at 12 months. As it turned out, the estimated a was not appreciably different from zero so the time trend could be ignored. At this stage one should in principle go to a more sophisticated from of cointegration.

However, all is fair in love, war, and forecasting so given the functional form I re-estimated the relationship in a regression based on:

(2) $PCPPIfood_t = c*PCCRB_t + d*PCCRB_{t-12} + e_t$,

where variables are defined as above except that t-12 refers to a 12 month lag on the percent change CRB. It turns out that the Ordinary Least Squares (OLS) estimated c was .148 and the estimated d was .046--a total slightly less than the above estimated .22.

Conclusion and Interpretation

As I would agree the above OLS results are not completely rigorous but we note that are broadly consistent with results generated using a Vector Autoregressive (VAR) on similar data. I point out that everything done greatly biases the results toward finding a large significant relationship of the CRB to PPIfood and still this is less than an overwhelming response. In the future we will attempt to test the above relationships using the maximum likelihood methods of Johansen and Jusilieus (1990).

References

1. Babula, R. "Dynamic Economic Relationships Among Grain-Based U.S. Food Prices," Journal of International Food & Agribusiness Marketing. Vol 5, No. 1, pp. 35-52. 1993.

2. Engle, R. and C. Granger. "Co-integration and Error Correction: Representation, Estimation and Testing," Econometrica. Vol.55, pp. 251-76. 1987.

3. Hall, R. E., R.E. Hall and D. M. Lilien. Micro-TSP User's Manual; Version 7.0. Quantitative Micro Software, Irvine, California, 1991.

4. Johansen, S., and K. Juselius. "Maximum Likelihood Estimation and Inference on Cointegration-with Application to the Demand for Money," Oxford Bulletin of Economics and Statistics.

METAISSUES

MODERATOR	Howard Fullerton, Bureau of Labor Statistics
PRESENTATION 1	George Wesley, Peg Young, Office of Inspector General, Department of Veterans Affairs
PRESENTATION 2	Karen Hamrick, Economic Research Service and Fred Mills, Federal Management Partners, Inc.

Metaissues

Forecasting As A Tool For Oversight

George Wesley and Peg Young, Office of Inspector General, Department of Veterans Affairs

As the federal government reinvents itself into the form of anticipatory government, the Office of Inspector General (OIG) moves, in concert, towards anticipatory review. The shift towards proactive review and inspection has opened, for the OIG, numerous avenues for forecasting. The difficulty lies in the fact that there exists no "standard" set of techniques, forecasting or otherwise, for inspections. Rather, each particular instance of inspection requires the development of a technique to solve that particular problem. The purpose of this presentation is to discuss the forecasting applications that have been used, or have been proposed for use, for the purposes of oversight within the Department of Veterans Affairs.

Introduction

The Office of Inspector General (OIG) was created " (1) to conduct and supervise audits and investigations relating to programs and operations ... ; (2) to provide leadership and coordination and recommend policies for activities designed (A) to promote economy, efficiency, and effectiveness in the administration of, and (B) to prevent and detect fraud and abuse in, such programs and operations; and (3) to provide a means for keeping the head of the establishment and the Congress fully and currently informed about problems and deficiencies relating to the administration of such programs and operations and the necessity for and progress of corrective actions." [1] As the direction of the OIG has shifted from reactive to proactive review, the definition of task for the Office of Inspector General has been expanded to include inspection, which is defined as "a process, other than an audit or an investigation, that is aimed at evaluating, reviewing, studying, and/or analyzing the programs and activities of a Department or Agency for the purposes of providing information to managers for decision making, for making recommendations for improvements to programs, policies or procedures, and for administrative action. The objectives of inspections include providing a source of factual and analytical information, monitoring compliance, measuring performance, assessing the efficiency and effectiveness of operations, and/or conducting inquiries into allegations of fraud, waste, abuse and mismanagement." [2]

Within the Department of Veterans Affairs' OIG, three distinct trends and influences (the first largely internal, and other two more general throughout the federal government) have led to the appropriateness of the employment of forecasting methods in the oversight role of the VA's OIG. The first trend is unique to the Department of Veterans Affairs (VA). In 1988, Congress, through the passage of "The Veterans' Benefits and Services Act of 1988," called for the VA's Office of Inspector General to establish oversight of the VA's clinical quality assurance programs. The rationale was that, in order to fulfill the OIG's statutory responsibility to oversee, monitor, and evaluate the Agency's programs, the OIG would require individuals with clinical expertise and analytic capability beyond the realm of classical audits and criminal investigation. As such, the Quality Assurance Review Division, now termed the Office of Healthcare Inspections, was established in 1989 in the VA's OIG. The second trend was the need for flexible and anticipatory analysis, i.e., forecasting techniques, to aid the OIG in keeping the VA's Secretary and the Congress currently informed as to the economy, efficiency, and effectiveness of VA's programs, particularly of its health care programs. The third trend has been the tremendous interest, in general, in health care reform and health care policy throughout the country.

In contrast to the field of audit, the inspection methodologies vary from case to case. Since there exists no standard set of inspection techniques, necessity requires the development of procedures on a case-by-case basis. For those instances that require projections into the future, or estimates of what should have occurred based upon previous patterns of behavior, techniques based upon forecasting procedures are deemed appropriate. This

[1] Inspector General Act of 1978

[2] QUALITY STANDARDS FOR INSPECTIONS, President's Council on Integrity and Efficiency, March 1993

presentation illustrates forecasting techniques that either have been used, or are planned for use, within the Office of Healthcare Inspections in the VA's OIG.

Forecasting Techniques

The evaluation of the complex and broad issues involved in health care inspections requires a wide variety of analytical tools. Numerous forecasting techniques have potential in inspections; this report illustrates how these forecasting techniques can aid in the function of oversight. In order to comprehend the techniques to be discussed, a diagram has been provided that illustrates, overall, the types of forecasting techniques generally employed (see Figure 1). Note that this chart is not meant to be comprehensive; rather, it is a tool for describing the general similarities and differences in the techniques.

Forecasting can be discussed in terms as either qualitative or quantitative techniques. The qualitative methods are designed to deal with soft, or non-numeric, data; the quantitative require hard numerical data gathered over several time periods. Obviously, the better the data, the more accurate the technique; but in those instances when no data exist, the qualitative techniques play a valuable role. Some techniques are quite common - expert opinion is used daily by upper management. Brainstorming allows several individuals to create ideas regarding the future; Delphi and panel consensus permit structured solicitation of group opinions. Scenario development is used to develop alternate plans, based upon potential occurrences in the future. Contextual mapping provides a pattern for technology transfer through its analysis of characteristics, and monitoring is a procedure for gathering information from journals, newspapers and other press material regarding changes in status quo.

Quantitative techniques can be divided into two categories - time series and modeling. Time series techniques utilize patterns of the data set under study to predict the future of that time series; modeling relates the behavior of the data set under study to other, independent, variables. The two fields overlap in multivariate intervention analysis and transfer functions.

For time series techniques, the driving force is the assumption that past behavior predicts future behavior; patterns from the past data in the time series are employed to estimate future patterns in the data. Under the additional assumption that the more recent data points are the more "informative" data, exponential smoothing takes advantage of the related error terms over time to make future predictions. Moving averages also work the emphasis of more current data by averaging recent data points to make forecasts. Both techniques are particularly accurate for short-term forecasting. Trend projections, be they linear, growth, logarithmic, etc., perform long-term projections of the long term trends. If seasonality variation occurs in the data, there exist time series techniques designed to handle such data. And, for the more complex designs, the ARIMA model is designed to handle almost any type of forecast characteristic (constant, trend, seasonal or cyclical, with moving average or autoregressive components).

Modeling provides techniques that relate the variable of interest, termed the dependent variable, with other, independent, variables in order to explain that behavior of the dependent variable. Multivariate regression, in its various forms, is the primary technique employed to combine independent with dependent variables; econometric modeling is a variation of multiple regression. Input/output analysis allows the creation of a matrix that analyzes how the various inputs and outputs in products and services relate to one another; this technique is primarily employed on a national level. Cross-impact also performs a type of sensitivity analysis of multiple variables to determine interrelated behaviors, whereas precursor analysis uses the behavior of a comparable variable to forecast the activity of the variable of interest. Intervention and transfer analysis falls into an overlap area of both time series and modeling, by using time series models to relate the dependent to independent variables.

Not included in the diagram, but worthy of note in the arena of federal forecasting, are the methods labeled as normative techniques, which first establish objectives to occur in a future period of time and then determine the path(s) necessary to achieve those objectives. The normative techniques make use of all the previously mentioned methods to some degree. The emphasis, however, is not where the future will take us, but rather how do we get to where we want to be. Variations of "backcasting" are employed to determine the viability of the goals and the potential paths to those goals.

Health Care Applications in Forecasting

The techniques mentioned in the previous section each have their potential in health care analysis from an oversight, as well as Continuous Quality Improvement (CQI), point of view (Kritchevsky & Simmons, 1991). This

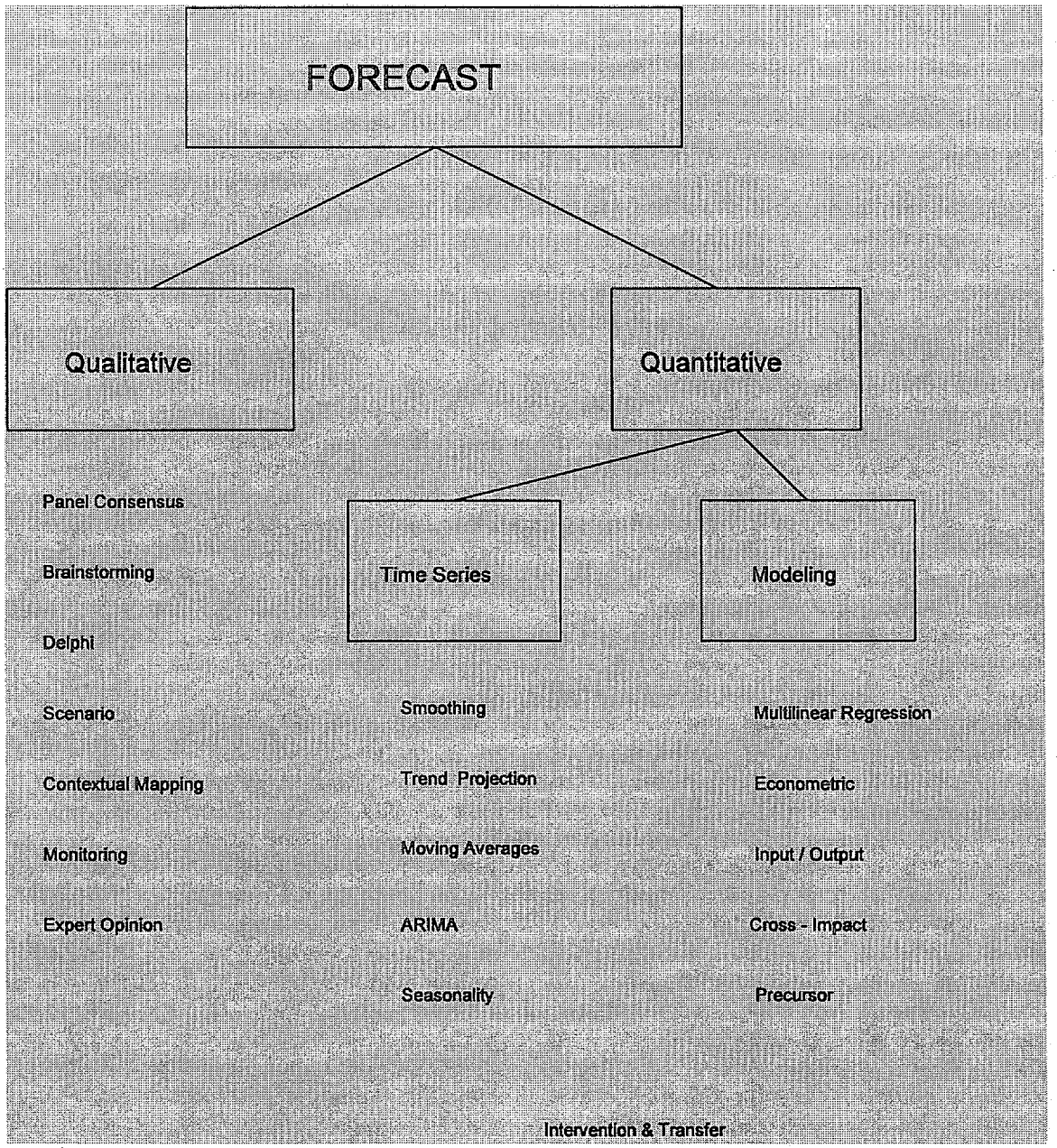


Figure 1. Diagram of forecasting techniques

section of the paper describes, in general, how some of the techniques have been, or are planned to be, used in OHI for heath care review.

With respect to the employment of qualitative techniques, scenario analysis plays a key role in determining how the Clinton national health care plan will impact the Veterans Affairs medical centers (VAMCs). Scenario analysis allows us to ask the questions, "If the plan provides all in the US with health care insurance, what is the impact on who will use the VAMCs?" Alternate futures can be hypothesized to permit speculation in a structured format. Another qualitative technique, expert opinion, is also used to elicit opinions where little to no data exist; another term for this is expert judgment. With expert opinion, question may be asked such as, " Given your knowledge of the workings of Congress, what is your opinion as to the likelihood of permitting non-veterans into the VAMCs?"

In the quantitative techniques, one of the simplest techniques is also one of the best techniques for providing forecasts - exponential smoothing. As will be illustrated later, the techniques show great promise in the monitoring of mortality rates, as well as other quality assurance measures. Linear and non-linear trend regressions also have great potential in oversight; such capabilities are already employed in statistical process control (SPC) within health care facilities (VanderVeen, 1992). Data can be reviewed to determine simply if a trend does exist, the most basic form of trend analysis. Growth curves, a specialized form of nonlinear curves, can be used to estimate special populations, such as the growth of the AIDS / HIV infection population, in order to determine if the VAMCs are prepared to handle such an increase in that population. A spin-off from the quantitative time series techniques is the genre of quality control or quality assurance (QA) methods. Some of these techniques are designed to monitor behaviors over time - such as mortality rates (Kritchevsky & Simmons, 1991) or changes in DRG's, or diagnosis-related groups (Cohen et al., 1987).

Also in quantitative forecasting, various modeling possibilities exist to aid in reviewing health care proposals. Modeling is presently being employed to forecast estimates of physician need within the VAMCs, based upon the changes in the demographics of the veteran population. Modeling has also been used to compare costs of the VAMCs to that of their affiliated university hospitals, again permitting a forecast of future requirements in funding. Modeling has also been employed by the Health Care Financing Administration (HCFA) to provide projected mortality rates among hospitals treating Medicare beneficiaries (Krakauer & Bailey, 1991). Intervention analysis has potential in numerous quality assurance analyses, in the attempt to determine which actions or external impacts have altered the pattern of QA measures.

Normative techniques are also employed in health care oversight. A study of health progress in the United States (McGinnis et al., 1992) details several normative goals in health care, along with the measures to determine the degree of achievement towards these goals. The OHI could employ comparable methods to monitor the actions of the VA towards the proposed restructuring of the health care towards the national program being established by the Clinton administration.

Mortality Rate Analysis

As mentioned in previous section, the technique to be studied in detail is the use of exponential smoothing, in this example on mortality rate analysis. Deaths per discharge are reviewed within a medical center to determine if unexpected deviations have occurred in the time series. The difficulty lies in the term "unexpected," when the data themselves are highly variable. When the data exhibit such high degrees of variability from time period to time period, then the task of identifying "unusual" deviations becomes onerous.

But such a technique does exist, as used in production and operations management. Tracking signals are used on smoothed data to signal deviations greater than would be expected from random fluctuations alone. The single exponential smoothing model

$$s_t = \alpha \cdot x_t + (1-\alpha) \cdot s_{t-1}$$, where

x_t is the time series under review (mortality rates, in this case) and s_t is the smoothed value at time t, and α is the smoothing constant with a value somewhere between 0 and 1 (Brown, 1963; Gardner, 1985; Gardner & Dannenbring, 1980). This model is appropriate for those instances in which the analyst does not expect to observe any long-term trends upward or downward in the data over time. The smoothed data can then be used to calculate the tracking signals over time to illustrate significant shifts in the data (Mandell & Bretschneider, 1984).

The smoothing model provides "smoothed" mortality rates against which the observed mortality rates can be compared to indicate significant movement not attributable to random error (Batty, 1969; Gardner, 1985a; Gardner, 1985b; Gardner & Dannenbring, 1980; McKenzie, 1978). Extreme deviations based upon variance from the smoothed data could be attributed to shifts in the mortality rate, which can take the form of one or more distinct forms of interventions within a time series. First, suppose an incident of severe food poisoning occurs at a facility; this change in mortality rates would be in the form of a spike (or pulse) in the time series. The spike or pulse intervention, as shown in Figure 2, indicates a short term intervention that causes a shift in the time series over a period of time, and then returns to the pervious state. Another change in death rates for a medical ward could be a policy change in sending the more severely ill patients to a different ward; this change in death rates would take the form of a step function, in which the level of the data shifts. The step intervention, as shown in Figure 3, also occurs in a short period of time, but the level of the time series remains at the new level, rather than return to the previous state. The third form of an intervention could occur due to a gradual degradation in quality of care; such a change would alter the slope of the underlying trend. The slope intervention as shown in Figure 4, occurs over a longer period of time as a long-term gradual change in the level of the time series; i.e., the slope changes. All of these interventions would be characterized by some deviation of the observed mortality rate from those forecast by the smoothing model. It is hypothesized in this study that significant deviations between the model and the observed data can be interpreted as indications of interventions in the underlying equilibrium (Trigg, 1964; Trigg & Leach, 1967; Mandell & Bretschneider, 1984).

To create a mechanism of identifying significant changes in the time series, the variance, σ^2_e , of the forecast error, $e_t = (x_t - \hat{x}_t)$, needs to be taken. For simple exponential smoothing, the standard deviation for a single period forecast error is

$$\hat{\sigma}_e(t) = ((2-\alpha)/2)^{1/2} \cdot \sigma_e(t)$$
$$= ((2-\alpha)/2)^{1/2} \cdot (1.25 \cdot \hat{MAD}_t), where$$
$$\hat{MAD}_t = \alpha|e_t| + (1-\alpha) \cdot \hat{MAD}_{t-1}$$

is the estimate of the Mean Absolute Deviation (Montgomery, Johnson & Gardiner, 1990). Dividing each observed forecast error by the above standard deviation, S.E. $\equiv e_t/\hat{\sigma}_\varepsilon(t)$, provides an indicator of the relative magnitude of the observed error. This term shall be referred to as the standardized error term (S.E.). If used as a type of confidence band, extreme deviations as noted by the indicator could be associated with any of the interventions mentioned, but primarily with a pulse or step intervention.

A tracking signal can also be employed to detect significant changes from the time series pattern. Several types of tracking signals are presently employed to detect underlying shifts in the time series data, such as CUSUM and smoothed-error forecast monitoring schemes (McClain, 1988). The smoothed tracking signal technique will be employed here to allow consistency in smoothing calculations with the previously designed signal.

Using the previous calculation of the estimate of the MAD, the smoothed error tracking signal can be calculated as

$$T.S. = |Q_t/\hat{MAD}_t| \; where$$
$$Q_t = \alpha \cdot e_t + (1-\alpha) \cdot Q_{t-1}.$$

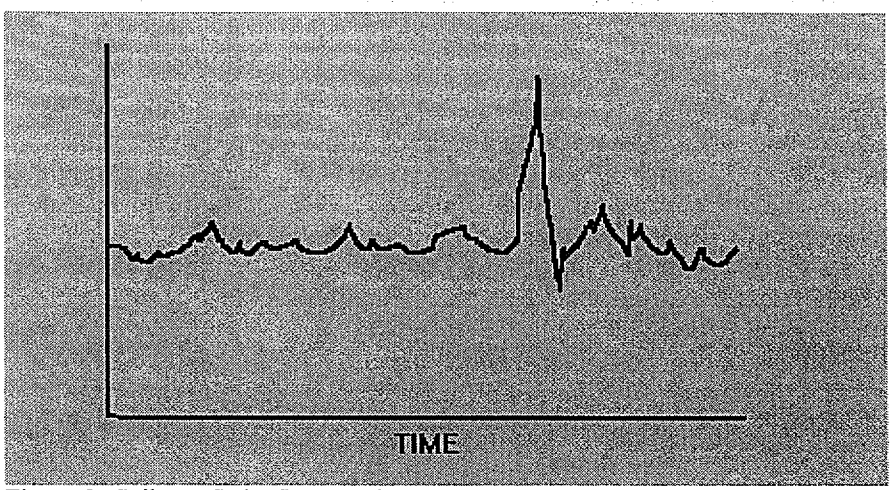


Figure 2. Spike or Pulse Intervention

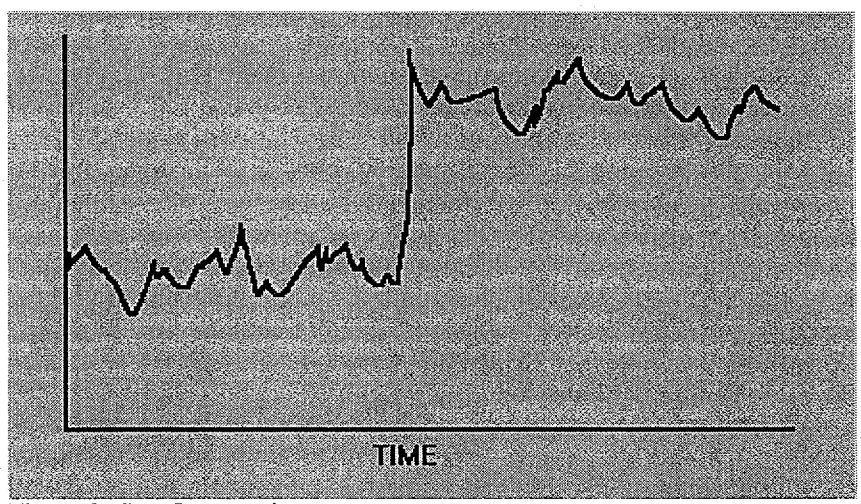


Figure 3. Step Intervention

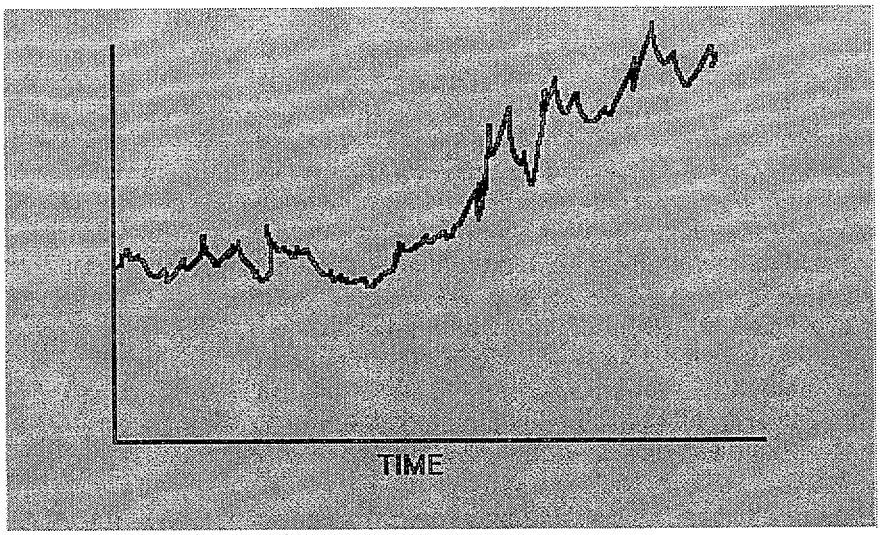


Figure 4. Slope Intervention

If we set

$$\sigma_Q^2 = (\alpha/(1+\beta)^2) \cdot \sigma_e^2 \text{ ,then}$$

$$|Q_t/M\hat{A}D_t|)(\frac{1.25 \cdot \sqrt{\alpha}}{1+B}) \cdot \kappa_1 = \kappa_3 \text{ , where}$$

κ_1 is a standardized score indicating the degree of confidence desired (usually a value between 2 and 5). When the tracking signal exceeds κ_3, which is usually between 0.2 to 0.5, this signal would be interpreted as indicating a movement of the observed values away from the smoothed values more than would be expected by random movement. Whereas the standardized error (S.E.) responds to abrupt changes in the time series, the smoothed error tracking signal (T.S.) has the capability of identifying longer-term changes in the data, such as changes in the slope.

The two tracking signals were applied to mortality rates at a medical center. The question being asked is: At what point in time did the intervention, if any, take place? Once the point in time is pinpointed, then the intervention that caused the change in the time series pattern can be possibly ascertained. The empirical analysis to establish the point of intervention consists of (1) the selection of the smoothing parameter for the mortality rate used to forecast the next month's rate, (2) the selection of the parameters for calculating the MAD and the tracking signal, and (3) the conduction of the actual test. The determination of the significant levels of the observed error and the tracking signal was set to that previously determined in a working paper which developed this procedure for the detection of interventions in stock price time series (DeSaix and Young, 1989).

In order to estimate an alpha value (α) for the smoothing process, an ARIMA(0,1,1) was run on the initial segment of the time series; the latter set of data points were those that were being scrutinized for potential changes in the data. The fitting of an ARIMA(0,1,1), which is the Box-Jenkins version of simple exponential smoothing, resulted in an alpha value that is the best fitting estimate of the smoothing parameter. Using the estimated alpha level, the two time signals were calculated - a standardized error (S.E.) and a smoothed tracking signal (T.S.). The standardized error indicates significant activity by some intervention when it becomes larger that $|2.0|$; the smoothed tracking signal indicates significance when it becomes larger than .35 (DeSaix & Young, 1989).

The results of the two signals are portrayed in Figures 5 and 6. The mapping of the standardized error (S.E.) in Figure 5 indicates a spike or step intervention in period 44. The horizontal arrow drawn at the value on the S.E. axes at +2.0 indicates the level up to which variability could be attributed to random error; the time series exceeds that level at period 44, which signals a potential step or pulse intervention. No S.E. value in the graph falls below the -2.0 S.E. value.

The mapping of the tracking signal (T.S.) in Figure 6 also has drawn a horizontal arrow at the value on the T.S. axis of .35, which is the maximum level of the tracking signal that could be attributed to random error. The figure indicates a movement above that maximum level beginning at period 36, as well as activity changes in period 44, which could be interpreted as an intervention that caused a change in slope.

The two tracking signals indicate two time periods of intervention. Whereas the standardized error indicates a pulse or step activity that initiates in period 44, the smoothed tracking signal highlights a slope change initiating at period 36 as well as period 44. What these two statistical measures signal are two potential time periods of change - a change in the slope of the death rate starting around period 36 and an additional intervention occurring near period 44. By being able to pinpoint the periods in time for the occurrence of interventions, an analysis into activity in the ward at those particular points in time could then determine the reasons for the interventions.

Results and Conclusions

Using signals such as the ones developed in this paper allows the time series data to 'inform' the analyst as to the time periods of potential interventions. This type of statistical quality control permits faster response to problems by driving the analyst in finding the interventions that caused the changes in the time series behavior, prior even to the knowledge of the intervention taking place. Time series analysis, as shown above, asks the question "Has any significant activity occurred?"

In general, forecasting techniques offer a new assortment of analytical tools to perform inspections within the Office of Inspector General. As each unique health care issue evolves, a forecasting technique could be devised to perform the analysis of that issue. A repertoire of analytical tools can now be developed to perform inspections through forecasting.

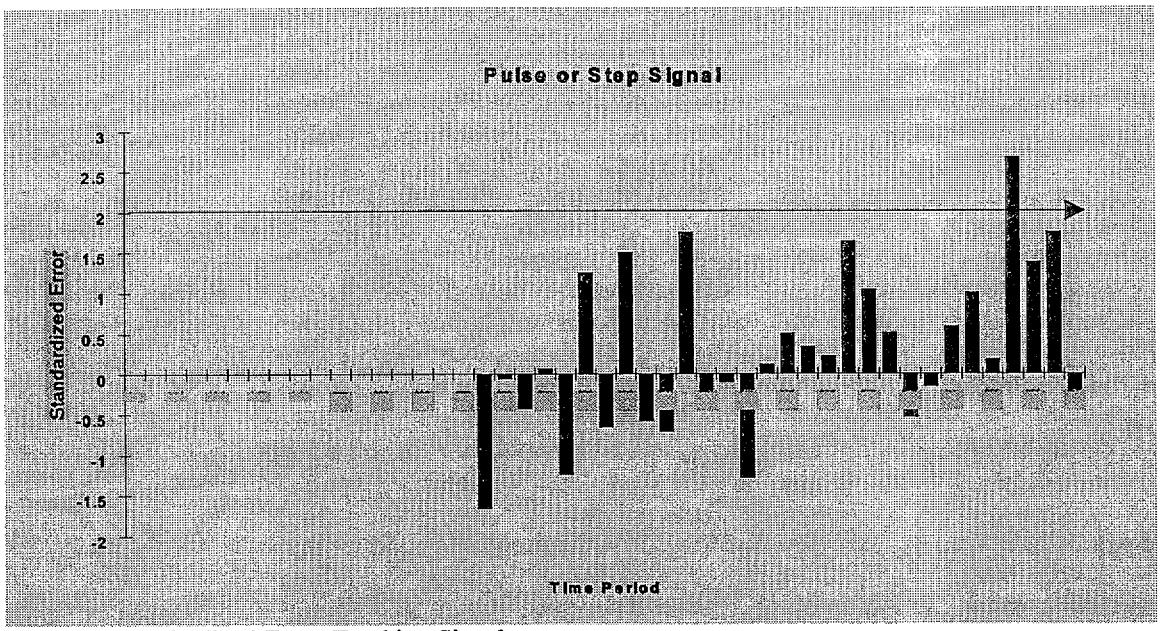


Figure 5. Standardized Error Tracking Signal

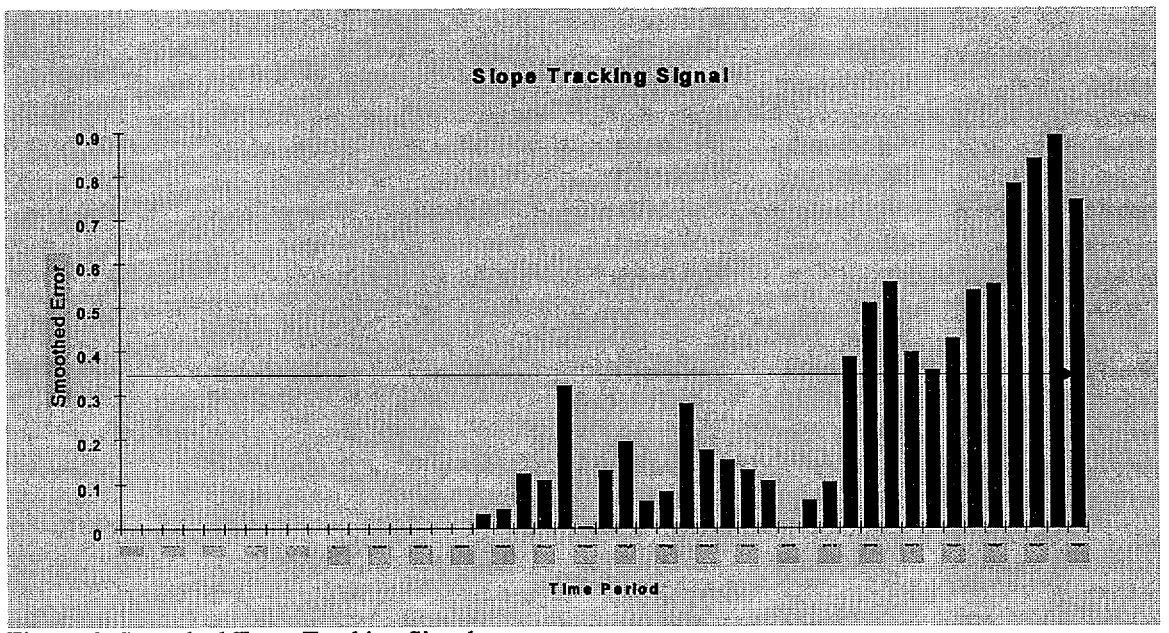


Figure 6. Smoothed Error Tracking Signal

References

Batty M. *Monitoring an Exponential Smoothing Forecasting System.* Operational Research Quarterly 1969;26:319-325.

Bretschneider S. *Estimating Forecast Variance with Exponential Smoothing.* International Journal of Forecasting 1986;2:349-355.

Brown RG. Smoothing, Forecasting and Prediction of Discrete Time Series. Englewood Cliffs, NJ: Prentice-Hall, Inc.; 1963.

Cohen BB, Pokras R, Meads MS, Krushat WM. *How Will Diagnosis- Related Groups Affect Epidemiological Research?* American Journal of Epidemiology 1987;126(1):1-9.

DeSaix JC, Young P. *Interventions in Stock Price Time Series.* Working paper. 1989 June.

Gardner ES, Jr. *CUSUM vs. Smoothed-Error Forecast Monitoring Schemes: Some Simulation Results.* Journal of the Operational Research Society 1985;36:43-47.

Gardner ES, Jr. *Exponential Smoothing: The State of the Art.* Journal of Forecasting 1985;4:1-28.

Gardner ES, Jr., Dannenbring DG. *Forecasting with Exponential Smoothing: Some Guidelines for Model Selection.* Decision Sciences 1980;11:370-383.

Krakauer H, Bailey RC. *Epidemiologic Oversight of the Medical Care Provided to Medicare Beneficiaries.* Statistics in Medicine 1991;10(4):521-540.

Kritchevsky SB, Simmons BP. *Continuous Quality Improvement: Concepts and Applications for Physician Care.* JAMA 1991; 266(13):1817-1823.

Mandell MB, Bretschneider SJ. *Using Exponential Smoothing to Specify Intervention Models for Interrupted Time Series.* Evaluation Review 1984;8:663-691.

McClain JO. *Dominant Tracking Signals.* International Journal of Forecasting 1988;4:563-572.

McGinnis JM, Richmond JB, Brandt EN, Windom RE, Mason JO. *Health Progress in the United States.* JAMA 1992;268(18):2545-2552.

McKenzie E. *The Monitoring of Exponentially Weighted Forecasts.* Journal of the Operational Research Society 1978; 29:449-458.

Montgomery DC, Johnson LA, Gardiner JS. Forecasting and Time Series Analysis, 2nd. Edition. New York: McGraw-Hill Book Company; 1990.

Trigg DW. *Monitoring a Forecasting System.* Operational Research Quarterly 1964;15:271-274.

Trigg DW, Leach AG. *Exponential Smoothing with an Adaptive Response Rate.* Operational Research Quarterly 1967;18:53-59.

VanderVeen. Loretta M. *Statistical Process Control: A Practical Application for Hospitals.* Journal for Healthcare Quality 1992;14(2):20-29.

Federal Forecasting: Occupation or Analytical Tool?

Karen S. Hamrick, USDA Economic Research Service and Fred Mills, Federal Management Partners, Inc.

Abstract

Should forecasting be recognized as an occupation in the Federal Government? Forecasting has developed into its own area of research in the last 10 years, yet the Federal personnel system does not recognize it as either an occupation or as a specialized discipline with unique requirements and qualification standards. Currently economists, statisticians, and other area specialists are hired by agencies as forecasters, with each agency applying its own definition of forecasting in establishing qualification, selection, and training criteria. With the increased emphasis on foresight in government is the potential for expanded roles and responsibilities for forecasters and will require a more rigorous, systematic approach. This paper outlines a method for formalizing forecasting in the Federal government.

Forecasting is more than a collection of techniques, it is a multi-step process. A broad spectrum of skills are needed for the process of forecasting. Setting standards for Federal forecasters would help insure that forecasters are well-versed in the many aspects of forecasting and could lead to better forecasts and improved credibility in Federal forecasts and Federal forecasters.

The Federal government provides a variety of forecasts to the public including those on agricultural markets, labor markets, the national economy, energy, and weather. These forecasts are produced by Federal employees from a variety of disciplines, including economics, statistics, demography, actuarial science, and meteorology. These forecasts are used by individuals, state and local governments, educational institutions, and private companies. For many of these forecasts, it would not in any one individual's interest or ability to forecast. Federally-provided forecasts facilitate planning for both individuals and organizations.

Federally-produced forecasts are also used for policy-making and planning purposes, such as planning Federal programs and budgets. In an anticipatory government, the need is even greater for more accurate forecasts. Indeed, Wright, et al. stated "Planning for the future at the strategic or any other level is bound up with forecasting."[1] Good forecasts are needed for successful policymaking or program implementation. Klein asserted, "Careful forecasts, as accurate as possible, are central to the successful implementation of policy. There are fundamental reasons why policy makers cannot 'play by ear,' adjusting policy quickly to each unexpected deviation in economic outcomes."[2] In addition, it is crucial that forecasting be integrated into the policymaking or planning process for success.[3]

Forecasting as an Area of Study

Forecasting has emerged as an area of study over the last 20 years with a unique body of knowledge. There are now professional organizations of forecasters, whose members come from a variety of fields. In addition to the Federal Forecasters Conference, now in its sixth year, there is the International Institute of Forecasters (13 years), and the International Association of Business Forecasting (12 years).[4] Four academic journals are devoted

[1]Wright, et. al. (1986), p. 149.

[2]Klein (1984), p. 1.

[3]From Wright, et al. (1986): "Our aim is to show the danger of separating the production of a forecast from the use of the forecast in decision making and to show how these two aspects can be integrated in practice." (p. 139) And to complete the thought cited in the text above: "Planning for the future at the strategic or any other level is bound up with forecasting. The conclusion of this paper indicate that this bond needs to be more tightly woven than is often the case." (p. 149)

[4]Futurist predictions are not included here as past of forecasting. Many futurists belong to the World Future Society, which publishes The Futurist and the Futures Research Quarterly.

to forecasting, the International Journal of Forecasting, the Journal of Forecasting, Technological Forecasting and Social Change, and the Journal of Business Forecasting. In addition, courses specifically on forecasting now exist at universities, usually in the economics department, the decision sciences program, or the business or engineering schools.

Be that as it may, forecasting has not yet reached the status of a discipline. There are still no degrees in forecasting. The business forecasting community has considered a certification program,[5] but that effort is in its infancy, and certification is controversial within the community of all forecasters. Many argue that forecasting should never be a separate field, for a forecaster must have substantive knowledge in order to make meaningful forecasts. Jenkins stated, "In trying to understand the relationships between the variables to be forecast and the policy and environmental variables, it is important that the forecaster understands the relationship between his system and its environment--and embodies this understanding in the conceptual model [of the system being studied]."[6]

Although forecasting is not a discipline, it clearly has become an interdisciplinary area of study. A consensus has emerged on what forecasting is and what forecasters need to know. This consensus defines forecasting as a process, not as a forecast product. See figure 1 for a representation of this process.[7] We have used this flow-chart representation of the process to define the critical elements of forecasting in table 1.[8] A forecaster would not be expected to know all the items listed under each element, but would need to have an awareness of them.

The Importance of Evaluation

Forecasting research was originally focused on existing techniques and the development of new techniques. Young commented:

> In the earlier years of forecasting research, journal articles were inundated with 'competitions' designed to find the ultimate forecasting technique. The classic example is the M-competition, an abbreviation for the Makridakis competition,[9] which provided over 1000 time series to a panel of forecasting experts. Seven experts in each of 24 methods essentially to determine which technique performed the best overall. What these competitions often failed to take into consideration were the characteristics of the data set, as well as the needs of the forecaster. The search was for the 'best' forecast procedure overall--not the best technique for a given set of time series of data and for a given forecasting need.[10]

[5]See Brown, et al. (1992) and Levenbach (1993).

[6]Jenkins (1982), p.6.

[7]This chart is based on the Box-Jenkins Algorithm [Box and Jenkins (1970), p. 19], and the Farnum and Stanton model selection flow chart [Farnum and Stanton (1989), p. 39].

[8]The authors used many references in compiling this table. Included among them are Armstrong (1985), Farnum and Stanton (1989), Granger (1989), Kennedy (1992), the International Journal of Forecasting, the Journal of Forecasting, the Proceedings of the Federal Forecasters Conferences (1990, 1991, and 1992), and discussions with Peg Young and Herman Stekler.

[9]Makridakis, A. et al. "The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition," Journal of Forecasting. 1982, Vol. 1, pp. 111-153.

[10]Young (1991), p. 35.

Figure 1--The process of forecasting

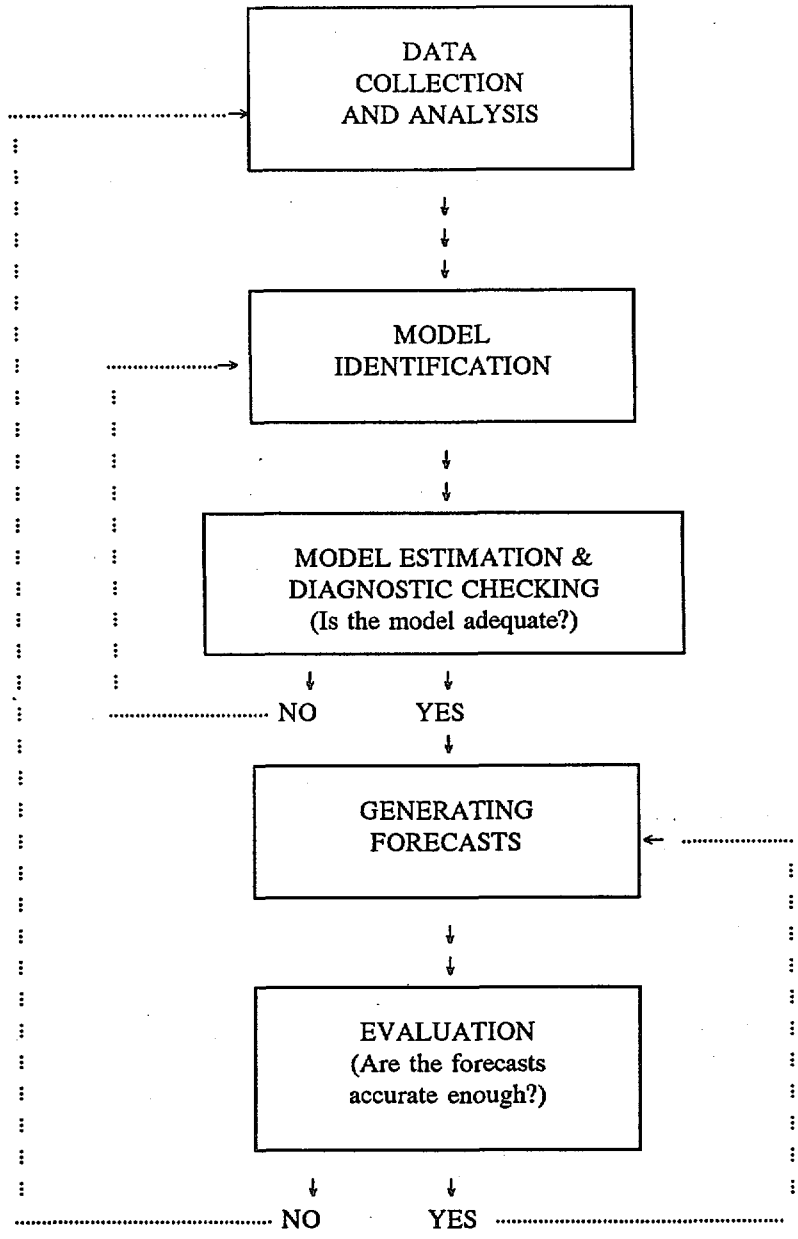


Table 1--The critical elements of forecasting

DATA COLLECTION AND ANALYSIS	MODEL IDENTIFICATION	MODEL ESTIMATION & DIAGNOSTIC TESTING	GENERATING FORECASTS	EVALUATION

- survey methods
  - understanding the psychology of the survey process
- knowledge of data sources
- using data/data analysis, including:
  - trend/no trend
  - seasonal adjustment
  - unit roots, stationarity
	- various techniques, including:
	  - time series analysis/ extrapolation methods
	  - causal/econometric models
	  - leading indicators
	  - Delphi method
	  - combining forecasts
	  - input-output analysis
	  - population forecasting
	  - actuarial methods
	  - segmentation methods
	  - bootstrapping/ expert systems
	  - judgmental approaches
	  - technology forecasting
	  - meteorological models
		- choosing an appropriate technique
		- idea of parsimony
		- awareness of use of judgement in forecasting process:
		  - assumptions
		  - choosing model
		  - choosing data
		- determining what can and cannot be forecast
		- cost/benefit considerations
		- diagnostic testing
			- point forecasts versus interval forecasts
			- scenario/ sensitivity analysis
			- confidence interval around the forecast
			- using forecasts
			- presenting and communicating forecasts
				- variety of forecast accuracy measures
				- evaluating a single forecast versus evaluating a group of forecasts
				- choosing appropriate benchmark for comparison of forecast accuracy
				- accepting/not accepting results
				- using evaluation to choose a forecasting technique and to improve future forecasts

Later, the forecasting community realized that there is no ultimate technique, and the "the data drives the forecast, not the forecasting technique."[11] With that understanding the focus shifted to the importance of evaluating forecasts. In 1989 Bretschneider and Gorr commented,

> Over the last ten years, forecasting has made great strides in developing its own research methods and a unique body of knowledge. During this time forecasting research has become less axiomatic and deductive and more empirical. Much of the emphasis in research has moved away from the development of new techniques towards the evaluation of existing techniques.[12]

Jenkins stated, that in order for forecasting to be effective in an organization, "forecasting should be seen as a learning process, so that when the monitored forecast errors are larger than expectation, remedial action can be taken to improve matters in the future."[13] The General Accounting Office (GAO) engages in the most pointed forecast evaluations of all. Their approach is as follows:

> Seven specific evaluation questions are used [by GAO] in assessing forecast accuracy. The methodology emphasizes comparing the forecasts to the actual subsequent event. The evaluation questions include the following:
> 1. What methodology is used for forecasting the event?
> 2. Who uses the forecasts?
> 3. How can forecast accuracy be measured?
> 4. How accurate are the forecasts?
> 5. Are the errors we identified "reasonable?"
> 6. What are the implications of forecast error on private sector as well as on the government policy, program, and budget decisions?
> 7. How can forecasts be improved?[14]

Clearly the forecasting community has adopted the view that forecasting activity is a process, and a crucial part of that process is evaluation. The recognition of evaluation can be considered a benchmark indicating that forecasting has attained status as a separate and distinct area of study and specialization. Evaluation requires that a field "close the circle," as in figure 1, and become self-contained. With evaluation, the field defines its own standards and criteria which make it separate and distinct.

Recognition in the Federal Government

If we accept that forecasting has emerged as a discrete area of study and specialization over the last 20 years, with its own body of knowledge and techniques, a logical question for Federal forecasters and managers to ask is: To what extent does the Federal personnel system recognize and reflect this new specialization? Specially, do the criteria and standards applied in evaluating and selecting candidates for Federal jobs which involve forecasting, as well as those used to establish the grade and pay levels for these positions accurately describe and assign value to the specialized knowledge and skills required for forecasting?

To help answer this question, a sample set of 22 Federal occupations was identified which reflects those job series whose members were thought most likely to be engaged in forecasting activities. (See table 2.) The sample set includes Economist and Statistician, the two occupations with the largest representation among participants in the Federal Forecasters Conference, as well as a number of other obvious candidates such as:

[11]Young (1991), p. 35.

[12]Bretschneider and Gorr (1989), p. 305.

[13]Jenkins (1982)2p. 17.

[14]Solenberger (1991), p. 25.

Table 2--List of 22 Federal Occupations Most Likely to Involve Forecasting

OCCUPATION	SERIES	REFERENCES TO FORECASTING IN OPM DOCUMENTATION	
		Qualification Standards	Classification Standards (year classification standard was written)
Actuary	1510	professional actuarial society training/ examination requirements and standards listed as qualifying at various grade levels	focus on programmatic applications; reference to actuarial science, but no real description of methods, techniques (1967)
Appraiser	1171	none	mentions multiple regression as technique employed by appraisers (1972)
Budget Analyst	560	none	reference to long-range cost estimation, industrial fund budget forecasting, multi-year (3-5) projection; no specifics (1981)
Community Planner	020	none	none (1973)
Economist	110	none	brief reference to forecasting in context of econometrics, identification of methodology as sub-specialty; no discussion of methods, techniques (1963)
Financial Analyst	1160	none	short-term/long-term financial plans, studies; no specifics (1966)
Financial Manager	505	none	none (1969)
Industrial Engineer	896	none	none (1969)
Industrial Specialist	1150	none	none (1961)
Intelligence Analyst	132	none	none (1960)
Management/Program Analyst	343	none	general reference to quantitative methods; no specifics (1990)

Table 2--List of 22 Federal Occupations Most Likely to Involve Forecasting (con't)

OCCUPATION	SERIES	REFERENCES TO FORECASTING IN OPM DOCUMENTATION	
		Qualification Standards	Classification Standards (year classification standard was written)
Mathematical Statistician	1529	none	none (1961)
Mathematician	1520	none	brief reference to forecasting as sample activity at GS-9 level (1965)
Meteorologist	1340	minimum of six semester hours in weather analysis/forecasting methods required	extensive reference to forecasting as a key job function, area of specialization; no discussion of methods, techniques (1972)
Operations Research Analyst	1515	none	discusses modelling in general terms, no specifics; notes that operations research does not have its own body of principles and laws (1967)
Policy Analyst (functional title)	various (e.g., 101)	forecasting listed as sample activity for one area; optional assessment of qualifications in survey, simulation, and regression methodology	reference to quantitative methods, simulation, modelling; forecasting as sample activity in one area, no specifics (1981)
Psychologist	180	none	description of research specialty refers to statistical/mathematical techniques; no specifics (1968)
Sociologist	184	none	list Demography as one of five specialty areas; regional growth forecasts, population projections noted as sample products, no specifics (1988)
Statistician	1530	none	none (1961)
Trade Specialist	1140	none	none (1963)
Transportation Industry Analyst	2110	none	none (1969)
Quality Assurance Specialist	1910	none	none (1983)

Meteorologist and Operations Research Analyst, many of whose practitioners perform forecasting work as part of their ongoing responsibilities. To determine if forecasting is recognized as a part of these occupations, the personnel qualifications standards (used for hiring) and position classification standards (used for setting the grade of a position) issued by the Office of Personnel Management (OPM) covering each were analyzed and reviewed for references or descriptions of forecasting functions, techniques, and methodology.

In general, the OPM documentation examined does not recognize forecasting either as a specialty area within individual occupations, or even as an analytical tool applied in accomplishing the various responsibilities associated with the sample occupations. With the exception of two occupations (Actuary and Meteorologist), the personal qualifications standards examined do not define even a minimum level of preparation or competence in forecasting as a prerequisite to hiring. The position classification standards--used by personnel specialists and managers to assign jobs to General Schedule (GS) grade levels--include some limited references to forecasting, but in no case do they provide any specific discussion of forecasting techniques, methods, or the overall forecasting process, i.e., those elements and criteria defined previously in this paper (figure 1 and table 1).

A partial explanation for this absence of recognition can be inferred from the issue dates of the classification standards reviewed. Most are of 1960's vintage, with the standards for the Economist and Statistician occupations, for example, issued in 1963 and 1961 respectively, predating the emergence of forecasting as a distinct specialty area. In addition, the personal qualifications standards, while generally much newer (three to five years old), are for the most part generic in nature and cover broad groups of occupations, e.g., all "professional" positions, with minimal occupation-specific guidance and criteria.

The net effect of this absence of recognition is that in many cases, Federal managers are evaluating and selecting candidates for jobs involving forecasting without any guidelines or criteria for ensuring that their new hires possess the requisite knowledge and skills to perform successfully. On the position side, jobs are established and grade levels defined absent any consideration of the forecasting work involved. In cases where forecasting represents a substantial portion or even a majority of a given position's duties, the position may be incorrectly graded. These problems are even more pronounced in cases where managers or selecting officials are not themselves forecasters.

Approaches to Formalization

Given this state of affairs, should efforts be made to create a new Federal occupation for forecasting, with a formally defined and officially sanctioned set of personal qualification and job grading standards applicable government-wide? Table 2 includes several examples of existing Federal occupations, such as Actuary and Operations Research Analyst, which are based on highly specialized, quantitative disciplines and whose practitioners--not unlike Federal forecasters--apply their expertise to a variety of policy, programmatic, and management issues.

Table 3 lists the four general criteria applied by OPM in determining whether or not a new Federal occupation should be established: significance of the population to be covered throughout the government; existence of a commonly recognized body of specialized knowledge; whether or not the proposed occupation is clearly separate and distinct from other, existing Federal occupations; and the extent to which the agencies with the largest numbers of covered employees agree that a new occupation is necessary and appropriate. Population significance reflects not only the potential size of the new occupation but also its distribution across agencies and departments, the organizational and grade levels at which covered positions are found, and the general visibility or importance of the group. Consensus among affected agencies, while an informal consideration, is also nevertheless a critically important element in establishing a new occupation.

Application of these criteria to Federal forecasting in its present state suggests several barriers to separate occupational status for forecasters. Most problematic at this point appears to be the separate and distinct issue--i.e., can forecasting logically be detached from the professional and subject matter knowledge associated with the (current) occupations of its practitioners--economists, statisticians, demographers, etc.? The likely answer, as already noted in this paper, is that it cannot; to paraphrase Jenkins, the forecasters must have an indepth understanding of the context and environment in which his or her skills are to be applied. Unlike the Actuary, who need not be a health policy specialist to project the impact on mortality rates of changes in access to primary medical care, or the Operations Research Analyst, who may require only a basic understanding of military science to

construct an effective battlefield simulation, the forecaster must be thoroughly grounded in the particular discipline associated with the forecasts he or she is expected to generate.

Table 3--Establishing a New Federal Occupation

OPM Criteria - New Occupations:

- Population

  How many positions/people will be covered by the new occupation government-wide, and does this constitute a "significant" population?

- Common Body of Knowledge

  Is there a commonly recognized body of (specialized) knowledge and/or skills required for all practitioners of this occupation across this government?

- Separate and Distinct

  Is this body of knowledge, and its various applications and outputs, clearly separate and distinct from other Federal occupations or groups?

- Consensus

  Do the agencies most directly affected, i.e., those with the largest number of covered employees/positions, concur in the establishment of a new occupation?

OPM Documentation - New Occupations:

- Classification Standard

  Occupational information, i.e., functions, terms, environment; factor level criteria for assigning covered positions to appropriate GS grades.

- Qualification Standard

  Description of education, experience, and/or certification and training criteria; entry level test/examination requirements.

And while some practitioners may disagree with this position, the lack of consensus reflected by such disagreement only further weakens the case for establishing forecasting as a separate occupation. Still, forecasting is more than just an analytical tool or statistical technique.

Separate occupational status is not the only approach to addressing the problems resulting from forecasting's current lack of recognition within the Federal personnel system. One possible alternative would be to pursue development of a functional classification guide covering forecasting activities as practiced by members of a variety of Federal occupations and series. Rather than focusing on a single discipline, such guides provide criteria for evaluating specialized functions across multiple occupational series and job titles. Functional guides describe the work performed in essentially generic form, but in sufficient detail to allow for thorough evaluation and appropriate recognition of the activity in determining grade levels and qualification requirements for covered positions.

A good example is the Policy Analyst Grade Evaluation Guide. This document, whose current issuance dates from 1981 (see table 2), provides criteria for evaluating professional and scientific positions which are primarily concerned with the analysis of public policy issues and the provision of expert, objective advice and guidance regarding these issues to policy makers and senior officials. The occupational titles of positions which may be covered by this guide include such diverse series as Social Science Analyst, Physical Scientist, Education Program Specialist, or Environmental Engineer. The common denominator, however, is that the work assigned to positions covered by this guide primarily involves the application of policy analysis skills and techniques, combined with subject matter expertise in a particular discipline or field, to produce a specialized end product or service which adds value to the policy-making process. The grade level evaluation of the analysis function, but retention of the subject matter/discipline title, recognizes the inseparability of these two elements in the practice of this craft within the Federal government.

There appear to be strong parallels in this example to the current state of Federal forecasting. Practiced by members of a variety of occupations and disciplines, forecasting has emerged in recent years as a separate functional area which can now be defined on the basis of a commonly recognized process and knowledge base (see table 1). Many Economist, Statistician, and other positions throughout the Federal government are primarily concerned with forecasting, and the importance of the forecast product can only be expected to increase as efforts to "reinvent" government and move towards a more proactive, "anticipatory" approach proliferate and gather momentum.[15] The role and value of forecasting in the policy-making process, as well as the importance of subject matter expertise to the practice of forecasting, are also well established.

Benefits from Formalization

Formalizing forecasting in the Federal government through development of a functional classification guide could produce a number of benefits. First and foremost, managers would be provided with an up-to-date description of forecasting work and criteria for evaluating candidates for forecasting positions. Over time, the availability of this information should produce higher quality selections, enhancing the credibility of the overall workforce as well as the forecasts it produces. Second, formalization would help insure that positions involving forecasting are accurately classified, and in turn, individuals who occupy those positions are appropriately compensated. In some cases, there may even be classification changes, i.e. grade increases, for particular positions once forecasting work is fully considered in setting grade levels.

Third, formalization may also draw attention to forecasting among agencies interested in establishing a forecast unit or enhancing their "anticipatory" capabilities, helping to expand the field and extending its presence and visibility throughout the government. Fourth, establishing a functional guide applicable to a variety of existing occupations, rather than creating an additional, new Federal occupation, would also avoid adding to the complexity of the current job classification system, which already includes more than 450 separate job series and titles. Rather,

[15]Such as that described in Osborne and Gaebler (1993).

a functional guide would provide a valuable tool to assist agencies in managing their current resources more effectively.

Finally, formalizing forecasting in the Federal government and recognizing its role and value is important simply in terms of acknowledging that the nature and complexity of the work performed by Federal employees, and in turn the level of expertise and sophistication required to perform that work, continues to grow and evolve. Recognition of forecasting on this basis, while intangible in its immediate impact, could nevertheless provide and important long-term benefit by communicating to Federal forecasters that their contributions and efforts are an integral and vital part of the process of governance.

Bibliography

Armstrong, J. Scott. Long-Range Forecasting: From Crystal Ball to Computer, second edition. John Wiley and Sons, New York, 1985.

Box, George E.P. and Gwilym M. Jenkins. Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco, 1970.

Bretschneider, Stuart I. and Wilpen L. Gorr. "Editorial: Forecasting as a Science," International Journal of Forecasting. 1989, Vol. 5, No. 3, pp. 305-306.

_____. "Forecasting Practices in the Federal Government," Federal Forecasters Conference/1991: Papers and Proceedings. September 12, 1991, pp. 17-24.

Brown, Robert G., Jeanne M. Doll, Hans Levenback, Len Tashman, and Lilian Shiao-yen Wu. "Business Forecasting: A Certifiable Profession?" A Focus for Bridging the Gap Between Practice and Research," The Forum (The Joint Newsletter of the International Association of Business Forecasting and the International Institute of Forecasters). Fall 1992, Vol. 5, No. 3, p. 5.

Farnum, Nicholas R. and LaVerne W. Stanton. Quantitative Forecasting Methods. PWS-KENT Publishing Company, Boston, 1989.

Fildes, Robert. "Editorial: Influencing Forecasting Practice," International Journal of Forecasting. June 1992, Vol. 8, No. 1, pp. 1-2.

Granger, C.W.J. Forecasting in Business and Economics, second edition. Academic Press, Inc., Boston, 1989.

Hamrick, Karen S., editor. Federal Forecasters Conference 1990: Proceedings. Staff Report No. AGES 9109, U.S. Department of Agriculture, Economic Research Service, March 1991.

Holden, K., D.A. Peel, and J.L. Thompson. Economic Forecasting: An Introduction. Cambridge University Press, New York, 1990.

Jenkins, Gwilym M. "Some Practical Aspects of Forecasting in Organizations," Journal of Forecasting. January-March 1982, Vol. 1, Issue 1, pp. 3-21.

Kennedy, Peter. A Guide to Econometrics, third edition. The MIT Press, Cambridge, Massachusetts, 1992.

Klein, L. R. "The Importance of the Forecast," Journal of Forecasting. January-March 1984, Vol. 3, No. 1, pp. 1-9.

Levenbach, Hans. "Business Forecasters See Benefits to Certification," The Forum (The Joint Newsletter of the International Association of Business Forecasting and the International Institute of Forecasters). Summer 1993, Vol. 6, No. 2, p. 3.

Lienesch, Thomas, editor. Federal Forecasters Conference-1992: Papers and Proceedings. September 17, 1992.

Makridakis, Spyros. "Editorial: Forecasting in the 21st Century," International Journal of Forecasting. 1991, Vol. 7, No. 2, pp. 123-126.

Monaco, Ralph M. "Developing an Effective Forecasting Program: An Economic Approach," Federal Forecasters Conference/1991: Papers and Proceedings. September 12, 1991, pp. 74-77.

Osborne, David and Ted Gaebler. Reinventing Government. Plume (Penguin Group), New York, 1993.

Saunders, Norman C. and Howard N Fullerton, Jr., editors. Federal Forecasters Conference/1991: Papers and Proceedings. September 12, 1991.

Solenberger, David. R. "U.S. Department of Agriculture Forecasts: Personal Comments on GAO'S Forecast Evaluation Methodology," Federal Forecasters Conference/1991: Papers and Proceedings. September 12, 1991, pp. 25-28.

United States General Accounting Office. "Status of Open Recommendations." Washington, DC, 1993.

United States Office of Personnel Management. "Handbook of Occupational Groups and Series." Washington, DC, October 1990.

_____. "Position Classification Standards for Positions Under the General Schedule Classification System." Washington, DC, 1990.

_____. "Qualification Standards for Positions Under the General Schedule." Washington, DC, 1990.

Wright, D.J., G. Capon, R. Page, J. Quiroga, A.A. Taseen, and F.F. Tomasini. "Evaluation of Forecasting Methods for Decision Support," International Journal of Forecasting. 1986, Vol. 2, No. 2 (1986), pp. 139-152.

Young, Peg. "Developments in Forecasting in the Past Decade," in Federal Forecasters Conference - 1992 Papers and Proceedings, Thomas Lienesch, editor. September 17, 1992, pp. 35-36.