Federal Forecasters Conference - 1992

Papers & Proceedings of the Fifth Annual Conference

Cosponsored by

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

The Fifth Annual
Federal Forecasters Conference

Papers & Proceedings of the Fifth Annual Conference

September 17, 1992
Washington, D.C.

TABLE OF CONTENTS

Morning Session

Federal Forecasters Conference Survey Results

Panel Discussion on Developments in Forecasting

Forecasting Supply and Demand for Health Personnel

Energy Modeling: The National Energy Modeling System (NEMS)

Issues in Forecasting I

Issues in Employment Forecasting

TABLE OF CONTENTS cont.

MORNING SESSION

Welcome and Opening

MR. LIENESCH: Welcome to the fifth annual Federal Forecasters Conference. I am Tom Lienesch from the Bureau of Economic Analysis and I'm the chairman of this year's Organizing Committee. This year we have close to 300 registered participants; we're offering 12 sessions; and we have 11 sponsors. So you can see, this event and organization continues to grow.

I'd like to individually thank the organizations and individuals sponsoring this year's conference--with financial contributions and through their creativity and efforts in organizing sessions and other aspects of the conference. This is definitely not a single-handed event. It takes the work and effort of a number of agencies -- I think our first meeting was in early April -- so this has been a long time in coming. In particular, I'd like to thank the Economic Research Service of the Department of Agriculture, specifically **Karen Hamrick**, for her work in securing this building. Without Karen we wouldn't be sitting in this very nice auditorium. Also from ERS is **Douglas Maxwell**; from the National Center of Education and Statistics, **Debra Gerald**; the Department of Census, who in addition to supplying our keynote speaker, gave us **Paul Campbell**; from the Bureau of Health Professions, **Herbert Traxler**; from the Bureau of Mines, **Pat Divine**; from the U.S. Geological Survey, we have **Tim Smith** who provided much help and guidance in organizing the Agency session. The Environmental Protection Agency joined this year with **Dave Rejeski** and **Joe Abe**. Both of whom have been a great help and have organized one of the more interesting sessions. From the CIA we have the Methodology Center and **Mai Nguyen** of the Research and Development Office. I would also like to thank last year's co-chairs of the conference, **Norm Saunders** and **Howard Fullerton** of BLS, who put on a conference that I can only hope to approach in quality. Finally, by way of a back door announcement, I would like to thank next year's conference chairman, **Ron Earley** from the Energy Information Administration.

Now, to welcome us to the Agriculture Building, I'd like to introduce Daniel Sumner, who since June of this year has been the Assistant Secretary for Economics at Agriculture. Now, I think that we can all grasp what a fairly large field that is. As Assistant Secretary, he has guidance and oversight responsibility for all the data collection, projections, policy analysis, and economic research that takes place in the Department of Agriculture. Before he became the assistant secretary he was the deputy secretary, and before that, a professor at North Carolina State. So, to welcome us to this building and to kick off this year's conference, I'd like to you to please welcome Daniel Sumner.

MR. SUMNER: Two years ago I had the privilege of welcoming the Federal Forecasters Conference to the Department of Agriculture. It really is a pleasure to provide meeting facilities for our Government forecaster colleagues. What I noticed about this conference the first time I participated was the breadth of forecasting in Federal Government that is represented. And, that is illustrated in the contest you have been running the last few years. Go down the list of forecast items. They cover a broad range of expertise, everything from corn prices to exchange rates to the win/loss record of the Baltimore Orioles.

Let me take this opportunity to provide a sense of forecasting at the Department of Agriculture (USDA). Our role is broad and includes everything from macroeconomic projections to hog numbers. But, we use broad economic forecasting mostly to indicate what the economic trends mean for agriculture in the United States not to duplicate the work done elsewhere. We also have our own weather forecasters at USDA, but again, not to duplicate what the National Weather Service does, but to indicate the effect of weather on crop conditions and what it means for the economics of agriculture. Several forecasting agencies work under the Assistant Secretary for Economics at the Department of Agriculture. The World Agricultural Outlook Board has forecasting in the title of the organization. The Economic Research Service (ERS) is a research organization, but it is also a projections organization. The flagship publication of ERS is Agricultural Outlook, which is a forward-looking outlet which includes projections and analysis that leads to projections. The other major economics agency is the National Agricultural Statistics Service (NASS). NASS collects data but also makes forecasts, particularly short-term projections. Indeed, one of the key activities for NASS is their crop harvest forecast. A few days ago NASS released their September crop harvest forecast for the fall crops. They released these forecasts at the same time as the World Agricultural Outlook Board released projections for prices and other crop conditions worldwide. The NASS projections garner interest in the agricultural media and throughout the industry because there is money to be made if one either anticipates or is able to react quickly to these projections.

The theme of your conference is "knowing your customer." That is certainly key to the work that we all do. In the case of forecasting and projections, our customers are often other government staff, but at the USDA we have an additional audience worldwide of customers, particularly for crop forecasts and similar projections. We have found that it is indeed important to know what our customers need and to tailor what we do to those needs. But, it is vital as well to make sure our customers know how to interpret our forecasts. Let me give you an illustration. When the USDA provides crop forecasts, the forecasts are not unconditional expected value forecasts of the size of the crop. Every December NASS forecasts the size of the orange crop in Florida. The forecast is based on actual examination of trees, fruit weight, combined with projection models based on both biological information and economic information. The published orange harvest forecast is explicitly conditional on the assumption that there will be no freeze in Florida until the crop is harvested. It is clearly not an unconditional expected value of the size of the crop. However, most data users know exactly what NASS is projecting. Based on their information, NASS reports that if a crop progresses to maturity with no freeze, X tons will be produced. People that use such a forecast need to add their own projections of the likelihood that we will have a freeze in Florida. NASS itself does not have the expertise to forecast weather and so they provide a conditional forecast. The Department of Agriculture could take the input from NASS, together with our own weather projections and produce an expected value of the crop. Let us say if the chance of a total killing freeze was 10 percent, we would multiply .9 times the NASS conditional forecast to publish on our unconditional forecast. What NASS does is tailor their forecasting efforts to their expertise. Rather than mixing the weather forecasts together with the current crop conditions, we report the current crop conditions expressed as a forecast, and let others take their weather forecasts from other sources. This example illustrates how it is important for forecasters to know the capabilities of their organization, and to make sure customers know the characteristics of the forecasts. For example, market participants should know that NASS does not make unconditional forecasts.

Let me mention one more example -- budget projections. In the academic community and elsewhere, there is a serious misunderstanding of Federal Government budget projections. Certainly that is true for the Department of Agriculture budget projections. At the USDA, we go through a very careful process to project the budget cost of agricultural subsidies. However, those projections are based on normal weather in the future years and also on current law projected into the future. Those baseline projections are used -- at least internally they are used -- to analyze alternative policies. They are not unconditional forecasts of budget outlays. Even if we think it is likely that there will be a different policy in place 6 months from now the projections in the baseline are based on current law. They must be so based to evaluate alternative policies. There are many observers who criticize our budget projections for not being accurate, even though the laws affecting outlays changed. But, we choose not to forecast policy changes. And, we probably have not done a good job in making sure our customers know that the budget projections are conditional on current laws.

My welcome for you this morning has emphasized the importance of not only knowing what your customers need, but also knowing the importance of making sure your customers understand your forecasts.

MR. LIENESCH: Thank you, Mr. Sumner. Now, for the fun part and for some deserving individuals, it's award time. Presenting the awards for this year's Forecasting Contest, or the most accurate forecaster, is Debbie Gerald from the National Center for Education Statistics, and Karen Hamrick from the Economic Research Service.

Award Presentations

MS. HAMRICK: Good morning. It's great to be here. It's so exciting to see so many people in the audience. We know this is the moment that 62 of you have been waiting for, the announcement of the Forecasting Contest winners. This is the second year that Debra Gerald and I have done the contest. We've had a lot of fun doing it, and we hope that you've had fun with it also. As you remember, we asked you to forecast five things. First the U.S. civilian unemployment rate for the month of August; and then four things for August 31st. They were the average bank prime rate, the cash price of number 2 yellow corn, the high temperature, and finally, the Baltimore Orioles win record.

I'll be announcing the runners up. Debbie will announce the winner. But before we do that, we decided to give an award this year for what we call the most courageous forecaster, for the earliest entry received. That award goes to Michael Lahr, Economic Research Service. If Mike is here? Come on down, Mike. Norm Saunders is giving out the awards, and we thank him for making these really beautiful certificates. Thank you, Norm.

Okay, the honorable mentions, if you'll come up when your name is called. John Cymbalsky, Energy Information Administration. William Miller, State Department. Ronald Trostle, Economic Research Service. None of those people are here? Clifford Woodruff, Bureau of Economic Analysis. Timothy Parker, Economic Research Service. Robert Gibbs, Economic Research Service. Thomas Snyder, National Center for Education Statistics. Don Kitchen, Council of Economic Advisors. Shelley Davis-Franklin, Bureau of Labor Statistics. Finally, the first runner-up, Patrick McCabe, Environmental Protection Agency. Now Debbie will announce the winner.

MS. GERALD: The winner of the FFC '92 Forecasting Contest is Larry Sink, Bureau of the Census. A poster with the names of the award recipients will be on display in the foyer of the auditorium.

MR. LIENESCH: Now we're going to do some place-switching, and Norm is going to give out the award from last year's conference for what was judged by a totally impartial panel of judges as the best paper of Federal Forecasters Conference 1991. In this case, I believe Karen Hamrick is going to pass out the awards. Norm?

MR. SAUNDERS: Thank you, Tom. Well, this was the last thing I had to take care of before I passed on the reins, as it were, of the conference to Tom. And it was certainly one of the more pleasurable events. We formed a committee. The committee looked at the 28 papers that were submitted. They looked at them with five criteria in mind: the significance of the paper to the audience's programs; the coherence of the paper; completeness and unity; the effective use of graphics; and the knowledge of the topic and of other research in the area.

Of these five criteria, the very first one that I mentioned, the significance to the audience's programs -- to your programs -- was judged probably the most important. Of the 28 we had five papers that surfaced that we felt were far superior to all the others. Four of those were selected as honorable mentions, and the people who were authors of those articles have nothing to be ashamed of; they were excellent papers. And we have certificates for each of the authors of those papers.

The first paper, the first runner-up, was "Using Dynamic Interactions To Aid Forecast the Case of Selected Urban, Rural Employment Measures." The author was Ron Babula, Economic Research Service, U.S. Department of Agriculture. The second runner-up, "Developing an Effective Forecasting Program and Economic Approach," by Ralph Monaco, also Economic Research Service, U.S. Department of Agriculture. Third runner-up -- "Neural Networks: An Exchange Rate Forecast" by David Stallings, also Economic Research Service, U.S. Department of Agriculture. The fourth runner-up, "Why Do Forecasters Fail to Predict the Big, Unusual Event?" Herman Stekler, Industrial College of the Armed Forces, U.S. Department of Defense. The bad news about the winner is that we only have one plaque. And it's bad news because there were co-authors of this paper. The paper that won the best technical paper for FFC '91, "Structural Models and Some Automated Alternatives for Forecasting Farmland Prices," by Carl Gertel and Linda Atkinson of the Economic Research Service, U.S. Department of Agriculture. Thank you.

Keynote Address

MR. LIENESCH: As you all know, the topic of today's conference is, "Forecasting and Total Quality Management." I don't think it requires a very large stretch of imagination to see that this is one of the more relevant topics we could have. It seems that virtually all agencies and certainly a large number of private organizations are exposed to TQM either through direct implementation at their workplace, or indirectly through relationships with other organizations that are involved in total quality management. It doesn't come without controversy. There are those who see TQM as dressed up common sense; those who see TQM as the only way to run an organization; and those who think TQM is the flavor-of-the-week fraud and a way for consultants to make money. Wherever you fall in that range, it's here. It has effects on everyone. And I think it's appropriate to ask those actively involved in TQM to come and talk to us; tell us what they think about TQM, their experiences in implementing it, and how it affects us as forecasters and as federal employees. So, with that in mind, I think you'll find our keynote speaker and the subsequent panel discussants to be eminently qualified to address this. Each of them, in their own way and level, are actively involved in TQM, and are mostly in statistical agencies. I think they'll be able to shed some light with their opinions on TQM.

Our speaker for the morning is Dr. Barbara Bryant. She has been the Director of the Bureau of the Census since 1989; in fact the first woman named to that position. Before she became the Director of the Census, she worked for 19 years as a marketing research executive in Detroit. She's been an editor, public relations consultant, professor. She's written numerous articles and three books. And perhaps more importantly, she has spearheaded Census Quality Management at the Bureau of the Census. So, would you please welcome Dr. Barbara Bryant, for our keynote address.

DR. BRYANT: I am humbled by being asked to keynote the conference of a group of forecasters. Somehow I feel as though you must be very visionary, clairvoyant people, and I'm more of a nuts and bolts, here's what life is like now, person. I did bring along my crystal ball to see if that would help me envision the future. It doesn't. It is just a fairly attractive crystal ball paperweight to hold down papers I work with here and now. So I decided, I will just have to be me, and let you be you. And given the theme of this conference of forecasting and total quality management, I will let you do the forecasting, including some of our people from the Census Bureau who do forecasting, and I will talk about the total quality management, or TQM, as the acronym goes.

TQM is something the Census Bureau is very deeply into, and something in which I am personally very much involved. In fact, my crystal ball paperweight has been holding down TQM papers and reports for nearly two years now. And while I've been thinking about total quality management on the Bureau-wide scale, there are forecasters at the Census Bureau who have been putting it into practice to improve their products and processes. And I might point out that some of them are nervy enough that we do population projections almost 100 years out, not just this fall's crop forecast. These are the people in the populations projection group of our Population Division.

So I thought this morning I would divide this talk into several parts: first, some background on the Census Bureau, and how we got into total quality management; second, how the Census Bureau has integrated strategic planning and total quality management; and finally, some concrete examples of projects resulting from process action teams in our Census Bureau population forecasting group, and what they've been doing using the principles and practices of total quality management. Just in case you think I misunderstand forecasting, let me add that the Census Bureau populations projection staffs are very research, quantitative, and data oriented. Like you and actually like me, none of us believe in crystal balls.

And now for a little background on the Census Bureau, and the history of our start up of total quality management. The Bureau of the Census, as many of you know, is a substantial part of the Department of Congress -- of Commerce. Boy, that was some sort of a Freudian flip, wasn't it? Worrying about budgets right now. The Census Bureau has a base employment of 10,500, and included in this are 3,500 part-time permanent field representatives who work out of their homes under the supervision of 12 regional offices. They do the interviews for the nearly 200 surveys we do every year. Some of these surveys are monthly, like the Current Population Survey; some quarterly; and some annual. In 1991 we actually fielded 682 separate waves of survey interviewing of samples of households or establishments by personal interview, computer assisted telephone interviews, and by mail. Now, the Census Bureau employment, as you might guess, is cyclical. It rises for the every five-year economic censuses of manufacturing, services, retail trade, wholesale trade, mineral industries, construction industries, transportation, governments, and agriculture. That's the phase we're just entering now to do the every five-year economic censuses. Of course our employment peaks in the years ending in zero, when every decade we field what up to then has been the largest peacetime army in history to conduct the census of population and housing. In May 1990, our employment peaked at about 340,000 temporary employees in addition to our permanent staff of over 10,000.

The decennial census is the project for which we're best know, and from which our name comes. Many think it's the only work we do, and so I really put in this little bit just to show you that we aren't like Rip Van Winkle, who goes to sleep for 10 years between population censuses. The decennial census is now 200 years old, which makes ours the longest running periodic census in the world today. There obviously were censuses in B.C. and in Biblical times, and China had one in 2 A.D. But nobody kept them up continuously, and so now we, who think of ourselves as a young country, have the oldest continuous census. I can't tell you whether the census of 1790 was done using total quality management principles. However, since Thomas Jefferson directed the first census, I'm inclined to believe it was managed according to the best known quality rules of the time. We can't say much about 1790, but what we can say about our history is that the Census Bureau pioneered statistical sampling in the 1930's and 1940's.

We've been using statistical quality control of processes in samples and surveys ever since. And I'll also point out that W. Edwards Deming, who taught it all to Japan, is a former employee of the Census Bureau. He learned about statistical sampling here. He acknowledged that last year when he came to give a lecture memorializing Morris Hanson, one of the pioneers of statistical sampling, for whom Deming had worked at the Census Bureau. Although we had not heard about total quality management at the time of planning the 1990 census, or acquired many of the techniques, we did build a great deal of quality assurance into the processes of the census. Quality assurance, however, was not done to enhance customer satisfaction, which is the driving force of total quality management. It was done instead to make the census fail-safe. The 1990 census was the most automated in history. We could not risk having software, computer hardware, management information systems, computerized mapping systems, laser sorters, microfilm processing, or high speed data capture, fail in the midst of the census. Our questionnaires were printed to exacting quality control standards to eliminate problems during automated processing. We built in quality assurance, which is different from quality control, in that you test everything throughout the process rather than fix up what didn't work at the end. Now, in one sense, this did have the customers in mind. We knew that if everything worked, the customers would have the data sooner than in prior censuses. And we knew this was something the customers wanted. But in general, the Census Bureau was a production oriented organization, rather than a customer focused one.

We were not, however, an organization without direction. We had begun strategic planning in the early 1980's, producing our first strategic plan in 1985 under my predecessor, Jack Keane, and our second in 1988. The developers of the second strategic plan examined the Census Bureau's strengths and weaknesses, identified opportunities and threats -- you've heard this language of strategic planning -- and laid out our goals for the next several years. These goals served us very well through the 1990 census. Staying on track, we planned to begin the third generation, the next cycle, of strategic planning, after that census was over in 1991. Well, in the meantime, along had come total quality management. It was toward the middle of the 1990 census year, while we were still under our second strategic plan, and less than two years ago now, that we started down the path to TQM with some skepticism that I'm sure all of you have experienced, and an enormous amount of arrogance because of this history of, well, Deming was our guy, you know. In our case, quite frankly, we felt we knew everything -- and I mean everything -- about statistical quality control and strategic planning. So what could TQM possibly have to offer the Census Bureau? We were open-minded enough to look, and so we appointed our Assistant Director for Administration, Cliff Parker, to explore TQM. And he had some training himself. He put together a small team, and naturally they put together a survey about TQM elements. At the Census Bureau, I must confess, our first reaction to anything we don't know anything about is to do a survey. Only this time, we also had to be the respondents. The teams surveyed both middle and upper level management on the TQM elements. You know the elements, or have heard them, top management and leadership support, strategic planning, focus on the customer, commitment to training and recognition, employee empowerment and teamwork, a measurement analysis of process and output, and quality assurance.

The results showed us that the cup was either half full or half empty. Our middle and top managers rated the Census Bureau neither excellent nor poor on these elements. On a scale of one to six, we put ourselves at a sort of a middling 2.7 to 2.8, and don't ask me what the plus or minus on that was because it was a fairly small sample -- very middle range scores. Well, what did the ratings show us? They showed us that, quite frankly, quality had not been an overriding priority. We had good TQM environment potential. Our strategic plan, though the top management was rather committed to it, had not really become a vision document for the total work force. Our customer focus was limited; it was not absent but it was limited. Our training and reward program did not support group performance, only individual performance. Teamwork was good, but empowerment needed improving. And the Census Bureau, as I've already showed, had quality control and not quality assurance, with the exception of the 1990 census processes. So clearly, there was something we could learn from total quality management. We put out a request for consultant proposals, and established an internal Quality Management Steering Committee. Our consultants helped us by providing advice and began our quality management training program. They provided quality management awareness training to our first group of upper level managers in January 1991, and they moved on fairly quickly to training a core group of facilitators, and next began providing problem-solving workshops.

Now -- pardon me. I'm going to get a drink of water. I went to the Hispanic Caucus dinner last night, and if you've ever tried to shout over really great Spanish music, you know it's worse than cheering at a football game in Ann Arbor, Michigan where I came from originally.

In an agency the size of ours, we soon realized that the Census Bureau had to take over the training, and our consultants agreed. Looking back, that was really one of the best decisions we made. It gives TQM a real chance for long range success. We took over the training for three reasons. First, it would have been prohibitively expensive -- well, cost is always your first reason, you know. It would have been prohibitively expensive to use consultants to train over 10,000 employees. Second, in-house trainers could make classroom examples much more relevant to Census Bureau activities. Thirdly, and I think the most important, the best way to learn something is to teach it. By creating our own staff of teachers, we built in a core group of believers. You cannot teach something if you don't believe in it. Converts do make the best evangelists.

We adapted the contractor training package to meet our own needs, and provided train-the-trainer sessions for volunteer trainers. By March 1992, 12-13 months later, virtually all of our general work force had been through some level of training, and many supervisors and managers had been through more than one level. The 3,500 part-time interviewers that I mentioned earlier, or field representatives as we call them, who work from their homes, had received audio tapes and a study manual. As they came into regional offices, as they do about every six months for update training on one of the surveys, we gave them an additional session on TQM.

I've been using the term "total quality management" as though we actually called it that at the Census Bureau. We do not. One of the first suggestions to come from an employee occurred after one of the earliest awareness training sessions. "TQM sounds just like another management fad," he said. "We've been through zero-based budgeting, MBOs, and strategic planning. If we want employees to take this seriously as a new way of running our business, we've got to give it a name that shows it is a process the Census Bureau expects to use for a long time, even if the technique changes and evolves, and the terminology shifts." The name he suggested was Census Quality Management, or CQM. We took the suggestion immediately. The first example, I guess, of employee empowerment in our TQM process.

Now, recall that when I described strategic planning, I said that the intent was to develop our third strategic plan after the 1990 census. Thus, in early 1991, we started resuming strategic planning meetings, after we were already into the initial quality management training. At the second meeting of the strategic planning committee, a number of us almost simultaneously threw up our hands -- this is ridiculous. We can't have Census Quality Management going this way, and strategic planning that way, or even have the two in parallel. There's not the time; there's not the energy; and it certainly is going to cause a lot of confusion. Now, it just happened that we were having our strategic planning meeting in a corridor, where a number of the awareness sessions were going for total quality management or Census Quality Management. And we realized that one of the things these awareness sessions was supposed to do was that the employees were supposed to be talking about what are the barriers to quality in this organization. We said, if we're sitting here talking about strategic planning and looking at what are the Census Bureau's strengths and weaknesses, we'd darn well better find out what the rest of the employees say the strengths and weaknesses are. We literally, several of us, just spread out and went down the hall -- the awareness session had been posting these posters all over the room with the magic marker stuff, tearing them off the easels. We just went and scooped up some of them off the walls and easels. From then on, each draft of our strategic plan or component was given to the CQM steering committee for a reaction. And so we began this interactive process developing the strategic plan and CQM.

Total quality management requires that an organization develop a vision and a quality policy to use in establishing strategic goals and action plans. At the time the first strategic plan had been developed seven years before, the Census Bureau had put together a mission statement. The strategic planning committee worked a short time trying to rewrite the statement into a vision, and the rewrite just didn't fly. Some of us felt that if we dumped this mission statement, which had been posted on the walls for seven years, that looks a little flaky. So we decided that we would have three things, not just the two. So, we kept our mission -- the Census Bureau's reason for being. We put together a vision -- what the Census Bureau wants to be. And then quality policy -- the cornerstone of the Census Bureau's commitment to quality management.

Now, a major function of strategic planning, of course, is the development of strategic goals. The Census Bureau's first two plans had very specific goals. They identified a series of projects for carrying out each. They assigned responsibility to specific personnel to do the projects, specific responsibilities to certain managers to oversee them. This top-down strategic planning to employees, with managers assigned monitoring responsibilities, no longer fit into the CQM mode. Former plans had neither customer focus nor employee empowerment. Like the Census Bureau itself, they were process-oriented. Our new strategic plan -- and there are copies of it on the back table out in the outer lobby for you, if you're interested -- we named Census Quality Management through Strategic Planning. It contains 10 very broad strategic goals -- down the center here. Each has a short description and defined target areas. There's not a word in the plan about how to achieve the goals. Employees at all level are empowered to develop the projects, to move the Census Bureau toward the goals. I say move the Census Bureau toward the goals, rather than achieve the goals, because most of the goals are moving targets. If you get better at them, then your expectations rise and you've got to keep on doing better.

Our 10 are: (1) meet or exceed customer expectations; (2) improve the product line to meet customer needs; (3) recognize and value respondents and other data suppliers, a particular type of customer that a survey research organization has; (4) enhance our own employees' career environment; (5) automate effectively. For an organization with the computer power of the Census, to get the goal in two words is something of an accomplishment, but it means there will have to be many projects to work on that goal; (6) improve administrative systems and management; (7) increase research capabilities and the relevance of research results; (8) provide an integrated international perspective for statistics and analysis, and we do international data as well as domestic; (9) improve the decennial and quinquennial censuses; and (10), the one I think the employees like the best, is consolidate headquarters employees in a modern facility. We're in a rather 50-year old monolith with five satellite facilities out in Suitland, Maryland.

To launch the strategic plan, we were very conscience of the fact that the prior strategic plans had really not gotten through to all levels of the Bureau. There wasn't a buy-in to them. So even while this strategic plan was in the draft -- as I say, we had been sharing it with the CQM Steering Committee, and with the CQM awareness sessions. During January of this year, early 1992, one year after the start-up of Census Quality Management, and seven years after the first strategic plan, we formally launched Census Quality Management through Strategic Planning. By then we had nearly 100 process action teams already working on quality improvement projects. And most employees had become conscience of identifying who their customers were, something about them, and something about what they want. That was an easier concept for those with external customers, like the people who work on the Current Population Survey who know they've got to please Tom Plewes at the Bureau of Labor Statistic, than those for whom their customer was just the next department over in the Bureau.

Our largest meeting space is a new 350-seat auditorium, about the size of this, a little smaller I think. In order to reach all our headquarters employees, it took us 12 back-to-back sessions, four per day, over three days, to meet with everyone at our Washington area headquarters. It was the first time, I think, since the Census Bureau had any size, that every employee met face to face with Executive Staff and the Strategic Planning Committee. These meetings were not one-on-one, but there was interactive Q&A time from the audience. Before the meeting, we had given every employee the folder I just showed you. Then we went out to our regional offices and some member of the executive staff did this in every regional office. And then we have a big data processing facility down in Jeffersonville, Indiana, and we assembled 500 employees at a time on folding chairs on a big empty warehouse space and had face-to-face sessions with them.

Well, when people ask me whether CQM through strategic planning is working, I tend to answer somewhat weasel-worded. We'll know better a year from now, or we'll know better a year after that, or 10 years from now, if this has proven to make a real change in the way the Census Bureau conducts its work. The proof of the pudding is in the eating. But that's really not quite fair to put off answering. We have instituted change, but we still have a long way to go. I think it's very important that we demonstrate results and provide feedback of this way of doing things if it's going to succeed. Poor communication is the biggest barrier to quality that our employees identified. And I think that barrier will be identified in most organizations, and particularly in large ones. We've tried to do something about that by instituting a newsletter, Census Counterparts, for all employees. We've held feedback and question and answer meetings with lower level supervisors with whom the Director and the Executive Staff have not met before. For this, we've been helped again by the new auditorium I mentioned.

But now I'd like to give you several examples of recent CQM activities in our population projections program, since that is where our forecasters reside. The three examples I will give you are really what has come out of process action teams. Somebody earlier in the introductions referred to quality management as organized common sense. I think that you'll see that these projects really are just sort of organized common sense, with a customer service focus to them.

The first is expanded race/ethnic information. We've been providing population projections by age and sex for the White, Black and "other" races groups since the mid-1960's. For the past few years, we've also been supplying our customers with data for Hispanics. Although they appreciated this enhancement, many of our customers were not happy with the little footnote attached to all our Hispanic origin data. That footnote reads, "Hispanic origin may be of any race." They were frustrated by this overlap between our race and our Hispanic data. In addition, many others wished to have "other" races, as we called it, disaggregated, separately, for two groups that have some size now, Asian and Pacific Islanders, which are about three percent of the population, and the American Indian, Eskimo, and Aleut, which is about one percent. Well, I'm pleased to announce that in October, we're releasing a new set of national population projections, which meet these needs of our customers. These projections will provide data for the four major race groups, and in addition, each race group has been separated into its Hispanic and non-Hispanic parts. Therefore, we're now producing statistics for eight distinct and non-overlapping race/ethnic groups. Later this year we plan to release new state population projections, which for the first time include separate information for the four major race groups, and the Hispanic origin population. And I should mention that all of these products are consistent with the 1990 Census as enumerated, so that people can look at the actual count in 1990 versus the projections, which, as I say, we go many, many years out on. We know many customers' needs are going to be met by these enhancements. However, we know that this work still did not satisfy all of our customers' needs for race or ethnic data. For example, some have asked that we begin providing estimates of the population of individual Asian and Pacific Islander groups: Chinese, Filipino, Japanese, Samoan, et cetera. Others would like us to provide data for groups within the Hispanic origin population: Mexican, Puerto Rican, Central American, et cetera. As you can see, we get into the customers sometimes wanting things that are a lot further than sample sizes will allow us to provide. Still others want separate data on the foreign-born, or about the population of the Commonwealth of Puerto Rico. Now, I can't promise that our implementation of CQM at the Census Bureau means projections for these groups will be available any time soon -- the conflict between customer needs and sample availability. But any such changes must be weighed against competing customer demands for other products, timeliness, or accuracy, as well as against the available resources.

A second TQM project from the projections group is the increased frequency of revision. Although we've been producing population projections since World War II, they've never been updated with any regularity. We've tended to say that they were revised when necessary--"when necessary" sounding suspiciously like when we get around to feeling like doing it. Many of the customers, on the other hand, revise their own projections on a set schedule of one or two years. And it's fair to say that our uncertain update schedules often created unhappy customers who felt that our definition of "when necessary" was unsatisfactory. Well, I'm pleased to tell you that we're now committed to a production cycle which responds to our customers' needs. Beginning in 1993, we intend to create and release new national and state projections every other year. We think this establishment of a regular schedule is going to benefit those who use the projections.

One final project I'll mention this morning. I told you that a year ago we had 100 process action teams going. We've now had some 140, some of which have finished. But there is not time in the conference for all that. But I'll do one third one, and that is return to a preferred series in state population projections. Again, this is a little bit organized common sense, when you go back to where you once were just because the customers were happier when you did. This actually happens to be one of my most dynamic examples of CQM in the population projections area. It's illustrative of the difficulties we all have in our attempts to maximize customer satisfaction. During most of the 1970's and 1980's, our state predictions contained only one series, or had several series but one was given as a preferred. In recent years, however, our most recent state projections contained four alternative series, none of which is designated as preferable to any other. This change to equally likely scenarios was made in response to complaints from some of our customers who made their own state population projections. The state population projections we are now preparing, however, return to our past practice of producing a preferred series. This is in response to customers who said, if you give us four, tell us which one you like best. Many customers publish some summary volume and just don't have room to run four series in it. We even found that some were taking our two middle series and averaging them to get down to three. We made this change, as we've made many others in the past six or eight months, after listening to customers. We're trying to do more and more of this, like all of you in Federal service. Why? The range of customers is so broad it's really hard to measure satisfaction. We can't just go and say, well, we sold that many of that item today.

Let me end now by talking about a few of the lessons and difficulties. During our work with quality management, we've learned some important lessons. Projects selected should have a narrow, well-defined focus, with a potential payoff. Just in time, training for teams is important so as not to forget the techniques learned. The team leader and facilitator should be trained before starting the team, but not too far in advance. Facilitators play an important role, and are most effective if detached from the specific project. That is, facilitators from outside the office or division that's actually doing the work seem to be the most effective. Recognition is important, no matter how small it is. We're using on-the-spot awards, certificates, little coffee parties to recognize the accomplishment of a project. Management needs to offer support and guidance without interfering. And managers must champion the cause for the employees to become disciples.

We've also encountered some difficulties. When the management steering team for the process action team questions the results, there is a little tendency to say, you are interfering and I thought you were empowering us, and now you're saying you didn't. When a process action team makes good recommendations that cannot be implemented because of budget or resource restraints, they feel as though they've wasted time and effort, even though they may have done a very good job and come up with some very good ideas. But to them, CQM becomes a sham because we can't put it in place. In that light, we've found that process action teams that work on improving space requirements, personnel matters, procurement matters, or other types of support activities, become more frustrated, or more easily frustrated than those who work on something more under their control. If you don't have the budget to redesign the building, don't get process action teams going on how we could redesign this building.

Measurement of results is a stumbling block, which I'm almost embarrassed to say because we consider ourselves the measurers and fact-finders of the nation. But we have trouble measuring results. For some projects, it's easy to decide. For our 1990 Census products it was very easy because we had another measure called the 1980 Census. Every time we got a product out four or five months earlier than for 1980, we clearly had made some gains. But for most projects, it's a lot more difficult than that. Supervisors, in their normal course of activities, are often viewed as impeding the CQM process either by not being willing to take the suggestions from the staff or making it difficult for the staff to participate or be a facilitator in a team project outside their own area. We're thinking of setting up sort of a CQM counselor system to get around this.

In summary, though, we do believe there are positive benefits to Census Quality Management. The staff do believe they can propose ideas and have a fair evaluation of them. Even though some managers felt they were operating in a CQM manner, the new word "empowered" seems to have emboldened some who sat on the sidelines before and it has shaken up a few managers. Our lines of communication do seem to be improving. There's more use of CQM in informal settings. I think most of all there's much more focus and consciousness about customer needs and customer satisfactions. People just in their day to day reporting on projects, are saying, this is the advantage to the customer if we do it this way. So I'm particularly upbeat on this last benefit. At the end of a year and a half, Census Bureau staff are talking about customers, who they are and what they need. Now, we still have a very heavy focus on process, getting the work out, the survey completed, the census tape delivered, the population projections made, or the data report published. However, we care more about what data the customer wants in that report, and how we can make the data products more useful to them. Customer satisfaction used to be the sole concern of our Data User Services Division. It was sort of like an afterthought -- now that we're produced the product, here it is for you to take to the customers. Increasingly, all of us understand that we have many customers with many needs, both within and outside of the Bureau.

To conclude, Census Quality Management provides the customer focus. Strategic planning provides the direction. Together, we think they are a winning combination.

Panel Discussion

MR LIENESCH: Thank you, Dr. Bryant. I would like to start the second part of our morning program now. Continuing the exploration of our theme, I'm going to introduce our panel discussion moderator who will run this portion of the program. Once our panel discussants are finished with their comments, we will open the program questions from the floor.

I'd like to introduce Suellen Hamby, who is a senior executive at Internal Revenue Service--nominally I should say--where she is the Director of the Resources Management Division. This is a division of over 300 employees responsible for all administrative support in the Internal Revenue Service. However, more to the point for our purposes, is she assisted in the formation of quality councils at IRS and helped implement TQM in that organization. Closer yet to the point as to why she is the perfect moderator for our purposes today, she is a founding member of the Federal Quality Institute, which was started in 1988. Since that time she has played a very active role in bringing TQM to various organizations. I know she's playing an active role in my organization, BEA. Please welcome Suellen Hamby.

MS. HAMBY: Thank you, Tom. When Hugh Knox and John Kort invited me to moderate this panel and said it was for the forecasting conference, I thought this is terrific. I just saw a wonderful special on PBS about hurricanes, and I really wanted to see how total quality was used in predicting the path of Andrew and Iniki, and they had to say no, no, Suellen--Economic forecasting.

And I said, you know, that's even better, because we at the FQI have worked with each of the agencies that are represented up here on the stage. Implementing total quality in their organizations, designing plans, and forming quality improvement teams; all the things that Dr. Bryant talked about earlier, and that we'll hear more about this morning. The bureaus represented here are in various stages of implementation. Some are relatively sophisticated. Others are just beginning to map out their strategy and form quality improvement teams of the employees to let them recommend how to do the work better. All, however, are headed in the direction of better customer satisfaction, employee empowerment, and continuous improvement of the way they do the work -- the three founding principles of total quality.

It's not an easy journey. We've all read the very recent articles on the demise of total quality. That may be so for those companies and agencies who are still succumbing to the pressure of the quick fix. Others -- and I think it's fair to say, those of us up here on the stage today -- believe it just makes good sense. It's here to stay. It may be evolving. It may be changing. It's certainly maturing. But it does provide a solid foundation for short-term, mid-term and long-term improvements. And the payoffs are coming back in terms of pleased customers, energized innovative employees, and streamlined simpler ways to do everybody's job every day.

The theme today is knowing your customers. Think about customers really on three levels. The first level will be those external customers that you all have in both the public and in industry. External customers in other agencies, or that second level, such as those who are represented here today. I know many of you work across bureau or across agency lines, and are customers and suppliers of information, data, and analyses to one another. But we can't overlook that third level of customers, the internal customers, our employees. That may be the toughest area for us to change our old paradigms and move to greater empowerment, sharing with our employees greater authority, responsibility, and accountability. I am forever having to explain to reluctant managers that empowering employees does not mean giving them permission to charter a helicopter to take a memo across the street. Along with empowerment comes those other two issues: accountability and responsibility. You're doing well at the hard stuff, the data gathering, the analytical problem solving. It's easier, especially for those of you who deal with these analytical tools and statistical tools in your professional everyday lives. The soft stuff, the trust, the mutual respect, the letting go of control -- it's hard, but it deserves our commitment. We like to say the hard stuff is easy but the soft stuff is hard in trying to implement total quality.

Before we begin, I'd like to just reinforce several of the points that Dr. Bryant made this morning. One of them is that management leads the effort. They lead it through training, through goal-setting. Dr. Bryant talked this morning about the importance of training and talked about using the in-house trainers at Census to cascade down that training to all the managers and employees. It struck a familiar chord because at IRS Fritz Scheuren and I were among the trainers at the executive level who subsequently trained all 10,000 managers at the IRS in a three-day quality leadership course.

Dr. Bryant also talked about the idea of focusing the goals that all can understand. Fritz will remember that when we first started getting our initial handle on strategic planning at IRS, we came out with no fewer than 58 strategic initiatives. We used to go around at executive meetings, saying, well, I'm working on the task force to implement number 23. And Fritz would say, well, you know, I'm on 17 and 41. It got to be like those old jokes of the prisoners telling jokes in the prison, all you had to do was shout out a number and everybody understood. The difference here was there were so many that none of us could keep track. None of us could remember. All of us lost focus. To IRS's credit, we've gotten that 58 down to a much more manageable, focused handful.

But the key point here is, focus in on what is under your control, or at least under your influence. Don't tilt at windmills, but zero in on your daily operational processes, on how you do your work, what you do when you hand your work off to the next person, and find out how that can be more streamlined, more simple, with more authority-delegated down to the lowest level.

Now, let's hear from our panelists. I'm going to ask them each to speak for about 10 or 15 minutes, and after all have spoken, ask them to share among themselves some lessons learned. And then we'll take questions from the audience.

The first one to speak will be Hugh Knox. Hugh is the Associate Director for Regional Economics at the Bureau of Economic Analysis. Before joining BEA he was a deputy assistant secretary at the Economic Development Administration. He has published widely in the areas of regional economic development and economic impact analysis, and he has completed graduate work in economics and regional science at the University of Pennsylvania, and taught regional economics and regional planning at the University of North Carolina at Chapel Hill. Hugh is also one of FQI's favorite customers. Hugh?

MR. KNOX: Thank you, Suellen. I have a few comments, not too many, because BEA, in terms of the agencies represented here, is much the junior partner in the TQM efforts. We have been involved with TQM at the Bureau, and only the regional program within the Bureau. It's not a Bureau-wide effort at this point. We've been involved for about a year, intensely for maybe nine months. In terms of where we are in the process, we have formed the BEA Quality Council. We have done some limited training of all managers. We have done even more limited training for all staff in the regional program at this point. We hope by November 1st to have our strategic plan put together, at which point we will share that with the regional program staff for their reaction before we move along. We have not yet had an action team put together. We have identified people, and we have identified a topic. But we are struggling now with the appropriate way to train those people so that they might become a core of trained people that we could use throughout the rest of the program.

For those of you not familiar with what we do within the regional program and who our customers are and who our suppliers are let me say briefly that we have three primary products to use to satisfy our customers. The first is a system that generates economic impact multipliers, called the RIMS system. Clients there are usually consultants who are working on one project or another. It's very site-specific. It requires fairly quick turnaround time. Another set of products we have come out of a system of state econometric models which provide annual projections for up to eight years. Our clients there range from other federal agencies to research institutes like the Urban Institute, to state governments, to anyone who has an interest in projections of what will happen in the various states. The third major product is a set of long-term regional projections. Very much like the long-term projections that Dr. Bryant mentioned this morning on the population side, these are projections which combine both population and economic forces. They are also similar to the kinds of projections that BLS produces that you've seen in the Monthly Labor Review, and I think they are a topic of a later session in the program.

So our clients range widely from individuals to consultants to federal agencies to state agencies. At the current time, we do not have a very structured way of gathering indicators of customer satisfaction. We do have a system where we exchange the econometric projections with anybody in the States who has their own set of projections and is willing to enter into a joint exchange. We have a similar kind of arrangement in our long-term projections with members of the Federal State Co-op on Population Projections, where we share our projections numbers with members of the co-op who are representatives of their states for their comments. Those two customer interactions have been very beneficial for us, and I think beneficial for the other people in the networks. But we have not yet sat down and looked at the issues in a different way. Instead of going to them and saying, here are our projections, let's see your projections, and let's discuss the differences, we would like to get to the point where we can say, well, what do you want to see? What would you like to have us do differently? And I think that is one direction in which we will go, certainly by the time we release our next round of long-term projections in 1995.

On the supplying side, the three other agencies represented here, are major suppliers for us. Without the Census and IRS and BLS providing us with source data, we would be dead in the water. There's no question about that. Yet we have no regular process to communicate with our suppliers either. The three agencies have been involved in TQM longer than BEA has, and I have seen a positive change in some of the communication processes. They were good to start with, but they've changed in character. When making calls from the regional program staff at BEA to our sister agencies, we get a different kind of response now. It's a response that we're calling as a customer, not as someone who is generating more work. Of course, we are generating more work, but it seems to be received differently, and I would attribute that to a change in focus on customers in those agencies. I would like to see the same kind of change happen at BEA once we have our training completed throughout the Bureau and have a better idea of how to go out and contact our customers on a regular and structured basis.

Finally, I might add that the BEA Quality Council has been in operation now for about nine months, and the staff, from my reading, is getting quite impatient with the eight of us going off to our meeting room three hours every other Wednesday and coming out with nothing for them to do, no way for them to change their behavior. They keep asking us when we're going to do something, when we're going to let them get involved in the process, and why don't we just get out of the way and let them achieve something. And we have a great deal of sympathy with that position, and we're looking forward to the completion of the strategic plan, at which time we hope to have as many of these process teams as we can. By the way, we've changed the acronym -- at least we think we have -- to Work Improvement Team so that we'll be able to keep our WITS about us. Thank you.

MS. HAMBY: Thank you. Our next speaker is Tom Plewes. He is the Associate Commissioner for Employment and Unemployment Statistics at the Bureau of Labor Statistics, where he has held a variety of positions since 1973. Before moving to the BLS, he was with the Department of Labor's Employment and Training Administration. He has a Bachelor of Arts in Economics from Hope College, a Master of Arts in Economics from George Washington University. He is also a Brigadier General in the Army Reserve. So, Tom.

MR. PLEWES: Thank you. There are lots of reasons that I'm on this discussion panel today instead of up here earlier today getting an award. The major reason is, I'm a lousy forecaster. I forecast the fact that, being on a panel with Fritz Scheuren, we'd have an opportunity to have lots of slides and so forth because Fritz always uses slides. Today Fritz chose not to use slides. So I have my slides here, but I have no overhead. I just assumed that, Fritz being here, we'd have an overhead. So what I'm going to do is ask everybody to move very close forward. My second assumption was that, since I was going to have the overheads, you'd be completely satisfied with overheads so I'd only have to bring 25 hard copies of my presentation because there would be a few people who might be interested in picking up this thing to fill up their books. I've only 25 copies of my presentation there. I'll follow my presentation as best as I can without the overheads, and you are certainly welcome to come forward after this session and pick up copies of the presentation.

The name of my presentation, if it has to have a name, is The Cost of Unquality in Federal Statistics. I think The Cost of Unquality in Federal Statistics makes a case for total quality management. I want to put that in the context of federal forecasters. That context causes me to recall one of the basic principles of total quality management-- addressing my customers needs; pleasing my customers, if you will. That's what I want to talk about today. I start, not as Hugh did, with a description of where we stand in terms of total quality management because I provided a full report on that in the back of the room. All the principles, the bases, those lessons that we learned and refined from the Federal Quality Institute are in here, along with a catalog of the kinds of programs and the kinds of process action teams that we have put together, with some indication of the success of those teams.

It's important to spend a moment to define what we're talking about in total quality management. We call it quality improvement program, or QIP. For us, QIP is a management technology for continuous improving performance at every level in every area of our responsibility to ensure customer satisfaction. Now that definition goes on at length, as definitions for TQM tend to, and I will not bore you with the rest of the definition. I think it's important to understand that there are some key principles that are involved in here. One of the key principles that we have to focus on and that we must indeed use as a primary generator of what goes on in TQM, is the idea of ensuring customer satisfaction. What we're doing in total quality management is captured in nine principles. I know that the Federal Quality Institute, being parsimonious, has three basic principles, and most people think of three principles. We've kind of expanded that a little bit to nine principles because they incorporate not only the outcome of total quality management but also how we would go about doing it.

Two of those nine principles have to do with customer satisfaction. The first of those principles, the most primary of the things that we are concerned with, is the idea of **understanding the needs of our customers.** We focus on outside customers -- the Secretary of Labor, the Congress, the states, the press, academia, business, labor, and the public -- and we focus on internal customers -- our matrix partners within the Bureau. One of our major customers, a number of whom are sitting here today, is our Office of Employment Projections. We try to get a thorough and systematic understanding of the needs of our customers, both internal and external because that helps us to establish our direction and goals. Sometimes we must help our customers clarify those needs, but the basic outcome is the understanding of the needs of our customers.

And the second principle is **meeting the requirements of our customers.** Our success in accomplishing what we are to do in the federal statistics business is finally measured, and ultimately surely measured, in the responses of our customers to our products and our services. We actively seek feedback on what we're doing to meet those needs.

We strive for **error-free work.** We understand that is not a reasonable goal. We aren't a zero-defects organization. There are defects that spring in, but we don't strive to have a 99 percent non-error rate. We strive for 99.9. We understand that is a constant goal; we need constant improvement.

Management commitment is focused in our quality council. All our senior people are on the thing.

Management by prevention -- trying to figure out what the process has in it that can fool us before it fools us.

Top-down implementation -- we do things at the management level first, then we involve other people as time goes on, solving problems at the appropriate level. Management cannot be involved in solving problems for people who are actually doing the work out there. They've got to solve their own problems, and we empower them to do that.

Teamwork -- we form teams, as most other people do. We form process action teams; facilitate those teams; train them; and set them loose on the problem and understanding the process. They come back to our quality council with their recommendations. We say yes, we resource them and they get out and solve the problems.

And finally, investment in people. We've got a very active training program that the FQI has helped us set up. We're really pleased with that.

Well, we have to start with a basic understanding of who our customers are. Certainly, our customers are policy-makers. You've seen that in recent weeks when we have put out, for example, our recent unemployment statistics and employment statistics. Three times in the last nine months, the Federal Reserve Board has changed the re-discount rate to try to change your interest rates on the day we have issued our data. Policy-makers, not only there within the executive branch, but also in Congress, are our customers. The financial markets are our customers. They have a different view of the kinds of needs for data, and different kinds of requirements. Program administrators, who want to help in setting up programs, and in evaluating those programs, are interested in a longitudinal look at what happens to programs over time. They have a different set of requirements. And finally, forecasters. Forecasters have a set of requirements that it is important for us to understand. Forecasters, as I say, both internally in our Office of Employment Projections, and externally, in the Bureau of Economic Analysis, the Federal Reserve, the Treasury, and then certainly in the private sector.

We have a lot of feedback, quite frankly, internally within the Bureau. We know the needs of those persons who are involved in projections based on our data, especially our industry, occupation, and our work force data. Hopefully, that translates into what others need. But there certainly are others out there with really different needs, and we've got to talk to them also. The questions we've got to ask are: Are we meeting your needs? Are you satisfied? What are we doing well? What do we need to focus on?

We're involved right now in doing what I call a customers satisfaction survey. And some of you have been involved, I would believe, and many of you -- those of you who have the chance to pick up the telephone and ask BLS for numbers over the next few months, will likely be involved in this customer satisfaction survey. We're trying to get a handle on those four questions I asked. So far, I think that the lessons that we've learned from the customer satisfaction survey can be broken down into three areas: areas that I think that we need to improve; areas that will have lesser priority, nice things to do; and areas we're doing a good job on. Our customers tell us that we need to improve on our standards of timeliness and currency of the data that we publish. We need to improve on the ease in which our customers can get in touch with someone who could answer the specific question. We need to improve -- and this is endemic to the Bureau of Labor Statistics unfortunately -- we need to improve on the staff's ability to explain conceptual and analytical issues without using overly technical language. We've got to lay data out in a way that users can understand it.

There are other things that we've got to work on. Demographic, geographic, and industrial coverage of statistics needs some fine-tuning. We need to improve giving referrals if we don't know the answers. The technical limitations of the data need some polishing. We're doing a pretty job, our customers perceive, in four other areas. The data do meet the standards of accuracy and reliability. The information is received promptly. The questions are answered willingly and promptly, and the staff generally are knowledgeable and competent, although you have to weigh that against the fact that they explain things in overly technical terms.

What do you need from us? We're in the process of understanding that. But I think that there are a few things that I can understand that you need from us. First of all, I assume you need from us a clear conceptual foundation for the statistics that we provide you. Now, we know that we violate that. We violate that in very basic ways. We have two definitions of employment floating around. One comes from the household survey, which is a measure of people, and the other comes from the establishment survey, which is a measure of jobs. And they often move in different ways. We've got dozens of concepts of earnings and wages, and those of you who use earnings in your forecasts have to do a lot of shopping and investigating, much more than you should be doing, or need to have to do, to get access to these data. We deviate in some important ways from the standard classification systems, from the Standard Occupational and the Standard Industrial Classification systems. We try to keep as close as we can, but there are practical reasons that we can't be exact, and we sometimes don't do a very good job of explaining it. We know forecasters need a consistent time series. In fact, you would like to have us sometimes stop, what I consider to be progress so that you can lock in a time series long enough so you can use it in a forecast. When we do make changes, we know that you need bridges and crosswalks to help explain the discontinuities, and do so in a very simple way. We know that you like minimal revisions. And yet here we are, revising our monthly employment series three times, and then we turn around and do an annual benchmark which has a subsequent revision. But I know that you don't want revisions because it's hard to keep up with the kind of data that I provide you for your models without having the revisions to keep up with too. I know you like rich detail, the more the better. And I also know that you like detailed geography, at least some of you do. I know some of you are satisfied with state data. A lot of you would like to have data for all counties, and much more frequently than we have. And I know that you like data with ready access. You like it in electronic bulletin board forms, and because BLS doesn't have one, you turn to the Commerce Department bulletin board. You like to have your data on diskettes. You like to have your data in CD-ROM format. And those of you who have used our publication, the Employment Earnings, have thrown up your hands saying that it's a very un-user-friendly publication. We know that those are your needs.

I want to shift now from consumer focus to the cost of un-quality. The price of poor quality data in the Federal Government, as the Boskin Group for the Federal Economic Indicators Panel took a look at several years ago, is paid by a lot of people. The price that you pay as forecasters for lousy data is in bad projections. And bad projections lead, unfortunately, to flawed decisions, and, sometimes, to litigation. So we know that one of the prices of un-quality is a bad projection, flawed decisions, litigation. We know also that there is a loss of confidence in government statistics--a loss of confidence in the government if the data are of poor quality. We know that there are missed schedules, and we know that there is an added cost of rework. Rework in your projections, rework in the statistical data. There is a cost to having data which are not of high quality.

We had a good example of the cost of unquality that I'll close with. Perhaps we can discuss it in the discussion period. We made a fairly large revision in our employment estimates for the first quarter of 1991.

Many of you are familiar with that issue and what happened. Basically, to recall the issue, our very large survey of establishments -- about 380,000 establishments monthly -- is a survey operation. We are very fortunate in the operation of the unemployment insurance system to have a benchmark capability. I think we're fortunate. We're the only country in the world that can actually benchmark your employment estimate to a really good count. Very often, other countries can't do that. I say "I think" because sometimes it's not a blessing. In the benchmark process we have an ability, five months after the end of the quarter, to see how well we're doing in the survey. Once a year, on a regularly scheduled basis, we re-anchor our survey to that benchmark, to that count. We had a very large discrepancy of about 620,000 nationally in the first quarter of 1991, following many, many years in which the discrepancy was about 200,000, something that nobody got excited about.

We know from whence the discrepancy came. One cause was an inaccurate estimation of what was going on in small businesses in the survey over the course of the recession. We didn't subtract enough, if you will, for a lot of the loss of small businesses that were underrepresented in our survey. That was un-quality. But we had two quality improvements that also were a cause. One was that we had better data from the economic censuses of what was going on in the non-covered day care sector. We found out that the employment of the non-covered day care sector was much less than we had originally estimated, so we subtracted 60,000 jobs based on that. Another stemmed from our investing millions of dollars in improving the unemployment insurance data over the years. The improvements are starting to come through, and many of our large reporters, especially payroll reporting firms -- those who do payrolls for other firms -- caught on to the fact that the data they had provided us were not correct. They were counting paychecks rather than people, and transactions. They made corrections in the first quarter of 1991, and managed to subtract 300,000 people who were never there. Now, what happened was that the result of this previous un-quality -- or shift of quality, if you will -- contributed mightily to a fairly large gross domestic product revision, which our friends in the Government Accounting Office are now investigating. It contributed to an upward revision in our estimates of the productivity of the economy. And it contributed to revisions in the forecasts, particularly as it translated back into many of the states. California saw downward revisions in the forecasts of the amount of revenue that the state could expect to receive. Unquality contributed to their overestimates of the amount of revenue they were going to get, and an underestimate of their budget deficits.

Possible solutions? Certainly better methods, we're working on that. Quarterly benchmarks -- should we revise our series on a quarterly basis? I'll bet there are those out there who will say, "no, don't do that, you've got too many revisions already." Or maybe we can wedge back our higher revisions. Now, that's a possibility, too. We don't have to show the world our revisions. We can just move it forward or move it back in time and not do that. But the fact of the matter is that these kinds of issues-- understanding what the data are used for, the needs of our customers, whether the customers are satisfied with it, and what the impacts of un-quality are-- are the things we've got to get on with if we are to work in a total quality environment. That's where we're going.

MS. HAMBY: Thank you Tom. When I work with different federal agencies in different stages of implementing total quality, one of the concepts that we teach and educate them in up front is this whole concept Tom was talking about, the cost of poor quality or un-quality. The way I like to phrase it is, all the things that we have to do -- all the time and all the money and all the staff power that we spend -- sweeping up after ourself when something goes wrong. And Tom has just given us a vivid illustration of just that.

Our next panelist is a colleague of mine from the IRS, Fritz Scheuren, where he is the Director of the Statistics of Income. He has held positions also with the Office of Economic Opportunity, and the Social Security Administration. He is a member of a humongously wide variety of professional, statistical, and scientific societies, has a Ph.D. from George Washington University, and goes back to his home turf and teaches statistics at GW as well. Fritz.

DR. SCHEUREN: Thank you. I also actually teach here at the USDA Graduate School. One of the principles of adult learning, as you know, is to find a teachable moment for people. And one of the things that happens in this quality process is that, for an organization, you also have to have a teachable moment. For IRS, that required that we take a considerable beating in 1985, when our computer system and some of our internal management systems really failed us miserably. If you filed in the Philadelphia Service Center, you probably recall having some personal experiences with that. In fact, if you filed anywhere, you may have had experiences with that.

We often talk about minding our Ps and Qs. I'd like to talk about Bs and Ps instead. (Also, I think I need to add an extra B for **Dr. Bryant**, if I might). The first B, at least in this context, is this notion of having a **beginning** or a **beating** -- a beating and then you have a beginning -- and the realization that you need to change. That's very important.

One of the things that I think is used at the Census Bureau that Dr. Bryant might have mentioned is a movie that Suellen Hamby showed to us on paradigms, by Joel Barker. It's a wonderful movie, and I recommend that you look it if you really want to get a little bit different sense of "change." The only way you're really going to understand what we're talking about with TQM is actually getting your hands into it. If you haven't done that, you probably will go away still puzzled by the deceptively simple, or common sense, nature of this when it's really not at all easy. Not at all easy.

Anyway, let me mention some of the other Bs quickly, in order to get on, because I don't have very much time.

One of the other Bs is **borrowing**. My handout is partly based on this principle; by now you should have it -- a single piece of paper that's printed on both sides. On one side there is a repeat of what Dr. Bryant gave you earlier (Figure A); it shows you Census' strategic approach

CENSUS BUREAU

Strategic Goals

Figure A

Goal	Description	Target Area
1. Meet or Exceed Customer Expectations	*Make customer satisfaction number one in setting Census Bureau priorities*	Communicate continuously with internal and external customers and obtain feedback from them. Improve the design, quality, and timeliness of products and services provided to internal and external customers. Develop measures of customer satisfaction for all major censuses and surveys. Continue to be a leader among world statistical agencies.
2. Improve Product Line to Meet Customer Needs	*Seek to serve new customers with a better product mix.*	Develop new partnerships to better serve Federal agency needs. Develop new products. Drop obsolete products.
3. Recognize and Value Respondents and Other Data Suppliers	*Emphasize key role of individual respondents/ organizations who participate in Census Bureau censuses and surveys.*	Form alliances with major suppliers of information. Implement systematic programs to improve survey and questionnaire design and effectiveness. Pursue least burdensome means of collecting necessary information. Protect confidentiality and reinforce the public's perception of that reality. Monitor and react to respondent's concerns about data collection and handling.
4. Enhance Employee Career Environment	*Make the Census Bureau a more desirable and fulfilling place to work.*	Establish career tracks that encourage employees at the Census Bureau to move across divisions and directorates. Encourage individual development to address Census Bureau needs through training, education, communication, and special assignments.
5. Automate Effectively	*Enhance the utility of automation technology. organizations who*	Automate to satisfy customer needs. Capitalize on and improve the Census Bureau's hardware, software, and ADP staff resources. Incorporate effective use of automation in all phases of our work.
6. Improve Administrative Systems and Management	*Cut red tape, speed decisions-making process, and improve customer service*	Develop and maintain a support infrastructure to provide more effectively and timely administrative services and information to Census Bureau operations and customers. Speed approval and decision-making process by decentralizing administrative authority and responsibility. In recruiting staff: • capitalize on the improved climate for public service. • demonstrate through diversity of our present work force that the Census Bureau is an organization that offers varied and worthwhile employment opportunities.
7. Increase Research Capabilities and Relevance of Research Results	*Create an environment hospitable to invention, innovation, and sharing of results.*	Increase research to update demographic/economic concepts, measurement methods, collection and processing technologies, and estimation and analysis of data. Incorporate in Census Bureau practice relevant innovations from the academic and commercial sectors. Encourage internal invention or developments of improved methods when no external resources exist. Form interdivisional teams to increase relevance and implementation of research.
8. Provide an Integrated, International Perspective for Statistics and Analysis	*Provide our customers an international frame of reference to expand understanding of U.S. statistics and statistics of other nations.*	Become a provider of international information to the American public. Refocus Census Bureau programs to meet emerging international needs of the 1990's, for areas such as Northern American Free Trade area and Eastern Europe.
9. Improve the Decennial and Quinquennial Censuses	*Adapt the Population and Housing, Economic, Governments and Agriculture Censuses to changing demographic and economic conditions.*	Research and develop design changes and innovative ways to improve coverage, response, methodology, processing, and timeliness and quality of products to increase customer satisfaction. Open channels to allow (1) broad external participation in the planning process so that major changes are given full consideration and (2) free and effective exchanges of information among those affected by these programs.
10. Consolidate Headquarters Employees in a Modern Facility	*Improve the physical working environment for Census Bureau headquarters staff.*	Develop and begin implementation of a long-term strategy to attract Departmental, Congressional, and other external support for new or renovated Census Bureau buildings. Ease logistical inconveniences and achieve balance among space, equipment, and personnel. Strengthen security environment.

to planning their quality improvement process. We saw this last summer, and we borrowed it. On the other side you'll see what we're doing with it at IRS (Figure B).

There are 10 principles in the Census proposal -- we had 10 plus one. We added a continuous improvement process -- which I'm sure is part of the Census process, too -- as an eleventh, i.e., to replan what we're doing. And, of course, I'm talking primarily about the statistical part of the Internal Revenue Service, not the whole IRS. I think it's in that context that most of you would have an interest in what we're doing.

Beyond borrowing, there's something called benchmarking. Benchmarking is not what Tom Plewes just talked about, not in this world of TQO or TQM. Benchmarking is a very thorough analysis of someone else's system -- and not a whole system; maybe just a little tiny piece of it, like how you open the mail, for example, which is a big deal for us, especially in April -- so that you can really get better at it. It's a very intense thing. Actually, you have to take some training in benchmarking. You put together a team similar to the Quality Improvement Process Teams mentioned earlier. BEA has the best name, calling their teams Work Improvement Teams (WIT). I'm going to borrow their acronym.

One of the problems with quality is that people think it's separate from the work; that's not true. Dr. Bryant talked to you about the notion of combining the strategic business planning process with the quality process. That was an essential step, and we have not done that well enough at IRS yet. We still have "Action 61" to worry about at IRS. Suellen talked about those. ("Action 61" is about how you answer the mail, by the way.)

After a period of internally looking at your processes, borrowing other people's ideas and formally benchmarking, you really start to get better! Trust me. But, also, you get harder on yourself as you go down this road. And you can get discouraged. In fact, generally speaking, there are always some people getting discouraged; hence, the notion of another B, which is to "catch your breath." That's going on all the time in this process. And it's hard to manage that. Many of you will be skeptics tomorrow and skeptics today, but some of you will be believers today and skeptics tomorrow, and you need to come back to being willing to take action. That's really difficult. I assure you that it's something that I deal with, personally, and I know that the organization deals with it, as well.

It is true that if you put in measures, you can get better and better. Anyway, you know when you're getting better and better. Eventually that gives you the kind of "yes, we're going the right way" reassurance that gets you moving from the Bs to the Ps.

One of the Ps is planning. That's why you have this handout I've provided. Good planning is really essential. The most important P, though, is people. People, people, people, people.

One other P is points or principles. Tom gave you nine that he is using at BLS. Deming has 14 -- you've heard Deming's name mentioned here. Deming is a truly great man, God bless him. He's among our most distinguished local residents here in Washington.

Anyway, the most important of Deming's 14 points may be his last, constancy of purpose. And constancy of focus-- but purpose. If you have that, and pay attention, then every mistake you make turns into a learning opportunity which can make you better.

Let me shift, finally, to talk about what you about Figure B, and tell you what our 10 + 1 strategic TQO goals are.

- The first two deal with customers, products, and services -- new products, better services, new services. I will come back to that later in the context of the world that you're in as people who do projections.

- The next two are about employees, including how to communicate. Communication is very important. One of the problems in government -- and it's really troubling me right now, particularly with my suppliers in the service centers -- is that communication systems are also command and control systems. Anyway, that's the way we treat them in government. The mail is controlled through the system. Somebody tells you how to answer the letter. That's really troubling, that "aliasing," to use the statistical term, of those two ideas -- the notion of communication linked to the notion of command and control.

  I'd like to emphasize the listening side of communication. Listening is a very difficult skill to acquire. Remember the comment that Dr. Bryant made about being arrogant -- that is one of the things that you start out with -- at least I did -- in this process. You think you know more than you do. You have to learn what you don't know. Then you end up learning that you really don't know very much about what you need to know in order to get there. It's essential that you are able to see beyond your own existing expertise, to push yourself down low enough in this process, so you can listen to the customer -- listen naively, as Tom Peters (another P), would say. Listen naively.

  That's a very profound idea, because everyone in this room is a really knowledgeable expert. A really knowledgeable expert already knows the answer before you ask the question. Yet, we have to listen, really listen. The listening side is the most important side of the communication system.

- Let me go back to the Figure and talk about the remaining objectives. The next three of these focus on process improvement. There's a phrase we're using, "lean production," that comes from a book called The Machine That Changed the World. It's a book written by three professors at MIT, James Womack, Daniel Jones and Daniel Roos. (It costs $11 at Olsson's, paperback.) Buy it. Read it. The book is about the automobile industry, where, as you know, some would say we're getting killed by Japanese competition. Be sure to read about a man named Taiichi Ohno from Toyota, who is perhaps at least Deming's equal.

  I'm getting off track, again, but you need to know, if you don't already, that most of the key

SOI 1993 TQO PLAN

Figure B

Vital Issue/Description	***Key Activities***
1. *Expand Customer Products - Develop new and improve existing products to delight the customer. Build a lasting relationship with the customer.*	☺ *Create at least one new public use file* ☺ *Develop system to produce data estimates on demand* ☺ *CDN access for OTA* ☺ *Improve format of existing data* ☺ *Develop meta-data systems (information about the data)* ☺ *Expand electronic bulletin board* ☺ *Pursue desk-top publishing capabilities*
2. *Enhance Customer Service - Provide greater access to our data in a more timely and flexible manner.*	☺ *Develop a system to evaluate customer satisfaction* ☺ *Pilot direct mailing of SOI publications* ☺ *Develop an electronic data contact list* ☺ *Expand SOI's mailing list*
3. *Enhance Employee Career Environment - Make Statistics of Income a more desirable, fulfilling and productive place to work.*	☺ *Create an environment where employees can spend 20% of their time working on assignments they choose* ☺ *Establish career tracks that encourage employees to move between Sections and Branches* ☺ *Strengthen performance through improved coaching* ☺ *Strengthen employee evaluations through structured implementation of n elements and standards*
4. *Improve Communication - Build a communication system that facilitates free exchange of information within SOI and between SOI and its customers and suppliers.*	☺ *Improve relationships with customers using systematic communication systems* ☺ *Enhance networking arrangements with suppliers* ☺ *Develop a general on-line project communication system* ☺ *Improve dialogue between managers and employees through use of job priority statements* ☺ *Improve communication within the N.O. via an Idea Bank*
5. *Adopt Lean Production Techniques - Develop a process which will maintain a steady work-flow as well as the ability to accept changes throughout the project's life cycle with a constant editor work-force.*	☺ *Customers order returns on basis of complexity and editors become specialized* ☺ *Implement projects in stages* ☺ *Conduct centralized and local "structured walk-throughs"* ☺ *Develop a SAT process which involves all editors*
6. *Optimize Edit Systems - Create edit systems that make use of the 'best of class' ideas from all SOI systems, are easy to use and provide the best product to the customer.*	☺ *Create a shared electronic library of program modules* ☺ *Develop unique coding lists, standardizing key usage* ☺ *Create common menu structures for on-line edit systems* ☺ *Create guidelines for the presentation of error messages* ☺ *Employ an outside consultant to advise on optimization and modularizati*
7. *Prevent Rework - Reduce the amount of rework or corrections/revisions needed at each processing stage of a project.*	☺ *Develop guidelines and improvements for SAT of on-line systems* ☺ *Select and expand use of CASE tools* ☺ *Explore ways to use existing programming in other projects* ☺ *Improve longitudinal and model-based testing of data* ☺ *Continue current planned initiatives*
8. *Manage Quality - Integrate and improve existing quality initiatives.*	☺ *Implement a double-edit quality review system for PRISM* ☺ *Complete implementation of QUIC Charts on all SOI studies* ☺ *Complete contract to review quality systems* ☺ *Involve customers and suppliers in development of measures*
9. *Integrate Return Inventory Management Systems - Implement automated return control systems for all SOI studies computer selected at MCC.*	☺ *Provide service centers with tools to order returns, based on complexity, "just in time"* ☺ *Produce summary level sample control reports* ☺ *Implement measurements of how long SOI holds returns*
10. *Better Manage Resources - Develop a system which will place management of resources in the hands of project teams.*	☺ *Train employees and managers to manage resources* ☺ *Build a Travel Tracking System* ☺ *Enhance system to manage overtime, travel, award and staffing budgets* ☺ *Develop a plan to manage outside contractors*
11. *Improve the Planning Process - Insure the use of a structured planning process including tools.*	☺ *Employ baseline assessment of existing quality efforts* ☺ *Benchmark quality efforts with those of other organizations* ☺ *Integrate the IRS/NTEU Quality Council into TQO* ☺ *Monitor and improve the 1993 TQO plan implementation* ☺ *Develop new and better measures of our quality progress*

figures in quality are Japanese. We have to learn from them, not just from Americans like Deming or Juran. One of the nice things about the Japanese is that they tend to be shorter than Americans, and that means that you have to bend down a little bit to listen. And that's important. That posture is important in this process. Anyway, read this book: The Machine That Changed the World; it's a great book.

o Back to the handout, again. The next three objectives are all linked to better Quality Measurement. These may be among our most challenging goals. To illustrate my point, let me mention a fine paper about service quality measurement, by Blan Godfrey, which was given two years (1990) ago at the American Statistical Association (ASA) meetings. We're in the service business in government. Service quality measurement is in its infancy. We have an enormous amount of work to do -- again, Dr. Bryant made this point -- in order to find the right measures.

  I'd love to have you ask some questions about what good measures look like. I have maybe two examples -- I should have hundreds -- but I have maybe one or two examples of where we may have found the right measure, and we're in search for the other 99. (It's not like the story in the Bible about the 99 that you have and the one you don't have. It's the other way around.)

o After the measurement goals is our last objective --to improve the planning process, itself. That's the 11th, or "lucky," step. If we didn't get it right, we can try, try again.

Earlier, I said I'd talk about a couple of things we're doing for you. Let me do that now. I'll give some examples of steps we've taken to improve access to our information, produce some great new data series and become generally more responsive.

o Access -- We have just recently established an electronic bulletin board. (You can get access to that by dialing (202) 874-9574.) During the filing season there will be weekly updates about what's going on at the IRS in terms of returns received. Eventually, all our publications will be on-line. So far, the Bulletin Board has primarily been focused on internal customers. It could, already though, have information you might want to know if you're a user of tax data.

o New Series -- Another thing that we're doing -- and we'll be putting out our first version this coming spring -- is what we call early economic estimates. They will be tax return based projections of income distribution statistics and will come out about a year earlier than our final estimates.

  As you know, if you use the Current Population Survey, in the spring and early summer the annual income estimates that come out of the March supplement become available. Our tax return income projections will be a companion series for you to use, along with the survey data.

o Greater Responsiveness -- The last thing I want to mention is a notion that we're really just playing with, but very seriously playing with -- the notion of projections on demand. Let me tell you what that means.

  As some of you may know, we at IRS have an ongoing data collection process in which we're compiling taxpayer information all the time, year-round, on different things. We have 60-plus programs that we're running -- corporate programs and many kinds of other kinds of business programs, individual programs, special programs on excise taxes, and international programs of various sorts. Because of the lag in filing with us, these programs are not as timely as our customers -- including some of you -- would like.

  To address this timeliness issue, we have set the goal of structuring our work to make earlier and earlier estimates. For this projection on demand strategy to work, we have begun to think of ways to reconceptualize how we process returns and how we estimate from them.

Eventually, if someone calls up and says, I want you to project to the end of a particular period, we hope to be able to do so quickly. Even if we fail, thinking about our work in this new way turns out to be a wonderful idea, because it retools the whole way we approach our processing and could change much of what we now do to be responsive to our customers. Thanks.

MS. HAMBY: Thank you, Fritz. Our last panelist, of course, has already been introduced, Dr. Bryant. I guess what I would ask you, Dr. Bryant, is for any additional comments you may have, now that you have heard the comments of the previous panelists--all of whom have plagiarized wildly from the Census Bureau.

DR. BRYANT: Well, as a matter of fact I was going to start out with a plagiarized comment, or as Fritz put it in terms of B's--borrowed. Chuck Wade at the Census Bureau says, "borrow shamelessly." He puts a little extra emphasis on it, which sort of reminds me of the Tom Lehrer song a few years back, you know, plagiarize, plagiarize, remember why the good Lord made your eyes, plagiarize, but please be calling it research. And I think, as Chuck says, we should be borrowing shamelessly from each other.

The sorts of things they said prompt me to talk about several tough things in CQM. One of them is when you can't satisfy the customer. Now, Mayor Dinkins of New York wanted 8 million people in New York in the 1990 census. We only counted 7.3. Now, we may be slightly under, but Mayor Dinkins is wildly and unrealistically over and claimed he had no forecast, kind of thing. Well, we need to get the customer to buy in to some of the things we do. We need to work with the customers, but there are 39,189 units of local government, or there were in 1990, and obviously we can't negotiate with each one of them. So there is this problem of when the customer's expectations are unrealistic and you're never going to satisfy them. And I don't think any of us have the answers for that.

The other thing is that benchmarking -- and you had it as one of your B's -- is very, very hard to do, particularly in the Federal Government. I mean, we can see how General Motors could benchmark against Toyota or someone with the same sort of a product/service line. But most of us in the Federal Government have rather unique products or services. And actually, when it comes to the census, we can't benchmark in the United States. We're the only one who does one. But we can go out and talk, as we've just had a two-day conference out at the Census Bureau with the United Kingdom and Canada, because they've both had the same sort of falloff in public response that we saw. And although they are counting fewer people there is a certain learning and interchange we can do. So you have to be a little innovative in finding where to benchmark, or you may be able to benchmark against some piece of something that's in the private sector.

The other thing that sometimes is difficult is when you have a really great TQM project but it doesn't come out quite as well as expected. Our first really large-scale quality management project was in our economic division, and this was to improve the design process of the 500 different questionnaires used in the economic surveys. And they set this up with all sorts of measurement, that, you know, one measure was to be how many times you had to recycle before you got that questionnaire off to OMB. They had the advantage of having done a customer study on recordkeeping practices. They were trying to redesign the questionnaires so that they matched the way the customers, the businesses, kept their books. Well, that whole project, actually, I think was a tremendous success, and yet there were some who said, yeah, but we didn't quite do it as well as we expected. I mean, other things hit them. Certain equipment didn't arrive in time and they had to work their way around it. Measurement kind of got loss in the process as some things slid behind, and so the benchmarking measurement had to be the thing that went. There were some on the team that said, well, this didn't turn out quite the way we thought it would, but we do have the 500 questionnaires redesigned, off to OMB, most of them being printed, and they will get mailed out on time. You have to say, well, what would that project have been like to redesign all those questionnaires if you hadn't used TQM? TQM doesn't always work perfectly.

MS. HAMBY: Thank you, Dr. Bryant. I'd like to pick up on two points that you just made. The first one is that we can't always satisfy our customers. And probably more so in the public sector than in the private sector, we have customers with conflicting requirements. I mean, you think about the spotted owl versus the loggers, and others that you read about in the Post from day to day. In other cases, there are just simply legislative or regulatory bars that say we can't do it. We either can't find an extra how many hundreds of thousands of people, or if you want to deduct the cost of the upkeep of your French poodle on your tax return, we're going to tell you that you can't do that either, even though you may say that's one of your requirements. Tom Peters would say, what do you do in that case? Well, total quality management is not a magic wand that's going to make all your tough strategic management issues go away. You listen. You explain. You listen to your customers. You explain your constraints. And in some cases you either may have to look for alternatives, or if none are available, then target the customers with the greatest impact, or the ones whom you know you can legitimately and rightfully meet their requirements.

The other comment you made about the great TQM project that doesn't come out well is in many cases a function of how management or the quality council assigns those projects. And Fritz will commiserate with me. I remember in our zeal once we first started at IRS after we'd all been trained by Dr. Juran in the quality improvement process -- in the return to the processing area, where Fritz and I worked. We said, "well, gee, what will we tackle?" And somebody, in their infinite wisdom, must have said, "I know let's put a quality improvement team on improving accounts receivable at the IRS." Think a little bit about the potential size of the scope of accounts receivable at the IRS. We formed a team. We gave them a room in the basement, and about six years later, we still see them every now and then as we're on our way to the credit union, but that's about it. The projects have to be manageable. They have to be bite-size. They have to be related to the business of the organization. But most of all, they have to be something that a team can, with training, tackle in six to nine months and show some results.

I'd like to ask our panelists, before we open up to questions from the floor, if they have any comments. Having heard their colleagues, if they have any echoes of, yeah, we had that problem, too, or gee, how did you do that? Any questions of one another, or any comments?

DR. SCHEUREN: I want to mention -- and I think you know this, Suellen, and you might want to elaborate, there is a benchmarking service that you can buy into. You really want to benchmark a little tiny piece of the business. There's no way you could benchmark what IRS does, or what the Census Bureau does, or any large complex structure. What you want to do is benchmark the sub-pieces of it. And you need to go to somebody who is good at that, like Xerox went to L.L. Bean to learn how to handle orders better. [L.L. Bean -- I'm sure you know the store in Maine. Any of you customers of L.L. Bean here? Thank you. I'm a New Englander too. I won't talk about Xerox because there's a conflict of interest there, since I do a lot of Xeroxing.] Do you want to elaborate on that?

MS. HAMBY: A couple of things with respect to the benchmarking is, as the FQI people work with customers in different government agencies, the first thing that we're asked is, well, show us another organization that has the same mission as we do and is the same kind of business, and show us how they have implemented total quality. Well, I've worked with the Travel and Tourism people at the Department of Commerce, and, aside from American Express, I can't find anybody else that does exactly what they do. And I worked with a group of lawyers recently who said, "Well, show me another firm in the government with lawyers doing our kind of law and are practicing total quality management." At some point you need to recognize that, it will be a leap of faith for some, that what you're taking a look at are management processes. And these may not be too different from the statistical forecasting area to the tax collection area to the analysis of information that the security agencies do as well. But going out and taking a look at some organization that has at least something similar in terms of paper flow or application processes, and seeing how they do and how they do it well -- whether it's in the private sector or the public sector -- can give you all sorts of valuable information.

One of the things that we did recently with the Bureau of Economic Analysis is take them down to Warner Robbins, one of the improvement prototype award winners. Well, it's a military base. So what do we want to hear from a military base? Well, I think that the people that went down there were thrilled when they found out that the people on the teams and who had looked at ways to make things better and of communication and of recognizing employees, had some wonderfully good ideas that the Bureau of Economic Analysis could pick up. For those reasons, we're also going to the IRS out in San Francisco, and to the Department of Labor, the Wage and Hour Division in San Francisco. So there's a lot of information available out there, not only for benchmarking, but also to learn from your colleagues who may be only a step or two ahead of yourselves.

What I'd like to do is open up the floor for questions. I would ask that if you have a question step up to the microphone and give your name and your organization.

MR. TURNER: Thank you. My name is John Turner and I work at the Bureau of Economic Analysis. Specifically, I'm concerned about interagency data exchange. Mr. Knox addressed this question somewhat in his comments, but I would like to ask the question of all the panelists and hopefully they might be able to elaborate a little bit about it. As you stated, in our work at BEA we use data obtained from various other government agencies, and in turn, other government agencies use our products. Sometimes this data sharing relationship can become, I'll say, cumbersome, stressful. But given that there seems to be a growing client-server relationship among agencies, and with the advent of TQM, do you foresee increasing cooperation among statistical agencies, or do you think that the relationship we have now is the best that we can expect?

MS. HAMBY: Okay, thank you. Who wants to handle that?

DR. SCHEUREN: We're in a hell of a lot of trouble if this is the best we can expect, I'll tell you that.

MR. KNOX: I touched a little bit on that in my remarks, that I'm seeing increasing cooperation. And it's a different kind of cooperation than I've seen in the past. I would look, when the regional program gets further along in the process, to try and construct some interagency teams. We already have a person that visits IRS now and then, and comes back and provides, to me, a big service by explaining what's going on at IRS and some of the forms and how they go together and what we might do in the future. I could see that happening with the 790 and 202 exchanges between BLS and the Bureau. There will always be some constraints about confidentiality and whatever, but I would like to see some interagency teams working on issues of exchange.

MS. HAMBY: Tom?

MR. PLEWES: I would like to address it both as a customer of the Census Bureau, and as a provider of information to BEA. As a customer of the Census Bureau, I have clearly noted that there has been an improvement in the willingness of the Census Bureau to communicate with its customers to identify needs and to work toward a fuller interchange of the information.

We have come to agreement over the past couple of years with Dr. Bryant on a protected exchange of certain of the microdata from the Current Population Survey -- not all of it, not the most confidential, but some of the more useful things that we needed to do some of our work.

So there is that. If an agency wants to go that mile, the TQM process works. That emphasis is clearly at the Census Bureau.

Looking at the BEA as a customer, we believe that we have tried -- again, over the past couple of years -- to get more close to the kind of information that they need and when they need it. And indeed, on the transmittal of some of the more detailed data, we have sped it by two to three months.

There's lots more things that we should be doing in terms of turning around the quality and the corrections much faster, but I think we're moving on the right track.

Always the issues of confidentiality crop up when you're talking about data sharing. Some day perhaps we will get some legislation which allows us to do some reasonable data sharing between the agencies to avoid the burden on our people and to have the higher quality data.

That's not there yet. We're working, I think, within total quality management, to make some of those things happen.

DR. BRYANT: Yes, I think we will see more interagency teams. Actually, Fritz and I were talking earlier that we need one right now to break down a few misunderstandings between IRS and the Census Bureau. And the advantage of teams is that a few more people and few more perspectives are involved. You don't get into some one-on-one personality situation, or something like that.

I think we have a beautiful model that really started before TQM but was done with TQM principles, and driven by necessity. That is the Bureau of Labor Statistics and the Census Bureau have had a project team working on the redesign of a survey instrument for the current population survey. It's many, many years since that survey instrument was redesigned, and it's being redesigned both for content and also to move it to computer-assisted interviewing. We can't say that we started it as a TQM process action team because it really was going before either of us heard about TQM. But it certainly has operated like a fairly large TQM team with some separate subcommittees on it, some of which are concerned with question wordings, and others which are concerned with how to get the thing up on the computer.

DR. SCHEUREN: Let me make a couple of observations, too. One of the things you'll see in the Census plan, under category number 3, is "form alliances with suppliers of information." That word "alliances" is a nice word. It's ambiguous. It's new. It has a lot of room for the kind of change that we need to make in the systems, as we are customers and suppliers of each other. I made a comment about Deming, and Dr. Bryant did too. Deming's fourth point is about managing suppliers.

There's a wonderful article -- I don't know if any of you are reading Quality Progress, which is a monthly magazine on quality -- it's something you ought to get. It's about $20 a year and it's worth looking at if you're a manager.

Anyway, there's a nice write-up in Quality Progress about how the supplier needs to understand the philosophy of the customer. This is not your "I'm-writing-a-term-paper-and-need-a-statistic" type of customer. This is the kind of customer that you have a long-term relationship with, like a BLS-Census relationship, or like the relationships that we in the Statistics of Income program have with the Congressional committees responsible for tax legislation.

Anyway, one of the things we've done the best at IRS -- at least our part of IRS -- is to develop strong supplier networks. That's where we've had our biggest gains. We're enormously more productive than we used to be. I haven't talked about that. We're enormously more productive because we've adopted a lot of "lean production" ideas, though we didn't know they went by that name at the time, because we hadn't read the book I mentioned earlier until about a year ago. The Japanese ideas worked for us, those we reinvented and those we simply borrowed as is. Once you start to work with your suppliers in a way that really, really makes a partnership or an alliance -- I like the word alliance better than partnership -- then you're really going to change.

MS. HAMBY: Any other questions? Yes, sir? I'll take the fellow over here and the fellow in the blue, just because I think one will get to the microphone faster. Go ahead. Your name and your organization, please.

MR. MACEK: Paul Macek, Bureau of Economic Analysis. I'm going to make two obvious statements before I ask my question. Number one, TQM requires funding. Number two, funding for government forecasting agencies is subject to exogenous factors, the worst of which we're about to see in November, the general election. My question is, to what extent can TQM teams influence the budget of their agency? The funding or the reallocation of resources within their agency, and also, how much money they will get from Congress.

MS. HAMBY: Okay. Who wants to address that?

DR. BRYANT: We have no line item in our budget for TQM. We did spend some money, you know, substantial sums, when we got started in the training process. We now have in our administrative area about an eight-person Census Quality Management staff that's funded in our regular administration. But really, the rest of it, we really are expecting to improve products and processes enough that it's going to pay for itself. So we have not asked for any funding for it.

MS. HAMBY: Any other experiences? Tom?

MR. PLEWES: Again, at the Bureau level we have a very small staff, about five people. And in each of the three major offices that are involved in TQM, we have one full-time facilitator. But most of it has worked within the organization, so there is, again, very little in the way you can identify a particular cost of TQM.

The question is, what happens now if these forecasts of tremendous budget cuts for these agencies, or at least stringencies, in this next year come about, what are you going to do about it? Are you going to drop TQM? The answer is no. I think we've got to push TQM even harder. It seems to me that having adopted that as a philosophy, that is the only way now that we can start to think about the kind of productivity, quick turnaround productivity, goals and gains that we need to operate in a more constrained environment.

And certainly as the teams come on with their costs and so forth, as they come on line saying we can make this improvement but it's going to cost this amount of money, I have to make those investments out of my more scarce resources so I can save more down the line.

And so we look at this as an absolute necessity now in a time of a more stringent budgets, and we fund those quality improvement programs before we fund many of the other kinds of things that we have been funding in the past.

DR. BRYANT: One of the slogans of TQM is: "Do the right thing right the first time." I think this is where we hope to save the money--doing the thing right the first time and not reworking.

MS. HAMBY: I'd like to jump in and add my two cents worth. There are really a couple of issues involved in this question. One is the up-front training that is required, whether it's awareness or team training in the analytical tools and structured problem-solving approach. And there is not going to be a cornucopia that's granted by Congress to any agency to do TQM training, whether it's awareness training or tools training. What's going to happen is the managers are going to make some tough decisions, where maybe last year we weren't going to have a management conference with this sort of theme, what we're going to do is take those dollars and redirect them to perhaps some awareness sessions for our employees. I don't know of any organization that has been blessed with extra allocation for training. I know a number, including that at the IRS, that said, I've got to make some tough decisions on my training priorities this year, and I've declared that training in TQM principles and practices, and some of the team training, comes before some other things that I thought I might have done when I was making this plan up this year.

I think the most difficult resource issue to deal with is just the issue of time. The time that the people on the teams are away from their desks, working on their problem-solving processes, and what happens to the in-boxes back home? The managers are going to have to find a way to deal, again, with priorities. And I remember one of the organizations I work with said, you know, everything comes in the door, but nothing ever falls off the table. I think that's very real in our environment and it's going to be something that managers need to deal with, and that is, how do we make sure that employees who volunteer their time on these teams to give us these good ideas to help us save the money are not penalized by when they get back to their desks, finding that nobody has done anything extra, and they've got to work a couple of hours late that evening or that weekend in order to catch up.

I don't know whether any of our panelists have had any solutions or have had any way of addressing this whole issue of time, but it is a very, very difficult thing for managers to grapple with.

There was another question over here. Sir?

MR. WALDO: My name is Dan Waldo. I'm from the Health Care Financing Administration. Our agency has recently begun a TQM process. I hate to sound a little Cassandra-like but, given the government's reputation for being on the blunt edge of management technology, I noticed in Newsweek magazine a couple of weeks ago an article entitled "Is TQM on the Way Out"? It seems that there are a number of private companies that started on the road and decided it was the wrong road for them to take. I'll ask you a question with two parts. One, is there a critical mass within an agency that gets one on a TQM road and keeps you there? And if so, how do we achieve that? And second, are there common false images or false notions of success that would lead an agency to think that they were embracing TQM when in fact they aren't? And if so, how do you pierce those things and get people to change?

MS. HAMBY: Let me make some introductory remarks and then I'll ask the panelists to join in. I referred to that Newsweek article when I opened up by saying that there are some that predict the demise of total quality. And I think that article also hit right on the head the fact that in some cases we were mistaking activity for results. There is a lot up-front work in planning that's involved, but at some point you need to say, okay, we've got our plan and let's start getting teams and let's start implementing some of these recommendations and get some results out of this process. So one of the common false images is if you've got a lot of people meeting, and they're using the analytical tools, and they've got a good plan, then by definition, quality is going to improve in the organization. You still need to look at what results -- are we commissioning them to work on the right problems? Are we having them take a look at really business-oriented things, rather than just some of the more recurring issues? We really need to get them looking at how do we do the work, how can we improve the organization.

The critical mass in the agency -- some will tell you it's top level leadership, and I'm the first one to say that if you have that, that's fine. You don't need it. Whoever you're working with, as long as they understand the goals and support them, I think can provide the needed influence and support. We found at IRS -- and Fritz, correct me if I'm wrong -- but my impression was that our move towards total quality in view of the crisis we experienced with the crash at the filing season and the crash of the world as we knew it around our ears back in 1985 was not precipitated by a vision on the part -- and I mean that in terms of enlightenment or a burning bush -- on the part of our then commissioner, but it was people such as Fritz, and such as one of our district directors, and others, who had been doing a lot of learning and understanding. They had gone underground, and they were very quietly championing these things in their organization. So that critical mass is really the people that can in their own district office or regional office or functional area, support it, get it going, and keep it going. I would ask, having expounded, what the reaction is. Do you see TQM as being long-lasting? I know, Tom, you addressed that in your remarks.

MR. PLEWES: My reaction to that is that, yes, TQM could die tomorrow in the Bureau of Labor Statistics, but the next day you'd have to replace it with something that looked just like it. And that's kind of where we've come.

I think that we bought into a focus on customer satisfaction. We've bought into involving employees, and our critical mass was really reached about a year ago when we started involving the bargaining unit employees in a major way and getting the unions to be a full partner in this with us. We hadn't done that before.

But the fact of the matter is that the basic notions of total quality management, customer satisfaction, continuous improvement, and measurement, are so important to anything you do that if you didn't do it through this thing called TQM, you'd do it through something else, if you were a good manager.

DR. BRYANT: I'll sort of say amen to that, but I also will say that it will last as long as it works. You know, the day it doesn't work, it will die.

DR. SCHEUREN: It is so basic, what we're talking about here. A lot of us work in white collar environments, even though we don't wear white shirts anymore. We have not flow-charted our office processes. We just haven't done it. There's lots of waste in those processes. If we would simply flow chart them, we might find a lot of resources to do other things with.

Another aspect, Dan, that I mentioned before is measurement. Juran says that the language of upper management is money. If you can show how you've saved money -- not necessarily in the same year, but in a reasonable amount of time -- you will, in fact, cement the TQO process. That's what Barbara Bryant said, "if it works." That's the measure. It will have to work. And that's a good measure.

You can fool yourself a long time, though. I admit that, and those of you who are economists understand the notion of bubbles. And that's possible too.

I do not believe that we're going to lose the TQO effort. We may continue to change the name, as Tom just said. Its demise is going to get predicted over and over again. I hope we do change the name, because one of the other quotes from Juran is, we need to fool the immune system of the organization.

It's nice that Barbara has titled it Census Quality Management. Incidentally, I'd even go further and say we need to get rid of the word "quality." That's why I like the "WIT" idea. That's a wonderful name. We need to focus on the work. It's just doing the work differently, being systematic. As the Japanese would say, "Ask the five why's." When something doesn't go right -- ask "Why?" and when you get an answer, ask why again and again.

Let me tell you a Honda story; it's a real simple story: When Honda has a supplier and they get a bad part from the supplier, they send the part back to the supplier. What do they expect the supplier to do? They expect the supplier to pay the postage and find out what went wrong, by asking the five why's.

You've heard the other story, the IBM story about the chip manufacturer, the Japanese manufacturer? IBM did not have a zero defects standard; they had a 98-99% standard. So, the chip manufacturer was sending them two boxes -- a big box and a little box. They found eventually that the chips in the little box never worked. So, they went back to the supplier and they asked him, and the supplier said, "Well you wanted 98 chips that worked and two chips that didn't work. We could give you 100 chips that work every time." And they did and so can we, or something similar in our own worlds.

MS. HAMBY: Well, I will say, too, that the Federal Quality Institute has just launched an action planning session with a number of people from HCFA where a cross-section of employees, 25 people at a crack, take the goals and the objectives and turn it into some really meaty action plans for how are we going to improve. And that is involvement all the way through every level of management, and including the employees. We're really pleased to see that happening at HCFA.

Do we have any other questions from the audience? Yes, sir.

MR. TRAXLER: Herbert Traxler, Bureau of Health Professions, U.S. Public Health Service. I'm lower to middle management-- at a working-level middle management.

MS. HAMBY: We won't take umbrage at that.

MR. TRAXLER: I didn't mean it that way. Our Bureau has been involved in TQM for the past two or three years, pretty heavily. They started one way -- and the resources and costs of TQM can be measured, to a certain extent, because we have about 300 employees who were put through a two-day workshop. So that comes to about 2-FTE years. In terms of initial investment, apart from the cash expenses for a consultant who is still on retainer, who is still holding various sessions and workshops --we have had steering committees, process improvement teams -- so a very heavy investment in expenses and in costs. One effect of this initial investment was that the middle management was at the same stage as the employees. We didn't have any more answers than the questions we were asked. So later on, with the steering committee and through a two-week session at the Executive Management Institute, for instance, I found out a little bit more about TQM.

So that's one way which it probably should not be approached. It's a little bit problematic when everybody is put in at the same ground. And then the managers are there in a management process without the answers. This is just a comment.

Another comment on which I would like the panel to react to is that we are faced with diminished resources and what you said, things don't fall off the table, they are added on. As an example, many of our prime customers -- Congress and the President. They have given us our appropriations and the money. One survey which we did was a survey of the States in terms of their constraints and their resources, a governor's survey. They said one of the things they wanted was additional funds for planning-- which we couldn't give them-- and they also wanted technical assistance. So I was charged with following up and putting on the technical assistance workshops in the States. I have done one and another one is coming. And in TQM, in the customer survey, as to what are their priorities, they said, well the states are not our prime customer really. Now, with the workshop, and identifying them in the initial lettering, the TQM spirit, said, well these are important customers, and now we are offering you technical assistance.

Now, resources are cut while certain expectations have been raised by customers for our services, which in the TQM process, will not be met. How do we meet those expectations when they answer in a national survey that they want technical assistance in terms of modeling and forecasting. You put on a workshop and in the feedback, they say "we want more of that." Having raised those expectations we can't do it because another process we are going through in the Bureau right now is streamlining. So we are streamlining our customers and our processes because of limited resources. So we are streamlining those customers whose expectations we have raised out of the TQM process. These are some of the constraints we are meeting and we are having to face. I would like some reactions to that.

MS. HAMBY: Okay. That's one of the things I was referring to when I said there's really no magic wand. Here we do have issues where expectations may be raised because we've been motivated and enabled to reach out to customers in a fashion in which we hadn't been able to do earlier, and then, because of resource cuts, those services or those contacts are forced to be cut, or at least come under a lower priority. How have those of you on the panel dealt with this issue if you had to at all? Taking candy away from the baby, huh?

DR. SCHEUREN: Oh, good. Yes, there you go. Yes, Americans are the second hardest working people in the world, already, and this is adding to our burden.

The problem is that, when you ask people their views, they are expecting you to take them and do something with them. That always happens. The time horizon for the quality process is not short. This is not a quick fix. So, the sort of paradox of rising expectations exists here, and I think it's an inevitable part of the process.

Clearly, too, you need to be prepared to respond to bad news or middle news. If you do a survey, you don't usually get good news, not at the beginning. Incidentally, we just did one last week; then we worked all weekend in order to compile the results, so we could have them on every employee's desk on Monday morning when they came in. It is a part of the baseline for the very plan that I gave you.

Another part of this is the business about resources. We have not had any new money in my part of IRS for a long time. However, we are doing probably four times the amount of work we were doing for those same resources.

There's lots of room to even do more within that, because lots of people, especially our white collar employees, have discretionary time available. They will tell you they are too busy if you give them an assignment that they don't want to do, but there's actually a lot of discretionary time in some of those jobs.

MS. HAMBY: Tom, do you wish to respond?
MR. PLEWES: I'll address three things--expectations, middle managers, and something I'll call vendor quality.

Expectations -- we purposely did not what the Census Bureau did. We did not bring everybody into a room and talk to them about TQM. What we have been doing is getting people involved in our team efforts, and then intensively training them. We haven't increased expectations of people who aren't involved in the process. Those people were involved in the process, and those processes were looking at, are intensively trained.

The next step has to be, however, to broadcast it more now that we've got some success stories. We have some success stories, so we have people say, okay, volunteer to take a look at your own process. And we're going to get involved in that next.

Middle management, as you mentioned, is a real issue for us. We've you've got a lot going on at the top in terms of leadership and commitment, and a lot going on where the work is being done in teams. The people who feel left out of that process are middle managers. You know, they are the ones to whom the hierarchy has always said, get the job done. Now somebody else is getting the job done, and they are to facilitate that process. And that's a very difficult change to put out.

We have trained middle managers to be the facilitators for teams, so they get involved in this process. And we have a large group of facilitator out there who are our good middle managers, who are doing a wonderful job. They buy into the process, and hopefully that works.

Vendor quality -- it's easy to get a quality product from an organization like the Census Bureau which has a commitment to quality and has its own quality environment. But we are heavily reliant on state governments to provide us information. They don't have the same kind of commitment to quality, nor the same kind of facilities, like the FQI within their states to talk to about quality management. So we have been going out and working with and training the state agencies that we work with in quality.

We brought employees of state governments on to our process action teams. It cost us quite a bit of money to pay for travel, but they are with us on our process action teams. We now have a representative of one of the state governments on our quality council to help us make decisions as to where the system ought to be going. So it's not an easy task to build quality into what you get from your vendors. And when government agencies like ours are so dependent on the kind of quality we get from others, its one of our biggest challenges.

MS. HAMBY: I think our time for the panel is about up. I'd like to thank our panel for helping us translate some these more abstract concepts into some good practical examples of what's going on in each of your bureaus and agencies in quality improvement. Thank you very much.

Mr. LIENESCH: Thank you to our distinguished panel and moderator. That concludes the Morning session.

FFC-91 and FFC-92 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-91 and FFC-92, we conducted surveys of conference participants. Our intent was to get basic demographic statistics on Federal Forecasters. For FFC-91, 50 out of 230 conference attendees who were currently working in forecasting in the Federal government completed the survey form. For FFC-92, 54 out of 255 conference attendees completed the survey form. Because there is no guarantee that these are representative samples, we cannot apply the results to all Federal forecasters, or even to all FFC-91 and FFC-92 registrants. We were pleased that so many people participated in the surveys, and that most of the survey questions were answered by all of the respondents. We also plan to do the survey at next year's conference.

Over 70 percent of the respondents at FFC-91 had a degree in economics. At FFC-92, over 60 percent of the respondents reported degrees in economics. Other fields represented at both conferences were biology, mathematics, operations research, statistics, geography, geology, sociology, psychology, public health, demography, and Spanish.

One-fourth of the FFC-91 and FFC-92 respondents were female. Also, one-fourth of the FFC-91 respondents were managers. Of FFC-92 respondents, 35 percent were managers. Between the two surveys, a notable increase in the participation of managers was observed. On average, the FFC-91 group had 11.2 years of Federal service, while the FFC-92 group had 15 years of Federal service. Both groups had been forecasting for 10 years. Almost half had a Master's degree in the FFC-91 group; 46 percent of the FFC-92 group had a Master's degree. Corresponding figures for those with a PhD were 32 percent and 35 percent, respectively. Of the FFC-91 and FFC-92 respondents, three-quarters published their forecasts. Nearly three-quarters of the FFC-91 group did evaluations of their forecasts, while two-thirds of the FFC-92 group did evaluations of their forecasts.

Most of the respondents include national forecasts in their scope of work. Regional/state forecasts concerned one-third of the FFC-91 group. The proportion was 38 percent for the FFC-92 group. Nearly 20 percent of both groups were concerned with international forecasts.

In terms of primary forecasting techniques, respondents cited a variety of methods, including trend analysis, regression models, time series methods (exponential smoothing and Box-Jenkins), macroeconometric models, demand analysis, dynamic simulation models, input/output models, and judgement.

Among the issues facing Federal forecasters listed by respondents were availability and quality of the data, staff, budget resources, reliability of forecasts, and coordination of forecasts among Federal agencies.

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

FFC-91 CHARACTERISTICS BY OCCUPATION

		Total respondents	Managers	Nonmanagers
Years of gov't service (average)		11.2 yrs	14.8 yrs	10.0 yrs
Distribution:	0-4 years	26	0	34
(percent of total)	5-14 years	44	50	37
	15-24 years	24	50	21
	25+ years	6	0	8
Percent male		74	92	68
Average grade (excl. Executive Service)		GS/GM 12.6	GS/GM 13.8	GS/GM 12.3
Percent GS/GM-13		39	25	43
Ed. highest degree:	Bachelor's	20	8	24
(percent of total)	Master's	48	42	50
	PhD	32	50	26
Years of forecasting (average)		9.6 yrs	12.0 yrs	8.8 yrs
Distribution:	0-4 years	42	33	45
(percent of total)	5-14 years	30	25	32
	15-24 years	18	33	13
	25+ years	10	8	10
Percent whose forecasts published		75	64	78
Percent, forecasts evaluated		71	67	73
of which, percent evaluation published		70	50	77

FFC-91 CHARACTERISTICS BY SEX

		Total respondents	Male	Female
Years of gov't service (average)		11.2 yrs	12.2 yrs	8.1 yrs
Distribution:	0-4 years	26	22	38
(percent of total)	5-14 years	44	40	54
	15-24 years	24	30	8
	25+ years	6	8	0
Percent managers		24	30	8
Average grade (excl. Executive Service)		GS/GM 12.6	GS/GM 12.9	GS/GM 12.0
Percent GS/GM-13		39	40	33
Ed. highest degree:	Bachelor's	20	16	31
(percent of total)	Master's	48	46	54
	PhD	32	38	15
Years of forecasting (average)		9.6 yrs	10.6 yrs	6.7 yrs
Distribution:	0-4 years	42	40	46
(percent of total)	5-14 years	30	24	46
	15-24 years	18	22	8
	25+ years	10	14	0
Percent whose forecasts published		75	74	77
Percent, forecasts evaluated		71	74	60
of which, percent evaluation published		70	67	80

FFC-92 CHARACTERISTICS BY OCCUPATION

		Total respondents	Managers	Nonmanagers
Years of gov't service (average)		15.0 yrs	20.0 yrs	12.6 yrs
Distribution:	0-4 years	11	0	17
(percent of total)	5-14 years	41	26	49
	15-24 years	30	47	20
	25+ years	19	26	14
Percent male		74	89	66
Average grade (excl. Executive Service)		GS/GM 13.4	GS/GM 14.6	GS/GM 12.6
Percent GS/GM-13		37	0	54
Ed. highest degree:	Bachelor's	19	21	17
(percent of total)	Master's	46	26	57
	PhD	35	53	26
Years of forecasting (average)		10.2 yrs	14.5 yrs	7.7 yrs
Distribution:	0-4 years	33	16	43
(percent of total)	5-14 years	38	37	39
	15-24 years	19	32	12
	25+ years	10	16	6
Percent whose forecasts published		73	72	74
Percent, forecasts evaluated		67	72	65
of which, percent evaluation published		49	46	50

FFC-92 CHARACTERISTICS BY SEX

		Total respondents	Male	Female
Years of gov't service (average)		15.0 yrs	17.0 yrs	10.1 yrs
Distribution:	0-4 years	11	5	29
(percent of total)	5-14 years	41	40	43
	15-24 years	30	30	29
	25+ years	19	25	0
Percent managers		35	42	14
Average grade (excl. Executive Service)		GS/GM 13.4	GS/GM 13.7	GS/GM 12.2
Percent GS/GM-13		37	30	50
Ed. highest degree:	Bachelor's	19	10	43
(percent of total)	Master's	46	48	43
	PhD	35	42	14
Years of forecasting (average)		10.2 yrs	11.3 yrs	7.1 yrs
Distribution:	0-4 years	33	31	38
(percent of total)	5-14 years	38	36	46
	15-24 years	19	21	15
	25+ years	10	13	0
Percent whose forecasts published		73	76	64
Percent, forecasts evaluated		67	64	77
of which, percent evaluation published		49	40	70

Federal Forecasters Conference Survey

Objective of Survey: The purpose of this survey is to obtain general information from Federal Forecasters on themselves. The survey results will be presented in the FFC Proceedings.

Demographic Information:

1. Are you currently working in forecasting employed by the Federal Government? _____Yes _____No (If yes, please answer the remaining questions. If no, do not fill out the survey.)

2. How many years have you worked for the Federal Government? _____years

3. What is your sex? _____M _____F

4. What is your pay schedule and grade? ____________________

5. What is your highest level of education? _____High School _____Associate _____Bachelor's _____Master's _____Ph.D.

6. What academic discipline is your degree in? ____________________

Job Information:

7. What is your job title? ____________________

8. Are you a manager? _____Yes _____No

9. How long have you been forecasting? __________years

Forecast Information:

10. What is the scope of your forecast? _____International
Domestic: _____National _____Regional/State _____Local

11. Are your forecasts published? _____Yes _____No

12. What are your primary forecasting techniques? ____________________
____

13. Do you perform forecast evaluations? _____Yes _____No
and if so, do you publish the results? _____Yes _____No

14. In your view, what is the single most important issue facing Federal forecasters today?
____

Please deposit completed survey in the designated box in the auditorium foyer. Thank you for your cooperation.

Developments in Forecasting: A Word of Caution for the 1990'S

Fred Joutz, Department of Economics, The George Washington University

I would like to thank Edward G. Gamber, Christopher Turner, and R. Clay Woods for data used in this paper. Any misuse of the data is my own.

Introduction

This paper discusses three recent important developments in the forecasting profession. I would like to begin with several quotes.

"I think there is a world market for about five computers."
Thomas J. Watson
Chairman of the Board-IBM, 1943

"Where a calculator on the ENIAC is equipped with 18,000 vacuum tubes and weighs 30 tons, computers in the future may have only 1,000 vacuum tubes and weigh perhaps only 1.5 tons."
Popular Mechanics, 1949

"A severe depression like that of 1920-21 is outside the range of probability."
Harvard Economic Society
November 16, 1929

"Forecasts are educated guesses; by definition they are made with error."
unnamed forecaster

The first two quotes relate to the first development. That is the enormous and rapid increase in computing power and more importantly access to that power. The second development involves the increasing role of time series analysis techniques in model building and forecasting. The final development is one whose outcome is still unfolding. This gives rise to the word of caution in the title of the paper. As the resources available to and demand for forecasters increase, there is a need to recognize and to educate the consumer(s) of the inherent uncertainty in any forecast.

Development 1: Computing Power and Access

In the last fifteen years there has been a rapid increase in the access to computing power and to the capabilities of hardware and software. The personal computer and the "invention" of the electronic spreadsheet stand out as the two major developments. The relative ease with which numbers could be generated and the emergence of the "information age" have increased the demand for projections and forecasts. Below is evidence on the rapid dissemination of (personal) computers have in the work place.

Table 1 presents benchmarks in the development of central processing units for personal computers. The first and second columns give the years that different Intel CPUs became commercially available. The third column gives the number of instructions the CPUs can execute per second (MIPS). Transistors, the "equivalent" of the vacuum tubes in the ENIAC are provided in column four. Finally the clock speed of the machines is given in the last column. Between 1979 and 1982 the MIPS executable increased tenfold. This fall Intel is introducing the P5 which has thirty-three times the "power" of the 80286 CPU in 1982.

It would be good to have a handle on the changes in the cost of this computing power over time. I was unable to get a reliable series for the conference. However, there were several classic studies by Chow (1967) and Triplett (1989) on constructing price indexes for computers. David Cartwright (1986) estimated that computer prices declined at an average annual growth rate of 13-14% from 1972 to 1984. Rosanne Cole et. al. (1986) constructed quality adjusted price indexes for computers and computer parts.

Table 2 reveals the rapid growth in the personal computer (PC) market. Ten years ago about 3.4 million units were sold in the U.S. This brought cumulative computer sales to 5 million units. This year, the Department of Commerce's International Trade Administration projects that about 11 million units will units will be sold. Over the past decade 90 million units have been purchased.

Figures 1 and 2 provide further evidence on the growth and importance of computers as a tool. Producer durable goods expenditures in 1987 dollars on computer equipment crossed the ten billion dollar threshold in 1980. At that time the expenditures represented 2 percent of the total expenditure. In the first quarter of this year expenditures on computers was 67 billion dollars on an annual basis, nearly 18 percent of producer durable goods expenditures. While most of these computers are purchased for non-forecasting applications, scientific research, word-processing, education, and graphic and engineering design, odds are that every planning and forecasting office is equipped with at least one PC for empirical work. Furthermore, a spreadsheet package and at least one statistical/econometric package will be on the hard disk of the PCs.

The technological advances in hardware and software have increased the capability of forecasters. Empirical work and report generation is much easier. Computers can provide forecasters with data answers to all kinds of questions. However, forecasters must guard against accepting whatever the computer "spits out". As Michael Hazilla, a mentor, once told me, "The computer is always right; it does exactly what your program told it do with the data." My own version of this advice is that as PCs get faster and more powerful it only allows modelers and forecasters to produce more errors and in a shorter amount of time. Modeler and forecasters must remember that computers and Pcs are only tools and not ends. Ultimately, they must communicate the empirical results and the meaning behind them.

Development 2: The Role of Time Series Techniques

The techniques of time series analysis have become an increasingly integral part of the toolkit of the applied economist and forecaster. Large structural macroeconometric models lost much of their appeal in the early 1970s for two reasons. The first was that their forecasting performance deteriorated particularly vis-a-vis simple time series models. Second, the emergence of rational expectations during the 1970's cast doubt on the validity of many of the exclusion restrictions used to identify large scale structural macroeconometric models.

One response to these criticisms regarding forecasting performance and ad hoc restrictions was proposed by Christopher Sims (1980). He developed an atheoretical approach to model building called vector autoregressions (VARs). Another approach has been to employ Kalman-filter techniques. These permit modelers to combine series of different frequencies, to adjust for measurement error in preliminary data, and to control for different states of the world. Stock and Watson (1989) have developed and are improving an leading indicator index. One ambitious goal of the index is to predict turning points.

My comments on the development of time series techniques will be more pedestrian and focus on a single variable or equation. The other techniques are the multivariate representations of the simple univariate approach. On a panel like this it is difficult to develop the more technical material. While it can be more powerful, it is easy to get lost in the detail and ultimate goal of the forecasting exercise when using these techniques. Furthermore, we can learn a great deal from careful analysis of the single variable models.

A common transformation in time series analysis of economic data is first differencing (and or seasonal differencing). The implication being that the variable(s) in level form are not stationary. First differencing assumes there is a unit root (coefficient of one) at the first lag. An alternative to the first difference model would include a constant and possibly a trend term; this implies the variable follows a trend stationary process. Kang and Nelson (1984) discuss the characteristics and problems of misspecification of trend stationary (deterministic trend) processes and difference stationary (stochastic trend) processes.

There are two important differences in these kinds of processes. The effect of a one time shock to the deterministic trend process dissipates with time since the series will return to its original path. However, the effect of a shock on the stochastic trend process will be permanent. Confidence intervals for the trend stochastic processes will be bounded. While those for stochastic trends are a positive function of time, implying increased uncertainty. For a good review of the pitfalls and opportunities presented by unit roots to modelers and forecasting see Campbell and Perron (1991) and Perman (1991).

I would like to focus my comments on the development of cointegration time series techniques in forecasting. While it currently is the sexiest topic in applied research, there have not been that many papers (in the U.S.) which have employed the technique in a forecasting environment. See Engle and Yoo (1987), Chambers (1992), and Clements and Hendry (1992). Also, it is related to my third development.

Forecasters are asked to produce models for predicting the short run and long run. Conventional practice involves the construction of separate models for the two horizons. The short run model explains demand as a function of seasonal or rapidly changing variables. The long run model explains demand as a function of slowly changing variables like demographic characteristics and income. In general these two models result in conflicting forecasts at overlapping horizon(s). Forecasters and planners must employ an ad hoc means of reconciling the difference(s) to produce a unified forecast.

A technique offering a possible solution involves cointegration testing which can lead to an error correction mechanism

(ECM). This produces a model which encompasses the information in both the traditional short-run and long-run models. One interpretation of this is that the traditional short run and long run models are subsets of the merged or "true" model. The presentation below directly follows Engle, Granger, and Hallman (1989). See Alogoskoufis and Smith (1991) for a discussion of specification, interpretation, and estimation of error correction models (ECM).

The variable of interest is y_t. The information set, Y_t, contains lagged values of y_t, other endogenous variables, exogenous variables and predetermined variables. Assume the variables are in natural logarithms.

If the series $(1-B)^d y_t$ is stationary it is said to integrated of order d, I(d). Here B serves as a backshift operator such that $B^i y_t = y_{t-i}$. When using monthly or quarterly seasonally unadjusted data it may be the case that the seasonal differencing and even multiplicative differencing, first and seasonal differences, is required to make the data stationary.

Suppose that d=1, then y_t is a random walk and may or may not have drift. A series integrated of order one, I(1), is smoother and slower changing than stationary I(0) series. The former has no affinity for the mean value, so that departures from the mean can be long.

Let w_t be a sub-set of Y_t and integrated of order d, the same as y_t. From the Granger representation theorem if there exists a stationary linear combination

$$z_t = y_t - \beta' w_t \tag{1}$$

then w_t is co-integrated with y_t. This implies the data generating process can be represented by an error correction model or mechanism, ECM, of the form

$$\Delta y_t = \mu - \pi * z_{t-1} + \gamma' x_t + \epsilon_t \tag{2}$$

where x_t is I(0) explanatory variables. Stationary lag polynomials of Δy_t and Δw_t may be included in x_t. The μ term can represent the intercept or "trend" growth and centered seasonal dummy variables. The random disturbance, ϵ_t is assumed to be white noise. This ECM model can be interpreted as the "true" or merged model.

The <u>long run (forecasting) model</u> is assumed to use the elements of Y_t which are I(1) and takes the form:

$$y_t = \beta_0 + \beta_1' w_t + \eta_t. \tag{3}$$

where the expected value(s) of β_1 are β in equation (1). These can be interpreted as the long run elasticity estimates. Again, the random disturbance, η_t, is assumed to be white noise.

The usual <u>short run (forecasting) model</u> does not incorporate the error correction mechanism; it omits information from the long run model.

$$\Delta y_t = \gamma_0 + \gamma_1' x_t + \epsilon_t; \quad where\ \Delta y_t\ and\ x_t \sim I(0) \tag{4}$$

The γ_1' represent the short run elasticity estimates. Notice that the short run elasticities in this expression can differ from those in (2). The ϵ_t could be white noise or follow a autoregressive process.

Thus a forecaster has a three potential forecasting models. The encompassing one as represented in equation 2, a long run model as in 3, or the short run model in equation 4. The first one is the merged model which makes efficient use of available information.

When monthly data is available for y, w, and x, the one step ahead forecast for the ECM or merged model is

$$y_{t+1} = f_{t,1}^{y} = (1-\pi) y_t + \pi \beta' w_t + \gamma' x_{t+1} + \epsilon_{t+1} \tag{5}$$

where x represents the forecasted value(s) of the explanatory variables. Here $f^{y}_{t,1}$ represents a forecast of y based on information available at time t out one period. Longer horizon forecasts are constructed from iterating the expression above out the desired number of periods.

As the forecasting horizon increases the x variables approach their monthly expected values. This results in a deterministic component to the forecasts, μ^*. We can express the long run or h step ahead forecasts two different ways:

$$f^{y}_{t,h} \approx \mu^* + (1-\pi)\, f^{y}_{t,h-1} + \pi\beta' f^{w}_{t,h-1}$$
$$or$$
$$f^{y}_{t,h+1} - f^{y}_{t,h} \approx \mu^* - \pi\,(f^{y}_{t,h} - \beta' f^{w}_{t,h}) \tag{6}$$

In the long run the forecasted change in the variable of interest is equal to the deterministic component(s) minus the difference between the predicted variable the previous period and the estimate from the long run predicted value using forecasts of the w variables. As the left hand side in the second expression approaches a constant, then the right hand side becomes

$$f^{y}_{t,h} = constant + \beta' f^{w}_{t,h} \; . \tag{7}$$

This approximates the long run model from equation (3). Furthermore, by implementing the ECM model short run forecasts are produced similar to those using the short run model, equation (4). These forecasts could even be improved, because of the inclusion of the information about the cointegrating relationship. Thus it appears as though the ECM framework provides a consistent bridge between the long run and short run forecasts.

Development 3: Handling Uncertainty and Consumers of Forecasts

It would seem that forecasters should be celebrating. They have access to powerful computing equipment, numerous software packages, and sophisticated techniques to build bigger and better models. Why am I suggesting caution?

Unfortunately, the number of forecast consumers, their demands, and expectations have risen faster than the capabilities of the forecasting profession. Furthermore, a number of fundamental problems remain and will continue to be with the community for a long time. (The classic example of this plague is data quality and availability.)

Federal model builders and forecasters are innovative problem solvers by nature. However they are too often and too easily convinced or volunteered to undertake grand projects. I would term these "global" models which are part of a grand scheme or political objective by policy makers in an agency. Alarm bells should go off in every model builder and forecasters head. That is impossible; they cannot have a 1.) theoretically consistent, 2.) empirically sound, and 3.) computationally manageable model to answer all the questions. The basic impression one gets from reading the hoopla announcing these modeling systems is that agency policy makers expect to have a universal (macro)economic modeling system which can answer every conceivable question.

A central problem in the design is that policy makers, consumers of these models, are unlikely to understand or accept the limits of realistic modeling. The notion of a confidence interval around a forecast is foreign. There is only one number. They treat forecasts as an end when in fact forecasts are tools and educated guesses in the decision making process. Forecasters must resist the demands and protest for a single answer or number. The government agency's modelers in these grand schemes are put in the unhappy position of providing data which they know will be wrong. The notion of uncertainty or confidence interval must be provided with every forecast. Agency heads, policy makers, forecasters need to decide which variables of interest can be reasonably forecast and still answer the more important policy questions. Further, the agency heads and policy makers need to ask themselves and the modelers what information is likely to be important in the future? Can it be realistically produced by models? At what level of confidence, accuracy, or believability could the information be supplied? What are the minimal requirements for timeliness, cost, inclusiveness? What resources are needed and where will they come from?

The task of policy makers is to plan for different contingencies and consider different options. This is particularly important with budget projections. The policy makers will demand a single number. Here is a classic opportunity for those charged with forecasting to resist the demand for a single number. They can give their best estimate(s) but must supply confidence intervals attached to them.

Suppose that legislators and the executive must come up with a deficit reduction package which is "perceived" or "sold" as solving the problem. There are costs to over-predicting and under-predicting revenues and expenditures. The costs are not symmetric. If policy makers are forewarned about the degree of confidence in "the" point estimate, their planning will incorporate the uncertainty and thus be improved. If uninformed, and the point estimate turns out be wrong, policy makers will not listen to excuses that what actually occurred was within the confidence range of the model. When surprised like this several times in a row the credibility of the modelers and forecasters falls. Thus it pays both society and the profession to educate the consumers of forecasts.

Conclusion

The economic forecasting profession has benefitted from technological advances and the application or merging of time series analysis in its work. Unfortunately the demands and expectations of forecasters have risen faster. If modelers and forecasters do not educate the consumers about their product(s), errors when they inevitably occur, will reduce the forecasters credibility and effectiveness. Now, is a particularly opportune and good time to communicate this advice. The demand for economists and econometric forecasts is acyclical. Policy makers, both public and private, seek them out in stagnant and recessionary periods. McNees (1990) has written that macroforecasts macroeconomic forecast errors have not been larger as volatility in macroeconomic activity has risen in the 1980s. However, uncertainty is ubiquitous with any forecast and this needs to be communicated to the consumers of forecasts. They can be convinced of the benefits of confidence intervals accompanying the point estimates, when they are recognize the potential costs of not having this information.

References

Alogoskoufis, George, and Ron Smith, (1991) "On Error Correction Models: Specification, Interpretation, and Estimation," Journal of Economic Surveys, vol. 5 no. 1, p97-128.

Campbell, J.Y. and P. Perron, (1991). "Pitfalls and Opportunities: What Macroeconomists Should Know About Unit Roots," NBER Macroeconomics Annual,vol. 6, pp141-219.

Cartwright, David W., (1986) "Improved Deflation of Purchasers of Computers," Survey of Current Business, 66:3, March, pp7-9.

Chambers, Marcus J., (1992) "A Note on Forecasting in Co-Integrated Systems," working paper Department of Economics, University of Essex.

Chow, Gregory C., (1967) "Technological Change and the Demand for Computers," American Economic Review, 57:5, December, pp 1117-1130.

Clements, Michael P. and David F. Hendry, (1992) "Forecasting in Cointegrated Systems," working paper Insititute of Economics and Statistics, Nuffield College, Oxford.

Cole, Rosanne, Y.C. Chen, Joan A. Barquinn-Stolleman, Ellen Dulberger, Nurhan Helvacian, and James P. Hodge, (1986) "Quality Adjusted Price Indexes for Computer Processors and Selected Peripheral Equipment," Survey of Current Business, 66:1, January, pp41-50.

Engle, R.F., and B.S. Yoo. (1987), "Forecasting and testing in cointegrated systems," Journal of Econometrics 35, pp143-159.

Engle, R.F., C.W.J. Granger and J.J. Hallman, (1989) "Merging Short- and Long-Run Forecasts: An Application of Seasonal Cointegration to Monthly Electricity Sales Forecasting," Journal of Econometrics, 40, p.45-62.

Kang, Heejoon and Charles R. Nelson, (1984) "Pitfalls in the Use of Time as an Explanatory Variable in Regression," Journal of Business and Economic Statistics, vol. 2, no. 1, p.73-82.

Perman, Roger, (1991) "Cointegration: An Introduction to the Literature," Journal of Economic Studies, 14:3, pp3-30.

Sims, Christopher, (1980) "Macroeconomics and Reality," Econometrica, 48, January, pp540-552.

Stock, J.H., and M.W. Watson, (1989) "New Indexes of Coincident and Leading Economic Indicators," in NBER Macroeconomics Annual, edited by O. Blanchard and S. Fischer, The MIT Press, pp351-409.

Triplett, Jack E., (1971) "Quality Bias in Price Indexes and New Methods of Quality Measurement," Chapter 6 in Zvi Griliches, ed. Price Indexes and Quality Change, Cambridge, MA, Harvard University Press, pp 180-214.

Table 1
Intel CPU Speed
5 Generations from 1979 to 1992

Year	CPU	MIPS	Transistors	MHZ
1979	8088	0.33	29000	5
1982	80286	3	134000	12
1985	80386DX	11	275000	33
1989	80486DX	41	1200000	50
1992	P5	100	3000000	100

"Trends" PC Magazine, September 15, 1992, p.31

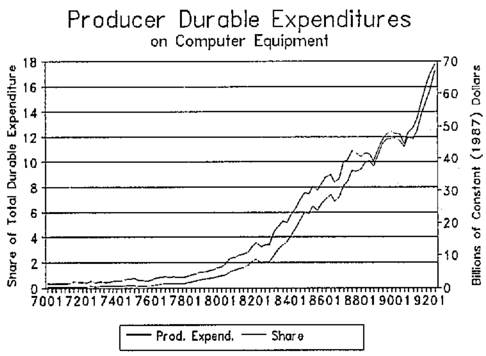


Figure 1

Table 2
U.S. Personal Computer Market
Total Sales, Millions

Year	Units	Cumulative
1980	0.52	0.96
1981	0.85	1.81
1982	3.40	5.21
1983	6.10	11.31
1984	7.50	18.81
1985	5.90	24.71
1956	6.80	31.51
1987	8.20	39.71
1988	9.10	48.81
1989	9.80	58.61
1990	10.20	68.81
1991	10.50	79.31
1992	11.00	90.31

Source: U.S. Department of Commerce
International Trade Administration
Data for 1992 is projected.

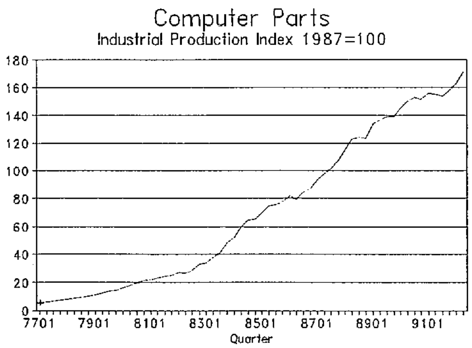


Figure 2

The Most Important Developments in Forecasting in the Recent Past

Herman O. Stekler, Industrial College of the Armed Forces, National Defense University

In thinking about what I consider the three most important developments in the field of forecasting which have occurred in the past decade or two, or three, I was struck by a paradox. This involves the fascination that our profession, which should be forward looking, has with the past. I will, therefore, try to show how these past developments have positive implications for our ability to make statements about the future. Briefly, the three most important developments in forecasting, in my view, are (1) the development of quantitative methods for forecasting, (2) the process of evaluating forecasts, and (3) the recognition that there are limits to our forecasting abilities.

I. Development of Quantitative Methods

For those of you who are immersed in PC, and quantitative methods, it may come as a shock that it wasn't always like this. Quantitative economic forecasts were not used at some of the Federal agencies, which directed economic policy, until the mid-1960s. Until then, the emphasis was on describing trends in imprecise delphi-like language which could be interpreted after the fact as having predicted whatever, in fact, did occur (Much like the stock market analyses which are still provided today.) The weather forecasters began the development of their models of weather systems in the 1950s. These models have definitely improved the quality of weather forecasts and the quality of economic forecasts has probably also increased.

It is not the fact that the forecasts are quantitative that makes this development important, rather it is the way that these numerical values are generated that is the most important feature. In generating a forecast, it is necessary to spell out all the assumptions that drive the forecasts and to insure that the results are consistent with each other. The use of models formal and informal, insures that these relationships are specified and that consistency is maintained.

Moreover, if forecasts are generated in this manner it is possible to describe the scenario for the end user in a meaningful way. This in and of itself, is an important contribution. It is also possible to evaluate quantitative forecasts and to determine the sources of error.

What about the atheoretical nonmodel, but still quantitative forecasts, such as Box-Jenkins' ARIMA? Again, these time series/procedures are relatively new-1970 for Box-Jenkins' major work. The development of modern computers has made it possible to generate hundreds of such series in the time that it took me, as a graduate student, to generate one multiple-regression using an electronic desk calculator.

Yes, these time series models are naive extrapolative procedures, but they serve other useful functions. They would show what could be expected if the past trends persisted. Thus they are a very useful check against the forecasts of more formal models to see whether all the dynamics of the system have been incorporated into the model. These time series models can be developed quickly and are excellent standards against which the more formal models should be compared. They are also very useful when many items such as sales of particular models or end-item inventories, must be predicted quickly.

Thus the development of quantitative forecasts enables the end users to receive more precise predictions, mandates that assumptions are clearly specified and permits forecast evaluations to be performed. These are all desirable qualities for our work.

II. Forecast Evaluation

The emphasis on forecast evaluation has increased over the recent past. This development can be attributed to two factors, the publication of quantitative forecasts (which has already been discussed) and the development and application of statistical techniques appropriate for this task.

We can then ask questions such as,

(1) How good is a method?
(2) Are the forecasts biased?
(3) is one method better than another?
(4) Does one forecast contain information not available in another, etc.?

But conducting such evaluations and answering such questions should not be an end in and of itself. Rather evaluations should be undertaken for other purposes. They could be used for choosing one technique, forecaster or forecasting service over another. (Personally, that's not likely to be useful since most forecasters do about equally well on average.)

Or the evaluations might be used for introspection to determine the sources of error. Once these sources of error have been found, and corrected, there should be an improvement in the quality of subsequent forecasts. As an example, consider the economic forecasts involving the current recession. Most economists have been surprised at the weak recovery. This despite the public's political outcries indicating that something was fundamentally wrong with the economy. Where did the forecasts go wrong? Were the data bad? Were our models incorrectly specified? Were the assumptions erroneous? or was this recession more similar to the pre-WWII decline than to those of the post-war period? These are the kinds of questions that must be answered if forecasts involving subsequent recessions are to be improved. Thus evaluations serve as a useful purpose if better techniques can be identified and if the right questions about the sources of errors are posed and the appropriate corrections are made to the procedures which generate the forecasts.

III. Limits to Forecast Accuracy

The knowledge obtained form these forecast evaluations serve another useful function. They establish the limits of forecast accuracy that our current state of knowledge and skill permit. Recognizing that there are limits to forecast accuracy is an important step in reducing the perennial conflicts between the supplier of forecasts (staff) and the end user of these predictions (policy or decision makers).

These conflicts occur because the end user frequently wants the kind of information that the forecaster cannot reasonably deliver, i.e., what is the likely state of the economy a year from now? What will the weather be like a week from Saturday?

If this recognition of the limits of forecast accuracy or plausibility is an important development from the recent past, what are the implications for our future work? First, since precise statements about likely outcomes probably will be inaccurate, we must devise techniques for presenting alternative outcomes. These might take the form of point predictions with probabilities attached. Alternatively, we might present the decision maker with a set of alternative scenarios.

The last approach opens up another set of issues. How does one recognize that one scenario is unfolding or that another has no chance of happening. The problem becomes one of identifying a sequence of events that will lead to one outcome rather than another and determining whether the observed data follow one of those sequences. This is likely to be a fruitful avenue of research for the academic forecasting profession.

IV. Summary

In summary, the three major developments that I have discussed, (1) quantitative forecasts, (2) evaluations and (3) recognition of accuracy limits, have implications for the future of forecasting. First we must continue to make and evaluate precise numerical estimates and we must improve our techniques for both tasks. We must ask the right questions in our evaluations and we must understand the forecasting process so that our errors and biases can be eliminated. Finally, we must deal with the issue of making forecasts for events in the distant future which cannot be predicted precisely. These are certainly challenges for both the practitioners and the academics concerned with forecasting.

Developments in Forecasting in the Past Decade

Peg Young, Office of Inspector General, Department of Veterans Affairs

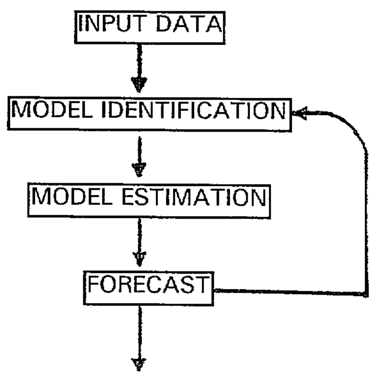
We forecasters are, by profession, often asked to discuss the events of the future. It is with an unusual, but great, pleasure that I respond to a request to discuss the events of the past - in this case, the events of the past decade - on the realm of forecasting techniques. I propose to discuss the past decade of forecasting along the lines of three particular developments. These are (1) the model selection process of finding the best forecast technique for a given set of data, (2) the realization that the 'best' technique may be the combination of techniques, and (3) a development of a forecasting procedure that we may be seeing much more of in the future.

I. Model Selection of Forecasting Technique

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 (Makridakis et al., 1982), 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.

During the last decade, progress has been made to study the characteristics of the data and to let these characteristics drive the selection of an appropriate forecasting technique - appropriate, that is, for the data under consideration. Such characteristics include the length of the data set, the length of the desired forecast horizon, the quality of the data, the seasonality of the data, the apparent error structure of the data, and the purpose of the forecast being conducted Kendall and Ord, 1990). The search for the optimal technique for all data sets has been superseded by the search for the technique that works best for the data of interest. The data drives the forecast, not the forecasting technique.

This is not a new concept; the classic Box-Jenkins paradigm provided the underlying procedure to perform such an analysis process (Vandaele, 1983). Figure 1 provides a flow diagram that portrays the process that allows the data to drive the model selection. Characteristics of the data (which, for ARIMA modelling, were usually displayed through autocorrelograms, or ACF's) were utilized to select the appropriate form of the ARIMA model. Note, in this instance, that the data characterization process pertained to the form of ARIMA model to be employed, so the earlier research still was geared towards the creation of the single best form of forecasting model - the Box-Jenkins model.



II. Combination of Forecasting Techniques

The assumption behind the model selection procedure is that there does exist a forecasting model that is 'best' for your set of data. This assumption may not hold - in some cases there seem to be several forecasting procedures that appear to be valid. If the handful of selected procedures each provides the same forecast, then from a practical point of view, any one of the techniques is appropriate. Other considerations, such as the cost of the procedure, would then dominate the selection of the technique. Suppose, however, that several techniques are deemed appropriate, and each results in different forecasts - what is an analyst supposed to do? Literature from the past decade proposes that the forecasts be combined into a single forecast.

The advantage of the procedure of combining forecasts is the ability of taking projections from essentially different points of view and compile them into a single, but complete, viewpoint. The process of weighting the estimates allows the forecaster to determine which estimate is deemed of greater value, and thereby deserving of a greater weight. The varying inputs to the forecast is not unlike the process undergone in regression modelling, in which each input is weighted according to its importance to the output, the ultimate forecast.

In practice, the combination of techniques encompasses forecasting procedures from varying degrees of objectivity. Often managers utilize the combination procedure to combine time series procedures with judgmental approaches. Subjective information can be incorporated with the more quantitative approaches, if information is known about changes in historical patterns that would not be reflected in that data set. Scenario development utilizes similar approaches to prepare strategic plans and alternative futures which incorporate the differing viewpoints.

III. Structural Modelling [The Harvey Model]

The previous two topics deal with the issue of multiple forecast procedures - how to choose among techniques or how to combine techniques. There still exists the desire for the 'optimal' procedure, and for those who prefer this direction in forecasting, there is a development that is appropriate for you.

It was originally thought the Box-Jenkins time series procedure, or ARIMA modelling, would be the 'be-all and end-all' of forecasting techniques. The model was designed to handle constant, trended, seasonal, cyclical, autoregressive, moving average, and random behavior - with or without interventions from external variables. Experience proved, however, that the process was rather unwieldy, if not just impossible, for all but the most experienced forecaster to use. Even in cases when the Box-Jenkins procedure proved to offer the best forecast, the explanation of what the model did proved to be evasive, if not an outright case of obfuscation. Managers appeared to be disenchanted with the description of time series behavior in terms of 'degrees of differencing', 'backshift operators', and 'autoregressive and moving average terms'.

The automation of model selection and application, as through 'expert systems', allowed for greater ease of use of the Box-Jenkins modelling, but there was still lacking the ability of explaining the results to any degree of satisfaction to the managers forced to use the results. For the sake of clearer explanations and greater compliance with the procedure, forecasters moved their analysis of univariate procedures back to the realm of regression analysis, which proved to be more acceptable to managers but not truly designed for the assumptions behind time series behavior. The forecaster was forced to decide whether she wanted forecast power with limited comprehension, or increased comprehension but with potential, if not outright, violations in assumptions.

Andrew Harvey, of the London School of Economics, provided a solution to the dilemma - structural modelling (Harvey and Peters, 1990). With the ease of explanation (as found in regression modelling), Harvey was able to incorporate time series data in a model which takes advantage of the time series nature of the data, rather than assume it away. The technique links together ARIMA models with state-space models through a class of structural models. These structural models, time series in nature, offer clearer interpretations through the decomposition into components of trend, seasonality, and so on. A detailed approach to this structural time series model is provided in Harvey's book, Forecasting, Structural Time Series and the Kalman Filter (1989, Cambridge University Press). For those who have been able to get a copy of his software, STAMP, or to program their own software to perform the calculations, they have found the procedure highly accurate in its analysis, able to incorporate a wide variety of behavior in the time series, and still be comprehensible in its results from a managerial point of view.

IV. Summary

The three suggested directions of forecasting taken in the past decade are not the only major changes in the field, as so noted by the other members of this panel. The field of forecasting has been undergoing a metamorphosis in the last few decades. Researchers from a variety of avenues - statistics, economics, engineering, psychology and so on, have all been providing new and valuable input in the field. As individuals from different realms of research begin their foray into forecasting, their impact on the manner in which forecasts are created will be felt. Such is the advantage of such a melding of knowledge. We forecasters look forward, with anticipation, to the next decade of change in forecasting.

References

Harvey, Andrew C. Forecasting, Structural Time Series and the Kalman Filter. Cambridge: Cambridge University Press, 1989.

__________,Harvey, A. C. and Peters, S. "Estimation Procedures for Structural Time Series Models." JOURNAL OF FORECASTING, 9, 1990, 89-108.

Kendall, Sir Maurice and Ord, J. Keith. Time Series, 3rd. edition. London: Edward Arnold, 1990.

Makridakis, S et al. "The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition." JOURNAL OF FORECASTING, 1, 1982, 111-153.

Vandaele, Walter. Applied Time series and Box-Jenkins Models. Orlando: Academic Press, Inc., 1983.

Developments in Forecasting: Comments

Karen S. Hamrick, Economic Research Service, U.S. Department of Agriculture

The question posed to the three panelists was: What are the three most important developments in forecasting in the last 10 years? In addition to asking the question, the panelists were urged to also explain why each development has had an impact on forecasting.

All three of the panelists have produced papers that are thoughtfully done and interesting to read. The ideas presented are all thought-provoking, and indeed, there was a lively discussion among the panel and with the audience at the conference session. Perhaps at the 15th Annual Federal Forecasters Conference (FFC 2002) we can review whether or not the developments of the 1980's presented here were actually historically significant.

Fred Joutz lists the phenomenal growth of computing power and access as one of the most important development to forecasting. He compiled statistics to back up this assertion, to quantify what most people assume as fact. That improvements in technology and computer availability have made a difference to forecasting is unquestionable; forecasters can do their work with an ease that was not possible just 10 years ago.

Joutz's second development is the role of time series techniques, and in particular, cointegration techniques and the error correction mechanism. His discussion of cointegration and his presentation of how to apply this technique to forecasting is useful. In addition, he provides a good bibliography for those who are new to the cointegration discussion. However, I question whether one particular technique is an important development. Techniques come and go in terms of popularity, and certainly cointegration has dominated the economic literature recently. I wonder if it will still be considered an all-purpose solution 10 years from now.

Joutz finishes off his list not with a development but with an issue for the 1990's--modelers and forecasters must educate their consumers on the uncertainty associated with a forecast. The perennial need to educate is even more important given the previously-mentioned computer advances. Just because a forecast is easier to generate does not mean it should be less thoughtfully made or communicated.

Herman Stekler chose to alter the question a bit and looked at developments over the last 20 to 30 years. His changing of the question makes his response no less insightful. First, he identifies the development of quantitative methods as significant. His justification is that by using a quantitative method, a forecaster must make assumptions explicit and must ensure that they are consistent. Not only does this make the forecast theoretically robust, but also it is more easily explainable to the end user. I agree with Stekler that it is not a specific technique, but instead the body of quantitative methods that is important. This is particularly important in fields that have been historically less quantitative, such as psychology and political science. The technique forces the forecaster to make explicit all the information that makes up the analyst's intuition. The process of doing this can, in some cases, be more useful than the forecast itself, since it forces the forecaster to analyze the forces affecting the forecast variable.

Second, Stekler asserts that the increased emphasis on forecast evaluation is also important. This development is possible because of the development of techniques for evaluation. I agree with him that the importance of evaluation is that the forecaster gains a better understanding of the events being forecast and can consequently improve future forecasts by evaluating previous forecasts.

Finally, Stekler stated that the recognition that there are limits to forecasting ability has been important. Although I agree, it is unclear to me that forecasters now know there is uncertainty attached to their forecasts, and did not know this years ago. Perhaps they now know more precisely what the uncertainty is, and are better able to communicate the caveats with their clients as Joutz recommended.

Peg Young provided improved model selection of forecasting technique as the first important development on her list. The crucial change in thinking has been from "What is the best technique?" to "What is the best technique for this data?" By using the process of identifying the model, estimating the model, forecasting, and then returning to the task of identifying the model, the forecaster generates a better result. This development is the same as Stekler's first two developments: use a quantitative technique to generate a forecast, then evaluate the accuracy of the forecast in order to better the forecast model.

Young's second important development of the 1980's is the combination of forecasting techniques. In the event that one particular quantitative method is not the clear choice for a forecast model, the forecaster can combine forecasts from two (or more) models, say, a structural economic model and a time series model, in order to produce one forecast utilizing information from both. Although the combination of forecasts has been getting quite a bit of attention in the literature, I respectfully submit that forecasters have always been combining forecasts, specifically, adjusting a model-generated forecast with judgement, however, not with the precision now provided by these combination of forecast techniques. Having said this, perhaps Stekler is correct in that by using techniques forecasters made their assumptions more explicit and consistent. So now, as always, forecasters adjust using judgement, but do the adjustment in a more clearly defined way.

Finally, Young cites the Harvey Model as an important development. The Harvey Model is similar to cointegration techniques in its attempt to combine the time-series characteristics of the data with structural models. Again I question if any one technique will prove significant over time in shaping how forecasts are done.

The questions now to consider are: Have these developments made a positive or negative impact on forecasting? Has the improved computer access made forecasting too easy and too fast, such that the art and substance of forecasting is lost? Has the proliferation of techniques helped forecasters or has it made them more tentative, reluctant to provide a forecast until all techniques are tried? Has the ability to evaluate forecasts truly helped in understanding the data or has it just made forecasting a horse race with forecasters jockeying to attain accurate short-run point estimates?

In thinking about these questions, two ideas touched on by all three panelists are relevant now more than ever. First, there is a need for forecasters to understand the data and the substance of their forecast. And second, forecasters must educate their forecast users and communicate the uncertainty associated with the forecast.

As a final comment, I suggest that here is one development over the last 10 years that was not mentioned by the panelists. That development is the existence of forecasting as a field in and of itself, and not just an obscure specialty within other disciplines. Not only is that evidenced by the three panelists themselves, all professors of forecasting, but also of the 12 years of the International Symposium on Forecasting and the five years of the Federal Forecasters Conference. This bringing together of forecasters from such a diverse collection of fields--economics, business, statistics, metrology, psychology, political science, demography, actuarial science, operations research, sociology, and planning--can only benefit forecasting. Both the cross-pollination of methods and techniques and the introspection resulting from evaluating these various techniques will strengthen the body of knowledge that is forecasting. The 1990's should prove to be both exciting and productive years for forecasting.

References

Holden, K., D. A. Peel, and J. L. Thompson. Economic Forecasting: An Introduction. Cambridge, England: Cambridge University Press, 1990.

Kennedy, Peter. A Guide to Econometrics, third edition. Cambridge, Massachusetts: The MIT Press, 1992.

The Nursing Demand-based Model (NDM) for Nurse Requirements Forecasting: Methodology, User Access and Interface

William A. Losaw, Division of Nursing, Bureau of Health Professions

I. General methodology

The Division of Nursing has employed a variety of modelling techniques over the past two decades in an effort to forecast the requirements for nursing resources - i.e. registered nurses, licensed practical/vocational nurses, and nursing assistive personnel (RNs, LP/VNs and Aides, respectively). These techniques included ratio estimating, expert judgement, system dynamics, historical trend analyses, and others.

The current approach uses the basic ordinary and weighted least squares regression techniques applied to cross-sectional data and is able to represent demand, as opposed to need or utilization. Additionally, the model incorporates and quantifies specification of the casual factors which are developed as the underlying forces responsible for change in the health care system. The explicit representation of the casual factors is important to the model user because they may be varied according to a scenario of future change and therefore easily converted into appropriate model behavior to answer "what if" questions.

Demand, as defined in the NDM, is the number of full-time equivalent nurses that employers would actually hire given prevailing market conditions, if not constrained by the availability of nurses. This definition does not allow goal or need-based specifications or standards unless employers would be forced to hire to the levels dictated.

The health care "system" is partitioned into thirteen sectors which represent important and distinct employers of nurses (Figure 1). The thirteen subsectors are related to five major employer sectors as shown, and, depending on the information objective, are aggregated upward as necessary. In addition to the employment sectors, the model is capable of forecasting annual demand levels for each State. Again, various aggregations of state level data may be achieved for reporting purposes.

The model attempts to deal with three levels of nursing, RNs, LP/VNs and Aides in all thirteen subsectors within each State and the United States. As is evident in Figure 2, the availability of data does not permit consistent model detail in the number of subsectors across levels of nursing nor the most appropriate nurse to service utilization ratios. While there is sufficient data to provide an analytic structure for RN utilization in all thirteen subsectors, the breadth of that support progressively diminishes for LP/VNs and Aides, decreasing successively to 10 and then 5, subsectors. Even for the RN utilizations, the lack of State level data forces the model to address utilization in terms of the general population for the ambulatory and public and home health subsectors, where a more suitable utilization basis would consider visits or encounters, depending on the nature of the subsector.

The model also forecasts several health care utilization variables (Figure 3), and considers a number of causal factors which represent the forces that influence and determine the behavior of the health care system. In addition to the utilization variables, these factors include general morbidity outcomes, insurance coverage - both general and, in one case, subsector specific (nursing homes), types of nurse employment (overtime, temporary contract personnel) as well as vacancy rates, HMO coverage and per capita income.

The connections or functional relationships embedded in the model are employed by the user to understand the effect a change in one or more influencing factors will have on the health care system as modelled. For example, the linear regression result presented in Figure 4 would be referenced by a user to understand that increases in LP/VN enrollments have a slightly greater impact on the demand for RN educators than does an equal increase in the RN school enrollment.

II. NDM computer model

The computer implementation of the NDM presents the user with the opportunity to:

1. Develop a scenario representing a possible future behavior mode of the health care system using the models input data as a point of departure,

2. Run the model with the input data sets developed by the user for the new scenario,

3. Develop data products which present and describe the forecast assumptions and results, and

4. Save the data representing the user's developed scenario which allows the user to regenerate the scenario at some future date and/or continue refinement/modification of the current scenario.

There are a large number of action sequences a user may need to employ in the course of generating a scenario and its information products. Figures 5 and 6 represent the flow of menu choices and the end operation content exercised at the end of the menu paths.

The menu system of the NDM acts as an interface between the computer based functions of the model and the user: it is entered immediately upon activation of the NDM program. There are nine sets or groups of menu "bars" or screens that make up the menu system (Fig 5). The G1 group is the overall control screen - this is a platform from which the user moves to specify a scenario (by operating within the G3 group), runs the model and specifies output products (the G2 group), or prepares to shutdown the model or start a new exercise (the remainder of the G1 group). The procedure for moving up and down through the menus is the same for all groups and is accomplished by using the right and left arrow keys to move to a selection and then pressing the enter key to select it, or the user can press the highlighted letter of a selection to execute it. Simply pressing the escape moves back one menu.

The menu sequence allows the user to move to different areas of the model operational venue and adjust/change/revise/etc. a large variety of model controls and specifications. While the procedure for moving down or back up through the menu system is relatively straight forward, the use of the model's controls and specifications requires a somewhat more deliberate and thoughtful approach.

The sequence of menu choices typically leaves the user at a menu endoperation or "endop" - i.e. literally the end of the line of a set of screen choices made by the user. As the user's action at these endops will determine the subsequent behavior of the model, the spectrum of responses needs to be understood by the user. While there are a few of the endops that are administrative in nature (e.g the BINARY, DUMP, QUIT, MEMORY, SAVE, LOAD, CLEAR, YEARS, and STATES endops), there are a great number which directly impact the model's forecast results or explicitly shape the content and form of the model's information products.

While Figures 5 and 6 represent a compact description of the NDM computer universe, the actual processes of it's operation must take place in the real world. The flavor of the operational environment can perhaps be sampled with the following examples.

The user moves either to the tasks of specifying a scenario (the G3 group) or determining the model's output products (the G2 group). Looking first at the scenario area, let the user specify that the managed care index of a given State is to be 10% above what is currently assumed in the model's input data base. The user will move through the menu hierarchy by selecting "Scenario", then "Edit", then "Variables". At this point a list of the independent variables in the model is presented to the user and "managed care index" would be picked. An editing screen would appear which accepts modification factors for the selected variable, in turn, for all states in the scenario. The user would move to the State(s) desired, enter the modifications that would cause a 10% elevation in the variable. When finished, the user must then back out through the menus to the point where further scenario specifications could be entered, or where the forecasting process could begin.

The forecasting process basically entails the selection of a type of output and then generating and/or presenting that output. Again, for example, if the user was interested in the number of Registered Nurses demanded in each of the states of the scenario over the years specified in the scenario, then the user would move through the menu hierarchy by selecting "Reports", then "Nurse", then "Demand", then "States". At this point a list of nurse types would appear and the user would pick "registered Nurses". The state by year demand matrix would be displayed for the user which could be written to a disk file by hitting the "F" key and entering a filename. As before, the user would then back out through the menus to the point where other forecasts could be made, the scenario revised, or the modelling session shut down.

III. User-Model Interaction and Reaction

A group of potential users representing planning agencies from eight states scattered across the U.S. and from three national nursing organizations were brought together by the Division for a two day workshop designed to provide an understanding of the NDM computer model capabilities and how it might be used as a forecasting tool. All of the individuals invited had responsibility for or were directly involved in the forecasting of nurse requirements. The session presented - in considerable detail - documentation, demonstrations and hands-on exercises covering the model features and characteristics summarized earlier in the presentation.

The users themselves represented a variety of backgrounds in terms of analytical expertise and computer literacy. The analytical backgrounds of the participants included some academic and practical statistical experience with most having encountered at least some rudimentary modelling and a few having been involved in some relatively complex modelling efforts. All had at least a basic level of computer literacy, most had a familiarity with Lotus 1-2-3 and a few had used more demanding statistical applications. Almost all represented at least middle management levels in their areas of their organizations, although such organizations are generally small. Participation in the workshop was completely voluntary and the entire group maintained a good level of enthusiasm throughout the workshop and did well with interpersonal communications. There were several results from this workshop process that appear to have value to other forecasters contemplating the development of computer based applications of this nature for their clients.

While the computerized NDM model is constructed along lines parallel to scenario and report development, and maintains the ability to demonstrate some of the models intermediate results and other quantitative aspects (e.g. the editing values used to change variables in the model), the level of computer literacy that is now exhibited by users imposes important design requirements on applications software. A general class of comments directed toward possible computer based model improvements addressed a set of service or support operations available on many of today's popular commercial spreadsheet, word processing and data base applications. Among those mentioned the most frequently and vehemently were:

1. A file list capability whereby the user could review the disk data files in any given directory. This would enable the user to review the scenario data files, output data file, report data files as well as other files important to the user's modelling effort.

2. At least a basic help facility to explain the situation the user is in and the actions available to the user.

3. A menu display which provided a concise operational title which clearly positions the user in the menu hierarchy without having to recall the user's path to that point in the menu.

An offshoot of the type of expectation demonstrated above can be further appreciated by the following example. The NDM allows the user to output reports by hitting the "F" key (for "FILENAME") and then responding to a request for a data set name. Some members of the workshop group were uncomfortable with the "hit the F key" type of instruction because they associated that with the ten or twelve "function" keys (the F1, F2, F3, etc) commonly used for various operations by current applications programs.

There were several other considerations evidenced by the group during the course of the workshop that are worth noting. The definitions used need be clear and the participants mind appropriately refreshed so that ambiguities or unfamiliar terms do not cause a gulf between the user and the applications. For example, the NDM model deals with full-time equivalent (FTE) positions - but many of the workshop participants habitually thought in terms of numbers of people and needed to be reoriented in terms of FTEs. Many thought that the terminology used in the documentation should be chosen so as to relate as closely as possible to the administrative/management experience of the audience. The need for consistency of computer (keyboard) operations within the applications program itself was explicitly emphasized - i.e. the same control key(s) should do the same thing throughout the program. Many other individuals offered suggestions for enhancing or facilitating the use of the software package.

A major consideration for nearly everyone in the workshop centered about the need for a comparative function available to the user which could, either in tabular or graphical form, contrast a designated baseline data set with a "what if" result. The user's differed in the type of quantitative comparison that was needed (e.g. ratio, difference, percent change, etc.) but were adamant in the fact one was needed. The variety and number of analyses undertaken by the users argue for such a contrast capability, but careful examination of the computational effort needed to support such a contrasting function should be determined before any effort is made to incorporate it into a computer based package.

The last point to be made is perhaps the most important for an organization to consider when contemplating the development of a computer application of this type. Almost every user was of the opinion that even if the entire package - consisting of computer program, manuals and technical documentation - had been made directly available to them, they would not have taken the time to find out if it held anything of value for them. The workshop/tutorial type of introduction was necessary to get them to evaluate the computer based model. While this is perhaps symptomatic of the lack of resources and time individuals in these positions have, it does vividly demonstrate that bringing forecasting (at an analytical level) to one's client base can be extremely resource consuming and an undertaking requiring careful evaluation before such an effort is initiated.

Figure 1
NDM HEALTH CARE SYSTEM SECTOR/SUBSECTOR STRUCTURE

```
    SECTOR                                                  SUBSECTOR

1. HOSPITAL ----------|- SHORT TERM --------|- 1.  INPATIENT
                      |                     |- 2.  OUTPATIENT
                      |                     |- 3.  EMERGENCY ROOM
                      |- LONG TERM ----------- 4.  LONG TERM

2. NURSING HOMES -----|- CERTIFIED ----------- 5.  CERTIFIED
                      |- NON-CERTIFIED ------- 6.  NON-CERTIFIED

3. AMBULATORY ----------- AMBULATORY ---------- 7.  AMBULATORY

4. COMMUNITY HEALTH --|- PUBLIC HEALTH ------- 8.  PUBLIC HEALTH
                      |- STUDENT HEALTH ------ 9.  STUDENT HEALTH
                      |- HOME HEALTH --------- 10. HOME HEALTH
                      |- OCCUPATIONAL HEALTH - 11. OCCUPATIONAL HEALTH

5. OTHER -------------|- NURSE EDUCATION ----- 12. NURSE EDUCATION
                      |- OTHER --------------- 13. OTHER
```

FIGURE 2
NDM HEALTH CARE SYSTEM
SUBSECTOR UTILIZATION DIMENSIONS

	SUBSECTOR	RN/(UNIT OF DEMAND)
1.	INPATIENT	10,000 INPATIENT DAYS
2.	OUTPATIENT	10,000 OUTPATIENT VISITS
3.	EMERGENCY ROOM	10,000 EMERGENCY ROOM VISITS
4.	LONG TERM	10,000 INPATIENT DAYS
5.	CERTIFIED	1000 RESIDENTS (CENSUS)
6.	NON-CERTIFIED	1000 RESIDENTS (CENSUS)
7.	AMBULATORY	10,000 GENERAL POPULATION
8.	PUBLIC HEALTH	10,000 GENERAL POPULATION
9.	STUDENT HEALTH	10,000 SCHOOL AGE POPULATION
10.	HOME HEALTH	10,000 GENERAL POPULATION
11.	OCCUPATIONAL HEALTH	10,000 WORKING AGE POPULATION
12.	NURSE EDUCATION	10,000 NURSING SCHOOL ENROLLMENT
13.	OTHER	10,000 GENERAL POPULATION

	SUBSECTOR	LP/VN/(UNIT OF DEMAND)
1.	INPATIENT	10,000 INPATIENT DAYS
4.	LONG TERM	10,000 INPATIENT DAYS
5.	CERTIFIED	1000 RESIDENTS (CENSUS)
6.	NON-CERTIFIED	1000 RESIDENTS (CENSUS)
7.	AMBULATORY	10,000 GENERAL POPULATION
8.	PUBLIC HEALTH	10,000 GENERAL POPULATION
9.	STUDENT HEALTH	10,000 SCHOOL AGE POPULATION
10.	HOME HEALTH	10,000 GENERAL POPULATION
11.	OCCUPATIONAL HEALTH	10,000 WORKING AGE POPULATION
13.	OTHER	10,000 GENERAL POPULATION

	SUBSECTOR	AIDES/(UNIT OF DEMAND)
1.	INPATIENT	10,000 INPATIENT DAYS
4.	LONG TERM	10,000 INPATIENT DAYS
5.	CERTIFIED	1000 RESIDENTS (CENSUS)
6.	NON-CERTIFIED	1000 RESIDENTS (CENSUS)
10.	HOME HEALTH	10,000 GENERAL POPULATION

FIGURE 3
NDM UTILIZATION OF CARE VARIABLES

1. SHORT TERM HOSPITAL DAYS, AGE < 65
2. SHORT TERM HOSPITAL DAYS, AGE 65+
3. SHORT TERM HOSPITAL OUTPATIENT VISITS
4. SHORT TERM HOSPITAL EMERGENCY VISITS
5. LONG TERM HOSPITAL DAYS
6. LONG TERM HOSPITAL DISPOSITIONS (DISCHARGES)
7. NURSING HOME RESIDENTS

FACTORS AFFECTING THE HEALTH CARE SYSTEM INDEPENDENT VARIABLES

1. ACTUAL DEATHS/EXPECTED DEATHS (65 AND OVER)
2. ACTUAL DEATHS/EXPECTED DEATHS (UNDER 65)
3. ACTUAL DEATHS/EXPECTED DEATHS (TOTAL)
4. ENROLLMENT (IN ALL LP/VN SCHOOLS) PER 10,000 POP.
5. ENROLLMENT (IN ALL RN SCHOOLS) PER 10,000 POP.
6. FRACTION OF NUR. HOME RES. IN NON-CERTIFIED HOMES
7. FRACTION OF NUR. HOME RES. WITH MEDICAID
8. FRACTION OF THE POPULATION WHO ARE UNINSURED
9. LP/VN HOSPITAL OVERTIME PROPORTION
10. LP/VN HOSPITAL VACANCY RATE
11. MANAGED CARE INDEX
12. MANUFACTURING EMPLOYMENT PER CAPITA
13. NURSING HOME COMPLEXITY INDEX
14. PER CAPITA INCOME
15. RN HOSPITAL OVERTIME AND TEMPORARY PROPORTION
16. RN HOSPITAL VACANCY RATE
17. RN NON-HOSPITAL VACANCY RATE
18. RNs PER 10,000 EMERGENCY VISITS
19. RNs PER 10,000 OUTPATIENT VISITS
20. POPULATION
21. (A/DHIED) FRACTION OF RNs DEMANDED WITH A/D HIED*
22. (BACCHIED) FRACTION OF RNs DEMANDED WITH BACC HIED*
23. (M+HIED) FRACTION OF RNs DEMANDED WITH MAST+ HIED*

* THE DISTRIBUTIONS OF RNs BY HIGHEST EDUCATIONAL PREPARATION REPRESENTS THE SITUATION REFLECTED IN THE 1988 NATIONAL SAMPLE OF REGISTERED NURSES AND NOT A RESULT OF DEMAND ESTIMATION.

NURSE EDUCATION RN DEMAND EQUATION

Nurse Education: Dependent Variable: RNs Per 10,000 Population (mean value: 1.210)

Explanatory Variables	Mean Values	Regression Coefficients
Constant		-.190
INCOME	15,875	.426E-4
LPNs ENROLL	1.604	.0852
RNs ENROLL	11.300	.0519*
R^2		.224
n		45

Excludes AK, CA, DC, HI, NJ, and UT

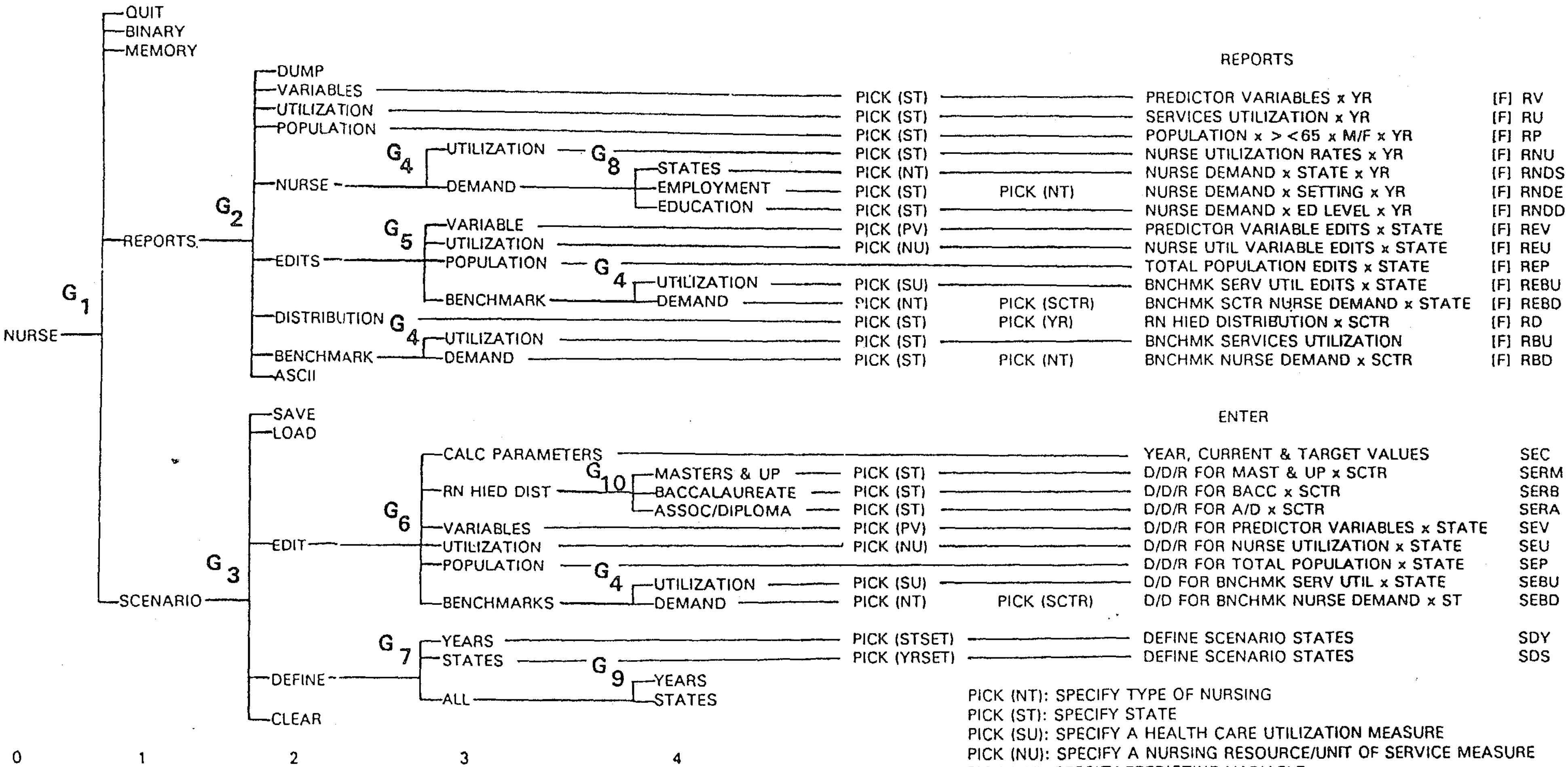
*Significant at 5% Level

FIGURE 4

NURSE DEMAND-BASED MODEL

MENU HIERARCHY

ENDOP SEQUENCES AND CONTENT



PICK (NT): SPECIFY TYPE OF NURSING
PICK (ST): SPECIFY STATE
PICK (SU): SPECIFY A HEALTH CARE UTILIZATION MEASURE
PICK (NU): SPECIFY A NURSING RESOURCE/UNIT OF SERVICE MEASURE
PICK (PV): SPECIFY PREDICTING VARIABLE
PICK (YR): SPECIFY REPORTING YEAR
PICK (SCTR): SPECIFY A HEALTH CARE SECTOR
PICK (STSET): SPECIFY A SET OF STATES FOR SCENARIO DEFINITION
PICK (YRSET): SPECIFY A SET OF YEARS FOR SCENARIO DEFINITION
[F]: PROVISION FOR FILE NAME SPEC FOR REPORT STORAGE
D/D/R: DISPLACEMENT, DILATION AND RATE EDITS
D/D: DISPLACEMENT AND DILATION EDITS

v1.1 09/14/92

FIGURE 5

FIGURE 6

Estimating the Supply of Health Personnel Serving Persons with HIV/AIDS in the United States

Daniel Gordon, New York State Department of Health
William Epple, New York State Department of Health

Introduction

This study, which is being conducted by the Division of Planning, Policy and Resource Development of the New York State Department of Health for the Bureau of Health Professions (BHPr), HRSA, is a first attempt to construct national estimates of the supply of health providers who are serving HIV/AIDS patients. The estimates are being constructed for the base year 1990, and will be used to project a likely scenario for the number of personnel who will be required to provide care in 1996/97.

The project began in September 1991 and will conclude with a report to BHPr in January 1993. This paper reflects the status of the project as of August 1992. The results presented in this paper are preliminary and are subject to change in response to modifications to the methodology and the receipt of additional data.

Personnel estimates are being constructed separately for four sectors of the health care system: hospital inpatient services, primary care, home health care and nursing homes (institutional long term care). Preliminary estimates for the hospital and primary care sectors have been produced and are presented in this paper. The home health care and nursing home estimates are still under construction.

I. Hospital Inpatient Services

Health personnel providing care to HIV/AIDS patients in hospitals in 1990 were estimated using discharge data from state-level hospital discharge data sets, staffing data from selected dedicated hospital AIDS units, staffing and utilization data from the American Hospital Association and AIDS prevalence data from the Centers for Disease Control.

Staff/patient ratios for each state were derived from general hospital data and were adjusted using information gathered from hospital AIDS units. The staff/patient ratios were applied to the state-level HIV/AIDS-related hospital average daily census to produce estimates of the full-time equivalent number of staff providing care to HIV/AIDS patients.

A. Data

State-level data on the number of HIV/AIDS-related hospital days and discharges and the total number of hospital days and discharges for all general acute care hospitals were collected from state Health Departments, Health Care Cost Containment Boards or private vendors (Appendix A). Data were obtained for 14 states which collectively accounted for over 75 percent of the national AIDS case census in 1990. State-level hospital discharge data were also obtained from the Veterans Administration.

HIV/AIDS-related hospital discharges were selected by applying a screen of ICD-9-CM criteria (Table 1) to primary and secondary discharge diagnosis codes in the data sets.

The Centers for Disease Control supplied state-level AIDS prevalence data for 1990. CDC estimates that only 85-90% of AIDS cases are reported and that the lag in transmission of death data caused the number of deaths recorded in 1990 to be undercounted by about 10%. The combined effect is that the true 1990 prevalence figures may be about 5% greater than those supplied by CDC. The CDC data were used without adjustment.

Hospital staff/patient ratios for general patient care were derived using data on the number of full-time equivalent staff by staff categories and the total number of adjusted hospital days from the American Hospital Association's (AHA) 1989 Annual Survey of Hospitals Data Base. AIDS-specific staffing data were obtained from 12 hospitals with dedicated AIDS units. The data collected include the average census of HIV/AIDS patients in the AIDS units, the number of HIV/AIDS patients in other hospital units and the full time equivalent staff associated with these units. The data were used to adjust the staff/patient ratios derived from the general AHA staffing data.

B. Methodology

The process of estimating hospital personnel providing services to HIV/AIDS patients is outlined in Figure 1.
Separate personnel estimates were produced for the staff categories of nurses, aides, therapists, technicians, physician assistants/nurse practitioners, physicians/interns/residents, social and human service workers, and pharmacists.

For states that supplied discharge data, HIV/AIDS community hospital days, Veteran's Affairs HIV/AIDS hospital days and military HIV/AIDS hospital days were summed to reach the total number of HIV/AIDS hospital days. The average HIV/AIDS hospital days per AIDS case for all reporting states was produced by dividing the total reported HIV/AIDS days by the total AIDS cases. To estimate the HIV/AIDS hospital days for non-reporting states, the average HIV/AIDS days per case was multiplied by the number of AIDS cases in each state.

General staff/patient ratios for each state were produced using the AHA survey data by dividing the full-time equivalent staff, by category, by the state's average daily census (hospital days/365).

The AHA data showed a large variation in staff/patient ratios across states. The rates for each staff type were trimmed by removing rates that were more than 50 percent greater than the next highest rate and substituting the highest remaining rate.

Table 1.
Hospital Discharge Diagnosis Screen
for Selecting HIV/AIDS-Related Discharges

Selection Criterion	ICD-9-CM code	Description
P or S	042.0 - 042.9	HIV infection with specified conditions
P or S	043.0 - 043.9	HIV infection causing other specified conditions
P or S	044.0 - 044.9	Other HIV infection
P or S	795.80	Positive serological or viral culture findings for HIV
P or S	279.10	Immunodeficiency with predominant T- cell defect, unspecified
P or S	279.19	Deficiency of cell-mediated immunity, other
P only	279.30	Unspecified immunity deficiency
P only	136.30	Pneumocystosis
Selection Criteria	- P: Primary diagnosis - S: Secondary diagnosis	

The 12 AIDS unit staffing reports indicated that staffing is more intensive than is reflected in the staff/patient ratios for the general hospital population. An adjustment factor for each staff type was derived by comparing the distribution of staff/patient ratios among the AIDS units to the national general staff/patient ratio. As a result, the number of social and human service workers serving HIV/AIDS patients was increased by a factor of 2.8, the number of physician assistants/nurse practitioners by a factor of 4, and the number of physicians by a factor of 1.5.

For each state the estimated HIV/AIDS hospital census was multiplied by the adjusted staff/patient ratios to yield the estimated number of staff. The state figures were combined to estimate the total number of full-time equivalent hospital staff providing inpatient care for patients with HIV/AIDS in the United States in 1990.

C. Assumptions

The estimation of HIV/AIDS-related hospital days in states without hospital discharge data is based on an assumption that the ratio of HIV/AIDS hospital days to AIDS cases is comparable to the average days/cases ratio in the states for which data were obtained. Variations in medical practice patterns, the availability of hospital versus alternative services and differences in the HIV-ill patient population of the states may cause this ratio to differ appreciably by state.

We assumed that staff/patient ratios derived for the general hospital patient population are applicable to HIV/AIDS patients, unless specific data are available to justify using a different ratio. Information on AIDS unit staffing was used to adjust the staff/patient ratios for physicians, physician assistants/nurse practitioners and social workers. Certain types of staff, such as technicians and therapists, serve patients in AIDS units but are not specifically attached to the units, and as a result are not included in the units' staffing reports.

We also assumed that staff hours of patient care for HIV/AIDS patients outside dedicated units are the same as for those in those units. One study (Van Servellen, 1990) indicates that nursing hours per patient are lower in integrated units than in dedicated units.

D. Data Issues

The ICD-9-CM screen (Table 1) that was used to extract HIV/AIDS-related hospital discharges accepts codes 136.3 (Pneumocystosis), 279.10 (Immunodeficiency with predominant T-cell defect, unspecified) and 279.19 (Deficiency of cell-mediated immunity, other) when these occur as primary diagnoses. These conditions may also occur in persons who are not HIV-infected. Analysis of California (DATIS,1990) and New York data showed that deleting these codes from the screen would reduce the number of discharges identified as HIV-related by under 5 percent. We also analyzed diagnosis codes of New York discharges who had a 136.3 diagnosis but no HIV-related code and found that about half of them had diagnosis codes related to conditions like drug use and candidiasis, which are common among HIV-ill patients.

Inaccuracies in the diagnosis codes on the discharge records can lead to the inclusion of non-HIV discharges ("false positives") or to the exclusion of true HIV-related discharges ("false negatives"). A 1988 study by the New York State Health Department found that for over 95 percent of records in the statewide hospital discharge data set that had HIV-related diagnosis codes, the associated hospital medical records confirmed that the patient was HIV-positive or HIV-ill (Smith, 1990). More recent studies in other states have found a somewhat higher rate of false positives, and show a substantial rate of false negatives (Hidalgo, 1992). We have not attempted to correct the hospital discharge data for this misreporting.

E. Preliminary Hospital Estimates
There is a large variation in the average ratio of HIV/AIDS hospital days per AIDS case in the reporting states. The reported hospital days per case vary from a low of 19.5 days per case for Washington State to a high of 67.4 for New York.

The 2.5 million days that HIV/AIDS patients spent in hospitals in 1990 translate into an average daily census of 6,800 patients. Their care was provided by the full-time equivalent of 1,500 physicians (including interns and residents), 7,900 nurses, 360 physician assistants/nurse practitioners, 2,700 technicians, 270 pharmacists, 580 therapists and 2,600 aides.

Overall, the provision of service to HIV/AIDS patients in hospitals in 1990 consumed the services of between 16,500 and 17,700 health professionals (depending on whether New York's hospital days per AIDS case are included in the calculation). The greatest proportion of these, about half, was made up of nurses. Technicians and aides each made up about a sixth of the total. The small number of estimated physician assistants/nurse practitioners is due to the limited or non-existent use of these professions in some states and in many hospitals.

II. Primary Care Personnel - Preliminary Estimates

Health personnel providing care to HIV/AIDS patients in primary care settings were estimated using **scenarios** of care. The scenarios, which were developed with the assistance of an advisory group of clinicians and clinic administrators, are representative patterns of care, and indicate the **visit frequency** with which HIV/AIDS patients receive primary care, and the staff time associated with each of the visit types. Two scenarios were constructed, representing high and low intensities of service use. Each scenario contains different visit frequency and staff time patterns for patients in each of four levels of illness.

Fractions of the estimated total U. S. HIV-infected population were assigned to each scenario, and the total staff time required to provide the services was calculated. The total staff times were converted to full-time equivalent staff using estimates of worker productivity that allow overhead for time spent outside patient contact.

According to these estimates, approximately 287,000 adults HIV/AIDS patients were receiving regular primary care in 1991, and their care was provided by the full-time equivalent of about 2,500 health personnel.

A. Methodology
For the purposes of this project, primary care refers to general medical care provided in an outpatient setting such as a physician's office, hospital-based or freestanding clinic or health center. Emergency room care that does not lead to an inpatient hospital admission is also included.

The primary care personnel estimates are based on a set of **scenarios** of care. The scenarios are a set of likely or plausible patterns of care, and indicate the frequency of service delivery (visits) for various types of primary care services and the staff time associated with those visits.

The framework of the scenarios was developed under the guidance of an advisory board consisting of providers and administrators at sites providing primary care to HIV/AIDS patients.

The scenarios divide HIV-infected primary care patients into four groups. Each of these groups was defined by a range of CD4 counts to help respondents in the survey visualize the service patterns of the groups. The CD4 ranges were also used in a later stage of the estimates to determine the number of people in each group. The four groups are asymptomatic (CD4 count over 500), early symptomatic (CD4 count 500-200), symptomatic (CD4 count 200-50) and late symptomatic (CD4 count under 50).

The scenarios address six types of service: initial evaluation and screening, monitoring visits, illness visits, emergency room illness visits, routine gynecological visits, and telephone consultation.

For each type of service, the scenarios specify the frequency of the service and the associated staff time. The staff included in the scenarios are physicians, physician assistants/nurse practitioners, nurses and social workers. Other types of staff are also involved in the provision of primary care, but only the four classes above appeared with regularity in the staff cited by respondents. Phlebotomists, pharmacists, dieticians were also mentioned frequently by the respondents, but not regularly enough include them in the staff categories. Formal case management workers are excluded, although it was clear from the respondents' comments that this function was frequently performed by social workers or nurses.

The staff times specified in the scenarios of care were combined with estimates of the number of persons in each patient group who received care to yield and estimate of total annual staff time for each staff type. The derivation of the number of patients is discussed below. The total staff time was then divided by the estimated productivity (hours worked per year, with an adjustment for administrative overhead) to yield the full-time equivalent staff providing primary care.

B. Data

The frequency of visits and the staff time associated with them were derived through a series of structured interviews with primary care clinicians and administrators. Information was obtained from 10 primary care practice sites in 7 states. Following the guidance of the advisory group, the survey results were arrayed into two scenarios (Appendix A).

The CDC estimates that approximately 1,000,000 million persons are infected with HIV in the United States. The proportion of the infected who have CD4 counts in the ranges used in the scenarios were based on a report by Brundage et al (1990). Brundage does not break the population with under 200 T cells into 200-50 and under 50 groups. For the purposes of these estimates, it was assumed that two thirds of the under 200 groups fall in the 200-50 range, and the remainder in the under 50 T cell range.

C. Assumptions

The primary care scenarios indicate the pattern of care received by individual patients, but to estimate the overall staff time involved it is necessary to estimate the overall number of patients who are in treatment. The number was calculated by developing a **participation rate** for infected persons in each T cell range. The participation rate is the fraction of persons who are in treatment at one point in time.

The participation rates cannot be measured directly, but they can be inferred from the T cell distribution of patients in treatment. If the participation rate is known, or is assumed, for persons in one of the four T-cell ranges, it can be calculated for the others, using a comparison of the number of infected persons, and the fraction of persons in treatment who fall into a given T-cell range. In order to do this we assumed that the participation rate for persons with CD4 counts under 50 is 90 percent.

The overall result of the patient estimation process shows that of the million persons believed to be infected, about 29 percent are active primary care patients at any given time (Table 2).

Table 2
Infected Persons and Number in Treatment, Estimated
United States, 1991

Group	CD4 Range	Total Persons	New Number in Treatment	Patients per Year
Asymptomatic	>500	390,000	67,744	16,523
Early Symptomatic	500-200	430,000	106,634	10,819
Symptomatic	200-50	120,000	58,369	5,085
Late Symptomatic	<50	60,000	54,000	3,573
Total		1,000,000	286,746	36,000

To convert staff time to full-time equivalents, it was assumed that staff work 8 hours per day for 48 weeks per year, and that administrative overhead was equal to 20 percent of the patient visit times.

D. Results

For these estimates the total estimated number of patients in treatment was allocated evenly between the two scenarios, and total staff times were calculated. The preliminary results of this method are that in 1991, primary care was provided to an average census of 287,000 patients by the full-time equivalent of 850 physicians, 520 physician assistants and nurse practitioners, 740 nurses, and 380 social workers (Table 3).

Table 3
Estimated Full-Time Equivalent (FTE) Selected Health Personnel
Providing Primary Care to HIV/AIDS Patients
United States, 1991

	Total Staff (FTE)	Staff/ Patient Ratio
Physicians	849	1/338
Physician Assistants/Nurse Practitioners	520	1/552
Nurses	737	1/339
Social Workers	377	1/760
Total estimated patients in treatment	286,746	

This preliminary estimate indicates that there was one full-time equivalent physician employed in providing care to every 338 patients. In contrast, crude staff and patient data from the clinics surveyed suggested a ratio of 450 - 675 patients per physician. The clinics reported a range of one nurse for every 166 - 450 patients, and our estimate is one nurse per 339 patients.

The visit frequencies in the scenarios yielded a total of 3.9 million primary care visits, or an average of 13.5 visits per patient per year. Data from two of the clinics and from the California Department of Health Services AIDS longitudinal file suggest a utilization rate in the range of 8 - 12 visits per year for patients who are in treatment.

III. Projections

A. Infected Population
The New York State Dept. of Health Bureau of Disease Control is currently working on a model that will estimate the T cell distribution of the population over time. This model uses a back calculation method (Brookmeyer and Damiano, 1989; Rosenberg et al, 1991,) but also incorporates data from other sources on the time course of T cell decline (Longini et al., 1991,; Brookmeyer and Liao, 1990). The results of this model will be used in the personnel estimates when they become available.

B. Hospital Care
The future use of inpatient hospital services will be affected by changes in the number of infected persons at various stages of illness and by medical practice patterns. In recent years, increasing use and effectiveness of outpatient services have caused a small but measurable decrease in AIDs/HIV hospital utilization. While this trend may continue, it is also possible that the movement of the epidemic into new population groups, and the emergence of new disease entities, such a drug-resistant tuberculosis, may tend to move the utilization rate back up again. Therefore, we will base our projections of hospital service utilization solely on the epidemiologic model described above, using the change in the numbers of HIV-infected persons at the lower CD4 levels to indicate the change in the population using hospital services.

C. Primary Care
Future use of services will depend on the number and CD4 distribution of the infected population, their participation rate and on the pattern of service use for patients in treatment.

Our survey respondents indicated that their patient populations are growing rapidly. This is due to the increasing number of HIV-ill persons, but it is also due to outreach programs and better coordination among providers, which increases the proportion of infected persons who are in treatment. Other factors, such as new treatment options or changes in reimbursement arrangements will also affect overall provision of primary care, but we will not attempt to model them in these estimates.

The projections will be based on the epidemiologic model described above, but will also incorporate changes in the participation rates if a consensus is reached among the project advisors on the size of the change.

IV. Data Issues

This project has encountered difficulties in assembling sufficient data to compile the personnel estimates in each of the four sectors. The data deficiencies relate to the quantity of health services utilization among HIV/AIDS patients and the relation between quantities of health services and the number of health services personnel involved in providing service.

The data issues appear to be tractable in the hospital sector. Most states have hospital discharge data bases, and while the diagnosis coding the identifies an HIV-related hospital stay is not always entirely reliable, the data appear to be basically sound enough for the purposes of this kind of estimate. The personnel involved in the provision of hospital services can be inferred from the staffing patterns of dedicated AIDS units, bearing in mind that not all HIV/AIDS-related hospital stays occur in such units and that the intensity of personal service may be substantially less outside the dedicated units.

In the primary care estimates the scenario approach was chosen in anticipation that comprehensive utilization information would be unavailable. Utilization information, in the form of simple visit counts, is available for a few select patient populations, such as Medicaid recipients, but it is difficult to link the visits to the staff who actually provide the services. Such data can be used to corroborate the results of the scenario approach but cannot form a basis for the estimates themselves.

Home health agencies frequently lack the data systems that can describe in detail the patterns of care that their HIV/AIDS patients receive. Even where adequate automated data systems are present, they are not linked together into a central data set, as in the case of the hospitals. The patterns of home health care for HIV/AIDS patients vary enormously from one location to another, and even from agency to agency within a city. This makes it hazardous to extrapolate from one agency's experience to another's and requires that many agencies be polled in order to assemble a comprehensive picture of home care in an area. The Maryland Department of Health's HIV Information System has developed a capacity to profile the health services received by HIV/AIDS patients in that state. Similar efforts could be put in place in other states, but for general policy-making and administrative purposes it would be more effective to encourage the development of home health care agencies' own reporting capabilities.

It is proving extremely difficult to produce a count of the number of AIDS patients receiving nursing home care. Centralized nursing home data sets do exist in some states, but the diagnostic information needed to identify these patients is not collected. Nursing home care for AIDS patients is not yet common, although designated units exist in several states and admissions to non-designated long term care beds may be getting easier to obtain as well. For this reason it is not possible to extrapolate from the experience of states that can be documented to those that cannot.

References

Brookmeyer, R. and A. Damiano <u>Statistical Methods for Short-Term Projections of AIDS Incidence</u>. Stat. in Med. 1989;8:23-34

________, and J. Liao. <u>Statistical Modeling of the AIDS Epidemic for Forecasting Health Care Needs</u>. Biometrics 1990;46:1151-1163.

Brundage, J. F., et. al. <u>The Current Distribution of CD4+ T-Lymphocyte Counts among Adults in the United States with Human Immunodeficiency Virus Infections: Estimates Based on the Experience of the U.S. Army</u>. JAIDS 1991;3:92-94

DATIS Corporation, San Mateo, Cal. Personal Communication, 1992.

Hidalgo, J. Personal communication. 1992

Longini, I.M., Jr., W.S. Clark, L.I. Gardner and J. F. Brundage. <u>The Dynamics of the CD4+ T-Lymphocyte Decline in HIV-Infected Individuals: A Markov Modeling Approach</u> JAIDS 1991;4:1141-1147.

Rosenberg, P.S., R.J. Biggar, J. J. Goedert and M. H. Gail <u>Backcalculation of the Number with Human Immunodeficiency Virus Infection in the United States</u>. Am. J. Epidemiology 1991;133(3)285.

Smith, P. Internal Report, Bureau of Communicable Diseases, New York State Dept. of Health, 1990

Van Servellen, G. <u>Quality and Cost of AIDS Nursing Care as a Function of Inpatient Delivery Systems</u>. J. Nurs. Admin. 1991;21(9)32-8.
(Brookmeyer and Damiano, 1989, Stat. in Med.

Appendix A
Example of a Completed Scenario

Patient Group	Service Type	Visits per Year	Staff Time per Visit (Minutes)			
			MD	PA/NP	Nurse	Social Worker*
Asymptomatic (CD4 >500)						
	Init. Eval. & Scrn.**	20	45	30	30	
	Mon. & Proph.	4	15	15	20	30
	Illness	2	15	15	15	
	Emergency Rm. Illness	0				
	Gynecology	2	20		20	
	Phone Consult.***		20	20	20	
Early Sympt. (CD4 500-200)						
	Init. Eval. & Scrn.		30	50	30	30
	Mon. & Proph.	6	15	20	20	30
	Illness	3	15	15	20	
	Emergency Rm. Illness	0				
	Gynecology	2	20		20	
	Phone Consult.		30	30	30	
Symptomatic (CD4 200-50)						
	Init. Eval. & Scrn.		35	60	30	45
	Mon. & Proph.	12	20	30	20	10
	Illness	6	20	20	20	15
	Emergency Rm. Illness	2	40	15	20	
	Gynecology	2	20		20	
	Phone Consult.		50	50	50	
Late Sympt. (CD4 <50)						
	Init. Eval. & Scrn.		45	60	30	45
	Mon. & Proph.	16	20	30	20	7.5
	Illness	12	20	20	20	7.5
	Emergency Rm. Illness	4	40	15	20	7.5
	Gynecology	2	20		20	
	Phone Consult.		120	120	120	

* In some cases, social worker contact only occurs in a portion of the visits within a service class. The reported figure is the average across all visits.
** Initial evaluation and screening occurs only once for a patient within a treatment site, and therefore no frequency is attached.
*** Estimated telephone consultation time is for the entire year.

Energy Modeling: The National Energy Modeling System (NEMS)

In two sessions, members of the Office of Integrated Analysis, Energy Information Administration, U.S. Department of Energy, provided overviews of the supply and demand components of the National Energy Modeling System (NEMS), directed by Ronald Earley. The following paper by Susan Shaw provides a synopsis of the two sessions. The discussion topics included: Design Objectives for the Sectoral Demand Modules; Macroeconomic and International Modules; Integrating Framework; The NEMS Oil and Gas Modules; The NEMS Electricity Market Module; and The NEMS Coal Market Module. The following participated in these sessions, Daniel Butler, Ronald Earley, Bob Eynon, Edward J. Flynn, James M. Kendell, Susan H. Shaw, and Scott Sitzer.

Overview of the National Energy Modeling System

Introduction

The Energy Information Administration (EIA) of the Department of Energy (DOE) has responsibility for the collection, analysis, and publication of data related to energy and energy's relationship to the economy. Part of the mission of EIA, as stated in the legislation creating DOE, includes the mandate to produce an annual report of long-term energy trends. Since 1974, EIA and its predecessor organization, the Federal Energy Administration, have fulfilled this mandate with a series of computer modeling systems representing domestic energy-economy markets and the projected trends. The major uses of these models include the annual reports of energy projections, currently titled the *Annual Energy Outlook*, and special studies requested by policy analysts and decisionmakers, such as the DOE Office of Domestic and International Energy Policy and the U.S. Congress.

From 1989 to 1991, DOE prepared the National Energy Strategy (NES), which contained long-term policy recommendations and analyses. The DOE Office of Policy, Planning and Analysis used a collection of existing models to support the development of the NES, including EIA's Intermediate Future Forecasting System (IFFS). IFFS was developed in the early-1980's and has been used since for the *Annual Energy Outlook* and related studies; however, IFFS alone could not support the analysis of many energy issues of the 1990's. Within the Department, the need was identified for a new comprehensive modeling system. This effort began in 1990 with a task to the National Research Council of the National Academy of Sciences to form a committee to review existing energy models and provide guidance on the development of a National Energy Modeling System (NEMS).

System Design of NEMS

The purpose of NEMS is:

> To illustrate the energy, economic, environmental, and energy security consequences on the United States of various energy policies and assumptions by providing forecasts of alternative energy futures in the mid and long-term periods, using a unified modeling system.

As its predecessor models, NEMS incorporates a market-based approach to energy analysis. NEMS balances the supply of and demand for energy for each fuel and consuming sector, taking into account the economic competition between energy sources. This competition is the foundation for most of the analysis performed by EIA.

NEMS is also designed as a modular system. The modules of the energy system represent each of the fuel supply markets, conversion sectors, and end-use consumption sectors. NEMS also incorporates interactive macroeconomic and international modules. The primary flows between these modules are the delivered prices of energy and the quantities consumed by product, region, and sector. The information flows are not, however, limited to prices and quantities and include other information such as economic activity, capital expenditures, and the impacts of demand-side management programs. The delivered prices of fuel encompass all the activities necessary to produce, import, and transport fuels to the end user.

The integrating methodology controls the independent execution of the component modules. The modules are executed from the integrating module, and, to facilitate modularity, the components do not pass information to each other directly but communicate through a central data file. This modular design provides the capability to execute modules individually or to substitute alternative modules, thus allowing decentralized development of the system and independent analysis and testing of individual modules. Furthermore, this modularity allows the flexibility to use the methodology and level of detail that is most appropriate to represent each energy sector.

Solution is achieved by equilibrating on the delivered prices of energy and quantities demanded, thus assuring an economic equilibrium of supply and demand in the consuming sectors. Each fuel supply, conversion, or end-use demand module is called in sequence by the integrating module and solves assuming all other variables in the energy markets are fixed. The modules are iteratively called until the end-use prices and quantities remain constant within a specified tolerance, a state defined as convergence. This equilibration is achieved annually through the midterm period to 2015.

The algorithm also checks for convergence on variables that represent petroleum product imports, crude oil imports, and several macroeconomic indicators.

Attributes of NEMS

A number of characteristics define the overall structure of NEMS and distinguish the system from its predecessor models. First and foremost, more of the energy market factors that were formerly analyzed offline are represented endogenously within NEMS, for example, the international oil market and the penetration of renewable energy sources. Also, the sectors incorporate greater structural detail, such as an embedded refinery model and fuel transportation networks in several modules. These enhancements are described in more detail in the later section on the individual modules.

Time Horizon

NEMS is planned to have both a midterm and a long-term modeling capability. The midterm model is the focus both of this paper and of the initial development effort. The horizon for the midterm model is 2015, covering that time period in which the structure of the economy, the nature of energy markets, and regional demographics are sufficiently well known that it is possible to represent considerable structural and regional detail. The majority of policies which are proposed today will have their greatest impacts during the midterm years.

The world has the potential to change dramatically over the next 40 years, with issues relating to capital investment and the penetration of new technologies, research and development programs, structural changes in the economy, the availability of energy resources, demographic shifts and immigration, shifts in transportation modes and manufacturing sectors, global trade issues, and environmental impacts. The long-term modules will require only a level of detail that is necessary to address the key issues, which in many cases will be different from that of the midterm modules. It will be a separate and distinct modeling capability from the midterm model. For the long-term model, EIA will develop a less-detailed representation of energy markets with the capability to analyze those energy issues with long-term impacts, such as technology penetration, resource depletion, and long-term economic growth, through a time horizon 40 to 50 years in the future.

Regional Structure

For many issues of interest in energy analysis, it is not sufficient for NEMS to represent the United States as a single, homogeneous region, but rather consider some regional structure. The diverse nature of energy supply, demand, and conversion in the United States and the desire of energy policy makers and analysts for information about specific parts of the country make it necessary for NEMS to support energy analysis at a regional level.

Regional representations are incorporated in an energy modeling system for several reasons: to portray transportation flows; to represent the regional differences in energy markets, such as proximity to supplies or differences in the infrastructure or demographics; and to provide impacts at the regional level. The definition of the regions, however, depends on many factors, among them are the analytic requirements for regional results, data availability, and computational considerations.

The level of regional detail for the demand sectors and for the integrating module is the 9 Census divisions. Other regional structures specific to the conversion and supply sectors are:

- Oil and Gas Supply - 12 supply regions, including 3 offshore and 3 Alaskan regions,
- Natural Gas Transmission and Distribution - 12 regions, based on the 9 Census divisions with some further subdivisions to represent key transportation issues,
- Coal - 16 supply regions, representing the major coal geologic formations,
- Electricity - 13 North American Electric Reliability Council regions and subregions,
- Refinery - 5 Petroleum Administration for Defense Districts.

Transformations between these regional structures and the central 9 Census division structure are accomplished by explicit mapping and sharing algorithms or by representations of aggregate transportation networks.

Foresight

Several of the NEMS modules require assumptions of future prices and demands for energy in order to make capacity expansion decisions. NEMS incorporates the capability to impose centralized control over these assumptions. Three alternative options for centralized foresight, which can be specified by the analyst, are:

- Myopic - assuming, within any forecast year, that the current prices will remain constant into the future for capacity expansion decisions,

- o Extrapolative - assuming, within any forecast year, that current prices and demands grow at a specified rate, and

- o Perfect - ensuring that the future prices and demands upon which decisions are based are the same as those realized by the model.

In addition, recognizing that there is valid evidence that decisionmaking varies by sector, an option allows each sector to use those assumptions deemed most appropriate for that sector. To accomplish perfect foresight, the normal year-by-year execution order is revised.

Reduced-Form Modules

The modular structure of NEMS allows for the easy substitution of alternative modules, as long as the inputs and outputs are aligned. The larger modules of NEMS will have reduced-form versions that simulate the aggregate response of the full module. A reduced-form model is a smaller version of a larger model, providing the same type of primary results as the larger model. The purpose of a reduced-form model is the estimation of results that are similar in magnitude with less computational resources.

One method for constructing a reduced-form model is the response-surface method in which the larger model is run over a range of values for a set of independent variables, yielding a set of model results. A reduced-form model can then be derived by a statistical estimation or interpolation of the results with respect to the independent variables. If the larger model is relatively smooth and if the sample of values for the independent variables is a fine enough grid, then the reduced-form model should yield results similar to those of the larger model within the domains of the values chosen for the independent variables.

Another method is non-parametric approximation. Again, the model is executed over a range of values for independent variables. These results then becomes the database for the reduced-form representation. The module uses those results where the occurrences of the independent variables most closely represent the current state of the model variables and computes a combination of the results of those occurrences.

Finally, a reduced-form model can be constructed by specifying a smaller, structural model. This smaller model would contain equations that represent engineering or economic concepts similar to the larger model; yet, it may be smaller in size by the aggregation of some dimensions of detail, such as product, regional, or sectoral detail. The primary difficulty inherent in this approach is the problem of maintaining reasonable consistency between the two models.

Uncertainty

The need for measures of the uncertainty associated with energy projections is frequently raised. For NEMS, alternative methods to quantify and analyze the uncertainty in the major model outputs is under research. This is not intended to explore all sources of uncertainty, such as unanticipated future events, but to consider the uncertainty associated with key model input parameters and assumptions and with critical model relationships.

Two possible techniques are under consideration, each requiring the identification of the inputs, parameters, and assumptions that drive the principal model results. An efficient fractional sampling plan will be devised and applied to quantify how uncertainties in the inputs translate to uncertainties in the outputs. Since NEMS is expected to require substantial computer resources to solve, the initial approach will be to develop and test the techniques on the individual modules of NEMS. The screening approaches and sampling plans will be tailored to the different module methodologies in order to take advantage of special features of the module structure. After developing suitable reduced-form modules for NEMS as necessary, the possibility of characterizing uncertainty in the integrated NEMS system, using the reduced-form modules, will be investigated.

By incorporating some measure of uncertainty into the analysis, it will be possible to develop error bounds around key model outputs, as well as some measure of the robustness of policy options.

Environment

Recognizing the importance of environmental issues associated with the use of energy, NEMS includes an environmental capability. Six emissions are accounted for in NEMS: SO_x, NO_x, CO, CO_2, carbon, and volatile organic compounds. These emissions are computed for energy production activities and fuel combustion. In addition, NEMS represent all current environmental regulations, such as the Clean Air Act Amendments of 1990. It is important to note that even though accounting is not incorporated for all possible toxic substances, all costs of regulating those toxics are included in NEMS.

NEMS also incorporates a capability to constrain emissions, an important feature for future policy analysis. Constraints will be implemented by raising the cost of producing emissions until the total system emissions are reduced to the level of the constraint.

Components of NEMS

The components of NEMS are modules representing individual sectors of domestic energy markets, an international energy module, and a macroeconomic activity module, directly interacting with one another through the integrating model. In general, the modules interact through values representing the prices of energy delivered to consuming sectors and quantities of end-use energy consumption. The international and macroeconomic modules provide information on international energy market activities that impact domestic energy activities and information on macroeconomic conditions related to activities in energy markets.

All NEMS components represent the impact and cost of environmental regulations that affect the sector and report emissions and other environmental impacts to a centralized location. The following descriptions summarize the key features of each of the modules of NEMS with an indication of the major enhancements over the current modeling capabilities.

Macroeconomic and International Components

Macroeconomic Activity Module. The macroeconomic activity module provides a set of essential macroeconomic drivers to the energy modules, provides a macroeconomic feedback mechanism within NEMS, and provides a mechanism to evaluate detailed macroeconomic and interindustry impacts associated with energy events. Industrial drivers are calculated for 35 industrial sectors. A capability to analyze the impacts of energy investment is included, as well as regional macroeconomic projections. This module is a response surface representation of the Data Resources, Inc., (DRI) Quarterly Model of the U.S. Economy.

International Module. The international module represents the world oil markets and the world oil prices endogenous to NEMS. International petroleum product supply curves, including curves for oxygenates, are incorporated, and an international refinery model is being added. This module defines crude oil categories that are consistent with those in the domestic refinery model.

Supply Components

Oil and Gas Supply Module. The oil and gas module represents domestic crude oil, natural gas liquids, and natural gas production within an integrated framework that captures the interrelationships between the various sources of supply: onshore, offshore, Alaska, conventional, and unconventional production. This framework analyzes cash flow and profitability to compute investment in each of the supply sources. Oil and gas market equilibration for production is computed at a regional level. The crude oil and natural gas liquids produced are input to refineries, a separate conversion module in NEMS, for conversion and blending into refined petroleum products.

Natural Gas Transmission and Distribution Module. This module represents the transmission, distribution, and pricing of natural gas, subject to end-use demand for natural gas, the production of domestic natural gas, and the availability and price of natural gas traded on the international market. The module tracks the flows of natural gas in an aggregate, domestic pipeline network. This capability allows the analysis of impacts of regional capacity constraints in the interstate natural gas pipeline network and the identification of pipeline capacity expansion requirements. There is an explicit representation of firm and interruptible markets for natural gas transmission services, and the key components of pipeline and distributor tariffs for transmission services are included for the pricing algorithms.

Coal Supply Module. The coal module represents the mining, transportation, and pricing of coal, subject to the end-use demand for coal differentiated by physical characteristics, such as the heat and sulfur content. The coal supply curves include a response to capacity utilization and fuel costs. Both the numbers of regions and of coal categories have been streamlined from previous modeling efforts. Additional transportation modes, such as trucks, are added. Coal transportation rates are reconstructed using more recent data, and the model includes a more comprehensive treatment of coal contracts.

The coal module incorporates an international component to calculate U.S. coal exports as part of the worldwide market for coal trade. A coal synthetics submodule is also included which evaluates the economics of the production of synthetic fuels, relative to conventional liquids and gases.

Uranium Supply Module. The uranium supply module calculates a levelized fuel cost of uranium fuel for nuclear generation which is directly incorporated into the electricity market module.

Renewable Supply Module. The renewable supply module includes several submodules, providing explicit representation of the supply of wood, municipal solid waste, wind, solar, hydropower, and geothermal technologies. The market penetration of renewable technologies used for centralized electricity generation is represented in the electricity module. The market penetration of dispersed renewables is incorporated within the end-use demand modules. Renewable supply curves from the renewable supply module provide costs and performance criteria to the modules. The renewables module also interacts with the refining module to represent the production and pricing of alcohol fuels.

Conversion Components

Electricity Market Module. The electricity module represents the generation, transmission, and pricing of electricity, subject to the delivered prices for coal, petroleum products, natural gas, and synthetic fuels, the costs of generation by centralized renewables, macroeconomic variables for costs of capital and domestic investment, and electricity load shapes and demand. The submodules include capacity planning, fuel dispatch, nonutility generation, finance and electricity pricing, transmission and trade, and demand side management (DSM) in conjunction with the demand models.

DSM programs and all Clean Air Act compliance options are explicitly represented in the capacity expansion and dispatch decisions. Both new generating technologies and renewable technologies compete directly in these decisions. Several options for wholesale pricing and the competition between utility and nonutility generation are included in the module.

Petroleum Market and Refinery Module. The refinery module includes the pricing of petroleum products, crude oil and product import activity in conjunction with the international module, and domestic refinery operations, subject to the demand for petroleum products, the availability and price of imported petroleum, and the domestic production of crude oil, natural gas liquids, and alcohol fuels. The module represents 5 crude oil types in the refining activities as does the international module. It explicitly models the requirements of the Clean Air Act Amendments of 1990 and the costs of new automotive fuels, such as oxygenated and reformulated gasoline, and includes oxygenated production and blending for reformulated gasoline. Costs include required capacity expansion for refinery processing units.

Demand Components

Buildings Sector Demand Modules. The residential module forecasts the consumption of residential sector energy by housing type and end-use, subject to delivered energy prices, the availability of renewable sources of energy, and macroeconomic variables representing disposable personal income, interest rates, and housing starts. The commercial module forecasts the consumption of commercial sector energy by building types and nonbuilding uses of energy and by category of end-use, subject to delivered prices of energy, the availability of renewable sources of energy, and macroeconomic variables representing Gross Domestic Product (GDP), employment, interest rates, and floorspace construction.

Both modules incorporate expanded assessments of advanced technologies, including endogenous representations of renewable energy technologies. Enhanced analyses of both building shell standards and new end-use services are included. Demand side management programs are incorporated in conjunction with the electricity market module.

Industrial Sector Demand Module. The industrial module forecasts the consumption of energy for heat and power and for feedstocks and raw materials, subject to the delivered prices of energy and macroeconomic variables representing GDP, interest rates, employment and labor cost, and the value of output for each industry. Simplified process models represent the use of energy for approximately 35 specific industries, with the capability to examine the boiler/steam, buildings, and process/assembly uses. A representation of cogeneration and a recycling component are also included.

Transportation Sector Demand Module. The transportation module forecasts the consumption of transportation sector fuels and electricity by transportation mode, including the use of renewables and synthetic fuels, subject to delivered prices of energy fuels and macroeconomic variables representing disposable personal income, GDP, population, interest rates, and the value of output for industries in the freight sector.

Fleet vehicles are represented separately to allow analysis of the Clean Air Act Amendments and other legislative proposals, and both the number of vintages and the number of size classes are increased. The analysis of the penetration of alternatively-fueled vehicles is endogenous to the model, and a capability for intermodal shifts has been added.

System Access

Initially, NEMS is being developed on the EIA mainframe computer with enhanced tools for providing results to users for personal computer analysis. Each individual module will be executable on a personal computer. The file-based data structure and the modularity of the system provide the capability for each module to be easily detached from NEMS for wider distribution, review, and use by the external community.

A system design is under development for a user interface that will allow for the execution of a module and include data review and editing features and display of results. In addition, research is underway on the feasibility of executing the entire system on an advanced personal computer, in which case the user interface would be extended for the system.

Development Schedule

Background work for the design of NEMS was accomplished in 1991 by a NEMS Project Office. Late in 1991, EIA reorganized to form the Office of Integrated Analysis and Forecasting with the mission of developing and maintaining NEMS and conducting all forward-looking analysis in EIA. This reorganization facilitates the coordinated model development effort.

Design and development plans were communicated in a series of 39 component design reports that were prepared prior to model implementation. These reports received wide distribution to the internal and external energy analysis community and were the subject of formal review through EIA's Independent Expert Review program. Additional guidance is received through a NEMS User Group, including representatives of Government agencies, industry trade associations, Congressional organizations, and environmental groups, an Energy Modeling Forum review group, and a public NEMS Conference in February 1993, presenting model designs and methodologies.

The development of NEMS proceeds in two phases. The Phase I, or prototype, NEMS was developed in 1992, establishing the basic model structures and implementing the linkages between the modules. The Phase II NEMS will be used for the *Annual Energy Outlook 1994* (*AEO94*), for which NEMS will be fully tested and evaluated during the Summer of 1993. Development of the *AEO94* projections using NEMS will occur in September 1993; however, further enhancements to NEMS will continue during and beyond the *AEO94* effort. Reduced-form versions of NEMS modules will be implemented in Spring 1994, and the long-term model is planned for the end of 1994.

The U.S. Bureau of Mines Metal Industry Indicators - Explaining Cyclical Forecasting to a Nontechnical Audience

George Swisko and Patricia R. Devine
U.S. Bureau of Mines, Department of the Interior

Introduction

In 1989 U.S. Bureau of Mines staff asked Dr. Geoffrey Moore, the Director of the Center for International Business Cycle Research (CIBCR) at Columbia University, to examine the feasibility of constructing coincident and leading indexes to predict the performance of major sectors of the U.S. metals industry. Dr. Moore has been recognized for many years as one of the world's leading authorities and researchers in the behavior and measurement of business cycles.

The ensuing investigation by the CIBCR and the U.S. Bureau of Mines led to the development of indexes that measure the current and future short-term performance of the following industries: (1) primary metals, (2) steel, (3) aluminum, and (4) copper production. The Bureau's objective in developing these indexes is to inform specialists from the minerals industry and other sectors of the economy of the current and expected economic status of the metals industry over the next 6 months, on average. Additionally, the Bureau decided to use the CIBCR Leading Index of Inflation to forecast major swings in metals prices.

For each of the four metal industries, a composite coincident index and a composite leading index were developed, based on data and procedures similar to those used in constructing the U.S. Department of Commerce's coincident and leading cyclical indicators of the national economy. For instance, the coincident index for primary metals includes an industrial production index and a series on total manufacturers' sales, and its leading index includes an index of stock prices and a series on the average weekly hours of production workers. In February 1992, we issued our first monthly report for all four metal industries. Presently, we give a forecast, each month, of what each metal industry will do in the next six months, based mostly on the recent behavior of each metal industry leading index and the leading inflation index. The attachment shows the components of the Bureau of Mines' metal industry indexes.

During development of the indexes, it became apparent to the project staff that the concept of a business cycle and much of the technical language describing cyclical indicators would not be easy for laymen or members of a nontechnical audience to understand. This was of concern since we are directing our efforts largely to individuals and organizations in the minerals industry whom we believe may have little or no formal knowledge of business cycle analysis. In fact, people employed in the U.S. minerals industry are very conservative and often suspicious of social scientists, especially economists and statisticians. Moreover, we were planning to interpret the latest changes in the metals leading and coincident indexes and the CIBCR leading index of inflation by relating these changes to trends and events affecting domestic and international markets, including international financial and political actions. Such interpretations, therefore, would attempt to go beyond simple descriptions of index movements and contributions to the monthly changes. We would try to explain WHY some of the indicators composing our indexes had changed and what these changes, if any, would mean to the metal industries.

Much of the work we undertook in regard to developing the metal indexes, then, concerned communicating the meaning of the movements of the indexes and our forecasts to our customers. Some of the major points we have tried to communicate include:

- why is knowledge of the business cycle important to the metal industries?
- is there a unique cycle for the metal industries?
- what is the relationship between the metal industry cycle and the business cycle?
- what are coincident and leading indexes?
- why are metal industry coincident and leading indexes useful?
- how do metals leading indexes forecast?
- how well have the Bureau of Mines' metals leading indexes forecast in the past?

Our efforts to explain business cycle analysis and the meaning of coincident and leading indexes can be broken down into four steps. I would like to briefly describe these steps, which are as follows:

1. designing a "user-friendly" publication,
2. letting our customers help direct the nature of the product,
3. writing easy-to-read monthly reports, and
4. planning a workshop to explain the **Metal Industry Indicators.**

Although we have been publishing our report since last February, we are still working to improve it, responding to reader comments, and looking for other ways to enable the reader to understand the business cycle approach to forecasting changes in the metals industry. The structure of our publication is still not final; it is evolving as we receive comments and suggestions from our readers and come up with new ideas ourselves.

Designing a "user-friendly" publication

The typical publication, whether government or private, that describes economic and statistical forecasts, as well as economic and statistical data, is, quite frankly, boring! Furthermore, its content sometimes is not clear to economists and statisticians, yet alone potential users outside these professions. Realizing this, we have attempted, and still are attempting, to design a publication that is attractive, readable, and informative to geologists, financial officers and planners, marketing professionals, and many others interested in the minerals industry. I'm afraid that we have some way to go. In looking back at the text of some past issues of our monthly report, I find some of them somewhat tedious and inconsequential, an indication that we may have some problems in not fully understanding the rationale for the movement of the indexes.

First of all, we tried to catch the reader's eye. Our first five monthly reports were printed in color. But we ran into two problems. The first problem was the Government Printing Office, which could not guarantee timely printing. The second problem was a recent change in Department of the Interior policy limiting the printing of color publications. We, of course, were disappointed that we could not publish in color. Right now, our monthly report is printed in black and white and can be viewed as composed of two parts:

1. a summary, and
2. an analysis with accompanying tables and charts.

We think that the most distinctive feature of our publication is the summary box. We rely on it as an efficient and simple way to give the reader our forecasts. We assume that most of our audience does not care to read a technical discussion of the cyclical behavior of the metal industries. We therefore designed the summary box to tell those readers what we think is going to happen to metals prices and activity for each of the four metal industries over the next 6 months.

We try to write our forecasts for each of the metal industries using simple language and in a such a manner that the reader will not have to refer to the rest of the report. The analytic portion of the report is geared to our more technically oriented readers, who seem to be employed mainly in banking, finance, commodity trading, and marketing. We also copy other good ideas. We've always liked the "contribution to the index" summaries that the Bureau of Economic Analysis provides when they release their composite cyclical indicators of the national economy. We developed contributions tables similar to those of the Bureau of Economic Analysis, which show not only how much each economic indicator was contributing to its specific metal index, but to let anyone who is interested know what indicators are included in the indexes and the sources of these indicators. Incidentally, a total of 43 indicators are included in our coincident and leading indexes for all four metal industries.

Letting our customers help direct the nature of the product

Like any other "dynamic, hard-charging organization" with a new product, we decided to have a trial test of the metal industry indicators with selected readers first. These individuals could be counted on for honest critiques of our report, and the report was printed and mailed to these individuals from July through October 1991.

The major objective of this trial report was to solicit comments from these experts regarding the design, usefulness, and content of the report, and to determine if some of our readers had trouble understanding the cyclical approach to analyzing the metal industries. Additionally, the trial period was used to help develop a production schedule for the monthly report, including how long it would take to produce a report once all the necessary data were available.

Most of the comments we received indicated an interest as well as suggested improvements in the publication. What most readers seemed to want more than anything else was a forecast of metals prices. Although the Bureau of Mines does not forecast metals prices, we decided that it would not be inappropriate to indicate if the direction of metals price movements was about to change, using a data series that predicts changes in inflation for the whole economy. This ultimately led to the inclusion in our monthly report of the CIBCR Leading Index of Inflation, which has been used by Dr. Moore in the past to estimate future changes in metals prices.

Other readers suggested that our publication be structured more like a newsletter, and we are trying to write the **Metal Industry Indicators** so that it looks and reads like a newsletter.

Writing easy-to-read monthly reports
Like many other technical and complex subjects, business cycle analysis has its arcane language that can sound somewhat confusing to those not familiar with the subject matter. Many of you, I'm sure, have heard terms such as "peak" and "trough," along with others such as "amplitude," "trend adjustment," etc. Even the most commonly used words for describing business cycles, "leading," "coincident," and "lagging," are difficult to comprehend by many people not familiar with the language of cyclical indicators.

Writing a description of the behavior of the indexes and of the forecasts is probably our most difficult job. We make an attempt to use simple precise English as much as possible. We try to avoid the technical language often associated with business cycle analysis. If we must use a business cycle term, we try to provide a short, simple definition that will make the term clear to our readers. For instance, we always remind the reader each month that the metal leading indexes signal changes in the coincident indexes an average of 6 months before actual changes occur and that our six-month smoothed growth rate is a measure of near term trend. We also remind our readers, each month, that each metal coincident index measures current metal industry performance as represented by production, shipments, and employee hours worked.

We realize that it is not possible nor practical to eliminate all words and phrases associated with business cycle and other economic analyses. However, we are determined to keep technical words such as these out of our report:

autoregressive moving average
stochastic
exogenous
endogenous
standardization

Our Chief Economist has been somewhat critical of some of our drafts lately. Because of the slow recovery occurring throughout the economy and within the metals industry, we have had difficulty interpreting where our leading indexes are pointing. The Chief Economist has suggested that we actually say that we don't know what is happening when we are unsure of what the indexes are telling us. In other words, write the interpretation as a "human being" would write it, and don't worry about making a forecast each month! With the way the economy is behaving, and given the major structural changes in our society this past 5 to 10 years, we may wind up doing this in one of our future issues. We are also planning to be candid about the quality of our product. For example, sometime during the next year we may publish a comparison of our monthly outlooks to the actual results that occurred in the metal industries.

Workshop on Metal Industry Indicators
Early this summer we decided to take another step in helping our readers to understand and benefit from the **Metal Industry Indicators.** The Bureau of Mines and the Center for International Business Cycle Research are planning a workshop to assist the minerals industry, consumers, and others interested in the minerals sector interpret the **Metal Industry Indicators** data and analysis, and the contributions of the **Metal Industry Indicators** to their own activities. This workshop will also offer those attending an additional opportunity to critique the present indicator approach, content of the report, and to recommend improvements.

The proposed workshop will be held for two days in November 1992, in New York City, with 50 to 75 in attendance. Those attending will be drawn from the mineral and mineral-consuming industries, the financial and research communities, and other planning, statistical and analytical groups. All individuals present will be invited to actively participate in the proposed sessions. We think the workshop is a great way to help the minerals industry get further acquainted with our product.

Conclusion
We have found that considering the interests of potential users for a cyclical forecast of the metals industry is just as demanding as developing indexes that will give the best possible forecast. We expect to have a continuing dialogue with our readers as we refine our current indexes and attempt to develop indexes for other minerals.

COMPONENTS OF U.S. BUREAU OF MINES METAL INDEXES

Primary Metals

Coincident Index

1. Industrial production index, primary metals (SIC 33)
2. Total employee hours, primary metals
3. Manufacturers' sales, primary metals, 1982$

Leading Index*

1. Average weekly hours, primary metals (SIC 33)
2. Weighted S&P stock price index, metals firms, (1941-3 = 10)
3. Ratio of price to unit labor cost (SIC 33)
4. Journal of Commerce metals price index growth rate (1980 = 100)
5. New orders, primary metals, 1982$

Steel

Coincident Index

1. Industrial production index, basic steel and mill products (SIC 331)
2. Value of shipments, steel works, blast furnaces, and rolling and finishing mills, 1982$
3. Total employee hours, blast furnaces and basic steel products

Leading Index

1. Average weekly hours, blast furnaces and basic steel products, (SIC 331)
2. New orders, steel works, blast furnaces, and rolling and finishing mills, 1982$, (SIC 331)
3. Contracts and orders for plant and equipment, 1982$
4. S&P stock price index, steel companies (1941-3 = 10)
5. Industrial production index for automotive products
6. Steel scrap price (#1 heavy melting, $/ton)
7. Total net new orders for machine tools, 1982$

COMPONENTS OF U.S. BUREAU OF MINES METAL INDEXES cont.

Aluminum

Coincident Index

1. Production of primary aluminum (thous. mt)
2. Recovery of aluminum from scrap (thous. mt)
3. Total employee hours, primary aluminum production
4. Shipments of aluminum ingot and mill products (mil.lbs)

Leading Index

1. Average weekly hours, primary aluminum production (SIC 3334)
2. Average weekly overtime hours, primary aluminum production
3. S&P stock price index, aluminum companies, (1941-3 = 10)
4. Aluminum LME spot price index (pounds sterling/mt)
5. Imports of aluminum, metal and alloys, crude, mt
6. Index of new private housing units authorized (1967 = 100)
7. Industrial production index for automotive products
8. Construction contracts, commercial and industrial (mil sq ft)
9. Net new orders for aluminum mill products (mil lbs)
10. Ratio of shipments to inventories, aluminum ingot and mill products (mil lbs)

Copper

Coincident Index

1. Industrial production index, primary smelting and refining of copper (SIC 3331)
2. Total employee hours, rolling, drawing, and extruding of copper (SIC 3351)
3. Copper refiners' shipments (short tons)

Leading Index

1. Average weekly overtime hours, rolling, drawing and extruding of copper (SIC 3351)
2. New orders, nonferrous and other primary metals, 1982$
3. S&P stock price index, miscellaneous metals (1941-3 = 10)
4. Construction contracts, commercial and industrial (mil sq ft)
5. Copper scrap price, N.Y. #2, (cents/lb)
6. Index of new private housing units authorized (1967 = 100)

---

*/The specific leading index for each of the metal industries is combined with the CIBCR short-leading and long-leading indexes of the U.S. economy. Each CIBCR index has been assigned a weight of 25, and the specific metal industry leading index has been assigned a weight of 50. Each indicator in the specific leading index has been given an equal weight.

Observations of A Novice

Robert E. Jarrett, Environmental Policy Institute, U.S. Department of the Army

I. Introduction

The U.S. Army Environmental Policy Institute (AEPI) is a relatively unique entity just two years old. AEPI came into existence as the result of a high level realization that the Army too frequently discovered itself to be "behind the curve" on environmental issues. Policy was too often obsolete by the time it appeared in the form of regulations. Undersecretary Shannon signed our charter in September 1990. It says:

> "The mission of the Army Environmental Policy Institute is to assist the Army Secretariat in the development of proactive policies and strategies to address environmental issues that may have significant future impacts on the Army."

It goes on to set several tasks among which are the following:

- Anticipate trends to allow for issuance of proactive policies to meet future challenges.
- Periodically assess and analyze future challenges in order to provide new policy approaches.
- Stay abreast of technology in order to provide options to minimize environmental impacts.

In other words, "Look into the future and tell us what we need to know early enough for us to be ready to do the right thing at the right time." This may be a familiar task for someone else at this conference.

The enormity of this job becomes clearer when one considers the Army in the light of some common images. The Army Science Board states the Army to be, "...the nation's largest industrial manufacturer" (Martin, et al, February, 1990). That speaks for the complexity of the industrial environmental issues. The Army also operates a large number of small and medium sized cities providing a full range of services to "citizens" and military units. And it manages over 20 million of acres of natural resources encompassing almost every kind of landform and biota. Add to that a peculiar type of agriculture in which the farmers run up and down the fields in heavy machinery and apply smokes and explosives, yet never take-in any crops! The Army is, by the definition of its mission supposed to be ready to wreak terrible harm on an enemy. At the same time it is involved with environmental laws, standards and issues of all kinds, including those that have yet to be recognized as important.

AEPI, currently at Champaign, Illinois, is a small organization of about 25 people, most of whom are temporary or short term personnel. A few are expert senior fellows and fellows. Others ore supporting graduate students and clerical staff. We do some policy analysis in-house and manage contracts for work efforts beyond our capabilities. We have contracts with a variety of private consultants and universities. We are gradually developing a network of nationally known experts who will work with us on an as needed basis. AEPI has taken a few tentative steps in the direction of futurist activity necessary to support the assigned mission and tasks and to provide information needed by research staff and contractors. To this point, each work effort has had to pretty much develop its own information, forecasts and conclusions.

II. Early Steps and Infrastructure Development

In its first year, AEPI commissioned some bibliometric work as an attempt to identify important, but less than obvious, issues in order to peer around the corner. That was helpful in limited respects, but clearly needed massive analytical work beyond the abilities of information science specialists lacking environmental subject matter training and intuition. A contractor using literature searches, expert interviews, poll analyses, and advocate group analyses developed a set of 41 environmental trends that seemed to have validity for describing key directions and intensities of environmental issues. AEPI issued the results in the form of a booklet for reference use, mainly among Army officials. A multi-disciplinary workshop of industrial, academic and government experts validated the list and added 16 more in August 1991.

The same contractor is now performing the first annual trend upgrade. (The criteria being used in deciding on trends worth following appear as an appendix.) Our intent is to eventually have a continuous environmental issue trend process yielding information to be used by our staff and external consultants. However, this is merely an informational infrastructure tool to aid in the actual forecasting and policy analysis work tasked by the charter. We're still experimenting and plagiarizing.

Internal staff read about and participated in a variety of forecasting procedures and concluded that scenario building techniques might be the most applicable. While scientific environmental information is quantitative, the largely socio-political determinants of how, why and when environmental issues ripen into culture change and legislation require qualitative approaches. Scenario building seemed to provide for both qualitative and quantitative inputs. Nevertheless, we chose not to leap into a commitment without further evaluation. The roads of applied behavioral science and management innovation are littered with countless hip-shot and faddist attempts to find the perfect miracle. The next section addresses more specifically what we've done to decide upon an approach to issues forecasting that we might be able to live with on a long term basis.

III. Scenario Building Explorations

The 1991 workshop that looked at the first trend analysis produced a recommendation: that AEPI should explore scenario building as a method of moving forward from merely looking for and at trends and on into identifying probable futures worthy of anticipatory policy attention.

For those not familiar with the jargon here are a few items of quasi-definition. Forecasting scenarios are basically fictional accounts of plausible future situations one can use; in our case, to develop sensible policy options from which to select and implement as real events unfold. The plausible futures grow out of disciplined analysis of observable trends, expert testimony, expert panel weightings and even projection of reasonably possible, but unexpected conditions and events.

Recognizing that many products are more attractive at a distance than close-up, and that recommendations are easier to give than to use, we held several internal meetings on the subject of forecasting methods. We concluded that scenario building had considerable promise. Taking an apparently unusual step, AEPI engaged a consultant to conduct a two day educational course outside of a formal process. The scope of work required presenters of three types: 1) proponents of several variants of scenario building, 2) industrial users, and 3) government users. We deliberately sought conflicting views and specified a need to hear about failures. We requested a kick-off session of a strictly descriptive nature devoid of advocacy. And we called for about four hours of pre-reading material to reduce classroom start-up lag. The course successfully met our goals of laying out pluses and minuses, identifying optional variants, clarifying resource needs (e.g. calendar time, person time, funds, and information). In contrast, the usual approach is to decide on a desired activity then have a consultant get on with it.

Scheduling and funding problems caused a bit narrower sampling than intended. Even so, we had an excellent set of speakers who covered the desired range and did well at separating romanticism from reality.

We received presentations including discussions of methods covering the spectrum from highly computerized manipulation of quantified qualitative input to relatively qualitative handling of qualitative input to conversion of highly quantitative input to obtain qualitative and quantitative insights. One Pentagon speaker described the expenditure of $100,000 to get a quick result in 90 days that took three times longer and netted them sets of scenarios from five sources. The products range from cute stories without supporting documentation to rather serious efforts with reasonable documentation to allow sensible interpretation. An industrial application of scenario building and analysis in one of the major oil companies went on to become the foundation of permanent strategic planning for the firm.

The course showed us the following:

- There are as many variants as consultants, as we had suspected.
- Some variants are disciplined, and some are not.
- There's no point in getting involved in such a process unless an organization has a real use for it and intends to suffer the considerable work costs needed to make it flourish.
- Quantification of the "unquantifiable," if attempted, must be done carefully and not as a magical smokescreen.
- The process is the essence; the documentation is the working tool for follow-on use; and the final scenario write-up is the cream.
- Consultant and client must be carefully matched.
- Disciplined scenario building is an expensive, long process and is rarely beneficial as a one-shot action.
- It isn't necessary to reinvent the wheel: many organizations have done basic fact gathering and scenario formation that can be analyzed for adaptation and continuation. (Is this blasphemy!?)

IV. Exploration Results, So Far

Now, here's how it turned out. A group of intelligent, educated and professionally motivated people who work well together met on several occasions to decide upon a path to take and have been unable to seriously address that question. The message seems to be that such a process is almost too hard to contemplate. Dan Gaske's paper at the 1990 Federal Forecasters' Conference offers several observations on why forecasting processes stumble. Since we meet them in an organization, like AEPI, charged with a futurist mission (yet trying to meet higher headquarter's fire of the month demands) then it will be very hard to design and conduct a continuous process involving many other agencies. That looks like a post-course learning that validates many authors' warnings to lay good groundwork, or be ready for failure. In our discussions, we discovered for ourselves something that consulting practitioners might prefer we hadn't. That is: there are at least three levels of abstraction from which a policy analyst may choose when using the basic processes of data gathering, work-shopping, expert validation, analysis and scenario writing. One can elect to do policy analysis and synthesis from a basis of facts and well researched trend analysis. One may go another step to disciplined forecasting. Or, one may go on to the final scenario(s). A danger one might anticipate in the latter could be a tendency for people doing individual policy issue analyses to forget to go back to the first principles embodied in the trend and forecast development documentation and to fail to update the information.

What comes next? One of the points that had been up for discussion was whether we should launch a scenario building process, including hiring a consultant. And, if the answer were yes, what should be the scope of work? Having failed to get meaningful intra-organization discussion, I was in a quandary. Finally the answer staring me in the face kicked me in the shin: despite a year of informal and formal preparation for its futurist role, this organization needs more institutional development before it can hope to be a credible proponent for and custodian of an expensive forecasting process. A significant failure would damage charter accomplishment far more than slow accomplishment. We are in preliminary stages of scoping a delivery order for a consultant to work us through a forecasting design process. Besides generating a design to be the basis for further contractor support and to guide the whole effort, that activity will serve to provide essential, further institutional development within our small group. Too often, in any activity, we all underestimate the difficulty of dealing with our friends! Use of a design consultant should provide these benefits: stimulate participation through the normal tendency to react to an expert outsider; apply expert help to the task: and stop drift by setting a formal deadline in the form of a contract closure date. We expect to specify that the design consultant can't get the implementation contract. That should help reduce any tendency to build personally preferred gimmicks into the design.

V. Closing Comments

Thank you for this opportunity to share a novice's observations. I've not meant to be critical of my organization, merely to be candid about some of the actual problems of introducing major work and thought process changes, even in an organization created to be flexible.

Here are the seven summary enlightenments from our experience, so far:

1. One should not be over-awed by the jargon - much that sounds different is the same, and much that sounds the same needs to be questioned.

2. One need not buy the whole ball of wax - caveat emptor.

3. One must be detailed and unambiguous as to what is needed - what the organization/audience will swallow and use.

4. One must be deeply involved - no turnkey projects or absentee landlords allowed.

5. Even players who seem most involved will waver badly in their participation intensity - the urgent will supplant the important in daily schedules.

6. Engaging in fads is expensive and too often destructive - short term exhilaration is a poor substitute for long-term effectiveness.

7. Expectations must not be allowed to outrun reality - relative failure can be as deadly as absolute failure.

Jarrett's law of the infinitude of errors states that there is always one more than the number found. By way of corollary one can ask, "If the list above was so easy to notice before getting deep into implementation, what truly evil monsters still lie hidden?"

References

1. Gaske, Dan, Improving the Utilization of Forecasts: Some Helpful Principles, Proceedings of the 1990 Federal Forecasters Conference, Washington, D.C., March, 1991.

2. Martin, Alexander, et al, Army Science Board Final Report of the Ad Hoc Subgroup on Toxic and Hazardous Waste Management, Assistant Secretary of the Army (Research, Development and Acquisition), Washington, D.C., February, 1990.

Appendix

Revised Criteria Considerations for Trends

1. Documented in multiple references rather than a single source.

2. Supported by hard data or findings of some type.

3. Reflected in actual or planned allocations (dollars, people, facilities, equipment within some major sector(s) (gov't, private sector, etc.))

4. Stated by persons/organizations with significant power, influence, responsibility or respect.

5. Documented repeatedly over longer time periods.

6. Documented in official rather than unofficial documents (EPA, GAO, OTA, NAS reports, etc.).

7. A consensus (the options/analysis of a group of people, not a single person, e.g., NAS report) or cited by a coalition of organizations, not just a single group (applies to environmental interest groups, trade associations, United Nations, NATO, EC, etc.) versus one person's view/observation.

8. Cited in peer review documents preferred over non-peer reviewed.

9. Primary reference sources over secondary reference sources.

10. National or international based observation versus regional or local.

11. Basis of support for trend (fear, religion, economics, law, etc.).

12. Trend does not appear to contradict other trends.

13. Number of courses or amount of training provided on subject.

14. Number of conference sessions or journal papers on subject.

15. Action at multiple levels (local, state, Federal, international) rather than action at only one level.

16. Establishment of offices companies or gov't. agencies on subject.

17. Cited in testimony to Congressional Committee or other gov't. body.

18. Reflected in passed or pending legislation or regulations.

Notes:
1. The order the criteria are listed does not reflect their relative priority.
2. Not all criteria necessarily apply to assessing all trends.

The Impact of Foreign Trade on U.S. Employment

Michael P. Botos, Office of Employment Projections, Bureau of Labor Statistics[1]

I. Introduction

Traditionally, one of the analytical steps in the presentation of a new set of industry-level employment projections by the Bureau of Labor Statistics (BLS) is a factoring of the basic employment equation. That is, projected industry employment growth is broken into the proportions of growth due to each of the pieces of the BLS employment identity. Standard factoring analysis is based on a disaggregation of a multiplicative relationship, changing one factor from the relationship at a time and examining the resulting growth path. This article presents such a factoring analysis at a 55-order level of detail for the 1977-90 and the 1990-2005 periods, the latter drawn from the latest BLS industry employment projections.

A further goal of this paper is to develop techniques for evaluating the employment growth factor shares of components of the employment identity which are fundamentally additive, rather than multiplicative, in nature. An example of this type of factor share would be changes in employment growth due to a change in a single component of industry final demand, such as personal consumption spending, or imports, rather than changes in the overall final demand distribution. This is a new development in the methodology of factor analysis. The new technique is here applied to changes in foreign trade shares and distributions over the historical period and as projected by BLS for the 1990-2005 period and that portion of employment change at an industry level of detail related to these components are detailed below.[2]

The discussion will proceed in two major sections. The first will describe the methodology followed and the results attained through a generalized factor analysis. This initial factor analysis was performed in order to identify the employment impact of each of the major factors contributing to U.S. employment growth. The second major section presents the methodology and results achieved through a more detailed factor analysis. This section describes the modifications necessary to analyze the impact of foreign trade on U.S. employment growth. The employment impact of the other factors, which were identified in the generalized factor analysis, remained unchanged by the modifications made in the more detailed factor analysis.

II. General Factor Analysis: Methodology and Results

The factors identified as contributing to U.S. employment growth over both the historical and projected periods were: the level of real Gross National Product (GNP), the distribution of real GNP, the distribution of final demand, the total requirements (Input-Output) table, and the employment-to-output ratio.[3] The equation used to relate these factors, or variables, to one another relies heavily on the input-output accounting structure derived by the U.S. Bureau of Economic Analysis (BEA)[4], and is given by the identity:

$$\text{ECalc} = \{[(\text{GNP} \times \text{GNPDist}) \times \text{FDDist}] \times \text{TReq}\} \times \text{E/O}.$$

The variables in this identity are defined as:

ECalc: The calculated employment level after one of the factors in the equation was allowed to vary while all other factors were held constant. This calculated employment level can be understood as the employment related to the varied factor.

GNP: The level of real Gross National Product. This factor identifies the employment change due to real economic growth.

GNPDist: The percent distribution of GNP by major demand category. The major demand categories identified for this analysis were: personal consumption expenditures (PCE), investment, exports, imports, federal defense purchases, and other government purchases. This factor identifies employment changes due to shifts over time between major components of GNP. For example, a shift in percentage share of GNP from investment to PCE.

FDDist: The percent distribution of final demand by sector. This factor identifies employment changes due to shifts over time within major final demand categories, such as a shift within PCE from the purchase of automobiles to the purchase of computers.

TReq: The Total Requirements (input-output) table. This factor identifies employment changes due to changes in the total input requirements necessary for production of a given level of output. It represents the technological state of the economy at a given point in time. The methodology followed in deriving the Total Requirements table appears in Appendix A.

E/O: The employment-to-output ratio. This factor identifies employment changes due to changes in labor productivity over time. An increase in labor productivity will result in a decrease in employment, a negative entry in this column. This factor, therefore, is inversely related to employment growth.

The individual employment impact of each factor was found by varying one of these factors at a time while holding all other constant. In this way, the individual contribution of each factor to the total employment change in each sector was isolated. The resulting factor impacts could be positive or negative, depending on the various correlations and economic relationships involved. Furthermore, the employment impact of each factor could assume a value greater than 100% when considered alone. The sum of the individual factor impacts, however, must equal the total employment change for each sector. In most cases, it did not due to the "interaction effect".

The interaction effect is that part of the total employment change not explained by the identified factors. The idea of an interaction effect is analogous to that of the "residual" described by Edward Denison. Simply put, the interaction effect results from the correlation of each of the identified factors with every other factor. The variation of only one factor at a time, therefore, creates second-order employment impacts due to its "interaction" with the other factors being held fixed. Although there are several methods of distributing this interaction effect back to the primary factors, the results attained using these methods were not reliable.[5] For this reason, the interaction effect remains in Table 1 and Table 2 as a separate factor.

The factor analysis methodology is sound, however, despite this interaction effect. The factor impacts that appear in Table 1 and Table 2 are the first-order impacts on employment in terms of percentage growth due to the variation of the identified factor. A few examples will further clarify this discussion. Table 1 shows the general results of the factor analysis for the historical period, 1977-1990. These results have been ranked by sector according to percentage growth. Table 1 indicates that the Business services sector was the fastest growing over the 1977-90 period. This sector added 4.6 million jobs from 1977 to 1990, a 182.8% employment increase over its 1977 level.

The individual factor contributions to each sector's employment growth also appear in Table 1. These individual impacts are the result of the variation of the identified factor to its 1990 level while all other factors were held constant at their 1977 levels. The percentage shown under each factor indicates the amount by which employment in that sector would have changed if only the identified factor were allowed to change over the 1977-90 period. For instance, if only the distribution of GNP were allowed to vary over the historical period, employment in the Business services sector would have increased 1.7%. In order to assess the impact of this variation in the GNPDist factor, the general equation presented above was modified in the following manner:

$$\text{ECalc} = \{[(\text{GNP1977} \times \text{GNPDist1990}) \times \text{FDDist1977}] \times \text{TReq1977}\} \times \text{E/O1977}.$$

Adjustments to the general equation were made in the same manner to calculate the employment impact of each individual factor. The results of these calculations for the historical period, which appear in Table 1, deserve further clarification. In order to fully understand the implications of these results, the intuition behind each of these individual factor impacts must be explained. For instance, what does it mean for the distribution of GNP to provide a 1.7% contribution to the total employment increase in the Business services sector?

To answer this question, recall that the GNPDist factor was defined above to be the impact on employment of shifts between major final demand categories. Interpretation, therefore, requires that we know how the percentage of GNP accounted for by each major final demand category has changed over time. In 1977, the majority of business services were purchased by two major final demand categories: personal consumption expenditures (PCE) and federal government nondefense expenditures. From 1977 to 1990, the percentage of GNP accounted for by PCE rose only slightly, while the percentage of GNP purchased by the federal nondefense category declined slightly. Holding all else fixed, the result of these minor shifts was a mere 1.7% positive contribution to the employment increase in the Business services sector over the 1977-90 period.

The Business services sector provides some other interesting results which are worthy of discussion. For instance, the 12.6% impact due to the FDDist factor is the result of the increased purchase of business services within all major final demand categories over the decade of the eighties. Much of this shift occurred because of an increased demand for computer and data processing services. As the use of computers grew over the 1977-90 period, demand within all major final demand categories shifted toward the purchase of computer and data processing services. The result of this demand shift was a positive impact on the employment of the Business services sector.

The largest factor impact in the Business services sector came from the Total Requirements (I-O) table. This factor, which captures the input-output relationships that exist in our economy, contributed a positive 64.7% to the employment increase in this sector. The TReq factor impact resulted from the increased level of "contracting out" of business services over the decade of the eighties. "As more private firms and government agencies found it effective to contract out for specialized services instead of performing them in-house," employment in the Business services sector increased.[6] Since labor is the primary input to the Business services sector, the Total Requirements (I-O) table provided a positive impact on employment growth.

Another side-effect of the increased demand for business services and subsequent employment growth in this sector was a decline in labor productivity over the 1977-90 period. This productivity decline was simply the result of employment growing faster than output in this sector. The industry leading this sector's employment growth was the Temporary help industry. "With more firms finding convenience and savings in contracting for temporary workers instead of hiring them, the number of employees on the payrolls of the personnel supply services industry swelled from 242,000 in 1975 to almost 1.6 million in 1990.[7] The resulting fall in labor productivity in this sector contributed 10.4% to employment growth in the Business services sector.

The Business services sector will continue to show strong growth throughout the next decade according to BLS projections. Table 2, which shows the general results of the factor analysis for the 1990-2005 period, indicates the Business services sector will grow by 50.6% over the projection period. Although this amounts to the addition of 3.6 million jobs, the Business services sector ranks only third in percentage growth terms, behind the Motion pictures sector and the Social services sector.

The Motion pictures sector is projected to grow fastest over the 1990-2005 period, increasing 73.7% by adding 418,200 jobs. This percentage growth is deceptive, however, since the Motion pictures sector began the projections period with a much smaller base than did the Social services sector and the Business services sector. For this reason, the following discussion will focus on the sector projected to grow second fastest over the 1990-2005 period: the Social services sector.

The Social services sector, which includes services such as child day care and residential care, will grow by 58.4% over the projection period, adding more than 1.2 million jobs. This projected job growth results primarily from the projected changes in the demographics of the U.S. population. For instance, the growth in child day care centers is projected to slow from the rates seen during the historical period due to a slower growth in the population of women of childbearing age.[8] Further growth to the Social services sector will result from the aging of the U.S. population, which is reflected in the projected growth of the residential care industry. In fact, according to Franklin and Carey (1992), the driving force for growth in the Social services sector over the projection period is the increase in the elderly population.

Most of the additional jobs will result from growth of the elderly population. The number of workers in residential care institutions, which provide around-the-clock assistance to older persons and others who have limited ability for self-care, is projected to increase 4.5 percent (annually), the fastest employment growth for any industry in the U.S. economy.[9] The increase in demand appears in the final demand distribution factor of the Social services sector. If the FDDist factor alone had been allowed to change over the 1990-2005 period, employment in the Social services sector would have increased 20.7%. This employment increase represents a shift within personal consumption expenditures (PCE) toward the purchase of social services. As the population ages further, demand for the services provided by this sector will increase and employment will grow in this sector.

III. Detailed Factor Analysis: Methodology and Results

Although the general results provide a wealth of information regarding the structure of the U.S. economy, the primary purpose of this study was to analyze employment impact of foreign trade. For this reason, the GNPDist factor became the focus of the detailed analysis. The distribution of GNP is the factor most often cited when claims are made that imports are eliminating U.S. jobs. The general factor analysis, therefore, was taken one step further by breaking the GNPDist factor into its component parts. By varying each of the individual components of this factor separately, the employment impact of each major final demand category could be isolated. The major demand components, which were mentioned above, include:

PCE: Personal consumption expenditures. PCE is represented by "PCE" in the equations that follow.

Investment: Investment includes producers' durable equipment purchases (PDE), residential construction, non-residential construction, and inventory change. Investment is represented by "Invest." in equations.

Exports: Exports are represented by "Exp." in equations.

Imports: Imports are represented by "Imp." in equations.

Defense: Federal defense purchases are represented by "Def." in equations.

Other Government: Other government purchases include Federal non-defense purchases and all State and local government purchases. Other government purchases are represented by "OthGov." in equations.

The factor analysis methodology was then conducted again using both the historical and projected data. The general equation, discussed earlier, was modified to incorporate the expansion of the GNPDist factor. Recall from above that the general equation was given by

$$\text{ECalc} = \{[(\text{GNP} \times \text{GNPDist}) \times \text{FDDist}] \times \text{TReq}\} \times \text{E/O}.$$

The major final demand components were substituted into the general equation for the GNPDist factor, so the modified equation took the form

$$\text{ECalc} = \{[(\text{GNP} \times \text{GNPDist}(\text{PCE} + \text{Invest.} + \text{Exp.} + \text{Imp.} + \text{Def.} + \text{OthGov.})) \times \text{FDDist}] \times \text{TReq}\} \times \text{E/O}.$$

The methodology followed from this point was analogous to that developed in the general analysis above. The individual employment impact of each major demand component was found by varying one of these components at a time while holding all other factors constant. In other words, the impact of a shift toward PCE would be identified by varying only the portion of the GNPDist factor related to PCE, instead of varying the entire GNPDist factor. The corresponding equation for this modification over the 1977-90 period would be

$$\text{ECalc} = \{[(\text{GNP77} \times \text{GNPDist}(\text{PCE90} + \text{Invest.77} + \text{Exp.77} + \text{Imp.77} + \text{Def.77} + \text{OthGov.77})) \\ \times \text{FDDist77}] \times \text{TReq77}\} \times \text{E/O77}.$$

This same procedure was followed for each major final demand component of the GNPDist factor individually. The sum of the resulting component impacts is exactly equal to the total factor impact for the GNPDist factor since the modifications made to the general equation were additive. In other words, there was no additional interaction effect. The major final demand components are independent, so there is no interaction amongst them. For this reason, the results of the detailed factor analysis fit neatly into the framework of the general factor analysis discussed above. The factor impacts resulting from the generalized factor analysis were unchanged by the modifications made for the detailed factor analysis, so these results are not reproduced in any of the remaining tables.

The results achieved through this more detailed analysis will be presented only for those industries most affected by foreign trade. This criterion does not necessarily limit the discussion to those industries in which the GNPDist factor had a large employment impact. On the contrary, employment impacts due to the major final demand components often neutralize each other since they are additive, leaving a small total GNPDist impact. It is informative, therefore, to identify these individual employment impacts by dissecting the GNPDist factor. In so doing, the effects of foreign trade on U.S. employment can be identified.

Imports grew from 10.6% of GNP in 1977 to 15.8% of GNP in 1990, an annual growth rate of 5.9%. This import growth has definitely had a negative impact on U.S. employment growth. By the same token, however, the growth of exports has had a positive impact on U.S. employment growth. Exports grew at an annual rate of 6.4% over the 1977-90 period, rising from 9.4% of GNP to 15.0% of GNP. The annual growth rate of exports, therefore, surpassed that of imports over the 1977-90 period. The employment impact resulting from this export growth is shown in Table 3.

Table 3 presents the results of the factor analysis calculations after varying each component of the GNPDist factor. These results have been sorted by total employment growth to facilitate comparison with the earlier discussion. Table 3 indicates that, in total, exports created more jobs than imports eliminated over the 1977-90 period. If only exports were allowed to adjust to their 1990 level, while all other factors were held fixed, U.S. employment would have increased by 3.6%, adding over 3.3 million jobs.[10] On the other hand, U.S. employment would have fallen by -2.6%, losing more than 2.4 million jobs over the historical period, if only imports had been allowed to change. This indicates the net employment impact of foreign trade over the 1977-90 period was positive.

A closer examination of the results presented in Table 3 indicate that the net employment impact of foreign trade was negligible for most of the fastest growing sectors over the 1977- 90 period. This is explained by the fact that service-producing sectors dominated the growth in U.S. employment over the last decade. Trade in "services" is typically more difficult than trade in "goods." In many cases, the provider must deliver the service in-person. Since this is not always possible, both exports and imports of services are inhibited. The U.S. might enjoy a comparative advantage in the production of certain services, but physical externalities hinder foreign trade in these services.

Although some services are more readily traded internationally, employment changes due to foreign trade often offset. The Business services sector, for example, produces services such as advertising which are heavily exported by the U.S. The Business services sector also produces computer and data processing services, however, which were heavily imported over the 1977-90 period. As a result, the 4.7% employment increase in this sector due to exports was neutralized by a -3.0% employment decline due to imports over the 1977-90 period. The net employment impact due to foreign trade in the Business services sector, therefore, was only 1.7%.

This offsetting employment impact provides additional evidence as to the growth in importance of foreign trade to the U.S. economy over the past decade. The results in Table 3 indicate that, in general, those sectors which experienced large employment increases due to exports also experienced large employment declines due to imports over the 1977-90 period. The industries of interest in Table 3 are those in which the net employment impact of foreign trade was not negligible. For this reason, the results shown in Table 3 were sorted on both exports and imports.

Table 4 lists the ten sectors which benefitted most from export growth over the 1977-90 period. At the top of this list is the Metal mining sector, which would have experienced a 19.0% employment gain if only exports had changed over the historical period. Exports, unfortunately, were not the only factor to have changed over the 1977-90 period. Table 1, which was discussed earlier, indicates that employment in the Metal mining sector fell -34.6% over the 1977-90 period. The GNPDist factor contributed -4.7% to this decline. This implies the large employment impact due to exports was more than negated by another component of the GNPDist factor. This component, as it turns out, was imports.

The growth in imports over the 1977-90 period contributed - 32.4% to the employment decline experienced by the Metal mining sector. The net impact of foreign trade on employment in this sector, therefore, was -13.4%. As discussed above, the growing importance of foreign trade to the U.S. economy often produced offsetting impacts over the 1977-90 period. This increased foreign trade had a negative net effect on the Metal mining sector. The same was not true for other sectors over the historical period. In fact, thirty-two of the fifty-five sectors included in this study experienced positive net employment impacts due to foreign trade while only fourteen experienced negative net impacts. The impacts of foreign trade on employment in the remaining nine sectors were either completely offsetting or irrelevant.

Strong foreign demand for U.S. industrial machinery and computer equipment led to a positive net employment impact due to foreign trade in the Industrial and commercial machinery and computer equipment sector. The growth of exports contributed 12.8% to employment growth, while import growth had a -6.3% dampening impact. The net effect of foreign trade on employment in this sector was 6.5%. Although the exports of many of the industries included in this sector grew over the 1977-90 period,much of the positive net employment impact due to foreign trade results from the growth in exports of computer equipment.

The growth in the output of computers over the historical period carried this sector and led to a positive net employment impact due to foreign trade. As Franklin and Carey (1992) indicate: In 1975, computer output was just under two percent of the real output of the major industry group Industrial machinery and equipment. In 1990, computers accounted for just over fifty percent of the real industrial machinery and equipment output, but only nineteen percent of the group's employment.[11]

Since computers accounted for such a small percentage of the employment in this sector, however, the output growth of the computer manufacturing industry could not offset the negative employment impacts due to other factors. As a result, employment actually fell -4.0% in the Industrial and commercial machinery and computer equipment sector over the 1977-90 period.

In fact, of the eight sectors listed in Table 4 that exhibit positive net employment impacts due to foreign trade, only three actually experienced employment growth over the 1977-90 period according to Table 1. These three sectors include the Transportation equipment sector, the Chemicals and allied products sector, and the Electronic and other electrical equipment sector. The positive net employment impact due to foreign trade in these three sectors, however, was quite small due to the offsetting employment impacts of exports and imports over the 1977-90 period. Moreover, many of the sectors listed in Table 4 which benefitted most from exports over the historical period were also hurt significantly by imports over this period. Imports provide the other side of the foreign trade picture. For this reason, the results of this factor analysis were sorted by imports. The sectors which were most hurt by import growth over the 1977-90 period appear in Table 5.

Table 5 indicates that five of the sectors which enjoyed the greatest benefits due to export growth over the 1977-90 period were significantly hurt by import growth over the period as well. The Metal mining sector, which topped the list of sectors that experienced export related employment growth, saw employment fall -32.4% due to imports over the historical period. Other sectors which appear in both Table 4 and Table 5 include the Primary metals sector, the Electronic and other electrical equipment sector, the Nonmetallic minerals sector, and the Chemicals and allied products sector.

Employment in the Primary metals sector would have fallen - 15.6% if only imports had changed over the 1977-90 period. This negative impact was offset partially, however, by the growth in exports over the period. The Primary metals sector, which includes the steel industry, experienced significant growth in foreign trade over the last decade. Imports grew dramatically in the early eighties as foreign competitors became the low-cost producers of basic steel products. This import growth slowed considerably in the latter half of the decade. U.S. mini-mills, which produced a customized products tailored to meet its purchasers' needs, took back market share from foreign producers and caused the level of U.S. exports to rise again. This rise in foreign trade amounted to a net employment impact of -3.4% in this sector.

The net employment impact of foreign trade in the Electronic and other electrical equipment sector, on the other hand, was a positive 0.6% over the 1977-90 period. Although this positive impact is not large, it reflects the export and import relationships that existed in 1977 for this sector. These relationships were held constant for the purposes of this analysis. In 1977, for instance, the U.S. was more competitive internationally in consumer electronics than it was in 1990. The offsetting employment impacts of these relationships are shown in Table 5.

The remaining two industries which were identified as appearing in both Table 4 and Table 5 also exhibit offsetting employment impacts due to foreign trade. For instance, employment in the Nonmetallic minerals sectors increased 1.2% over the 1977-90 period, while employment in the Chemicals and allied products sector grew 2.9% due to foreign trade. These small net employment impacts due to foreign trade exemplify the primary conclusion of this paper: Foreign trade is beneficial to U.S. employment growth. For the most part, the net impact of foreign trade is positive in those sectors in which the employment impacts of exports and imports do not offset.

This result will continue to hold over the 1990-2005 period according to BLS projections. Table 6 presents the results of the factor analysis after splitting the GNPDist factor over the projection period. Once again, these results have been sorted by total employment growth to facilitate comparison with the earlier discussion. Table 6 indicates that, in total, exports created more jobs than imports eliminated over the 1990-2005 period. The growth of exports will contribute nearly 3.8 million jobs over the projection period, a 3.1% increase in employment, while import growth will eliminate 2.3 million jobs, a -1.9% impact. This is due to the fact that exports are projected to grow faster than imports over the 1990-2005 period. The BLS projections indicate exports will grow from 15.0% of GNP in 1990 to 20.5% of GNP by 2005, an average annual growth rate of 2.1%. Over the same period, imports are projected to grow at a rate of only 1.4% per year. The net employment impact of foreign trade resulting from these differing projected growth rates, therefore, was positive over the 1990-2005 period.

As in the historical period, the net employment impacts in each sector due to foreign trade tended to offset over the 1990- 2005 period. Moreover, forty-one of the fifty-five sectors included in this study are projected to experience positive net employment impacts due to foreign trade over the 1990-2005 period, while only six were projected to experience negative net employment impacts. The net foreign trade impact the remaining eight sectors was either completely offsetting or negligible. The trends discussed with regard to the historical period, therefore, continued to hold over the projections period.

In order to further analyze the results over the projection period, the values shown in Table 6 were sorted by exported impact. The sectors which benefitted most from the projected growth of exports are shown in Table 7. Nine of the ten industries listed in Table 7 have positive net employment impacts due to foreign trade. The one sector projected to exhibit a negative net employment impact due to foreign trade is the Oil and gas extraction sector. This sector is projected to experience an 8.5% positive employment impact due to exports and a -19.0% negative employment impact due to imports. The net employment impact of foreign trade, therefore, is projected to be -10.5%. Demand for the output of this industry has been falling over the past two decades. As Norman Sanders points out, this trend is expected to continue into the future.

Consumer energy use-- gasoline and motor oil for our automobiles and fuel oil, natural gas, and electricity for heating and air-conditioning our homes - has grown at a relatively slow pace since 1972, a reaction to higher energy costs and a reflection of the economy-wide move toward energy conservation. More efficient automobiles and appliances and better insulated homes have led to declining consumer energy use, from a 10.4 percent share of overall personal consumption expenditures in 1975 to 7.2 percent by 1990. The moderate growth projections assume that many of these trends will continue, leading to energy use accounting for only 6.1 percent of consumption spending by 2005.[12]

The sector most helped by the projected growth in exports over the 1990-2005 period was the Metal mining sector. The structure of exports, coupled with the projected export growth, led to an employment impact of export growth in this sector of 20.4% over the 1990-2005 period. This large positive impact was offset, however, by the -18.1% negative impact of imports over the period. As a result, the net impact of foreign trade contributed only 2.3% to the employment change in the Metal mining sector. This positive net employment impact contrasts with the historical period, where the large negative impact due to imports left the net employment impact of foreign trade down -13.4%.

Another contrast from the historical period is given by the Primary metals sector. This sector also experienced a negative net employment impact due to foreign trade over the historical period, but a positive net employment impact over the projected period. The growth in exports contributed 12.8% to the projected employment growth, while import growth negated only -11.3% of this gain. The projected net positive employment impact due to foreign trade, 1.5%, results from the continuation of the trend toward customized products in this sector.

Similarities with the historical period also are represented in these results. For instance, many of the sectors whose employment was helped most by exports over the historical period are projected to benefit significantly over the projections period as well. Those industries which are projected to experience large employment impacts due to export growth will also feel large employment impacts as a result of import growth. In fact, six of the ten industries projected to be most helped by export growth over the 1990-2005 period also appear below in Table 8, the list of the ten industries most hurt by import growth over the projections period.

Table 8 presents the results of the factor analysis for the projection period after the results have been sorted by imports. The results are displayed only for those ten sectors projected to be most hurt by the growth in imports over the 1990-2005 period. The positive impact of the overall growth in foreign trade is apparent in that only five of the ten sectors listed are projected to have negative net employment impacts due to foreign trade. Additionally, many of the sectors in the list of the top ten sectors most hurt by import growth were also in the list for the historical period.

For instance, the sector projected to be most hurt by import growth over the 1990-2005 period, the Leather and leather products sector, also appeared in Table 5, the historical period list. The Leather and leather products sector is projected to be significantly hurt by import growth over the 1990-2005 period. The employment impact of the growth in imports on this sector will be -47.7%. Exports are projected to offset this decline by only 8.2%, so the net employment impact of foreign trade will be -39.5%. Imports were the primary source of supply for the Leather and leather products sector in 1990. The projected growth in imports for 1990-2005, therefore, accounts for this large negative employment impact.

Two sectors which are included in Table 8, but did not make the list of sectors hardest hit by import growth during the historical period, were the Apparel and other finished products sector and the Textile mill products sector. Employment in these two sectors is closely tied to the U.S. foreign trade position. In the Apparel and other finished products sector, for instance, the growth in imports over the 1990-2005 period is expected to push employment down -13.5%, while export growth over the period will only offset this decline by 2.9%. The net employment impact of foreign trade in the Apparel and other finished products sector, therefore, is projected to be -10.6%.

The aging U.S. population will contribute to the increased demand for the output of this sector. According to Franklin and Carey (1992), "Demand for the products of the apparel industry is expected to increase faster than the growth of the population, but more of that demand is projected to be met through imports."[13] This expected increase in market share coming from imports is partially due to the 807 classification in the U.S. tariff schedule. This classification reduces the import duties paid by U.S. manufacturers. According to Anne Clymer, it allows domestically produced and cut parts to be assembled offshore. Because assembly is the most labor-intensive portion of production, 807 exports assembly to low-wage countries - usually Caribbean or Latin American. Import tariffs are imposed only for value added offshore. This provision adversely affects the apparel assembly industry but could help strengthen the textile industry.[14]

The strengthening of the textile industry would come through export growth over the 1990-2005 period. Employment in the Textile mill products sector is projected to experience a 6.7% positive impact due to the growth of exports over the projection period. This positive impact will be negated, however, by the - 10.0% employment impact due to the growth of imports over the period. The resulting net employment impact of foreign trade in this sector over the 1990-2005 period, therefore, will be -3.3%. The projection of a net negative employment impact due to foreign trade might be affected by several trade agreements which are now pending legislative approval.

The North American Free Trade Agreement (NAFTA), for instance, would require both Canada and Mexico to purchase at least 80% of their textiles within North America. "Since neither Canada nor Mexico has a domestic textile industry, the U.S. textile industry will benefit from the expanded market."[15] Furthermore, as more apparel manufacturers move production to low-wage Mexico, textiles from U.S. manufacturers will become more economically appealing due to their quicker service and lower transportation costs.[16]

The beneficial effect of NAFTA on exports could be nullified by the impending phase-out of the MultiFiber Arrangement (MFA). According to Clymer (1992), the MFA allows countries to impose import quotas and tariffs on textiles that would be illegal under the General Agreement on Tariffs and Trade (GATT). As this agreement is slowly abandoned, imports to the U.S. could increase. Due to the uncertainty surrounding the passage of these trade arrangements, however, the net effect on employment in the Textile mill products sector is unclear at this point.

IV. Conclusion

The results of this factor analysis provide clear evidence that foreign trade is indeed beneficial to the U.S. economy. This conclusion, which most proponents of comparative advantage would argue is obvious, tends to get mired in the protectionist bantering of many labor leaders. While the industries represented by these labor leaders are typically the hardest hit by import growth, these same industries are often buoyed by strong export growth. As a whole, the U.S. economy benefits from increased foreign trade. The net employment impact of foreign trade, in most cases, is positive. For this reason, the U.S. must not succumb to protectionist pressures, but rather encourage growth in foreign trade.

Footnotes

1. Michael P. Botos is an economist for the U.S. Department of Labor, Bureau of Labor Statistics, Office of Employment Projections.

2. The historical period covers the years 1977-1990, while the projected period extends from 1990-2005. The projected data was provided by the U.S. Bureau of Labor Statistics and appeared in the Monthly Labor Review, November 1991. The 228 industries identified by the BLS were aggregated to 55 sectors for the purposes of this analysis. These 55 sectors correspond roughly to a two-digit SIC aggregation and are shown in Appendix B of this paper.

3. For the purposes of this analysis, GNP was used rather than GDP because the most recent BLS long-term employment projections were based on the GNP concept. The revision to GDP, however, would not alter the overall conclusions of this paper. 4/ Bureau of Economic Analysis, "Mathematical Derivation of the Total Requirements Tables for the 1977 Input-Output Study," Survey of Current Business, Vol. 64, N. 5, May 1984, pp. 42-78.

4. The problem of the interaction effect and some possible approaches to dealing with the phenomenon were first raised in H. S. Levine, "A Small Problem in the Analysis of Growth," Review of Economics and Statistics, May 1960, pp. 225-228, and in B. F. Massell, "Another Small Problem in the Analysis of Growth," Review of Economics and Statistics, April 1962, pp. 330-332. Later studies, which include S. J. Feldman et al, "Sources of Structural Change in the United States, 1963-78: An Input-Output Perspective," Review of Economics and Statistics, August 1987, pp. 503-510, and D. Fujimagari, "The Sources of Change in Canadian Industry Output," Economic Systems Research, Vol. 1, No. 2, 1989, pp. 187-201, described the interaction effect more thoroughly and provided more sophisticated approaches to the redistribution of this effect to each of the primary factors.

5. Carey, Max L., and Franklin, James C., "Industry Output and Job Growth Continues to Slow into the Next Century," Outlook 1990-2005, U.S. Department of Labor, Bureau of Labor Statistics, BLS Bulletin 2402, May 1992, p. 54.

6. Ibid., p. 54.

7. Ibid., p. 55.

8. Ibid., p. 55.

9. Table 9 and Table 10 at the end of this paper show the actual employment change for each sector over both the historical and projected period.

10. Ibid., p. 47.

11. Saunders, Norman C., "The U.S. Economy into the 21st Century," Outlook 1990-2005, U.S. Department of Labor, Bureau of Labor Statistics, BLS Bulletin 2402, May 1992, p. 16.

12. Carey, Max L., and Franklin, James C., "Industry Output and Job Growth Continues to Slow into the Next Century," Outlook 1990-2005, U.S. Department of Labor, Bureau of Labor Statistics, BLS Bulletin 2402, May 1992, p. 49.

13. Clymer, Anne W., "Industry Study, Textile Mill Products (SIC 22)," Internal Study for the Office of Employment Projections, U.S. Department of Labor, Bureau of Labor Statistics, July 1992, p. 4.

14. Ibid., pp. 3-4.

15. Bacon, Kenneth H., "Quick Reaction: Trade Pact is Likely to Step Up Business Even Before Approval," The Wall Street Journal, Dow Jones & Company, Inc., August 13, 1992, p. A4.

Appendix A.: Derivation of the Total Requirements Table

This appendix will describe the mathematical derivation of the Total Requirements tables used in this analysis. According to the Bureau of Economic Analysis, the original derivation of the tables used in this analysis follows the System of National Accounts recommended by the United Nations.[1] The matrix algebra will be described verbally, as well as through equations, in order to facilitate understanding. All of the following calculations were performed on 55-order matrices. For further detail, please refer to the sources cited.

The first step in deriving a Total Requirements table is to divide the columns of the Use table by each industry's total output giving a Direct Requirements table. The Direct Requirements table shows the amount of a commodity "used" by an industry to produce each dollar of that industry's output. In addition to a Direct Requirements table, however, a Market Share matrix is required. The Market Share matrix shows the percentage of each commodity's output "made" in each industry. The Market Share matrix, therefore, is calculated by dividing the Make table (adjusted for scrap) by each commodity's total output.

The only other matrices needed for the calculations are the Non-Scrap Production matrix and an Identity matrix. The Non-Scrap production matrix represents that portion of an industry's output which is not scrap. This matrix is derived by first finding the ratio of scrap production to total industry output in each industry using the Make table. This ratio is subtracted from the Identity matrix giving a matrix of non-scrap production percentages. The non-scrap production ratio matrix is inverted and multiplied by the Market Share matrix giving the Non-Scrap Production matrix.

The Total Requirements table is then attained through another series of matrix calculations. The Non-Scrap Production matrix is multiplied by the Direct Requirements table. The product of this calculation is then subtracted from the Identity matrix and inverted, which yields the commodity-by-commodity Total Requirements table. The industry-by-commodity Total Requirements table, which was the desired table for this analysis, is simply the product of the multiplication of this commodity-by-commodity table by the Non-Scrap Production matrix.

The equations that correspond to the matrix algebra discussed above appear below:

$$B = U/g$$
$$D = Madj/q$$
$$W = D * (I - p)^{-1}$$
$$TReq = W * [I - (B * W)]^{-1}$$

Where the variables include:

B:	Direct Requirements table, 55 x 55
U:	Use table, 55 x 55
g:	Industry output, 1 x 55
D:	Market Share matrix, 55 x 55
Madj:	Make table (adjusted for scrap production), 55x55
q:	Commodity output, 1 x 55
W:	Non-Scrap Production matrix, 55 x 55
I:	Identity matrix, 55 x 55
p:	Matrix of the ratio of scrap to total industry output, 55 x 55
TReq:	Industry-by-commodity Total Requirements table, 55 x 55

Footnotes

1. United Nations, "A System of National Accounts," Studies in Methods, Series F, No. 2, Rev. 3, United Nations, New York, 1968. In addition see, Stone, R., Bacharach, M., & Bates, J., "Input-Output Relationships, 1951-1966," A Programme for Growth, Volume 3, London, Chapman, and Hall, 1963.

Appendix B.: 55-Order Sectoring Plan

Sector #	Sector Title	1987 SIC
Sector 1	Agriculture, livestock, and animal specialties	01,02
Sector 2	Agricultural services, forestry, and fishing	07,08,09
Sector 3	Metal Mining	10
Sector 4	Coal mining	12
Sector 5	Oil and gas extraction	13
Sector 6	Nonmetallic minerals except fuels	14
Sector 7	Construction	15,16,17
Sector 8	Lumber and wood products, except furniture	24
Sector 9	Furniture and fixtures	25
Sector 10	Stone, clay, glass, and concrete products	32
Sector 11	Primary metal industries	33
Sector 12	Fabricated metal products exc. machinery	34
Sector 13	Industrial, commercial mach., computer equip.	35
Sector 14	Electronic & other electrical equipment	36
Sector 15	Transportation equipment	37
Sector 16	Measuring, analyzing, and controlling instr.	38
Sector 17	Miscellaneous manufacturing industries	39
Sector 18	Food and kindred products	20
Sector 19	Tobacco products	21
Sector 20	Textile mill products	22
Sector 21	Apparel and other finished products	23
Sector 22	Paper and allied products	26
Sector 23	Printing, publishing, and allied industries	27
Sector 24	Chemicals and allied products	28
Sector 25	Petroleum refining and related industries	29
Sector 26	Rubber and miscellaneous plastics products	30
Sector 27	Leather and leather products	31
Sector 28	Transportation	40-47
Sector 29	Communications	48
Sector 30	Electric, gas, and sanitary services	49
Sector 31	Wholesale trade	50-51
Sector 32	Retail trade, exc. eating and drinking places	52-57,59
Sector 33	Eating and drinking places	58
Sector 34	Finance, insurance, and real estate	60-67
Sector 35	Hotels and other lodging places	70
Sector 36	Personal services	72
Sector 37	Business services	73,873*,874
Sector 38	Automotive repair, services, and parking	75
Sector 39	Miscellaneous repair services	76
Sector 40	Motion pictures	78
Sector 41	Amusement and recreation services	79
Sector 42	Health services	80
Sector 43	Legal services	81
Sector 44	Educational services	82
Sector 45	Social services	83
Sector 46	All other services	**
Sector 47	Private households	88
Sector 48	Federal government enterprises, exc CCC	
Sector 49	State and local government enterprises	
Sector 50	Noncomparable imports	
Sector 51	Scrap, used, and second-hand goods	
Sector 52	Government compensation	
Sector 53	Rest of the world industry	
Sector 54	Inventory valuation adjustment	
Sector 55	Commodity Credit Corporation	

* Excluding SIC 8733
** Includes SICs 84,86,871,872,8733,89

Table 1. General Results: 1977-1990

[Percentage Change in Employment]

Sector #	Title	TOTAL	GNP	GNPDIST	FDDIST	TReq	E/O	INTERACT
				FACTOR IMPACT				
Sector 37	Business services	182.8%	40.5%	1.7%	12.6%	64.7%	10.4%	53.0%
Sector 45	Social services	130.2%	40.5%	1.4%	73.4%	2.1%	-7.9%	20.7%
Sector 2	Agricultural services, forestry, and fishing.	105.4%	40.5%	-4.6%	-1.9%	-3.2%	53.5%	21.0%
Sector 40	Motion pictures	97.9%	40.5%	5.1%	27.4%	11.9%	-6.7%	19.6%
Sector 43	Legal services	94.1%	40.5%	2.2%	8.7%	8.8%	15.4%	18.3%
Sector 38	Automotive repair, services, and parking	73.1%	40.5%	1.6%	0.2%	7.8%	12.2%	10.7%
Sector 42	Health services	67.6%	40.5%	1.2%	22.3%	-0.3%	-3.3%	7.2%
Sector 33	Eating and drinking places	60.9%	40.5%	2.0%	-2.7%	-0.4%	16.2%	5.3%
Sector 35	Hotels and other lodging places	58.3%	40.5%	1.8%	-5.1%	-11.0%	33.3%	-1.3%
Sector 41	Amusement and recreation services	55.3%	40.5%	1.6%	21.8%	4.6%	-14.0%	0.8%
Sector 34	Finance, insurance and real estate	53.0%	40.5%	0.9%	-0.6%	1.4%	6.3%	4.5%
Sector 46	All other services	52.9%	40.5%	-0.0%	-1.5%	14.9%	-1.1%	0.1%
Sector 44	Educational services	51.5%	40.5%	-0.6%	10.1%	-0.4%	-4.0%	5.8%
Sector 23	Printing, publishing, and allied industries..	40.2%	40.5%	1.1%	7.2%	1.9%	-5.9%	-4.6%
Sector 7	Construction	36.9%	40.5%	-15.4%	3.0%	3.2%	7.8%	-2.3%
Sector 39	Miscellaneous repair services	36.7%	40.5%	1.9%	-7.0%	1.1%	1.1%	-0.8%
Sector 49	State and local government enterprises	36.7%	40.5%	0.8%	7.2%	-3.5%	-7.8%	-0.6%
Sector 36	Personal services	31.4%	40.5%	1.3%	2.7%	-1.4%	-8.8%	-3.0%
Sector 31	Wholesale trade	30.9%	40.5%	8.0%	-5.2%	-6.4%	-4.7%	-1.3%
Sector 30	Electric, gas, and sanitary services	29.0%	40.5%	-0.6%	-3.8%	-23.5%	24.3%	-7.9%
Sector 32	Retail trade, exc. eating and drinking places	28.0%	40.5%	1.2%	5.7%	6.7%	-20.1%	-6.0%
Sector 28	Transportation	27.8%	40.5%	5.4%	-7.0%	-7.7%	-0.9%	-2.4%
Sector 48	Federal government enterprises	22.3%	40.5%	1.6%	2.9%	6.2%	-21.2%	-7.7%
Sector 52	Government compensation	20.3%	40.5%	-0.5%	-12.7%	-0.0%	-0.3%	-6.8%
Sector 26	Rubber and miscellaneous plastics products	18.3%	40.5%	0.3%	5.1%	7.2%	-17.2%	-17.6%
Sector 16	Measuring, analyzing, and controlling instr..	12.2%	40.5%	14.5%	8.7%	11.4%	-41.7%	-21.3%
Sector 29	Communications	11.3%	40.5%	3.9%	11.3%	9.3%	-36.0%	-17.7%
Sector 9	Furniture and fixtures	10.8%	40.5%	5.6%	-10.6%	-0.2%	-18.3%	-6.2%
Sector 15	Transportation equipment	6.8%	40.5%	11.3%	13.4%	-6.4%	-10.4%	-41.6%
Sector 5	Oil and gas extraction	4.9%	40.5%	-34.6%	-6.4%	-39.3%	20.2%	24.4%
Sector 22	Paper and allied products	2.4%	40.5%	-1.7%	-4.3%	-3.4%	-23.2%	-5.5%
Sector 24	Chemicals and allied products	2.1%	40.5%	1.4%	-11.2%	-12.2%	-14.8%	-1.7%
Sector 8	Lumber and wood products, except furniture	0.3%	40.5%	-15.0%	3.9%	-9.2%	-14.1%	-5.8%
Sector 14	Electronic and other electrical equipment	0.3%	40.5%	6.3%	13.9%	12.4%	-39.9%	-32.9%
Sector 50	Noncomparable imports	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Sector 51	Scrap, used and secondhand goods	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Sector 53	Rest of the world industry	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Sector 54	Inventory valuation adjustment	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Sector 55	Commodity credit corporation	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Sector 18	Food and kindred products	-2.6%	40.5%	1.2%	-11.1%	-4.4%	-19.8%	-9.0%
Sector 13	Industrial, commercial mach., computer equip.	-4.0%	40.5%	19.6%	17.8%	0.1%	-48.0%	-34.0%
Sector 6	Nonmetallic minerals, except fuels	-6.0%	40.5%	-7.3%	-21.8%	-8.5%	-17.1%	8.2%
Sector 17	Miscellaneous manufacturing industries	-7.9%	40.5%	-6.6%	-36.1%	-5.7%	-16.2%	16.2%
Sector 12	Fabricated metal products, except machinery..	-9.6%	40.5%	1.9%	-0.5%	-17.2%	-13.6%	-20.8%
Sector 10	Stone, clay, glass, and concrete products	-11.9%	40.5%	-8.6%	3.9%	-16.9%	-14.9%	-16.0%
Sector 21	Apparel and other finished products	-19.5%	40.5%	-12.8%	-12.0%	-1.9%	-21.9%	-11.4%
Sector 25	Petroleum refining and related industries	-21.7%	40.5%	-6.2%	-25.3%	-15.3%	-25.6%	10.1%
Sector 1	Agriculture, livestock and animal specialties	-21.8%	40.5%	7.8%	-25.6%	-5.6%	-31.8%	-7.3%
Sector 20	Textile mill products	-23.4%	40.5%	-8.1%	-12.5%	-4.7%	-32.8%	-5.9%
Sector 47	Private households	-28.3%	40.5%	1.3%	-26.0%	0.0%	-31.9%	-12.2%
Sector 19	Tobacco products	-30.6%	40.5%	7.3%	-32.9%	-2.9%	-31.9%	-10.6%
Sector 3	Metal mining	-34.6%	40.5%	-4.7%	-30.0%	-29.1%	-47.0%	35.7%
Sector 4	Coal mining	-34.6%	40.5%	5.3%	-20.0%	-2.9%	-55.5%	-2.0%
Sector 11	Primary metal industries	-35.8%	40.5%	-0.6%	11.0%	-47.6%	-20.9%	-18.3%
Sector 27	Leather and leather products	-47.8%	40.5%	-18.2%	-18.6%	-5.7%	-16.6%	-29.3%
Total		31.1%	40.5%	0.9%	0.1%	1.0%	-7.8%	-3.6%

Table 2. General Results: 1990–2005

[Percentage Change in Employment]

Sector #	Title	TOTAL	FACTOR IMPACT GNP	GNPDIST	FDDIST	TReq	E/O	INTERACT
Sector 40	Motion pictures..............................	73.7%	40.5%	1.9%	-3.3%	13.7%	7.8%	13.0%
Sector 45	Social services..............................	58.4%	40.5%	-0.5%	20.7%	1.1%	-7.0%	3.6%
Sector 37	Business services............................	50.6%	40.5%	-0.3%	5.9%	16.6%	-11.1%	-0.9%
Sector 42	Health services..............................	47.4%	40.5%	-0.3%	16.0%	0.3%	-9.5%	0.4%
Sector 43	Legal services...............................	42.2%	40.5%	1.0%	4.9%	0.2%	-4.2%	-0.3%
Sector 44	Educational services.........................	41.6%	40.5%	1.8%	-5.3%	-0.2%	5.1%	-0.2%
Sector 2	Agricultural services, forestry, and fishing.	33.5%	40.5%	-2.8%	5.4%	-7.1%	-3.0%	0.4%
Sector 38	Automotive repair, services, and parking.....	32.9%	40.5%	0.4%	-3.3%	-1.5%	-0.8%	-2.4%
Sector 35	Hotels and other lodging places..............	31.2%	40.5%	-0.0%	1.5%	-9.8%	2.5%	-3.6%
Sector 41	Amusement and recreation services............	30.9%	40.5%	-0.3%	14.3%	1.3%	-19.1%	-5.8%
Sector 33	Eating and drinking places...................	29.5%	40.5%	0.3%	-0.3%	-2.1%	-5.5%	-3.4%
Sector 49	State and local government enterprises.......	27.7%	40.5%	-0.3%	4.8%	-11.0%	-3.2%	-3.1%
Sector 46	All other services...........................	27.1%	40.5%	-1.2%	-2.2%	-2.3%	-3.0%	-4.7%
Sector 28	Transportation...............................	24.7%	40.5%	3.0%	-1.3%	-0.8%	-11.7%	-5.0%
Sector 39	Miscellaneous repair services................	21.2%	40.5%	-0.5%	-7.8%	-6.9%	2.1%	-6.2%
Sector 23	Printing, publishing, and allied industries..	20.8%	40.5%	1.7%	-2.4%	7.5%	-17.8%	-8.6%
Sector 9	Furniture and fixtures.......................	20.7%	40.5%	1.5%	4.2%	-0.2%	-20.0%	-5.3%
Sector 34	Finance, insurance and real estate...........	20.7%	40.5%	0.3%	-2.5%	-0.4%	-11.8%	-5.3%
Sector 36	Personal services............................	19.5%	40.5%	-0.2%	-6.6%	-0.7%	-8.0%	-5.6%
Sector 32	Retail trade, exc. eating and drinking places	18.4%	40.5%	-0.2%	1.5%	2.7%	-18.8%	-7.3%
Sector 7	Construction.................................	17.8%	40.5%	-7.2%	1.3%	1.1%	-11.8%	-6.2%
Sector 26	Rubber and miscellaneous plastics products...	17.1%	40.5%	6.8%	-2.3%	22.1%	-30.1%	-19.8%
Sector 52	Government compensation......................	17.1%	40.5%	-10.2%	-4.7%	0.0%	-2.2%	-6.2%
Sector 30	Electric, gas, and sanitary services.........	16.0%	40.5%	-1.1%	-4.1%	-2.2%	-11.0%	-6.1%
Sector 31	Wholesale trade..............................	14.5%	40.5%	4.4%	-4.3%	-2.7%	-17.1%	-6.4%
Sector 48	Federal government enterprises...............	12.9%	40.5%	-0.0%	-2.3%	3.5%	-20.3%	-8.4%
Sector 3	Metal mining.................................	8.9%	40.5%	-34.8%	11.0%	-26.3%	-9.4%	28.0%
Sector 22	Paper and allied products....................	3.9%	40.5%	-0.4%	-2.7%	-1.7%	-23.7%	-8.1%
Sector 6	Nonmetallic minerals, except fuels...........	2.7%	40.5%	-14.3%	-2.5%	-5.6%	-18.1%	2.7%
Sector 17	Miscellaneous manufacturing industries.......	2.4%	40.5%	-33.1%	-6.6%	-1.4%	-15.5%	18.5%
Sector 16	Measuring, analyzing, and controlling instr..	1.1%	40.5%	2.7%	5.6%	2.0%	-34.1%	-15.6%
Sector 24	Chemicals and allied products................	0.4%	40.5%	0.2%	-2.4%	-4.6%	-26.4%	-7.0%
Sector 50	Noncomparable imports........................	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Sector 51	Scrap, used and secondhand goods.............	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Sector 53	Rest of the world industry...................	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Sector 54	Inventory valuation adjustment...............	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Sector 55	Commodity credit corporation.................	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Sector 8	Lumber and wood products, except furniture...	-3.9%	40.5%	-2.1%	-3.7%	3.9%	-28.2%	-14.4%
Sector 15	Transportation equipment.....................	-4.5%	40.5%	18.6%	-9.9%	0.3%	-25.4%	-28.6%
Sector 5	Oil and gas extraction.......................	-5.0%	40.5%	-29.0%	-6.9%	-17.3%	3.4%	4.3%
Sector 14	Electronic and other electrical equipment....	-6.4%	40.5%	8.0%	5.6%	11.0%	-41.7%	-29.7%
Sector 18	Food and kindred products....................	-6.7%	40.5%	0.4%	-11.3%	-2.5%	-24.2%	-9.6%
Sector 10	Stone, clay, glass, and concrete products....	-6.9%	40.5%	0.5%	-3.0%	-5.2%	-23.5%	-16.1%
Sector 13	Industrial, commercial mach., computer equip.	-7.4%	40.5%	15.1%	31.0%	5.3%	-58.1%	-41.3%
Sector 29	Communications...............................	-12.6%	40.5%	0.9%	-1.9%	2.7%	-38.3%	-16.4%
Sector 12	Fabricated metal products, except machinery..	-13.1%	40.5%	3.1%	-4.8%	-6.7%	-27.9%	-17.4%
Sector 20	Textile mill products........................	-13.8%	40.5%	-5.4%	-4.0%	-0.4%	-34.2%	-10.4%
Sector 11	Primary metal industries.....................	-14.9%	40.5%	1.0%	2.4%	-20.4%	-23.6%	-14.8%
Sector 21	Apparel and other finished products..........	-19.0%	40.5%	-9.1%	4.9%	6.2%	-44.8%	-16.6%
Sector 25	Petroleum refining and related industries....	-22.5%	40.5%	-13.1%	-15.4%	-7.3%	-27.9%	0.7%
Sector 4	Coal mining..................................	-23.7%	40.5%	-6.2%	-8.2%	-8.5%	-38.2%	-3.3%
Sector 1	Agriculture, livestock and animal specialties	-25.5%	40.5%	-0.5%	-9.7%	-10.3%	-38.9%	-6.6%
Sector 47	Private households...........................	-31.1%	40.5%	-0.3%	-27.0%	0.0%	-32.6%	-11.7%
Sector 19	Tobacco products.............................	-31.8%	40.5%	5.7%	-35.6%	-0.1%	-31.0%	-11.4%
Sector 27	Leather and leather products.................	-45.7%	40.5%	-30.6%	-4.9%	-2.1%	-36.3%	-12.3%
Total..		20.1%	40.5%	-1.1%	0.6%	0.7%	-13.9%	-6.7%

Table 3. **Component Impact of GNPDist Factor: 1977-1990**

[Percentage Change in Employment]

Sector #	Title	GNPDIST	PCE	Invest.	Exp.	Imp.	Def.	OthGov.
			COMPONENT IMPACT					
Sector 37	Business services	1.7%	0.9%	-0.3%	4.7%	-3.0%	0.7%	-1.3%
Sector 45	Social services	1.4%	1.3%	0.0%	0.1%	-0.1%	0.1%	-0.0%
Sector 2	Agricultural services, forestry, and fishing.	-4.6%	1.1%	-3.6%	6.5%	-8.7%	0.2%	-0.1%
Sector 40	Motion pictures	5.1%	1.0%	-1.3%	6.3%	-1.0%	0.5%	-0.4%
Sector 43	Legal services	2.2%	1.1%	0.0%	2.6%	-1.0%	0.1%	-0.6%
Sector 38	Automotive repair, services, and parking	1.6%	1.1%	-0.2%	2.1%	-1.3%	0.2%	-0.3%
Sector 42	Health services	1.2%	1.3%	-0.0%	0.0%	-0.0%	0.0%	-0.2%
Sector 33	Eating and drinking places	2.0%	1.3%	0.1%	1.6%	-1.0%	0.2%	-0.1%
Sector 35	Hotels and other lodging places	1.8%	1.1%	-0.0%	2.2%	-1.4%	0.4%	-0.5%
Sector 41	Amusement and recreation services	1.6%	1.3%	0.1%	0.6%	-0.3%	0.1%	-0.1%
Sector 34	Finance, insurance and real estate	0.9%	1.2%	-1.0%	2.2%	-1.2%	0.1%	-0.3%
Sector 46	All other services	-0.0%	0.8%	-3.4%	4.5%	-1.4%	0.3%	-0.8%
Sector 44	Educational services	-0.6%	1.2%	0.0%	0.3%	-0.2%	0.1%	-1.9%
Sector 23	Printing, publishing, and allied industries.	1.1%	0.9%	-0.7%	4.0%	-2.3%	0.4%	-1.2%
Sector 7	Construction	-15.4%	0.2%	-14.4%	0.9%	-0.7%	0.4%	-1.7%
Sector 39	Miscellaneous repair services	1.9%	1.0%	-0.8%	4.5%	-2.9%	0.7%	-0.6%
Sector 49	State and local government enterprises	0.8%	1.1%	-0.5%	3.6%	-3.2%	0.4%	-0.6%
Sector 36	Personal services	1.3%	1.3%	-0.0%	0.5%	-0.4%	0.0%	-0.1%
Sector 31	Wholesale trade	8.0%	0.8%	1.5%	7.2%	-1.5%	0.4%	-0.5%
Sector 30	Electric, gas, and sanitary services	-0.6%	1.1%	-0.4%	3.6%	-4.6%	0.4%	-0.7%
Sector 32	Retail trade, exc. eating and drinking places	1.2%	1.2%	-0.1%	0.5%	-0.3%	0.1%	-0.2%
Sector 28	Transportation	5.4%	0.9%	-0.7%	8.3%	-3.2%	0.7%	-0.6%
Sector 48	Federal government enterprises	1.6%	1.1%	-0.3%	3.3%	-2.2%	0.4%	-0.6%
Sector 52	Government compensation	-0.5%	-0.0%	-0.0%	-0.0%	-0.0%	3.9%	-4.4%
Sector 26	Rubber and miscellaneous plastics products	0.3%	0.8%	-1.1%	8.3%	-8.0%	0.8%	-0.6%
Sector 16	Measuring, analyzing, and controlling instr.	14.5%	0.4%	9.9%	9.6%	-6.0%	1.6%	-0.9%
Sector 29	Communications	3.9%	1.0%	1.5%	3.2%	-1.4%	0.4%	-0.7%
Sector 9	Furniture and fixtures	5.6%	0.8%	6.2%	1.8%	-2.9%	0.1%	-0.5%
Sector 15	Transportation equipment	11.3%	0.6%	6.3%	10.6%	-8.0%	2.1%	-0.4%
Sector 5	Oil and gas extraction	-34.6%	1.4%	-7.1%	7.3%	-36.2%	0.9%	-0.9%
Sector 22	Paper and allied products	-1.7%	1.0%	-2.0%	7.7%	-8.0%	0.4%	-0.8%
Sector 24	Chemicals and allied products	1.4%	0.9%	-2.4%	11.1%	-8.2%	0.6%	-0.6%
Sector 8	Lumber and wood products, except furniture	-15.0%	0.4%	-12.6%	5.9%	-7.7%	0.4%	-1.3%
Sector 14	Electronic and other electrical equipment	6.3%	0.6%	4.6%	10.0%	-9.4%	1.2%	-0.7%
Sector 50	Noncomparable imports	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Sector 51	Scrap, used and secondhand goods	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Sector 53	Rest of the world industry	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Sector 54	Inventory valuation adjustment	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Sector 55	Commodity credit corporation	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Sector 18	Food and kindred products	1.2%	1.3%	-0.1%	3.4%	-3.3%	0.1%	-0.2%
Sector 13	Industrial, commercial mach., computer equip.	19.6%	0.3%	12.6%	12.8%	-6.3%	0.7%	-0.4%
Sector 6	Nonmetallic minerals, except fuels	-7.3%	0.6%	-8.5%	9.8%	-8.6%	0.5%	-1.0%
Sector 17	Miscellaneous manufacturing industries	-6.6%	1.2%	-2.2%	5.4%	-10.7%	0.2%	-0.5%
Sector 12	Fabricated metal products, except machinery.	1.9%	0.5%	-1.2%	8.1%	-5.9%	1.1%	-0.8%
Sector 10	Stone, clay, glass, and concrete products	-8.6%	0.6%	-7.9%	5.6%	-6.0%	0.5%	-1.2%
Sector 21	Apparel and other finished products	-12.8%	1.3%	-8.9%	2.5%	-7.9%	0.3%	-0.1%
Sector 25	Petroleum refining and related industries	-6.2%	1.1%	-3.2%	5.1%	-9.2%	0.7%	-0.7%
Sector 1	Agriculture, livestock and animal specialties	7.8%	1.1%	0.7%	10.3%	-3.5%	0.1%	-0.9%
Sector 20	Textile mill products	-8.1%	1.1%	-7.1%	6.0%	-8.1%	0.3%	-0.3%
Sector 47	Private households	1.3%	1.3%	0.0%	0.0%	0.0%	0.0%	0.0%
Sector 19	Tobacco products	7.3%	1.1%	-2.7%	10.2%	-1.4%	0.0%	-0.0%
Sector 3	Metal mining	-4.7%	0.7%	7.8%	19.0%	-32.4%	1.1%	-0.9%
Sector 4	Coal mining	5.3%	0.8%	-0.7%	12.2%	-6.9%	0.6%	-0.7%
Sector 11	Primary metal industries	-0.6%	0.6%	1.8%	12.2%	-15.6%	1.2%	-0.8%
Sector 27	Leather and leather products	-18.2%	1.7%	-3.2%	3.7%	-20.5%	0.1%	-0.0%
Total		0.9%	0.8%	-0.7%	3.6%	-2.6%	0.9%	-1.1%

Table 4. Component Impact of GNPDist Factor: 1977–1990

[Ten Sectors Benefitting Most From Export Growth]

Sector #	Title	GNPDIST	COMPONENT IMPACT PCE	Invest.	Exp.	Imp.	Def.	OthGov.
Sector 3	Metal mining................................	-4.7%	0.7%	7.8%	19.0%	-32.4%	1.1%	-0.9%
Sector 13	Industrial, commercial mach., computer equip.	19.6%	0.3%	12.6%	12.8%	-6.3%	0.7%	-0.4%
Sector 4	Coal mining.................................	5.3%	0.8%	-0.7%	12.2%	-6.9%	0.6%	-0.7%
Sector 11	Primary metal industries.....................	-0.6%	0.6%	1.8%	12.2%	-15.6%	1.2%	-0.8%
Sector 24	Chemicals and allied products................	1.4%	0.9%	-2.4%	11.1%	-8.2%	0.6%	-0.6%
Sector 15	Transportation equipment.....................	11.3%	0.6%	6.3%	10.6%	-8.0%	2.1%	-0.4%
Sector 1	Agriculture, livestock and animal specialties	7.8%	1.1%	0.7%	10.3%	-3.5%	0.1%	-0.9%
Sector 19	Tobacco products.............................	7.3%	1.1%	-2.7%	10.2%	-1.4%	0.0%	-0.0%
Sector 14	Electronic and other electrical equipment....	6.3%	0.6%	4.6%	10.0%	-9.4%	1.2%	-0.7%
Sector 6	Nonmetallic minerals, except fuels...........	-7.3%	0.6%	-8.5%	9.8%	-8.6%	0.5%	-1.0%

Table 5. Component Impact of GNPDist Factor: 1977–1990

[Ten Sectors Hardest Hit By Import Growth]

Sector #	Title	GNPDIST	COMPONENT IMPACT PCE	Invest.	Exp.	Imp.	Def.	OthGov.
Sector 5	Oil and gas extraction.......................	-34.6%	1.4%	-7.1%	7.3%	-36.2%	0.9%	-0.9%
Sector 3	Metal mining................................	-4.7%	0.7%	7.8%	19.0%	-32.4%	1.1%	-0.9%
Sector 27	Leather and leather products.................	-18.2%	1.7%	-3.2%	3.7%	-20.5%	0.1%	-0.0%
Sector 11	Primary metal industries.....................	-0.6%	0.6%	1.8%	12.2%	-15.6%	1.2%	-0.8%
Sector 17	Miscellaneous manufacturing industries.......	-6.6%	1.2%	-2.2%	5.4%	-10.7%	0.2%	-0.5%
Sector 14	Electronic and other electrical equipment....	6.3%	0.6%	4.6%	10.0%	-9.4%	1.2%	-0.7%
Sector 25	Petroleum refining and related industries....	-6.2%	1.1%	-3.2%	5.1%	-9.2%	0.7%	-0.7%
Sector 2	Agricultural services, forestry, and fishing.	-4.6%	1.1%	-3.6%	6.5%	-8.7%	0.2%	-0.1%
Sector 6	Nonmetallic minerals, except fuels...........	-7.3%	0.6%	-8.5%	9.8%	-8.6%	0.5%	-1.0%
Sector 24	Chemicals and allied products................	1.4%	0.9%	-2.4%	11.1%	-8.2%	0.6%	-0.6%

Table 6. **Component Impact of GNPDist Factor: 1990-2005**

[Percentage Change in Employment]

Sector #	Title	GNPDIST	PCE	Invest.	Exp.	Imp.	Def.	OthGov.
			COMPONENT IMPACT					
Sector 40	Motion pictures	1.9%	-0.3%	-0.8%	4.9%	-0.5%	-1.3%	-0.1%
Sector 45	Social services	-0.5%	-0.3%	0.0%	0.1%	-0.1%	-0.2%	-0.0%
Sector 37	Business services	-0.3%	-0.2%	0.9%	4.2%	-2.2%	-2.8%	-0.2%
Sector 42	Health services	-0.3%	-0.3%	-0.0%	0.1%	-0.0%	-0.0%	-0.0%
Sector 43	Legal services	1.0%	-0.3%	0.6%	2.2%	-0.9%	-0.6%	-0.0%
Sector 44	Educational services	1.8%	-0.3%	0.1%	3.0%	-0.3%	-0.3%	-0.3%
Sector 2	Agricultural services, forestry, and fishing.	-2.8%	-0.3%	-2.4%	5.3%	-4.9%	-0.6%	0.2%
Sector 38	Automotive repair, services, and parking	0.4%	-0.3%	0.5%	2.1%	-1.0%	-0.9%	-0.0%
Sector 35	Hotels and other lodging places	-0.0%	-0.3%	0.9%	1.8%	-1.1%	-1.4%	0.1%
Sector 41	Amusement and recreation services	-0.3%	-0.3%	0.1%	0.5%	-0.2%	-0.3%	-0.0%
Sector 33	Eating and drinking places	0.3%	-0.3%	0.5%	1.4%	-0.7%	-0.6%	0.0%
Sector 49	State and local government enterprises	-0.3%	-0.3%	0.0%	2.8%	-1.6%	-1.2%	-0.0%
Sector 46	All other services	-1.2%	-0.2%	-1.2%	2.7%	-1.2%	-1.2%	-0.1%
Sector 28	Transportation	3.0%	-0.2%	-0.2%	7.2%	-1.6%	-2.2%	-0.1%
Sector 39	Miscellaneous repair services	-0.5%	-0.2%	0.1%	4.4%	-2.2%	-2.6%	-0.0%
Sector 23	Printing, publishing, and allied industries..	1.7%	-0.2%	1.9%	3.7%	-1.8%	-1.7%	-0.2%
Sector 9	Furniture and fixtures	1.5%	-0.2%	4.4%	2.1%	-4.1%	-0.6%	-0.1%
Sector 34	Finance, insurance and real estate	0.3%	-0.3%	-0.4%	2.1%	-0.8%	-0.4%	0.0%
Sector 36	Personal services	-0.2%	-0.3%	0.1%	0.4%	-0.2%	-0.1%	-0.0%
Sector 32	Retail trade, exc. eating and drinking places	-0.2%	-0.3%	0.1%	0.6%	-0.3%	-0.4%	-0.0%
Sector 7	Construction	-7.2%	-0.1%	-6.5%	1.0%	-0.5%	-1.1%	0.0%
Sector 26	Rubber and miscellaneous plastics products...	6.8%	-0.2%	8.0%	9.1%	-7.1%	-3.0%	-0.0%
Sector 52	Government compensation	-10.2%	0.0%	0.0%	0.0%	0.0%	-9.5%	-0.7%
Sector 30	Electric, gas, and sanitary services	-1.1%	-0.3%	0.1%	3.0%	-2.4%	-1.5%	-0.1%
Sector 31	Wholesale trade	4.4%	-0.2%	1.1%	6.1%	-1.0%	-1.5%	-0.0%
Sector 48	Federal government enterprises	-0.0%	-0.3%	0.6%	2.7%	-1.6%	-1.5%	-0.0%
Sector 3	Metal mining	-34.8%	-0.2%	-32.7%	20.4%	-18.1%	-4.1%	-0.1%
Sector 22	Paper and allied products	-0.4%	-0.2%	-0.6%	7.8%	-5.7%	-1.5%	-0.1%
Sector 6	Nonmetallic minerals, except fuels	-14.3%	-0.1%	-14.9%	6.6%	-4.4%	-1.5%	-0.0%
Sector 17	Miscellaneous manufacturing industries	-33.1%	-0.4%	-21.6%	4.8%	-15.0%	-0.8%	-0.1%
Sector 16	Measuring, analyzing, and controlling instr..	2.7%	-0.1%	6.6%	7.5%	-4.7%	-6.5%	-0.2%
Sector 24	Chemicals and allied products	0.2%	-0.2%	-2.4%	11.5%	-6.6%	-2.0%	-0.1%
Sector 50	Noncomparable imports	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Sector 51	Scrap, used and secondhand goods	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Sector 53	Rest of the world industry	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Sector 54	Inventory valuation adjustment	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Sector 55	Commodity credit corporation	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Sector 8	Lumber and wood products, except furniture...	-2.1%	-0.1%	-3.0%	6.7%	-4.4%	-1.3%	-0.0%
Sector 15	Transportation equipment	18.6%	-0.1%	25.4%	9.9%	-7.3%	-9.2%	-0.1%
Sector 5	Oil and gas extraction	-29.0%	-0.4%	-15.7%	8.5%	-19.0%	-2.4%	-0.0%
Sector 14	Electronic and other electrical equipment	8.0%	-0.2%	12.4%	13.4%	-12.0%	-5.6%	-0.1%
Sector 18	Food and kindred products	0.4%	-0.3%	-0.3%	3.3%	-2.1%	-0.2%	-0.0%
Sector 10	Stone, clay, glass, and concrete products	0.5%	-0.1%	1.7%	6.0%	-5.2%	-1.9%	-0.0%
Sector 13	Industrial, commercial mach., computer equip.	15.1%	-0.1%	13.2%	15.0%	-9.6%	-3.4%	-0.1%
Sector 29	Communications	0.9%	-0.2%	0.9%	2.7%	-1.0%	-1.3%	-0.1%
Sector 12	Fabricated metal products, except machinery..	3.1%	-0.1%	6.3%	7.5%	-5.4%	-5.1%	-0.0%
Sector 20	Textile mill products	-5.4%	-0.3%	-0.4%	6.7%	-10.0%	-1.3%	0.0%
Sector 11	Primary metals industry	1.0%	-0.1%	4.9%	12.8%	-11.3%	-5.2%	-0.1%
Sector 21	Apparel and other finished products	-9.1%	-0.5%	3.2%	2.9%	-13.5%	-1.4%	0.1%
Sector 25	Petroleum refining and related industries	-13.1%	-0.3%	-12.2%	6.2%	-5.1%	-1.7%	-0.0%
Sector 4	Coal mining	-6.2%	-0.2%	-11.8%	10.3%	-2.8%	-1.5%	-0.1%
Sector 1	Agriculture, livestock and animal specialties	-0.5%	-0.3%	-5.6%	8.3%	-2.7%	-0.2%	0.1%
Sector 47	Private households	-0.3%	-0.3%	0.0%	0.0%	0.0%	0.0%	0.0%
Sector 19	Tobacco products	5.7%	-0.3%	-1.4%	8.1%	-0.7%	-0.0%	-0.0%
Sector 27	Leather and leather products	-30.6%	-0.9%	10.2%	8.2%	-47.7%	-0.5%	0.0%
Total		-1.1%	-0.2%	0.5%	3.1%	-1.9%	-2.5%	-0.1%

Table 7. Component Impact of GNPDist Factor: 1990–2005

[Ten Sectors Benefitting Most From Projected Export Growth]

			COMPONENT IMPACT					
Sector #	Title	GNPDIST	PCE	Invest.	Exp.	Imp.	Def.	OthGov.
Sector 3	Metal mining.................................	-34.8%	-0.2%	-32.7%	20.4%	-18.1%	-4.1%	-0.1%
Sector 13	Industrial, commercial mach., computer equip.	15.1%	-0.1%	13.2%	15.0%	-9.6%	-3.4%	-0.1%
Sector 14	Electronic and other electrical equipment....	8.0%	-0.2%	12.4%	13.4%	-12.0%	-5.6%	-0.1%
Sector 11	Primary metals industry......................	1.0%	-0.1%	4.9%	12.8%	-11.3%	-5.2%	-0.1%
Sector 24	Chemicals and allied products................	0.2%	-0.2%	-2.4%	11.5%	-6.6%	-2.0%	-0.1%
Sector 4	Coal mining..................................	-6.2%	-0.2%	-11.8%	10.3%	-2.8%	-1.5%	-0.1%
Sector 15	Transportation equipment.....................	18.6%	-0.1%	25.4%	9.9%	-7.3%	-9.2%	-0.1%
Sector 26	Rubber and miscellaneous plastics products...	6.8%	-0.2%	8.0%	9.1%	-7.1%	-3.0%	-0.0%
Sector 5	Oil and gas extraction.......................	-29.0%	-0.4%	-15.7%	8.5%	-19.0%	-2.4%	-0.0%
Sector 1	Agriculture, livestock and animal specialties	-0.5%	-0.3%	-5.6%	8.3%	-2.7%	-0.2%	0.1%

Table 8. Component Impact of GNPDist Factor: 1990–2005

[Ten Sectors Hardest Hit By Projected Import Growth]

			COMPONENT IMPACT					
Sector #	Title	GNPDIST	PCE	Invest.	Exp.	Imp.	Def.	OthGov.
Sector 27	Leather and leather products.................	-30.6%	-0.9%	10.2%	8.2%	-47.7%	-0.5%	0.0%
Sector 5	Oil and gas extraction.......................	-29.0%	-0.4%	-15.7%	8.5%	-19.0%	-2.4%	-0.0%
Sector 3	Metal mining.................................	-34.8%	-0.2%	-32.7%	20.4%	-18.1%	-4.1%	-0.1%
Sector 17	Miscellaneous manufacturing industries.......	-33.1%	-0.4%	-21.6%	4.8%	-15.0%	-0.8%	-0.1%
Sector 21	Apparel and other finished products..........	-9.1%	-0.5%	3.2%	2.9%	-13.5%	-1.4%	0.1%
Sector 14	Electronic and other electrical equipment....	8.0%	-0.2%	12.4%	13.4%	-12.0%	-5.6%	-0.1%
Sector 11	Primary metals industry......................	1.0%	-0.1%	4.9%	12.8%	-11.3%	-5.2%	-0.1%
Sector 20	Textile mill products........................	-5.4%	-0.3%	-0.4%	6.7%	-10.0%	-1.3%	0.0%
Sector 13	Industrial, commercial mach., computer equip.	15.1%	-0.1%	13.2%	15.0%	-9.6%	-3.4%	-0.1%
Sector 15	Transportation equipment.....................	18.6%	-0.1%	25.4%	9.9%	-7.3%	-9.2%	-0.1%

Table 9. **Component Impact of GNPDist Factor: 1977-1990**

[Actual Employment Change (Thousands)]

Sector #	Title	PCE	INVEST.	COMPONENT IMPACT EXPs.	IMPs.	DEF.	OTH. GOV.
Sector 1	Agriculture, livestock and animal specialties	30.4	20.1	288.9	-97.5	2.0	-24.1
Sector 2	Agricultural services, forestry, and fishing.	5.7	-19.2	34.3	-46.0	1.0	-0.3
Sector 3	Metal mining................................	0.7	7.2	17.3	-29.6	1.0	-0.8
Sector 4	Coal mining.................................	1.8	-1.6	27.7	-15.6	1.3	-1.6
Sector 5	Oil and gas extraction......................	5.7	-28.2	28.9	-143.4	3.5	-3.7
Sector 6	Nonmetallic minerals, except fuels..........	0.7	-10.1	11.7	-10.2	0.6	-1.2
Sector 7	Construction................................	9.4	-699.7	41.6	-31.6	17.2	-83.5
Sector 8	Lumber and wood products, except furniture...	3.2	-102.9	48.3	-63.0	3.1	-10.5
Sector 9	Furniture and fixtures......................	3.7	30.0	8.8	-14.0	0.7	-2.3
Sector 10	Stone, clay, glass, and concrete products....	3.6	-51.5	36.1	-39.1	3.2	-7.8
Sector 11	Primary metal industries....................	6.9	21.8	144.1	-184.5	14.7	-9.9
Sector 12	Fabricated metal products, except machinery..	8.5	-19.3	129.9	-93.5	17.7	-13.4
Sector 13	Industrial, commercial mach., computer equip.	6.1	279.7	284.6	-140.6	14.8	-9.7
Sector 14	Electronic and other electrical equipment....	10.1	78.0	167.7	-157.5	19.5	-11.5
Sector 15	Transportation equipment....................	11.3	118.1	198.1	-148.7	39.5	-6.9
Sector 16	Measuring, analyzing, and controlling instr..	3.5	89.3	86.1	-54.4	14.3	-8.0
Sector 17	Miscellaneous manufacturing industries.......	5.5	-10.4	25.3	-50.5	1.0	-2.4
Sector 18	Food and kindred products...................	22.3	-2.1	59.5	-57.1	1.1	-3.7
Sector 19	Tobacco products............................	0.8	-1.9	7.2	-1.0	0.0	0.0
Sector 20	Textile mill products.......................	9.8	-65.2	54.6	-74.1	3.2	-2.5
Sector 21	Apparel and other finished products..........	17.2	-119.1	33.3	-105.2	4.1	-0.9
Sector 22	Paper and allied products...................	6.9	-13.8	52.9	-54.7	2.6	-5.6
Sector 23	Printing, publishing, and allied industries..	11.2	-8.9	48.0	-27.6	5.1	-14.7
Sector 24	Chemicals and allied products................	9.8	-25.6	119.2	-88.2	6.5	-6.8
Sector 25	Petroleum refining and related industries....	2.2	-6.6	10.3	-18.7	1.5	-1.3
Sector 26	Rubber and miscellaneous plastics products...	6.3	-8.0	62.8	-60.2	5.8	-4.6
Sector 27	Leather and leather products................	4.5	-8.3	9.7	-53.6	0.2	-0.1
Sector 28	Transportation..............................	26.0	-20.3	250.1	-96.0	20.7	-19.1
Sector 29	Communications..............................	11.6	17.6	38.4	-16.9	4.5	-8.6
Sector 30	Electric, gas, and sanitary services.........	8.0	-3.2	27.1	-34.4	3.1	-5.3
Sector 31	Wholesale trade.............................	39.7	74.6	359.8	-73.1	21.5	-23.2
Sector 32	Retail trade, exc. eating and drinking places	136.4	-13.7	57.7	-37.0	7.3	-18.3
Sector 33	Eating and drinking places..................	53.3	3.2	66.7	-40.9	6.5	-3.0
Sector 34	Finance, insurance and real estate...........	56.5	-49.0	105.0	-59.2	4.9	-16.8
Sector 35	Hotels and other lodging places.............	14.1	-0.3	27.9	-17.7	5.4	-6.2
Sector 36	Personal services...........................	16.7	-0.3	5.9	-4.6	0.6	-1.5
Sector 37	Business services...........................	22.7	-8.4	118.5	-75.5	17.6	-32.5
Sector 38	Automotive repair, services, and parking.....	8.4	-1.7	15.4	-9.8	1.5	-2.2
Sector 39	Miscellaneous repair services...............	4.5	-3.6	20.1	-13.1	3.3	-2.6
Sector 40	Motion pictures.............................	2.9	-3.6	18.2	-2.9	1.3	-1.2
Sector 41	Amusement and recreation services...........	9.6	0.4	4.2	-2.5	0.7	-0.8
Sector 42	Health services.............................	64.9	-0.2	2.2	-0.9	0.1	-8.0
Sector 43	Legal services..............................	6.3	0.2	15.1	-5.8	0.8	-3.7
Sector 44	Educational services........................	13.8	0.1	3.2	-2.8	1.5	-22.2
Sector 45	Social services.............................	11.9	0.1	1.0	-1.0	0.8	-0.1
Sector 46	All other services..........................	22.1	-89.4	117.0	-37.9	8.1	-21.2
Sector 47	Private households..........................	19.2	0.0	0.0	0.0	0.0	0.0
Sector 48	Federal government enterprises..............	9.0	-2.5	28.4	-18.5	3.1	-5.5
Sector 49	State and local government enterprises.......	7.3	-3.1	24.6	-21.9	2.6	-3.8
Sector 50	Noncomparable imports.......................	0.0	0.0	0.0	0.0	0.0	0.0
Sector 51	Scrap, used and secondhand goods............	0.0	0.0	0.0	0.0	0.0	0.0
Sector 52	Government compensation.....................	0.0	0.0	0.0	0.0	528.5	-592.4
Sector 53	Rest of the world industry..................	0.0	0.0	0.0	0.0	0.0	0.0
Sector 54	Inventory valuation adjustment..............	0.0	0.0	0.0	0.0	0.0	0.0
Sector 55	Commodity credit corporation................	0.0	0.0	0.0	0.0	0.0	0.0
Total..		772.7	-661.3	3343.4	-2432.5	829.6	-1036.0

Table 10. Component Impact of GNPDist Factor: 1990-2005

[Actual Employment Change (Thousands)]

Sector #	Title	PCE	INVEST.	COMPONENT IMPACT EXPs.	IMPs.	DEF.	OTH. GOV.
Sector 1	Agriculture, livestock and animal specialties	-6.1	-123.1	181.0	-58.9	-5.2	1.6
Sector 2	Agricultural services, forestry, and fishing.	-3.1	-26.2	57.0	-53.2	-6.2	1.7
Sector 3	Metal mining................................	-0.1	-19.5	12.2	-10.8	-2.5	0.0
Sector 4	Coal mining.................................	-0.3	-17.6	15.2	-4.2	-2.2	-0.1
Sector 5	Oil and gas extraction......................	-1.5	-65.1	35.4	-79.2	-9.9	-0.2
Sector 6	Nonmetallic minerals, except fuels..........	-0.2	-16.6	7.4	-5.0	-1.7	0.0
Sector 7	Construction................................	-4.3	-429.4	63.7	-36.3	-70.0	1.6
Sector 8	Lumber and wood products, except furniture...	-0.9	-24.5	54.4	-35.6	-10.4	-0.1
Sector 9	Furniture and fixtures......................	-1.0	23.4	11.0	-22.2	-3.1	-0.4
Sector 10	Stone, clay, glass, and concrete products....	-0.8	9.6	34.3	-29.6	-10.6	-0.1
Sector 11	Primary metal industries....................	-1.1	37.1	97.2	-85.5	-39.4	-0.5
Sector 12	Fabricated metal products, except machinery..	-2.0	90.2	108.2	-77.8	-73.2	-0.6
Sector 13	Industrial, commercial mach., computer equip.	-1.1	281.3	319.6	-204.0	-71.6	-1.5
Sector 14	Electronic and other electrical equipment....	-2.8	209.4	225.3	-201.6	-94.1	-1.7
Sector 15	Transportation equipment....................	-2.9	507.1	198.0	-146.7	-183.1	-1.1
Sector 16	Measuring, analyzing, and controlling instr..	-0.9	66.6	76.3	-47.4	-65.8	-1.6
Sector 17	Miscellaneous manufacturing industries.......	-1.7	-94.3	21.0	-65.5	-3.3	-0.3
Sector 18	Food and kindred products...................	-5.3	-5.8	56.1	-34.9	-3.6	-0.2
Sector 19	Tobacco products............................	-0.1	-0.7	4.0	-0.4	0.0	0.0
Sector 20	Textile mill products.......................	-2.3	-2.5	46.6	-70.0	-9.4	0.1
Sector 21	Apparel and other finished products..........	-4.9	34.3	31.3	-144.6	-14.7	0.6
Sector 22	Paper and allied products...................	-1.7	-4.5	54.3	-40.3	-10.3	-0.5
Sector 23	Printing, publishing, and allied industries..	-3.9	32.0	62.1	-30.6	-27.9	-2.9
Sector 24	Chemicals and allied products...............	-2.5	-26.3	125.8	-72.3	-21.4	-0.6
Sector 25	Petroleum refining and related industries....	-0.4	-19.4	9.8	-8.0	-2.7	0.0
Sector 26	Rubber and miscellaneous plastics products...	-1.9	71.1	81.5	-63.1	-26.7	-0.4
Sector 27	Leather and leather products................	-1.2	13.9	11.2	-65.0	-0.6	0.1
Sector 28	Transportation..............................	-7.7	-8.2	277.8	-60.0	-85.3	-2.3
Sector 29	Communications..............................	-3.3	11.5	35.6	-13.8	-17.4	-1.3
Sector 30	Electric, gas, and sanitary services.........	-2.6	1.0	28.6	-23.4	-14.2	-0.7
Sector 31	Wholesale trade.............................	-11.7	70.0	396.6	-66.8	-98.9	-2.3
Sector 32	Retail trade, exc. eating and drinking places	-42.7	19.4	93.9	-48.5	-53.6	-2.8
Sector 33	Eating and drinking places..................	-21.0	35.4	97.0	-51.1	-40.7	1.9
Sector 34	Finance, insurance and real estate...........	-21.2	-26.9	158.1	-60.6	-26.5	0.2
Sector 35	Hotels and other lodging places.............	-5.6	18.1	35.8	-21.4	-28.5	1.4
Sector 36	Personal services...........................	-5.4	2.1	6.0	-3.9	-2.3	-0.3
Sector 37	Business services...........................	-15.5	65.5	299.0	-157.5	-199.6	-16.2
Sector 38	Automotive repair, services, and parking.....	-3.5	6.0	26.8	-13.1	-11.2	-0.1
Sector 39	Miscellaneous repair services...............	-1.5	0.9	27.1	-13.6	-15.9	-0.3
Sector 40	Motion pictures.............................	-1.4	-4.6	27.7	-2.9	-7.4	-0.4
Sector 41	Amusement and recreation services............	-3.7	1.6	5.7	-2.8	-4.0	-0.3
Sector 42	Health services.............................	-26.5	-1.0	7.2	-0.7	-0.3	-3.6
Sector 43	Legal services..............................	-3.0	6.5	25.2	-9.8	-6.9	-0.5
Sector 44	Educational services........................	-4.9	1.8	52.8	-5.8	-6.1	-6.0
Sector 45	Social services.............................	-6.7	1.0	2.4	-2.5	-4.6	0.0
Sector 46	All other services..........................	-8.2	-48.0	109.2	-49.9	-49.2	-2.8
Sector 47	Private households..........................	-3.4	0.0	0.0	0.0	0.0	0.0
Sector 48	Federal government enterprises..............	-2.7	6.6	28.2	-16.7	-15.2	-0.5
Sector 49	State and local government enterprises.......	-2.5	0.3	25.7	-14.6	-11.1	-0.4
Sector 50	Noncomparable imports.......................	0.0	0.0	0.0	0.0	0.0	0.0
Sector 51	Scrap, used and secondhand goods............	0.0	0.0	0.0	0.0	0.0	0.0
Sector 52	Government compensation.....................	0.0	0.0	0.0	0.0	-1555.9	-117.7
Sector 53	Rest of the world industry..................	0.0	0.0	0.0	0.0	0.0	0.0
Sector 54	Inventory valuation adjustment..............	0.0	0.0	0.0	0.0	0.0	0.0
Sector 55	Commodity credit corporation................	0.0	0.0	0.0	0.0	0.0	0.0
Total..		-259.7	659.5	3766.3	-2332.1	-3024.4	-162.1

Reconciling Conflicting Data on Jobs For College Graduates

Daniel Hecker, Office of Employment Projections, U.S. Department of Labor

The work that I'm going to discuss today, along with Tina Shelley's, has been published in the July 1992 Monthly Labor Review and the Summer 1992 Occupational Outlook Quarterly. The conflicting data are for the 1980's. The article is an evaluation and a defense of projections done in 1979. It was prompted by a very critical evaluation published by John Bishop, a respected Cornell University labor economist and his graduate assistant, Shani Carter. The conflict cited in the title is: rising relative wages, Bishop and others contend, suggest a shortage of college graduates. However, BLS projected and later reported that one-fourth of graduates entered jobs that traditionally didn't require a degree, clear evidence of a surplus.

Background

The Bureau has been projecting the outlook for college grads as a group every two years since the early 1970's. This work supplements the outlook for individual occupations in the Occupational Outlook Handbook. While it is possible to generally discuss the supply-demand situation for engineering, nursing, teaching and similar occupations where there is one major field of study that provides most entrants and most with that major seek entry. However for many occupations generally requiring a degree, a wide range of majors may be acceptable, so a supply analysis is not possible. Similarly, many graduates do not seek work in an occupation directly related to their major--and for liberal arts grads, as well as general business and a few other majors, there are few occupations directly related. So no demand analysis is possible for these majors. Their job prospects depend largely on the outlook for graduates in general.

In the Winter 1979 Occupational Outlook Quarterly we estimated that a fourth of all graduates who entered the labor force during the 1970's had entered a job not generally requiring a degree or was unemployed--based on responses to the Current Population Survey (CPS) showing years of school and occupation. We had counted in the quarter those who reported 4 years or more of college and work in a retail sales, clerical, service or blue collar job--jobs for which employers didn't usually seek graduates nor ones that called on skills learned in college. (See Chart 1.) This was a radical change from the 1960's. And we projected that this surplus of grads would continue during the 1980's--based on NCES projections of bachelor's degrees compared to job openings from growth in managerial, professional, technician, and non retail sales jobs--those generally requiring a degree, plus the number of graduates expected to leave the labor force.

Our analysis of 1988 data indicated we had been correct--a quarter of graduates were in fact in clerical and similar jobs. However, in 1991 we were informed by John Bishop, not only that our overall occupational projections for the 80's were glaringly in error, but that our college graduates projections were too. Bishop and Carter later published an article in the Fall 1991 Educational Evaluation and Policy Research stating "..the (1980) BLS effort to project the supply-demand for college graduates has been a failure. Such a judgment is possible because changes in the ratio of young college graduates wages to young high school graduates wages provide an ex post criterion. In fact the increased demand... substantially outstripped increased supply... the ...1980's were clearly a period of a growing shortage of college graduates. The BLS approach ...is invalid because of the unreliability of CPS coding of occupations..". They also forecast a worsening shortage and recommended government policies to increase the supply. (See Chart 2, with data we prepared.) The decline during the 1970's, everyone agrees, represents a surplus; the upturn they claim proves a shortage.

Academics may regularly attack each other's work, but our projections are rarely criticized. They either are treated with great respect, or taken with a grain of salt. However, Bishop's challenge, and the problem he raised, clearly made this project more exciting than a typical end-of-decade evaluation. For the same issue of the Journal, we had prepare a hasty response, but now needed to prepare a more thorough one, and also convince ourselves that our method was valid, before we did further projections.

The Relative Earnings Problem

Bishop's case for shortages (and against surpluses) of college graduates rests entirely on their relative earnings increase. He says the reason we were wrong is that the CPS, upon which our analysis rested, misclassifies many respondents' occupations. Furthermore our results are meaningless because it is impossible to define and identify what jobs require a college degree and which don't--for example, most managerial jobs require a degree, but some don't and most clerical don't but some do, and that requirements vary from employer to employer. Clearly our major concern was to evaluate Bishop's relative earnings analysis. In our preliminary response, we had argued that either the base against which college graduates earnings were measured--the earnings of high school graduates--had fallen and driven up the relative ages of college graduates, or that the market for college graduates is segmented, with shortages and bidding up of wages of engineers, computer scientists, and other high demand occupations, while at the same time, there were not enough jobs for liberal arts, communications and similar graduates, who accounted for the surplus.

Now, it is an axiom among economists that relative wage increases prove a shortage. This is an extension of the proposition in Econ. 101 that the price for any item increases only when the demand exceeds supply and buyers are forced to bid up prices (in this case wages) to get the quantities they want. Since all wages increase over time, the operant figure is wages in the population under study relative to wages in other groups. So while we believed our CPS-based analysis, we weren't sure we could discredit the criticism.

As a first step, we examined the economic literature and found that a number of researchers Bishop had cited, and others, including the Council of Economic Advisors had noted the rising relative wages of college graduates, and they concluded that demand had grown faster than supply--that is close to saying shortage, but no one besides Bishop had actually used the word. Others noted that college was an increasingly good investment. Clearly there was substantial support for Bishop's position.

We also found several articles on the sharp real wage declines for male high school graduates--due to economic restructuring--which was encouraging since it suggested high school graduates might not be an appropriate base for determining relative wage changes. Also, the earliest work on relative wage analysis had compared engineers wage increases both to those of all other workers (about 99 percent of the total) and to selected other professional occupations. Bishop and others had used only one base, and that made up only 40 percent of the total, rather than 99 percent--a possible shortcoming. And I recalled a small book published 20 years ago by 3 respected labor economists saying that a perennial problem with relative earnings work is finding a suitable (i.e. stable) base against which to make the comparison--one more useful piece.

We also noted NCES surveys of recent college graduates showing similar proportions of graduates in clerical, service and blue collar jobs, and College Placement Council data suggested only modest on-campus recruiting by employers--behavior clearly inconsistent with shortages. Next, we disaggregated the earnings data as much as possible. Bishop and other researchers had shown just the relative earnings ratios for men and for women. We first tabulated CPS data to show earnings growth for individual occupations, by level of education and sex, 1983-90 (1983 was the earliest year with consistent definitions) to see if all of the relative increase in earning of college graduates was taking place in only some (shortage) occupations, with others not increasing or even declining.

We found that in fact, earnings in some occupations--teachers, nurses, college professors, and lawyers grew faster than average for college graduates and much faster than the average for high school graduates. So did a few others filled by low-demand general business or liberal arts graduates. Except for nurses, we had no anecdotal or survey evidence showing shortages in these occupations. However, earnings increases for engineers, an occupation Bishop had cited as being critically scarce, were lower than for the occupations just cited. There was no logical pattern in the increases and little evidence for a shortages in some segments of the market and surpluses in others.

Next, we examined whether the base was declining and what that meant. Income data from the Census P-60 series back to the mid 1960's confirmed that the relative earning of graduates had increased sharply after 1979. We calculated a simple percent increase in income by sex and years of school for less than 4 years of high school, 4 high school, and 4 or more of college (see Chart 3). Women college graduates had the largest earnings increase. College graduate men were next, closely followed by high school graduate women, and high school dropout women (who were still one point above the average for all workers). Males with years of high school and then male dropouts were far lower. Male high school graduate earnings, as noted in earlier research, the base against which Bishop measured male college graduates' earnings, were clearly out of line.

The very fast growth for women college graduates' earnings, and the rise in their earnings relative to high school graduate women, we hypothesized, was not due to a shortage, but to a growing proportion of women college graduates in high wage occupations such as lawyer, physician and business manager, to salary adjustments intended to bring their earnings in line with male coworkers, and sharp earning increases in female-dominated nursing and school teaching. (Despite bigger increases, women still averaged much less than men.)

Seeking a way to show the weakness in Bishop's use of a declining base, we asked, why evaluate only male college graduates against male high school graduates? Why not female high school graduates. Obviously wages of these women rose relative to their male counterparts--proof, using Bishop's relative wage logic, of a shortage of women high school graduates. This seems unlikely. How about another inter-gender comparison--male college graduates against female high school graduates? Presto, no shortage of male college graduates! Encouraged, we asked, why are male high school graduates the appropriate standard, why not male high school dropouts? So we "proved" that every other category had a shortage, including male high school graduates--since their relative wages rose. This is an obviously absurd conclusion. We dramatically showed the problem with Bishop's declining base, a problem noted 20 years ago. We concluded that over this period, there was no stable base against which to compare college graduates earnings, that relative earnings analysis can not be used in this situation, and that our finding of surpluses stood. And we supported our position with NCES and on campus recruiting data. Earnings of male high school graduates.

Next, we looked for reasons earnings of male high school graduates grew so slowly. BLS 790 survey data showed that manufacturing employment peaked in 1979, mining in 1981; and their production worker employment dropped from 18 to 12 percent of all jobs from 1979 to 1990. Other high wage and predominantly male industries such as railroads and telephone communication also declined since 1979. This suggested that average wages of high school graduate men grew slowly because many jobs at the upper end of the wage distribution disappeared. Also, CPS data showed earnings of those with at least some college increased faster than those with only 4 years of high school in virtually every occupation. Either college-educated workers got bigger periodic wage increases, or more likely, college-educated workers took a growing proportion of the high wage jobs. As top-paid high school graduates left their jobs, they were replaced by those with at least some college. High school graduates were increasingly limited to the lower paid jobs, another reason for their average to rise slowly.

Interestingly, in this situation, the surplus of college graduates had directly raised their wages relative to high school graduates by making it more difficult for high school graduates to move into high paid jobs. This is exactly the opposite effect expected from a surplus. Since few high school graduate women had held high wage manufacturing and mining or white collar jobs, these phenomenon did not much affect their wages, which grew at just about the average.

The CPS Data

Also, based on Bishop's criticism that lots of jobs could be classified either way, we tightened the definition of jobs not requiring a degree--deleting more sales jobs and putting them in the "requires a degree" category; this cut the estimated surplus from 1\4 to 1\5. I then created a 23 year time series based on this definition and noted the sharp, smooth rise in the proportion underemployed from 10% in 1969 to 20% in 1980 (consistent with 1969-78 trends from the 1979 report), then a steady 20% thereafter. This made a strong case that classification errors hadn't caused the data to incorrectly show a 20% surplus. CPS earnings averages for college graduates reported in clerical and retail sales jobs were so low, equivalent roughly to GS-3, that few could possibly have been misclassified college level jobs, as Bishop charged.

What we learned by disaggregating the CPS earnings data was also noteworthy. It had been used for many years to show that college pays, because average earnings rise sharply with education. "On average "has been in fine print or not even mentioned. While college graduates average much more than high school graduates, there is substantial overlap. (See table 4.) A third of those over age 25 working full time with 4 years of college earned less than $500 a week and a fifth of those with at least 5 years of college did also. On the other hand, a third of high school graduates and half of those with 1-3 years of college earned more than$500 a week. Some of the spread can be explained by years of experience, gender, and geographic location, but even controlling for these, the differences would be substantial. In the OOQ we explored these differences more fully, showing the wide spread in median weekly earnings by occupation and level of education. Also, we presented data from a 1987 Census report on earnings by degree level, including associate, masters, doctorate, and professional degrees and by major field of study.

Publicity

I'd like to finish by telling you about our unusual press coverage, since Federal forecasts don't usually get a lot of publicity--nor do forecasters. My work, along with Tina Shelley's, had a 2 column write-up in an August issue of the New York Times. This piece was the subject of Bob Samuelson's op-ed column in the Washington Post and in the August 31 Newsweek--following the cover article on Woody and Soon-Yi, on the overleaf from a picture of Fergie topless. Samuelson's slant on the controversy was more than I could have hoped for. Bear with me while I crow a little. He said the expected benefits of college haven't fully materialized and "Now comes economist Daniel Hecker of the Labor Department to help explain why. In a new study he convincingly demolishes the notion that there's a scarcity of college graduates... What had persuaded many economists to think otherwise is the hefty wage premiums enjoyed by college graduates over high school graduates." There's more ...but I won't go on.

Chart 1.
Jobs Entered by College Graduates, by Major Occupational Group
1962-69 and 1969-78

Occupation Group	1962-69	1969-78
Operatives, laborers, service farm and unemployed	1.0%	11.6%
Craft workers	2.6%	3.1%
Clerical workers	2.9%	10.1%
Sales workers	3.0%	9.0%
Managers and administrators, except farm	17.2%	20.3%
Professional and technical workers	73.2%	45.9%

Source: Bureau of Labor Statistics

Table 1
Usual Weekly Earnings of Full-time Wage and Salary Workers,
25 Years of Age and Older, by Level of Education, 1990

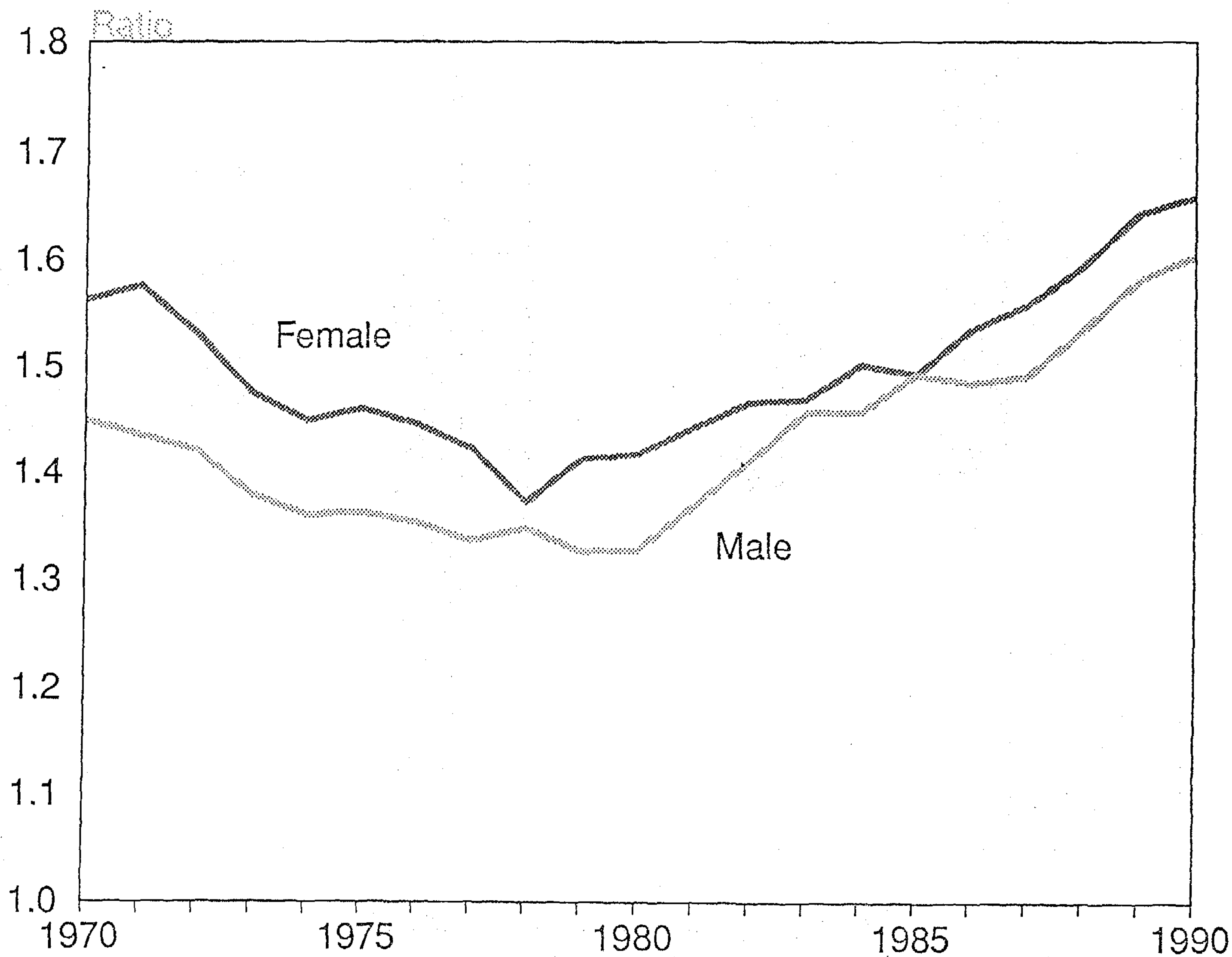
Years of school completed	Percent with Usual Weekly Earnings					Median weekly earnings
	Total	Under $250	$250-$499	$500-$999	More than $1,000	
4 years of high school	100	18	51	29	3	$386
1-3 years of college	100	9	44	40	6	$476
4 years of college	100	4	30	51	16	$595
5 or more years of college	100	3	17	52	28	$726

Note: Percents may not add to 100 due to rounding.

Source: Current Population Survey.

Chart 2.

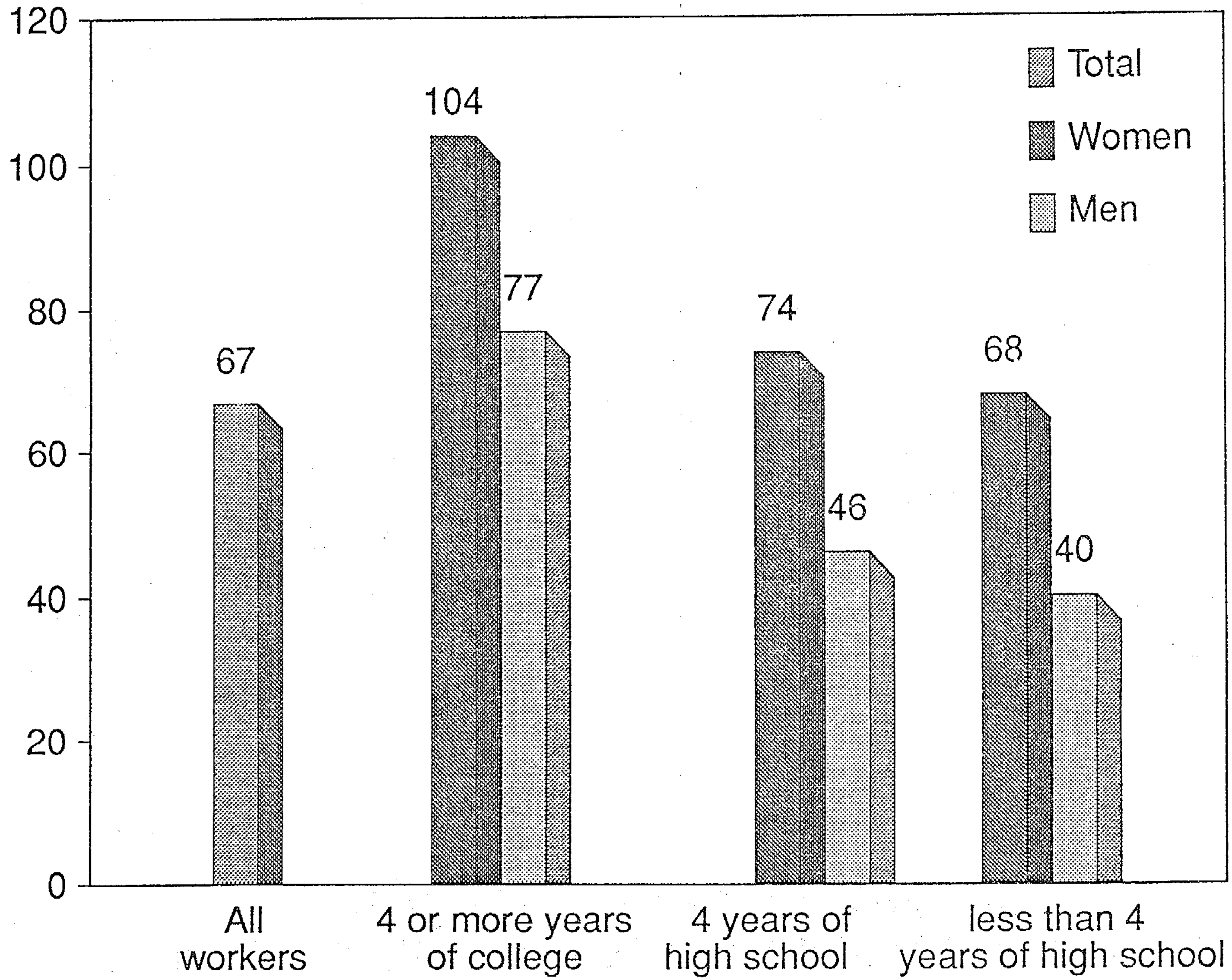
Ratio of earnings of college graduates compared with high school graduates, 1970-90



SOURCE: Bureau of Labor Statistics

Chart 3.

Percent change in earnings of men and women by years of school completed, 1979-90[1]



SOURCE: Bureau of the Census and Bureau of Labor Statistics

[1]Median annual income, full time workers 25 years and older

A Review of the Methodology for Projecting Supply and Demand for College Graduates

Kristina Shelley, Office of Employment Projections, Bureau of Labor Statistics

The July issue of the Monthly Labor Review contains the results of the Bureau of Labor Statistics most recent supply and demand analysis for college graduates. To very briefly summarize, the analysis shows that about 30% of the supply of college graduates entering the labor force between 1990 and 2005 will end up in jobs which traditionally do not require someone with a 4-year college degree--or in other words, will be underemployed. This indicates a more competitive environment for college graduates in the future job market than that faced by college graduates during the 1980's, when only 20% of the supply were underemployed.

This is not the first time the Bureau of Labor Statistics has published the outlook for college graduates nor is this the first time BLS has projected a "surplus" of college graduates. Following the publication of the 1988-2000 outlook for college graduates, some debate arose regarding BLS's projections of underemployment. To be more specific, in 1990, John Bishop, an economics professor from Cornell University, charged that the BLS occupational employment projections and the college graduate analyses were flawed, and as a result, the demand for college graduates throughout the 1980's was consistently underprojected. Bishop further contended that the 1980's were actually a period of shortages of college graduates, evidenced by rising wages of college graduates relative to those of less educated workers. And finally, Bishop suggested that the shortages will continue throughout the 1990's as the traditional college-age population will number even less than during the 1980's.

Although some of Bishop's criticisms appeared invalid at first glance, BLS decided it was time to reexamine closely its analytical procedures, and if necessary, make changes in the methodologies employed. Consequently, Neal Rosenthal, chief of the Division of Occupation Outlook in the Office of Employment Projections, evaluated the 1978-1990 employment projections, and at the same time addressed the difficulties associated with trying to compare the projections for that period with the actual data. Dan Hecker, also of the Office of Employment Projections, examined the issue of relative wages, and found an alternative explanation for what rising relative wages for college graduates mean. I was given the task of preparing the college graduate supply and demand analysis for the new projection period, focusing on making any necessary adjustments in the analysis and presentation to account for the recent criticisms. This paper will outline the analytical procedure used to develop the outlook for college graduates, note some of the changes that were made in an attempt to improve the accuracy of the forecast, and also note some of the limitations of the analysis.

The entire procedure involves three major steps: first, demand for college graduates is projected; second, the supply of college graduates is projected; and third, the actual supply and demand conditions of a recent period of time (in this case, 1984-1990) are estimated, in order to compare the actual conditions with the projected conditions.

Demand

There are three sources contributing to demand for college graduates: occupational employment growth, educational upgrading of jobs which formerly did not require someone with a college degree, and replacement needs.

The first step of the demand process is to determine the number of new job openings due to employment growth. Base year requirements for college graduates must be calculated, then requirements in the target year projected; the difference between the estimated requirements for college graduates in 1990 and the projected requirements for college graduates in 2005 comprises openings due to employment growth. At the heart of projecting employment growth of jobs for college graduates are the BLS occupational employment projections, which are derived from a series of demographic, economic models and the industry-occupational requirements matrix. The matrix is based on establishment data from the Current Employment Statistics program and on staffing patterns of industries from the Occupational Employment Statistics survey. However, projecting opportunities for college graduates specifically requires the use of educational attainment data, which are only collected through the Current Population Survey, which is a household survey. In order to account for differences in occupational structure, the two data sets must be matched via a crosswalk, which converts the CPS data on occupations into a matrix occupational classification basis.

Because employment levels of the two data sources also differ, with matrix employment levels lower in many of the occupations employing the most college graduates, the traditional process of adjusting the CPS educational attainment data to a matrix basis results in fewer college graduates than are reported in the CPS. Therefore, to avoid possible undercounting of college educated workers, the actual CPS numbers were used as the most accurate portrayal of the employment of workers having four or more years of college education in 1990. These numbers were then used to calculate college graduates as a proportion of matrix employment in each occupation.

Having incorporated the CPS data on college graduates into the matrix, the number of college graduates holding jobs in 1990 were easily counted. However, the purpose of the BLS analysis is to project demand for college graduates in jobs actually requiring someone of that educational level, which means jobs must be classified into those requiring a college degree and those not requiring a college degree. This was done by means of requirement ratios, which are the number of workers in each occupation in jobs requiring a degree as a proportion of all workers in the occupation having a degree. They were developed based on the Bureau's in-depth knowledge of occupational educational requirements and on a supplemental CPS survey in which workers in all occupations indicated whether or not their jobs needed a college degree.

To simplify the process, but also to ensure that no college level jobs were left out of the analysis, requirement ratios were set at 100% for all occupations in the executive, administrative, and managerial; professional specialty; and technician occupations, as well as these six marketing and sales occupations: insurance sales workers; real estate agents, brokers, and appraisers; securities and financial services sales workers; advertising sales agents; sales agents within selected business services; and sales representatives of scientific products and services. In these cases, all employed college graduates were assumed to be in jobs actually requiring a college graduate. Likewise, requirement ratios for a number of occupations were set at 0; that is, of the jobs held by college graduates in these occupations, none were counted as jobs really requiring someone with a college degree. These types of occupations include most non-supervisory laborer and production occupations, private household workers, and some others. The remaining occupations, such as bookkeepers, secretaries, and farm operators, comprise a "grey" area, where judgments regarding the necessity of a college degree are too subjective; in these cases the requirement ratios were set based on the response in the previously mentioned CPS survey. Having determined the number of college graduates in the matrix by occupation, and applying requirement ratios to those numbers, it is possible to count the number of college level jobs in 1990.

To project future requirements for college graduates, it is first necessary to project changes in those proportions of jobs in each occupation requiring someone with a college degree, since increases in jobs for college graduates stem from both employment growth in occupations generally requiring a degree and from upgrading of jobs. Presumably, educational upgrading occurs when changes in technology, business practices, or other factors render the skills required to perform jobs more complex, thereby requiring workers with more education. Although the proportions of college level jobs may also increase due to more widespread availability of college graduates seeking jobs, it is not possible to know when this is the case, or whether there was a real increase in skill level. Therefore, all increases in proportions were counted as educational upgrading in those occupations for which analytical evidence shows that college training is or may be needed.

In earlier analyses, projecting the changes in proportions of college level jobs in each occupation was done by using the analyst's judgement as to how quickly jobs will be upgraded in the future. Examining historical CPS data on college graduate employment as a percent of employment of all workers in each occupation and extrapolating the trends to 2005 instead eliminates the more subjective aspects of using the analyst's judgement. (The trends were based on data for the 1983-1990 period in which the CPS incorporated the Standard Occupational Classification used in the 1980 census. Prior to 1983, the CPS used the 1970 census classification.)

To relate the CPS trends to matrix employment, the CPS-based change between 1983 and 1990 in the percent of jobs in occupations that required a college degree in 1990 were converted to average annual rates of change, which were multiplied by 15 to determine projected rates of change for each occupation. Applying the projected rates of change to the proportion of jobs in an occupation requiring college graduates in 1990 yielded the projected proportion of jobs in an occupation requiring college graduates in 2005. These proportions were then multiplied by the projected employment of all workers for each occupation and summed to derive total requirements for college graduates in 2005.

The difference between 1990 requirements and 2005 projected requirements constitutes the number of new job openings for college graduates during the projected period due to both employment growth and educational upgrading. To measure the number of job openings due to educational upgrading alone, total employment in the target year was multiplied by the percent of college level jobs in the base year--thereby producing the educational requirements that would have been, if the proportions of college level jobs had not increased in the interim. The difference is upgrading.

The final piece contributing to college level job demand is replacement needs, which are a measure of additional job openings resulting as workers leave existing college level jobs. A critical difference between the most recent analysis and earlier analyses concerns the use of a net separation rate to measure replacement openings rather than a total, or gross rate. The projected net separation rate, applied to 1990 requirements for college graduates, indicates the approximate number of college graduates who leave jobs in which they are employed, and who are not expected to ever return--in other words, permanent separations from college level jobs. This rate was recently developed by Alan Eck, of the Office of Employment Projections (see the technical note in 'The Future of Jobs for College Graduates' in the July Monthly Labor Review), and is considerably lower than the total rate used previously, which included all separations from the labor force--temporary as well as permanent. The impact of using net rather than total separation rates was a huge one, which can be seen by comparing the 1988-2000 replacement needs with the recently revised replacement needs. In 1988-2000, average annual replacement openings were estimated at 1,075,000 compared with the new estimate of about 300,000.
(The net rate used in the recent analysis is not a rate for all college graduates, but one representing separations of

graduates from college level jobs only. It does not count separations of underemployed college graduates because the point of this analysis is to project college level job openings, not job openings for workers with any amount of education.)

Supply

The procedure for projecting the supply of college graduates is more straightforward than projecting demand, although not entirely without complications. New graduates have been, and will continue to be, the major source of entrants to the college level job market. The BLS analysis relies on the National Center for Education Statistics' middle projections of the number of bachelor's degrees conferred to do this. (The number of bachelor's degrees awarded is used as a proxy for the number of new entrants qualified to seek jobs requiring at least a bachelor's degree because it is assumed that anyone receiving a masters or Phd. degree has already obtained the bachelor's and should not be counted a second time.) Since all new bachelor's degree recipients are not expected to enter the labor force during the fifteen year period, they should not all automatically be counted in the supply and demand analysis. Calculating the number expected to enter the labor force was done by applying a labor force entrance rate of 97.2% to all new graduates. This rate was derived from an April 1985 NCES survey on the labor force status of 1983-84 bachelor's degree recipients, 88% of whom reported being in the labor force one year after obtaining a college degree. An additional 9.2% were in graduate school, presumably to enter the labor force at some point during the projected period.

Although new college graduates make up the major part of the supply in the analysis, there is a second source of supply that must be considered--that of other entrants such as immigrants, recently discharged military personnel, or persons recently released from institutions. Other entrants were assumed to average around 200,000 per year, the same as during the 1984-1990 period. Other entrants were estimated by combining the labor force increase each year ("growth entrants") between 1984 and 1990 with annual labor force separations ("replacement entrants") for a total number of labor force entrants each year. Subtracting yearly new graduate entrants yielded the average annual number of other entrants. The separation rate used to determine the number of "replacement entrants" measures labor force separations of all workers with four or more years of college, not just separations of college educated workers in college level jobs, as was the case in the demand portion of the analysis.
Again, the separation rate used in the 1990-2005 analysis was a net rate, as opposed to a total rate, resulting in far fewer other entrants than the 1988-2000 analysis projected. Consequently, excess entrants as a percent of supply increased, leading to large projections of underemployment.

The third major step in the process of analyzing the outlook for college graduates was to estimate the actual supply and demand conditions for the 1984-1990 period for comparison purposes. This allows future job market expectations to be placed in perspective with conditions that existed in the recent past. (In developing a historical trend, 1984 was used as the base year because occupational employment data are classified consistently between 1984 and 1990. Prior to 1984, the industry-occupation matrix used in the Bureau's projection program was based on a different occupational classification system.)

The described procedure which has, for the most part, been used to project supply of and demand for college graduates for many years, has also been periodically reviewed and revised over the years in an effort to improve the forecast. The most recent reevaluation resulted in additional revisions which, it is believed, reduce the likelihood of mistakes in the projection. Nevertheless, there are some limitations and/or weaknesses which cannot be entirely eliminated from the analysis.

As with any forecast, assumptions are made regarding future conditions which may change unexpectedly. For example, dramatically faster increases in the use of sophisticated technology or certain business practices could result in a much faster rate of educational upgrading than was projected by extrapolating the historic rates of change.

In addition, the accuracy of the projections for college graduates are largely dependent on the accuracy of the middle economic growth projections and the middle bachelor's degree projections that I used. Therefore, the traditional presentation of the analysis' conclusions was altered somewhat to account for the possible variances in the rates of economic growth, bachelor's degree growth, and educational upgrading growth. Combining various alternative BLS and NCES projections enabled BLS to illustrate the worst-case and best-case scenarios for college graduates in the July Monthly Labor Review. Even the best-case scenario, though, indicated conditions similar to those of the 1980's, when 20% of college graduate entrants were underemployed, while the worst case showed underemployment at 45% of total entrants to the supply.

Aside from the possibility of errors in the employment growth or the bachelor's degree projections, there remain some concerns about the reliability of the CPS occupational coding system. Mismatches between education and occupation might affect BLS's ability to classify jobs into those requiring a degree and those not requiring a degree accurately. Although errors in reporting occupation do occur, mismatches between high and low skill occupations do not necessarily invalidate the analyses. These mismatches are equally likely to occur in both directions and with the same frequency, thus roughly balancing out.

There are also some difficulties associated with measuring educational attainment. Until 1992, CPS respondents who said they had completed four or more years of college were counted to have a college degree, since the CPS did not collect data on whether or not an individual had a degree. Consequently, the number counted as college graduates in the analysis includes some who actually are not college graduates. This problem should be rectified during the next analysis. The Current Population Survey now (1992) asks respondents if they have actually received a college degree, which should improve the accuracy of the educational attainment data.

The final limitation of the analysis is the general nature of the projection. Opportunities are projected for graduates as a whole, not just those with certain skills or backgrounds. It is not possible to determine the level of academic preparedness (or lack thereof) among what BLS calls the underemployed college graduates. While it is possible that some of the underemployed are actually undereducated, that is a completely different issue and one that BLS did not address.

In summary, a substantial effort was made during the past year to look critically at all aspects of the supply and demand analysis for college graduates. Where changes were warranted, they were made. There may always be questions regarding the credibility of the data used, but such is the case for many forecasters and data users, and should not invalidate the broad conclusions of the college graduate analysis of supply and demand.

How Large Are Economic Forecast Errors?

Stephen K. McNees, Federal Reserve Bank of Boston

Virtually everyone follows some forecaster's views, analyzing each pronouncement and eagerly awaiting the next. Opinion about the reliability of economic forecasts ranges widely, however--some argue that they are literally worthless, even though most forecasters typically can point to a sequence of predictions that virtually replicate the eventual outcome. How much confidence should one place in economic forecasts?

The answer would seem straightforward: To measure a forecast's reliability one need simply compare it with what "actually" occurs. The diversity of opinion on reliability indicates the answer is not so simple. Two problems arise immediately, one philosophical and one practical. The philosophical problem is one of induction: Forecast accuracy cannot be measured until what actually happened is known, but the main interest typically lies in the accuracy of current forecasts for which, necessarily, no actual outcome is available. Despite many attempts to make headway with this problem, some form of assumption must be made that the future will resemble the present. Neither logic nor econometrics can provide assurance that this assumption will hold. In fact, the future is almost certain to differ at least somewhat from previous experience. Nevertheless, no alternative exists to blithely assuming that the reliability of today's forecasts will resemble the reliability of previous forecasts--that some forecaster (or model) has captured the essential lasting features of past and future behavior.

The practical problem in measuring the accuracy of past forecasts is that so many different forecasts are available--and, in some cases, so many different measures of what actually happened--that millions of different errors can be calculated, and this varied experience can be summarized in many different ways. The problem, in other words, is not the paucity of measures of reliability but their multiplicity or, more precisely, their variety. The errors vary with many factors, including (1) the economic series or variable predicted, (2) the forecaster, (3) the time period being forecast, (4) the horizon of the forecast, and (5) the choice of "actual" data to measure what really happened.

Much attention focuses on the first three factors--the economic variable, the forecaster, and the forecast period. To illustrate the importance of the fourth and fifth factors, consider the accuracy of one prominent forecaster's predictions over the last 10 years of real GNP growth in the current quarter. The top panel of Table 1 describes the accuracy of the predictions as measured against the first official estimate of real growth ("preliminary actual data," or "advance" actual data); the bottom panel, the accuracy of the predictions when measured against the final revised estimate of real growth (prior to the benchmark revision). The first column shows the accuracy of forecasts made late in the first month of each quarter, just after the preliminary estimate of the prior quarter became available; these are called "early-quarter" forecasts. The second column shows "mid-quarter" forecasts, those made in the middle month of each quarter. The final column shows the errors of the forecasts made in the last month of the quarter, or "late-quarter" forecasts. These forecasts are customarily the expectations against which the press and financial market participants judge the preliminary GNP data release.

The table documents two obvious points: (1) The forecasts are much more accurate predictions of the preliminary data, which are based largely on information also available to the forecaster, than they are of the final revised data, which are based on information that does not become available until much later. (2) Forecasts made later in the quarter, when the forecaster has more information, are more accurate than earlier forecasts. Note, however, that the improvement in forecast accuracy is much greater compared to the preliminary than to the revised actual data. For example, 10 percent of the forecasts of real growth made in the first month of the quarter were off the mark by more than 3 percentage points, while none of the forecasts made during the last month of the quarter missed the preliminary estimate by more than 2.1 percentage points. The elimination of the large outliers, through the incorporation of incoming high-frequency data, cuts the root mean square error (RMSE) in half between the first and third months. In contrast, relative to the revised actuals, the proportion of errors exceeding 3 percentage points falls only from 22 percent of the forecasts made in the first month of the quarter to 15 percent of the forecasts made in the last month of the quarter; the proportion of forecast errors exceeding 1 percentage point was actually somewhat larger for the forecasts made in the last month of the quarter. The RMSE falls only by about 20 percent over the quarter. Thus, while the incoming high-frequency data shed a lot of light on what the preliminary estimate of real GNP will be, they provide relatively little new information on what the final revised number will be.

Table 2 presents comparable information for forecasts of the current-quarter rate of growth of the consumer price index (CPI). Note first that little difference can be seen in the accuracy of the predictions whether compared to the preliminary or the revised data. Unlike real GNP, where additional information is collected to improve the estimates, the CPI is based on a survey conducted each month, which cannot be repeated; all revisions come solely from changing the seasonal adjustment factors. Note also that the timing of the forecast is even more important for the CPI than for real GNP; this reflects the fact that CPI data are collected and released monthly so that by the time the late-quarter forecast is made, forecasters know the actual outcome for two of the three months of the quarter.

Forecasters have often been accused of bias. However, none of these forecasts shows a systematic tendency to either overestimate or underestimate the actual outcome. The mean errors are essentially zero, whatever the forecast horizon and whichever actual data are used.

Should forecast accuracy be assessed relative to the preliminary or to the revised actuals? The answer depends entirely on the purposes of the forecast. If the objective is to understand what influences behavior at the time--for example, if one is interested in the reaction of investors in financial markets--the preliminary data are the obvious choice, as the revised data are not available until much later. However, if the objective is to measure how close the forecast comes to what actually occurred--what nonfinancial decisionmakers, modelbuilders, and policymakers presumably would want to know--it is equally clear that the revised data, based on the most complete information set, provide a better estimate of reality.

This is particularly true for comparative evaluations: if forecaster A provides the most accurate predictions of what was initially thought to have happened (preliminary data), but forecaster B provides the best forecasts of what turns out to have actually occurred, once all the facts are in, it would seem odd to call A the better forecaster of the economy, even though forecaster A clearly is a superior forecaster of the social accountants who produce GNP estimates. Fortunately, the distinction between preliminary and revised data becomes less important for forecasts of longer time spans, such as one-year-ahead forecasts, and for variables other than the National Income and Product Accounts and the monetary aggregates, such as the CPI and the unemployment rate. For example, prices in financial markets (stock prices and interest and exchange rates) are measured precisely and thus are not subject to revision.

Variations in Forecast Accuracy over Time

A crucial determinant of the size of forecast errors is the forecast period; some periods are very difficult to predict while others are relatively easy. Figure 1 shows the errors of one-quarter-ahead and four-quarter-ahead forecasts, made by one prominent forecaster, of growth in real GNP from 1971:I to 1991:III. The errors for the different time spans follow different patterns: The four-quarter-ahead forecasts are dominated by the overestimates of the two major recessions, 1974-75 and 1981-82, and the underestimates of the early recoveries from the 1980 and 1981-82 recessions. The only other errors in the four-quarter-ahead forecasts that exceeded 2 1/2 percentage points were a 3.2 percentage point underestimate of the rate of real growth in the year after the October 1987 stock market crash and a 2.9 percentage point overestimate for the 1990:I to 1991:I period, which included the 1990-91 recession.

The one-quarter-ahead forecasts are not so clearly linked to business cycle turning points, even though the largest errors were the overestimates in 1974 and the underestimates of the early recovery from the 1980 recession. In addition, large errors occurred in 1978, 1979, 1983, and 1984. But because the one-quarter-ahead errors, although large, were offsetting, the errors of forecasts covering multi-quarter time spans were not especially great.

Forecasters' reputations probably reached the nadir in 1979-80, when for six quarters in a row virtually all one-quarter-ahead forecasts were in the wrong direction--when forecasts expected positive real growth, it was negative and vice versa. And in the only quarter (1980:II) when everyone's forecast was of the correct algebraic sign, the size of the decline was vastly underestimated.

Figure 2 shows corresponding information for CPI forecast errors. By far the largest errors were the sustained underestimations of the acceleration of inflation in 1973-75 and again in 1978-80. From these experiences forecasters gained the reputation of systematically underestimating inflation. These shortfalls were followed by large overestimates of the rate of inflation in 1983, which was undoubtedly associated with the underestimation of the severity of the 1981-82 recession. Since 1983, the record of forecasting the CPI has been much improved. The one-quarter-ahead forecast errors have exceeded 2 percentage points only in 1990:I and in 1990:III, when the forecasts were made just prior to the sharp increase in oil prices associated with Iraq's invasion of Kuwait. These errors resulted in the 2.1 percentage point underestimate of the inflation rate for the year 1990, the first error that large since the overestimates in 1983. The fact that CPI forecasting errors have declined in absolute terms does not necessarily indicate that forecasting ability has improved, however. The variability of the inflation rate has also been much smaller in the last 10 years. Relative to the 1970's, the 1980's have been an easy time to forecast inflation.

Large variations in forecast accuracy over time have several important implications. First, in terms of comparing different forecasters, it is critically important to compare identical forecast periods. The best forecaster's errors in the 1970's would be far larger, in absolute terms, than an inferior forecaster's errors in the 1980's. More fundamentally, the fact that accuracy varies over time poses a challenge to the constancy assumption needed to make inferences about future periods. Is it possible to know whether the current "easy" period will last or whether we will revert to the hectic 1970's? In the former case, only recent experience would be relevant for estimating the accuracy of current forecasts. But in the latter case, recent experience would be deceptive; it will be important to look at a longer sweep of history to remind us of how much uncertainty there can be.

Has Forecast Accuracy Improved?

The figures clearly suggest forecast accuracy has improved over the past 20 years. Since the four-quarter period ending in 1984:I, no four-quarter real GNP forecast error has exceeded 3 1/4 percentage points and only two have exceeded 2 1/2 percentage points. The record for inflation forecasts has been more impressive: since the four-quarter period ending in 1983:IV, no four-quarter-ahead CPI forecast error has exceeded 2 1/4 percentage points and only one (1989:I to 1990:I) has exceeded 2 percentage points.

These facts undoubtedly overstate the degree of improvement that has been achieved. History shows a close association between business cycle turning points and the size of forecast errors. Much of the improvement merely reflects the fact that no turning point occurred for the 92 months between November 1982 and July 1990. Forecast errors did increase during the 1990-91 recession, when real growth was overestimated by nearly 3 percentage points and inflation underestimated by about 2 percentage points. Even errors this large, far larger than average, pale in comparison with those from earlier recessions.

In order to try to distinguish genuine improvement from a string of good luck, it is helpful to examine a longer time period. Table 3 summarizes the longest consistent forecasting record available--the forecasts of real GNP growth in the following year made each November since 1952 by the Research Seminar in Quantitative Economics (RSQE) at the University of Michigan. The distribution of errors has been fairly stable over time: About half of the errors were less than one percentage point, ranging only from a low of 40 percent in the 1970's to 60 percent in the 1960's and 1980's; about one-fifth of the errors exceeded two percentage points, ranging only from a low of 10 percent in the 1980's to a high of nearly 30 percent in the 1950's. In absolute terms, the largest errors, underestimates of the first years of expansions, occurred in the 1950's. Errors were far smaller in the relatively tranquil 1960's but rose somewhat in the turbulent 1970's; errors in the 1980's were about the same as the 1960's. The 1990's are off to a poor start: The errors for 1990 and 1991 are both larger than the average for the entire period, nearly double the average error in the 1980's.

A long-term trend toward greater accuracy is more apparent when the errors are judged relative to standards, in order to account for varying degrees of difficulty over time. Column (3) in the table compares the MAE of the RSQE forecast with that of a naive rule of thumb that predicts real growth each year to be equal to its average rate in the four previous years. (This rule is more accurate than the simple rule that predicts next year's growth will be the same as this year's growth.) The RSQE errors were 40 to 30 percent smaller than those of the naive rule in the 1950's and 1960's, respectively, and improved to a level nearly 60 percent smaller in the 1970's and 1980's. Column (5) shows that the RMSE of the Michigan forecast has declined steadily relative to the standard deviation of real GNP in each forecast period. The standard deviation of real GNP is a direct measure of the difficulty of forecasting in each period. Alternatively, it can be viewed as the RMSE of a forecaster who knew in advance the average actual growth rate in the forecast period but knew nothing about the yearly deviations from that true average. The Michigan forecasts have improved steadily relative to that hypothetical straw man.

Thus, forecast accuracy seems to have improved, whether viewed from the perspective of several decades or by comparing the recent performance with the rather dismal record in the 1970's and early 1980's. Continuing improvement is not inevitable; the performance in the 1990-91 recession was distinctly worse than average. Future improvement (deterioration) depends on whether forecasting techniques improve more rapidly (slowly) than changes occur in the structure of the economy.

Variations in Accuracy among Variables

It is commonly asserted that particular economic variables are "unpredictable." Because it is easy to find someone who will gladly predict anything, such statements are intended to refer to the accuracy of predictions and not the difficulty of making some prediction, no matter how reliable. It is obvious that some variables can be predicted more accurately than others, but not at all obvious how to compare errors in forecasts of different variables. Is a $10 billion error in GNP better or worse than a 50 basis point error in interest rates? Is a 1 percentage point error for the CPI the same as a 1 percentage point error for the unemployment rate? Clearly, forecast errors for different variables cannot simply be added up. Some kind of standardization is required if a comparison of different variables is even to be attempted.

Although perfection is the goal of forecasting, we know that the future is unknown and we do not expect forecasts to eliminate all uncertainty. A forecast is useful if it can reduce uncertainty. But to measure a reduction presumes some estimate of the level of uncertainty that prevailed initially. Forecast evaluation cannot be done in absolute terms but only relative to some standard, because no unique estimate of the level of uncertainty exists, no totally obvious standard of comparison. The only sensible standard of comparison is some alternative forecasting technique. Traditionally, forecasts have been evaluated relative to simple rule-of-thumb forecasts, such as no change or same change (as in some past period). A no-change standard of comparison is a sensible, even a surprisingly stringent, standard of comparison for several variables--primarily ratios of two variables, such as unemployment rates, profit rates, foreign exchange rates, and interest rates. Most economic variables, however, grow exponentially over time. For these variables, a same-change standard is a more stringent and sensible basis of comparison.

Variations in the difficulty of predicting different variables can be illustrated by examining the forecasts published twice a year in The Wall Street Journal from a survey conducted by Tom Herman. Interest rate forecasts for the next six months have been collected since 1982, and for the next year since 1984; forecasts of real GNP, the CPI, and the unemployment rate have been collected since 1986. Although 68 different individuals have submitted at least one forecast, more than half (36) of these have participated in fewer than 10 of the surveys, and only three have participated in all surveys. We have already seen that forecast accuracy varies over time, so that the infrequent forecasters would benefit from skipping difficult periods and suffer if they missed the easy periods. In order to try to control for these missing forecasts, each forecaster's performance is compared, not to those of the other forecasters but to a straw man--a no-change forecast for interest rates and the unemployment rate, and a same-change forecast for the CPI and the real GNP growth rate. Difficult (easy) periods presumably would also be more (less) difficult for the straw man, so that individuals' performance relative to the straw man would be affected less by missing forecasts or gaps.

The results, summarized in Table 4, show drastic differences among variables in the forecasters' ability to outperform the straw man. At one extreme, none of the forecasters could predict the long-term interest rate a half-year into the future as well as the simple assumption that the rate would not change; 83 percent (10 of the 12 forecasters) were more than 20 percent less accurate than the naive straw man. Only one forecaster, a different individual for the half-year and the full-year horizons, could predict short-term interest rates more accurately than the straw man, and neither forecaster was more than 5 percent more accurate.

At the other extreme, everyone could predict the CPI better than the simple straw man forecast, which predicted that future changes will be the same as the most recent change. Only 14 percent (four of 29 forecasters) were unable to beat the straw man by more than 20 percent in forecasting CPI growth over the next year.

The real GNP growth and unemployment rates are more difficult to estimate than the CPI but not as difficult as interest rates. Only about one-third of the forecasters were unable to outperform the no-change straw man for the unemployment rate. Nearly half of the forecasters could beat the straw man by more than 20 percent for the half-year horizon, and 20 percent of the forecasters were over 20 percent more accurate in the year-ahead forecast.

Real GNP forecasts are compared to two straw men. The first, GNP lag, is the simple idea that real GNP will continue to grow at the same rate as it grew in the last observed half-year. One-third of the forecasters could not improve upon this forecast of the next half-year, while all but one could improve upon this forecast of the following half-year and of the entire year after the forecast is made.

The forecasts were made during the first few days of January and July, a few weeks before the initial estimate of actual growth in the prior quarter was released. Although they did not yet know the preliminary official estimate of the previous quarter, the forecasters had a considerable amount of information on that quarter. A second straw man--GNP lead--compares the forecasts with the preliminary estimate of real GNP growth in the half-year before the forecasts, which is released a few weeks after the forecasts were made. Only a few forecasters slightly outperformed this straw man for the first half-year period, but a majority were more accurate in forecasting real growth in the subsequent half-year and in the full year after the forecast.

This contrast reinforces the earlier observation concerning the importance of forecast release dates. It also illustrates the importance of choosing a straw man as a standard of comparison. Although the no-change and same-change standards applied here seem reasonable, other standards could alter the results. These results are not sensitive, however, to the summary error measure or the actual data employed. Similar results hold for the RMSE instead of the MAE, or for revised actual data in place of the preliminary actual data used in the table.

Variations in Forecast Accuracy among Forecasters

Much of the interest in forecast accuracy stems from the wish to know "Who is the best forecaster?" Appendix A presents the mean absolute errors of nine different forecasters for 24 different variables over the period from 1986:I through 1991:III, corresponding to the period when the National Income and Product Accounts were based to 1982, and prior to the December 1991 benchmark revision to a 1987 base year.[1]

Even a cursory examination of the information in Appendix A shows that no single forecaster dominates all outliers for all, or even most, of the variables.[2] In light of the importance of the time within the quarter when the forecast was made, consider only the early-quarter forecasts, those made in the first month. For most variables, the most accurate forecaster varies depending on the horizon of the forecast. Even for the few exceptions (gross domestic final sales, housing starts, state and local government purchases, and the unemployment rate), three different forecasters were "the best." One of the two remaining forecasters was best in predicting the GNP deflator up through seven quarters ahead. However, different forecasters have different interests; to deem one of these forecasters the best, based on a few variables, runs the risk of misleading those forecast users whose primary interest is in some other variable.

Suppose attention is confined to the concept of the inflation rate; Appendix A shows one forecaster who excels for the CPI measure while a different forecaster excels for the GNP deflator. Assume a forecast user cares only about one specific variable and one specific horizon. Appendix A can be used to determine which forecaster has been the most accurate for that particular variable and horizon, but this does not imply that this particular forecaster will continue to be the most accurate in the future. The reason is that the differences in accuracy are typically fairly small; the "best" forecaster's errors were, on average, less than 10 percent smaller than those of the second best forecaster. These differences are of doubtful economic or statistical significance.

The fact that the accuracy of the most prominent group of forecasters is similar does not imply that all forecasters are equally accurate. A few of the individuals whose performance was summarized in Table 4 commonly made errors that were large multiples of the simple straw man used as a standard of comparison. It is as easy to make poor forecasts as it is difficult to consistently make the best forecasts.

Conclusion

With so much variability in forecasting accuracy, it is easier to disprove any generalization than to offer a valid one. Nevertheless, it seems clear that a major factor in forecast accuracy is the time period to be forecast. Errors were enormous in the severe 1973-75 and 1981-82 recessions, much smaller in the 1980 and 1990-91 recessions, and generally quite minimal apart from business cycle turning points. Because turning points also tend to be periods when simple rule-of-thumb forecasts fare poorly, the moral for the forecast user seems to be not to ignore the forecasts but rather to think carefully about plausible outcomes far from the consensus view.

Clearly, accuracy also varies among variables. For good theoretical reasons, it is difficult to forecast a financial variable where genuinely unique knowledge presents an opportunity to profit. These reasons do not hold as forcefully for standard nonfinancial variables--real GNP, inflation, and unemployment rates--where the opportunities for profit are less apparent. Nevertheless, some nonfinancial variables are also extremely difficult to predict. A prominent example is the change in business inventories, where forecasts are often inferior to a no-change rule of thumb.

The interplay between forecast accuracy and the length or span of the forecast is also important. Forecast accuracy obviously tends to improve as the horizon of the forecast declines. But, at least for real GNP, the improvement is relatively slow over time until the forecast period actually starts, when some actual high-frequency data can be incorporated into the forecast. At the same time, longer time spans are often easier to forecast, as aberrations in the economy and/or noise in the measurement procedures "average out." The variability of four-quarter or eight-quarter cumulative changes is generally smaller than that of quarterly changes.

Finally, the importance of the forecaster, as a determinant of accuracy, is often exaggerated, perhaps by the forecasters themselves. Some forecasters have much to fear from a clear statement of the accuracy of their forecasts. But the vast majority of prominent forecasters, including those who have invited public scrutiny of their performance, have much to gain from disclosure of how accurate their forecasts have been. First, although it may be disappointing to learn that others' performances have been similar, it must be comforting to learn that others cannot document a clearly superior performance. Second, and more importantly, there has been much disillusionment with macroeconomic forecasting. Some of this is justified, but some of it may reflect forecasters' failure to educate forecast users in how much (little) confidence to place in their forecasts. In forecasting, an explanation of how much (little) the forecaster knows can be more useful to the user than a single best guess of what the future will be. Only with some understanding of how large forecast errors are likely to be does the forecaster's message become valuable.

FOOTNOTES

1. Additional summary error measures (RMSE's, Theil coefficients, and mean errors) are available on request from the author.

2. Further information on the participating forecasting organizations is provided in Appendix B.

Table 1
Accuracy of Current Quarter Forecasts of Annual Growth Rate of Real GNP
1981:III to 1991:III

(Percentage Points unless Otherwise Specified)

Relative to PRELIMINARY Actual Data	Early (First month of quarter)	Mid (Second month of quarter)	Late (Third month of quarter)
RANGE	-5.2 to 4.7	-3.2 to 2.8	-2.1 to 1.9
> 1	59%	41%	29%
> 2	15%	17%	2%
> 3	10%	5%	0%
MAE	1.4	1.0	.8
RMSE	1.8	1.4	.9
MEAN	-.2	-.1	-.1
Relative to REVISED Actual Data			
RANGE	-5.8 to 4.4	-4.0 to 3.8	-4.0 to 4.2
> 1	61%	51%	68%
> 2	37%	34%	24%
> 3	22%	20%	15%
MAE	1.9	1.6	1.5
RMSE	2.4	2.0	1.9
MEAN	-.4	-.3	-.3

Note: Preliminary Actual Data are the first estimates released in the month immediately following each quarter's end and are equivalent to what the U.S. Department of Commerce terms "advance" actual data. Revised Actual Data are the last estimates made prior to the benchmark revision. MAE = Mean Absolute Error, RMSE = Root Mean Squared Error, MEAN = Mean Error.

Table 2
Accuracy of Current Quarter Forecasts of Annual Growth Rate of CPI
1980:I to 1992:I

(Percentage Points unless Otherwise Specified)

Relative to PRELIMINARY Actual Data	Early (First month of quarter)	Mid (Second month of quarter)	Late (Third month of quarter)
RANGE	-5.0 to 4.8	-2.7 to 2.2	-3.5 to 1.2
> 1	35%	20%	4%
> 2	22%	8%	2%
> 3	12%	0%	2%
MAE	1.2	.7	.3
RMSE	1.8	1.0	.6
MEAN	.1	.1	-.0
Relative to REVISED Actual Data			
RANGE	-5.2 to 4.0	-2.9 to 2.8	-1.7 to 2.1
> 1	39%	22%	14%
> 2	22%	10%	2%
> 3	10%	0%	0%
MAE	1.2	.8	.5
RMSE	1.7	1.1	.7
MEAN	.1	.1	-.0

Note: See Table 1.

Table 3
Accuracy of RSQE Forecasts of Real GNP 1953 to 1991

(Percentage Points unless Otherwise Specified)

Years	MEAN (1)	MAE (2)	MAE/N4 (3)	RMSE (4)	RMSE/SD Actual (5)
All	-.1	1.3	.51	2.0	.70
1953-71	-.8	1.4	.62	2.2	.84
1972-91	.5	1.2	.43	1.6	.57
1950's	-1.5	2.1	.59	3.2	.90
1960's	-.7	1.0	.71	1.4	.85
1970's	.6	1.4	.39	1.9	.55
1980's	.2	.9	.44	1.3	.51

Note: MEAN = Mean Error, MAE = Mean Absolute Error, RMSE = Root Mean Squared Error, N4 = naive "same as four-year average" forecast, SD Actual = standard deviation of actual real growth in forecast period.
Source: Forecasts: Research Seminar in Quantitative Economics, University of Michigan, *The Economic Outlook for 1992*, Table 1, p. 4.

Table 4
Mean Absolute Errors of Forecasts Relative to Naive Straw Man

Variable (Straw Man) Forecast	High (1)	Low (Ratio) (2)	Median (3)	>1.2 (4)	>1.1 (Percent) (5)	>1 (6)	>.9 (7)	>.8 (8)
Short-term interest rate								
Next half-year	1.48	.97	1.28	67	92	92	100	100
Next year	1.67	.95	1.20	50	67	92	100	100
Long-term interest rate								
Next half-year	1.59	1.04	1.28	83	83	100	100	100
Next year	1.57	.89	1.20	50	75	83	92	100
Unemployment rate								
Next half-year	2.26	.57	.84	3	14	28	31	55
Next year	2.71	.63	.97	14	24	31	62	76
CPI growth rate								
Next half-year	.98	.59	.72	0	0	0	7	21
Following half-year	1.02	.56	.68	0	0	7	10	14
Next year	1.11	.38	.54	0	3	3	10	14
GNP (Lag*)								
Next half-year	2.05	.63	.86	21	31	34	48	72
Following half-year	1.21	.56	.75	3	3	7	17	34
Next year	1.54	.56	.69	3	3	7	14	31
GNP (Lead**)								
Next half-year	3.25	1.00	1.30	69	86	93	100	100
Following half-year	1.49	.74	.96	17	28	38	66	76
Next year	2.09	.78	.99	28	38	48	69	97

Note: Short-term interest, long-term interest, and unemployment rates are relative to a no change straw man. CPI and GNP growth rates are relative to a same-change straw man.
*Lag: Last observed half-year growth rate prior to forecast.
**Lead: Next half-year growth rate after forecast.
Source: Twelve individual forecasters' interest rate forecasts, 1982-91; other variables, 29 individual forecasts, 1986-91, as published in The Wall Street Journal.

Appendix A
Mean Absolute Errors 1986:I to 1991:III

Forecaster	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
	Forecast Horizon (Quarters)							
Change in Business Inventories (Billions of Current Dollars)								
Early Quarter								
DRI	19.7	26.0	31.2	28.5	29.8	31.2	27.7	25.4
GSU	15.8	22.3	28.0	27.5	30.2	37.5	32.9	26.7
LHMA	17.3	21.3	28.3	27.5	29.2	32.8	29.3	28.5
RSQE	17.9	19.7	25.4	26.5	28.3	31.4	30.1	.
WEFA	21.2	20.9	26.3	30.5	34.9	34.8	31.3	31.9
Mid Quarter								
DRI	16.3	23.7	31.2	29.8	31.6	32.6	32.8	27.8
LHMA	15.9	23.2	27.8	29.3	30.9	33.1	31.9	27.5
WEFA	19.4	21.9	23.9	28.7	33.3	34.6	33.4	30.6
Late Quarter								
DRI	15.5	24.6	31.6	29.7	32.6	35.2	36.0	29.0
LHMA	13.2	21.5	24.5	28.2	31.8	33.3	32.9	26.0
Real Change in Business Inventories (Billions of 1982 Dollars)								
Early Quarter								
DRI	16.8	19.3	22.9	22.0	21.5	23.5	17.9	17.9
GSU	13.2	17.3	20.8	20.2	22.1	27.7	23.6	18.7
LHMA	13.4	16.6	20.6	20.4	22.2	26.4	21.0	23.0
RSQE	14.1	14.3	18.9	20.4	20.4	23.8	20.0	.
WEFA	17.8	17.2	20.8	23.1	24.0	26.9	21.0	23.0
Mid Quarter								
DRI	12.2	17.8	21.4	21.7	22.0	24.0	19.6	19.3
LHMA	13.7	17.3	20.4	21.8	22.2	26.4	22.0	19.9
UCLA	13.1	22.4	28.3	29.2	27.9	24.6	.	.
WEFA	14.9	16.4	18.9	21.2	22.5	26.2	21.7	22.4
Late Quarter								
DRI	12.2	20.0	22.5	22.3	22.7	26.1	21.5	20.7
LHMA	11.9	16.2	19.6	20.6	21.5	26.1	23.5	19.9
SPF	13.3	17.0	20.7	21.2	21.7	.	.	.
Total Civilian Employment-Household Survey (Percentage Points, Annual Rates of Growth)								
Early Quarter								
DRI	0.6	0.5	0.5	0.6	0.6	0.6	0.5	0.4
GSU	0.4	0.6	0.7	0.8	0.8	0.9	0.8	0.8
LHMA	1.0	0.8	0.9	0.9	0.9	1.0	0.9	0.9
WEFA	0.9	0.7	0.8	0.7	0.7	0.7	0.7	0.7
Mid Quarter								
KEDI	2.2	1.8	1.6	1.4	1.4	1.4	1.3	.
LHMA	0.6	0.7	0.7	0.7	0.8	0.9	0.8	0.8
UCLA	0.5	0.6	0.6	0.7	0.8	0.7	.	.
WEFA	0.6	0.6	0.6	0.7	0.7	0.7	0.7	0.6
Late Quarter								
DRI	0.4	0.5	0.5	0.5	0.6	0.6	0.5	0.4
LHMA	0.5	0.7	0.7	0.8	0.8	0.9	0.9	0.8

Forecast Horizon (Quarters)

Forecaster	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
				Civilian Unemployment Rate				
Early Quarter								
DRI	0.1	0.2	0.3	0.5	0.6	0.7	0.7	0.7
GSU	0.1	0.3	0.4	0.5	0.6	0.7	0.8	0.9
LHMA	0.2	0.3	0.4	0.6	0.7	0.9	1.0	1.1
RSQE	0.1	0.3	0.4	0.6	0.6	0.7	0.7	.
WEFA	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Mid Quarter								
DRI	0.1	0.2	0.3	0.5	0.6	0.7	0.7	0.6
KEDI	0.1	0.3	0.4	0.5	0.6	0.7	0.8	.
LHMA	0.1	0.2	0.3	0.5	0.6	0.7	0.8	0.9
UCLA	0.1	0.3	0.4	0.6	0.7	0.7	.	.
WEFA	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Late Quarter								
DRI	0.1	0.2	0.3	0.4	0.5	0.7	0.7	0.7
LHMA	0.1	0.2	0.3	0.5	0.6	0.8	0.9	0.9
SPF	0.1	0.3	0.4	0.5	0.6	.	.	.
				Consumer Price Index (Percentage Points, Annual Rates of Growth)				
Early Quarter								
DRI	0.9	0.9	0.8	0.7	0.6	0.5	0.4	0.3
GSU	1.0	0.9	0.8	0.6	0.5	0.4	0.2	0.2
LHMA	0.8	0.9	0.8	0.7	0.5	0.5	0.4	0.4
RSQE	1.2	1.1	1.0	0.8	0.8	0.7	0.6	1.3
WEFA	1.0	1.0	0.9	0.8	0.6	0.5	0.4	0.3
Mid Quarter								
DRI	0.4	0.7	0.7	0.6	0.5	0.5	0.4	0.3
KEDI	1.9	1.1	0.9	0.7	0.7	0.7	0.7	.
LHMA	0.5	0.7	0.7	0.6	0.6	0.5	0.4	0.4
UCLA	0.7	0.7	0.7	0.6	0.5	0.5	.	.
WEFA	0.6	0.8	0.8	0.8	0.6	0.5	0.4	0.3
Late Quarter								
DRI	0.2	0.6	0.6	0.6	0.5	0.5	0.4	0.3
LHMA	0.3	0.6	0.7	0.6	0.5	0.4	0.4	0.4
				Federal Government Purchases, Nominal (Percentage Points, Annual Rates of Growth)				
Early Quarter								
DRI	7.4	5.4	4.9	4.1	3.5	3.2	2.6	1.9
GSU	7.5	6.3	5.1	3.3	3.5	3.0	2.8	2.8
LHMA	8.3	6.5	4.5	3.8	3.5	3.2	2.8	2.3
RSQE	6.9	5.4	4.6	4.5	4.0	3.7	3.2	.
WEFA	8.1	5.6	4.6	3.5	3.2	2.6	2.2	1.9
Mid Quarter								
DRI	7.6	5.1	4.2	3.8	3.3	3.2	2.6	2.0
LHMA	7.6	6.2	4.5	3.9	3.6	3.3	2.9	2.4
WEFA	7.1	4.7	3.3	2.8	2.8	2.5	1.7	1.5
Late Quarter								
DRI	6.7	4.9	3.9	3.8	3.5	3.3	2.8	2.1
LHMA	7.2	5.7	4.2	3.8	3.6	3.2	2.9	2.5

Forecaster	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
	Forecast Horizon (Quarters)							
Federal Government Purchases, Real (Percentage Points, Annual Rates of Growth)								
Early Quarter								
DRI	9.0	4.7	5.4	4.6	3.7	2.9	2.1	1.8
GSU	10.4	5.5	4.4	3.2	2.5	2.1	2.1	1.9
LHMA	10.1	5.3	4.7	3.7	3.3	2.9	2.6	1.8
WEFA	9.0	4.5	4.7	3.4	2.8	2.2	1.8	1.5
Mid Quarter								
DRI	8.6	4.8	4.7	4.3	3.5	3.0	2.2	2.0
LHMA	8.8	5.3	4.3	3.7	3.3	2.8	2.5	1.9
WEFA	8.9	4.1	3.6	3.1	2.4	1.9	1.5	1.4
Late Quarter								
DRI	8.5	4.5	4.4	4.1	3.5	2.9	2.2	1.9
LHMA	9.2	5.3	4.5	3.7	3.2	2.6	2.5	2.1
Federal Surplus (Billions of Current Dollars)								
Early Quarter								
DRI	20.1	29.1	33.0	34.3	32.8	30.9	30.9	34.8
GSU	22.3	27.5	28.1	31.5	31.8	39.2	39.1	44.4
LHMA	28.0	29.8	31.0	37.9	39.8	42.5	43.8	41.8
WEFA	22.4	27.6	25.9	34.7	35.4	39.6	40.2	42.8
Mid Quarter								
DRI	14.8	26.2	34.2	33.2	33.0	31.0	33.2	35.0
KEDI	31.9	28.1	27.2	29.1	35.8	41.6	40.9	.
LHMA	21.7	34.8	32.8	34.9	40.7	42.5	41.7	37.9
WEFA	17.8	28.7	28.5	31.1	33.3	35.4	39.0	43.2
Late Quarter								
DRI	13.8	24.4	30.2	33.0	31.4	29.6	32.2	31.9
LHMA	21.0	30.2	30.9	35.0	40.0	40.9	42.3	39.7
Final Sales, Real (Percentage Points, Annual Rates of Growth)								
Early Quarter								
DRI	1.6	1.2	0.9	0.9	0.9	0.9	0.9	0.8
GSU	1.8	1.3	1.0	1.1	1.1	1.3	1.3	1.2
LHMA	1.5	1.1	1.0	1.0	1.0	1.0	1.0	1.0
RSQE	1.6	1.1	1.1	1.1	1.0	0.9	0.9	.
WEFA	1.3	1.1	1.0	0.9	0.9	1.0	0.9	0.8
Mid Quarter								
DRI	1.5	1.2	0.9	0.9	0.9	0.9	0.8	0.8
LHMA	1.5	1.2	0.9	0.9	0.9	1.0	1.0	0.9
WEFA	1.4	1.1	1.0	0.9	0.9	0.9	0.9	0.8
Late Quarter								
DRI	1.5	1.2	0.8	0.8	0.8	0.8	0.8	0.8
LHMA	1.4	1.2	0.8	0.9	0.9	1.0	1.0	1.0
SPF	1.1	0.9	0.8	0.9	1.0	.	.	.

Forecaster	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
	Forecast Horizon (Quarters)							
Gross Domestic Purchases, Real (Percentage Points, Annual Rates of Growth)								
Early Quarter								
DRI	2.1	1.6	1.5	1.6	1.5	1.4	1.2	1.1
GSU	1.8	1.8	1.6	1.5	1.5	1.6	1.5	1.5
LHMA	1.9	1.6	1.4	1.4	1.4	1.5	1.4	1.3
RSQE	2.0	1.9	1.7	1.6	1.5	1.3	1.1	.
WEFA	2.1	1.5	1.4	1.3	1.3	1.2	1.1	1.0
Mid Quarter								
DRI	1.7	1.6	1.4	1.5	1.5	1.5	1.3	1.2
LHMA	1.6	1.6	1.4	1.4	1.3	1.4	1.3	1.2
WEFA	1.9	1.6	1.5	1.4	1.3	1.2	1.1	1.0
Late Quarter								
DRI	1.7	1.6	1.4	1.4	1.4	1.4	1.3	1.2
LHMA	1.4	1.3	1.2	1.3	1.4	1.4	1.4	1.3
Gross Domestic Final Sales, Real (Percentage Points, Annual Rates of Growth)								
Early Quarter								
DRI	1.8	1.3	1.3	1.3	1.2	1.2	1.0	1.0
GSU	1.8	1.2	1.1	1.2	1.2	1.4	1.3	1.3
LHMA	1.6	1.1	1.1	1.1	1.1	1.2	1.1	1.1
RSQE	2.0	1.4	1.2	1.2	1.1	1.0	0.9	.
WEFA	1.6	1.0	0.9	0.9	0.9	0.9	0.9	0.8
Mid Quarter								
DRI	1.6	1.3	1.3	1.3	1.2	1.2	1.1	1.0
LHMA	1.6	1.1	1.0	1.1	1.1	1.1	1.1	1.0
WEFA	1.6	1.1	1.0	1.0	1.0	0.9	0.9	0.8
Late Quarter								
DRI	1.4	1.1	1.2	1.2	1.1	1.2	1.1	1.0
LHMA	1.1	1.0	0.9	1.1	1.1	1.2	1.1	1.0
Gross Domestic Final Private Sales, Real (Percentage Points, Annual Rates of Growth)								
Early Quarter								
DRI	1.8	1.7	1.5	1.6	1.4	1.3	1.2	1.0
GSU	2.1	1.7	1.6	1.7	1.6	1.7	1.7	1.6
LHMA	2.1	1.6	1.5	1.6	1.5	1.5	1.5	1.4
WEFA	1.8	1.2	1.2	1.2	1.3	1.3	1.2	1.1
Mid Quarter								
DRI	1.6	1.7	1.5	1.5	1.4	1.4	1.2	1.1
LHMA	2.0	1.6	1.4	1.5	1.5	1.5	1.4	1.3
WEFA	1.7	1.3	1.3	1.3	1.3	1.3	1.1	1.1
Late Quarter								
DRI	1.5	1.5	1.4	1.5	1.4	1.4	1.2	1.1
LHMA	1.6	1.4	1.4	1.5	1.6	1.6	1.5	1.4

Forecast Horizon (Quarters)

Forecaster	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
	Gross National Product, Nominal (Percentage Points, Annual Rates of Growth)							
Early Quarter								
BMARK	1.8	1.8	1.7	1.8	.	.	.	.
DRI	1.7	1.3	1.3	1.2	1.2	1.1	1.1	1.1
GSU	1.6	1.7	1.6	1.6	1.6	1.5	1.4	1.3
LHMA	1.8	1.5	1.5	1.5	1.5	1.5	1.4	1.4
RSQE	2.1	1.6	1.5	1.4	1.3	1.1	1.0	.
WEFA	2.1	1.5	1.5	1.4	1.3	1.2	1.1	1.0
Mid Quarter								
DRI	1.4	1.3	1.3	1.2	1.2	1.2	1.1	1.1
KEDI	2.5	2.3	2.4	2.3	2.1	1.8	1.6	.
LHMA	1.5	1.5	1.5	1.4	1.4	1.4	1.4	1.3
UCLA	1.5	1.4	1.4	1.4	1.4	1.2	.	.
WEFA	1.6	1.5	1.5	1.4	1.3	1.2	1.1	1.0
Late Quarter								
DRI	1.3	1.2	1.2	1.1	1.2	1.2	1.1	1.1
LHMA	1.2	1.3	1.3	1.3	1.4	1.5	1.4	1.4
SPF	1.5	1.3	1.4	1.4	1.3	.	.	.
	Gross National Product, Real (Percentage Points, Annual Rates of Growth)							
Early Quarter								
BMARK	1.9	1.7	1.7	1.6	.	.	.	.
DRI	1.7	1.2	1.2	1.2	1.2	1.1	1.0	1.0
GSU	1.6	1.6	1.5	1.5	1.5	1.5	1.4	1.4
LHMA	1.9	1.3	1.3	1.3	1.3	1.4	1.3	1.3
RSQE	1.6	1.6	1.6	1.5	1.4	1.3	1.1	.
WEFA	2.1	1.5	1.5	1.4	1.3	1.2	1.2	1.0
Mid Quarter								
DRI	1.5	1.3	1.2	1.2	1.2	1.1	1.0	1.0
KEDI	1.8	1.7	1.6	1.6	1.6	1.4	1.2	.
LHMA	1.6	1.4	1.3	1.3	1.3	1.3	1.2	1.1
UCLA	1.6	1.6	1.6	1.5	1.4	1.1	.	.
WEFA	1.7	1.6	1.5	1.4	1.3	1.2	1.2	1.0
Late Quarter								
DRI	1.5	1.3	1.2	1.1	1.2	1.1	1.0	1.0
LHMA	1.4	1.2	1.2	1.2	1.3	1.3	1.2	1.2
SPF	1.4	1.2	1.2	1.3	1.2	.	.	.
	Housing Starts (Millions of Units)							
Early Quarter								
DRI	0.1	0.1	0.2	0.2	0.2	0.3	0.3	0.3
GSU	0.1	0.1	0.1	0.2	0.2	0.2	0.2	0.2
LHMA	0.1	0.1	0.1	0.2	0.2	0.2	0.3	0.3
Mid Quarter								
DRI	0.0	0.1	0.1	0.2	0.2	0.3	0.3	0.3
KEDI	0.1	0.2	0.2	0.3	0.4	0.4	0.5	.
LHMA	0.0	0.1	0.1	0.2	0.2	0.2	0.3	0.3
UCLA	0.1	0.1	0.1	0.2	0.2	0.3	.	.
Late Quarter								
DRI	0.0	0.1	0.1	0.2	0.2	0.2	0.3	0.3
LHMA	0.0	0.1	0.1	0.2	0.2	0.2	0.2	0.3
SPF	0.1	0.1	0.1	0.2	0.2	.	.	.

Forecaster	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
	Implicit GNP Price Deflator (Percentage Points, Annual Rates of Growth)							
Early Quarter								
BMARK	1.1	0.9	0.7	0.6	.	.	.	.
DRI	1.2	0.9	0.7	0.6	0.4	0.3	0.3	0.2
GSU	1.1	0.8	0.6	0.5	0.4	0.4	0.4	0.5
LHMA	0.8	0.7	0.5	0.4	0.4	0.3	0.3	0.3
RSQE	1.4	1.1	1.0	0.8	0.7	0.7	0.6	.
WEFA	1.1	0.8	0.6	0.5	0.5	0.4	0.3	0.4
Mid Quarter								
DRI	0.8	0.7	0.6	0.4	0.4	0.3	0.2	0.1
KEDI	2.1	1.4	1.4	1.1	0.8	0.7	0.7	.
LHMA	0.8	0.7	0.5	0.4	0.3	0.2	0.2	0.3
UCLA	0.9	0.6	0.5	0.4	0.3	0.2	.	.
WEFA	1.1	0.8	0.5	0.5	0.5	0.5	0.4	0.5
Late Quarter								
DRI	0.8	0.7	0.7	0.5	0.4	0.3	0.3	0.2
LHMA	0.9	0.7	0.5	0.5	0.4	0.3	0.3	0.2
SPF	1.0	0.7	0.5	0.5	0.5	.	.	.
	Investment in Residential Structures, Real (Percentage Points, Annual Rates of Growth)							
Early Quarter								
DRI	7.5	6.4	6.2	6.0	5.7	5.3	5.1	4.6
GSU	5.0	6.0	5.7	6.5	5.9	5.5	5.3	5.0
LHMA	6.7	5.9	5.4	5.5	5.5	5.5	5.5	5.7
RSQE	7.3	6.7	7.0	8.1	8.1	7.9	7.9	.
WEFA	7.0	6.4	6.1	6.2	6.1	5.9	5.6	5.2
Mid Quarter								
DRI	5.9	5.5	5.7	5.4	5.6	5.5	5.5	5.0
LHMA	6.3	5.5	5.2	5.0	5.1	5.0	5.1	5.2
WEFA	5.8	5.2	4.8	5.3	5.7	5.7	5.4	5.3
Late Quarter								
DRI	5.5	5.2	5.5	5.8	6.2	5.9	5.4	4.8
LHMA	6.5	5.5	5.5	5.8	5.9	5.6	5.4	5.3
	Net Exports of Goods and Services (Billions of Current Dollars)							
Early Quarter								
DRI	10.7	14.8	23.6	26.6	28.3	29.3	25.5	25.1
GSU	11.7	15.6	20.0	23.2	23.7	26.7	27.4	27.5
LHMA	10.8	14.5	16.3	22.4	28.6	34.3	38.8	41.1
RSQE	14.0	21.1	25.6	30.0	36.5	44.3	48.6	.
WEFA	10.0	14.0	19.3	25.1	31.5	38.3	40.4	40.8
Mid Quarter								
DRI	9.8	17.8	24.3	29.6	31.3	31.4	29.7	27.5
LHMA	9.8	13.7	28.2	24.1	30.4	36.5	42.8	45.3
WEFA	10.1	16.8	21.4	28.5	35.0	40.7	43.7	45.0
Late Quarter								
DRI	11.1	17.9	22.6	28.9	31.9	31.1	29.8	27.1
LHMA	10.1	16.3	18.4	20.8	27.0	32.9	39.8	42.2

Forecaster	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
	Forecast Horizon (Quarters)							
	Real Net Exports of Goods and Services (Billions of 1982 Dollars)							
Early Quarter								
DRI	12.6	17.1	21.9	24.0	27.8	28.0	30.0	29.7
GSU	13.2	17.3	16.7	15.3	17.5	19.5	20.1	19.0
LHMA	12.4	16.7	14.1	15.5	21.6	22.1	24.5	25.9
RSQE	14.2	15.7	14.3	14.9	21.4	22.0	20.3	.
WEFA	11.0	12.4	15.6	14.7	18.8	22.9	20.6	19.0
Mid Quarter								
DRI	11.5	16.5	19.1	22.4	24.6	27.6	28.4	27.4
LHMA	11.5	15.2	13.8	16.5	22.0	22.5	25.7	25.7
WEFA	11.6	14.5	16.6	19.2	22.1	27.4	27.7	24.8
Late Quarter								
DRI	12.1	16.6	18.0	22.9	25.6	28.9	30.0	28.9
LHMA	11.1	16.0	14.0	14.7	21.8	24.6	24.4	25.4
	Nonresidential Fixed Investment, Real (Percentage Points, Annual Rates of Growth)							
Early Quarter								
DRI	6.1	4.8	4.1	3.7	3.6	3.3	3.0	2.4
GSU	7.5	5.8	4.6	4.3	4.3	4.2	3.6	3.7
LHMA	8.0	5.1	4.4	3.6	3.5	3.3	2.9	2.6
RSQE	7.7	5.9	4.7	4.3	4.2	3.7	3.5	.
WEFA	7.3	4.8	3.8	3.2	2.9	2.7	2.6	2.3
Mid Quarter								
DRI	5.8	4.8	3.9	3.6	3.5	3.4	3.0	2.6
LHMA	7.7	5.1	4.0	3.4	3.5	3.1	2.9	2.7
WEFA	6.3	4.7	3.6	3.2	2.9	2.6	2.2	2.0
Late Quarter								
DRI	6.2	4.2	3.5	3.5	3.4	3.4	3.1	2.6
LHMA	7.4	4.7	3.9	3.3	3.5	3.1	2.8	2.8
	Personal Consumption Expenditures, Durable Goods, Real (Percentage Points, Annual Rates of Growth)							
Early Quarter								
DRI	7.4	5.2	4.6	4.4	3.7	3.7	3.5	3.3
GSU	8.8	5.4	4.5	4.3	3.9	4.4	4.1	3.7
LHMA	7.1	4.6	3.8	3.6	3.2	3.6	3.3	3.3
RSQE	8.1	5.8	4.7	4.2	3.5	3.1	2.8	.
WEFA	7.3	4.9	3.7	3.5	2.7	2.8	3.0	2.7
Mid Quarter								
DRI	7.0	4.5	3.9	4.1	3.5	3.7	3.5	3.3
LHMA	7.7	5.5	4.1	3.7	3.0	3.5	3.2	3.0
WEFA	7.1	4.9	4.0	3.5	2.8	2.9	2.9	2.6
Late Quarter								
DRI	4.3	4.1	3.9	4.1	3.6	3.6	3.4	3.2
LHMA	4.9	4.2	3.6	3.5	3.1	3.5	3.3	3.0

Forecaster	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
	Personal Consumption Expenditures, Nondurable Goods and Services, Real (Percentage Points, Annual Rates of Growth)							
Early Quarter								
DRI	1.3	1.0	0.9	0.7	0.6	0.5	0.5	0.5
GSU	1.2	1.0	0.9	0.8	0.8	0.8	0.7	0.7
LHMA	1.4	1.2	1.0	0.9	0.9	0.9	0.8	0.7
RSQE	1.5	1.3	0.9	0.7	0.7	0.7	0.6	.
WEFA	1.1	0.9	0.7	0.7	0.7	0.6	0.6	0.6
Mid Quarter								
DRI	1.2	1.1	1.0	0.7	0.6	0.7	0.5	0.5
LHMA	1.2	1.0	0.9	0.9	0.8	0.8	0.7	0.7
WEFA	1.2	1.0	0.8	0.7	0.6	0.7	0.6	0.6
Late Quarter								
DRI	1.2	1.2	0.9	0.7	0.6	0.6	0.5	0.5
LHMA	1.2	1.2	0.9	0.9	0.8	0.8	0.8	0.7
	State and Local Government Purchases, Real (Percentage Points, Annual Rates of Growth)							
Early Quarter								
DRI	2.5	2.2	1.9	1.8	1.9	1.9	1.9	1.8
GSU	2.2	1.7	1.4	1.4	1.3	1.3	1.3	1.3
LHMA	1.8	1.5	1.2	1.0	1.0	1.0	1.0	1.0
WEFA	1.6	1.2	0.9	0.9	0.9	0.9	0.8	0.8
Mid Quarter								
DRI	2.0	1.8	1.7	1.6	1.7	1.6	1.7	1.6
LHMA	1.8	1.4	1.3	1.0	0.9	0.9	0.9	0.9
WEFA	1.7	1.2	1.0	0.9	0.9	0.8	0.8	0.8
Late Quarter								
DRI	2.0	1.8	1.7	1.7	1.7	1.7	1.7	1.6
LHMA	1.8	1.4	1.2	1.0	0.9	0.9	0.8	0.9
	90-Day Treasury Bill Rate							
Early Quarter								
DRI	0.1	0.4	0.8	1.0	1.1	1.2	1.3	1.5
LHMA	0.2	0.5	0.8	1.0	1.1	1.3	1.4	1.6
RSQE	0.1	0.4	0.8	1.1	1.3	1.4	1.5	.
WEFA	0.2	0.5	0.7	0.8	0.9	0.9	0.9	0.9
Mid Quarter								
DRI	0.1	0.3	0.7	0.9	1.0	1.1	1.3	1.5
KEDI	0.3	0.6	0.8	0.9	1.1	1.3	1.4	1.4
LHMA	0.1	0.4	0.7	0.8	1.0	1.1	1.2	1.4
WEFA	0.1	0.4	0.7	0.8	0.8	0.8	0.8	0.9
Late Quarter								
DRI	0.0	0.3	0.6	0.8	1.0	1.2	1.4	1.5
LHMA	0.0	0.3	0.6	0.8	0.9	1.1	1.2	1.3

Forecast Horizon (Quarters)

Note: . = more than two forecasts not available.

Appendix B

Summary Information on Forecasting Organizations Studied

Forecasting Organization (Abbreviated Title), Contact for Further Information	Number of Macroeconomic Variables Forecast[a]	Typical Forecast Horizon, Quarters	Frequency of Release, per Year	Date Forecast First Issued Regularly
1) Benchmark Forecast (BMARK) George Washington University, Frederick Joutz (202) 994-4899	30	8	4	1976
2) Data Resources, Inc. (DRI), Roger Brinner (617) 863-5100	1,200	10 to 12	12	1969
3) Georgia State University (GSU) Economic Forecasting Project, Donald Ratajczak (404) 651-3282	540	8	4	1973
4) Kent Economic and Development Institute, Inc. (KEDI), Vladimir Simunek (216) 678-8215	1,700	10	12	1974
5) Laurence H. Meyer & Associates, Ltd. (LHMA), Larry Meyer (314) 721-4747	450	7 to 11	12	1983
6) Research Seminar in Quantitative Economics (RSQE), University of Michigan, Saul Hymans (313) 764-3299	200	8	8	1969
7) Survey of Professional Forecasters (SPF), Federal Reserve Bank of Philadelphia, formerly ASA/NBER, Dean Croushore (215) 574-3809	20	5	4	1968
8) University of California at Los Angeles (UCLA), School of Business, David Hensley (310) 825-1623	1,000	8 to 12	4	1968
9) Wharton Econometric Forecasting Associates, Inc. (WEFA), Kurt Karl (215) 660-6357	1,000	12	12	1963

[a]Approximately.

Alternatives to Forecast Error Based Evaluation: Communicability, Manipulability, Credibility and Policy Relevance

Stuart Bretschneider, Center for Technology and Information Policy, The Maxwell School, Syracuse University
Wilpen Gorr, H. John Heinz School of Public Policy and Management, Carnegie Mellon University

Introduction

The past decade has witnessed an expansion in the amount of research generated on forecasting as well as a more focused perspective on evaluation (Armstrong 1986). Several authors have noted that earlier forecasting research followed an 'advocacy' model (Armstrong and Collopy 1992), where researchers generated new approaches or innovations in forecasting and 'tested' them, usually on a single data sample in a very limited setting. Much of this early work can be faulted on purely technical grounds associated with experimental research designs, including biased samples and the use of limited or incomplete controls. These technical deficiencies led to inappropriate comparisons across alternative techniques and inaccurate results.

Since the path breaking work by Makridakis and Hibon (1979) most forecasting research has adopted more structured approaches to evaluation of new forecasting techniques. Most of the advances in forecasting research have come by tightening up study 'designs', usually through the adoption of standard 'scientific' and 'experimental' designs. This new wave of forecasting research has made extensive application of large samples, careful designs and an emphasis on alternative measures of accuracy and bias (Armstrong and Collopy 1992).

Nevertheless, all of these advances have focused on a single underlying basis of evaluation--laboratory-based reproducible forecast errors. Though a significant advance and a necessary step to moving forecast research beyond simple innovator advocacy, these measures fail to consider the organizational context of forecasting. Forecasting occurs within large complex organizations. The phenomenon being forecasted, the organizational forecasting process, the use of forecasts, and consequently the value of forecasting varies tremendously across organizational settings. Many of these factors influence the efficacy of forecasting and many of these factors are manipulable. Some have argued that these less often studied factors may, in fact, be greater sources of forecast error than technical factors such as model selection and estimation technique (Bretschneider and Gorr 1987; Bretschneider and Gorr 1991a). Why then have forecasting researchers not examined additional concepts of valuation for forecasts beyond those based directly on forecast error?

The purpose of this paper is to first understand why forecasting research is stuck on the use of reproducible forecast error as the basis for all evaluation, and secondly to propose and demonstrate alternative and supplemental metrics for evaluation. The next section discusses the importance of evaluation to advancing knowledge of forecasting. This is followed by a discussion highlighting the reasons forecast errors are the dominant basis for evaluating alternative forecasting approaches. The next section provides a multidimensional view of evaluation for forecasts which considers organizational context. This is followed by an empirical example based on survey data from practicing forecasters. The final section of the paper considers the next steps necessary to advancing the quality of forecasting research.

The problem of evaluation of forecasts: why it's important and how it's currently done

Forecasting, like many other types of research, reflects not only a basic human need to understand but also a need to control. Unlike a great deal of other scientific research, though, forecasting research is motivated to improve the actual performance of people and organizations as they carry out forecasting. A consequence of this basic motivation is to force most forecasting research into a comparative mold. Current practice must be compared with proposed alternatives and various alternatives must be compared with each other. Not surprisingly, even the most extreme 'advocacy' oriented forecasting research maintained some level of comparative analysis. For example, the work by Box and Jenkins (1976) though mainly a textbook approach or 'how to do it manual' still maintained elements of comparison within it (pp. 167-170).

For a comparison to have any value, particularly to those responsible for running organizations, there are two necessary considerations. The first concern is that the comparison be relevant and the second is that the results from the comparison seem credible to those reading the work. Focusing on the second consideration--credibility of results--we note that researchers tend to espouse scientific norms. Hence the application of scientific criteria in the design of studies leads to enhanced credibility among scientists and their peers. This goes a long way to explain why more and more forecasting research has been concerned with sampling, measurement, and experimental design. Enhancing each of these dimensions of a study will enhance the credibility (as well as the publishability) of the research. Unfortunately, there is some evidence to suggest that the scientific factors which enhance credibility among researchers do little to enhance credibility among practitioners (DeRoeck 1992; Mahmoud et. al. 1992). In fact, increase concerns over scientific methodology tends to make the published research less accessible to practitioners by making it overly technical and may have a negative effect on credibility.

Relevancy is the other major concern when carrying out comparison in forecasting research. This is much more important to practitioners, who consider any innovation or change in terms of one or more organizational objectives. Relevancy in a forecasting comparison is directly related to the extent a study demonstrates a link between how forecasting is done and specific organizational objectives. In the case of business, increasing revenues or decreasing costs are the most common objectives, though other important goals such as maintenance of quality in outputs are also used. Public sector organizations apply a similar approach though organizational goals are often different (such as maintenance of procedural equity and maximizing public benefits).

Once objectives have been established, attributes of forecasts or the forecasting process must be linked to those objectives so that comparisons can then be made. This process of linking forecast attributes to goals is necessary to obtain any level of relevance, yet it is currently one of the least developed parts of forecasting research. In formal decision theory and operations research, the link between actions and objectives is referred to as an objective or value function. This can be expressed in the following form:

$$V_i = F_i(x_1,\ldots,x_p) \qquad (1)$$

where V_i is the value to the organization (e.g. revenues) of its ith objective and x_1 through x_p are various factors which describe different attributes of forecasts or forecasting processes. Techniques in operations research and decision theory focus on how to make use of these functions in various settings, such as when there are multiple and conflicting objectives ($i=1,2,\ldots m$). The use of this framework for formal decision making is beyond the scope of this paper. Nevertheless, the framework is useful in identifying that relevancy in forecast research requires measurement of organizational objectives, measurement of multiple dimensions of forecasts and the forecasting process, and some linking of these variables through either an implicit or explicit relationship like an objective function. Figure 1 attempts to make use of this framework to characterize the bulk of current forecasting research.

```
Organizational                          Forecast
Objectives                              Attributes

V1<--+                                  +----x1
     |                                  |
V2<--+----------Forecast          <-----+----x2
.    |           Error                  |     .
.    |                                  |     .
.    |                                  |     .
Vm<--+                                  +----xp

      Implicit                   Explicit
      Assumed                    Empirical
      Relationship               Relationship
```

Figure 1: Decision Making Framework Applied to Forecasting Research

Generally empirical forecasting research does not directly link forecast attributes to organizational objectives. Rather the valuation is focused almost entirely on intermediary variables that are almost exclusively based on forecast error (Stekler 1991; Armstrong and Collopy 1992; Fildes 1992; for an exception see Gardner 1990). Forecast error can be used to construct several different measures of forecasting performance (e.g. bias, accuracy) but these are not direct measures of organizational goals. Instead the link between forecast error based measures and organizational objectives is implicit and often simply assumed. There are a few studies that provide more explicit links (for example Gardner 1990). The most common implied link is one where the researcher assumes cost is directly related to forecast error, usually referred to as a loss or cost function. Though there is little doubt that such links exist, other factors influence the nature of those links. For example, although reducing forecast error in final demand for goods directly reduces both production and inventory costs, the magnitude of the effect differentially depends on numerous other variables related to production costs, relative importance of the product line to the overall profitability of the organization, current status of product with regard to its life cycle, etc. Many of these factors and relationships are generalizable and even reasonably well known, but typically not dealt with by researchers. Another criticism of the characterizations provided in Figure 1 is that the link between forecast characteristics and organizational goals does not only occur through intermediary variables based on forecast error. Other relevant intermediary variables include cost of generating the forecast, and timeliness of the forecasting process.

Why the over-emphasis on forecast error?

Those producing research on forecasting and those most in need of the results form two distinct groups. Academics generate the bulk of published forecasting research while practitioners in business and government represent the major market for such work. Even applied research organizations providing support for many organizations, tend to employ

Ph.D. level researchers (e.g. Economic Research Service of the U.S. Department of Agriculture). As noted above, credibility and publishability concerns are reasons for published forecasting research to become more 'scientific' and potentially less accessible to practitioners (DeRoeck 1991). These same influences have also led to the over-emphasis of forecast error based measures of outcome.

Forecast errors and forecast error based measures are easily reproduced and replicable measures of forecast performance. Replicability is an important criteria of scientific research, one that enhances the credibility of the work. Access to data and model specification are typically sufficient conditions for replication of forecast errors. Having direct access to the value measures used by other researchers makes it possible for new research to, not only replicate prior work, but also to design studies which build incrementally on that prior work. The data from 1,001 time series used in a comparison of over 20 different univariate forecasting methods plus forecast errors generated by application of those methods has been made available to researchers (Makridakis et. al. 1982). Several of the major technical advances in forecasting research came through direct access and use of that data and the forecasting errors generated by the earlier study (Makridakis and Winkler 1983; Gardner and McKenzie 1985; Makridakis 1990).

Another major argument for using forecast errors and forecast error based measures for evaluation is parsimoniousness. As pointed out earlier there is little argument that forecast errors are directly related to at least some of the major organizational objectives. Empirical work has demonstrated that most of the major forecasting and forecasting process variables can be successfully related to forecast errors as well (Makridakis et.al. 1982, Bretschneider and Gorr 1991a). It is therefore reasonable to view forecast errors as a parsimonious, though intermediate, measure of organizational value. The fact that using forecast errors is reasonable and parsimonious again reflects the value system inherent in 'scientific' research. Parsimonious representations of reality are generally preferred. Unfortunately, the level of simplicity represented by using error-based valuation increases the likelihood that the results are less applicable in practice. This is not an argument for highly idiosyncratic models of reality, Rather, it is an argument for more contingency-based models of reality that investigate obvious prior factors affecting the performance of forecasting models and procedures.

The final factor influencing the over-emphasis of error-based valuation is the cost of generating measures of valuation. This argument has two components. Firstly, generation of forecast errors is a relatively inexpensive activity for most researchers. Forecasting researchers have access to data (even forecast errors generated by prior research). They have access to sophisticated computers and the skills to make use of them for generation of forecast errors. They also tend to have the necessary capacity in mathematics to generate alternative techniques or models as well as to understand similar work by others. The second part of the argument is that alternative measures of forecast valuation are relatively expensive to obtain. Since many of the alternative measures require some organizational context, measurement of non-error based valuation requires field work in the form of case study, survey research or document collection. It will also require investments in learning either new techniques of research (e.g. survey research, interviewing, etc.) or new substantive knowledge (e.g organization theory, sociology and political science) or both. The bottom line--error-based valuation is easier and cheaper to do than the alternative.

An Alternative Approach

What are the alternatives to focusing solely on forecast errors? Again using a decision making orientation, Figure 2 provides an alternative model for framing forecasting research.

```
Organizational        Intermediate   Forecast
Objectives            Variables      Attributes
     V1<--+           +----I1        +----x1
          |           |              |
     V2<--+     <-----+----I2   <----+----x2
     .    |           |     .        |     .
     .    |           |     .        |     .
     .    |           |     .        |     .
     Vm<--+           +----In        +----xp

     |<---Explicit------>|<---Explicit---->|
          Empirical                Empirical
          Relationship        Relationship

     |<------------Explicit-------------->|
                  Empirical
                  Relationships
```

Figure 2: Model for Prescriptive Forecasting Research

There are essentially three major modifications present in the alternative model: increasing the number of intermediate variables that are considered as outcomes; an increased focus on understanding how the intermediate variables relate to specific organizational objectives and increased efforts at directly linking forecast descriptors with organizational objectives.

In order to make these ideas more explicit, Table 1 provides some suggestions for potentially useful intermediate variables for the study of forecasting. Many of these measures, not unlike forecast error measures, can be simultaneously viewed as both attributes of a forecast and intermediate measures. For example, it is possible to have forecasters or decision makers within organizations characterize the extent to which a particular forecasting approach makes use of 'reasonable' assumptions or not. Alternatively, forecasters or decision makers can subjectively assess the importance placed on assumptions as a valuation metric. For some situations, such as forecasting the effect of a tax rate change, many univariate models make unreasonable assumptions about future conditions, while a regression model using tax rate to predict revenue might be predicated on more 'reasonable' assumptions. This discussion also demonstrates that the valuation of reasonableness of assumptions is not independent of who is providing the assessments--forecasters or decision makers. Alternatively, applications of the reasonableness of assumptions as a criteria for valuation might be less significant when considering alternative univariate forecasting approaches to forecasting demand for a products at the SKU level.

Many of these variable can be measured either in terms of subjective attitudes or direct measurement. For example, all of the cost variables can be directly measured in terms of dollars or subjectively by attitude scales or importance scales. Regardless of the approach to measurement, access to organizations is a necessary condition for measurement. Attitudes and subjective valuation can be carried out through survey or interview techniques but access to forecasts or users of forecasters within an organization is necessary. To obtain direct measures such as dollar cost also requires direct access to organizations. This requirement dramatically increase the cost of doing forecasting research.

Though the major cost is that of data collection, usually in the form of survey costs or carrying out on-site visits, there are other problems associated with obtaining measurements from operating organizations. Obtaining the cooperation of an organization is in part a data collection cost. Nonresponse problems in survey research illustrates how cooperation can be directly measured and costed. Sometimes cooperation is attainable but a constraint is imposed on the research effort in terms of publishing results based on using the organizations data. Concerns over competitiveness or unwanted public attention can also impose costs on this type of research.

Once data on these variables are collected there are several types of analyses that make sense. Prior studies that have collected data on forecasting techniques and forecast errors have attempted to model explicit relationships (Makridakis et al 1982; Bretschneider and Gorr 1991a). Similar work is possible from both objective and subjective measure of intermediate variables (Bretschneider and Gorr 1991b). Both relationships between forecast methods and/or process can be related to subjective valuations provided by decision makers in a variety of dimensions. It is likely that these relationships will be contingent on what is being forecasted and how the forecasts are used in the decision process.

Figure 2 also calls for enhanced efforts at measurement of final organizational objectives. Once measures of organizational outcomes, such as profits, cost, and service levels are available, modeling of relationships between intermediate variables and objectives is possible. It seems important that such relationships be more formally understood before effective prescription is possible.

An Empirical Example

A survey of professional forecasters in the U.S. Federal Government was conducted to test the feasibility and potential for measuring and studying alternative intermediate forecasting variables. The sample frame for this study consisted of 259 members drawn from the 1989 and 1990 directories of a professional organization known as the Federal Forecasters. Members of the Federal Forecasters have either had forecasting as part of their job responsibilities in the federal government or have had an interest in forecasting. The Association is voluntary, and the principle activity of the organization has been an annual conference where ideas are exchanged and discussed through the use of keynote speakers and presentation of papers.

As the above description implies, the use of the Federal Forecasters yields a sample frame of convenience, and cannot be construed to be a purely random sample of forecasters working in the federal government. Nevertheless, the sample does represent several federal agencies and has value for an exploratory study.

Survey Process

The survey process consisted of an alert letter, and an initial mailing of the survey with a cover letter two weeks later. There were two follow-up mailings for non-respondents. After the first follow-up mailing a 53% response rate was reached. After the second the overall response rate 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, as well as individuals who referred to others in their organizations who had already filled out a survey. Of the total number of respondents, over 75% are actively involved with developing or reviewing forecasts. In total, 115 useable survey responses were obtained for a 45% sample.

Some individuals felt that the survey process was not sufficiently blind due to the use of control codes that identified individual respondents. This may have led to some problems of sample selection.

Responding Individuals and Organizations
In terms of educational background, 78% of the respondents have a Masters degree or higher and 47% have a background in economics. 75% of the respondents had at least five years of work experience in their current organization. In terms of experience in forecasting, 57% had taken a formal course in forecasting and the median respondent has eight years of experience in the forecasting area.

Finally, the nature of the forecasting groups represented varied significantly. Not only was a wide spectrum of agencies represented in the sample, but these organizations varied significantly in terms of size. Though the mean group employed 13 full time equivalent staff members and had a budget of over $800,000, the median organization only employed five staff with a budget of $300,000. Clearly the sample contains a few very large forecasting units, while most employ 10 or fewer technical staff.

Measurement of Intermediate Variables and Underlying Dimensions:
The survey instrument included a list of the 25 intermediate variables from Table 1. Respondents were instructed to indicate the importance of each 'criteria' for evaluating the quality of a forecast. The objective was to develop a quantifiable view of the objectives for these professional forecasters. The scale was a 7-point scale which ran from 1 for "not important" to 7 for "essential." Table 2 provides some descriptive statistics for each of the variables.

In order to determine the extent to which general underlying evaluative criteria exist, factor analysis was used. Since the data are ordinal in nature, a matrix of Kendall's Tau-B correlations was calculated. This matrix was used as input to estimation of a five factor model using unconditional least squares estimation. The resulting factor weights were then rotated using a orthogonal varimax rotation. The results are provided in table 3. Factor weights above 0.46 were deemed large enough to indicate substantive importance and are signified in Table 3 by an asterisk. The results are encouraging.

The first factor captures several dimensions including both measures of comparative evaluation (e.g. benchmarking), all four criteria associated with the management of uncertainty, two of the three measures associated with manipulability of forecasts related to adjusting data or inputs, and both manpower and computer costs. Though this dimension captures several components, the dominant dimension is in managing uncertainty either in the form of presenting alternative futures or through the ability to adjust and manipulate inputs (as part of the process for generating alternative estimates of the future). The second, third and fourth factors are much more straight forward. The second factor captures all the components we have termed "explainability components" such as defensibility of method, reasonableness of assumptions, and linking assumptions and data to forecast outcomes. The third dimension is dominated by concerns over coordination across units while the fourth captures the traditional error-based criteria, such as accuracy and bias. The final factor captures both questions under utility criteria (indicating forecasts should have multiple uses) and the use of forecasts in evaluation of policy alternatives.

The most significant point in these findings in that forecast error measure metrics do not dominate the evaluation criteria, at least for those that participated in the survey. The technical criteria were associated with the fourth factor and accounted for only 17% of the total variation in responses. If we focus on the communalities associated with each of the original variables in the final factor model, the single most influential variable was the reasonableness of assumptions. Of the top five variables, three were related to the explainability dimension while only one was related to error based norms.

There are, of course, a number of plausible explanations for these results. First, there is the potential for sample selection in response as well as bias resulting from the overall definition of the sample frame. The Federal Forecasters undoubtedly represent a select group within the forecasting profession, but then the same could be said of studies that focus solely on individuals and organizations forecasting for inventory management of private firms. Despite the potential for sample selection bias some useful results emerge:

1) Evaluation based on criteria other than forecast error metrics seems not only plausible, but empirically observable.

2) Explainability seems to dominate forecast error and pure accuracy measures, at least for this sample.

3) The capacity to represent and manage uncertainty in a variety of forms is important, particularly the potential

to manipulate inputs, carry out scenario analysis, and study the impact of alternative policies.

4) Coordination of forecasting across groups is also important--for example coordination through the use of standardized data and standardized forecasting reports.

Though the emphasis on policy analysis is related to the nature of the sample, some of these factor could have a significant bearing on the evaluation of forecasting in private sector organizations as well. Coordination is often cited as a general management problem. When forecasts are used by different elements of the same business organizations for different purposes, there is a strong likelihood that coordination factors will weigh heavily in assessing forecasting systems for business.

Conclusions

The major strides in forecasting research achieved in the 80's came through a 'tightening up' of research designs and a heavy emphasis on the use of forecast error norms for comparative evaluation. The next step is to broaden our view of forecasting research to include a broader range of intermediate variables for measuring forecast outcomes and more attention to linking all the various measures of forecast outcomes to organizational objectives. These new directions should not be viewed as a move toward highly idiosyncratic and applied research which would fail to measure up to a variety of scientific norms, such as replicability and generalizability.

These new directions require some modification in the research skills currently represented among the forecast research community. It requires that forecasting research pick up tools used by social and organizational scientists as they study forecasting as an organizational process.

References

Armstrong, J. Scott, "Research on Forecasting: A Quarter-Century Review, 1960-1984," Interfaces, Vol. 16 (January-February, 1986), pp. 52-66.

Armstrong, J. Scott, Long Range Forecasting: From Crystal Ball to Computer, Second Edition, New York: John Wiley and Sons, 1985.

Armstrong, J. Scott, and F. Collopy, "Error Measures for Generalizing about Forecasting Methods: Empirical Comparisons," International Journal of Forecasting, Vol. 8, No. 1 (1992), pp. 69-80.

Ascher, William, "Beyond Accuracy," International Journal of Forecasting, Vol. 5, No. 4 (1989), pp. 469-484.

Box, G.E.P., and G. Jenkins, Time Series Analysis: Forecasting and Control, Revised Edition, San Francisco: Holden-Day, 1976.

Bretschneider, S. and W. Gorr, "Economic, Organizational, and Political Influences on Biases in Forecasting State Sales Tax Receipts," International Journal of Forecasting, Vol. 7, No 4 (1991a), pp. 457-466.

Bretschneider, S. and W. Gorr, "Forecasting Practices in the Federal Government," Federal Forecasters Conference 1991: Papers and Proceedings, (1991b), pp 17-24.

Bretschneider, S. and W. Gorr, "State and Local Government Revenue Forecasting," in Makridakis, S. and Wheelwright, S. eds., The Handbook of Forecasting, 2nd Edition, New York: John Wiley and Sons, 1987.

DeRoeck, R., "Editorial: Is there a gap between forecasting theory and practice?", International Journal of Forecasting, Vol. 7, No. 1 (1991) pp. 1-2.

Fildes, R., "The Evaluation of Extrapolative Forecasting Methods," International Journal of Forecasting, Vo. 8, No. 1 (1992), pp. 81-98.

Fildes, R., "Evaluation of Aggregate and Individual Forecast Method Selection Rules, Management Science, Vol. 35, No. 9 (1989), pp. 1056-1065.

Gardner, Everette S., and E. McKenzie, "Forecasting Trends in Time Series," Management Science, Vol. 31, No 10 (1985), pp. 1237-1246.

Gardner, Everette S., "Evaluating Forecast Performance in An Inventory Control System," Management Science, Vol. 36, No. 4 (1990), pp. 490-499.

References continued

Mahmoud, Essam, et. al., "Bridging the Gap between Theory and Practice in Forecasting", International Journal of Forecasting, Vol. 8, No 2. (1992) forthcoming.

Makridakis, S., "Sliding Simulation: A New Approach to Time Series Forecasting," Management Science, Vol. 36, No. 4, (1990), pp. 505-512.

_________, and M. Hibon, "Confidence Intervals: An Empirical Investigation of the Series in the M-Competition," International Journal of Forecasting, Vol. 3, No. 3/4 (1987), pp. 489-508.

_________, and R. Winkler, "Averages of Forecasts: Some Empirical Results," Management Science, Vol. 29, No. 9 (1983), pp. 987-996.

_________, and M. Hibon, "Accuracy of Forecasting: An Empirical Investigation," Journal of the Royal Statistical Society, Series A, Vol. 142, part 2, (1979), pp. 97-145.

_________, A. Andersen, R. Carbone, R. Fildes, R. Lewandowski, J. Newton, E. Parzen, and R. Winkler, "The Accuracy of Extrapolation (Time Series) Methods, Journal of Forecasting, Vol. 1 No. 2 (1982), pp. 111-154.

Stekler, H.O., "Macroeconomic Forecast Evaluation Techniques," International Journal of Forecasting, Vol. 7, No 3 (1991), pp. 375-384

Table 1
Intermediate Measures of Forecast Outcomes

Explainability
- Importance of defendablity of method to others
- Importance of reasonableness of assumptions
- Importance of using sensible and accurate data
- Importance of input data and assumptions being linked to output forecasts

Technical Forecast Accuracy
- Importance of systematic bias (e.g. mean error)
- Importance of accuracy (e.g mean square error)
- Importance of existence of serial correlation
- Importance of sophistication of technique (State-of-the-art)

Comparative Criteria
- Importance of simple benchmarking
- Importance of multiple benchmarking

Managing Uncertainty
- Importance of the ability to present low, middle and high forecasts
- Importance of the ability to evaluate scenarios
- Importance of the ability to identify outliers
- Importance of the ability to identify structural change

Manipulability of Forecasts
- Importance of the ability to aggregate and disaggregate forecasts
- Importance of the ability to adjust inputs
- Importance of the ability to adjust forecasts

Cost Criteria
- Importance of manpower costs
- Importance of computer costs

Coordination Criteria (intra- and inter-agency)
- Importance of use of standard data
- Importance of use of standard method
- Importance of coordination and timing data collection and forecasting
- Importance of standard forecast reporting

Utility Criteria
- Importance of multiple uses for forecasts
- Importance of policy alternative evaluation

Table 2
Descriptive Statistics for Importance Measures

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
CREXDEF	Importance of defendable methods	93	5.5	1.3	1.0	7.0
CREXREAS	Importance of reasonable assumptions	92	5.7	1.1	3.0	7.0
CREXSENS	Importance of sensible & accurate data	93	5.6	1.1	3.0	7.0
CREXLINK	Imp. of output forecasts link to data	91	5.5	1.2	2.0	7.0
CRACBIAS	Importance of systematic bias (Mean Err)	86	4.1	1.5	1.0	7.0
CRACMSE	Importance of accuracy (Mean Squared Er)	86	4.4	1.7	1.0	7.0
CRACCORR	Importance of ex. of serial correlation	85	3.3	1.6	1.0	7.0
CRACSOPH	Imp. of sophistication of technique	87	3.3	1.6	1.0	7.0
CRCMBNCH	Importance of simple benchmarking	90	3.3	1.5	1.0	7.0
CRCMMLBN	Importance of multiple benchmarking	89	3.0	1.7	1.0	7.0
CRUNRNGE	Imp. of ability to present range of vals	93	4.1	2.0	1.0	7.0
CRUNEVAL	Imp. of ability to evaluate scenarios	93	4.8	1.8	1.0	7.0
CRUNOUTL	Imp. of ability to identify outliers	92	3.9	1.8	1.0	7.0
CRUNSCHG	Imp. of ability to id. structural change	92	4.5	1.6	1.0	7.0
CRMNAGG	Imp. of ability to agg & disagg f'casts	93	4.4	1.7	1.0	7.0
CRMNINP	Imp. of ability to adjust inputs	92	4.4	1.7	1.0	7.0
CRMNFOR	Imp. of ability to adjust forecasts	92	4.8	1.6	1.0	7.0
CRCTMANP	Importance of manpower costs	94	3.7	1.7	1.0	7.0
CRCTCOMP	Importance of computer costs	93	3.0	1.7	1.0	7.0
CRCODATA	Importance of use of standard data	94	4.4	1.6	1.0	7.0
CRCOMETH	Importance of use of standard methods	93	3.8	1.5	1.0	7.0
CRCOTIME	Imp. of coordination of timing	94	4.2	1.8	1.0	7.0
CRCOREPT	Imp. of standard forecast reporting use	93	3.7	1.7	1.0	7.0
CRUTMULT	Imp. of multiple uses for forecasts	94	4.4	1.5	1.0	7.0
CRUTALT	Imp. of policy alternative evaluation	93	4.7	1.6	1.0	7.0

Table 3
Varimax Rotated Factor Weights
(Summary Statistics)

Variance explained by each factor

FACTOR1	FACTOR2	FACTOR3	FACTOR4	FACTOR5
3.637725	2.769085	2.380141	2.102308	1.672691

Final Communality Estimates: Total = 12.561950

CREXDEF	CREXREAS	CREXSENS	CREXLINK	CRACBIAS	CRACMSE	CRACCORR
0.602200	0.716368	0.591863	0.572935	0.529240	0.661999	0.465865
CRACSOPH	CRCMBNCH	CRCMMLBN	CRUNRNGE	CRUNEVAL	CRUNOUTL	CRUNSCHG
0.474959	0.394791	0.458660	0.417039	0.561131	0.567028	0.423178
CRMNAGG	CRMNINP	CRMNFOR	CRCTMANP	CRCTCOMP	CRCODATA	CRCOMETH
0.460218	0.550531	0.518570	0.364248	0.291986	0.559933	0.545691
CRCOTIME	CRCOREPT	CRUTMULT	CRUTALT			
0.519479	0.587679	0.427704	0.298653			

Table 3
Varimax Rotated Factor Weights

	FACTOR1	FACTOR2	FACTOR3	FACTOR4	FACTOR5
CREXDEF	0.06530	0.71483*	0.09920	0.24855	0.12385
CREXREAS	0.04082	0.76102*	-0.01897	0.14932	0.33599
CREXSENS	0.20980	0.70067*	0.13045	0.15421	0.12690
CREXLINK	0.16572	0.70808*	0.18966	0.04072	0.08040
CRACBIAS	0.23657	0.14583	0.06015	0.65661*	-0.13136
CRACMSE	-0.03383	0.18866	-0.00898	0.74856*	0.25462
CRACCORR	0.03059	0.11303	0.27999	0.59627*	0.13498
CRACSOPH	0.32199	0.16333	0.19389	0.49287*	0.25318
CRCMBNCH	0.57767*	-0.00002	0.10499	0.22101	-0.03494
CRCMMLBN	0.60450*	0.09353	0.11493	0.26348	0.04308
CRUNRNGE	0.57036*	0.25594	0.10205	-0.12561	0.00525
CRUNEVAL	0.57352*	0.27722	0.21161	-0.32009	0.09014
CRUNOUTL	0.54206*	0.33187	0.36894	0.04839	-0.15684
CRUNSCHG	0.46117*	0.36819	0.26606	-0.00372	-0.06436
CRMNAGG	0.54458*	0.19536	0.08402	-0.04545	0.34111
CRMNINP	0.66643*	0.12005	0.07126	-0.03510	0.29271
CRMNFOR	0.26932	0.20678	0.15497	0.07976	0.61066*
CRCTMANP	0.52082*	-0.05973	0.13188	0.22123	0.15196
CRCTCOMP	0.48871*	-0.00757	0.12351	0.16963	0.09521
CRCODATA	0.14829	0.17813	0.67311*	0.06763	0.22037
CRCOMETH	0.07622	0.16761	0.64855*	0.16684	0.25166
CRCOTIME	0.25664	0.07110	0.62973*	0.11144	0.19896
CRCOREPT	0.33253	0.02895	0.67929*	0.07167	0.09845
CRUTMULT	-0.01028	0.09985	0.26563	0.14431	0.57118*
CRUTALT	0.10517	0.13734	0.19972	0.07641	0.47223*

An Evaluation of BLS Aggregate and Industry Employment Projections to 1990

Norman C. Saunders, Office of Employment Projections, Bureau of Labor Statistics

The BLS regularly prepares projections of alternate future growth paths of aggregate economic activity and the employment by industry generated by that aggregate projection of the economy.[1] These projections form the basis for occupational employment projections which underlie occupational outlook information prepared by BLS for use in career guidance and education planning. They also stand in their own right as projections used by other Federal government agencies, State Employment Security agencies, and firms in the private sector. Because of the wide use made of the projections, either directly or indirectly, BLS regularly evaluates those projections as historical data for the projected years become available. This is the final stage of the projection process, effectively closing the books on a given year's projections. The evaluation allows BLS to identify both strengths and weaknesses in the process of preparing the projections and users to gain insights into the accuracy of the projections. This article examines projections of 1990 economic activity and employment, in total and by industry, and is part of a continuing effort to improve current projections work by carefully evaluating past projections.[2] The article also compares the results of the 1990 projections with those done for earlier years.

Over the period 1978 to 1985, the BLS published four sets of aggregate economic and industry employment projections for the year 1990.[3] In each of these sets of projections, one--usually the moderate--is used as the basic alternative and the others are variations on this basic set. In part for ease of presentation, the evaluation at the detailed level will concentrate on this basic alternative in each projection set. At the aggregate level, however, all alternatives will be shown.

For the first of the four sets of 1990 projections, two alternative growth projections were developed. For the last three projections, three alternatives were developed for each set. For the most part, the alternatives were aggregate in nature, i.e. aggregate economic assumptions were varied to arrive at a range of GNP and employment projections and detailed industry relationships derived for the basic projection were then applied more or less unchanged to the aggregate alternatives.

In terms of levels, GNP projections improved across the entire set of projections, while estimates of employment as a count of jobs generally worsened as they got closer to 1990. However, the projected distribution and growth ranking of employment by industry improved with each set.

Publication Year	Alternative	GNP (billions of 1982 $)	Percent Error	Total Employment (millions)	Percent Error
1978	Base*	4543.7	9.3	118.6	-3.2
	High	4723.7	13.6	125.6	2.5
1981	Low	4091.2	-1.6	122.0	-0.5
	High-I	4672.3	12.4	130.7	6.6
	High-II*	4670.5	12.3	124.0	1.1
1983	Low	3995.5	-3.9	116.9	-4.6
	Moderate*	4119.4	-0.9	118.3	-3.5
	High	4310.1	3.7	119.4	-2.6
1985	Low	3999.1	-3.8	112.8	-8.0
	Moderate*	4192.3	0.8	116.9	-4.7
	High	4437.9	6.7	119.0	-2.9
1990	Actual	4157.3		122.6	

* Basic alternative

GNP was overestimated by an average, across the two projections, of 11.5 percent in the 1978 publication and by an average of almost 8 percent in the three 1981 estimates. In both cases, the relatively large error was due primarily to overly optimistic assumptions regarding the potential for labor productivity growth over the decade of the 1980's. The 1983 projection of GNP was in error by an average of only -0.4 percent while the 1985 projection was off by an average of 1.2 percent. The major source of error in the 1985 projection of GNP was an underestimate of the civilian labor force.

Total employment, on the other hand, was in error in the 1978 projection by only -0.4 percent, on average. This error grew in absolute terms in each subsequent release, reaching -5.2 percent on average in the 1985 projection. The primary source of error in the employment projections appears to be related to errors in the potential labor productivity assumption and errors in accounting for the conceptual difference between the household employment series (a count of persons) and the establishment employment series (a count of jobs).[4]

This article will examine in more detail the level and sources of error in the aggregate projections--GNP, the demand composition of GNP, and the aggregate level of employment--and at the industry level of detail. Careful attention will be paid at the industry level to three types of errors: level errors, distribution errors, and errors in growth rate ranking. Two other articles in this evaluation series are presented elsewhere in this issue of the Review, one examining the detailed projections of the labor force to 1990 and the other aimed at explaining the errors in the projections of occupational employment.

Framework for the Evaluation

Each of the projections was based on a specific macroeconomic model being run under a specific set of assumptions and targets. The following section describes the major assumptions and results underlying each of the published alternative projections.

1978/1979 Publication
The projections published in 1978/1979 were carried out at an aggregate level in the context of the Thurow econometric model, a small-scale model of demand activity maintained and periodically re-estimated within the Office of Employment Projections. The projections were carried out in 1972 prices and industry level estimates were based on a 1972 Standard Industrial Classification (SIC) definition. Two aggregate economic alternatives were developed in this round of projections:

Base: Prepared in 1977, this alternative assumed a smoothly- growing economy characterized by moderate inflation and labor force growth, a declining unemployment rate, a strong comeback for labor productivity growth, and a Federal government generally becoming less important over the entire decade of the 1980's. The major impediment to growth anticipated in this set of projections was the high cost of imported oil and the subsequent impacts on the production process necessitated by material input substitution of relatively less expensive inputs for petroleum- based products. Budget deficit problems from the early 1970's appeared to be well under control and no adverse economic impacts were anticipated from this quarter. Neither were any major problems anticipated with the exchange rate for the U.S. dollar. As a result, foreign trade was expected to continue in a "business as usual" mode, with nominal balance of trade figures near zero and moderate real trade surpluses. The five industries projected to have the fastest job growth rates over the decade were other medical services, other transportation equipment, miscellaneous business services, synthetic fibers, and computers and peripheral equipment. The five industries with the largest absolute increase were projected to be retail trade, State and local government other than education, miscellaneous business services, other medical services, and hospitals.

High employment: This alternative assumed higher labor force participation rates than in the base projection, much faster growth in Federal grants-in-aid to state and local governments, and consequently faster employment growth due primarily to more intense spending at the state and local government level.

1981 Publication
Very few changes in techniques or underlying data definitions took place in this round of projections. What did change was that the economy was three years closer to 1990, a new Administration was in the White House, and the economy was recovering from a relatively mild recession in 1980 and heading into another more serious one in 1982 (which, incidentally, was not anticipated in any of the alternatives for this set of projections). Three aggregate alternatives were prepared for this round of projections:

Low: The low alternative was characterized by assumptions of continued high inflation, low productivity growth, and only moderate expansion in real GNP. The assumption of Federal expenditures accounting for progressively smaller shares of GNP over the decade was still in place and several significant personal tax cuts were assumed for the period, as well. The five industries projected to exhibit the fastest rates of employment growth to 1990 were other medical services, typewriters and other office equipment, computers and peripheral equipment, coal mining, and hospitals. Largest job gains were projected for eating and drinking places, retail trade except eating and drinking places, hospitals, miscellaneous business services, and other medical services.

High-I: The first high alternative assumed marked improvements in both inflation and labor productivity growth, greater labor force growth, sharp reductions in the unemployment rate from the 1981 recession peak, and higher production levels.

High-II: Finally, the second high alternative embodied labor force growth consistent with that assumed in the low alternative, but much more marked improvements in labor productivity and inflation, leading to GNP levels commensurate with that in the High-I alternative and an employment level similar to that in the low projection.

1983 Publication

By the time the next round of projections to 1990 was published, several major changes in the projection procedures had taken place. Prior to the 1981 publication date, all of the projections efforts of the BLS had been merged, offering new challenges in terms of the preparation timing of the projections but also offering the prospect for a more broadly-based internal review of results. The more comprehensive review process instituted as a result of this merger, however, was not fully in place until the 1983 projection was being prepared. The forecast horizon, though, had been moved forward by 5 years, to 1995. Although estimates for 1990 were still published, they were not subject to the same critical review that was true for the 1995 data. The aggregate projections were prepared in the context of the Chase Econometric Model, a detailed quarterly model of economic activity. Most important for the projections, though, was the fact that it was becoming increasingly apparent that significant slowdowns were occurring in the manufacturing sector, that Federal deficits were growing at a rate not anticipated in prior projections, and that our balance of trade was deteriorating at a rate not experienced in the post-World War II economy. Three alternative aggregate projections were developed for the 1983 projections:

Moderate: General fiscal restraint was assumed in this projection as both defense and nondefense federal spending were assumed to grow very slowly or to decline in real terms over the latter half of the decade, which, when combined with tax policy assumptions, led to a steadily declining Federal budget deficit. Productivity growth was assumed to return to near post-World War II highs, and recovering markets for exports were expected to remove the growing trade deficits of the early 1980's. The result was moderate-to-strong growth in production and good growth across the board in employment.

Low: This scenario assumed higher budget deficits over the entire decade and a generally more sluggish economy as durable goods purchases were affected by higher interest rates. Slower income growth led to lowered import levels and declining employment in manufacturing.

High: The high scenario assumed a less restrictive monetary authority, leading to more robust GNP growth, albeit accompanied by higher inflation. Since the Federal spending assumptions were virtually unchanged from the moderate growth alternative but incomes were higher, the impact on the deficit was to reduce it to zero near the end of the decade. Improved economic conditions led to higher rates of investment, more competitive domestic industries, and consequent improvements in the balance of trade.

1985 Publication

The projections published by BLS in 1985 and which also focused on 1995 were the last look taken at 1990. The economy had climbed out of the 1982 recession by this time but was troubled on two fronts, the Federal budget deficit and the balance of trade with foreign countries. Investment growth had been very sluggish over the first half of the decade and labor productivity growth had been relatively dismal, as well. The aggregate projections were developed using the Wharton Long-Term Annual Model and industry-level data were now being developed in the context of the newly-released 1977 input-output table for the U.S. economy. As with the 1983 publication, three scenarios were developed in 1985:

Moderate: Sharp increases in real defense spending were coupled with moderate increases on the nondefense side and relatively stable tax rates, leading to high Federal deficits throughout the projection period. Anticipated improvements in the late 1980's in the value of the dollar led to some resumption in export growth but not enough to cure the trade deficits of the early 1980's. As a consequence, good growth in GNP was predicted, more than offsetting slowing labor force growth, due primarily to very optimistic assumptions regarding labor productivity potential.

Low: Lower labor force growth combined with lower savings rates and lower investment spending lead to GNP growth that is sharply lower, with more sluggish productivity behavior. Deficits, both Federal budget and balance of trade, are worse, and employment growth well below the prior ten year period is projected.

High: Stronger labor force growth accompanied by a generally more robust economy leads to higher employment and production levels, offset somewhat by higher productivity, due to relatively strong investment spending. Growth well above the prior ten-year historical period is predicted for both GNP and employment.

Projections Methodology

A short note is in order here regarding the methods BLS uses to prepare projections. At an aggregate level, a macroeconomic model is driven by assumptions regarding certain key variables in the areas of fiscal and monetary policy, demographics, foreign economic conditions, and energy. Perhaps more important than these exogenous assumptions, however, are variables which are determined by the model in use but for which certain acceptable ranges have been reviewed in detail and determined in advance by BLS analysts to be within an acceptable range. These variables include the civilian labor force, the unemployment rate, the rate of growth of labor productivity, and, in a very rough way, the resulting rate of growth of GNP. It is these target variables which are generally subject to the greatest amount of uncertainty when long-run growth prospects are examined, and which are, therefore, the variables most subject to wide variations in the process of generating alternative economic projections and the variables we will focus on in this evaluation of 1990 projections at the aggregate level of detail. The projected values for these key target variables are presented along with actual 1990 values and their percent errors in table 1. Also important to later stages of the projections process is the predicted distribution of GNP by major demand categories, and these results are presented in table 2.

Once a level and distribution of real GNP, and an associated aggregate employment level are settled on, the analysis then turns to the industry level of detail. BLS first distributes the various demand categories of GNP more finely (personal consumption spending, for example, is disaggregated to 82 spending types such as autos or banking services). Then, each category of GNP is broken into its commodity content. The resulting disaggregation of GNP by detailed commodity is then used to drive a projected interindustry goods and services total requirements matrix, the result being the total output required from each detailed industry in order to produce the projected GNP. This step is taken because certain commodities are generally not sold directly to final consumers (steel, for example) but rather, are embodied in other products which are sold to consumers (automobiles, for example). It would thus be very difficult to determine employment requirements in these so- called intermediate industries if the analysis were based only on sales to final consumers.

The third stage of the projections process is to determine for each industry in the economy the level of employment necessary to produce that industry's output. This determination is made in the context of a detailed econometric model of labor productivity by industry and average annual hours by industry. Given industry output from the prior stage of the projections, the productivity estimates (output per hour) are translated to total hours paid by industry. Dividing this figure by average annual hours yields an estimate of the number of jobs in each industry. Employment projections for major sectors of the economy in 1990 are presented in table 4. Average errors for industry-level projections are presented in table 6.

Initially the projections are developed from the top down, from aggregate control values for production and employment to disaggregated industry results. The final stage of the process is for all analysts involved in the projection estimation to review critically all phases of the work for consistency and meaningfulness. This aspect of the work really came into its own during the preparation of the 1983 and later projections. Prior to this the review was more limited in scope and generally did not involve analysts from all phases of the process. With the more intensive review process, all analysts were involved, from those preparing detailed labor force projections to those working on very specific projections of occupational employment. The review process can have ramifications back to all levels of the projections and, as will be demonstrated later, appears to have markedly improved the accuracy of detailed BLS estimates since its inception with the 1983 set of projections.

Aggregate Results

As noted earlier, for each set of projections published in a given year, one of the set was considered the basic alternative. For the four basic alternative projections of 1990 real GNP, the worst was the 1978 estimate. From that point, the projections of GNP steadily improved, reaching an 0.3 percent error by 1985, an error of less than $13 billion. Employment projections improved from 1978 to 1981 and then diverged sharply from realized employment in 1983 and 1985, reaching a 4.7 percent error in the last projection year and underestimating actual 1990 employment by almost 6 million jobs.

Where do these errors come from? It is possible to express the GNP estimate as an identity based on the civilian labor force, the civilian unemployment rate, and the level of real GNP per employee, a proxy for labor productivity. All of these factors are key target variables, as noted earlier, in the process of preparing and evaluating the aggregate projections. Table 3 presents a factoring of the error in real 1990 GNP into its shares attributable to errors in each of the three target variables noted above. In examining the factor shares across the four basic alternative projections, it becomes immediately apparent that the largest source of error in the GNP projection comes from errors in assumed rates of growth of labor productivity. With only a few exceptions, in fact, labor productivity has been over-estimated in almost all of the alternatives developed in each of the four publication years.

Errors attributable to a mis-specified unemployment rate have a relatively small effect on projected GNP. The civilian labor force, because of its relatively large weight in the GNP identity, will lead to relatively large percentage errors in GNP, even if the labor force itself is in error by only a small amount. Nonetheless, in every case, the major error contributing to mis-estimated GNP is a mis-specification of labor productivity.

The same type of error factoring can also be carried out for the estimate of establishment employment. In this case, the productivity component of the identity is replaced with the conceptual difference, that variable which embodies the statistical and conceptual differences between employment as a count of persons (the household series) and employment as a count of jobs (the establishment, or payroll, series). As noted above, while the GNP estimate for 1990 was improving from one projection to the next, the employment estimate was getting worse--the error increased from a 4-million underestimate in 1978 to a 5.7-million job undercount in 1985. In each of the basic alternative projections of establishment employment, the largest share of the error was accounted for by errors in estimating the conceptual difference. Because the macroeconomic model is primarily oriented to the derivation of household employment necessary to produce aggregate GNP, the conversion factor between the two concepts of employment is actually derived from the adding-up of industry-level estimates of employment. The main source of error in establishment employment was the failure to anticipate trends at industry and sector levels of detail. This will be explored more fully in the next section of the article.

Structure of Demand

The last area which should be examined before moving along to the industry detail projections is the demand structure of GNP. Where did BLS expect growth in expenditures and where did BLS err in those expectations?

Personal Consumption: Over the decade of the 1980's, personal consumption expenditures (PCE) increased their share of GNP from 62 percent in 1978 to a high of 65.8 percent in 1986. The major reason for this share increase was a long-term decline in the personal savings rate and a consumer spending splurge, particularly on new and expanded services. These increases were possible because of tax cuts over the period and a lowering in the amount of income allocated to savings. The 1978 projections of PCE actually anticipated the surge in consumer spending but overstated the share runups (see table 2). The 1981 projection returned the PCE share of GNP to levels consistent with the early 1970's and consequently missed on the low side. Both the 1983 and the 1985 projections were quite close to what actually did transpire in 1990.

What none of the four projections did well was to anticipate the big drop in the share of GNP accounted for by purchases of nondurable goods, primarily food, clothing, fuels, and pharmaceuticals. Virtually all growth in real disposable income over the decade of the 1980's flowed into spending for new or expanded financial and medical services, a development not at all anticipated in the early BLS projections for 1990 and only partially accounted for in the later projections. This represents a major structural change in the way consumers allocated their income.

Investment: Fixed investment--purchases of plant and equipment by businesses and the construction of new residential dwellings-- makes up the most volatile component of GNP. Projections of 1990 business spending on equipment were low in 1978, high in 1981, low again in 1983, and, finally, very close to the right answer in 1985. At the same time, projections of residential construction were too high over the first three sets of projections, due to a failure to predict the slower growth of average home values over the decade, and, again, right on the money in the 1985 projection. The category of investment that the BLS failed most consistently in predicting was nonresidential construction--construction of new office buildings, commercial structures, warehouses, and the like. Expectations for growth in this category of GNP were, without exception, well above actual behavior of construction investment over the decade of the 1980's. BLS analysts failed to appreciate the amount of overbuilding that took place in the late 1970's and early 1980's, exacerbated by slowdowns in export growth. The resulting oversupply of business structures is still today in the process of being worked off. The failure of BLS analysts to anticipate that manufacturing would become a declining sector of the economy over the 1980's was the primary cause of the error in projecting investment needs.

Foreign Trade: Exports of goods and services underwent a roller- coaster ride in the 1980's. Two decades of steady export growth peaked in 1980 with exports accounting for 12.2 percent of GNP. During the first half of the decade of the 1980's, the exchange rate soared, peaking in 1985, as a result of generally deteriorating business conditions in the U.S. The steadily growing Federal budget deficit, combined with lower savings and investment rates and the movement of many heretofore domestic manufacturing operations to offshore locations, generated the exchange rate surge which led to cheaper imports and more expensive exports.

By 1985, the export share of GNP had dropped to 10.1 percent while imports of goods and services had increased their share from 10.4 percent in 1980 to 13.0 percent. Over the remainder of the decade, imports have continued to account for increasing shares of GNP, reaching a 16.0 percent share by 1990. A fall in the exchange rate and a push on the part of many U.S. manufacturing firms to regain lost markets and to become competitive in emerging new markets led to a resurgence in export growth. They accounted for 15.2 percent of GNP by 1990, up sharply from the 10.1 percent low point in 1985.

How did the BLS projections fare against this turbulent backdrop? Imports were, without exception, underestimated in all four sets of 1990 projections. The BLS failed to anticipate the exchange rate anomalies of the 1980's and the subsequent impacts on relative prices. The 1981 and the 1985 projections were the best of the four, underestimating imports by less than 4 percent in each case.

Export projections were more interesting. In 1978, exports for 1990 were underestimated an average of 31 percent. The BLS clearly had missed the fact that foreign trade was fated to become an increasingly important factor in U.S. economic growth. Realizing that the earlier projections had been unrealistic, the BLS raised the projected growth of exports considerably in the 1981 projection, the closest we were to get to predicting 1990 exports with any accuracy in this entire set of projections. The 1981 projection underestimated actual exports by only about 3 percent on average. Following the 1981 projection, the big exchange rate runup was becoming very apparent. Over- compensation for this factor led to an average underestimation of exports in 1983 and 1985 of 22 percent. Clearly the models in use by the BLS and the understanding of BLS analysts were not adequate to capture foreign trade behavior over the confusing decade of the 1980's.

Government: Spending on national defense grew from a low of 5.4 percent of GNP in 1977 to a peak of almost 7 percent in 1987. By 1990, the share had fallen back to 6.2 percent of GNP. Clearly BLS analysts did not foresee such a spending boost in 1978. Rather, they assumed that defense spending would continue to decline in importance over the decade, reaching a 4.2 percent share of GNP by 1990. By 1981, President Reagan was in the White House and had publicized his plan to bolster the military preparedness of this country. BLS projections of defense spending published in that year were somewhat higher--5.5 percent of GNP, on average, but still a small real decline from the late 1970's. By 1985, at last understanding the intent of the Reagan Administration but not yet anticipating the adverse reactions of many U.S. citizens to the surging budget deficit or the disappearance of the Soviet Union as a world military power, BLS assumed strong growth in defense spending over the remainder of the decade, in fact overstating actual 1990 defense purchases by almost $16 billion.

At the same time that the BLS was at first understating and then overstating defense growth, assumptions regarding growth in Federal nondefense spending were relatively accurate, within 3 percent of the actual spending level in every case. In the area of State and local government purchases of goods and services, BLS estimates were generally quite close to actual 1990 spending levels with the exception of the projections published in 1983. Reports at that time of major taxpayer revolts at the state level were taken quite seriously by the BLS and led to underestimates of actual growth in State and local government spending of about 10 percent.

In summary, some demand categories proved more tractable than others when it came to applying past trends to future expectations. Where the BLS projections process fell into the worst errors were where, using perfect hindsight, major changes in the underlying economic relationships were taking place, areas where no econometric approach to forecasting does well. At an aggregate level, GNP projections improved noticeably the closer we drew to our target year, attaining an error of only 0.3 percent in the 1985 moderate growth projection of real GNP in 1990.

Industry Results

Level Errors. At the industry level of detail there are three possible approaches to assessing a given set of projections--level errors, distributional errors, and rank errors. The first of these is generally a result of mis-specified control levels. That is, errors at the aggregate level of detail will inevitably be carried over to the industry level. Because of the way aggregate employment was determined in the BLS approach to the 1990 projections, however, there was also an opportunity for analyst error at the industry level to affect the accuracy of the projections. The aggregate model, as noted earlier, determined the level of household employment, a count of persons. The industry employment projections, on the other hand, are developed from the establishment concept of employment, a count of jobs by industry. In order to ensure consistency between the two approaches to employment measurement, the conceptual difference between the two series is computed and evaluated over the projection period for consistency with past trends.

The jobs-to-persons conceptual difference was a relatively well-behaved data series during the 1950's and 1960's. The ratio declined sharply between 1970 and 1975 and then proceeded to behave quite erratically for the next ten years. Beginning in 1987, the series appears to be returning to its average level of the 1960's. The 1978 conceptual difference projection was exactly on target but the situation degenerated in later projections.

The 1981 estimate was much too high and led to a further exaggeration of the employment overestimate already flowing from the aggregate projections model. Since the conceptual difference ratio is not explicitly estimated until the industry job estimates are completed, it provides an index of the amount of level error brought to the industry projections by the analyst preparing those projections.

Since the industry employment estimates were overstated in the 1981 projection, it is useful to examine the percent errors of those projections to determine what sectors or industries were the most seriously mis-perceived at that time. Table 4 presents the projected employments by major sector for all of the alternatives, along with sector percent errors. Table 5 presents, at the industry level of detail, the average error across all alternatives published in a single year. The largest overstatement error in the 1981 projections occurred, not surprisingly, in the area of durable manufacturing, but was also accompanied by serious over-estimates of mining and construction employment. These overestimates were offset by large underestimates of services employment, primarily in the miscellaneous business services industry, a failure to anticipate the contracting-out phenomenon of the 1980's.

In both the 1983 and 1985 projections of 1990 industry employment, overall employment levels were underestimated, by an average of 4.4 million jobs and 6.3 million jobs, respectively. Almost the entire error in the 1983 projection was due to industry bias, as reflected in the conceptual difference ratio. Employment in transportation, utilities, trade, services, and government were all understated, while the overestimates in mining and manufacturing were not nearly large enough to offset the low estimates. The largest errors occurred, again, in miscellaneous business services and in general government employment. These two industries accounted for nearly three-fourths of the undercount in the 1983 projection.

In the 1985 projection, about 40 percent of the level error was accounted for by industry bias, the remainder being due to errors at the aggregate level of detail. As with the 1983 projections, the major errors occurred in services and in government.

Industry Distribution Errors. Once the level errors have been determined, it becomes interesting to examine the extent to which BLS' industry analysts were able to anticipate correctly the distribution of employment by industry. Two measures of goodness-of-fit have been used to evaluate the various projections to 1990, the index of dissimilarity, and Theil's information statistic.[5] Both are presented in table 6.

The index of dissimilarity is a measure of average absolute errors between the actual industry percent distribution of employment and the projected percent distribution. A perfect estimate would yield an index value of zero. Unfortunately, tests of the statistical significance of this index do not exist, but it is instructive to note that for each successive set of projections, the index of dissimilarity is decreasing in size, implying that BLS projections improved steadily as we approached the target year.

In a like manner, the information statistic measures the amount of variation between two sets of percent distributions. Here we are looking at a projected percent distribution of employment and attempting to quantify the amount of extra information imparted by gaining access to the actual percent distribution. If no new information is gained, that is, if the projected distribution is identical with the actual distribution, then the value of the information statistic will also be zero. A quick examination of the information statistic values across the projections of 1990 employment distributions confirms the story told by the index of dissimilarity--the projections get progressively more accurate the closer we get to the target year. The Theil statistic has the added benefit of being normally distributed when it is computed with the large number of degrees of freedom involved in assessing the all-industry percent distribution. Knowing this, we can formulate and test hypotheses regarding the value of the statistic. In this case we wish to know if the information statistic is significantly different from zero or, alternatively, if we are unable to reject the null hypothesis, namely, that the statistic is equal to zero. In the first three sets of projections, those published in 1979, 1981, and 1983, we find we must accept that a significant distributional difference existed between actual and projected data. The 1985 projection, on the other hand, has an information statistic that is insignificantly different from zero. In short, we projected the distribution of employment imperfectly, but in an increasingly perfect way over the four projections.

Ranking Errors. The third and final way of evaluating a set of projections is to compare projected growth rankings to those which actually occurred. This is especially important in the context of the BLS industry employment projections because, traditionally, the descriptions of the projections, and the implied usefulness of those projections, has relied heavily on presentations of the fastest growing industries or industries with the largest changes in levels. In fact, one could reasonably argue that if we do a good job of projecting rankings, in spite of level and distributional errors, then we have accomplished one of our main purposes in preparing industry-level projections.

The industry employment projections are typically ranked ac-cording to growth rate and according to changes in the levels of employment, in both cases from the largest positive change to the largest negative change. The first type of ranking tends to capture the smaller, more volatile industries--generally industries which are newly emerging or those in accelerating decline. The second approach generally focuses attention on the bigger, more stable industries. Both types of rankings were computed for each of the alternative sets of projections and compared to actual 1977-90 growth rankings by means of a rank correlation coefficient[6] and are presented as the final two columns in table 6.

In both cases, the correlation between actual and projected growth rankings started at dismally low levels and improved markedly as we drew nearer to our target year. The rank correlations on growth rates, as might be expected, were much lower initially than were the corresponding coefficients on level changes. One anticipates that it would be easier to make predictions about large, solid industries than about small, volatile ones. What is most interesting is that projections of both types of rankings improve to almost the same level of accuracy--by 1985 BLS analysts were able to anticipate relative growth rates and relative level changes with an almost 88 percent accuracy in both cases.

The rank correlations presented in table 6 cover the entire spectrum of industries. Another way of assessing how BLS did at projecting rankings is presented in table 7. Here the twenty fastest growing industries and the twenty industries with the largest level changes are presented, based on actual data between 1977 and 1990. For each of these industries, their relative rank in one of each year's projections are also presented. How many of the top twenty industries actually ended up projected to be in the top 20?

Number projected in the top twenty:

	1979	1981	1983	1985
Fastest growing	6	9	12	14
Largest change	19	19	17	17

This implies an accuracy of only 30 percent in the 1979 top growth industries, increasing steadily to a 70 percent accuracy by 1985. The industries with the largest changes, on the other hand, were obviously easier to predict over the entire period, with accuracy rates in the 90 percent range.

In summary, BLS projections of industry employment evolved over time as new information on industries became available, as new Administrations began to have their impact on the economy, and as structural changes in the economy began to be felt. As one would hope, they evolved in a positive way, toward an increasingly accurate estimate of employment distribution and growth. Errors in levels were the single factor which did not show marked improvement over time, an indication, perhaps, of the extent to which BLS' methods and models are wedded to a rigorous examination and extrapolation of historical trends in the data.

Past BLS Projections

So far in this article we have examined only the errors of BLS projections to 1990. In order to judge the overall effectiveness of the evaluation program and to put this set of projections into the context of all of BLS' earlier projections, it is useful to compare errors in 1990 with those identified in earlier projections. BLS has now evaluated twelve employment projections. The error in the projection of total employment growth ranges from -0.8 percent in the 1985 projection of 1990 employment to 0.6 percent in the 1973 projection of 1975 employment (see table 8). The first two projections for 1990 employment, those published in 1979 and 1981, were very close to being the best projections, in total employment level terms, ever prepared by BLS. Conversely, the last two projections for 1990, those published in 1983 and 1985, were the worst ever published by BLS. The primary sources of the employment level error were noted earlier. However, we feel that it is not coincidental that for these last two projections, the primary year of interest, and thus the one receiving the most attention during the review process, was 1995. Data for 1990 were published but considerably less effort was given over to analyzing that particular year's results.

The average absolute errors for projected industry employment trends have ranged from 1.3 to 2.9 percentage points per year. The spread of error is slightly smaller when the errors are weighted for industry size, ranging from 1.0 to 2.1 percentage points per year. The 1980 projections prepared in 1970 were the most accurate while the 1975 projections published in 1973 were the least accurate. Projections for 1990 published in 1979 and 1981 were both in error by 1.4 percentage points, close to the middle of the error range of past BLS projections, while the two 1990 projections published in 1983 and 1985 were near the worst end of the error range, again due to factors noted previously.

Summary and Recommendations

What has been learned by this examination of BLS projections to 1990? Several major points are immediately apparent. First, perhaps not surprisingly, the projections improve at both an aggregate and industry level of detail as we move closer to our target year. For this reason, intermediate projection years should continue to be published, even when the terminal year is pushed forward by five years. Thus projections to 2005 should also include projections to 2000. The review process should be modified to include not only the terminal year but this earlier year, as well, and a careful review of the time path of the economy to the terminal year should be carried out.

Second, GNP estimates by the BLS are generally much more accurate than the aggregate employment projections related to that GNP. This is due primarily to a tendency on the part of BLS analysts to overestimate potential labor productivity growth. The implication is that future projections should very carefully assess the reasons underlying slower productivity growth and assign reasonable probabilities to the continuation of those factors in the future. To a certain extent this recommendation has already been taken in projections published by BLS since 1985.

Third, the estimates at the industry level of detail, in terms of level, distribution, and growth ranking, have improved noticeably as a direct result of the intensive review process instituted between the 1981 and 1983 projections. That this review process should be continued goes without saying. Care must be taken, however, on the part of the industry employment analyst to prevent the distortion of the jobs-to-persons conceptual difference as the review process proceeds. In other words, adequate feedback from the industry employment through industry industrial activity and to the aggregate level of detail must be carefully provided for. Again, this factor has been better taken into account in later BLS projections.

Finally, it is obvious from the anomalies of the 1980's that the BLS does not do well at predicting major structural shifts in economic relationships, such as the decline of the manufacturing sector and the large shifts in foreign trade balances over the 1980's. To the greatest extent possible, the models used by the BLS rely on past data and the trends implicit in that past data. Since structural change implies abnegation of past trends, the methodology of the projections would tend to imply that there is not much to be done about this last problem noted with the projections. However, alternative projections are now prepared surrounding the basic alternative. Their purpose is to address those areas of the projections which analysts feel are most open to question regarding future adherence to historical trends. A careful examination of past projection errors may well suggest further alternatives which should be explored in any given round of projections estimates.

The final conclusion reached is that while BLS projections do contain serious errors, their usefulness and accuracy have been markedly improved over time and continue to provide a valuable information resource about possible future courses for the U.S. economy as a whole and for employment at the industry level of detail in particular.

Footnotes

1. The latest projections published by BLS are entitled "OUTLOOK: 1990-2005," and appeared as a series of articles in the November 1991 Monthly Labor Review. Such medium-term projections are prepared on a two-year cycle for detailed labor force, for aggregate economic activity, for industry-level output and employment, and for detailed occupational demand by industry. A full statement of the methodology underlying BLS projections is included in OUTLOOK: 1990-2005, BLS Bulletin 2402, April 1992.

2. BLS has published evaluations of projections for the years 1970, 1975, 1980, and 1985. For the latest labor force and aggregate/industry evaluations see Howard N Fullerton, Jr., "An evaluation of labor force projections to 1985," Monthly Labor Review, November 1988, pp. 7-17, and John H. Tschetter,"An evaluation of BLS projections of the 1985 economy," Monthly Labor Review, September 1988, pp. 24-33. The last published evaluation of occupational projections was in Max L. Carey and Kevin Kasunic, "Evaluating the 1980 projections of occupational employment," Monthly Labor Review, July 1982, pp. 22-30.

3. Aggregate economic projections for 1990 were published in Norman C. Saunders, "The U.S. economy to 1990: two projections for growth," Monthly Labor Review, December 1978, pp. 36-46, in Norman C. Saunders, "The U.S. economy through 1990--an update," Monthly Labor Review, August 1981, pp. 18-27, in Arthur J. Andreassen et al, "Economic outlook for the 1990's: three scenarios for economic growth," Monthly Labor Review, November 1983, pp. 11-23, and in Betty W. Su, "The economic outlook to 1995: new assumptions and projections," Monthly Labor Review, November 1985, pp. 3-16. Industry employment projections were published in Valerie A. Personick, "Industry output and employment: BLS projections to 1990," Monthly Labor Review, April 1979, pp. 3-14, in Valerie A. Personick, "The outlook for industry output and employment through 1990," Monthly Labor Review, August 1981, pp. 28-41, in Valerie A. Personick, "The job outlook through 1995: industry output and employment," Monthly Labor Review, November 1983, pp. 24-36, and in Valerie A. Personick, "A second look at industry output and employment trends to 1995," Monthly Labor Review, November 1985, pp. 26-41.

4. The household survey employment concept (a count of persons) is that which underlies the aggregate projection of employment in the BLS projections process while the establishment, or payroll, survey data (a count of jobs) underlies the industry-level employment projections. In order to ensure consistency between the aggregate and industry employment projections, careful attention must be paid to the conceptual difference between the two employment series. The historical differences between the household and establishment employment surveys are examined by the BLS approximately once each decade. For an analysis of these differences for the 1960's, confer Gloria P. Green, "Comparing Employment Estimates from Household and Payroll Surveys," Monthly Labor Review, December 1969, pp. 9-20. The period of the 1970's was examined in Alexander Korns, "The Difference Between the Payroll and the Household Measures of Employment, 1975-79," Survey of Current Business, December 1979, pp. 44-49. The period of the 1980's is covered in Paul Flaim, "How many new jobs since 1982? Data from two surveys differ," Monthly Labor Review, August 1989, pp. 10-15.

Footnotes continued

5. The index of dissimilarity is defined as

$$D = \sum_i \frac{|A_i - P_i|}{2}$$

where, A_i is the actual employment in industry i as a percent of total actual employment and
P_i is the projected employment in industry i as a percent of total projected employment.

The information statistic is described in Henri Theil, Principles of Econometrics, John Wiley and Sons, 1971, pp. 641-644 and an application of the information statistic to the analysis of projection accuracy is contained in R.A. Kolb and H.O. Stekler, "The Information Content of Long-term Employment Forecasts," Applied Economics, forthcoming 1992.

6. A description of the rank correlation coefficient, developed by Spearman, may be found in G. W. Snedecor and W. G. Cochran, Statistical Methods, 6th ed., Iowa State University Press, 1967,pp. 193-195.

Table 1. Major Economic Variables, Actual and Projected, 1990

	Actual 1990	12/78 Base	8/81 High-II	11/83 Moderate	11/85 Moderate
Civilian labor force	124.787	119.367	122.375	124.951	122.653
Unemployment rate	5.5	4.5	4.5	6.0	6.3
Employment	117.924	113.996	116.868	117.454	114.926
GNP/Employee	35.25	39.40	39.29	34.74	36.29
GNP	4157.3	4491.9	4591.5	4080	4170.1
Employment (persons)	117.924	113.996	116.868	117.454	114.926
Employment (jobs)	122.571	118.615	123.960	118.315	116.865
Conceptual difference	1.039	1.041	1.061	1.007	1.017
			Percent errors:		
Civilian labor force		-4.3	-1.9	0.1	-1.7
Unemployment rate		-18.2	-18.2	9.1	14.5
Employment		-3.3	-0.9	-0.4	-2.5
GNP/Employee		11.8	11.4	-1.5	2.9
GNP		8.0	10.4	-1.9	0.3
Employment (persons)		-3.3	-0.9	-0.4	-2.5
Employment (jobs)		-3.2	1.1	-3.5	-4.7
Conceptual difference		0.1	2.0	-3.1	-2.2

Table 2. GNP by Major Demand Category, Actual and Projected, 1990

[billions of 1982 dollars]

	Actual 1990	12/78 Base	8/81 High-II	11/83 Moderate	11/85 Moderate
Gross national product	4157.3	4491.9	4591.5	4080.0	4170.1
Personal consumption expenditures	2681.6	3041.6	2912.6	2648.4	2709.3
Durable goods	427.4	473.3	415.3	376.2	391.2
Nondurable goods	911.1	1103.7	1030.6	952.4	956.3
Services	1343.1	1464.5	1466.7	1319.7	1361.7
Gross private domestic investment	688.7	751.2	948.6	695.7	789.7
Fixed investment	692.3	704.1	897.3	665.5	735.5
Nonresidential	515.5	504.1	675.0	470.1	561.7
Structures	120.9	160.1	154.7	151.5	165.9
Producers' durable equipment	394.6	344.0	520.3	318.6	395.8
Residential	176.8	200.0	222.3	195.4	173.8
Exports	631.5	415.1	597.4	485.1	500.3
Imports	665.3	457.8	660.6	479.7	643.6
Government purchases of goods & services	820.8	741.8	793.6	730.4	814.4
Federal	343.7	271.8	320.4	305.9	347.1
National defense	258.7	188.0	249.0	214.8	273.0
Nondefense	85.0	83.8	71.4	91.0	74.1
State and local	477.1	470.0	473.2	424.5	467.3
Percent distributions:					
Gross national product	100.0	100.0	100.0	100.0	100.0
Personal consumption expenditures	64.5	67.7	63.4	64.9	65.0
Durable goods	10.3	10.5	9.0	9.2	9.4
Nondurable goods	21.9	24.6	22.4	23.3	22.9
Services	32.3	32.6	31.9	32.3	32.7
Gross private domestic investment	16.6	16.7	20.7	17.1	18.9
Fixed investment	16.7	15.7	19.5	16.3	17.6
Nonresidential	12.4	11.2	14.7	11.5	13.5
Structures	2.9	3.6	3.4	3.7	4.0
Producers' durable equipment	9.5	7.7	11.3	7.8	9.5
Residential	4.3	4.5	4.8	4.8	4.2
Exports	15.2	9.2	13.0	11.9	12.0
Imports	16.0	10.2	14.4	11.8	15.4
Government purchases of goods & services	19.7	16.5	17.3	17.9	19.5
Federal	8.3	6.1	7.0	7.5	8.3
National defense	6.2	4.2	5.4	5.3	6.5
Nondefense	2.0	1.9	1.6	2.2	1.8
State and local	11.5	10.5	10.3	10.4	11.2
Percent errors:					
Gross national product		8.0	10.4	-1.9	0.3
Personal consumption expenditures		13.4	8.6	-1.2	1.0
Durable goods		10.7	-2.8	-12.0	-8.5
Nondurable goods		21.1	13.1	4.5	5.0
Services		9.0	9.2	-1.7	1.4
Gross private domestic investment		9.1	37.7	1.0	14.7
Fixed investment		1.7	29.6	-3.9	6.2
Nonresidential		-2.2	30.9	-8.8	9.0
Structures		32.4	27.9	25.3	37.2
Producers' durable equipment		-12.8	31.9	-19.3	0.3
Residential		13.1	25.8	10.5	-1.7
Exports		-34.3	-5.4	-23.2	-20.8
Imports		-31.2	-0.7	-27.9	-3.3
Government purchases of goods & services		-9.6	-3.3	-11.0	-0.8
Federal		-20.9	-6.8	-11.0	1.0
National defense		-27.3	-3.7	-17.0	5.5
Nondefense		-1.5	-16.0	7.1	-12.8
State and local		-1.5	-0.8	-11.0	-2.0

Table 3. Sources of Error, GNP and Employment Projections, 1990

Levels	Actual 1990	12/78 Base	8/81 High-II	11/83 Moderate	11/85 Moderate
GNP	4157.3	4491.9	4591.5	4080.0	4170.1
Civilian labor force	124.787	119.367	122.375	124.951	122.653
Unemployment rate	5.5	4.5	4.5	6.0	6.3
GNP/Employee	35.254	39.404	39.288	34.737	36.285
		GNP percent error due to:			
Total GNP Error		8.0	10.4	-1.9	0.3
Labor Force Error		-4.3	-1.9	0.1	-1.7
Unemployment Rate Error		1.1	1.1	-0.5	-0.8
Productivity Error		11.8	11.4	-1.5	2.9
Interaction		-0.4	-0.1	0.0	-0.1
Employment (jobs)	122.571	118.615	123.960	118.315	116.865
Civilian labor force	124.787	119.367	122.375	124.951	122.653
Unemployment rate	5.5	4.5	4.5	6.0	6.3
Conceptual difference	1.039	1.041	1.061	1.007	1.017
		Employment percent error due to:			
Total employment error		-3.2	1.1	-3.5	-4.7
Civilian labor force		-4.3	-1.9	0.1	-1.7
Unemployment rate		1.1	1.1	-0.5	-0.8
Conceptual difference		0.1	2.0	-3.1	-2.2
Interaction		0.0	0.0	0.0	0.1

Table 4. Sector-level Employment, Actual and Projected, 1990
[thousands of jobs]

	Actual 1990	4/79 Base	8/81 High-II	11/83 Moderate	11/85 Moderate
Total employment	122571	118615	123960	118315	116865
Agriculture	3276	3046	3253	3354	3164
Mining	527	787	959	781	659
Construction	6842	6033	7104	6963	6189
Manufacturing	19561	23882	23905	22236	20913
Durables	11386	14693	14872	13550	12872
Nondurables	8175	9189	9033	8686	8041
Transportation	3843	3332	3671	3451	3507
Communications	1321	1473	1567	1688	1485
Utilities	1091	853	1003	1063	1111
Trade	27843	27370	27445	26355	27106
Finance, insurance, and real estate	7390	6695	7108	7113	6991
Services	31654	26330	26694	27161	28142
Government enterprises	1846	2017	1778	1607	1543
Special industries	17377	16797	19473	16543	16055
			Percent errors:		
Total employment		-3.2	1.1	-3.5	-4.7
Agriculture		-7.0	-0.7	2.4	-3.4
Mining		49.3	82.0	48.2	25.0
Construction		-11.8	3.8	1.8	-9.5
Manufacturing		22.1	22.2	13.7	6.9
Durables		29.0	30.6	19.0	13.1
Nondurables		12.4	10.5	6.3	-1.6
Transportation		-13.3	-4.5	-10.2	-8.7
Communications		11.5	18.6	27.8	12.4
Utilities		-21.8	-8.1	-2.6	1.8
Trade		-1.7	-1.4	-5.3	-2.6
Finance, insurance, and real estate		-9.4	-3.8	-3.7	-5.4
Services		-16.8	-15.7	-14.2	-11.1
Government enterprises		9.3	-3.7	-12.9	-16.4
Special industries		-3.3	12.1	-4.8	-7.6

Table 5. Industry Employment, Actual 1990 and Projected Errors

[thousands of jobs]

	Actual 1990	April 1979 Average Error	April 1979 Percent of Actual	August 1981 Average Error	August 1981 Percent of Actual	November 1983 Average Error	November 1983 Percent of Actual	November 1985 Average Error	November 1985 Percent of Actual
Total employment	122571	-452	-0.4	2960	2.4	-4352	-3.6	-6343	-5.2
Agriculture	3276	-220	-6.7	-104	-3.2	81	2.5	-112	-3.4
Dairy and poultry products	372	53	14.4	14	3.9	11	3.0	-49	-13.2
Meat animals and livestock	470	22	4.8	24	5.1	4	0.9	-23	-5.0
Cotton	46	110	240.2	84	184.1	8	18.8	-12	-26.1
Food and feed grains	601	87	14.5	41	6.8	-12	-2.0	-50	-8.4
Other agricultural products	1149	-278	-24.2	-269	-23.4	1	0.1	8	0.8
Forestry and fishery products	77	-31	-40.9	1	2.2	3	4.3	-1	-1.3
Agricultural, forestry, and fishery services	561	-184	-32.9	-1	-0.3	64	11.5	14	2.6
Mining	527	277	52.7	468	88.8	245	46.5	129	24.5
Iron and ferroalloy ores mining	18	9	52.8	17	94.4	6	33.3	-3	-16.7
Copper ore mining	17	41	244.1	18	109.8	9	56.9	-1	-5.9
Nonferrous metal ores mining except copper	25	4	18.0	15	62.7	8	34.7	-2	-10.7
Coal mining	149	212	142.6	283	190.2	137	92.2	48	32.7
Crude petroleum and natural gas, except drilling	206	7	3.6	108	52.8	76	37.1	83	40.6
Stone and clay mining and quarrying	90	10	11.1	14	16.3	-2	-2.2	4	5.2
Chemical and fertilizer mineral mining	22	-8	-36.4	10	45.5	9	40.9	-1	-6.1
Construction	6842	-643	-9.4	335	4.9	169	2.5	-717	-10.5
New construction	5405	-733	-13.6	301	5.6	-114	-2.1	-601	-11.1
Maintenance and repair construction	1437	90	6.3	34	2.4	284	19.8	-115	-8.0
Manufacturing	19561	5002	25.6	4739	24.2	2624	13.4	1204	6.2
Durables	11386	3735	32.8	3773	33.1	2160	19.0	1395	12.3
Ordnance	75	-6	-8.7	29	38.7	13	17.8	32	42.7
Guided missiles and space vehicles	134	-46	-34.3	-61	-45.5	-5	-3.7	13	10.2
Logging	127	-8	-6.3	-13	-10.2	4	3.4	-16	-12.9
Sawmills and planing mills	217	-13	-6.0	6	2.8	-17	-8.1	-26	-12.3
Other millwork, plywood, and wood products	451	-0	-0.1	-87	-19.4	-43	-9.7	-74	-16.4
Wooden containers	22	-7	-34.1	-0	-3.0	-9	-43.9	-10	-45.5
Household furniture	307	209	68.1	86	28.0	42	13.8	7	2.3
Furniture and fixtures, except household	229	-39	-17.2	-32	-14.3	-30	-13.1	18	8.0
Glass	164	67	41.2	80	49.2	37	22.8	7	4.3
Cement and concrete products	229	46	20.1	31	13.8	8	3.6	14	6.4
Structural clay products	37	2	5.4	7	18.9	0	0.0	-2	-7.2
Pottery and related products	40	3	7.5	17	43.3	5	12.5	6	15.8
Other stone and clay products	102	58	57.4	77	76.1	62	61.1	48	47.1
Blast furnaces and basic steel products	276	293	106.3	308	111.7	152	55.2	35	12.7
Iron and steel foundries and forging	200	166	83.3	180	90.2	53	26.7	1	0.8
Primary copper and copper products	101	81	80.7	65	64.7	59	58.7	35	35.0
Primary aluminum and aluminum products	86	90	104.7	89	103.5	86	100.0	66	77.1
Primary nonferrous metals and products	136	-31	-22.8	-24	-18.1	-51	-38.0	-62	-45.6
Metal containers	50	41	83.0	45	90.0	18	37.3	7	14.0
Heating apparatus and plumbing fixtures	61	18	30.3	42	68.9	14	23.0	1	2.2
Fabricated structural metal products	440	275	62.5	168	38.3	129	29.3	54	12.3
Screw machine products	96	29	30.7	49	51.0	18	19.4	8	8.7
Metal stampings	187	123	65.8	91	48.7	58	31.2	39	21.0
Cutlery, handtools, general hardware	133	103	77.4	98	74.2	50	37.6	26	20.1

Table 5. Industry Employment, Actual 1990 and Projected Errors (Cont.)

[thousands of jobs]

	Actual 1990	April 1979 Average Error	April 1979 Percent of Actual	August 1981 Average Error	August 1981 Percent of Actual	November 1983 Average Error	November 1983 Percent of Actual	November 1985 Average Error	November 1985 Percent of Actual
Other fabricated metal products	360	38	10.6	99	27.6	45	12.5	24	6.7
Engines, turbines, and generators	89	60	68.0	72	81.6	62	70.4	30	34.5
Farm machinery	107	80	75.2	120	112.5	62	57.9	23	21.8
Construction, mining, and oilfield machinery	147	198	134.7	257	175.3	173	117.9	60	41.0
Material handling equipment	82	60	73.2	78	95.9	32	39.4	20	25.2
Metalworking machinery	337	112	33.2	124	36.9	47	13.9	20	6.0
Special industry machinery	164	73	44.8	67	41.1	44	26.8	20	12.4
General industrial machinery	257	145	56.6	148	57.6	83	32.4	47	18.5
Other nonelectrical machinery	327	63	19.4	39	12.1	4	1.4	9	2.8
Computers and peripheral equipment	397	97	24.4	177	44.6	191	48.2	237	59.8
Typewriters and other office equipment	43	3	8.1	36	85.3	16	38.8	5	13.2
Service industry machines	183	86	47.0	28	15.5	17	9.3	2	1.3
Electric transmission equipment	98	186	189.8	155	158.8	144	146.9	128	131.0
Electrical industrial apparatus	170	129	76.2	156	92.0	93	55.1	58	34.3
Household appliances	125	95	76.4	68	54.7	58	46.9	26	21.3
Electric lighting and wiring	190	120	63.2	132	69.8	48	25.3	28	15.1
Radio and television receiving sets	84	55	65.5	27	32.5	19	23.4	7	8.7
Telephone and telegraph apparatus	128	97	75.8	92	72.4	59	46.1	42	33.3
Radio and communication equipment	138	176	127.5	287	208.2	303	220.0	414	300.0
Electronic components	585	38	6.6	83	14.2	169	28.9	199	34.1
Other electrical machinery and equipment	167	15	9.0	20	12.2	3	2.2	17	10.4
Motor vehicles	812	379	46.7	158	19.5	6	0.8	32	4.0
Aircraft	760	-227	-29.9	36	4.7	-73	-9.6	-70	-9.3
Ship and boat building and repair	198	70	35.6	87	44.1	56	28.3	19	9.9
Railroad equipment	33	33	101.5	42	129.3	13	40.4	4	12.1
Motorcycles, bicycles, and parts	13	26	203.8	15	120.5	5	38.5	3	28.2
Other transportation equipment	46	270	588.0	83	181.2	51	110.9	55	121.0
Scientific and controlling instruments	609	-402	-66.0	-344	-56.5	-316	-51.9	-340	-55.9
Medical and dental instruments	247	-64	-26.1	-48	-19.4	-41	-16.6	-32	-13.2
Optical and ophthalmic equipment	43	50	117.4	54	125.6	43	100.0	36	84.5
Photographic equipment and supplies	100	92	92.0	53	53.7	69	69.7	32	32.3
Watches and clocks	11	28	259.1	15	136.4	11	103.0	5	45.5
Jewelry and silverware	86	15	18.0	5	6.2	-4	-5.0	-7	-8.9
Musical instruments and sporting goods	129	75	58.1	42	32.8	10	8.0	12	9.3
Other manufactured products	221	0	0.2	44	19.9	-5	-2.3	-9	-4.4
Nondurables	8175	1267	15.5	966	11.8	464	5.7	-190	-2.3
Meat products	430	-54	-12.7	-44	-10.4	-71	-16.7	-86	-20.2
Dairy products	159	-13	-8.5	-0	-0.4	-13	-8.4	-25	-15.9
Canned and frozen foods	304	75	24.8	3	1.0	31	10.4	-19	-6.5
Grain mill products	128	32	25.0	29	22.9	16	12.8	3	2.9
Bakery products	217	1	0.5	-6	-2.9	-9	-4.5	-21	-9.7
Sugar	25	10	40.0	8	33.3	5	21.3	-3	-12.0
Confectionery products	79	10	12.7	-6	-8.0	-0	-0.8	-8	-11.0
Alcoholic beverages	65	17	26.9	-1	-2.1	19	30.3	-2	-3.1
Soft drinks and flavorings	123	46	37.8	35	29.0	44	35.8	15	12.2
Other food products	159	27	17.0	-5	-3.1	11	6.9	-10	-6.3
Tobacco manufacturing	49	12	25.5	17	34.7	13	27.2	11	23.1
Fabric, yarn, and thread mills	377	179	47.6	159	42.4	78	20.7	24	6.5
Floor covering mills	64	43	67.2	0	1.0	-7	-10.9	-17	-27.1

Table 5. Industry Employment, Actual 1990 and Projected Errors (Cont.)

[thousands of jobs]

	Actual 1990	April 1979 Average Error	April 1979 Percent of Actual	August 1981 Average Error	August 1981 Percent of Actual	November 1983 Average Error	November 1983 Percent of Actual	November 1985 Average Error	November 1985 Percent of Actual
Other textile mill products	51	38	75.5	25	49.7	21	41.2	0	1.3
Hosiery and knit goods	208	138	66.6	36	17.5	6	3.0	-24	-11.9
Apparel	857	510	59.5	382	44.7	206	24.0	58	6.8
Other fabricated textile products	217	31	14.3	23	10.9	6	3.1	-25	-11.7
Paper products	490	58	11.9	57	11.6	27	5.6	-9	-1.8
Paperboard	210	68	32.4	18	8.9	-10	-4.8	-23	-11.0
Newspaper printing and publishing	497	8	1.7	30	6.2	-4	-0.9	13	2.7
Periodical and book printing and publishing	358	-105	-29.3	-45	-12.6	-58	-16.4	-60	-16.8
Other printing and publishing	823	-226	-27.5	-131	-15.9	-76	-9.3	-3	-0.4
Industrial inorganic and organic chemicals	293	117	40.1	130	44.4	64	22.1	7	2.4
Agricultural chemicals	56	4	7.1	17	30.4	27	48.2	6	11.3
Other chemical products	104	19	18.8	14	13.5	9	8.7	-3	-3.5
Plastic materials and synthetic rubber	104	8	7.7	-0	-0.6	10	9.9	-17	-17.0
Synthetic fibers	80	109	136.9	18	23.3	36	45.8	2	2.5
Drugs	238	-5	-2.3	-1	-0.7	15	6.3	-9	-3.9
Cleaning and toilet preparations	161	12	7.5	-8	-5.0	5	3.5	-8	-5.0
Paints and allied products	62	34	54.8	9	15.1	8	13.4	-2	-4.3
Petroleum refining and related products	159	23	14.5	30	19.3	24	15.3	23	14.7
Tires and inner tubes	86	71	83.1	41	47.7	16	18.6	3	4.3
Rubber products except tires and tubes	176	27	15.3	5	2.8	-24	-13.8	-31	-18.0
Plastic products	630	-147	-23.4	27	4.3	-12	-1.9	21	3.4
Leather tanning and industrial leather	15	1	10.0	-0	-2.2	0	4.4	-0	-4.4
Leather products including footwear	121	84	69.8	96	79.6	48	39.9	31	26.2
Transportation	3843	-417	-10.9	-80	-2.1	-388	-10.1	-364	-9.5
Railroad transportation	287	177	61.8	188	65.6	98	34.1	33	11.7
Local transit and intercity buses	402	-70	-17.5	-49	-12.2	-58	-14.5	-77	-19.3
Truck transportation	1827	-196	-10.7	134	7.4	-118	-6.5	-91	-5.0
Water transportation	183	11	6.0	11	6.2	24	13.1	32	17.9
Air transportation	759	-232	-30.6	-252	-33.3	-231	-30.5	-226	-29.8
Pipeline transportation	19	-1	-5.3	3	15.8	4	24.6	1	5.3
Transportation services	366	-106	-29.0	-116	-31.7	-106	-29.0	-36	-9.8
Communications	1321	179	13.6	293	22.2	378	28.7	154	11.7
Radio and television broadcasting	238	21	9.0	32	13.4	62	26.2	23	9.8
Communications except radio and television	1083	157	14.5	261	24.2	316	29.2	130	12.1
Utilities	1091	-215	-19.7	-36	-3.4	-32	-2.9	15	1.4
Electric utilities, public and private	694	-157	-22.7	-6	-1.0	10	1.4	65	9.5
Gas utilities, excluding public	207	-34	-16.4	43	20.9	12	5.8	19	9.2
Water and sanitary services,except public	190	-23	-12.4	-73	-38.6	-54	-28.4	-69	-36.3
Trade	27843	202	0.7	59	0.2	-1546	-5.6	-908	-3.3
Wholesale trade	6551	-507	-7.7	29	0.5	-268	-4.1	118	1.8
Eating and drinking places	6841			111	1.6	-901	-13.2	-422	-6.2
Retail trade, except eating and drinking	14451	709	4.9	-81	-0.6	-376	-2.6	-605	-4.2

Table 5. Industry Employment, Actual 1990 and Projected Errors (Cont.)

[thousands of jobs]

	Actual 1990	April 1979 Average Error	April 1979 Percent of Actual	August 1981 Average Error	August 1981 Percent of Actual	November 1983 Average Error	November 1983 Percent of Actual	November 1985 Average Error	November 1985 Percent of Actual
Finance, insurance, and real estate	7390	-498	-6.7	-196	-2.7	-456	-6.2	-444	-6.0
Banking	2282	-155	-6.8	-298	-13.1	-323	-14.2	-501	-22.0
Credit agencies and financial brokers	1098	104	9.5	171	15.6	244	22.3	355	32.3
Insurance	2297	-125	-5.5	-148	-6.4	-122	-5.3	-164	-7.2
Real estate	1713	-321	-18.8	78	4.6	-254	-14.9	-134	-7.8
Services	31654	-4598	-14.5	-4728	-14.9	-4496	-14.2	-3692	-11.7
Hotels and lodging places	2561	-760	-29.7	-544	-21.3	-654	-25.5	-425	-16.6
Personal and repair services	1163	99	8.6	257	22.1	372	32.0	368	31.6
Barber and beauty shops	744	-260	-34.9	-26	-3.6	-78	-10.5	-84	-11.3
Miscellaneous business services	7284	-2760	-37.9	-2756	-37.8	-2132	-29.3	-1116	-15.3
Advertising	274	-50	-18.4	-73	-26.6	-56	-20.7	-24	-8.9
Professional services, n.e.c.	2925	-696	-23.8	-629	-21.5	-314	-10.7	-124	-4.3
Automobile repair	1260	-102	-8.1	-85	-6.8	-228	-18.1	-194	-15.4
Motion pictures	567	-318	-56.2	-250	-44.2	-248	-43.8	-210	-37.1
Amusements and recreation services	1163	-133	-11.4	-133	-11.4	-104	-9.0	-125	-10.7
Doctor's and dentist's services	2509	-584	-23.3	-591	-23.6	-572	-22.8	-562	-22.4
Hospitals	3553	917	25.8	490	13.8	363	10.2	-341	-9.6
Medical services, except hospitals	2201	922	41.9	222	10.1	-9	-0.4	229	10.4
Education services (private)	2590	-752	-29.1	-482	-18.6	-388	-15.0	-548	-21.2
Nonprofit organizations	2860	-118	-4.1	-126	-4.4	-445	-15.6	-533	-18.6
Government enterprises	1846	270	14.7	-30	-1.6	-220	-12.0	-316	-17.1
Post Office	819			-134	-16.4	-212	-25.9	-129	-15.8
Other federal enterprises	182	-5	-2.7	33	18.1	-1	-0.7	-49	-26.9
Local government passenger transit	205			-13	-6.5	5	2.6	-8	-3.9
Other state and local enterprises	640	275	43.0	84	13.1	-12	-2.0	-129	-20.3
Special industries	17377	209	1.2	2241	12.9	-710	-4.1	-1290	-7.4
General Government	16354	-74	-0.5	1678	10.3	-1099	-6.7	-1410	-8.6
Private households	1023	283	27.7	562	55.0	388	38.0	119	11.7

Table 6. Various Forecast Error Measurements, Industry-level Employment Projections, 1990

Simulation Title	Published In	Average Error	Average Absolute Error	Mean Square Error
Base	1979	-26.9	152.3	110569
High-II	1981	9.3	124.2	93016
Moderate	1983	-28.4	109.5	61018
Moderate	1985	-38.0	93.2	43445

Simulation Title	Published In	Index of Dis-similarity	Theil's Information Statistic	Rank Correlation Coefficients On-- Growth Rates	Rank Correlation Coefficients On-- Absolute Changes
Base	1979	9.127	10.293	29.3	49.9
High-II	1981	7.582	8.039	52.0	63.5
Moderate	1983	6.224	5.253	76.6	76.5
Moderate	1985	4.915	3.628	87.5	87.6

Table 7. Industry Employment Rankings, 1977-90 Change, Actual vs. Projected

Industry	Actual 1977-90	4/79 Base	8/81 High-II	11/83 Moderate	11/85 Moderate
Fastest Growing Industries:					
Miscellaneous business services	1	3	5	2	2
Credit agencies and financial brokers	2	22	6	8	7
Transportation services	3	9	12	7	4
Water and sanitary services,except public	4	25	56	18	30
Guided missiles and space vehicles	5	108	87	3	3
Motion pictures	6	88	106	85	40
Medical services, except hospitals	7	1	2	6	6
Air transportation	8	36	65	30	29
Professional services, n.e.c.	9	32	24	9	8
Advertising	10	120	42	19	10
Doctor's and dentist's services	11	23	18	15	12
Hotels and lodging places	12	24	38	37	19
Automobile repair	13	73	14	24	20
Local government passenger transit	14	4	19	10	13
Education services (private)	15	60	48	27	35
Computers and peripheral equipment	16	6	1	1	1
Medical and dental instruments	17	15	30	12	11
Eating and drinking places	18	150	9	26	18
Real estate	19	55	21	23	25
Periodical and book printing and publishing	20	89	37	29	31
Largest Level Change:					
Miscellaneous business services	1	3	4	2	1
Retail trade, except eating and drinking	2	1	2	1	2
General Government	3	4	1	4	5
Eating and drinking places	4	2	3	3	3
Wholesale trade	5	8	7	7	4
New construction	6	7	5	6	8
Professional services, n.e.c.	7	10	10	8	6
Doctor's and dentist's services	8	12	11	10	10
Hotels and lodging places	9	13	13	18	11
Hospitals	10	6	6	5	9
Medical services, except hospitals	11	5	8	9	7
Education services (private)	12	16	17	13	14
Nonprofit organizations	13	11	9	17	17
Insurance	14	14	15	15	13
Credit agencies and financial brokers	15	15	16	14	12
Real estate	16	20	14	16	15
Banking	17	9	12	12	16
Automobile repair	18	26	20	25	21
Truck transportation	19	19	18	24	22
Amusements and recreation services	20	17	21	21	23

Table 8. Errors in projecting employment trends, selected projections

[percentage points]

Year published	Year projected	Difference between projected and actual trends: Total Employment	Industry trends (Average absolute errors): Unweighted	Industry trends (Average absolute errors): Weighted by size of industry
1966	1970	-0.2	1.4	1.1
1973	1975	0.6	2.3	1.3
1970	1980	-0.3	1.3	1.0
1973	1980	-0.2	2.7	2.1
1976	1980	-0.4	1.5	1.2
1973	1985	-0.2	2.0	1.5
1976	1985	-0.1	1.9	1.4
1979	1985	0.2	2.9	1.5
1979	1990	-0.2	2.6	1.4
1981	1990	0.2	2.6	1.4
1983	1990	-0.5	2.8	1.7
1985	1990	-0.8	2.8	2.0

Evaluation of the Labor Force Projections to 1990

Howard N Fullerton, Jr., Office of Employment Projections, Bureau of Labor Statistics

I. Introduction

The final step in the projection process is evaluation. Such evaluations help persons making projections better understand the types of problems and errors that could occur and allows users to focus on the accuracy of projections for a specific group in the labor force or on overall accuracy. Because the labor force projections are used in a variety of ways, several criteria are used to evaluate the projections.

This paper examines the errors in the labor force projections to 1990 and their sources. It examines projected levels and the rates of labor force participation errors within specific age groups for men and women; and for white and black and other. Where appropriate, the accuracy of the 1990 labor force projections is compared with the 1975, 1980, and 1985 labor force projections. The Bureau of Labor Statistics assesses its labor force projections -- evaluations of the projections to 1975, 1980, and 1985 have been published. See Swerdloff (1969), Ryscavage (1979), and Fullerton 1982 and 1988). The evaluations of projections for 1980 concluded that the BLS projections of the labor force had been too low, with the level of the male labor force being projected to be too high and that of women too low; in fact so low that the overall level of the projected labor force was too low. By 1985, the projections, though generally low, also included some cases where the overall projections, including those for women, were too high. Indeed, the conclusion was that BLS had improved the accuracy of its labor force projections.

Six projections of the labor force to 1990 were published over the 1973 to 1985 period. See the following *Monthly Labor Review* articles: Johnston (1973), Fullerton and Flaim (1976 and 1978), Fullerton (1983 and 1985), Fullerton and Tschetter (1983). Each of the six projections had three alternatives; for most of this analysis, the focus will be on the middle or "moderate" growth projection in each series. (See Appendix table 1.) In this analysis, we compare the projected labor force numbers for 1990 with the annual average estimates of the labor force derived from the Current Population Survey, using weights from the 1980 Census. We call such estimates "the actual." Table 1 shows the 6 projections to 1990 (in millions) and their errors. The overall error became progressively smaller through the 1983 projection, when it was 0.1 percent, or fewer than a quarter of a million persons, but increased in the next projection (1985) to 1.7 percent, to near that of 1980. What were the sources of labor force error and why, with one exception, did the error fall as time passed?

A closer look at the 1990 labor force projections rounds for men and women provides a clue. For most of the rounds, labor force levels for both women and men were projected too low. The 1983 projection of women in the labor force was too high, but in 1980, there was no difference between the actual and projected number of women in the labor force. Men had the most accurately projected labor force estimates in 1983, the year their labor force was slightly overprojected. In the 1985 projection, the error was about the same size for men and for women.

It is to be expected that the earlier projections are less accurate than the more recent ones. The table 2 displays the growth rates for the total civilian labor force historically with the projected annual rate and the actual annual rate of change. The historic rate is calculated over the same number of years *before* the date of the projection as 1990 is *after* the date of the projection. The historic rate gives a standard of comparison--a naive projection.

The error in the projected growth rate for the 1985 labor force projection was greater than the error in the 1980 projection. Still, the more recent projections are the more accurate. The 1983 projected labor force growth rate is the only one that exceeded the actual growth rate. This table also allows us to characterize the labor force projections: All six projections reflected a view that the labor force would grow more slowly in the future than it had in the past. This did not happen over the 1973 to 1990 period, but did hold for the remainder of the projections.

Labor force projections are prepared by BLS by developing for each specific age-sex (and in the more recent projections, race) group a projected labor force participation rate. Then, using population projections by the Bureau of the Census for the same specific age-sex group, total labor force levels are developed. Consequently, in reviewing the labor force projections, there are two possible sources of error--the population projection and the participation rate projection. Before the 1980 Census, population projections were considered to be a trivial source of error and their potential contribution to the errors in the labor force projections was ignored. However, after the 1980 Census, there was a significant upward revision in the estimated civilian noninstitutional population that resulted in a similar upward revision in the labor force estimates for the 1971-82 period. The current labor force estimates are consistent with those revisions. The labor force projections made to 1985 were low by some 3.4 million because of errors associated with estimating the population size and making population projections. For the 1970 and 1973 projections for 1985, this amounted to a third of the error. For the 1976 and later projections made to 1985, the error due to participation rate projection dropped, so the share of error attributed to population projection increased.

In this paper, the consequences of errors in the population projections will be examined first, then the effects of labor force participation rate errors. This will include an examination of the errors in the age-sex specific labor force participation rates and the resulting errors in labor force composition.

2. Population Errors

Table 3 shows 1990 projections for the civilian, noninstitutional population aged 16 and older for men and women (in millions) and the errors associated with the total population projections. As indicated, the error in the population projection fell over the 1973-78 period, was steady in 1980 and 1983, and then dropped sharply in 1985. The population projections were published in U.S. Bureau of the Census (1972; 1975; 1977; 1982; and 1984). To determine further the effects of the population projection error, the projected age-sex specific labor force participation rates for 1990 were multiplied by the actual 1990 civilian, noninstitutional population. Had the actual civilian noninstitutional population been known or projected correctly, all the projections except that made in 1983 would have been more accurate. The 1983 and 1985 projections' errors would have been trivial (less than one percent). The size of the errors because of population projection errors varied, unlike the case for the projections prepared for 1985. For the 1976 through 1980 projections, population projection errors added more than 5 million to the labor force error.

The error due to low population projections affected the male labor force projections more than the female labor force projections. This reflects the cause of the projection error: under estimation and projection of immigration. For men 25 to 34 there was an error of over a million due to population projection errors for the projections made over the 1976-80 period. With adjustments for immigration reflected in more recent projections, the error due to population projections dropped sharply for this age group and overall.

There are four elements of a population projection: the base year estimate, projections of births, of deaths, and of net immigration. Should the estimated structure or size of the population in the base year be incorrect, this will be extended through the early years of the projection. If projected net immigration is too low or too high, both the level and the age composition of the projected population would be affected. Errors in the base year estimates and projected net immigration significantly affect the labor force projections. For the period of time over which BLS makes labor force projections, the fertility and mortality assumptions have only a minor effect.

Although base year estimates and net immigration were the components of population projection error that significantly affected the labor force projections, the error was essentially the same: under estimates or under projections of immigration. More specifically, the sources of these errors were undocumented and refugee immigration. The base year estimates for projections using the 1970 census reflected under-enumeration of immigrants in the 1970 census as well as under estimation of immigration during the 1970's. The 1980 census also differed significantly from the 1970 census in coverage. Much of this, but not all, can be attributed to immigration over the period.
Until 1989, the Census Bureau did not incorporate any estimate of undocumented immigrants into their middle series population projections because they were not in their current estimates. Thus, the base year estimates were too low because of under-enumeration in the census and because undocumented immigrants were not included in the population estimates for the intervening years. Further, between 1985 and 1990, there were a major revisions of immigration law; the likely effect on the level and composition of immigration could not have been incorporated since the projections do not anticipate major changes in policy. Currently, the Census Bureau and BLS use more than one immigration scenario to reflect the effect of alternative assumptions about immigration on the size and composition of the population and labor force.

Although the population projection errors cannot be allocated between the base year errors and the specific immigration projection errors, it is possible to determine the share of overall error in the projection of the labor force due to population and the share due to participation rate error (table 4). This was done by using the projected labor force participation rates and the estimated 1990 annual average civilian noninstitutional population.

The errors due to the population projection dropped across the sets of projections, from a high of 4.6 million for the 1973 projection to a low of 1.2 million for the 1985 projection. Because the errors due to participation rates dropped for the first three projections, the population errors became a greater proportion of the overall error in the labor force projections in each succeeding projection. Two-thirds of the error in the 1973 projection may be attributed to the participation rate errors; by the 1985 projection, that had dropped to less than one-half of the error.

According to this analysis, the size of errors in the population projection varied over the 5 projection years. Population accounted for a small proportion of the error in the earliest projection. As participation rate error decreased in later projections, population projection error accounted for an increasing proportion of the error in the projected labor force level. By 1980, the errors in projecting labor force participation and in projecting population offset each other. The primary source of error for the population projections was underestimates of immigration, in particular, lack of any accounting for undocumented immigrants. This was anticipated by the authors of the 1978 projections who suggested that "The population projections might have to be revised to reflect a better knowledge of net migration trends, particularly with regard to the inflows of undocumented aliens," see Fullerton and Flaim.

3. Measures of Errors

For each of the six projections of the 1990 labor force, there are 20 combinations of age-sex groups and, therefore, 20 possible errors. The errors in participation rates can be examined. One can either look at each error or calculate a statistic to summarize the error for a specific projection. Different summary statistics emphasize different problems with the projections.

Summary measures of errors. This measure is calculated using the mean of the absolute value of percent errors in the age-sex specific labor force participation rates. The percent or relative errors attach more significance to errors in groups with low participation. The MAPE's for the projected participation rates ranged from 6.8 to 11.8, with the 1973, 1978, and 1980 projections having by far the greatest values. See appendix table 2. The other three remaining projection years (1976, 1983, and 1985) had the same MAPE's, around 6.9, with the 1983 projection having the lowest error. This is consistent with the earlier finding that the growth rate projected in 1983 had the least error. It appears that the 1976 projection is more like the 1983 and 1985 projections than like the earlier ones. Using this measure, the range of error is down from that for the projections to 1985. Further, the greatest MAPE for the projection to 1990 (11.6) is much smaller than the greatest for the projections to 1985 (17.0).

Regression. Another summary measure of the errors in the labor force projection is the regression of projected labor force participation rates against actual 1990 labor force participation rates. If the projections were perfect, the actual labor force participation rate plotted against the projected rate would yield a straight line through the origin with a slope of 1. The table 5 presents estimates of the slope and intercept of these lines for each projection with a test of the hypothesis that the intercept is zero and the slope 1. Except for the 1985 projection, the hypothesis of "perfect forecast" cannot be rejected. (Since the errors in the projections are found not to be normally distributed, the reader may ask why an F-test is used because the normal distribution is required for such a test. A short answer is that it still provides a useful indication. For a discussion of the problem and methods of handling the problems, see Scheffe', (1959). Generally, the slopes are consistent with an interpretation of the errors being widely diffused among groups--no specific groups were overprojected or underprojected. The large values for the intercept reflect the errors in the participation rates. Tests of the hypothesis that the intercept is zero are not rejected. Thus, we conclude from these tests that the projections are unbiased, but have sizable errors.

Three of the projections are displayed in charts 1 through 3. The charts display projected labor force participation rates plotted against the actual for 1990 for 20 age groups. The dashed diagonal line from corner to corner shows the "line of perfect forecast:" the line where the markers would be on if the projection were perfect: values above this line are underprojected, below it, overprojected. The solid diagonal line summarizes how well the projected values, taken together, approximate the "line of perfect forecast." For the 1973 projection (chart 1), the fitted line is not parallel to the line of perfect forecast. It is pulled up by the cluster of rates projected to be 50 to 60 percent but which were in to 60 to 70 percent range. The value most over projected was for men 60 to 64. Labor force participation rates for women 25 to 34 were under projected the most. For the 1978 projection the lines are close. The observations are not as far from the line as in the 1973 projection. For this projection, there were over projections of the rates for teenagers, while the rates for women 45 to 54 and men 60 to 64 were under projected. The 1983 projection exhibits more precision--the values are even closer to the fitted line, but again the line does not coincide with the line of perfect forecast. The charts as a group suggest that the projections improved over time; the errors being equally likely to be extremely positive or negative.

Median error. The errors in the projection of participation rates for the various age-sex groups range from 22.5 percent too low for women 35 to 44 in the projection made in 1973 to 14.4 percent too high for men 60 to 64, also in the 1973 projection (appendix table 2). For the other projections to 1990, one of the teenager groups had the greatest over projection, reflecting the drop in their participation that occurred at the end of the projection period. The most recent BLS labor force projection assumes that participation for these ages will return to their levels of the late 1980's. See Fullerton (1991). Table 6 indicates the median error for each year a projection was made to 1990, the dispersion of the error and their extreme values.

If BLS is improving its projections, the median error would be closer to zero in 1985 than in 1973. This pattern does not appear, but all the median errors are less than one percentage point, suggesting a random drift with a small error. A median error near zero indicates that the projection was unbiased. That is not helpful if large positive and negative errors tended to cancel each other The dispersion, here measured by the mean absolute deviation (MAD), also became smaller. A low measure of dispersion indicates that there were few large, offsetting errors. Another way to verify this is to look at the greatest over projection and lowest under projection. We see that these numbers did get closer together in the more recent projections. The projections made in 1983 and in 1985 had their greatest errors less than 10 percent. This contrasts with the projection made in 1973, with errors greater than twenty percentage points. By comparison, an evaluation of projections of the 1985 labor force shows a 25-percent greatest error--more than any error in the projections for 1990.

Shapiro-Wilk test. Generally, it is assumed that errors are distributed according to the normal or Gaussian law. We can test for this using the Shapiro - Wilk test. Values for the test ranged from .90 to .92 for the 1973 to 1983 projections. By these test values, the hypothesis of normality would be rejected. The 1985 projection's test value was .96, which is consistent with normally distributed errors. Departures from normality could occur because the errors were not symmetric, for example more negative than positive errors, or because the tails of the distribution were too "fat" (there were several errors with very large positive or negative values) or too "thin." The kurtosis statistics indicate that the errors are grouped more closely around the mean than a Gaussian distribution making significance tests, such as regression tests, conservative. However, it appears that the distribution of errors did become more symmetric in the more recent projections.

Age, sex, and race errors. In the first two projections to 1990, there were large errors in the participation of women 25 to 54, reflecting assumptions that the participation rates of mothers would not grow sharply. They did. The pattern of errors reflect problems in projecting the participation rates of women born in the 1940's. Thus, BLS moved from an under projection of 10 percent in the 1976 projection of participation rates of women 25 to 34 to an over projection of 4.5 percent in 1978. The over projection grew to 7.1 percentage points before dropping to an over projection of 2.6 percentage points in 1985.

The pattern of groups with greatest errors shifted from women 25 to 54 in the 1973 projection to teenagers in the 1985 projection. Given the cyclical responsiveness of the teenage groups and the small number of these people in the labor force, it is not surprising that this is where the larger error is found. Of greater concern are the errors in participation of older workers, men ages 60 to 74 and women 65 to 69—because these errors may reflect a change in the long term trend in labor force participation for older workers. To illustrate this, the error in the participation rate for men 60 to 64 made in 1973 was 14 percentage points too high; by 1985, it was projected 5 percentage points too low. The 1973 to 1985 period was a time of rapid decreases in participation at these ages. Since 1985, participation has barely dropped. The same pattern of projecting participation too high at the beginning of the 1973-85 period and too low at the end also applies to women 60 to 64, though the percentage point error is lower.

In general, participation for men was projected higher than the actual--the overall rates were too high for five projections, with the lowest error in 1978 and the greatest in 1973. For women, the first three and the last projections of participation were too low--by 11 percentage points in the 1973 projection. The 1980 and 1983 projections had participation too high for women, as measured by their overall rate. (See appendix table 2.) This suggests that as time passed, the projections of women and men's participation were adjusted to reflect the changes in participation observed. Because the errors in participation for women were greater than those for men in all six projections, overall participation was under projected or over projected according to the pattern for women.

Starting with 1978, the labor force was projected by two race groups independently: whites and blacks and others. Because the white labor force is still the much larger component, errors in the projection of this group has a greater effect on the overall error. Overall white participation was over projected in 1978 through 1983. Participation for both white men and women was over projected in all projections, with the greatest error in 1980. In 1978, participation of both black men and women was under projected. The errors were much greater for blacks and others than for whites. In the 1980 and 1983 projections, rates for black men were more accurately projected than for whites, men or women. However, the rates for women were projected too high. The overall rate for black men was very near their actual 1990 rate. The errors were equivalent in participation rates by sex and race for 1985. Given that the black participation rates as measured are more variable than those for whites, the relative accuracy of black labor force participation is a surprise.

Relative errors. As noted earlier, the errors in participation of older women are small. That is not surprising as their participation is low. Relatively, their participation error is larger. For example, the 1.3 percentage point error for women 70 and older is a 26.6 percent relative error. Men in the prime working years have participation well over 95 percent, their relative errors are roughly the same size as their percentage point errors; women's participation is lower, their relative errors will be larger than their percentage point errors.

The earliest characterization of 1973 being by far the least accurate and 1978 being the most accurate holds for the relative error in overall participation. Overall, participation was more accurately projected for men than for women. Men's participation was equally accurate in 1978, 1983, and in 1985, whereas women's participation was projected most accurately in 1978. There was an improvement in the projection of both women and men's rates over the last two projections. The relative errors by race were higher for blacks and others than for whites. Black women had the highest relative error, black men the lowest.

To summarize the findings for detailed age groups; for the early years, the largest relative errors were for women 25 to 44. Starting in 1978, the relative errors for women 25 to 44 were no longer large, but were for teenagers and those 65 and older. These errors approached the size of the earlier relative errors for women 25 to 44. For women 20 to 34—principle ages of childbearing—the relative error was least for the 1985 projection. Since 1978, there has been an over projection of participation rates for women these ages. The 1976, 1983, and 1985 projections had about the same accuracy; the 1978 was worst.

Composition errors. Much of the interest in the labor force projections centers on its size and growth. To understand these, we must also consider labor force participation rates. However, there also is interest in the composition or age-sex structure of the labor force (appendix table 3). The index of dissimilarity measures how much the projected composition would have to change to be like the 1990 actual. For example, the 1980 projected composition would have to change by 3.7 percent to have the same composition as the 1990 estimates. Although the projected composition was worst in 1973; it improved with each projection, with the greatest improvement between the 1973 and 1976 projections. The errors in distribution for the 1973 projection were concentrated in men and women 25 to 44. For other projections, the error is widely distributed with small errors for any group.

4. Alternative Labor Force Projections

For each of these projections, two alternative projections were made. This raises at least two questions: did the range from low to high span the actual, and was one of the alternatives closer to the actual than the middle reviewed in the earlier sections. For evidence, we turn to chart 4. The last four projections had a range that did indeed cover the actual 1990 level. The 1978 high alternative was closer to the actual than the middle; the low alternative was closest in 1985.

The first two projections in chart 2 are striking. Not only did the 1973 and 1976 projections fail to cover the actual line, but the range was much smaller. At the time the projections were made, women in the 25- to 44-age group were a small part of the labor force. Their labor force participation rate though low, was growing rapidly. Although these women were the most significant source of error for the projections, they were too small to yield a large variation in the overall labor force. BLS changed its methodology in 1978 to have variations in labor force participation for all age groups.

For any year, BLS alternatives plotted through time have a "fan" shape; they are further apart the further from the take-off-year. It would then appear that these plots of alternatives should exhibit a "funnel" shape, the closer the alternative projections got to the target year, the more certain the projections should be about the actual. Over the 1978 to 1985 period, BLS was interested in making the range of projections approximate a confidence or credible interval. By the time the 1985 labor force projection was made it was apparent that economic variables could not be used to account for the variability in the labor force that a confidence interval approach implied. The alternative labor force projections are used in the aggregate economic projections, thus there must be some economic content in the alternatives. Starting with the 1985 labor force projection, the "fan" of alternatives did not spread more widely with each successive projection. Thus, in later evaluations, we should observe some of the "funnel" shape.

5. Assumptions

One of the questions of concern in the evaluation of projections is, "why one set has less error than another, particularly if the reason yields information which could improve future projections?" The BLS labor force projection method involves a high level of disaggregation followed by extrapolation of the labor force participation rate. The refinement of the methodology over time has included using five-year-of age data (1972 to present), use of parental status for women (1972 to 1978), and disaggregation by race (1978 to present). The extrapolation technique developed for the 1972 projection dampened the estimated growth rates for women rapidly. For the 1976-85 projections, tapering of rates was greatest toward the end of the projection period. Because the projections improved with later projections, the question arises whether the improvements result from changes in methods or simply later data.

For the labor force projections made over the 1972-85 period, the change in participation rates was projected. These changes were applied to a "take-off" (or base) participation rate and then successive participation rates were projected. To project the changes, past changes in participation rates were estimated. It was *assumed* that participation rate changes would ultimately cease. For the 1972 projection, when the drop in fertility rates had just begun, it was assumed that the rapid growth in women's labor force participation would soon cease as fertility increased. The opposite occurred and fertility dropped to the levels prevailing in the early 1930's and remained there. If a behavioral model relating fertility and women's labor force participation had been developed and used, the expectation that fertility would rise also would have led to participation lower than that which actually occurred. For the remaining projections, it was also assumed that changes in participation would also end, but that the greatest slowdown would be towards the end of the projection period--for the 1976 projection, between 1990 and 1995, for later projections, after 1995. For the 1980 projection, it was assumed that participation rates for women aged 20 to 44 would increase at an increasing rate then increase at a decreasing rate. The problems involved with selecting a take off point have been discussed by Ryscavage (1979) and Kok and de Neuborg (1986) and by Armstrong (1978) as the problem of estimating the current level. Especially in the short run, a projection's accuracy can be affected by the choice of a take off point. Because the 1972 projection was to be made for the years 1980, 1985, and 1990, 1970 was used as a takeoff point. This affected the accuracy of 1973 projection. The 1976 projection used the average of the last three years; later projections have used the last year in the sample period. If the rate of change is under estimated because a linear estimate is made when change is actually growing non-linearly, then every year the take off year is moved back compounds the problem. The effect of not using the most recent year is to shift the entire projection down (or up) for the entire period.
The 1973 through 1978 projections explicitly used the fertility assumptions to derive the number of women with young children. The use of the assumptions overstated the number of women with young children for the 1973 and 1976 projections and understated it slightly for the 1978 projection.

6. Summary

Overall comparison. Eleven explicit tests of the 1990 labor force projections were made. Which projection was best? Table 7 lists the number of times a specific projection was best or worst. The 1973 projection was worst 6 times. The 1985 projection ranked best on 5 tests, but was *worst* once. In considering this, there are several ways a projection can be best. For example, if errors offset, the projected level of the labor force would be nearly the actual level, yet the participation rates and the projected population would be incorrectly projected. However, if the main use of the projected labor force was the level or the growth of the overall labor force these details would not matter.

These tests help the user evaluate the projections in terms of their own needs--for accurate level of the total, for accurate participation rate projections, or for accurate projections of the composition. Different tests of the accuracy of the participation rate projections allow the user to focus on overall accuracy or accuracy of specific groups.

Earlier evaluations. As a group, the projections to 1990 were more accurate than the projections to 1975 and 1985.

References

Armstrong, J. S., *Long-Range Forecasting*, (New York, John Wiley, 1978).

Fullerton, H. N, Jr., (1980), "The 1995 labor force: a first look," *Monthly Labor Review*, December, p. 11-21.
_____, (1982), "How accurate were the projections of the 1980 labor force?" *Monthly Labor Review*, July, p. 15-21.
_____, (1985), "The 1995 labor force: BLS's latest projections," *Monthly Labor Review*, November, pp. 17-25.
_____, (1988), "An evaluation of labor force projections to 1985," *Monthly Labor Review*,. November, p. 7-17.
_____, (1991), "Labor force projections: the baby boom moves on," *Monthly Labor Review*, November, pp. 31-44.
_____, and Flaim, P. O., (1976), "New labor force projections to 1990," *Monthly Labor Review*, December, p. 3-13.
_____, and John Tschetter, (1983), "The 1995 labor force: a second look," *Monthly Labor Review*, November; pp. 3-10.

Johnston, D. F., (1973), "The U.S. labor force: projections to 1990," *Monthly Labor Review*, July, p. 3-13.

Kok, L., and de Neubourg, C., (1986), *Projecting labour supply, Methods, theory and research: an international comparison*, (The Hague, Organisatie voor Strategisch Arbeidsmarktonderzoek).

Ryscavage, P., (1979), "BLS labor force projections: a review of methods and results," *Monthly Labor Review*, April, p. 14-22.

Scheffe', Henry, (1959) "The Effects of Departures from the Underlying Assumptions," *The Analysis of Variance*, (New York, John Wiley & Sons), pp. 331-69.

Swerdloff, S., (1969),"How good were manpower projections for the 1960's," *Monthly Labor Review*, November, pp. 17-22.
U.S. Bureau of the Census, (1972), "Projections of the Population of the United States by Age and Sex: 1972 to 2020", *Current Population Reports*, Series P-25, No 493.

U.S. Bureau of the Census, (1975), "Projections of the Population of the United States: 1975 to 2050", *Current Population Reports*, Series P-25, No 601.
_____, (1977), "Projections of the Population of the United States: 1977 to 2050", *Current Population Reports*, Series P-25, No 704.
_____, (1982), "Projections of the Population of the United States: 1982 to 2050", *Current Population Reports*, Series P-25, No 922.
_____, (1984), "Projections of the Population of the United States: 1983 to 2080", *Current Population Reports*, Series P-25, No 952.

According to these summary measures, the worst projection to 1990 was worse than the worst projection to 1985, but the best projection to 1990 was often significantly more accurate than the best to 1985. When adjusted for the actual population, four projections to 1990 were more accurate. Generally, the more recent projections were more accurate, with the 1985 projection the most accurate.

Table 1. Actual and Projected Labor Force for 1990
Error and Percent Error

Projections for 1990 published in:	Labor Force Millions	Error	Percent Error
1973	110.6	-14.2	11.4
1976	113.8	-10.9	8.8
1978	119.4	-5.4	4.3
1980	122.4	-2.4	1.9
1983	125.0	0.2	0.1
1985	122.6	-2.1	1.7
1990 (actual)	124.8		

Table 2. Historical and projected annual growth rates, selected periods and to 1990

Percent

Projection for 1985 published in:	*Historical rate (1)*	*Projected rate (2)*	*Actual rate (3)*	*Error (2) - (3)*
1973	1.75	1.34	2.02	-0.68
1976	2.00	1.30	1.92	-0.62
1978	2.36	1.45	1.80	-0.35
1980	2.65	1.41	1.59	-0.18
1983	2.29	1.58	1.57	.02
1985	1.76	1.29	1.59	-0.29

Table 3. Total population 1990 actual and projected, by sex

Millions

Projection for 1990 published in:	*Total*	*Men*	*Women*	*Error of total*
1973	179	85	94	-9.5
1976	179	84	95	-9.1
1978	180	85	95	-7.8
1980	180	85	95	-7.9
1983	180	85	95	-7.9
1985	187	89	98	-1.4
1990 (estimated)	188	90	98	

Table 4. Division of projection error between participation rate and population errors

Millions

		Error attributed to:	
Projection for 1990 published in:	*Total error*	*Participation*	*Population*
1973	-14.2	-9.6	-4.6
1976	-10.9	-5.6	-5.3
1978	- 5.4	-0.2	-5.2
1980	- 2.4	3.0	-5.4
1983	0.2	1.4	-1.3
1985	- 2.1	-0.9	-1.2

Table 5. Regression coefficients for test of "perfect forecast"

Projection for 1990 published in:	*Intercept*	*Slope*	*F-test*	*Probability > F*
1973	3.2	1.0	0.52	0.60
1976	1.2	1.0	0.30	0.74
1978	3.2	0.9	0.98	0.39
1980	1.3	0.9	2.30	1.13
1983	1.6	1.0	1.49	0.25
1985	3.7	1.0	5.29	0.02

Table 6. Summary of labor force participation projection errors

	1973	*1976*	*1978*	*1980*	*1983*	*1985*
Median	-0.05	0.15	-0.80	0.45	-0.20	-0.55
MAD	6.1	3.7	4.9	4.4	3.1	2.1
Greatest	14.4	7.1	11.6	12.8	7.3	3.2
Lowest	-22.5	-13.5	-10.7	-6.9	-4.2	-6.0

Table 7. Number of times a projection was best or worst

Projection	Best	Worst
1973	1	6
1976	1	...
1978	1	3
1980	1	1
1983	2	...
1985	5	1

Table 8. Comparison of projections to 1985 and 1990 for specific characteristics

Projection to:		**1990**		**1985**	
Errors in level (millions):	Best (year):	.2	(1983)	-0.5	(1980)
	Worst (year)	-14.2	(1973)	-11.0	(1970)
Error in growth rate (percent)	Best (year):	.02	(1983)	-0.07	(1978)
	Worst (year)	-0.68	(1973)	-0.61	(1970)
Mean Absolute Percent Error	Best (year):	6.8	(1985)	6.0	(1980)
	Worst (year)	10.8	(1973)	17.0	(1970)
Index of dissimilarity	Best (year)	2.6	(1985)	1.4	(1980)
	Worst (year):	7.6	(1973)	7.5	(1970)

Chart 1. Labor force participation rates for 1990, actual and as projected 1973

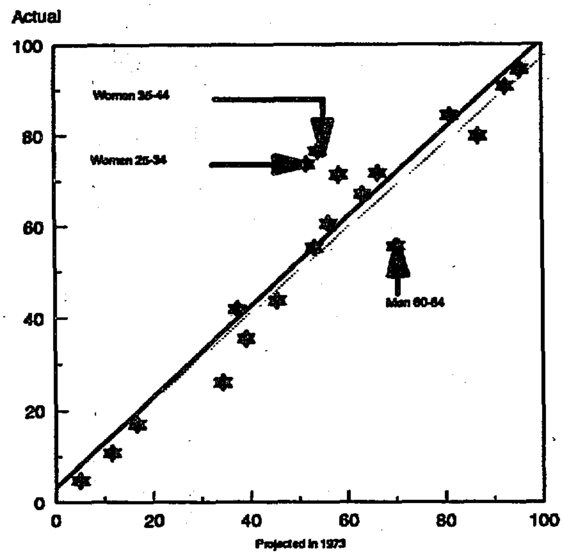


Chart 2. Labor force participation rates for 1990, actual and as projected in 1978

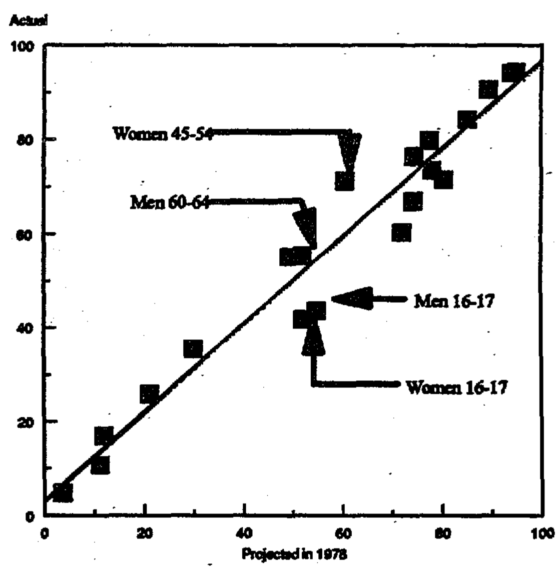


Chart 3. Labor force participation rates for 1990, actual and as projected 1983

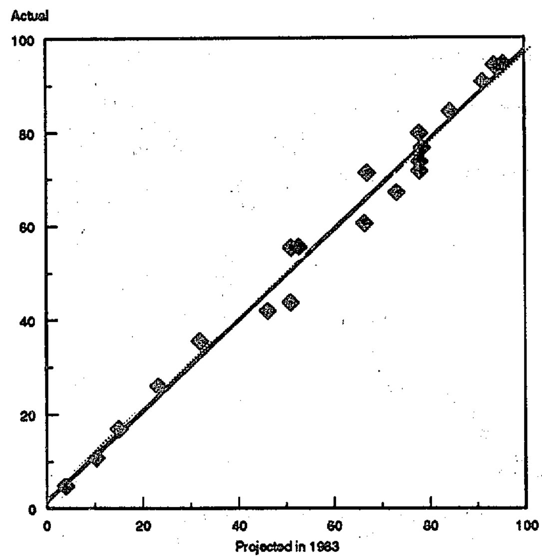
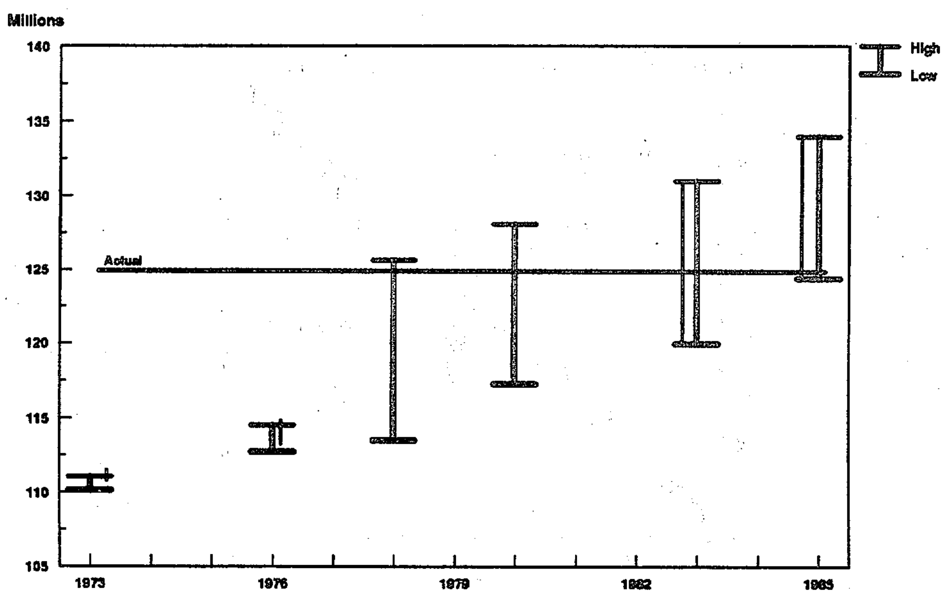


Chart 4. Range of projections for 1990, made 1973-85



Appendix Table 1. Characteristics of the 1990 labor force, and participation rates, actual and as projected in 1973, 1976, 1978, 1980, 1983, and 1985

Labor force group	Labor force (in thousands)							Participation rate (in percent)						
	As projected in --						Actual	As projected in --						Actual
	1973	1976	1978	1980	1983	1985	1990	1973	1976	1978	1980	1983	1985	1990
Total	110,576	113,839	119,366	122,375	124,951	122,653	124,787	62.0	63.6	66.2	67.9	66.9	65.7	66.4
Men, 16 and older	66,947	65,220	65,115	65,880	67,701	67,146	68,234	79.1	77.3	76.4	77.2	76.5	75.8	76.1
16 and 17 years	1,511	1,612	1,740	1,733	1,664	1,453	1,477	45.6	50.8	54.9	54.5	51.0	44.4	43.7
18 and 19 years	2,159	2,364	2,459	2,483	2,459	2,387	2,389	63.2	71.4	74.4	74.3	73.2	70.2	67.0
20 to 24 years	6,462	6,671	6,957	7,066	7,151	7,323	7,291	81.1	82.1	85.0	86.4	84.4	86.3	84.3
25 to 34 years	19,382	18,545	18,401	18,453	19,569	19,665	19,813	95.4	94.7	93.9	94.3	93.7	94.1	94.2
35 to 44 years	17,131	16,571	16,593	16,672	17,469	17,318	17,268	95.6	94.8	94.8	95.2	95.6	94.7	94.4
45 to 54 years	10,863	10,901	10,851	11,022	11,142	11,096	11,177	92.5	90.2	89.4	90.8	91.3	90.8	90.7
55 to 59 years	4,109	3,990	3,870	3,922	3,842	3,849	4,014	86.9	81.6	77.6	78.7	78.1	78.3	79.8
60 to 64 years	3,195	2,714	2,513	2,703	2,577	2,446	2,771	69.9	57.7	52.0	55.9	52.8	50.2	55.5
65 to 69 years	1,365	1,125	932	1,019	1,019	873	1,192	34.4	26.6	21.2	23.2	23.3	20.0	26.0
70 years and older	770	727	799	807	809	736	841	11.6	10.7	11.2	11.3	10.3	9.4	10.8
Women, 16 and older	43,629	48,619	54,253	56,495	57,250	55,507	56,554	46.5	51.4	57.1	59.6	58.3	56.6	57.5
16 and 17 years	1,205	1,448	1,608	1,685	1,461	1,309	1,356	37.4	46.9	52.1	54.7	46.2	41.4	41.9
18 and 19 years	1,975	2,201	2,531	2,509	2,317	2,139	2,188	56.2	62.5	72.1	72.1	66.5	61.3	60.5
20 to 24 years	5,808	6,656	7,086	7,131	7,035	6,641	6,552	66.3	75.2	80.4	81.4	78.1	73.8	71.6
25 to 34 years	10,669	13,077	16,063	16,568	16,804	16,366	15,990	51.6	63.5	78.1	80.7	78.1	76.2	73.6
35 to 44 years	10,216	11,678	13,820	14,581	14,974	14,458	14,576	54.0	63.0	74.5	78.6	78.6	75.9	76.5
45 to 54 years	7,362	7,795	7,830	8,320	8,718	8,808	9,316	58.3	60.3	60.5	64.3	67.1	67.8	71.2
55 to 59 years	2,853	2,703	2,642	2,650	2,791	2,779	3,059	53.3	51.0	49.5	49.7	51.1	51.0	55.3
60 to 64 years	2,150	1,811	1,628	1,826	1,821	1,869	2,016	39.2	33.7	30.1	33.8	32.1	33.0	35.5
65 to 69 years	864	768	649	772	829	705	941	16.7	14.2	11.9	14.1	15.1	12.9	17.0
70 years and older	527	482	394	453	500	433	561	5.0	4.4	3.5	4.0	4.0	3.5	4.8
Whites	--	--	103,751	105,867	107,734	105,467	107,177	--	--	66.9	68.3	67.3	65.9	66.8
Men	--	--	57,185	57,800	59,201	58,524	59,298	--	--	77.4	78.1	77.4	76.5	76.9
Women	--	--	46,586	48,067	48,533	46,943	47,879	--	--	57.4	59.3	58.1	56.2	57.5
Blacks and others	--	--	15,615	16,508	17,217	17,186	17,610	--	--	62.0	65.8	64.8	64.5	63.7
Men	--	--	7,930	8,080	8,500	8,622	8,936	--	--	69.9	71.5	71.0	71.7	71.1
Women	--	--	7,683	8,428	8,717	8,564	8,674	--	--	55.6	61.1	59.7	58.6	57.6

NOTE: Dash indicates data not available.

Appendix Table 2. Difference between the 1990 labor force participation rates and the projections made in 1973, 1976, 1978, 1980, 1983, and 1985

Labor force group	Percentage-point difference						Absolute percentage-point error					
	1973	1976	1978	1980	1983	1985	1973	1976	1978	1980	1983	1985
Total	-4.4	-2.8	-0.2	1.5	0.5	-0.7	6.6	4.2	0.3	2.3	0.8	1.1
Men, 16 and older	3.0	1.2	.3	1.1	.4	-.3	3.9	1.6	.4	1.4	.5	.4
16 and 17 years	1.9	7.1	11.2	10.8	7.3	.7	4.3	16.2	25.6	24.7	16.7	1.6
18 and 19 years	-3.8	4.4	7.4	7.3	6.2	3.2	5.7	6.6	11.0	10.9	9.3	4.8
20 to 24 years	-3.2	-2.2	.7	2.1	.1	2.0	3.8	2.6	.8	2.5	.1	2.4
25 to 34 years	1.2	.5	-.3	.1	-.5	-.1	1.3	.5	.3	.1	.5	.1
35 to 44 years	1.2	.4	.4	.8	1.2	.3	1.3	.4	0.4	.8	1.3	.3
45 to 54 years	1.8	-.5	-1.3	.1	.6	.1	2.0	.6	1.4	.1	.7	.1
55 to 59 years	7.1	1.8	-2.2	-1.1	-1.7	-1.5	8.9	2.3	2.8	1.4	2.1	1.9
60 to 64 years	14.4	2.2	-3.5	.4	-2.7	-5.3	25.9	4.0	6.3	.7	4.9	9.5
65 to 69 years	8.4	.6	-4.8	-2.8	-2.7	-6.0	32.3	2.3	18.5	10.8	10.4	23.1
70 years and older	.8	-.1	.4	.5	-.5	-1.4	7.6	.8	3.6	4.6	4.5	13.0
Women, 16 and older	-11.0	-6.1	-.4	2.1	.8	-.9	19.1	10.6	.7	3.7	1.4	1.6
16 and 17 years	-4.5	5.0	10.2	12.8	4.3	-.5	10.7	11.9	24.3	30.5	10.3	1.2
18 and 19 years	-4.3	2.0	11.6	11.6	6.0	.8	7.1	3.3	19.2	19.2	9.9	1.3
20 to 24 years	-5.3	3.6	8.8	9.8	6.5	2.2	7.4	5.0	12.3	13.7	9.1	3.1
25 to 34 years	-22.0	-10.1	4.5	7.1	4.5	2.6	29.9	13.7	6.1	9.6	6.1	3.5
35 to 44 years	-22.5	-13.5	-2.0	2.1	2.1	-0.6	29.4	17.6	2.6	2.7	2.7	0.8
45 to 54 years	-12.9	-10.9	-10.7	-6.9	-4.1	-3.4	18.1	15.3	15.0	9.7	5.8	4.8
55 to 59 years	-2.0	-4.3	-5.8	-5.6	-4.2	-4.3	3.6	7.8	10.5	10.1	7.6	7.8
60 to 64 years	3.7	-1.8	-5.4	-1.7	-3.4	-2.5	10.4	5.1	15.2	4.8	9.6	7.0
65 to 69 years	-.3	-2.8	-5.1	-2.9	-1.9	-4.1	1.8	16.5	30.0	17.1	11.2	24.1
70 years and older	.2	-.4	-1.3	-.7	-.7	-1.2	5.2	7.5	26.6	15.7	15.1	26.1
Whites	--	--	.1	1.5	.5	-.9	--	--	.1	2.2	.7	1.3
Men	--	--	.5	1.2	.5	-.4	--	--	.7	1.6	.7	.5
Women	--	--	-.1	1.8	.6	-1.3	--	--	.2	3.1	1.0	2.3
Blacks and others	--	--	-1.7	2.1	1.1	.8	--	--	2.7	3.3	1.7	1.3
Men	--	--	-1.2	.4	-.1	.6	--	--	1.7	.6	.1	.8
Women	--	--	-2.0	3.5	2.1	1.0	--	--	3.5	6.1	3.6	1.7
Mean absolute percent error	--	--	--	--	--	--	10.8	7.0	11.6	9.5	6.9	6.8

Appendix Table 3. **Distribution of the 1990 labor force and as projected in 1973, 1976, 1978, 1980, 1983, and 1985**

Labor force group	Distribution as projected in --							Percentage-point difference from 1990					
	1973	1976	1978	1980	1983	1985	1990	1973	1976	1978	1980	1983	1985
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0						
Men, 16 and older	60.5	57.3	54.6	53.8	54.2	54.7	54.7	5.9	2.6	-0.1	-0.8	-0.5	0.1
16 and 17 years	1.4	1.4	1.5	1.4	1.3	1.2	1.2	.2	.2	.3	.2	.1	.0
18 and 19 years	2.0	2.1	2.1	2.0	2.0	1.9	1.9	.0	.2	.1	.1	.1	.0
20 to 24 years	5.8	5.9	5.8	5.8	5.7	6.0	5.8	.0	.0	.0	-.1	-.1	.1
25 to 34 years	17.5	16.3	15.4	15.1	15.7	16.0	15.9	1.7	.4	-.5	-.8	-.2	.2
35 to 44 years	15.5	14.6	13.9	13.6	14.0	14.1	13.8	1.7	.7	.1	-.2	.1	.3
45 to 54 years	9.8	9.6	9.1	9.0	8.9	9.0	9.0	.9	.6	.1	.0	.0	.1
55 to 59 years	3.7	3.5	3.2	3.2	3.1	3.1	3.2	.5	.3	.0	.0	-.1	-.1
60 to 64 years	2.9	2.4	2.1	2.2	2.1	2.0	2.2	.7	.2	-.1	.0	-.2	-.2
65 to 69 years	1.2	1.0	.8	.8	.8	.7	1.0	.3	.0	-.2	-.1	-.1	-.2
70 years and older	.7	.6	.7	.7	.6	.6	.7	.0	.0	.0	.0	.0	-.1
Women, 16 and older	39.5	42.7	45.5	46.2	45.8	45.3	45.3	-5.9	-2.6	.1	.8	.5	-.1
16 and 17 years	1.1	1.3	1.3	1.4	1.2	1.1	1.1	.0	.2	.3	.3	.1	.0
18 and 19 years	1.8	1.9	2.1	2.1	1.9	1.7	1.8	.0	.2	.4	.3	.1	.0
20 to 24 years	5.3	5.8	5.9	5.8	5.6	5.4	5.3	.0	.6	.7	.6	.4	.2
25 to 34 years	9.6	11.5	13.5	13.5	13.4	13.3	12.8	-3.2	-1.3	.6	.7	.6	.5
35 to 44 years	9.2	10.3	11.6	11.9	12.0	11.8	11.7	-2.4	-1.4	-.1	.2	.3	.1
45 to 54 years	6.7	6.8	6.6	6.8	7.0	7.2	7.5	-.8	-.6	-.9	-.7	-.5	-.3
55 to 59 years	2.6	2.4	2.2	2.2	2.2	2.3	2.5	.1	-.1	-.2	-.3	-.2	-.2
60 to 64 years	1.9	1.6	1.4	1.5	1.5	1.5	1.6	.3	.0	-.3	-.1	-.2	-.1
65 to 69 years	.8	.7	.5	.6	.7	.6	.8	.0	-.1	-.2	-.1	-.1	-.2
70 years and older	.5	.4	.3	.4	.4	.4	.4	.0	.0	-.1	-.1	.0	-.1
Whites	--	--	86.9	86.5	86.2	86.0	85.9	--	--	1.0	.6	.3	.1
Men	--	--	47.9	47.2	47.4	47.7	47.5	--	--	.4	-.3	-.1	.2
Women	--	--	39.0	39.3	38.8	38.3	38.4	--	--	.7	.9	.5	-.1
Blacks and others	--	--	13.1	13.5	13.8	14.0	14.1	--	--	-1.0	-.6	-.3	-.1
Men	--	--	6.6	6.6	6.8	7.0	7.2	--	--	-.5	-.6	-.4	-.1
Women	--	--	6.4	6.9	7.0	7.0	7.0	--	--	-.5	-.1	.0	.0
Dissimilarity index	--	--	--	--	--	--	--	6.4	3.6	2.6	2.5	1.8	1.5

NOTE: Dash indicates data not available or not applicable.

Evaluation of Labor Market Forecasts: Comments

Herman O. Stekler, Industrial College of the Armed Forces, National Defense University

These papers are three very professionally executed evaluations of labor market forecasts. Consequently most of my comments will be extremely positive. I will, however, note that these are long-run forecasts and that evaluation procedures designed specifically for evaluating the important characteristics of these types of predictions must be developed and then implemented. To the extent that such evaluation procedures have been available, these studies have utilized them, but not in as prominent a way as I, personally, would have liked.

First, what should an evaluation do? It should describe the forecasts and outcomes, examine the errors and if possible determine the sources of the errors. The last step is essential if subsequent forecasts are to be improved. Evaluations should also compare the results against some standard. This is useful to determine whether there is any value to the forecasts regardless of the magnitude or sources of the errors.

All three papers do most of these things. Thus we know how the forecasts were generated, the data problems, the issues involved in replicating the results, etc. The statistical procedures were all appropriate. In addition we discover that the forecasts (in most cases) improve the shorter the forecasting horizon, that errors in forecasting output produce misestimates of industry employment which in turn yield occupational errors. We learn that errors in projecting population and participation rates influence the labor force estimates, etc.

While self-criticism is helpful in analyzing forecasting performance, these authors may have been too harsh on themselves. The problem is that these evaluations do not consistently compare the BLS published results against some standard. (Fullerton does compare the projected growth rates of the labor force against a naive standard, but doesn't undertake a full analysis).

To provide evaluation standards for such long-run forecasts it is first necessary to determine what the desirable characteristics of such projections might be. Since the profession has not yet settled on a list of desirable characteristics, what is presented here is one person's opinions. Consequently, a long-run forecast should provide a picture of the structure of the labor market at a distant point in the future. Thus the accuracy of quantitative forecasts of individual industry or occupational growth rates is not as relevant as is an accurate depiction of major trends or basic structural changes.

Thus the use of Theil's information statistic on Spearman's correlation coefficient of the rankings between those industries which were expected to grow the fastest and those which actually did provide such pictures of the state of the economy. Some of the evaluations use these measures, but perhaps not as prominently as I would have advocated.

However by not showing how badly the naive method would have performed using these same evaluation measures, the BLS economists may have understated the value of their own projections.
Yes, as the BLS economists indicate, structural changes are hard to predict. However, naive methods cannot predict these changes at all. So, any improvement over the naive methods should be considered an accomplishment.

The Recent Consensus Forecasting Record for Monthly Macroeconomic Indicators

Jennifer L. Beattie, Economic Research Service, U.S. Department of Agriculture
John Kitchen, Council of Economic Advisers
R. M. Monaco, Economic Research Service, U.S. Department of Agriculture

The authors would like to thank Annette Dargan for her careful attention in collecting and posting the data.

Introduction

Economic forecasts are an important component of government and business policy-making. Forecasts spanning several quarters or years influence decisions about current Federal tax and spending proposals. Businesses use these longer term forecasts to support strategic planning for marketing and product development. Forecasts spanning shorter periods are also sometimes used by government to help make policy decisions, while businesses use them to make market-timing and investment decisions. Short-term forecasts are especially important to financial market participants like banks and brokerage houses, which operate in markets where conditions are quick to change and where small deviations from expectations can result in large profit making (or loss making) opportunities.

The role that expectations play in the functioning of the economy has been a key focus of economic theory over the past several decades. Muth (1961) originally advanced the concept of rational expectations which eventually became the basis of an entire school of thought (see Lucas and Sargent). According to Muth, an expectation is rational if the subjective expectation of a variable is equal to its mathematical expectation. Rational expectations are optimal predictions that incorporate all available information that could affect the distribution of possible outcomes. Hence, expectations that are rational have two key properties: (1) unbiasedness; and, (2) efficiency. In practice, a measure of central tendency--the mean or median--of the distribution of the subjective expectations of market participants (in this case, a consensus forecast) serves as a proxy for the rational expectation of a variable.

In this paper we focus on whether forecasts are unbiased, and, specifically, test whether short-term, consensus forecasts of monthly macroeconomic indicators were unbiased over the past four and a half years. Examining forecast performance during this period is particularly interesting because it contains business cycle turning points, during which economic forecasting is unusually difficult. The results provide evidence that consensus forecasts were unbiased for about half of the indicators we examined. The evidence for the other indicators does not strongly support the unbiasedness hypothesis. In addition, when we tested for changes in the relationships associated with different phases of the business cycle, significant differences were observed for forecasts of labor market indicators. In particular, while market forecasts of nonfarm payroll jobs were unbiased in the late expansion period before July 1990, the forecasts were biased upward in the recession/recovery period following July 1990. This suggests that, even on a very short-term basis, forecasts of the recovery tended to be overly optimistic.

The Data

We collected two data series for 10 monthly indicators: the consensus forecasted value from MMS International, Inc. and the actual value. Each week, MMS surveys economists about their forecasts for indicators that will be released during the next week. The median of these responses is then reported as the MMS consensus value. The forecasted values are very short-term forecasts of the monthly series, that is, they are made only a few days before the statistics are released.

The flow of macroeconomic data released during the month follows a regular pattern. The data released during a month refer either to the month immediately preceding the month in which the data are released, or 2 months prior to the month in which the data are released. For example, the unemployment rate for June is released on the first or second Friday of July, but the merchandise trade release in July contains May data.

The actual values used are those that are reported at the time the data are initially released. It is important to emphasize that the actual series to which we compare the forecasts does not contain revisions made after the initial release. Thus, for example, while the volume of retail sales for any month is revised several times in the months subsequent to its initial release, we compare the forecasts only to the initial release. In general, the data series begin in early 1988 and extend through the middle of 1992. The list of monthly indicators used are shown in Table 1, listed in the order in which they are typically released each month with a description of what is forecasted (level, percentage change from the previous level, or change from the previous level), and summary statistics for the actual and forecast series.

Table 1 also shows that there is a considerable difference in the volatility in the actual data series. For example, among the series measured in percent changes, the standard deviation of the percent change in advance durable goods orders is more than 15 times its mean, while the standard deviation of the percent change in consumer prices is only about 2/3 its mean. It is also worth noting that the standard deviations of the forecasts are uniformly smaller than the standard deviations of the actuals. This conforms with the notion that the variance of an optimal forecast is smaller than the variance of the series being forecast (see Granger and Newbold, page 131 for details).

Statistical Test 1: Were The Forecasts Unbiased over the Entire Sample Period?

To test for unbiasedness, we ran the following regression:

$$\text{Actual} = a + b(\text{Forecast}) + \text{error}$$

and tested the joint hypothesis that $a = 0$ and $b = 1$. Assuming the errors are normally distributed, the test statistic is distributed F with 2, N-2 degrees of freedom, where N is the number of observations.

Table 2 lists the results of running the regressions and testing the joint hypothesis for each monthly indicator. The table reports the value of the F statistic and the p-value (also known as the prob-value). The p-value is the probability of obtaining the value of the reported F statistic given that the joint hypothesis, $a=0$ and $b=1$, is true. High F values are consistent with low p-values. P-values can be interpreted as a measure of the plausibility of the null hypothesis (see, for example, Wonnacott and Wonnacott, pages 246-255 for details). Thus, a low p-value suggests that the $a = 0$, $b = 1$ hypothesis is not very credible for that indicator. In addition to the F statistics and p-values, Table 2 reports the R-Squared, a measure of the goodness of fit of the regression. Ideally, the forecasts should not only be unbiased, but they should account for most of the variation in the actuals series.

Discussion

According to the test statistics, the null hypothesis that the forecasts of monthly macroeconomic indicators are unbiased is fairly credible. If we had adhered to a critical value of 0.01, the null hypothesis would have been accepted for all of the indicators except producer prices, where the p-value is well below 0.01 (0.01 is a typical choice for the size of the Type I error; the probability of rejecting the unbiasedness hypothesis given that it is true). Eight of the ten regression intercepts were negative, while seven of the ten slope coefficients were greater than 1. One interpretation of the regression coefficient on the forecast series is that it shows how much a change in the forecast correctly reflects a change in the actual. When the coefficient is greater than 1, it suggests that a change in the forecast must be "blown up" to correctly predict the full change in the actual. In six of the ten regressions, the forecast series accounted for at least 73 percent of the variation in the actual data.

The evidence that forecasts of monthly percent changes in producer prices are biased is strong (p-value 0.004). Somewhat surprisingly, given the producer-price results, the F statistic for consumer prices does not reject unbiasedness, and has the highest p-value of all the indicators (0.961). The fit of the producer price relationship is somewhat better than the fit of the consumer price relationship, however. Differences in the construction of the indicators may partly explain these results.

First, although they are both measures of aggregate price changes, producer prices are more affected by commodity price movements than are consumer prices. For example, the overall producer price index for finished goods contains no component for services, which are mostly wage costs and are much less volatile than goods prices, while consumer prices are heavily influenced by movements in services prices. Also, consumer prices are sampled throughout the month in several cycles with an approximately equal amount of the index prices in each of the cycles. In contrast, producer prices are collected from one day in the middle of the month. The averaging that results for consumer prices compared to the single observation day for producer prices also likely increases the volatility of producer prices relative to consumer prices. The standard deviation of producer price changes is more than twice that of consumer price changes.

P-values were relatively low for the leading index, the change in the number of nonfarm payroll jobs, the percent change in industrial production, and the percent change in durable goods orders. This suggests that unbiasedness is not very credible for these indicators, even though if one were using the conventional critical value of 0.01 the null hypotheses of unbiasedness could not be rejected. In the case of the leading index, the fit of the regression is particularly good, suggesting that the forecast series captures nearly 90 percent of the variation in the actual series. For the change in nonfarm payroll jobs, the forecast series accounts for less than 60 percent of the variation in the actual series.

The relationship between the actual and forecasted values of the percent change in durable goods orders is very noisy, that is, the fit is relatively poor. The percent change in durable goods orders is very volatile (the average percent change over the sample period is 0.3 percent but the standard deviation is over 4 percentage points). Orders for aircraft and transportation equipment fluctuate considerably and it is sometimes difficult to predict in which month an order will be counted. The results for retail sales present an interesting contrast with durable goods orders. While both have fits of less than 0.5, the p-value for durable goods is fairly low (0.070), the p-value for retail sales is reasonably high (0.417). This suggests that while forecasts of the percent change in retail sales are unbiased, they are not very reliable.

In sum, for 5 of the 10 indicators examined, there appeared to be relatively strong evidence that consensus forecasts were unbiased. For the 5 other indicators, the evidence is somewhat mixed. Although in 4 of the latter 5 cases the null hypothesis could not be rejected at the 0.01 level, the relatively low p-values suggest that the unbiasedness hypothesis is not well supported by the data. There appears to be little relationship between good fits and unbiasedness.

One useful implication of rejecting the null hypothesis that the consensus forecast is unbiased is that a better forecast could be obtained by combining the consensus forecast with the estimated regression equation. It appears that, for producer prices and the leading index especially, the estimated equations are probably more useful to use than just the consensus forecast alone.

Statistical Test 2: Is There a Discernible Difference Between the End of an Expansion and a Recession\Recovery Period?

We also tested whether there was a discernible difference in the relationship between forecasted and actual values before and after the onset of the recent recession. To do this, we split the sample into pre-recession and post-recession parts and tested for a statistically discernible difference in the relationship. We used tests for structural change in a regression equation, also known as Chow tests. When the error terms in the regressions are normally distributed, the test statistics are distributed F with K and N degrees of freedom, where K is 2 and N is the number of observations less 4 (see Gujarati, page 444 for details). We split the sample at two alternative points to test for possible recession effects. In the first case, we split the sample at July 1990, which the National Bureau of Economic Research (NBER) has dated as the beginning of the recession. In the second case, we split the sample at April 1991, when the NBER announced that the recession began in July 1990. The choice of April 1991 is also interesting, since at least some observers suggest it is around the likely end of the recession (see for example, the Congressional Budget Office report, page 4). Thus, overall, our tests are aimed at discovering whether the onset of the recent recession affected unbiasedness, or whether the announcement of the recession or the recognition of its likely end affected unbiasedness. Table 3 shows the results of these tests.

Discussion

For most of the indicators, the onset of recession or the announcement of the recession's beginning point did not appreciably change the statistical relationship between actual and forecasted values. Specifically, for all 10 indicators, the null hypothesis that the pre- and post-recession relationship was the same could not be rejected at the 0.01 level. However, the p-values were relatively low for the change in nonfarm payroll jobs and the unemployment rate using the July 1990 sample split, and the unemployment rate using the April 1991 split.

For the change in nonfarm payroll jobs, there is fairly strong evidence that the relationship between the actual and forecasted values differs between the two sub-sample periods. In Table 3, the p-value of the Chow test is 0.014. In regressions run over the two sub-periods, the p-value of the null hypothesis of unbiased forecasts was 0.321 for the late expansion period, but fell to 0.018 in the recession/recovery period. In the latter regression, the consensus forecast tended to overestimate the actual changes. Figure 1 also illustrates the tendency to over-predict, showing that most of the forecasts of the change in nonfarm payroll jobs over-predicted the subsequent actual change over from July 1990 through July 1992.

For the unemployment rate, the unbiasedness test results are opposite the nonfarm payroll job results. The null hypothesis of unbiased forecasts has a p-value of 0.035 in the pre-recession period, but a p-value of 0.821 in the recession/recovery period. A similar result was found using the April 1991 sample split.

Conclusions

There are several conclusions that can be drawn from these tests.

- For 5 monthly indicators, the evidence that consensus forecasts are unbiased is relatively strong. Those indicators are: the unemployment rate, consumer prices, retail sales, merchandise trade, and nonfarm payroll jobs.

- For the 5 remaining indicators, there is evidence that the consensus forecasts were biased over the sample period. These indicators are: the leading index, producer prices, industrial production, housing starts, and durable goods orders. In general, there appeared to be little relationship between the goodness of fit of the regression and evidence for unbiasedness.

- For 8 of the 10 indicators studied, there was little evidence to suggest that the relationship between the forecast and actual values changed with the onset of recession. The exceptions were the unemployment rate and the change in the number of nonfarm payroll jobs. The statistical evidence for such a shift occurring was strongest for the change in nonfarm payroll jobs.

Directions for Future Research

The tests described here are the first results using the data that we have collected and are continuing to collect. Currently, we plan to address at least 3 other issues. First, we intend to investigate whether the monthly forecasts are efficient. That is, we intend to test whether the forecasts incorporate all the available information, or whether they can be improved on by accounting for other data that was also available at the time the forecast was made. The results of this paper suggest that at the 0.01 significance level, we can already rule out efficiency in producer price forecasts, since they are biased. Secondly we will address the issue of data revisions. It is possible that forecasters "look through the revisions" when forecasting the indicator. In that case, a better test of forecast unbiasedness would be conducted using the most recently available actual series, as opposed to the initial release. Thirdly, we intend to try to measure the effects that unexpected developments -- forecast errors -- had on interest rates, exchange rates, and commodity prices during over the last four-and-a-half years.

Table 1. Description of the Data

Indicator	When Released	Form of Forecast	Data Span	Averages		Standard Deviation	
				Actual	Forecasts	Actual	Forecasts
Nonfarm payroll employment	1st or 2nd Friday of month	Monthly change in thousands	3/4/88-7/2/92	95.38	115.04	187.87	126.90
Civilian unemployment rate	1st or 2nd Friday of month	Percent of labor force	1/8/88-7/2/92	5.93	5.94	0.77	0.75
Producer prices for finished goods	Near middle of month	Percent change from previous month	1/15/88-7/10/92	0.28	0.30	0.51	0.31
Consumer prices	Near middle of month, after PPI is released	Percent change from previous month	1/20/88-7/14/92	0.35	0.36	0.21	0.16
Retail Sales	Middle of month	Percent change from previous month	1/14/88-7/14/92	0.19	0.27	0.62	0.44
Industrial production	Middle of month	Percent change from previous month	1/15/88-7/15/92	0.09	0.10	0.52	0.36
Merchandise Trade Balance	Middle of month	Value of surplus in billions of $	1/15/88-7/17/92	-8.48	-8.61	2.67	2.55
Housing starts	Near end of month	Number of units started in millions	1/20/88-7/16/92	1.27	1.28	0.20	0.20
Durable goods orders, advance release	End of month	Percent change from previous month	1/26/88-7/24/92	0.27	0.10	4.24	1.65
Index of Leading Indicators	End of month or beginning of next month	Percent change from previous month	3/1/88-6/30/92	0.11	0.10	0.65	0.53

Table 2. Are Consensus Forecasts Unbiased? Testing for a=0, b=1 in Regression Equation: Actual = a + b(Forecast)

Indicator	Number of observations	Data Span	Intercept (a)	Slope coefficient (b)	F Value (a=0,b=1)	P-value	R-Squared
Nonfarm payroll	53	3/4/88-7/2/92	-33.11	1.112	2.10	0.133	0.57
Unemployment rate	55	1/8/88-7/2/92	-0.066	1.010	0.10	0.905	0.96
Producer prices	55	1/15/88-7/10/92	-0.141	1.404	6.05**	0.004	0.73
Consumer prices	55	1/20/88-7/14/92	-0.012	1.026	0.04	0.961	0.61
Retail sales	55	1/14/88-7/14/92	-0.072	0.963	0.89	0.417	0.46
Industrial production	55	1/15/88-7/15/92	-0.033	1.227	2.57	0.086	0.73
Merchandise Trade Balance	55	1/15/88-7/17/92	-0.581	0.917	1.00	0.375	0.77
Housing starts	55	1/20/88-7/16/92	0.088	0.928	0.78	0.460	0.82
Durable goods orders	54	1/26/88-7/24/92	0.105	1.634	2.74	0.070	0.41
Leading Index	53	3/1/88-6/30/92	-0.007	1.160	3.61*	0.034	0.88

* Significant at the 0.05 level   ** Significant at the 0.01 level

T ble 3. Are Forecasts Made During Recession\Recovery Different Than Those Made During Expansions?

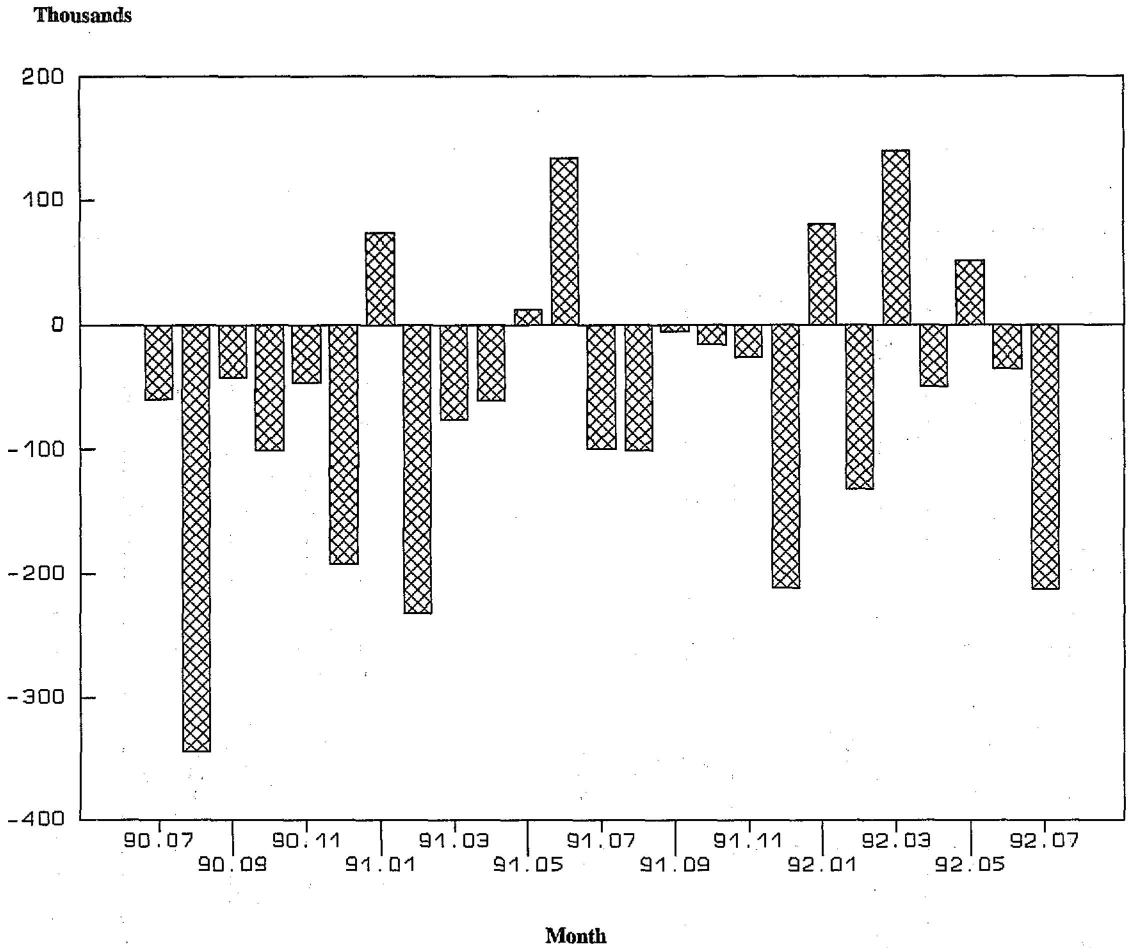
	Sample Split after July 1990		Sample Split after April 1991	
Indicator	F Statistic	P-value	F Statistic	P-value
Nonfarm payroll jobs	4.69*	0.014	1.06	0.354
Unemployment rate	2.91	0.064	2.50	0.092
Producer prices	1.02	0.368	0.83	0.442
Consumer prices	0.58	0.564	1.16	0.322
Retail sales	0.83	0.442	0.78	0.464
Industrial production	0.39	0.679	0.12	0.887
Merchandise Trade Balance	1.20	0.310	0.46	0.634
Housing starts	0.45	0.640	0.09	0.915
Durable goods orders	0.09	0.914	0.62	0.542
Leading Index	0.18	0.836	0.38	0.686

Table 4. Are Forecasts Unbiased In Sub-samples?

Indicator	Sample Split after July 1990	
	P-value Pre-recession	P-value Recession/recovery
Nonfarm payroll jobs	0.321	0.018*
Unemployment rate	0.035*	0.821

* Significant at the 0.05 level ** Significant at the 0.01 level

Figure 1
NONFARM PAYROLL JOBS FORECAST ERROR
July 1990 - July 1992



References

Congressional Budget Office, The Economic and Budget Outlook: An Update, August 1992.

Cook, Timothy and Steven Korn, "The Reaction of Interest Rates to the Employment Report: The Role of Policy Anticipations," Economic Review. Federal Reserve Bank of Richmond, (Sept/Oct 1991). p. 3-12.

Granger, C. W. J. and P. Newbold, Forecasting Economic Time Series, Academic Press Inc., Orlando, 1986.

Gujarati, Damodor N., Basic Econometrics, McGraw-Hill Inc., New York, 1988.

Lucas, Robert E. and Thomas J. Sargent (editors), Rational Expecations and Econometric Practice Volumes 1-2, University of Minnesota Press, Minneapolis, 1981.

Muth, John F., "Rational Expectations and the Theory of Price Movements", Econometrica, Vol. 29, Number 6, (1961). pp. 315-335.

Wonnacott, T. H. and R. J. Wonnacott, Introductory Statistics for Business and Economics (2nd edition), John Wiley & Sons, New York, 1977.

The Accuracy of USDA's Export Forecasts

Stephen MacDonald, Economic Resarch Service, U.S. Department of Agriculture

Abstract. USDA's quarterly forecasts of fiscal year agricultural exports by commodity and region were examined for their reliability in predicting annual changes during 1977-89. Most of the forecasts were strongly correlated with actual exports. Most obvious exceptions probably stemmed from rounding errors. Bias was not a problem for the forecasts of total exports in any quarter, nor for most of the commodity forecasts. There was some upward bias in the forecasts for less developed countries, and downward bias for some developed countries. The USDA forecasts were conservative; they were more likely to underestimate the magnitude of change than to overestimate it.

The U.S.Department of Agriculture's (USDA) short-term forecasts are probably the most widely disseminated agricultural forecasts in the world. They are generally believed to be accurate, and, while specific forecasts are occasionally questioned, they remain the benchmark against which alternative forecasts are compared. The accuracy of USDA's forecasts is, therefore, of vital concern. This report examines the accuracy of the fiscal year forecasts of U.S. agricultural exports that USDA publishes quarterly in its Outlook for U.S. Agricultural Exports.

In this report, USDA's forecasts were determined to be upwardly biased for cotton (value), Eastern Europe, South Asia, East and Southeast Asia, the Middle East, and North Africa. Forecasts were determined to be downwardly biased for rice (volume), animal fats, sugar and tropical products, the USSR, and Latin America. USDA's forecast for livestock products was also biased downward, but that was probably due to a tendency to underestimate change coinciding with a period of rapidly rising exports. Although the result of this tendency was a set of forecasts that averaged significantly lower than actual exports, such bias would probably not have appeared if livestock exports had trended downwards or tended not to grow. A tendency to underestimate change was also detected in USDA's forecasts of grain exports and total agricultural products.

It is desirable to eliminate systematic errors from any forecast, but increasing forecast reliability is likely to entail costs. Any desire to improve forecast accuracy must be balanced by considerations of how costs compare with benefits. In any case, the first step is discovering systematic errors. Unforeseeable events will always result in some forecast error, but when errors fall into discernible patterns, they represent behavior that can be altered to improve forecast accuracy.

USDA's Forecasts

Each month USDA publishes forecasts of the annual levels of production, consumption, trade, and stocks for key commodities and countries. The forecasts are produced by an interagency process that includes the Agricultural Marketing Service (AMS), the Agricultural Stabilization and Conservation Service (ASCS), the Economic Research Service (ERS), the Foreign Agricultural Service (FAS), National Agricultural Statistics Service (NASS), and the World Agricultural Outlook Board (WAOB). Annual U.S. average farm prices are also forecast for key commodities each month. These are forecasts of annual totals on crop marketing years or calendar years, and they are published in the World Agricultural Supply and Demand Estimates.

Every 3 months USDA publishes forecasts of U.S. fiscal year exports for a smaller group of commodities in the Outlook for U.S. Agricultural Exports. Forecasts for most of these commodities are also published monthly, but some forecasts are published only quarterly. Total agricultural export value and volume forecasts for the United States are available only quarterly. The same is true of all the other forecasts of annual export value. These include forecasts of the total value of U.S. agricultural exports to selected countries and regions during the current fiscal year and forecasts of commodity export values. The forecasts analyzed included 15 different commodities by value, 10 commodities by volume, and 21 regional aggregations. Forecasts published during 1977-89 were studied.

Quarterly forecasts are produced through an interagency process, but with only ERS, FAS, and the WAOB participating. As much as possible, these quarterly forecasts are intended to be consistent with the most recent monthly forecasts, which precede the quarterly forecasts by a few weeks. Quarterly forecasts are consistent with the monthly forecasts even when more current information indicates that conditions determining the monthly forecasts have changed. This is unusual and is acknowledged in the Outlook for U.S. Agricultural Exports when necessary. Updates are published between quarters if circumstances warrant. These updates were not included in this analysis.

The first forecast of the fiscal year is published in late November or early December when no official trade data are available (table 1). A revised forecast is published 3 months later in late February with actual export data available for the first 3 months of the fiscal year. When the next revised forecast is published, 6 months of actual data are available, and when the final forecast is published, 9 months of actual data are available. In other words, the final forecast of total fiscal year exports is actually made with only 3 months of exports unknown. These later forecasts are therefore very accurate.

Table 1--Timing of USDA's quarterly export forecasts

Order of publication	Month published	Export data available
First quarter	November or December	None
Second quarter	February	October-December
Third quarter	May	October-March
Fourth quarter	August	October-June

An important point is that each quarter's forecasts are conditional on different information sets. Throughout this report, four sets of quarterly forecasts are compared. These are not necessarily forecasts made through different processes (for example, using different models) but forecasts made with increasing amounts of information. The question of what sort of models produce the forecasts is a good one. In general terms, the process is best described as using a "delphic" method, averaging the judgment of a large number of experts.

Forecast Analysis

In this report, USDA's quarterly forecasts of fiscal year exports were tested by regressing actual exports (A_i) on export forecasts (F_i). These regressions yielded measures of correlation that are used to measure efficiency. The regressions also provided estimated coefficient values that were tested to determine bias and consistency. Regression analysis was used rather than decomposition of mean squared error (MSE) because of regression's superior ability to separate the effects of bias and consistency (see Appendix). Regression analysis also lends itself to statistical testing to determine the significance of the results.

"Efficiency's" general econometric meaning refers to an estimator's "spread" around its expected value. In this report, the term efficiency is used in a somewhat similar sense and is measured by the correlation between a forecast (F) and its actual variable (A). Correlation is used rather than a measure of variance between F and A ($\sigma_{F\text{-}A}$) in order to allow comparison between different forecasts because correlation is always between 0 and 1 regardless of the magnitude of the variables. Correlation is also affected by the randomness of the spread. A forecast with low $\sigma_{F\text{-}A}$ might be less accurate than one with a high $\sigma_{F\text{-}A}$. To completely understand forecast error, knowledge of the pattern of error is as critical as that of its size.

Bias refers to whether we can expect the forecast to exceed the actual variable or perhaps vice versa. The expected value of the difference between the forecast and the actual variable must be zero for the forecast to be unbiased:

$$E(F - A) = 0$$

Consistency generally refers to the asymptotic property of an estimator: the increasing accuracy of an estimator's ability to approximate the parameter as sample size increases. In this report, consistency also refers to a parameter value. Consistency is perhaps best understood here as the forecast's ability to predict the magnitude of change in a variable. Consistency refers to a parameter value because when a series of annual forecasts are compared with the respective actual values of the variable through a regression equation,

$$dA_i = \alpha + \beta dF_i + \varepsilon_i ,$$

then a consistent forecast is one where the estimated $\beta = 1$. Inconsistency and bias may be indistinguishable in some cases. If an otherwise perfect forecast were decreased by 10 percent every year, there would be a downward bias equal to 10 percent of the average value of the actual data over the sample period, and β could equal 1.11.

Consistency and bias also determine if a forecast is rational. Forecasts have been used to measure expectations and tested to see if these expectations conform to the rational expectations hypothesis. If a test of α and β estimated in the above regression $\alpha = 0$ and $\beta = 1$ cannot be rejected, then the forecast is described as rational. This is a weak-form test for rationality (7).[1] This report does not explicitly explore whether USDA's export forecasts are rational, but all the forecasts described in this report as either biased or inconsistent also fail the above weak-form test for rationality.

In this report, most of USDA's forecasts examined were determined to be efficient. The exceptions were largely confined to commodities or regions where exports were so small that rounding played a significant role in determining the forecast. In the first quarter (the forecasts published in November or December), only two forecasts had average errors above 22 percent, and by the last quarter (published in August) the forecast for total U.S. exports was wrong by less than 2

[1]Underscored numbers in parentheses are cited in References.

percent on average. Few commodity forecasts appear biased, but both upward and downward bias were more common in the regional forecasts.

Some key forecasts showed signs of inconsistency. The magnitude of change in total U.S. export value was typically underestimated. The same was true of grain exports, particularly when exports were falling. Underestimating the magnitude of change was more common than overestimating it.

Methodology

Mean squared error (MSE) is perhaps the most frequently used measure of forecast accuracy. It is particularly appealing when comparing various models predicting a common dependent variable. If the various models' equations are estimated using ordinary least squares (OLS), then each model's parameters will be estimated so as to minimize the sum of squared errors (SSE). Taking a mean of the squared errors simply normalizes SSE by sample size:

$$MSE(F_i) = \frac{1}{N}\sum_{i=1}^{N}(F_i - A_i)^2$$

where F is the series of forecasts and A the actual data.

My goal is to measure the reliability of forecasts of many different variables rather than one variable with different models; MSE is not an appropriate statistic because the MSE of a variable averaging $30 billion would not be comparable to the MSE of a variable averaging $300 million. However, MSE can be "decomposed" into components that provide more specific characterizations of forecast reliability. The simplest decomposition,

$$MSE(F) = (\overline{F} - \overline{A})^2 - \sigma^2_{(F-A)}$$

separates MSE into a statistic measuring bias and another measuring the variance of the forecast errors. The effect of bias on forecast accuracy is clear, but the inevitable variation of the errors can take many forms. Therefore, a simple measure of forecast error variance is inadequate in characterizing reliability. Also, the variance of forecast errors (as with MSE) is not independent of the magnitude of the variables in question.

Correlation, however, is independent of magnitude. Granger and Newbold (3) demonstrate through a further decomposition of MSE that correlation between a series of forecasts and a series of matching actual data is a good measure for analyzing these further errors in variation.

$$MSE(F) = (\overline{F} - \overline{A}) + \sigma^2_F + \sigma^2_A - 2\rho_{FA}\sigma_F\sigma_A$$

$$\sigma_j = \textit{standard deviation of } j$$

$$\rho_{FA} = \textit{correlation of } A \textit{ and } F$$

This equation shows that MSE(F) is minimized by a smaller bias and a larger correlation. Equivalence between the two series' variances would not minimize MSE(F), except when bias is zero and correlation is perfect:

$$\frac{\partial MSE(F)}{\partial \sigma_F} = 2(\sigma_F - \rho_{FA}\sigma_A)\text{, and}$$

$$\sigma_F = \rho\sigma_A$$

Kost (6), Maddala (8), and others point out that correlation is a poor measure of forecast reliability because it does not account for bias or some other systematic linear error. Correlation, however, is not the sole measure of reliability in this report, but it is combined with further measures of bias and consistency.

The accuracy of USDA's quarterly forecasts of U.S. agricultural exports by commodity and region were analyzed by regressing series of actual data on their respective forecasts.

$$dA_i = \alpha + \beta dF_i + \varepsilon_i$$

These regressions yielded coefficients of determination (R^2) that were used to measure the efficiency of the forecasts and coefficient values that were tested to determine bias and consistency.[2] As the "d" preceding A_i and F_i implies, forecasts were examined as forecasts of the amount of change.

Expressing the forecast as a difference removes the effect of a trend that is irrelevant to understanding the pattern of errors in these particular forecasts. Long-term trends will always be embodied in the F_i forecasts. However, the forecasts studied here are always for one period ahead. Therefore any trends of the preceding years will spuriously appear to be correctly forecasted. That is, including long-term trends will raise measured forecast accuracy to no useful end.

Efficiency has a more restrictive meaning in this report. Efficiency is measured with little reference to the regression coefficients, using correlation. Consistency refers to the ability to correctly forecast the magnitude of change and is determined by testing $\beta = 1$.

Another summary statistic presented for each forecast series is mean absolute percent error (MAPE) (tables 2-4). MAPE provides comparison between forecasts of series with different average values. The errors are put into absolute values to ensure that over- and under-estimates do not offset into a mean of zero. Squaring errors for a MSE serves the same purpose, and a square root of a mean squared percent error (RMSPE) would provide a statistic similar to MAPE. The difference between RMSPE and MAPE is that RMSPE gives larger weight to larger errors.

Results

Examination of 48 series of U.S. agricultural exports forecast by USDA found that forecasts made during the first quarter have MAPE generally ranging from 6 to 22 percent (tables 2-4). Two forecasts exceeded this range, the forecasts of U.S. exports to China and the former USSR, with MAPE of 66 and 37 percent, respectively. China and the former USSR were generally the world's largest grain importers during 1977-89. They were also the world's largest grain producers, and their imports fluctuated widely with shortfalls in production. The first quarter's forecasts are published well in advance of the period when grain supplies in these countries are determined for the year and, therefore, well in advance of events determining their highly variable demand for grain. A relative lack of reliable information concerning events within these countries also hampered forecast accuracy during the period studied.

Much of the 37-percent average error in the forecasts for the former USSR stems from a partial U.S. embargo on grain sales to the Soviets following the Soviet invasion of Afghanistan. If that year were excluded, the average first quarter error would be 27 percent, much closer to the normal range. This illustrates an important point about forecast accuracy. A forecast can be inaccurate because either the forecaster does not understand how circumstances can affect trade, or because these circumstances change. The former error should be avoided, while the latter is often unavoidable.

The first-quarter forecast of total export value tended to be off by 10 percent, and volume by 8 percent. By the last quarter, the forecast for total export value had a 1.4 percent MAPE, and the forecasts' MAPE's for most major commodities were below 5 percent. The regional forecasts were only slightly less accurate. Most of the forecasts for individual commodities and regions were only slightly less accurate than the forecasts for total value and volume. Better accuracy in the total forecasts is not surprising because they are aggregations of the individual commodity forecasts. Offsetting errors among the commodity and regional forecasts probably improve the accuracy of the totals.

The accuracy of the forecasts in most cases improved with each subsequent quarter, in other words,

$$MAPE_1 > MAPE_2 > MAPE_3 > MAPE_4$$

and,

$$R^2_1 < R^2_2 < R^2_3 < R^2_4.$$

The only exceptions were the value forecasts for rice, dairy products, Oceania, and Canada, and the volume forecasts for animal fats, tobacco, and rice. The exceptions number less than 10 percent of all forecasts examined.

Most of these exceptions are probably the result of random rounding errors. Forecasts are published in rounded numbers: to the nearest $100 million and 100,000 tons. If a commodity's exports never vary by more than these amounts, forecast errors are inevitable. Tobacco volume forecasts were particularly vulnerable: exports were always between 200,000 tons and 300,000 tons during 1977-89, often closer to 250,000 tons than to any publishable forecast. Similarly, exports to Oceania have slowly fluctuated between $200 million and $300 million.

Canada is a special case due to chronic reporting errors. During 1977-89, U.S. agricultural exports to Canada were underreported by as much as $1 billion annually. Canadian import data showed about 50 percent more U.S. agricultural products entering Canada than did the U.S. export data reported by USDA and the U.S. Department of Commerce. For

[2] In a regression with one independent variable, the square root of the R^2 equals the simple correlation between the dependent and independent variables.

the forecasts to correctly align with the underreported export data, it would have been necessary to both estimate what Canada imported and what went uncounted.

There is no such simple explanation for rice. The rice forecasts are examined more carefully in the discussion of consistency.

Correlation
Efficiency was naturally weakest for the first-quarter forecasts. Less than half the R^2 values for the first-quarter commodity forecasts were above 0.50 (tables 5-7). Total export value had an R^2 of 0.38, while volume reached only 0.20. A total of 14 first-quarter forecasts were so explicitly inefficient that β was not significantly different from zero, including total export volume.[3] The regional forecasts generally have slightly better R^2 values than the commodity value forecasts, and the commodity value forecast R^2 values were better than those for volume. This was true in every quarter.

By the fourth quarter, it was unusual to find R^2 values below 0.90, and common to observe values above 0.95. Out of 20 regional forecasts analyzed, 13 showed R^2 of at least 0.95. The same was true with the most of the other forecasts.

The regional forecasts were more efficient in all quarters, and value was also more efficient than volume. (These are very general statements based on the frequency any forecast's R^2 in one group significantly exceeded any forecast's R^2 in another group.[4]) When accuracy is measured by MAPE, the regional forecasts are less accurate than the commodity forecasts. One implication is that there is more systematic error in the regional forecasts and more random error in the commodity forecasts. Bias is found more frequently among the regional forecasts, as noted earlier. This may be the source of the higher MAPE, since inconsistency does not seem particularly more common among the regional forecasts.

The higher R^2 values for the regional forecasts may also reflect differences in the characteristics of global versus regional trade. An individual country's demand for imports varies more than global demand; therefore, there is somewhat more variation in U.S. exports by region than by commodity. Given two forecasts--one commodity and one region, for example--equally accurate in terms of MAPE, but with more variation in exports to the region, then the regional forecast will show higher correlation with actual exports.

The commodity value forecasts were more accurate than the volume forecasts both in terms of MAPE and R^2. USDA generally has enough information about global supply and demand to correctly forecast the direction of change. This makes the value forecasts more accurate because they embody the correct direction in prices as well as volume.

The greater ease of forecasting export value is also demonstrated by differences in how accuracy increased from one quarter to the next. The R^2, as noted above, rises almost invariably. However, some of these changes were too small to be statistically significant. Statistically significant improvements in R^2 are more common for the regional and volume forecasts than for the commodity value forecasts in any quarter-to-quarter comparison.

No forecast's R^2 showed a significant deterioration, but some forecasts never improved from one quarter to another.[5] The forecasts failing to improve their correlation were generally the same forecasts that showed anomalous quarter-to-quarter changes in MAPE. These forecasts also failed to reach the R^2 threshold of 0.80, along with the forecasts for

---

[3]The other forecasts were: Western Europe, North Africa, the Middle East, Other Latin America, developed countries, oilseeds and products, soybean value, soybean meal value, livestock products, coarse grains volume, tobacco volume, animal fats volume, and other volume.

[4]The further the population correlation is from zero, the more skewed is the sample distribution of its estimator. We assume all correlations between these forecasts and their respective variables are substantially different from zero. However, it is possible to transform such a sample correlation into a variable that is normally distributed, assuming that the sample size is larger than 10 (4):

$$Z_r = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right), \quad with$$
$$\sigma_r = \frac{1}{\sqrt{n-3}}$$

The transformed variable lends itself to statistical testing.

[5]Tobacco value and volume, dairy, and animal fats.

other volume and Oceania.

Most forecasts by the third quarter had R^2 exceeding 0.80. Exceptions included the forecasts for high-value products (HVP). The third quarter forecasts for livestock, poultry, dairy, and horticultural products all had an R^2 below 0.80. The below-average accuracy of these forecasts may reflect the effect of the data problems with Canadian trade. Canada is one of the largest buyers of HVP's from the United States. Horticultural products, however, performed the best of this group, with a third quarter $R^2=0.76$, and horticultural exports were the most underreported of any commodity group.

Relatively poor efficiency for the HVP forecasts may have stemmed from a concentration of USDA's resources on program commodities. The differentiated nature of these goods multiplies the number of markets that would have to be monitored to anticipate events. Furthermore, USDA's intelligence gathering and analysis are geared toward low-value crops because of several priorities. One priority is that during much of the past 20 years, low-value products have dominated U.S. agricultural exports. Also, a substantially larger proportion of U.S. production is exported for low-value crops than for most high-value products. Perhaps most important, domestic commodity programs involving substantial Government expenditures exist for most low-value crops, necessitating increased vigilance by and for policymakers.

Bias

As forecast efficiency improves in the later quarters, one can measure bias and consistency. The ability to measure bias does not necessarily depend on strong correlation, but the weaker the bias, the greater the correlation must be to prove it is not a random occurrence.

In the regression equations shown earlier, the average difference between the forecasts and actual exports (or bias) can be found and tested by estimating with β restricted to 1. If the forecast is biased, then the estimated value of α will be significantly different from zero. This is essentially a "matched pair" test where the differences between the forecasts and actual trade are averaged across the sample (4). For example, the first-quarter forecasts for exports to North Africa averaged $124 million higher than actual exports during 1977-89 (table 8). This is the estimated bias resulting either from a restricted regression or from the matched-pair test.[6]

Bias occurred more frequently in the regional forecasts than in the commodity forecasts, and upward bias was more often found to be statistically significant than downward bias. Upward bias was generally found among the forecasts for the less developed countries and downward bias in the forecasts for developed countries.

The forecasts for North Africa and South Asia were upwardly biased in every quarter, although by smaller amounts in later quarters (table 8). During the first quarter, these forecasts were the only two that demonstrated statistically significant bias, with North Africa $124 million too high and South Asia $186 million too high. In the last two quarters, Eastern Europe, the Middle East, and East and Southeast Asia were biased upward.

Downward bias was more frequent among the commodities. However, cotton value forecasts were upwardly biased during the third and fourth quarters (by $77 and $61 million, respectively) and coarse grain value was upwardly biased by $353 million in the third quarter. Rice volume forecasts were downwardly biased during the last three quarters, and livestock products during the second quarter. The downward bias for tobacco volume and animal fats is less significant because these forecasts show such poor efficiency.

Coarse grain value's upward bias was large, but the size of the sample examined was below 10 observations. The smaller the sample, the less likely is one to find a normally distributed mean if the population is not normally distributed, and the less likely these tests are appropriate. Most of the forecasts analyzed here had been published for 13 years, every year since the U.S. Government switched its fiscal year to October-September. But some forecasts were published for fewer years. The first forecast in the Outlook for U.S. Agricultural Exports for the value of U.S. coarse grain exports was published in May 1981. Note that the coarse grain volume forecast, with a sample size of 13, does not demonstrate bias. If only forecasts since May 1981 are used, then the volume forecasts are also biased upward (see section on Consistency).

Consistency

Consistency means to correctly forecast the magnitude of change. Export forecasts were examined for inconsistency by testing the ordinary least squares estimates of
for $\beta = 1$. If a forecast fails the test, and there is no bias present, it is usually considered inconsistent. How to interpret

---

[6]The standard deviation of this average bias is calculated differently in the matched pair test than is the deviation for α in the restricted estimations. Usually both tests give the same results. The t-statistic for α in the restricted equation is used in this report.

$$dA_i = \alpha + \beta dF_i + \varepsilon_i$$

the test when bias is present is discussed below.[7]

The earlier discussion of A_i and F_i versus dA_i and dF_i noted that using A_i and F_i amounted to imposing a severe restriction on the model. One can further argue that the dA_i and dF_i model also is restricted and that loosening the restriction permits an intuitive interpretation of the resulting β's. If one postulates that the relationship between dA_i and dF_i differs depending on whether exports are rising or falling, then $dA_i = \alpha + \beta dF_i$ is a restricted model, and the estimated β is affected by aggregation:

$$dA_i = \alpha_r + \beta_r dF_i \quad if \quad dA_i > 0$$
$$dA_i = \alpha_f + \beta_f dF_i \quad if \quad dA_i < 0$$

In presenting the results of testing for consistency, I first present results based on the restricted model: $\alpha_r = \alpha_f$, $\beta_r = \beta_f$.

Another important factor affecting the estimated value of β is autocorrelation. Autocorrelated residuals are important in their own right in tests of weak-form rationality (7), and in the tests here for consistency they are important because they introduce bias into the OLS estimates of β. A handful of the OLS equations had Durbin-Watson statistics that indicated autocorrelated residuals.[8] Before testing for $\beta = 1$, the data for these forecasts were transformed using the Prais-Winsten method. Prais-Winsten was chosen to preserve degrees of freedom.

Consistency Estimates for the Restricted Model

Underestimates seem more common among the forecasts than overestimates. That is, $\beta > 1$ is observed more frequently than $\beta < 1$. The magnitude of change in total U.S. agricultural exports was typically underestimated about 25 percent by the third-quarter forecast (table 12). About a 7-percent underestimate was typical during the fourth quarter. As with bias, significant inconsistency is more common during the third and fourth quarters. This does not necessarily mean that the forecasts published in the first half of the year were more accurate, quite the opposite according to R^2 and MAPE. Instead, the early quarter forecasts have a greater degree of random error, which could conceal a systematic error like inconsistency.

A lack of consistency may lead to forecast bias or vice-versa. If actual exports are tending to increase or decrease, then bias and inefficiency are likely to coincide (if actual exports rise every year and the forecasts are biased upward, then clearly the magnitude of change is overestimated). If actual exports show no trend, and if the change in actual exports during the period studied averages to zero, then bias and inefficiency are more likely to be distinct.

Cotton value was significantly overestimated in the fourth quarter, while cotton volume was insignificantly overestimated. The cotton estimates leaned toward overestimation in all quarters, but reached significance in only one case. In addition, cotton value forecasts were biased upward in the last two quarters. Similarly, forecast changes for South Asia were consistently overestimated, and the forecasts were biased upward.

The only other biased forecast to demonstrate inefficiency (excluding Oceania) was Eastern Europe. Eastern Europe's bias was in the same direction as South Asia, upward, but its inefficiency was in the opposite direction.

---

[7] Seemingly Unrelated Regression (SUR) is appropriate for the sets of equations used in this report since the error terms are so highly correlated across equations. Our ability to use SUR was constrained by possible linear relationships among the error terms and lagged correlations across equations. Evidence that at least one of these conditions holds was found in the instability of some coefficient estimates from one SUR system to the next. The instability is never such that an equation determined to be inconsistent with OLS tests for consistency under SUR, but some equations consistent under OLS test either way under SUR, depending on what other equations are included in the system.

[8] In the first quarter: total export value, Asia, horticultural products, and grains and feeds. In the second quarter: sugar and tropical products. In the third quarter: China, livestock, Oceania, and coarse grains volume.

With cotton and Eastern Europe, the direction of trend, bias, and inefficiency agree in a manner that makes it impossible to use the restricted equation to tell which came first, bias or inefficiency. With cotton, overestimated growth possibly led to upward bias; with Eastern Europe, underestimated declines possibly led to upward bias. South Asia strikes a discordant note--it does not seem logical that an overestimated decline would led to upward bias. This would imply a bias distinct from inefficiency. Relaxing the constraint that $\alpha_r = \alpha_f$ and $\beta_r = \beta_f$ leads to the conclusion that the forecasts for Eastern Europe are definitely biased, and the forecasts for cotton are likely biased as well.

Consistency Estimates for the Unrestricted Model

Dividing the forecasts into two groups corresponding to years of rising and falling exports incorporates additional information into the equations. The results provide evidence that upward bias is more frequently the cause of overestimated change than vice-versa. The results also provide evidence that USDA's forecasts are less accurate in years when exports decline than in years they increase.

When exports trend upward, then upward bias would inevitably lead to an apparent overestimate of change ($\beta<1$), and an overestimate of change would inevitably lead to an apparent bias. With the restricted equation used above, it is impossible to determine which is the cause and which is the result. A tendency to overestimate change seems counterintuitive, while a tendency to underestimate change seems plausible ($\beta>1$). The difficulty economists have in predicting turning points in the economy has been widely documented. It is not too surprising that models, which must be estimated with historical data, fail to anticipate changing circumstances. Also, since time series are by nature strongly correlated with past values, a similar tendency for forecasts of time series data is rational. One would be suspicious of evidence that implies that USDA overanticipates events. However, some of the estimates for β in table 12 suggest USDA overestimates the amount of annual change by 100 percent.[9] This can be reconciled by more closely examining how β can vary from 1.

A forecast's bias could easily be proportional to amount of change. An upwardly biased forecast might be produced by increasing forecast change by 10 percent when exports are forecast to rise, and decreasing it by 10 percent when exports are forecast to decline. What is important here is that the first case is consistent with $\beta = 0.91$ and that the second case is consistent with $\beta = 1.11$. If export increases and declines occur with similar frequency and magnitude, then the estimated β of $dA_i = \alpha + \beta dF_i$ will be 1. If exports trend upward, however, this bias will lead to $\beta<1$.

To discern such cases, the following equation was estimated:

$$dA_i = \alpha_r + \alpha_f + \beta_r dF_{ri} + \beta_f dF_{fi} + \varepsilon_i$$

$$dF_{ri} = dF_i \quad when \quad dA_i > 0$$
$$dF_{fi} = dF_i \quad when \quad dA_1 < 0$$

This equation has no constant term, with α_r and α_f being dummy variables corresponding to rising and falling exports. The proportional bias described above would imply $\beta_r = 2 - \beta_f$.

There are several issues associated with estimating such a model. One is that if one can reject the pair of restrictions, $\alpha_r = \alpha_f$ and $\beta_r = \beta_f$, then it may be necessary to estimate rising and falling exports separately. The difference in the parameters may mean differences between the two sets of residuals and the associated variances. Unfortunately, given an overall sample size of 13 observations, dividing the samples into two sets makes for extremely small samples.[10]

A simple case of bias or inconsistency is one where α_j and β_j (where j = r or f) both imply the same relationship be dA_i and dF_i: for example, $\alpha_r<0$ and $\beta_r<1$, or $\alpha_f<$ and $\beta_f>1$. Each member of each pair above implies upward bias. The

---

[9]Intuitively, $\beta = 0.5$ means that 50 percent of the forecast is excess. Restating β so that the excess of the forecast is stated as a percentage of the actual: $(1/\beta) - 1$. Thus, when $\beta = 0.5$, half the forecast is excess, and the forecast is 100 percent greater than the actual value. Most of the estimated β's are closer to 1, and the difference between β and $1/\beta$ is substantially smaller.

[10]The pair of restrictions is rejected quite frequently: all but three of the first quarter forecasts studied rejected it with 10-percent significance. The restrictions were rejected less frequently in subsequent quarters; the restrictions could not be rejected for a majority of the fourth quarter forecasts.

closest to this simplest case of bias is the forecast for Eastern Europe in the fourth quarter (t-statistics for the difference from zero in parentheses):

α_r = 0.024 (0.632) β_r = 0.78 (11.262)
α_f = -0.011 (0.364) β_f = 1.19 (14.137)

This bias is not ideal because α_r has the wrong sign. However, this bias is of little concern because each α_j is so clearly insignificant and each β_j is so clearly different from 1. The relationship between dA_i and dF_i is a simple one of the forecast being biased upward 20 percent.

Upwardly Biased Forecasts

The fourth-quarter forecast for Eastern Europe is an example where relaxing the restrictions $\alpha_r = \alpha_f$ and $\beta_r = \beta_f$ provides some insight into the appearance of bias. The first section on inconsistency highlighted this forecast's underestimated change and declining trend as a possible cause of upward bias. However, because the tendency to underestimate change did not extend to periods when exports rose and was replaced by a tendency to overestimate change, bias seems to be a factor.

Cotton value was an example of an apparent overestimate possibly causing bias according to the restricted coefficient estimates. Relaxing the restrictions gives results that imply that upward bias was confined to years when exports rose. Earlier, when $\alpha_r = \alpha_f$ and $\beta_r = \beta_f$ were imposed, the estimated β was less than 1 because of the effect of aggregating $\beta_r < 1$ and $\beta_f = 1$ with exports rising in more years than falling. Therefore, the presence of $\beta < 1$ seems to be a result of upward bias rather than its cause. This also seems to describe the forecasts for South Asia in all four quarters.

Conclusions

USDA's quarterly export forecasts were largely efficient and unbiased, although they showed signs of being consistently cautious. The forecasts for grain exports were the most accurate of the group. They generally had the smallest percentage error and the best correlation, but the magnitude of change was conservatively forecast. Given the importance of grain to total exports, this led to conservative forecasts of change for total U.S. agricultural exports. Cotton exports were also accurate, matching grains in correlation, but showing bias and larger average errors. Cotton exports also varied from grain exports in that the forecasts were not conservative: the magnitude of change was overestimated on average. Since the overestimation was confined to the years exports rose, bias probably caused the overestimated change. Oilseed and product forecasts were less accurate than grain and cotton forecasts, probably due to the concentration of trade among a small number of countries.

Upward bias occurred in the forecasts of exports to a number of less developed countries that chiefly imported food grains and also received U.S. Government assistance in their purchases. Conclusions regarding the causes of upward bias in the regional forecasts can come only after further research. In some cases, the bias seemed concentrated in years when exports to a given region rose; in other cases the bias seemed concentrated in years when exports to the region fell.

The upward bias found for a number of regional forecasts does not necessarily reflect a bias by analysts responsible for concentrating on any of these regions. The regional forecasts published in the Outlook for U.S. Agricultural Exports are based on unpublished commodity forecasts that receive interagency review. Each month USDA publishes forecasts of expected marketing year U.S. export volume for a number of crops produced through a process of interagency review. To reach a consensus regarding the total for U.S. exports, unpublished forecasts of U.S. exports to each U.S. customer are formulated. These unpublished forecasts are then combined with ERS price forecasts to form the foundation of the published forecasts for the total value of U.S. agricultural exports to various regions. The regional forecast bias found in this report may stem from errors in either the interagency or the ERS component of these published forecasts.

Downward bias occurred in forecasts to some developed regions and the largest high-value commodity group, livestock products. Japan was the only developed-country forecast that was close to being biased and also important and otherwise accurate. But, its possible bias was less than 2 percent of the value of fiscal 1989 exports to that market. Livestock's bias was of equivalent absolute magnitude and totaled little more than 3 percent of fiscal 1989 exports. In both cases, underestimates of rapid export growth probably led to downward bias.

While it is of course desirable to eliminate such systematic errors, increasing forecast reliability is likely to entail costs. Any desire to improve forecast accuracy must be balanced by considerations of how costs compare with benefits. USDA is unlikely to reorient its intelligence-gathering efforts toward high-value products simply to increase the accuracy of its export forecasts for these products.

The first step is discovering systematic errors. Unforeseeable events will always result in some forecast error, but when errors fall into discernible patterns they represent behavior that can be altered to improve forecast accuracy.

References

1. Dietrich, J. Kimball, and Alfredo D. Gutierrez. "An Evaluation of Short-Term Forecasts of Coffee and Cocoa," American Journal of Agricultural Economics, 32(1) Feb. 1973, pp.93-9.

2. Gardner, George R., and David W. Skully. The Conduct of Wheat Marketing in North Africa. Staff Report No. AGES860808, Econ. Res. Serv., U.S. Dept. Agr., Sept. 1986.

3. Granger, C.W.J., and P. Newbold. "Some Comments on the Evaluation of Economic Forecasts," Applied Economics, 5(1) 1973, pp. 35-47.

4. Harnett, Donald L. Introduction to Statistical Methods. Addison-Wesley Publishing Company Inc., Reading, MA, 1975 (2d ed.).

5. Hiemstra, Stephen W., and Stephen A. MacDonald. "A Trade Estimates System to Assist Forecasting U.S. Agricultural Exports," NCR-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management, St. Louis, MO, Apr. 1988.

6. Kost, William E. "Model Validation and the Net Trade Model," Agricultural Economics Research, 32(2) Apr. 1980, pp. 1-16.

7. Lovell, Michael C. "Tests of the Rational Expectations Hypothesis," American Economic Review, 76(1) March 1986, pp. 110-124.

8. Maddala, G.S. Econometrics. McGraw-Hill Inc., New York, 1977.

9. Thomson, James M. "Analysis of the Accuracy of USDA Hog Farrowings Statistics," American Journal of Agricultural Economics, 56(5) December 1974, pp. 1213-17.

10. U.S. Department of Agriculture. Outlook for U.S. Agricultural Exports. All issues 1976-1989.

11. U.S. General Accounting Office. Accuracy of USDA's Meat Forecasts and Estimates. GAO/PEMD-91-16 May 1991.

Table 2--Average percentage error in value forecasts, by commodity, region, and quarter, 1977-89

Commodity/ Region	First quarter	Second quarter	Third quarter	Fourth quarter
		Mean absolute percent error		
China	66	38	18	11
Former USSR	37	22	13	6
South Asia	31	21	16	10
Eastern Europe	23	16	11	8
Sub-Saharan Africa	21	17	13	6
Oceania	20	15	18	17
Cotton (value)	19	13	7	5
Middle East	19	14	13	4
Tobacco (volume)	17	16	16	18
Cotton (volume)	17	12	5	3
Tobacco (value)	6	5	5	6
Horticultural products	7	6	4	3
Canada	7	6	6	3
Soybeans (volume)	8	8	6	2
Japan	9	7	5	3
East and Southeast Asia	9	7	4	4
Other Latin America	9	7	7	5
Wheat and flour (volume)	10	7	6	4
Asia	10	6	4	2
Total export value	10	7	4	1

Table 3--Correlation of forecast change with actual change in value, by commodity and quarter, 1977-89: Regression coefficient of determination (R^2)

Commodity/ Region	First quarter	Second quarter	Third quarter	Fourth quarter
		Percent		
Tobacco (volume)	0	11	11	17
Animal fats (volume)	0	46	68	81
Middle East	1	15	42	94
Oilseeds and products (value)	5	45	81	98
Western Europe	5	49	90	97
Poultry and products (value)	6	45	45	91
Livestock products (value)	9	8	53	93
Other Latin America	15	67	89	97
North Africa	16	47	76	80
Other Asia	25	60	90	95
Cotton (volume)	70	86	97	99
Soybeans (volume)	70	77	87	97
Sub-Saharan Africa	67	79	85	95
Cotton (value)	60	87	95	98
Grains and feeds (value)	60	81	95	98
Canada	59	57	58	91
South Asia	58	78	83	94
Wheat and flour (volume)	55	76	86	97
Eastern Europe	54	87	81	96
Japan	54	73	89	97
Total (value)	38	70	92	98

Table 4--Failure of test of weak-form rationality: rejection of $\alpha = 0$ and $\beta = 1$

Commodity/ Region	First quarter	Second quarter	Third quarter	Fourth quarter
		F-statistic		
Dairy	6.56	10.45	14.10	
Sugar and tropical	3.10	13.14		
South Asia	12.14	5.59	4.96	4.22
Eastern Europe		4.08		
Sub-Saharan Africa		3.42		
Rice (volume)		3.21		10.21
Total (value)			2.98	
North Africa			3.07	
Other Asia			4.13	6.46
Grains and feeds			3.21	
Coarse grains			3.07	
China			10.65	
Oceania			43.00	
Latin America				3.80
Cotton (value)				3.67
East & Southeast Asia				4.64
Middle East				3.32

Table 5--Forecasts rejecting β = 1, by quarter, 1977-89[1]

Commodity and region	First quarter	Second quarter	Third quarter	Fourth quarter
		Estimated coefficient value		
Commodity:				
Total value			1.26 (2.385)**	1.07 (1.77)*
Grains and feeds			1.21 (2.464)**	
Dairy	.47 (3.447)**	.46 (4.140)**	.36 (4.740)**	.72 (1.819)*
Cotton value				.95 (1.915)*
Sugar and tropical products	2.21 (2.349)**	1.38 (4.780)**		
Total volume			1.31 (1.933)*	
Rice				1.21 (3.196)**
Coarse grains		1.820 (1.942)*	1.24 (1.954)*	
Cotton				.94 (1.735)
Region:				
Western Europe			1.24 (2.000)*	
Eastern Europe		1.33 (2.179)*		
China			1.18 (1.972)*	
South Asia	.53 (3.167)**	.73 (2.044)*		.85 (1.998)*
Sub-Saharan Africa	1.59 (1.746)*	1.61 (2.47)**	1.34 (2.010)*	
Canada				.86 (1.774)
Oceania	.31 (5.025)**	.63 (2.846)**	.41 (6.644)**	.41 (7.205)**
Less developed countries			1.23 (1.794)*	1.11 (2.135)*
Centrally planned countries		1.36 (2.590)**	1.23 (2.610)**	

[1]T-statistics for difference from zero in parentheses:
* = significant at 10 percent
** = significant at 5 percent.

Table 6--Forecast bias, by quarter, 1977-89[1]

Commodity and region	First quarter	Second quarter	Third quarter	Fourth quarter
	1,000 metric tons			
Commodity:				
Rice		-155 (2.276)*	-140 (1.955)*	-147 (3.754)**
Tobacco				-22 (1.980)*
Animal fats		-60 (1.851)*		-35 (1.915)*
	Million dollars			
Coarse grains[2]			353 (2.050)*	
Livestock products		-187 (1.953)*		
Cotton			77 (1.839)*	61 (2.037)*
Sugar and tropical products		-50 (2.007)*		
Region:				
Eastern Europe		88 (1.611)	80 (2.033)*	57 (2.236)*
Former USSR		-200 (1.915)*		
Japan		-207 (1.679)	-138 (1.698)	
Other Asia			185 (2.992)**	163 (3.755)**
East and Southeast Asia[3]			117 (2.223)*	135 (3.080)**
South Asia[3]	186 (2.736)**	95 (2.304)**	95 (2.995)**	41 (1.853)*
Middle East			129 (2.117)*	44 (2.226)*
North Africa	124 (2.029)*	85 (1.940)*	70 (2.271)*	47 (1.733)
Latin America				-127 (2.314)*
Oceania		-26 (3.797)*		
Centrally planned countries			115 (1.800)*	115 (2.405)**

[1]T-statistics for difference from zero in parentheses:
* = significant at 10 percent
** = significant at 5 percent.

[2]Data are for 1981-89.
[3]Data are for 1977-87.

Table 7--Failure of test of weak-form rationality: presence of serial correlation

Commodity/ Region	First quarter	Second quarter	Third quarter	Fourth quarter
		Durbin-Watson statistic		
Total (value)	.82			
Grains and feeds	.87			
Horticultural products	.99			
Sugar and tropical		.90		
Livestock products			.93	
Coarse grains (volume			.95	
China			3.20	
Oceania			.95	

Forecasting Farm Commodity Program Participation Rates[1]

Douglas Maxwell, Economic Research Service, U.S. Department of Agriculture [2]

This preliminary study of farm commodity program participation rates is presented here at the Federal Forecasters Conference not only because the participation rates greatly affect one of the more variable programs of the federal government, but also because an issue came up that might be of interest to others doing forecasting or projecting in other areas: a method for dealing with a needed time series that changes part way through the estimation period.

This paper will tell about the importance of the farm commodity program participation rate forecasts for the government budget, one piece of previous work in the area and how to improve it, the importance of the cost of production in estimating the participation rate and the difficulties in extending the previous piece of work to the present. An extension and improvement of this work will be given.

This version of the paper is primarily directed to people who are interested in the process of forecasting rather than in farm programs per se. Therefore the discussion includes some background to understand the problem, but is somewhat too simplified for someone wishing to learn the intricacies of farm commodity programs.

Background

The farm commodity program participation rate is important, but variable, in its effect on government spending. This makes it important for the Department of Agriculture to study the participation rate carefully. The amount that the federal government spends on farm commodity programs depends not only on the governmental policy parameters and the season average price that farmers receive for their crops, but also on whether the farmer agrees to participate in the program.

The rate of participation is affected by governmental policy, which sets several policy parameters of which the main ones are: loan rate (at which the participating farmer can surrender the crop to the government), the target price (which results in deficiency payments to participants if the season's average price is lower), and the proportion of land that must be set-aside if the farmer participates.

These parameters affect the main payment to farmers, the deficiency payment, which is not known in advance since it also depends on the seasons average price, SAP. The amount of the deficiency payment per bushel is generally the amount by which the SAP falls short of the target price. However, if the SAP is less than the loan rate, then the deficiency payment is limited to the difference between the target price and the loan rate and the participating farmer may forfeit his crop to the Commodity Credit Corporation and receive the entire loan rate per bushel with some correction for carrying and storage charges. If the SAP exceeds the target price, then there is no deficiency payment. (I have left out refinements such as the cap on payments received per farm and the diversion payment sometimes paid on part of the idled land.)

Each of the government policy parameters affects the participation rate. If the other policy parameters remain constant (and the expected season's average price doesn't change: a big if in some cases), then the participation rate:

* will increase if the target price is raised since the participating farmer will expect to receive an increased deficiency payment (unless the expected market price already exceeds the target price, in which case there will be little effect.)

* will decrease as the set-aside is increased constant, since the farmer cannot plant as much land, which will lower the amount of crop that can be produced and hence the expected profits of the participating farmer.

* may increase or decrease as the loan rate increases since the participants receive a higher guarantee, but the non-participants may expect that the loan rate will act as a floor to them also (since the participants would forfeit their crop to the government rather than sell it on the market at a price lower than the loan rate.)

In addition a higher expected seasons average price will reduce participation since participating farmers would receive less from the government and forfeit more profits from the set-aside land.

If more farmers participate then prices will rise and production will decline, since acreage planted will decline due to more set-aside land. (This will be partially offset since the government payments may induce farmers to add more inputs, such as fertilizers, which will increase yield. In recent years the use of a "program yield" has reduced the incentive of farmers to increase actual yield, but farmers still tend to set-aside less productive land.)

Previous Work and the Possibilities of Extending It
There has been previous work through 1987 by Keith Menzie and Lawrence Van Meir and this study seeks to extend that work.

Their work used a composite variable (discussed in the section on results below) which included both the net returns to participants and the net returns to non participants. In their article these net returns are constructed from the raw data series for a number of program parameters set by USDA and the expected price and costs. The cost of production is a key variable.

Menzie and Van Meir estimated only the linear portion of their equation for selected certain values of the non-linear parameter and chose the non-linear parameter that gave the best fit.

It is important to use cost of production data in estimating participation rates. In order to estimate the participation rate, it is necessary to view the participation decision from the point of view of the decision maker, the farmer. He will usually participate in the commodity program if he thinks he can make more money than by not participating. If he does not participate, he can plant all his acreage and sell it at the market price, but if he participates, he can only plant on his base acres less the set-aside amount, but his revenues may include government payments as well as market receipts. The government payments will be lower as the market price increases and the lost revenues from the set-aside land will be more important. However, the farmer is also vitally interested in the cost of production, since costs must be subtracted from revenue to find income. The cost of production will be different for participating farmers since they grow the program commodity on fewer acres and generally plant a cheaper cover crop on the remaining acres. Variable costs (seed, fertilizer, herbicides, pesticides, fuel, machinery wear, irrigation, etc.) will decline if some of the land is set-aside--although a lesser amount of variable cost must be expended on the set-aside land to establish a cover crop (which, if it is a legume, will reduce the need for nitrogen fertilizer the following year) to reduce weeds and erosion. Possible complications include: The farmer may set-aside land that is significantly poorer in quality where the cost of production per bushel (or even per acre) is higher, the farmer may increase the use of inputs on the remaining land to boost yield (especially since the cost of some inputs may decline if planted acreage declines), and since the 1990 farm bill there have also been "flex acres" which could be planted to a variety of crops but received no governmental support.

There is some difficulty extending previous work due to inconsistency in the data needed. Since the study by Menzie and Van Meir stopped with 1987, it is important to extend it to the present for the purposes of forecasting. However, the cost of production, a key variable, is now constructed according to different procedures than before. The cost of production has always been constructed based on survey results that ask a series of underlying questions. For example, the cost of seed may be constructed from seeding rates and the price of seed and the cost of machinery use may be based on variables such as questions on the number of times the field was worked, types of machinery used, and the cost of fuel, or, if some was contracted out (custom work), the amount paid for what service. The ways in the cost of production is calculated are now significantly different, with parts of the calculation now done farm by farm rather than state by state. This makes it important to use a methodology which minimizes the errors due to the change in the method of calculating.

Extending a Time Series
How to deal with inconsistent data series. In estimating behavioral equations using time series, it is necessary to use data that is available over an estimation period that is long enough so that one does not run out of degrees of freedom. For example, assume that one is estimating the number of people who voluntarily participate in a government program, and that an important economic variable a assumes the value of a_p for participants and a_{np} for non participants. However, assume that the methodology used to construct the value of a has changed over time. One may run out of data constructed using the latest method before having enough degrees of freedom to be confident in the method used.

For example, for the purposes of estimating the participation rate of farmers in the US agricultural commodity programs, the variable that is most problematical is the cost of production of the crop in question. The cost of production data has been produced using varying techniques over any reasonable period that one could use in estimating a participation rate equation. Therefore, it is best to use some functional form which minimizes the influence of an incorrect cost of production estimate. For example, if pc is the cost of production per acre assuming compliance with the relevant government programs, and pn is the cost of production for a non-participant, then it is important to use a formulation in which errors tend to cancel out. For example, if pc = A + B and pn = A' + C, where A and A' are the same variable (or set of variables) calculated for different farmers, but the methodology for constructing them has changed over the time period, then it is best to use a formulation where the errors can cancel out before the estimation process. Thus an equation which uses pc - pn or pc/pn is preferable to an equation in which pc and pn have different coefficients.

Results
This work improves the previous work by Keith Menzie and Lawrence Van Meir and extends the length of the data series used. Although simple, their analysis is surprisingly robust when some sophications are added and the data series lengthened.

The present study also uses their basic independent variable **G** = **(nri-nro)/nro** (where **nri** is net returns to participants and **nro** is net returns to non participants--see Appendix A) which allows the maximum cancellation of cost of production values prior to estimation. Thus on those acres which are planted by both participants and non participants, the variable costs of production are cancelled out prior to the estimation since only the difference in cost enters the composite variable **G**. Even the cost of the cover crop (planted by participants on the set-aside acres) and the cost of the program crop (planted by non-participants on the same acres) would be estimated by the same methodology--and hence both would probably be either overestimates or underestimates. This maximum cancellation of costs is important because the cost of production data is now estimated with more of the data being calculated from the individual data provided by the farm in question with less use of calculated or statewide data in the calculation of the individual farm's cost of production.

It was necessary to reconstruct the calculation of the composite variable **G** since there were a few errors. (See Appendix A for the details.)

In the exploratory study presented here, equations were estimated for corn. (Other program commodities with target prices and a similar use of the loan rate include wheat, sorghum, barley, and oats.)

I did not attempt to use their procedure for estimating the equation **ln(P-K)** = a + **bG^c** where P is the participation rate and they imposed the K (in their opinion, the highest possible participation rate, about .95) and substituted various values of c until they found the highest R-squared value. Instead I assumed that **K = 1** and estimated **ln P** = a+**bG^c**.

I first estimated the equation with linear regression setting the value of c at 0.5, the value arrived at by Menzie and Van Meir. I used both their data series for **G** using the statistical package TSP (Time Series Processor). Interestingly, the fit of their equation **improved** when the data were corrected, thus offering additional evidence that their overall approach was correct and not an artifact of the miscalculations. I did not save the results with the miscalculated **G** variables, but the results with the corrected data are in Appendix B [with the coefficients c(1) and c(2) in the place of a and b.]

Next, the data series was lengthened to provide an additional degree of freedom. Since they only estimated through 1987 due to the unavailability of the 1988 participation rate (which they were forecasting), I also estimated with the extra observation. This improved the Durbin-Watson statistic, which went from bad to quite respectable. The results are shown in Appendix C.

Thus both the corrections of some clerical errors and the lengthening of the estimation period improved the results of their approach, which confirms the basic soundness of their composite variable.

Finally, in contrast to Menzie and Van Meir, this study estimated all of the parameters in the equations using non-linear regression, rather than estimating only the linear portion for certain values of the non-linear parameter and choosing the non-linear parameter that gave the best fit. The results with the corrected data are in Appendix D [with the coefficients c(1), c(2), and c(3) in the place of a, b, and c.] The obvious next step would be to try other specifications in which the participation rate approaches the value of 0 asymptotically (for unattractive government programs) as well as the value of 1. Similar work remains to be done on other commodities.

Footnotes

1. This paper represents the views of the author only and not the views of USDA.

2. Economic Research Service, ATAD, Europe Branch, Room 628, 1301 New York Avenue, NW, Washington, DC 20005. Phone: (202) 219-0620.

BIBLIOGRAPHY

Menzie, Keith L. and Lawrence W. Van Meir, "Farmers' Response to Program Incentives" in **Feed, Situation and Outlook Report** May, 1988, Economic Research Service, USDA, FdS-306

Appendix A

In the Menzie article **nri** and **nro** are constructed from the raw data series ARP, PLD, target price, loan rate, program yield, expected yield, expected deficiency payment, diversion payment, expected price, production cost, and idled production cost.

The exact formulation is:

nro = P*Y*A - C*A

where * represents multiplication

nro is the net return for non-participants
P is the expected price
Y is the yield per acre
A is the eligible base acreage
C is the variable cost of production per acre

nri = (1-ARP-PLDR)*A*PrY*DP + PLDR*A*PY*PLD + (1-ARP-PLDR)*P*Y*A
- (1-ARP-PLDR)*C*A - (ARP+PLDR)*A*CI

where nri is the net return for participants
ARP is the proportion of land that must be idled without any payment
PLDR is the proportion of land that must be idled to receive the diversion payment
PrY is the program yield (often based on historical yields
DP is the deficiency payment per bushel which is 0 if the seasons average price is above the target price, or the amount by which the target price exceeds the seasons average price or the loan rate
PLD is the rate per bushel for paid land diversion
CI is the variable cost per acre for idled land

I found that I had to reconstruct the calculation of **nri** and **nro** from the raw data that they had provided since there seemed to be some inconsistencies. It turned out that three of the g variables had been miscalculated, and the correct figures follow:

Year	Original	Corrected
1978	0.0989	0.0989
1979	0.0159	0.0159
1982	0.1949*	0.0797
1983	0.4622	0.4622
1984	0.2540	0.2540
1985	0.3535*	0.3167
1986	1.8267*	1.9558
1987	2.4235	2.4235
1988	1.0245	1.0245

When the new figures were substituted in the regression, the fit of the equation improved.

Appendix B

```
                                Results without 1988, imposing .5 for c(3)
NLS // Dependent Variable is LPART
SMPL range: 1978  -  1979    1982  -  1987
Number of observations: 8
LPART = C(1) + C(2)*GNEW^.5
Equation is linear - 1 iteration
==========================================================================
                    COEFFICIENT   STD. ERROR      T-STAT.   2-TAIL SIG.
==========================================================================
        C(1)        -0.0632476     0.1012093    -0.6249192     0.5550
        C(2)        -1.8210115     0.1209067    -15.061298     0.0000
==========================================================================
R-squared                 0.974232    Mean of dependent var     -1.297990
Adjusted R-squared        0.969937    S.D. of dependent var      0.968159
S.E. of regression        0.167867    Sum of squared resid       0.169075
Log likelihood            4.075904    F-statistic                226.8427
Durbin-Watson stat        3.753483    Prob(F-statistic)          0.000005
==========================================================================

==========================================================================
              Residual Plot              obs RESIDUAL  ACTUAL   FITTED
==========================================================================
|          :         | *       :         | 1978  0.03809 -0.59784 -0.63593
|          :        *|         :         | 1979 -0.00824 -0.30111 -0.29287
|          :         |         *         | 1982  0.16183 -0.41552 -0.57734
|          : *       |         :         | 1983 -0.12585 -1.42712 -1.30127
|          :         |     *   :         | 1984  0.08941 -0.89160 -0.98101
|    *     :         |         :         | 1985 -0.25903 -1.34707 -1.08804
|          :         |         : *       | 1986  0.20198 -2.40795 -2.60993
|          :   *     |         :         | 1987 -0.09819 -2.99573 -2.89754
==========================================================================
```

Appendix C

```
                                       Results with 1988, c(3) = .5

NLS // Dependent Variable is LPART8
SMPL range: 1978  -  1979    1982  -  1988
Number of observations: 9
LPART8 = C(1) + C(2)*GNEW^.5
Equation is linear - 1 iteration
==========================================================================
                    COEFFICIENT   STD. ERROR      T-STAT.   2-TAIL SIG.
==========================================================================
        C(1)        -0.0671594     0.1600420    -0.4196364     0.6873
        C(2)        -1.9117567     0.1864626    -10.252761     0.0000
==========================================================================
R-squared                 0.937566    Mean of dependent var     -1.434406
Adjusted R-squared        0.928647    S.D. of dependent var      0.993805
S.E. of regression        0.265464    Sum of squared resid       0.493299
Log likelihood            0.296941    F-statistic                105.1191
Durbin-Watson stat        2.098467    Prob(F-statistic)          0.000018
==========================================================================

==========================================================================
              Residual Plot              obs RESIDUAL  ACTUAL   FITTED
==========================================================================
|            :       | *     :           | 1978  0.07054 -0.59784 -0.66838
|            :       *       :           | 1979  0.00712 -0.30111 -0.30822
|            :       |     * :           | 1982  0.19136 -0.41552 -0.60687
|            :     * |       :           | 1983 -0.06024 -1.42712 -1.36687
|            :       |   *   :           | 1984  0.13906 -0.89160 -1.03065
|            : *     |       :           | 1985 -0.20405 -1.34707 -1.14302
|            :       |       : *         | 1986  0.33280 -2.40795 -2.74075
|            :       |*      :           | 1987  0.04696 -2.99573 -3.04269
|    *       :       |       :           | 1988 -0.52354 -2.52573 -2.00219
==========================================================================
```

Results with 1988, c(3) estimated

```
NLS // Dependent Variable is LPART8
SMPL range: 1978  -  1979    1982  -  1988
Number of observations: 9
LPART8 = C(1) + C(2)*GNEW^C(3)
Convergence achieved after 6 iterations
========================================================================
                    COEFFICIENT   STD. ERROR      T-STAT.   2-TAIL SIG.
========================================================================
        C(1)         0.4535076    0.7491536    0.6053600     0.5671
        C(2)        -2.5039497    0.8185213   -3.0591136     0.0223
        C(3)         0.3479825    0.1479557    2.3519371     0.0569
========================================================================
R-squared                   0.947108    Mean of dependent var    -1.434406
Adjusted R-squared          0.929478    S.D. of dependent var     0.993805
S.E. of regression          0.263915    Sum of squared resid      0.417908
Log likelihood              1.043287    F-statistic               53.71953
Durbin-Watson stat          1.826214    Prob(F-statistic)         0.000148
========================================================================

========================================================================
              Residual Plot                 obs RESIDUAL  ACTUAL   FITTED
========================================================================
|          :        | *      :         | 1978  0.06802 -0.59784 -0.66586
|          :  *     |        :         | 1979 -0.16204 -0.30111 -0.13907
|          :        |     *  :         | 1982  0.16935 -0.41552 -0.58486
|          :        |*       :         | 1983  0.03360 -1.42712 -1.46071
|          :        |      * :         | 1984  0.20913 -0.89160 -1.10073
|          :    *   |        :         | 1985 -0.12232 -1.34707 -1.22475
|          :        |        : *       | 1986  0.30083 -2.40795 -2.70878
|          :       *|        :         | 1987 -0.04247 -2.99573 -2.95326
|   *      :        |        :         | 1988 -0.45411 -2.52573 -2.07162
========================================================================
```

Using an ARIMA Model in an Unconventional Setting: Forecasting The Demand For Inpatient Hospital Services

John J. Hisnanick*, U.S. Dept. of Veterans Affairs, Biometrics Division

Abstract. The methods of the autoregressive integrated moving average (ARIMA) modeling technique are used extensively in forecasting and predicting such economic phenomena as corn yield, livestock production and short-term interest rates. The issue of using these methods to look at the utilization (or demand) of services within the service related industries, such as inpatient hospitalizations, has received minimal attention. It is an established fact that increased use of inpatient hospital services is driven by such factors as an extension of third-party payments through private insurance and public programs, the aging of the general population, and advances in medical technology. However, it is often not possible to abstract objective information that can be used in a multivariate model to assess these factors relative to inpatient use. ARIMA methods are a way to get around this problem, in that forecasts and projections are based upon a univariate time-series. Using monthly inpatient discharges from the US Department of Veterans Affairs (VA) hospitals, an ARIMA specification was developed that provided a thirty-six month ahead forecast of inpatient utilization. The times-series used in the development of the ARIMA specification consisted of 132 monthly observations that spanned fiscal years 1981-91. The development of the ARIMA specification, the resulting model and its validation, and the thirty-six month ahead projections are presented.

Introduction

In the discipline of economics, the study of the demand for health, the demand for health care and the allocation and distribution of resources needed for their production has become an area of increasing concern and study. The quantitative evaluation of this field has concentrated on the econometric estimation of certain important relationships. Among the most important of these are demand functions for health care/hospital services. It is important to recognize that the demand for health and demand for health care are two separate entities. That is, the demand for health is determined by many factors and health care is just one part. However, for this study, the emphasis will focus on the demand for health care. Particularly the demand for inpatient hospital services provided by facilities run by the U.S. Department of Veterans Affairs (VA), as defined as discharge episodes. The terms health care, health services and hospital services are used interchangeably throughout.

Traditionally, demand models for health services have looked at either the length of stay or physician visits relative to such independent variables as health insurance status, health condition status, age, and sex, to name but a few. These demand studies are based upon a secondary analysis of cross-sectional data taken from various surveys. Several extensive review articles have been written which look at the empirical work that has been done on the demand for health care (or health services),[Feldstein(1966,1974), Joseph(1971) and Culyer(1981)].

In looking at the demand for health services provided by the facilities run by the VA, several studies are of notable interest. Using data from the 1978 first National Survey of Veterans, Page (1982), looked at 1260 veteran responses and constructed a multivariate logistic regression. This study investigated the choice of hospital used relative to the existence of health insurance, age, income group and service-connected disability. He discovered that health insurance status was the most important factor affecting a veteran's choice of hospital. Other studies have looked at the demand for health services provided by the VA, but have used non-VA data. Specifically, the work done by Wolinsky, et al (1985,1987), used the 1978 National Health Interview Survey. These studies were a comparative analysis of veteran and non-veteran uses of health services, and an investigation of length of stay at VA facilities. In addition, specific studies addressing the needs (or demand) of aging veterans on VA facilities have been done. The study of Horgan, Taylor and Wilensky (1983) used the 1977 National Medical Expenditure Survey to look at the implication of an aging user population. Similarly, the work of Romeis, et al (1988) used the 1982 Survey of Aging Veterans to investigate this groups potential demand for health services. All the above mentioned studies used cross-sectional data to evaluate a structural specification of the demand for health services provided by the VA. While cross-sectional data is appropriate to test causal relationships, as well as estimating the sensitivity to change in your outcome variable relative to other independent variables, it does have certain limitations. Specifically, cross-section data does not permit one to investigate the dynamic behavior of the demand for health care. This type of analysis involves the use of time-series data.

In a recent study by Arron(1991), the following factors were outlined and discussed as influencing the demand for health care/hospital services: the aging of the population; shifts in the inpatient use patterns; a changing mix in demand for inpatient and outpatient services; and the influence of technology. As is often the case in empirical work, a series to measure such effects are either not available or are not easily quantifiable for a structural demand model specification. Therefore, to test the behavior over time of the demand for health care/hospital services, one often only has the time-series which describes an outcome measure for inpatient hospital services, such as discharges. This should not be treated as a shortcoming in that appropriate statistical methods exist which allow for univariate time-series forecasting.

The type of methods referred to are autoregressive-integrated-moving average (ARIMA) techniques. They are founded on statistical concepts and principals which form the basis of its power and wide applicability. While these methods are considered naive for program planning purposes in health care, their application in forecasting the future behavior of an observed time-series, such as hospital inpatient discharges, seems to be quite appropriate. Since the administrative databases maintained by the VA can reasonably generate such data, these techniques have the potential to become tools in the planning process.

The focus of this study is twofold; first, outline the steps that can be followed in developing an ARIMA forecasting model, and second, present an ARIMA specification developed to investigate the short-term and mid-term demand for future inpatient use at VA facilities.
About the first objective, the forecasting process can be thought of as being composed of six separate steps [Hoff(1983)]. These steps involve defining the forecasting problem; collecting and organizing the data; selecting and applying a forecasting method; reviewing and adjusting the preliminary forecast; tracking the forecast accuracy; and updating the forecast. Each step will be expanded upon in the development of the specific ARIMA model. On the second objective, this work was done using the fourth generation language SAS, within a mainframe environment. The modeling and estimation was done using the SAS/ETS (Econometric and Time-Series) library;[SAS Institute (1984)].

Methods

The Forecasting Problem: The forecasting problem inherently involves predicting the outcomes in unobserved states of nature with some measure of certainty. Policy and program planners within the VA are constantly faced with the problem of allocating resources for the efficient delivery of health care to it's service population that is both aging and declining. Taking the case that an aging population, is hypothesized to result in an increased consumption of health care/hospital services the forecasting problem can be stated:

> What will be the short and mid-term forecast in the demand for hospital (inpatient) services at VA run facilities, given an aging and declining service population?

An investigation of this problem will center around a forecasting model based upon monthly inpatient discharges from VA operated facilities.

Collecting and Preparing the Data: The data used in the construction of the time-series was the VA Patient Treatment File. This database is episode specific and contains both administrative and demographic information related to the hospitalized individual at the time of discharge. The data are maintained on computer tapes in a SAS format and are most easily accessed and manipulated within a mainframe environment. The respective monthly time-series was generated by the aggregation of daily discharge episodes and contained 132 observations for federal fiscal years 1981-1991. Figure 1 provides a graphical representation of the generated univariate time-series.

Selecting and Applying a Forecasting Method: A univariate time-series represents a collection of data with a specific temporal ordering and is assumed to be generated by a stochastic process with a structure that can be characterized and described accordingly. In general, a time-series is composed of three distinct components: a trend component; a seasonal component; and an irregular or random component. The trend component represents the overall long-run movement in the time-series. The seasonal component reflects a possible seasonal pattern that repeats itself over some regularly space time interval. And, the irregular or random component reflects the non-systematic movements that occur in the series. These components can be related in either an additive or a multiplicative fashion.

A major question that needs to be addressed before the selection of a forecasting model is whether the generated univariate time-series is stationary. Stationarity is an important characteristic of a stochastic process; it implies that the time-series under investigation is invariant with respect to time and that it wanders more or less uniformly about some fixed level. There are several techniques that can be used to check the stationarity of time-series. The most intuitive approach is to graphically observe the series for stationarity. However, the more accepted approach is to look at the plot of the autocorrelation function (ACF) of the time-series and check it's behavior.

The ACF of a time-series is a statistical measure that has a possible range covering the interval [-1, +1] and measures how strongly the time-series values, a specific number of periods apart, are correlated to each other. If the values of the ACF do not fall off quickly or within the prescribed range of statistical insignificance as the number of lags increases, this shows that the series is nonstationary.

If a nonstationary series is encountered, it can be made stationary by differencing one or more times, until stationarity is obtained. The number of times that the series is differenced (until stationarity is achieved) is referred to as the "order of homogeneity" of the series. To assess whether a more stationary series could be obtained, further differencing should be done on the series. If the successive differencing does not result in a significant qualitative change in the decline of the ACF, the proper order of homogeneity for the series is the level of differencing where no further qualitative changes occurred in the plot of the ACF.

Figures 1 and 2 are the graphical representations of the generated monthly time-series and the series associated ACF.

By observation of Figure 1, the series is not uniformly distributed about some fixed value, thus implying nonstationarity. This is further supported by Figure 2, were the observed ACF does not fall off quickly. These results imply that the series should be differenced to see if stationarity occurs.

Figure 3 is the graphical representation of the first difference of the original series and by inspection there appears to be a uniformed distribution around zero. Upon the inspection of the ACF associated with the first differenced series, there is a decline towards zero (or within the area of statistical insignificance). However, there are spikes that occur at the first, twelveth and thirteenth lags. This observation leads to the possible conclusion there may be seasonality in the data. To investigate this possible affect and remove the cycle a twelve month difference of the series was done. Figure 5 is the graphical representation of the ACF for the seasonally adjusted differenced series and the problem of significant spikes occurring at lags 12 and 13 have been corrected accordingly. The seasonally adjusted differenced series will be used in the estimation of the proper forecasting model.

Pindyck and Rubinfeld (1976) referred to the specification of an appropriate ARIMA model more of an art than a science. However, in the setting up an ARIMA forecasting model, the ACF and partial autocorrelation function (PAC) can aid in selecting the proper model components. The determination of the appropriate moving average (MA) portion of the model can be done by checking the behavior of the ACF. In general, the ACF of the MA(q) process has "q" positive and negative values and is zero for lags greater than "q". In Figure 5, looking at the ACF of the seasonally adjusted differenced series, there is a significant spikes that occurs with an erratic, declining behavior of the ACF thereafter, as well as the affect of seasonality in the data. This observation suggest that the MA portion could be of order two.

For the autoregressive (AR) part of the model, the patterns are somewhat more complicated than those of the MA part. For the AR(p) component the current values of the time-series are related to the past values of the series. As the lag values increase back in time, the level of association with current values of the series becomes weaker until at some value greater than "p" the level of association is zero. In a pure AR structure this type of behavior is reflected through the ACF in that it will dampen as the number of lags increases. This observed dampening can take the form of an exponential rate of decrease, a constant rate of decrease, or an alternating pattern of decrease due to the sign change (positive to negative) of the AR level of association. For mixed ARIMA models using the ACF to determine the "p" components for the AR portion can be often confusing and misleading. This is quite clear when looking at the behavior of the ACF in Figure 5. The PAC tends to be a more appropriate tool for evaluating the respective components of the AR portion. That is, the AR portion can be checked by using the PAC, similarly to the way that the MA portion is checked using the ACF. Therefore, by observing the PAC, the AR(p) portion was determined to be of order two.

For the time-series under investigation first differencing resulted in stationarity and a twelve month differencing adjusted for the seasonality in the series. The model chosen to forecast the monthly behavior of discharges at VA facilities was an ARIMA$(2,1,2)_{12} \times (2,1,2)$; a seasonally adjusted model with first differencing and two AR and MA components. Table 1 presents the respective parameter estimates, T-ratios, and the 95% confidence interval for the parameters of the proposed ARIMA specification.

Testing the Model Specification: In dealing with the adequacy of the proposed ARIMA specification there are several steps that can be taken. These include assessing the estimated parameters of the model for significance and testing for how well the overall model fits in forecasting future values of the series. For the adequacy of the estimated parameters for the model, from Table 1 it can be seen that the first MA estimated parameter is marginally insignificant at the 0.05 level of significance, while all others are extremely significant. However, including the first MA parameter and all other parameter estimates into the model are supported by examining the matrix of correlations of the estimates in Table 2, Part A. All estimated parameters are relatively uncorrelated with each other.

In looking at the adequacy of the overall model, this can be tested by observing the sample ACF of the model's residuals. In theory if the model specification is adequate the estimated residuals should be nearly uncorrelated with each other. A very convenient test to check the adequacy of the model is the Box-Pierce Test, [Box and Pierce (1970)]. This test assumes that for large displacements or lags, the residual autocorrelations are uncorrelated, normally distributed random variables with a mean of zero and a variance of 1/T; T being the number of observations in the series. The resulting test statistic, named the "R" statistic, is composed of the first "K" residual correlations of the series and is approximately distributed as a chi-square with (K-p-q) degrees of freedom. While the Box-Pierce test can give an indication of an inadequate model specification, it does not provide any insights into how to improve the model specification.

Table 2 (part B), summarizes the results of the Box-Pierce test. Under the assumption of a 95% level of confidence it can be concluded that the sample residual autocorrelations are uncorrelated at least for lags less than eighteen.

Overall, in looking at the adequacy of the individual parameter estimates of the model and the model's structure, the results suggest that the proposed ARIMA specification should be adequate for modeling short-term and mid-term forecasts for VA inpatient discharges.

Tracking the Forecasting Accuracy: The value of any forecasting exercise rests heavily upon how well the proposed model will reflect the best guess at what the future holds for the univariate time-series under investigation. Using the parameter estimates from Table 1 a forecast was done. Table 3 and Figure 7 present a forecast for twelve periods back and thirty-six periods ahead, with the 95% lower and upper confidence interval estimates.

Assessing the accuracy of the forecast requires looking at the forecast limits and an after-the-fact forecast using some periods before the series ends. The forecast limits are essentially a confidence band about which statistical statements can be made regarding the validity of the range of an estimate. The 95% level is the most commonly used in time-series forecasting and indicates that for 95 out of 100 forecasts, the corresponding actual series value should fall within the confidence band.

An after-the-fact forecast provides a technique that allows one to compare forecasts computed from the current model with actual series values that are already known. This approach provides a way to measure the accuracy that can be expected in the forecast for the real future. In addition, it allows for the computing of closeness of fit statistics based upon the after-the-fact forecast errors. These are the forecast mean percent error and the forecast average absolute percent error.

In looking at the forecast limits for the twelve periods back the actual series values all fall within the 95% confidence limit band. Turning to the after-the-fact forecast the calculated mean forecast error was -1.52% and forecast absolute percent error was 2.97%. The mean forecast error implies that for the twelve periods before the end of the series the model overestimated the series value on average by 1.52%. The calculated forecast absolute percent error provides a way to measure the fit of the model for the same twelve periods back and the value of 2.97% suggests that the forecast values are on average only slightly off from the fitted values.

Therefore in assessing the accuracy of the forecasts, the proposed model appears to behave appropriately and yields reasonable, somewhat overestimates, of the monthly VA inpatient discharges.

Updating the Forecasts and The Forecasting System: The last step in the forecasting process concerns the updating of both the model's forecasts and the forecasting system. This step should be done periodically as new actual values for the time-series becomes available. It is important to remember that forecast modeling is a dynamic activity. It is rarely a one-time activity and as time moves on, the assumptions, requirements and constraints that were initially proposed and developed for the system will most likely change. Lastly, it is important for program planner to realize that ARIMA forecasts should be used in conjunction with intuition, general knowledge and common sense in developing policy recommendations.

Summary and Conclusions

The focus of this study was to find an appropriate ARIMA forecasting model following the six steps outlined for developing such models. Working under the assumption that the VA service population are both aging and declining, the proposition under investigation focused on the short and mid-term forecasts in the demand for hospital inpatient use at VA run facilities.

The developing of an adequate forecasting model to look at this question involved a univariate time-series. The series values consisted of monthly measures of VA inpatient discharges covering federal fiscal years 1981-91. The resulting ARIMA model from the series was a seasonally adjusted model, with one order of differencing and two autoregressive and moving average components; ARIMA$(2,1,2)_{12} \times (2,1,2)$.

The proposed model specification was checked for both adequacy and accuracy using established statistical techniques. These methods included an evaluation of individual estimated model parameters and checking of a 95% confidence limit band in an after-the-fact forecast. These techniques lead to the conclusion that the proposed model specification would be adequate for forecasting the future demand for VA inpatient hospital services.

The value of any forecasting model rests heavily upon how well it projects future values of the series under investigation. From the proposed model specification a thirty-six month ahead forecast was generated with it's associated 95% confidence limit band. The generated series values reflect projections for VA inpatient discharges for federal fiscal years 1992-94. When the monthly projections are totaled to yearly estimates, the VA inpatient discharges are declining modestly ,-0.14%, over the short-term, but decrease more sharply over the mid-term; about -1.5% on average.

Lastly, while the proposed model appears both adequate and accurate in explaining the short and mid-term behavior of VA inpatient discharges, the final forecasts are contrary to what one would expect. Arron (1991) hypothesized that as a population ages there is an increase in the demand for medical care and hospital services. However, the projected declines, and the reasons for them, need further investigation. As is the case with most empirical econometric studies, the final results often provide the point of departure for future studies.

References

* The opinions expressed are solely those of the author and do not reflect those of the U.S. Department of Veterans Affairs.

Arron, H.J.(1991). Serious and Unstable Condition: Financing America's Health Care. Washington, DC, The Brooking Institution.

Box, G.E.P and D.A.S. Pierce (1970). "Distribution of Residual Autocorrelations in Autoregress-integrated Moving Average Time Series." Journal of the American Statistical Association, 65.

Culyer, A.J. (1981). "Health, Economics and Health Economics", in J. van der Gaag and M. Perlman, eds., Health, Economics and Health Economics, Amsterdam, North-Holland.

Feldstein, M.S. (1974). "Econometric Studies of Health Economics", in: M.D. Intriligator and D.A. Kendricks, eds., Frontiers in Quantitative Economics, Vol II. North-Holland, Amsterdam.

__________ (1966). "Research on the Demand for Health Services." Milbank Memorial Fund Quarterly, 43.

Hoff, J.C. (1983). A Practical Guide to Box-Jenkins Forecasting, Lifetime Learning Publications, Wadsworth, Inc., Belmont CA.

Joseph, H. (1971). "Empirical Research on the Demand for Health Care." Inquiry, 8.

Horgan, C., A. Taylor and G. Wilensky (1983). "Aging Veterans: Will They Overwhelm The VA Medical Care System." Health Affairs.

Page, W.F. (1982). "Why Veterans Choose Veterans Administration Hospitalizations: A Multivariate Model." Medical Care, Vol 20, No.3, March

Pindyck R.S. and D.L. Rubinfeld (1976). Econometric Models and Economic Forecasts, McGraw-Hill Book Co., New York, NY.

Romeis, J.C., Gillespie, K.N. and R.M. Coe (1988). "Older Veterans' Future Use of VA Health Care Services." Medical Care, September, Vol. 26, No. 9.

SAS Institute (1984). SAS/ETS User's Guide, 5th edn., SAS Institute, Cary, North Carolina.

Wolinsky, F.D., R.M. Coe, R.R. Mosely, and S.M. Homan (1987). "Length of Stay in the VA: Long-Term Care in Short-Term Hospitals." Medical Care, March, Vol. 25, No.3.

__________ (1985). "Veterans' and Nonveterans' Use of Health Services: A Comparative Analysis", Medical Care, Vol. 23, No.12, December.

TABLE 1

Parameter Estimates, T-Ratios, And 95% Confidence Intervals
For The Proposed ARIMA$(2,1,2)_{12} \times (2,1,2)$
To Forecast Short-Term And Mid-Term
Overall VA Inpatient Discharges

Parameter	Estimates	T-Ratio	95% Confidence Interval
MA1	0.28005	2.24	[-0.067, 0.626]
MA2	0.71202	7.48	[0.447, 0.977]
AR1	-0.99044	-12.53	[-1.210,-0.771]
AR2	-0.54162	- 6.41	[-0.776,-0.307]

TABLE 2
Associated Measure Used To Assess
The Inclusion Of Estimated Parameters
For The Proposed ARIMA Specification

Part A
<u>Correlations of the Estimates</u>

Parameter	MA1	MA2	AR1	AR2
MA1	1.000	-0.057	-0.318	0.480
MA2	-0.057	1.000	0.201	0.003
AR1	-0.318	0.201	1.000	0.370
AR2	0.480	0.003	0.370	1.000

Part B
<u>Autocorrelation Check of Residuals</u>

Number of Lags	R Statistic	Critical Chi-Squared	DF	Decision
6	5.40	5.99	2	R
12	14.12	15.51	8	R
18	25.94	23.68	14	A
24	41.35	31.35	20	A

TABLE 3
Twelve Period Back And Thirty-Six Period Ahead
Forecast Values And 95% Confidence Limit Bands
From Proposed ARIMA Specification

Obs.	Forecast	Lower 95%	Upper 95%	Actual
121	82640	77225	88055	82873
122	77736	72321	83151	77358
123	76978	71487	82471	73201
124	79521	73280	85762	76464
125	76702	70444	82856	74020
126	86973	80453	93494	82793
127	82371	75598	89144	80957
128	82320	75460	89181	84774
129	81117	74018	88216	75539
130	78450	71195	85706	78805
131	84620	77225	92014	84157
132	76235	68654	83817	79921
133	81016	72852	89179	
134	75534	67225	83844	
135	77002	68490	85514	
136	77653	68822	86485	
137	75502	66510	84494	
138	86136	76909	95363	
139	80813	71360	90266	
140	81280	71644	90915	
141	79954	70101	89808	
142	77128	67080	87177	
143	82522	73285	93758	
144	75002	64569	85436	
145	79795	68783	90807	
146	74374	63182	85566	
147	75775	64358	87192	
148	76460	64698	88222	
149	74311	62360	86263	
150	84924	72712	97137	
151	79621	67153	92088	
152	80080	67402	92757	
153	78751	65828	91675	
154	75933	62786	89079	
155	82320	68957	95684	
156	73803	60214	87391	
157	78597	64442	92752	
158	73174	58810	87538	
159	74576	59960	89192	
160	75261	60284	90238	
161	73112	57916	88307	
162	83725	68245	99205	
163	78421	62663	94180	
164	78880	62885	94876	
165	77552	61286	93817	
166	74733	58218	91247	
167	81121	64364	97877	
168	72603	55596	89610	

TABLE 4

The 1991 Actual VA Inpatient Hospital Discharges
Thirty-Six Month Ahead Forecast For VA Inpatient Discharges

1991 Month	1992 Actual	1993 Forecast	1994 Forecast	Forecast
October	82873	81016	79795	78597
November	77358	75534	74374	73174
December	73201	77002	75775	74576
January	76464	77653	76460	75261
February	74020	75502	74311	73112
March	82793	86136	84924	83725
April	80957	80813	79621	78421
May	84774	81280	80080	78880
June	75539	79954	78751	77552
July	78805	77128	75933	74733
August	84157	82522	82320	81121
September	79921	75002	73803	72603
Yearly Total	950862	949542	936147	921755
Annual % Change		-0.14	-1.41	-1.54

Figure 1

Monthly Time-Series For FY 1981-91

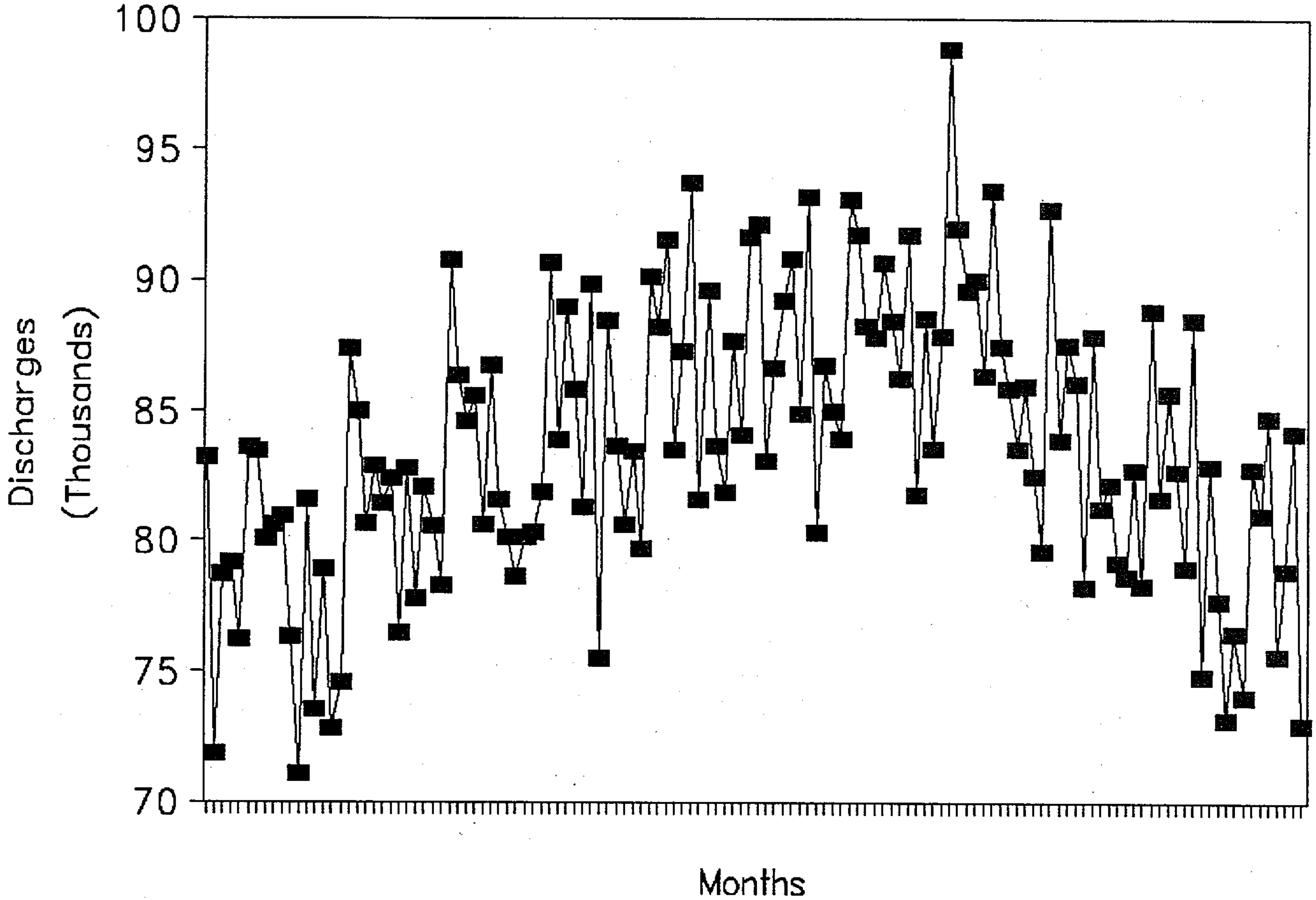


Figure 2

Correlogram For Monthly Time-Series

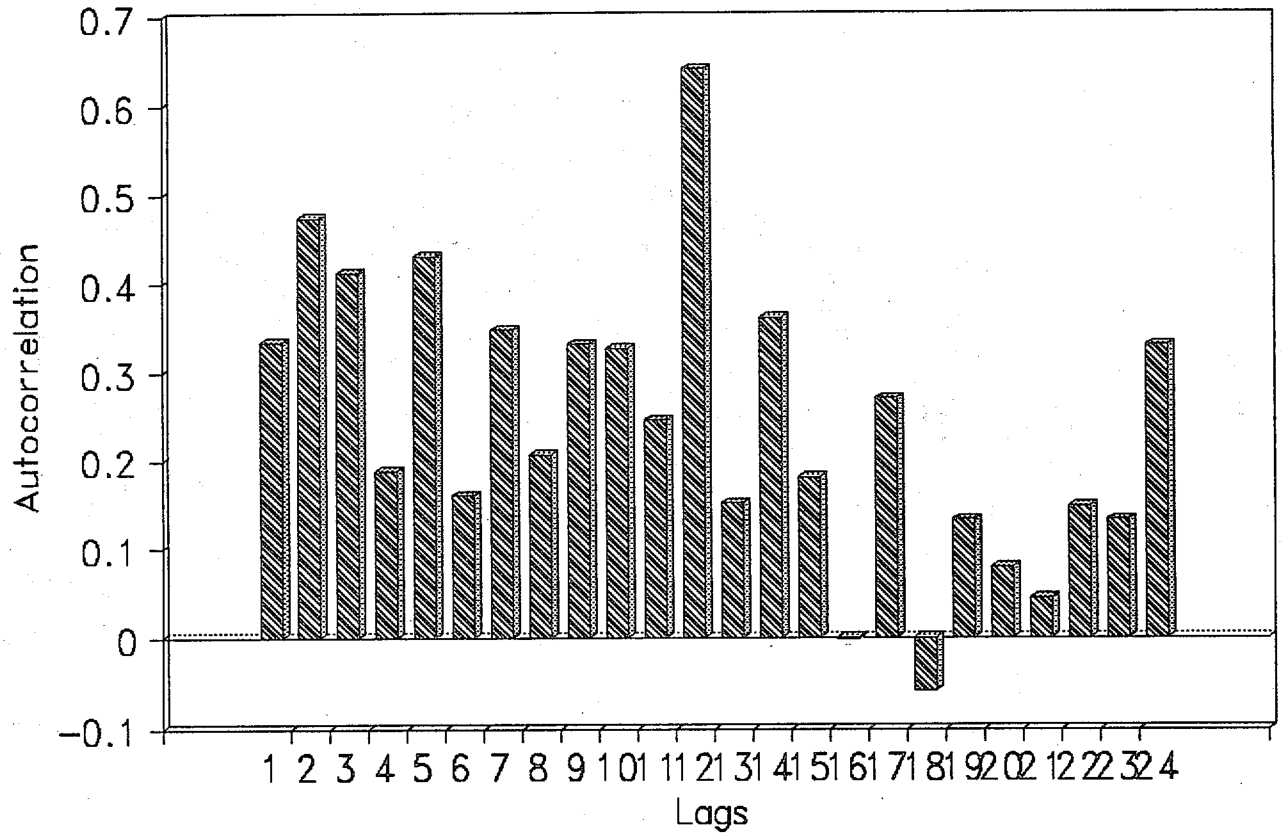


Figure 3
First Difference Of Monthly Time-Series

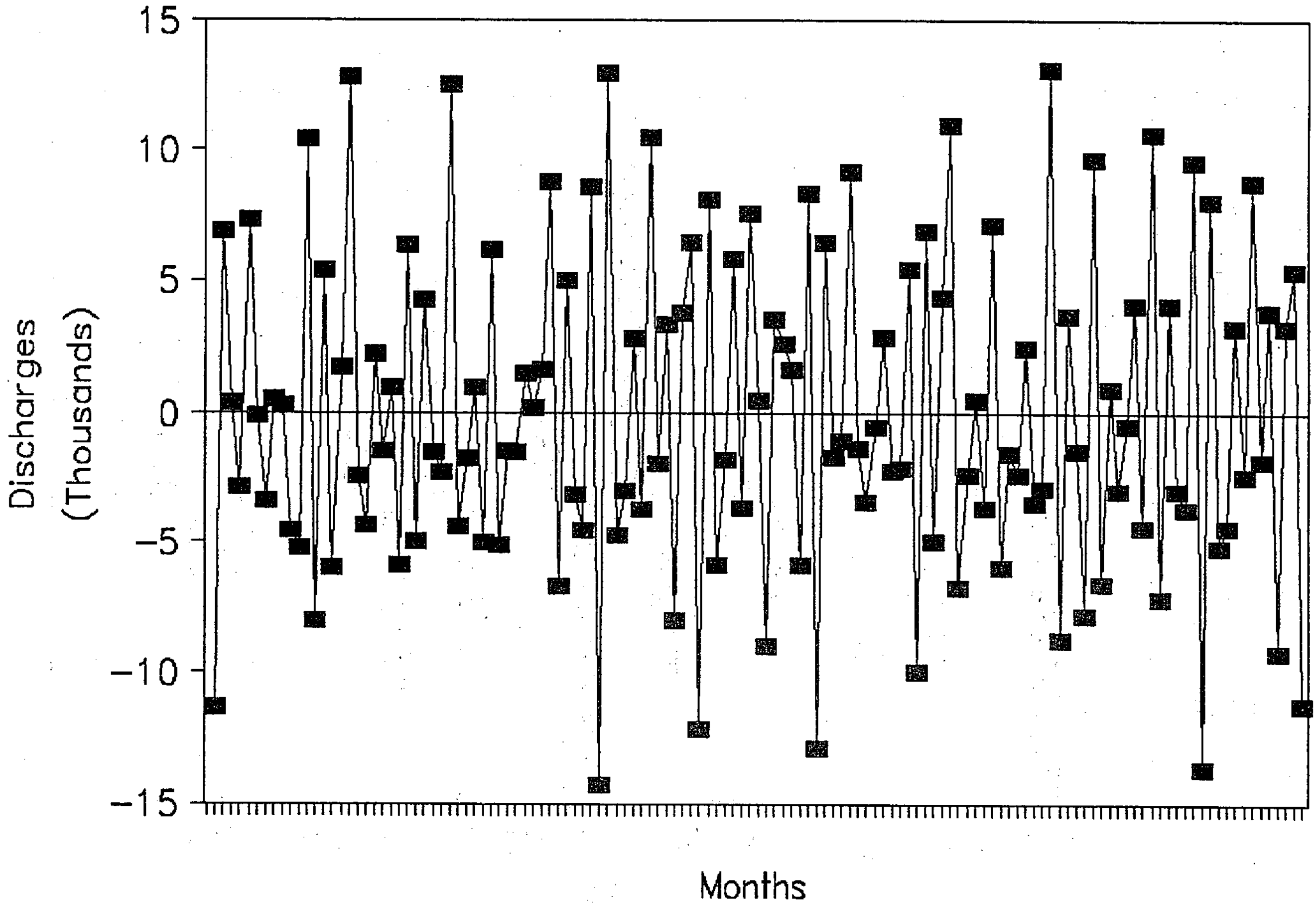


Figure 4

Correlogram For 1st Difference of Ser.

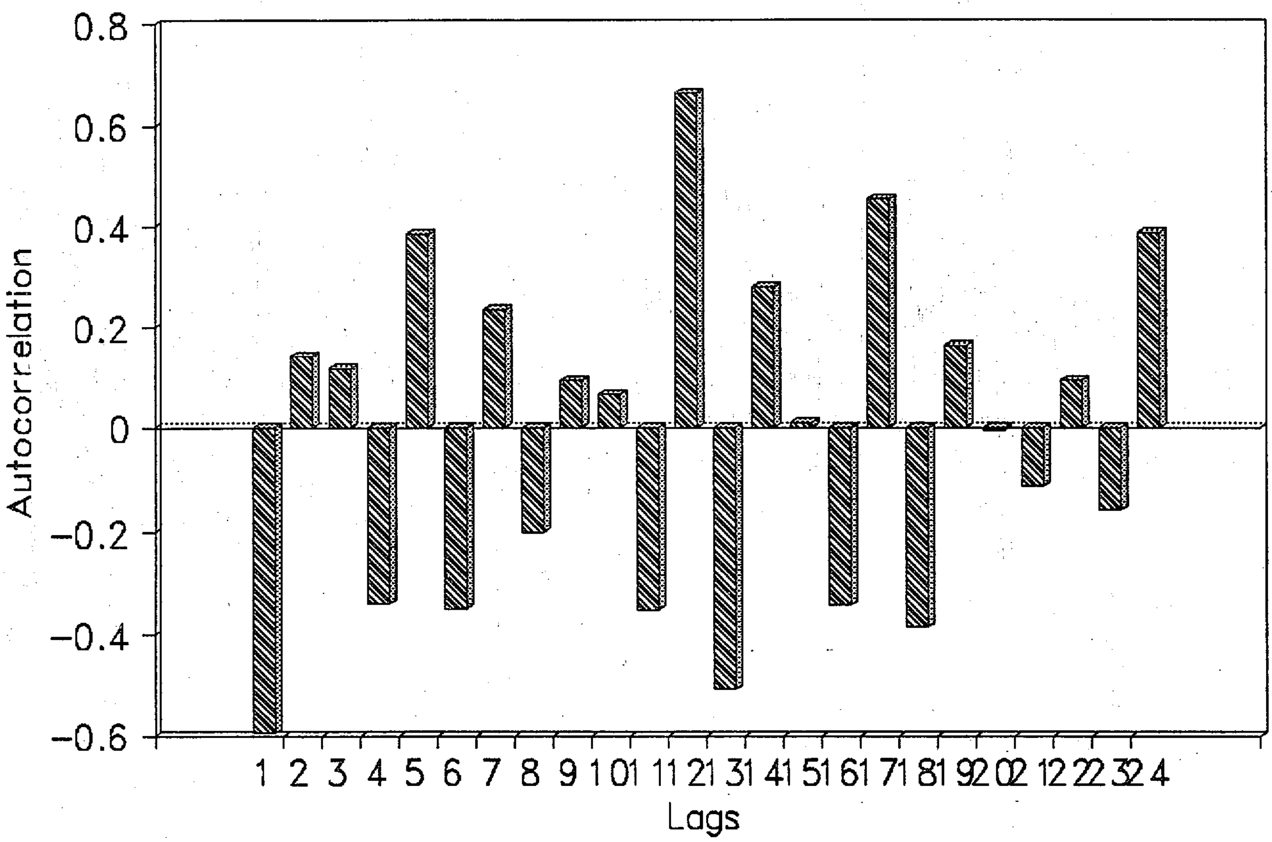


Figure 5
Correlogram For Seasonal Diff. Ser.

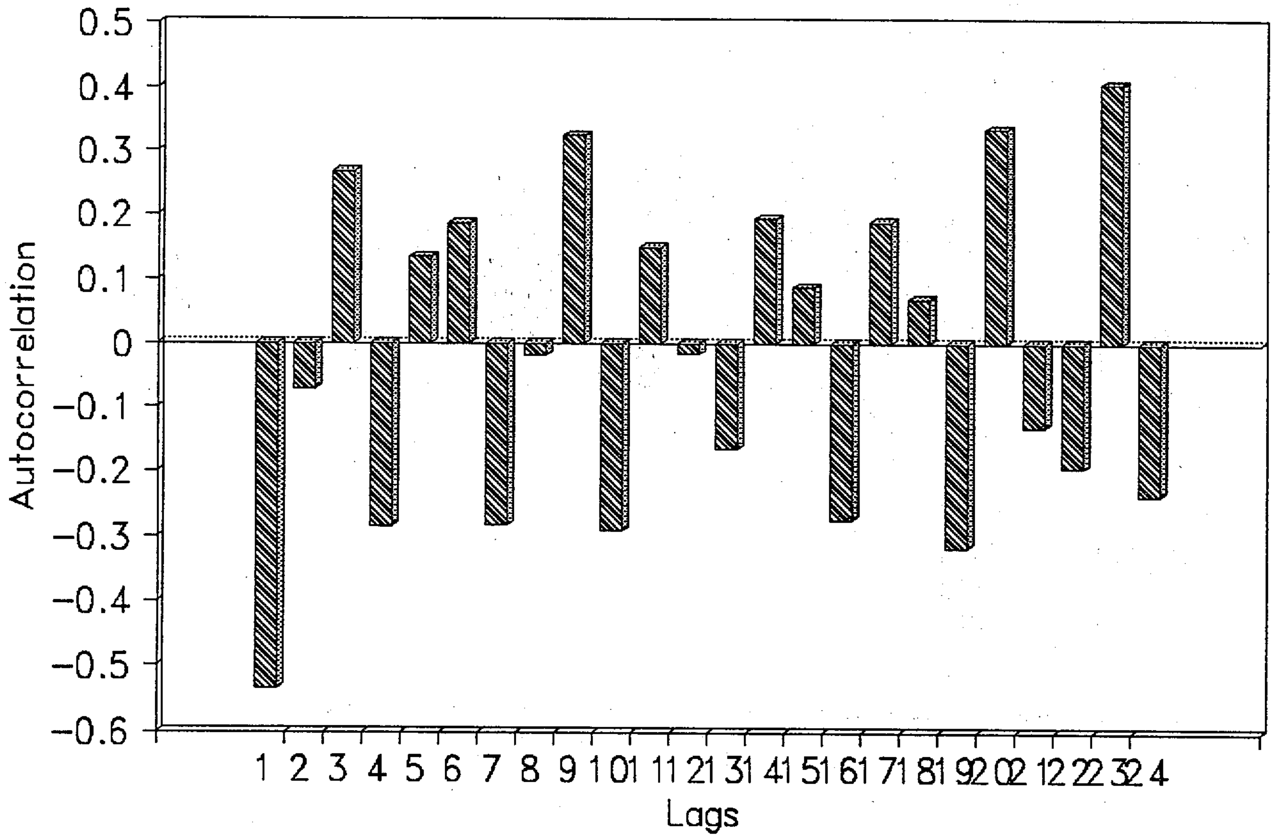


Figure 6

PAC Correlogram For Seasonal Diff. Ser.

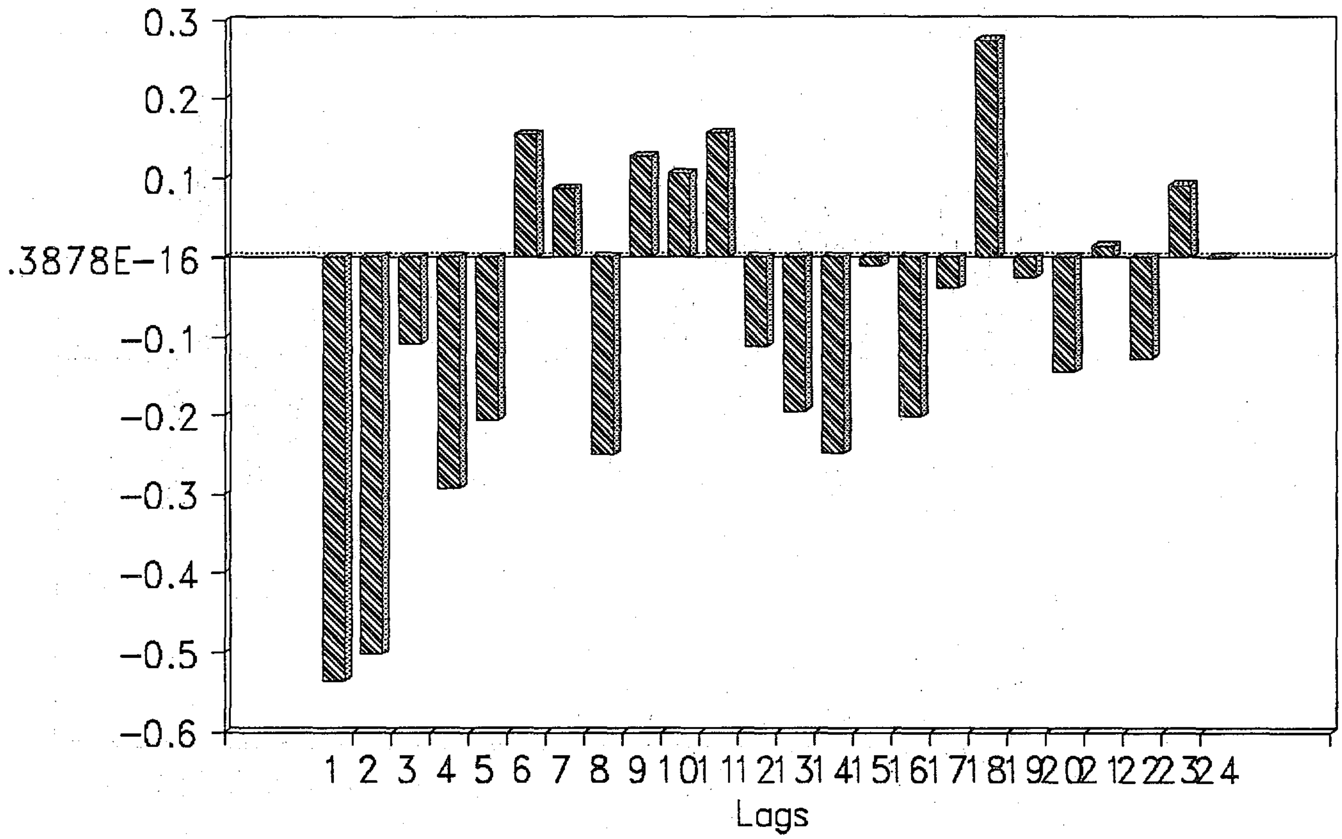
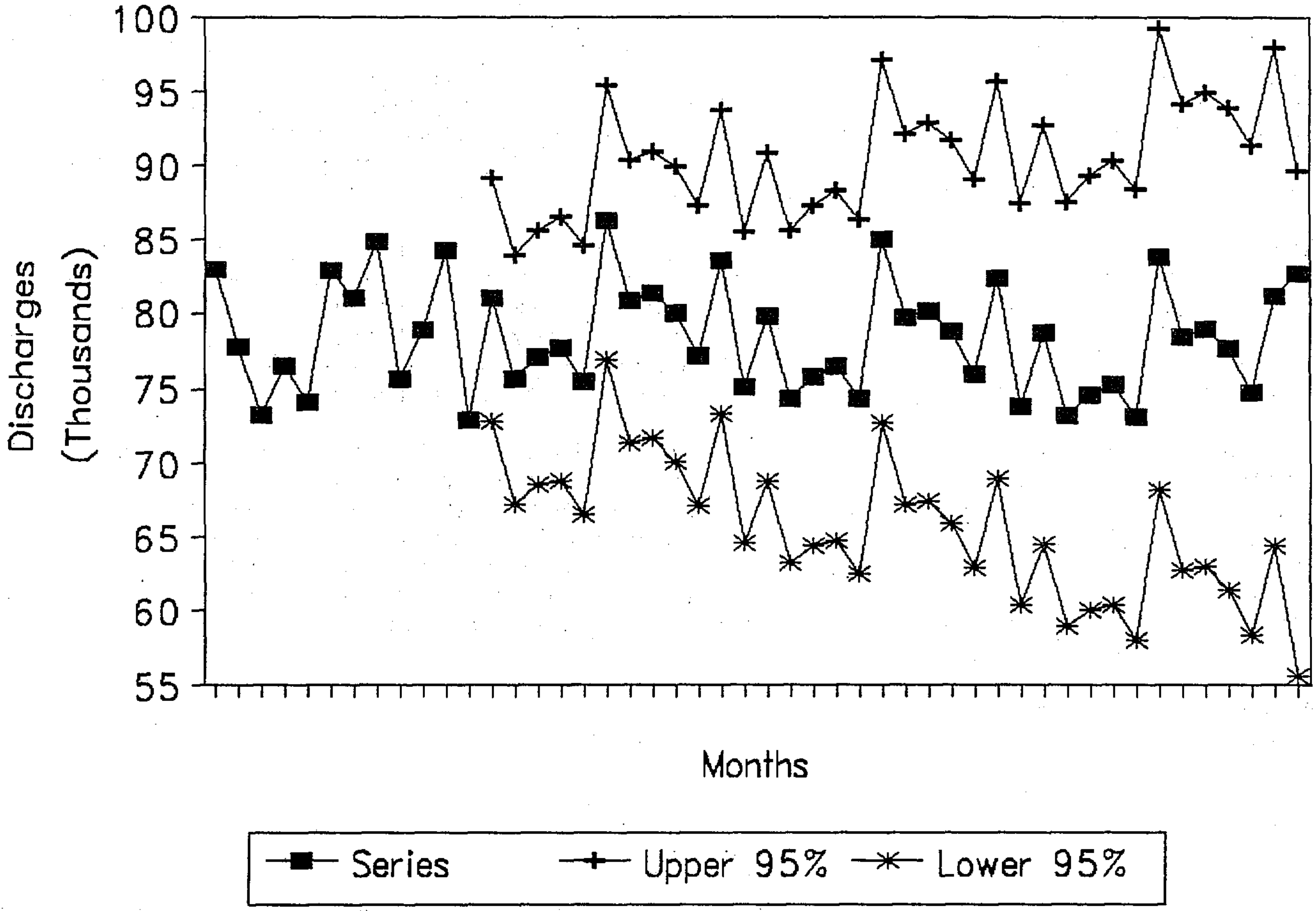


Figure 7

36 Month Ahead Forecast & 95% C.I.



Unit Roots and Fractional Differencing of Time Series Implications for Forecasting

Jeffrey S. Butler, IRS Research Division

This paper was prepared for the Fifth Annual Federal Forecasters Conference in Washington, D.C., 1992. All views expressed in the paper represent those of the author and not the Internal Revenue Service.

I. Introduction

Over the past decade, there has been much debate over whether a given nonstationary time series is made up of a deterministic or stochastic trend. Studying such a question has important implications for policy making in a broad number of fields. Traditional models of economic phenomena, for example, have favored a deterministic trend representation, which implies that any unpredictable shock to a process in a given time period will only be temporary, and that the path of the process will eventually revert back to its long-run trend. If an economic process is thought to respond to discretionary fiscal or monetary policy, this means that such policies can be conducted in the presence of such shocks because they are only temporary. If the long-run trend of a process is fundamentally stochastic, a given shock is permanent; discretionary policy under such circumstances may be counterproductive. If effective policy making depends on short-run forecast accuracy, it is shown below that post-sample forecast errors are significantly affected by misspecificaton of the trend component.

Nelson and Plosser (1982) cannot reject the hypothesis that many U.S. economic phenomena are made up of a stochastic trend. They use results from Dickey (1976), Dickey and Fuller (1979, 1981), Chan, Hayya, and Ord (1977), Plosser and Schwert (1977), Nelson and Kang (1981), and Beveridge and Nelson (1981), to discuss identification of such processes. Stochastic trends imply the presence of a unit root in the autoregressive representation of a time series, with Dickey and Fuller (1979, 1981) providing the most common method for testing such a hypothesis. The work of Evans and Savin (1981, 1984), Solo (1984), Said and Dickey (1985), and Phillips (1987) has also been at the forefront of this research. The analysis of time series with a unit root plays a major role in the theory and testing of cointegrated systems, as discussed in Granger (1986). Related work can be found in Sargan and Bhargava (1983) and Bhargava (1986).

More recently, the trend-break analysis of Christiano (1992), Peron (1989), Rappaport and Reichlin (1989), Christiano and Eichenbaum (1990), Chu and White (1992), and Banerjee, Lumsdaine, and Stock (1992), has suggested that the impact of shocks to national income is significantly overstated when income is represented as having a stochastic trend. This is because "big events" that cause a one-time structural change in trend are permanent and should not be confounded with non-structural shocks, which may only be temporary. To the extent that discretionary fiscal or monetary policy is concerned with the persistence of shocks over time, these studies may have broad implications.

A similar look at measures of persistence can be found in the literature on fractional differencing. Introduced by Mandelbrot and Van Ness (1968), and Mandelbrot and Wallis (1968, 1969), and developed in McLeod and Hipel (1978), Hosking (1981), Granger and Joyeux (1981), Geweke and Porter-Hudak (1982), Li and McLeod (1987), Diebold and Rudebusch (1989), and Sowell (1989, 1992), fractional differencing challenges the unit root hypothesis by suggesting that the order of integration of a stochastic trend need not be restricted to unity, but instead could take on a whole range of possible fractional orders of integration. Like trend-break analysis, measures of persistence may be overstated for time series that are of a fractional order of integration but incorrectly specified as having a unit root. These and related issues are discussed later in this paper.

II. Deterministic vs. Stochastic Trends

Define the trend stationary (TS) representation of a time series by

$$x_t = \beta t + u_t \tag{2.1}$$

$$\phi(B)u_t = \Theta(B)e_t,$$

where $e_t \sim \text{iidN}(0,\sigma^2)$, and B is the backshift operator such that $z_tB^k = z_{t-k}$. The usual boundary conditions are placed on the roots of the polynomials $\phi(B) = (1-\phi_1B-\cdots-\phi_pB^p)$ and $\Theta(B) = (1-\Theta_1B-\cdots-\Theta_qB^q)$ to ensure stationarity and invertibility. Thus, $E(u_t)=0$, which implies that the variance of x_t is bounded. This means that any change to x_t through u_t will be temporary; its effect will eventually die off and the path of x_t will revert back to its long-run mean of $E(x_t) = \beta t$.

Now consider a random walk without drift, which belongs to the class of difference stationary (DS) processes:

$$x_t = x_{t-1} + u_t, \tag{2.2}$$

with disturbances generated from the same stationary ARMA model as the TS specification. Setting $x_1=0$ (without loss of generality), the solution $x_t = \Sigma u_{t-j}$, $j=1,2,\ldots t$, implies that the DS series is simply an accumulation of shocks over time.

Therefore, unlike the TS model, the variance of x_t in the DS representation is unbounded; uncertainty grows in proportion to the forecast horizon. This is the fundamental difference between the two models.

Most nonstationary time series can be decomposed into a trend and cycle component, with the latter representing the short-run dynamics of the system which are often of most interest to policy makers. Obtaining an accurate representation of the stationary cyclical component for estimation and forecasting crucially depends on the method of detrending. Three methods of identifying the proper trend specification are discussed below.

First, let x_t be the DS process in (2.2), and suppose we incorrectly assume a TS representation and run the regression

$$x_t = \alpha + \beta t + e_t.$$

This misspecification will result in sample autocorrelations of $a_t = x_t - E(x_t)$ that will be characteristic of a random walk, i.e., a large lag-one correlation followed by a slow decay function. Although use of the sample autocorrelation function for identification purposes is not without problems, it can be used as a first approximation to the existence of any potential misspecification. Second, consider the following DS process:

$$x_t = \alpha + x_{t-1} + e_t.$$

With $x_1 = 0$, repeated substitution gives $x_t = \alpha t + \Sigma u_{t-j}$, $j = 1,2,\ldots,t$. This is the random walk with drift model, and suggests a direct test for deterministic versus stochastic trend by forming the regression

$$x_t = \delta x_{t-1} + \alpha t + e_t$$

with H_o: $\delta = 1$. Equivalently, we can write

$$x_t = (1 + \delta)x_{t-1} + \alpha t + e_t,$$

or

$$(1-B)x_t = \delta x_{t-1} + \alpha t + e_t \qquad (2.3)$$

with H_o $\delta = 0$. This is the essence of the Dickey-Fuller test for a unit root in the autoregressive representation of the series, and is therefore a test of deterministic versus stochastic trend. The usual t-statistic cannot be used here, however, and Dickey and Fuller (1979) have tabulated critical values associated with this test. For the regression model in (2.3), failing to reject the null corresponds to the presence of a unit root.

Finally, although it is now standard to take the DS specification as the null, suppose that we start with the TS process given in (2.1), but incorrectly assume that the process has a DS representation. Upon taking first differences we obtain

$$(1-B)x_t = \beta(1-B)t + (1-B)e_t$$

or

$$(1-B)x_t = \beta + e_t - e_{t-1}$$

which leaves a unit root in the moving average representation. This suggests that (1) the lag-one autocorrelation coefficient will be close to the boundary region, with inverse autocorrelations following a pattern characteristic of a random walk, or (2) convergence problems will occur during estimation due to non-invertibility.

While the three cases above discuss the theoretical consequences of misspecification and how to identify them, the identification stage in practice should seek to employ as many methods as possible to ensure proper representation. This is especially important when applied research is limited by small samples, or when new data for which no a priori assumptions about trend representation exist.

There are two primary reasons why misspecification should be avoided. First, it would seem intuitively obvious that errors from forecasting would be larger on average for misspecified models. To test such intuition for the rival trend models, a small Monte Carlo experiment was conducted to analyze one-step ahead forecast errors from a DS process that was incorrectly specified as TS. For sample sizes of 50 and 100 observations, 500 replications for each of three ARMA(1,0), ARMA(0,1), and two ARMA(1,1) models were generated. Estimation using maximum likelihood was performed for both the DS and TS specifications, and one-step ahead root mean square errors (RMSE) calculated. Table 1 reports results from the experiment.

Table 1.
One-step RMSE for DS and TS specifications of $\phi B)(1—B)_t = \Theta(B)e_t$.*

	T=50		T=100	
ARMA(1,0)	DS	TS	DS	TS
ϕ=.25	1.003	0.997	0.974	0.994
ϕ=.50	1.008	1.135	1.012	1.141
ϕ=.75	0.984	1.418	0.983	1.443
ARMA(0,1)	DS	TS	DS	TS
Θ=-.25	0.978	1.531	1.001	1.864
Θ=-.50	0.979	1.647	1.006	2.229
Θ=-.75	1.012	1.886	0.994	2.374
ARMA(1,1)	DS	TS	DS	TS
ϕ=.5,Θ=-.8	1.026	1.085	1.019	1.107
ϕ=.8,Θ=-.5	1.015	1.336	0.995	1.378

* For the DS specification, first differences were taken and ARMA parameters were estimated using maximum likelihood. For the TS specification, maximum likelihood was used to simultaneously estimate all parameters in the model $x_t = \alpha + ßt + u_t$, where $\phi()u_t = \Theta(Be_t$.

Because the short-run, forecastable momentum in either a TS or DS process is its stationary ARMA representation, this simulation looks only at one-step ahead errors for answers concerning the effects of misspecification. Table 1 shows that forecasting a DS process using a TS specification generally results in larger one-step RMSEs than those from the correctly specified DS process. This seems to be especially true in the case of first order moving average disturbances. There is little evidence to suggest that differences in errors vary with sample size, although a more comprehensive simulation should be conducted before drawing such a conclusion.

The second reason that misspecification should be avoided relates to the notion of persistence, which is a measure of the total change in output of a process due to a unit input change. For example, a given shock to real U.S. GNP (unexpected U.S. layoffs in heavy industry, unanticipated interest rate cut in Germany, etc.) in time period will ultimately affect real GNP in period t+k. The questions are: (1) by how much, and (2) over what period of time? The answer to the second question can be found by observing the decay process of the coefficients of the moving average representation, but in practice may never be known precisely. The answer to the first question is given by the sum of the coefficients of the moving average representation of the process. For example, consider the stationary process

$$\phi(B)u_t = \Theta(B)e_t$$

or

$$u_t = \phi^{-1}(B)\Theta(B)e_t.$$

The mean, or expected value, of u_t is $E(u_t) = \Theta(1)/\phi(1)$, which is also its equilibrium solution If the "surprise" through e_t was exactly one unit, the new equilibrium solution for u_t would be

$$E(u_t) = [\Theta(1)/\phi(1)](E(e_t)+1) = \Theta(1)/\phi(1),$$

whose sum can be determined numerically by either equating coefficients of powers of B or by expanding the ratio in partial fractions when the order of the respective lag polynomials is known.

As shown in Chan, Hayya, and Ord (1987), and discussed in Watson (1986) and Schwert (1987), inappropriate detrending methods can significantly alter the parameterization of the resulting stationary ARMA process. Terms that should not be present or are absent from the true ARMA representation will mean differences in the measure of persistence, as discussed above. If discretionary policy is sensitive to the persistence of "surprises" that affect a policy variable, decisions based on measures of persistence may be less effective when models are misspecified.

In summary, any priori assumptions that a given nonstationary time series is made up of a deterministic trend component should be tested on the basis of one of three identification procedures outlined above. The consequences of misspecifying the trend component are (1) larger short-run forecast errors, (2) inaccurate measures of persistence, and (3) the potential for less effective policy action.

III. Fractional vs. Unit Roots

The discussion thus far concerning tests for deterministic vs. stochastic trends have assumed that, under the null hypothesis of a difference stationary process, the order of integration (differencing) is integer. Although it may be intuitively more appealing, as well as operationally easier to work with, the assumption of an integer order of integration fails to account for the wide variety of important stochastic behavior that results from allowing fractional orders of integration.

To introduce the nature of fractional differencing, consider the stationary ARIMA(0,d0) process

$$(1—B)^d x_t = e_t,$$

with $e_t \sim iidN(0,\sigma^2)$ and d=0. This implies that $x_t$is white noise, whose spectrum has equal contributions to variance at every frequency. Now define $0 < d < 1/2$. For this range of d, x_t has the autoregressive representation:

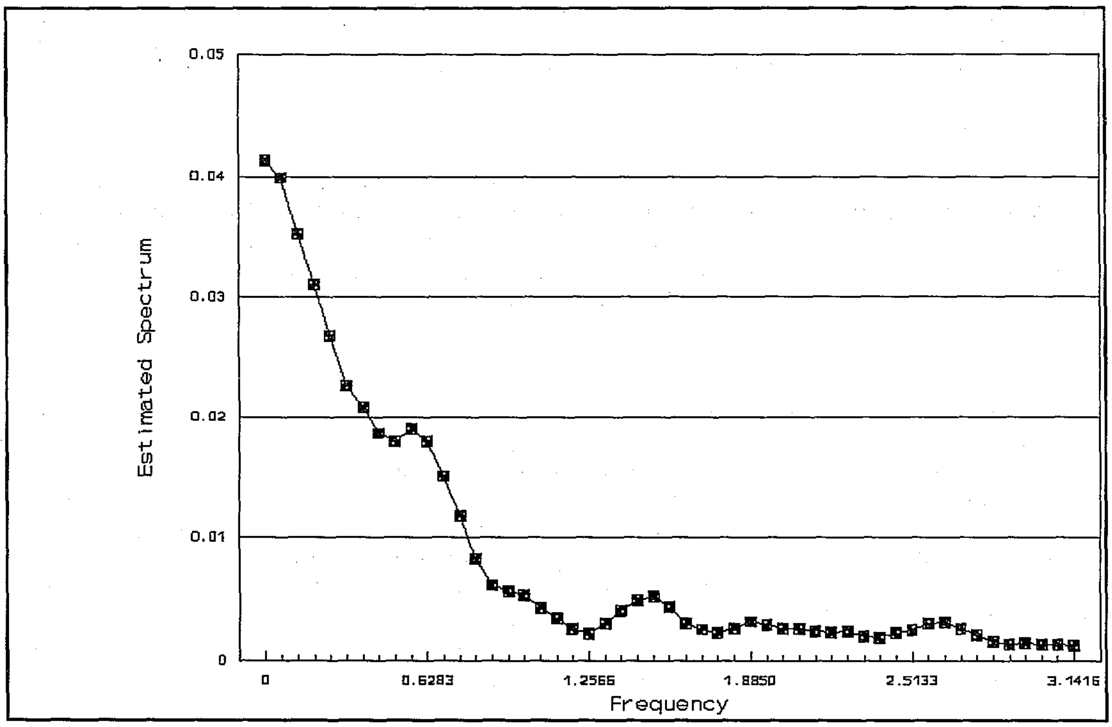
$$(1—dB+(1/2)d(d—1)B^2—\cdots)x_t = e_t$$

upon expansion of $(1—B)^d$ Solving for x_t gives $x_t = (1—B)^{-d}e_t$, which is stationary with moving average representation

$$x_t = \Sigma\pi_j e_{t-j}, \quad \pi_j = (d—1+j)!/j(d—1)! \qquad (3.1)$$

where j=1,2,.... When $0 < d < 1/2$, the process is said to have long memory and is characterized by autocorrelations r(k) that decay hyperbolically, $r(k) \sim k^{2d-1}$, for large k, instead of the exponential decay of an ordinary ARMA(p,q) process. However, the process in (3.1) is ARMA(0,d,0), i.e., fractional Gaussian noise, and will possess this property regardless of its paramterization. An autocorrelation function (ACF) that decays hyperbolically implies that correlations of the series at long lags will be stronger than those from a series with exponentially decaying ACF. To make this even more meaningful, consider the sample spectrum of a stationary ARIMA(0,.2,0) process shown in Figure 1.

Figure 1. Estimated Spectrum of an ARIMA(0,.2,0) Process



The estimated spectrum is dominated by low frequencies, characteristic of a nonstationary process, or one with positive long-lag correlations. The ARIMA(0,d0) process, however, is stationary for $0 < d < 1/2$. Therefore, fractionally integrated processes are capable of exhibiting a much richer variety of long-run dynamics than a process of integer order. More importantly, by allowing for fractional orders of integration in an ARIMA(p,d,q) model, one is able to capture both the long-run dynamics through fractional as well as short-run dynamics through the ARMA(p,q) parameterization.

The above discussion suggests that specifying the correct order of integration can provide more flexibility in modeling certain economic phenomena. Therefore, misspecifying the order of integration should be avoided. To analyze the consequences of misspecification, consider the following nonstationary process:

$$\phi(B)(1-B)^d y_t = \Theta(B)e_t,$$

or

$$(1-B)^d y_t = C(B)e_t$$

where $C(B) \equiv \phi^{-1}(B)\Theta(B)$, and $e_t \sim iidN(0,\sigma^2)$. Factor $(1-B)^d$ into $(1-B)(1-B)^{d-1}$. If $d=1$, we have a unit root, but if $1/2 < d < 1$ (for the nonstationary series) and we take first differences, the result is over-differencing. The effects of over-differencing are discussed in Plosser and Schwert (1977), and Plosser, Schwert, and White (1982). In general, over-differencing results in convoluting to the original process with unwanted information and tends to give rise to spurious and pseudo-periodic behavior in the autocorrelation function. As pointed out in Chatfield (1984), the first-difference filter has poor spectral cut-off properties as it is, so that compounding this problem with the issues raised above may only complicate the use of differencing as a diagnostic tool during the identification process.

A second consequence of misspecification has to do with post-sample forecast performance. Specifically, while short-run forecast errors will tend to be similar for stationary ARMA(p,q) and ARIMA(p,d,q) models, $0 < d < 1/2$, forecast errors over longer post-sample horizons will be smaller if a fractionally integrated model is correctly specified. See, for example, Mandelbrot and wallis (1969), McLeod and Hipel (1978), or Granger and Joyeux (1981). As pointed out above, this is due to the fact that correlations at long lags are much stronger for fractionally integrated processes.

Forecasting in cointegrated systems will also be affected by misspecification. Suppose you want to test whether two processes, x_t and y_t, are cointegrated. If these two processes form an equilibrium system, then such a system will satisfy equilibrium conditions if $x_t - \alpha y_t = 0$, i.e., a linear combination of the processes is zero. Economic systems often exhibit disequilibria in the short-run, but in the long-run the mean relationship may have an equilibrium solution. Satisfying such modeling criteria requires only the definition of a stationary, mean-zero process, z_t, such that $x_t - \alpha y_t = z_t$. If a vector $(1-\alpha)$ exists such that z_t is integrated of order zero, denoted $z_t \sim I(0)$, then x_t and y_t are said to be cointegrated. Determining whether $_t \sim I(0)$ is similar to using Dickey-Fuller procedures to test whether a process x_t has a unit root, i.e., if $x_t \sim I(1)$. See Engle and Granger (1987), and Engle and Yoo (1987) for details. An obvious problem arises, however, if the two processes, x_t and y_t, are not integrated of the same order: cointegration tests are invalid. Therefore, assuming they are of the same integer order of integration when, say, $x_t \sim I(1)$ and $y_t \sim I(d)$, for fractional value of d implies that hypothesis testing and forecasting will be affected.

Finally, measures of persistence may be distorted under conditions of misspecification. Both Diebold and Rudebusch (1989), and Sowell (1992) find evidence that the persistence of changes to real quarterly U.S. GNP have been significantly overstated when modeling is based on either the deterministic or stochastic trend hypothesis. These empirical findings support the above discussion of over-differencing and its effects on the autocorrelation function (and hence the parameterization of a stationary ARMA representation).

IV. Conclusion

If a nonstationary time series can be decomposed into trend and cycle components, the extent to which the trend is properly specified plays a key role in estimating and forecasting the stationary, cyclical momentum of the series. Much debate over the past decade has focused on two rival models for trend representation, one based on a deterministic function of time (trend stationary) and the other based on an accumulation of random shocks (difference stationary). It was shown that misspecifying a stochastic trend as one that is deterministic results in one-step forecast errors which are larger than those for the true model. Measures of persistence will also be affected by misspecification. Conversely, estimation and forecasting can be dramatically affected when the order of integration of a stochastic trend is misspecified as integer if it is truly fractional. A much richer variety of long-run dynamics can be captured by fractionally integrated processes, due to its long memory. As such, post-sample forecast errors over a long forecast horizon will tend to be smaller for fractionally integrated processes that are properly specified.

References

Banerjee, A., Robin L. Lumsdaine, and James H. Stock, 1992, Recursive and sequential tests of the unit-root and trend-break hypotheses: Theory and international evidence, Journal of Business and Economic Statistics, 10, 271-287.

Beveridge, Stephen, and Charles R. Nelson, 1981, A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle,' Journal of Monetary Economics, 7, 151-174.

Bhargava, A., 1986, On the theory of testing for unit roots in observed time series, mimeographed, University of Pennsylvania.

Chan, K. Hung, Jack C. Hayya, and J. Keith Ord, 1977, A note on trend removal methods. The case of Polynomial versus variate differencing, Econometrica, 45, 737-744.

Chatfield, C., 1984, The Analysis of Time Series, New York, Chapman and Hall.

Christiano, L. J., 1992, Searching for a break in GNP, Journal of Business and Economic Statistics, 10, 237-250.
______, and M. Eichenbaum, 1990, Unit roots in real GNP: Do we know, and do we care?, Carnegie-Rochester Conference Series on Public Policy, 32, Spring, 7-61.

Chu, James Chia-Shang, and Halbert White, 1992, A direct test for changing trend, Journal of Business and Economic Statistics, 10, 289-299.

Diebold, F. X., and G. D. Rudebusch, 1989, Long memory and persistence in aggregate output, Journal of Monetary Economics, 24, 189-209.

Dickey, David A., Estimation and hypothesis testing in nonstationary time series, Unpublished doctoral dissertation (Iowa State University, Ames, IA).
Dickey, David A., and Wayne A. Fuller, 1979, Distribution of the estimators for autoregressive time series with a unit root, Journal of the American Statistical Association, 74, 427-431.
________________________, 1981, Likelihood ratio statistics for autoregressive time series with a unit root, Econometrica, 49, 1057-1072.

Engle, R. F., and C. W. J. Granger, 1987, Cointegration and error correction: Representation, esimation and testing, Econometrica, 55, 251-276.
___, and B. S. Yoo, 1987, Forecasting and testing in co-integrated systems, Journal of Econometrics, 35, 143-159.
___, and N. E. Savin, 1981, Testing for Unit Roots: I, Econometrica, 49,
753-779.
___, 1984, Testing for Unit Roots: II, Econometrica, 52,
1241-1269.

Geweke, J. F., and Susan Porter-Hudak, 1982, The estimation and application of long memory time series models, Journal of Time Series Analysis, 4, 221-238.

Granger, C. W. J., 1986, Developments in the study of cointegrated economic variables, Oxford Bulletin of Economics and Statistics, 48, 213-228.
____, and Roselyne Joyeux, 1981, An introduction to long-memory time series models and fractional differencing, Journal of Time Series Analysis, 1, 15-30.

Hipel, K. W., and I. A. McLeod, 1978, Preservation of the rescaled adjusted range, part II, Simulation studies using Box-Jenkins models, Water Resources Research, 14,509-516.
________, 1978, Preservation of the rescaled adjusted range, part III, Fractional Gaussian noise algorithms, Water Resources Research, 14, 517-518.

Hosking, J. R. M., 1981, Fractional differencing, Biometrika, 68, 165-176.

Janacek, G. J., 1982, Determining the degree of differencing for time series via the log spectrum, Journal of Time Series Analysis, 3, 177-183.

Kashyap, R. L., and Eom, K., 1988, Estimation in long memory time series model, Journal of Time Series Analysis, 9, 35-41.

References cont.

Li., W. K., and A. I. McLeod, 1986, Fractional time series modelling, Biometrika, 73, 217-221.

Mandelbrot, B. B., and J. W. Van Ness, 1968, Fractional brownian motions, fractiona noises and applications, SIAM, 10, 422-437.

Mandelbrot, B. B., and J. R. Wallis, 1968, Noah, Joseph, and operational hydrology, Water Resources Research, 4, 909-918.

______, 1969, Computer experiments with fractional Gaussian noises, part I, averages and variances, Water Resources Research, 5, 228-241.

______, 1969, Computer experiments with fractional Gaussian noises, part II, rescaled ranges and spectra, Water Resources Research, 5, 242-259.

______, 1969, Computer experiments with fractional Gaussian noises, part III, mathematical appendix, Water Resources Research, 5, 260-267.

______, 1969, Some long run properties of geophysical records, Water Resources Research, 5, 321-340.

______, 1969, Robustness of the rescaled range R/S in the measurement of non-cyclic long-run statistical dependence, Water Resources Research, 5, 967-988.

Mandelbrot, B. B., 1971, A fast fractional Gaussian noise generator, Water Resources Research, 7, 543-553.

McLeod, A. I., and K. W. Hipel, 1978, Preservation of the rescaled adjusted range, part I, A reassessment of the Hurst phenomenon, Water Resources Research, 14, 491-508.

Nelson, Charles R., and Heejoon Kang, 1981, Spurious periodicity in inappropriately detrended time series, Econometrica, 49, 741-751.

____, and Charles I. Plosser, 1982, Trends and random walks in macroeconomic time series, Journal of Monetary Economics, 10, 140-162.

Peron, p., 1989, The great crash, the oil price shock, and the unit root hypothesis, Econometrica, 57, 1361-1401.

Phillips, P. C. B., 1987, 1987, Time series regression with a unit root, Econometrica, 55, 227-301.

Plosser, Charles I., and G. William Schwert, 1977, Estimation of a noninvertible moving average process: The case of overdifferencing, Journal of Econometrics, 6, 199-224.

_____, G. W. Schwert, and H. White, 1982, Differencing as a test of specification, International Economic Review, 23, 535-552.

Rappaport, P., and L. Reichlin, 1989, Segmented trends and non-stationary time series, Economic Journal, 99, Conference Supplement, 168-177.

Said, Said E., and David A. Dickey, 1984, Testing for unit roots in autoregressive moving-average models of unknown order, Biometrika, 71, 599-607.

___, 1985, Hypothesis testing in ARIMA(p,1,q models, Journal of the American Statistical Association, 80, 369-374.

Sargan, J. D., and A. Bhargava, 1983, Testing residuals from least squares regression for being generated from a Gaussian random walk, Econometrica, 51, 153-174.

Schwert, G. W., 1987, Effects of model specification on tests for unit roots in macroeconomic data, Journal of Monetary Economics, 20, 73-103.

Solo, V., 1984, The order of differencing in ARIMA models, Journal of the American Statistical Association, 79, 916-921.

Sowell, F. B., 1990, The fractional unit root distribution, Econometrica, 58, 495-505.

____, 1992, Modeling long-run behavior with the fractional ARIMA model, Journal of Monetary Economics, 29, 277-302.

Watson, M. W., 1986, Univariate detrending methods with stochastic trends, Journal of Monetary Economics, 18, 49-75.

Augmenting Statistical Forecasting Techniques with Neural Networks

Mansur Arbabi, Ph.D and Scott Fischthal, IBM Federal Systems Company
Jeffrey S. Butler, IRS Research Division

ABSTRACT: Forecasting of tax returns by form type for geographical districts is typically performed using traditional statistical methods. We assess the impact of neural networks on forecasting and how they contribute to traditional linear and nonlinear statistical methods. Our results show that combining neural networks with these statistical methods may provide better forecasts for this particular problem than statistics alone. Models combining neural networks and other statistical methods promise to use the best of both techniques. Several networks are provided which compare various architectures, preprocessing of data and training methods; our discussion highlights why some of these models produce better results than others. We also discuss other nonlinear statistical techniques and their relationships to our neural networks.

I. Introduction

The objective of this applied research is to investigate the applicability of the neural network methodology to the forecasting of tax returns.

The U.S. Internal Revenue Service (IRS) projects resource needs at each of 63 district offices as much as 15 years in advance. Such projections make it possible to anticipate shifts in the population of return filers, for example, thereby allowing sufficient time to plan and allocate necessary resources associated with such shifts. The more accurate the resource projections, the more efficient the allocation process.

A useful and effective method for defining the level of resource needs is to measure the level of tax returns being filed in a given IRS district. That is, the level of filing activity is a proxy for resource needs. Filing activity is typically decomposed into specific categories. For example, individuals have three primary returns at their disposal: Forms 1040, 1040A, and 1040EZ. The volume of filing in any category provides valuable information about the makeup of taxpayers in a given geographic area, and thus the resource needs of that area as well.

The IRS Research Division is currently responsible for making projections of tax return filing for over 150 individual and business return categories at each of the 63 IRS districts. Both parametric and nonparametric statistical methods are used to estimate and forecast the thousands of time series: ordinary and generalized least squares; ARIMA; stepwise autoregression; exponential smoothing; weighted moving averages; subjective methods, and so on. Many factors affect forecast accuracy. Regression models rely on third party economic and demographic data as inputs. New tax legislation is passed frequently, contributing to structural breaks in the data. Recently introduced tax forms typically have short base periods. Other factors that can affect forecast accuracy include model misspecification, forecasting deadlines, and accounting relationships in the data that must be preserved. It is not surprising, therefore, that new methods are constantly being pursued in an attempt to more fully automate and improve the accuracy of this forecasting problem.

We introduce the concept of using neural networks to help offer better solutions to these problems. We demonstrate that neural nets perform well in this environment. We also explore the combination of neural networks with traditional statistical methods. Such a hybrid model can use the advantages of both methods -- statistical results are easier to explain and analyze, while neural networks consider a wider range of variables and require less prior knowledge of the functional form of the trends being explored.

II. Comparison of neural networks and traditional statistical models

Single equation regression and time series analysis methods, although useful in many applied problems, have many operational drawbacks. One of the most significant drawbacks of these methods as they relate to forecasting tax return filing volumes, for example, is that they do not explicitly use interrelationships between filing variation across geographic areas. That is, each district's filing pattern is treated as a single equation--unaffected by trends in other districts. Where regression analysis is concerned, this implies that interrelationships between economic and demographic variables across geographic areas are also ignored. (Interestingly, the filing data available across geographic locations over time form an ideal panel, but as yet no panel analysis, estimation, or forecasting has been attempted.) Because many neural networks are based on clustering algorithms, they may be better suited in exploiting information present in these interrelationships, and thereby increasing the likelihood of improved forecast accuracy.

The appropriate statistical model must be specified by the statistical analyst before using regressions and discriminant analysis to perform any numerical computation on the data. An approach using neural networks is more general; there is no need to specify a model ahead of time. The neural network effectively discovers the appropriate model that suits the data.

One advantage the statistical approach has over neural networks is that it is more explicit and thus easier to interpret its results. Statistical models have assumptions of an underlying distribution of data not necessarily required by neural networks. Neural nets are often more robust and thus generalize better when underlying processes are nonlinear and distributions are strongly non-Gaussian. Another advantage that neural networks have over the statistical approach is that they can be made to adapt when more data becomes available.

Statistics	Neural Networks
Often linear relationship between variables assumed	Any function (Kolomogorov's Theorem)
No interaction between data values assumed	All interactions allowed
Underlying model must be specified (nonlinear and linear methods)	Underlying model generated as part of network training
Distribution of data is usually assumed	Distribution-free
Fixed model	Can be adaptive
Significance tests for input, confidence intervals for forecasted values	No such tests available
Difficult to handle noisy or incomplete data	Robust enough to handle noisy or incomplete data

There are many differences between statistical methods and neural networks because of the existence of numerous models in each field. To limit the scope, for the purposes of this paper we will compare regression, exponential smoothing and stepwise autoregression to feed-forward, backpropagation networks as described in [Rumelhart 86].

A feed-forward neural network can be viewed as an extension of a statistical regression model. Nonlinear neural network models have been shown to be universal approximators of any measurable, continuous function by Kolomogorov [Hecht-Nielsen 87]. Neural networks should be capable of producing the same result as a linear regression on data that has no nonlinearities. We now briefly describe the two methods in a network formulation.

A typical linear model consists of estimating parameters $(a_0,\ldots,a_n)$ in the equation

$$Y=a_0+a_1x_1+\ldots+a_nx_n$$

where $(x_1,\ldots,x_n)$ is an observed (input) vector and Y is the predicted value (Figure 1).

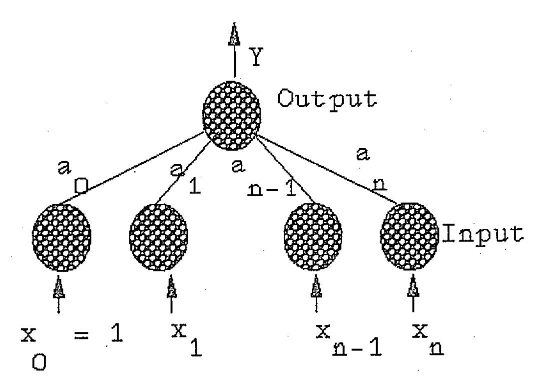


Figure 1 Linear neural network equivalent to a linear regression.

The model used to forecast the number of returns is a three layer feed-forward network consisting of an input layer that distributes the weighted input to the hidden layer, which then transforms that input and passes it to an output layer, which further transforms the vector and produces an output (forecast). In this example, the hidden layer contains three nodes (See Figure 2).

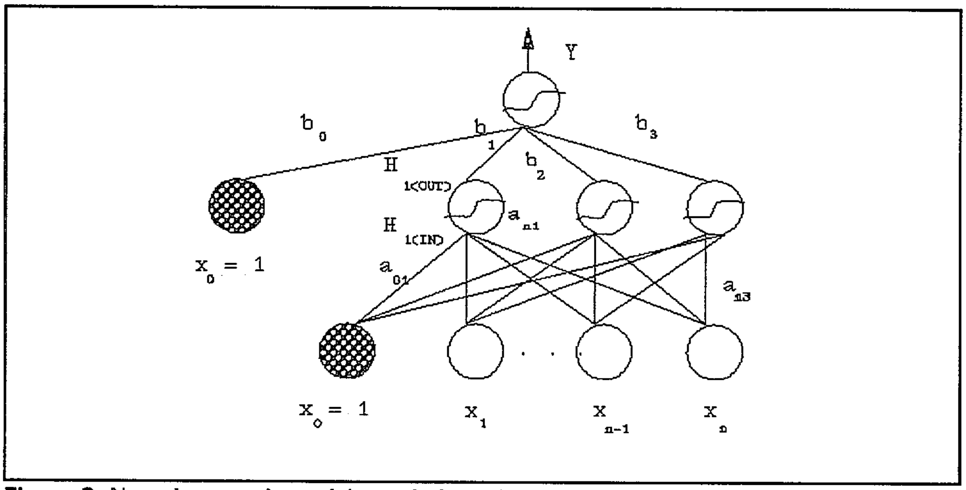


Figure 2. Neural network model consisting of multiple linear regressions with nonlinear sigmoidal transformations.

Each node acts as a regression equation:

$$H_{I(IN)}=a_{01}+a_{11}x_1+\ldots+a_{n1}x_n$$

In turn, each hidden node transforms this input using a sigmoidal activation function:

$$H_{1\,(OUT)}=\frac{1}{1+e^{-H_{1\,(IN)}}}$$

The output of each hidden node is multiplied by the weight of its connection to the output node. The results of these multiplications are summed to provide the input to the output layer node.

$$Y_{IN}=b_0+b_1H_{1\,(OUT)}+b_2H_{2\,(OUT)}+b_3H_{3\,(OUT)}$$

The predicted value Y is obtained by a sigmoidal transformation of this input.

$$Y_{1\,(OUT)}=\frac{1}{1+e^{-Y_{IN}}}$$

The connection weights are adjusted interactively through the successive presentation of training data (the observed vectors).

The power of the neural network is due to the use of multiple linear regression followed by two levels of nonlinear transforms. This process allows better approximation of these parameters ($a_{01},\ldots,b_3$) and in turn a better prediction of the output *Y*.

III. Methodology

For purposes of this analysis, the individual Form 1040 has been selected for forecasting. Three separate methodologies currently used by the IRS--regression, stepwise autoregression, and exponential smoothing--will be compared with a neural network to forecast Form 1040 return filing at 54 district offices. Post-sample forecast errors for one, two, and three periods will be tabulated.

Statistical Models

Three statistical models were used to estimate and forecast Form 1040 filing. First, ordinary and generalized least squares were used in a regression of filing volume in levels against civilian employment and a linear time trend. Employment variation is perhaps the strongest theoretical covariate for individual return filing; a time trend was employed as a proxy for unobservable influences. Although major tax legislation in 1981 and 1986, as well as a recession in 1992, contribute to filing variation, no intervention (dummy variable) terms were found to be statistically significant. Second, exponential smoothing using a grid search algorithm was used. Estimation and forecasting were performed on stationary filing. Three methods of detrending were used: simple growth, first differences, and residuals from a linear time trend regression. Mean square forecast errors were smallest by using the last method of detrending. Finally, stepwise autoregression was used. Of the three methods of detrending mentioned above, first differences produced the smallest mean square forecast errors.

Neural Network Models

Two neural network models were used to forecast form 1040 data: cascade correlation [Fahlman 90] and classical backpropagation [Rumelhart 86]. These models work best on data normalized either in the range [-1,1] or [0,1]. The technique which worked best for normalizing the information for our neural network was to take the natural log of the input and then normalize in the range [0.2,0.8] so as to use the "heart" of the sigmoidal activation function, thus ameliorating the problems implicit in values that lie at the edges of the function, near 0 and 1 [Fahlman 88]. We used a logarithmic scale because it compacts the large data values more than the smaller values. This is especially effective when only a relatively small segment of the data lies out on the "right" portion of the graph (see Figure 3). Our tax return regional data fits this example; most of the districts are small or medium sized. The data for forms 1120 and 1040 were similarly skewed.

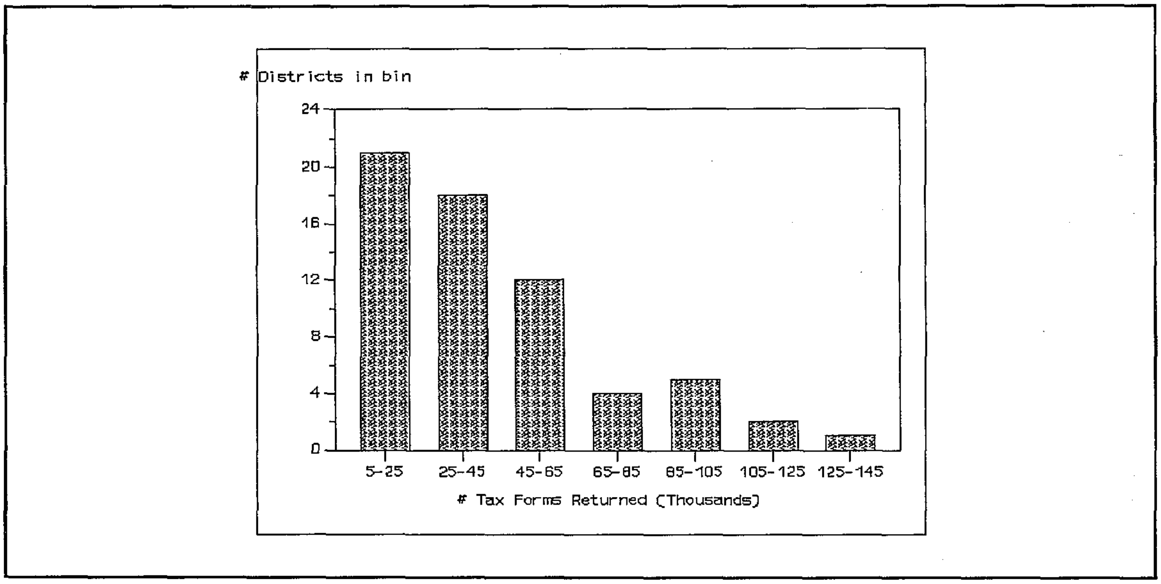


Figure 3: Distribution of targets before preprocessing.

Note how skewed the raw data is; a long tail exists on the high end of the distribution. Linearly normalizing would map most of the data into less than half of the input space. Worst of all, such a distribution causes the activation function to map a relatively small portion of the input space to outputs along the steeply linear portion of the sigmoid function's curve (see Figure 4).

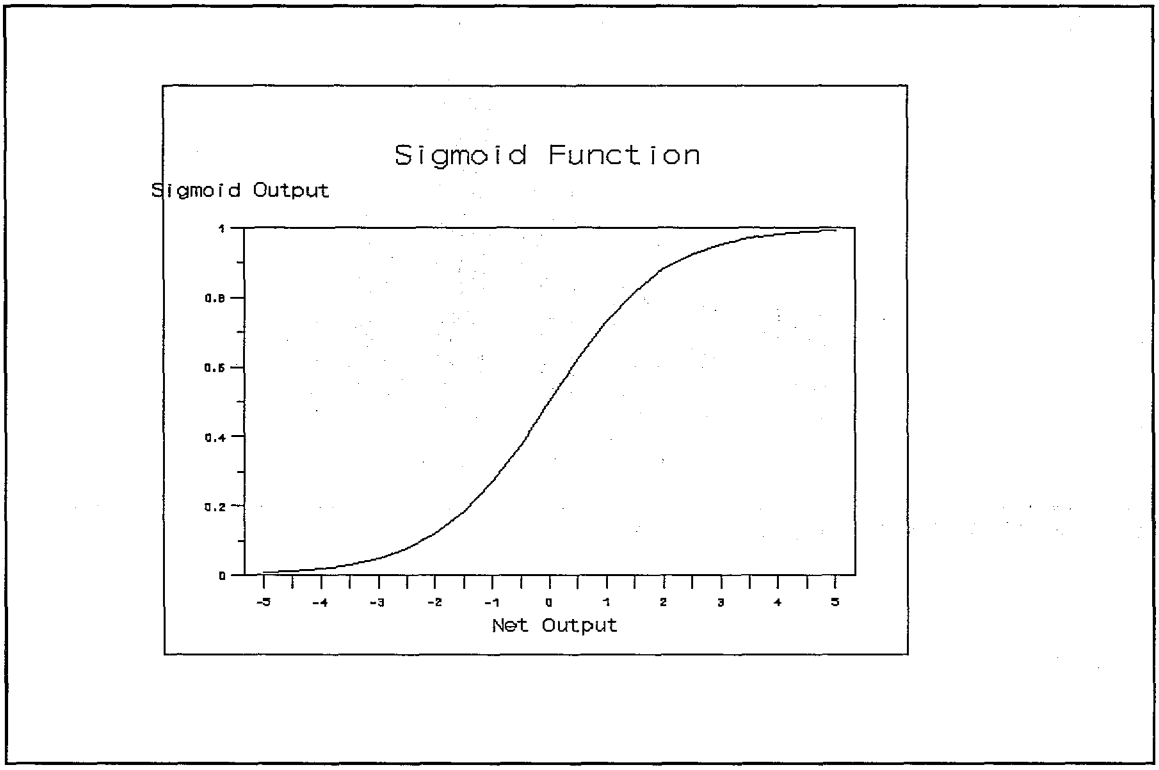


Figure 4: Standard sigmoid function; note "flat spots".

The logarithmic scale bring in these large districts, which are effectively outliers, and creates a distribution of data for the network which appears approximately normal, if somewhat skewed (see Figure 5). Most of the points now fall in the region of the sigmoid function which is approximately linear.

Our neural networks all use 11 inputs and one output. For form 1120, we use only classical backpropagation. The best results are produced by a network with one hidden layer of three nodes. Our inputs were 3 lag years of returns and 4 lag years each of employment and population on a districtwide basis. After training this network, we added another input node which represents the forecast of the standard IRS statistical method drawn from published projections [IRS 90]. The network underwent further training, using a very small learning rate (.001) on the connections from the original eleven inputs and a larger learning rate on the (.1) on the connections from the statistical forecast input. This produced improved results. For form 1040, we had the same inputs but one hidden layer of 8 nodes. In some cases, random noise was added to training set inputs as a method of focusing on low frequency trends; different random noise was added on each epoch. We also tried several cascade correlation networks with the same inputs.

For form 1120, we also used a competitive network [Rumelhart 85] to perform a simple form of cluster analysis on the districtwide data. 54 of the 63 districts clustered together, while the other nine districts clustered into a second class. We trained a neural network on just these 54 regions using the same architecture as described above, resulting in a lower maximum error rate as the districts that were most poorly forecasted in the 63 district network ended up in the nine district class.

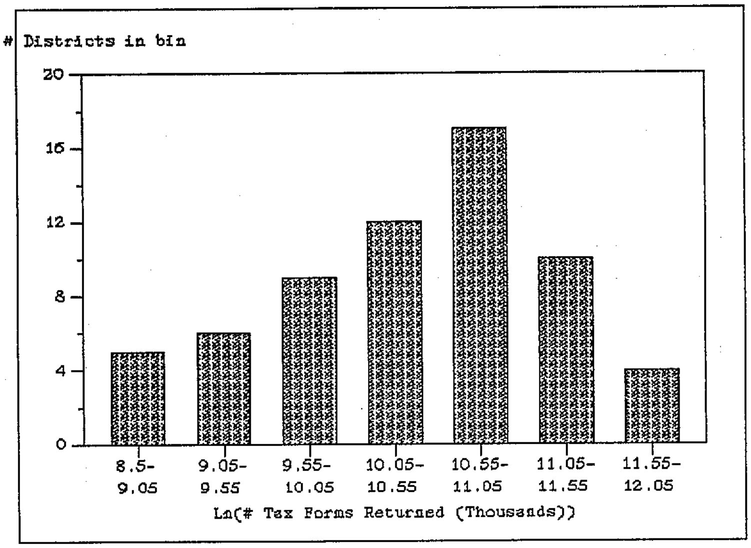


Figure 5: Distribution of targets after preprocessing.

IV. Empirical Results

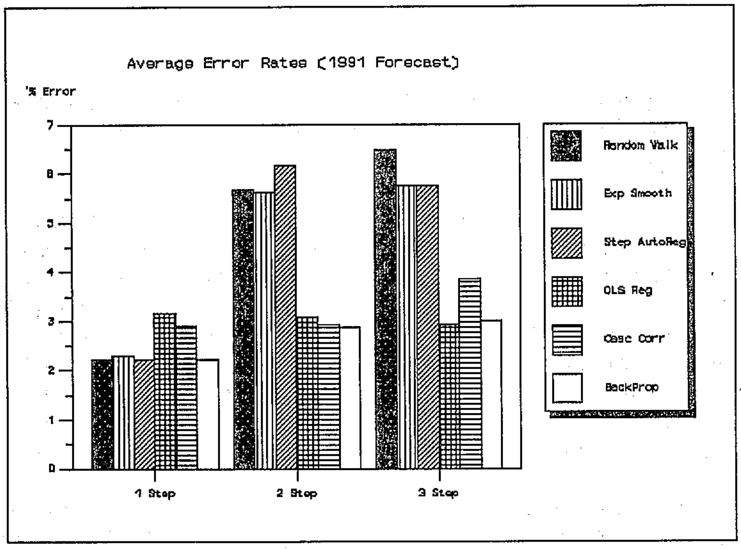


Figure 6: Average percent error per district.

The primary factor in determining the number of returns forecast by the neural network was simply how many returns came in the previous year. However, it was not only time-series trends that made the smaller adjustments to get the correct forecasts. Conditional variables for the trends in district employment and population were crucial to providing good projections. This was also the case for the statistical forecasts. In the longer range projections, which are the fundamental goal of this research, it was backpropagation, cascade-correlation and ordinary least squares regression that provided the best results (see Figure 6). While exponential smoothing and stepwise autoregression did better than OLS regression and cascade-correlation in the short term case, they were unable to beat classical backpropagation.

Backpropagation also did well in terms of maximum error rates (see). Both average and maximum percent error rates are useful in examining the effectiveness of a forecast; a large miscalculation in a single region would have deleterious cost effects, perhaps larger even than slightly worse results across the board.

On form 1120 (the corporate tax form), backpropagation was far superior to the other forecasting methods we tried. It is also interesting to note that adding a statistical input to the neural network gave the best results of all for this problem.

V. Conclusions

The goal of this applied research is to improve the forecasting of tax returns. The following approaches need to be considered:

> Improve the preprocessing of historical data and legislative data.
>
> Provide a unique model which couples an expert system, traditional statistical methods and neural networks. We expect an expert system could provide insight into the effect of legislative changes on historical trends.
>
> Investigate other neural network architectures such as multiple parallel networks.

We preprocessed our data for the neural networks by taking the natural log and then normalizing in the range [0,1]. This is discussed in detail above. Other potentially applicable methods exist. We took z-scores of the raw data, but initial results were disappointing. Ratios between the various inputs or fractional powers could also be useful. The goal of preprocessing is to make the data as linear as possible, which eases the job of the neural network in approximating a function.

Neural network architectures and paradigms must be investigated further to find the most appropriate solution to this forecasting problem. Adding more hidden layers or using multiple networks are possible approaches to investigate. Our objective is to find the network that most reliably provides robust forecasts.

Hybrid networks present another important angle to investigate. Coupling neural networks with other modeling methods could be done at various levels. We found, for example, that coupling statistical methods with neural networks improved forecasting of Form 1120. We envision an expert system coupled with a neural network addressing legislative changes.

What we did find is that neural networks are complementary to the standard statistical methods being used for forecasting today. Classical backpropagation was the only method in the group we compared that was consistently as good as or better than any other method, for both shorter and longer range forecasts. Clearly, neural networks have a role to play in improving tax return volume forecasts.

We recommend further work be done in this area to provide a method of forecasting that is accurate, robust and explainable. The potential for improved forecasting is quite promising and we have really only scratched the surface of neural networks' uses in this field.

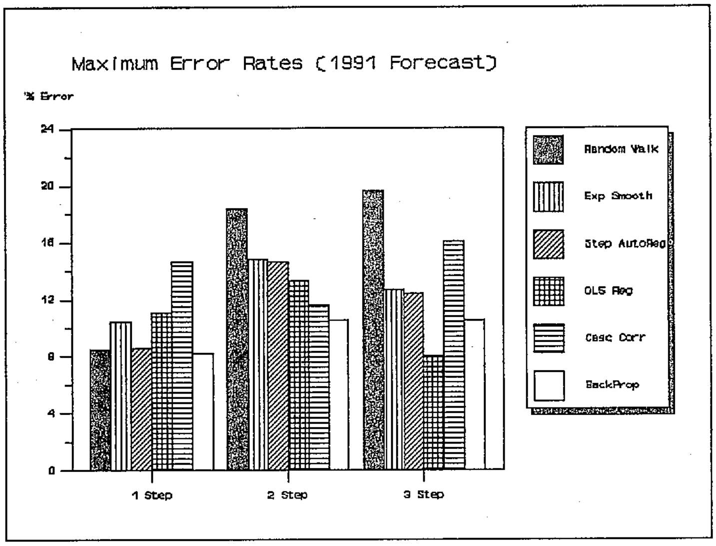


Figure 7: Maximum district error rates.

Bibliography

[Fahlman 88] Fahlman, Scott. *Faster learning variations of backpropagation: an empirical study.* In **Proceedings of the Connectionist Models Summer School.** San Mateo, CA: Morgan-Kaufmann, 1988, pp. 38-51.

[Fahlman 90] Fahlman, Scott and Lebiere, Christian. *The Cascade-Correlation Learning Architecture.* In **Advances in Neural Information Processing 2.** Ed. David Touretzky. Morgan-Kaufmann, 1990, pp. 524-32.

[Harston 90] Harston, Craig T. *Business with Neural Networks.* In **Handbook of Neural Computing Applications.** Eds. Alianna Maren, Craig Harston and Robert Pap. San Diego: Academic Press, 1990, pp. 391-400.

[Hecht-Nielsen 87] Hecht-Nielsen, Robert. *Kolmogorov's mapping neural network existence theorem.* In **Proceedings of the International Conference on Neural Networks.** Vol. III. New York: IEEE Press, 1987, pp. 11-3.

[IRS 90] Internal Revenue Service. **Projections -- Number of Returns to be Filed, Districts (Rev 11-90).** Washington, DC.

[Rumelhart 85] Rumelhart, David and Zipser, D. *Feature Discovery by Competitive Learning.* In **Cognitive Science, #9.** Ablex Publishing, 1985, pp. 77-112.

[Rumelhart 86] Rumelhart, David, Hinton, Geoffrey and Williams, R.J. *Learning Internal Representation by Error Backpropagation.* In **Parallel Distributed Processing.** Vol. I. Cambridge: MIT Press, 1986, pp. 318-62.

[Tang 91] Tang, Zaiyong, de Almeida, Chrys and Fishwick, Paul A. *Time series forecasting using neural networks vs. Box-Jenkins methodology.* In **Simulation.** 57:5, 1991, pp. 303-10.

Combining Short Run and Long Run Forecasts: Modeling the Demand for Electricity Using Seasonal Cointegration Techniques to Produce Consistent Forecasts

Fred Joutz, Department of Economics, The George Washington University

This research was partially supported by the U.S. Energy Information Administration (EIA). All statements are the authors own and do not reflect views of EIA. The results presented in the paper are preliminary and should not be quoted without the authors permission.

1. Introduction

Forecasters are asked to produce models for predicting the short run and long run. Conventional practice involves the construction of separate models for the two horizons. The short run model explains demand as a function of seasonal or rapidly changing variables. The long run model explains demand as a function of slowly changing variables like demographic characteristics and income. In general these two models result in conflicting forecasts at overlapping horizon(s). Forecasters and planners must employ an ad hoc means of reconciling the difference(s) to produce a unified forecast.

This paper incorporates the information from short run models and a long run model for aggregate U.S. electricity demand into a single model using the error correction framework. Recent developments in the cointegration and seasonal cointegration literature are exploited in the model's construction and estimation. The forecast performance (from the merged models) is compared against the separate models predictions. The evidence from one and two year ahead forecasts of national residential electricity sales is mixed.

2. The Forecasting Model

The objective in this section is to demonstrate a means for combining or merging a short run model and a long run model (of energy demand). I begin by discussing a cointegration model which can lead to an error correction mechanism (ECM). This is a model which encompasses the information in both the traditional short-run and long-run models. One interpretation of this is that the traditional short run and long run models are subsets of the merged or "true" model. The presentation follows Engle, Granger, and Hallman (1989). See Alogoskoufis and Smith (1991) for a discussion of specification, interpretation, and estimation of error correction models (ECM).

The variable of interest is y_t. The information set, Y_t, contains lagged values of y_t, other endogenous variables, exogenous variables and predetermined variables. Assume the variables are in natural logarithms.

If the series $(1-B)^d y_t$ is stationary it is said to integrated of order d, I(d). Here B serves as a backshift operator such that $B^i y_t = y_{t-i}$. When using monthly or quarterly seasonally unadjusted data it may be the case that the seasonal differencing and even multiplicative differencing, first and seasonal differences, is required to make the data stationary.

Suppose that d=1, then y_t is a random walk and may or may not have drift. A series integrated of order one, I(1), is smoother and slower changing than stationary I(0) series. The former as no affinity for the mean value, so that departures from the mean can be long.

Let w_t be a sub-set of Y_t and integrated of order d, the same as y_t. From the Granger representation theorem if there exists a stationary linear combination

$$z_t = y_t - \beta' w_t \tag{1}$$

then w_t is co-integrated with y_t. This implies the data generating process can be represented by an error correction model or mechanism, ECM, of the form

$$\Delta y_t = \mu - \pi * z_{t-1} + \gamma' x_t + \epsilon_t \tag{2}$$

where x_t is I(0) explanatory variables. Stationary lag polynomials of Δy_t and Δw_t may be included in x_t. The μ term can represent the intercept or "trend" growth and centered seasonal dummy variables. The random disturbance, ϵ_t is assumed to be white noise. This ECM model can be interpreted as the "true" or merged model.

The <u>long run (forecasting) model</u> is assumed to use the elements of Y_t which are I(1) and takes the form:

$$y_t = \beta_0 + \beta_1' w_t + \eta_t. \tag{3}$$

where the expected value(s) of β_1 are β in equation (1). These can be interpreted as the long run elasticity estimates. Again, the random disturbance, η_t, is assumed to be white noise.

The usual short run (forecasting) model does not incorporate the error correction mechanism; it omits information from the long run model.

$$\Delta y_t = \gamma_0 + \gamma_1' x_t + \epsilon_t; \quad where\ \Delta y_t\ and\ x_t \sim I(0) \tag{4}$$

The γ_1' represent the short run elasticity estimates. Notice that the short run elasticities in this expression can differ from those in (2). The ϵ_t could be white noise or follow a autoregressive process.

Thus a forecaster has three potential forecasting models. The encompassing one as represented in equation 2, a long run model as in 3, or the short run model in equation 4. The first one is the merged model which makes efficient use of available information.

When monthly data is available for y, w, and x, the one step ahead forecast for the ECM or merged model is

$$y_{t+1} = f^y_{t,1} = (1-\pi) y_t + \pi \beta' w_t + \gamma' x_{t+1} + \epsilon_{t+1} \tag{5}$$

where x are the forecasted value(s) of the explanatory variables. Here $f^y_{t,1}$ represents a forecast of y based on available information at t out one period. Longer horizon forecasts are constructed from iterating the expression above out the desired number of periods. As the forecasting horizon increases the x variables approach their monthly expected values. This results in a deterministic component to the forecasts, μ^*. We can express the long run or h step ahead forecasts two different ways:

$$\begin{gathered} f^y_{t,h} \approx \mu^* + (1-\pi) f^y_{t,h-1} + \pi \beta' f^w_{t,h-1} \\ or \\ f^y_{t,h+1} - f^y_{t,h} \approx \mu^* - \pi (f^y_{t,h} - \beta' f^w_{t,h}) \end{gathered} \tag{6}$$

In the long run the forecasted change in the variable of interest is equal to the deterministic component(s) minus the difference between the predicted variable the previous period and the estimate from the long run predicted value using forecasts of the w variables. As the left hand side in the second expression approaches a constant, then the right hand side becomes

$$f^y_{t,h} = constant + \beta' f^w_{t,h} . \tag{7}$$

This approximates the long run model from equation (3). Furthermore, by implementing the ECM model short run forecasts are produced similar to those using the short run model, equation (4). These forecasts could even be improved, because of the inclusion of the information about the cointegrating relationship. Thus it appears as though the ECM framework provides a consistent bridge between the long run and short run forecasts.

3. The Data

The sample period for the study is January, 1978 through December, 1991; all series are not seasonally adjusted. The U.S. residential electricity consumption and price data used in this study comes from the Historical Monthly Energy Review published by the U.S. Energy Information Administration (EIA). Heating and cooling day series are population weighted by states; this series was provided by Mr. David Costello of the Integrated Analysis and Forecasting Office of the EIA. CITIBASE was the source for the urban consumer price index 1982 = 100, unemployment rate, and the household estimates series. A real price of electricity per KWH is derived from the nominal price deflated by the consumer price index. This series will be used in the models and forecasts.

Figure 1 plots residential electricity sales in billions of KWH per month. Sales grow fairly steady from an average of 55 billion KWH per month to approximately 80 billion KWH per month. There are seasonal fluctuation(s) representing the heating needs in the winter months and the cooling need in the summer. The winter "spike" is larger than the summer one. Heating and cooling degree days are plotted in Figure 2.

Nominal residential electricity prices in cents per KWH and monthly growth rates are graphed in Figures 3 and 4. Again, there is a seasonal pattern, but only one a single "spike" during winter. The price rose from 4 cents in 1978 to 8 cents 1985, about 1.3 percent a month. Thereafter, the price stabilized fluctuating about one cent per KWH over the seasonal cycle.

In real terms residential electricity prices rose until about 1984 as price increases outpaced the inflation rate. The increase was from 6.3 cents per KWH to 7.2 cents per KWH. However the price declined to just below 6 cents by the early 1990s. See Figures 5 and 6.

The civilian unemployment rate was chosen as a proxy for economic activity. There is no monthly personal income series available on a seasonally unadjusted basis, because it is calculated as a residual. Figure 7 show that unemployment is rising through the 1982-83 recession and falls during the expansion of the mid to late 1980s before rising again in the downturn and recession of 1990 and 1991.
Figure 8 plots the estimates for number of households in thousands by the Bureau of Census. Household formation is generally used as an indicator of energy demand. It can proxy for the stock of energy using durable capital stocks. While the series is seasonally unadjusted there is little variation over the year.

4. Model Results
In this section I present results from the monthly data series used in producing the error-correction model. The first part examines the time series properties of the series. Second, the different models are presented. All empirical work was performed using the extended memory version of SHAZAM (White, 1990).

A common transformation in time series analysis of economic data is first differencing (and or seasonal differencing). The implication being that the variable(s) in level form are not stationary. First differencing assumes there is a unit root (coefficient of one) at the first lag. An alternative to the first difference model would include a constant and possibly a trend term; this implies the variable(s) follows a trend stationary process. Kang and Nelson (1984) discuss the characteristics and problems of misspecification of trend stationary processes and difference stationary processes. The data plots suggest that the series might not be stationary.

The stationarity problem becomes more complex with data subject to seasonal variation. Engle, Granger, and Hallman (1989) and Ilmakunnas (1990) caution the unwary modeler using seasonally unadjusted data against conducting only unit root tests at lag one. There could be unit roots at first order and at seasonal lags, the multiplicative seasonal difference model. One alternative is the seasonal dummy variable model with trend; this is just a deterministic seasonal model with trend. In addition there could be a unit root at lag one and fixed seasonal dummy variables, the first difference seasonal dummy variable model.

Following Franses (1991) I test for the presence of unit roots against the alternative models. The following auxiliary regression is estimated. Tests are performed on the significance of the parameters for the null hypothesis of unit roots in the monthly time series.

$$\begin{aligned}\Psi(B)\,y_{8,t} = {} & \beta_1 y_{1,t-1} + \beta_2 y_{2,t-1} + \beta_3 y_{3,t-1} + \beta_4 y_{3,t-2} + \beta_5 y_{4,t-1} \\ & + \beta_6 y_{4,t-2} + \beta_7 y_{5,t-1} + \beta_8 y_{5,t-2} + \beta_9 y_{6,t-1} + \beta_{10} y_{6,t-2} \\ & + \beta_{11} y_{7,t-1} + \beta_{12} y_{7,t-2} + \mu + \delta\ trend + \epsilon_t\end{aligned}$$

The y variables are constructed as follows

$$\begin{aligned}
y_{1,t} &= (1+B)(1+B^2)(1+B^4+B^8)\,y_t \\
y_{2,t} &= -(1-B)(1+B^2)(1+B^4+B^8)\,y_t \\
y_{3,t} &= -(1-B^2)(1+B^4+B^8)\,y_t \\
y_{4,t} &= -(1-B^4)(1-\sqrt{3}B+B^2)(1+B^2+B^4)\,y_t \\
y_{5,t} &= -(1-B^4)(1+\sqrt{3}B+B^2)(1+B^2+B^4)\,y_t \\
y_{6,t} &= -(1-B^4)(1-B^2+B^4)(1-B+B^2)\,y_t \\
y_{7,t} &= -(1-B^4)(1-B^2+B^4)(1+B+B^2)\,y_t \\
y_{8,t} &= (1-B^{12})\,y_t
\end{aligned}$$

where B is the backshift operator ie. $(1-B^i)y_t = y_t - y_{t-i}$. Lagged dependent variables enter through the $\Psi(B)$ lagged polynomial term. The number of lagged dependent variables is chosen by minimizing the Aikaike Information Criterion (AIC) and the Bayesian Schwarz Criterion following Granger and Newbold (1986). The μ term can include possible deterministic components like a constant, seasonal dummy variables, or a trend. The deterministic variables depend upon and condition the alternative hypothesis being considered.

Four variables are examined in the tests for a first difference deterministic seasonal model versus the multiplicative seasonal difference model. They are the residential electricity sales (KWH) and real price series (RPKWH), the number of households (POH), and the civilian unemployment rate (UE).

Estimation is performed by ordinary least squares including seasonal (monthly) dummy variables with and without a trend. If $\beta_1 = 0$, then the null of a unit root cannot be rejected. If $\beta_2, \ldots, \beta_{12}$, are significantly different from zero, there are no seasonal unit roots. When $\beta_1 = 0$ and $\beta_2, \ldots, \beta_{12}$ are significantly different from zero, a first difference seasonal dummy variable model is appropriate. However, if β_i i=1,...12 are all zero, then a multiplicative seasonal difference model would be the model of choice. The critical values for individual t-tests and F-tests with a sample size of 120 are found in Franses (1990). The sample here is from 1978.01-1991.12 yielding an effective size of 166 observations. The critical value for β_1 at 10% in a model including a constant term, trend and seasonal dummy variables is -2.92 without the trend it is -2.35. The respective critical values at 5% for the F-test are 4.45 and 4.46.

The evidence in Table 1 suggests that the first difference seasonal dummy model is the most appropriate for the electricity sales data (KWH), the number of households(POH), and the unemployment rate(UE). A trend term and one lag of the dependent variable were used in each of the regressions. Real electricity prices appear to be stationary, but with deterministic components. The coefficients for the seasonal roots and at lag one all exceed the critical values in a regression with no trend and a third order autoregressive process for the dependent variable. Although, not reported here the tests for stationarity of the nominal price and the consumer price index suggested that the two components could be integrated of order one. The results for the real price appear to imply that the series are cointegrated of order zero.

Seven different models were estimated for the forecast comparison. One long run model was estimated to generate the error correction term. Monthly Electricity Sales are regressed on the Total Number of Households and the Civilian Unemployment Rate. The regression results are provided in Table 2. The two explanatory variables have positive and negative effects on residential electricity sales respectively. The estimated residuals, E1, from this regression are used in the Error Correction Models (ECM). The ECM does not have a unit root given the value of the Durbin Watson Statistic.

Three short run models are estimated. The first two are in levels and the third is in first differences. Short Run Model 1 specifies the level of sales (ESRCPUS) as a function of lagged sales for the last two months, the current and lagged values of real prices (RPRICE), heating degree days (ZWHDPUS) and cooling degree days (ZWCDPUS), plus eleven seasonal dummy variables and a constant. The seasonal dummies are centered on December. The same variables except the three month average real price (ARPRICE) is substituted for the current and lagged real prices are used in short run model 2. Sales are modeled in first differences in short run model 3. The explanatory variables include lagged dependent variables, the current first difference of real prices, and current and lagged levels for heating degree days and cooling degree days. A constant and seasonal dummy variables are included too. The regression results for short run model 1 are found in Table 3, short run model 2 in Table 4, and short run model 3 in Table 5.

In the first two short run models the two lagged dependent variables are significant and sum to about 0.9, a common property for a series with a unit root. The two price variables are significant in short run model 1, but appear to cancel each other out. In short run model 2 the average price term is negative, however insignificant. The weather related variables all have positive and significant effects on sales. In the model with first differences the two lagged dependent variables have negative and significant effects. The first difference of real prices is negative and significant. The heating and cooling degree day variables continue to have positive explanatory power.

Three error correction models (ECM) are estimated. The results are presented in Tables 6, 7, and 8. They correspond to the three short run models, only augmented by the lagged error correction mechanism variable (E1) derived from the long run model.

In each ECM model E1 contributes to the explanatory power of the model. It has a negative impact as theory would suggest. If actual sales are above their equilibrium values one period, they should move in the opposite direction in the next period. The estimated coefficients for the remaining variables are stable and their significance increases somewhat. This supports the theoretical assertion that the short run model estimates are inefficient, because they omit important long run information.

5. Forecast Comparisons

The forecasting ability of the different models are compared based on their predictive power over the one and two year horizon. Six one year ahead forecasts are made and five two year ahead forecasts are made. The first set of forecasts are made based on data through December, 1985 and predict out 24 months. The forecasts for the first twelve months are aggregated to make the one year ahead annual electricity sales forecast. The second twelve months are aggregated to make the two year ahead forecast. Then the data set is updated through December, 1986 and predictions for the next 24 months are generated. This process is repeated through December, 1990. At that time only a one year ahead forecast is made, because of the sample size. The forecasts and forecast errors are presented in Table 9.

The top half of the table presents the forecasts at the one and two year horizons. The bottom half gives the forecast errors and three summary statistics. They are the average error or bias, the mean absolute error (MAE), and the mean absolute percentage error (MAPE). At the one year horizon all three ECM models have smaller average errors, MAEs, and MAPEs. The results are just the opposite at the two year horizon. It is difficult to attach much weight to these results, because of the small sample size.

In all cases except one (ECM 3 at the two year horizon) the ECM models have smaller average errors and MAPE than the long run model. This could be explained by the definition of the long run model or equilibrium relationships between residential electricity sales, households, and unemployment.

In Section 2 it was shown that the ECM would produce models and forecasts which were consistent in the short run and the long run. The forecasts from the ECM or merged models are closer to the long run model forecast at the two year horizon than at the one year horizon relative to the short run model forecasts. At the one year horizon the ECM models are closer to the long run models in only 1 case compared to the short run models. However at the two year horizon the ECM models produce forecasts which are closer in three of the five cases. This result seems consistent with the notion of the ECM cointegrating model's purpose, but the one year ahead results are not intuitive. Again, the sample size for the analysis is too small to make any strong conclusion(s).

6. Conclusion

Short run and long run econometric forecasting models can be merged using error correction mechanism (ECM) models based on co-integration theory. This allows the forecaster to use a single model and possibly improve the forecast. It overcomes the practical difficulty of having conflicting forecasts at some horizon(s). For example a monthly model can depend on rapidly changing variables due to seasonal variability. A long run model using quarterly or even annual data depends on slowly changing variables like demographic characteristics and income.

This paper presents the results from a preliminary study using error correction models to predict residential electricity sales. Merging long run information through the ECM in a short run forecasting model improves the fit and appears theoretically consistent. However, a comparison of one and two year ahead sales forecasts using monthly data produces mixed results.

The empirical findings are initial results of a research agenda into cointegration and forecasting. Three extensions are envisioned. First, annual data on electricity sales and its long run determinants are available back to the 1940's. In this paper the "long run" model was constructed based on monthly data for 13 years. The frequency of the data and the relatively short historical span may not accurately portray the long run relationships. A different long run model based on the annual data can be constructed for the ECM model(s) and the results compared. Second, this paper employs the cointegration and ECM techniques in a single equation case. Johansen (1988) and Johansen and Juselius (1990) have developed techniques for multivariate cointegration modeling. Their approach could be used to forecast as well. Third, other data sets and variables of interest can be considered. These include industrial electricity sales, home heating oil consumption, and motor gasoline to name a few.

BIBLIOGRAPHY

Alogoskoufis, George, and Ron Smith, (1991) "On Error Correction Models: Specification, Interpretation, and Estimation," Journal of Economic Surveys, vol. 5 no. 1, p97-128.

CITIBASE: Citibank economic database (machine readable magnetic data file). 1946 - present, New York, Citibank, N.A., 1978.

Engle, R.F., C.W.J. Granger and J.J. Hallman, (1989) "Merging Short- and Long-Run Forecasts: An Application of Seasonal Cointegration to Monthly Electricity Sales Forecasting," Journal of Econometrics, 40, p.45-62.

Frances, Philip Hans, (1991) "Seasonality, Non-Stationarity and the Forecasting of Monthly Time Series," International Journal of Forecasting, vol. 7, p199-208.

Granger, C.W.J. and Paul Newbold, (1986) Forecasting Economic Time Series, Second Edition, Academic Press, Inc. San Diego.

Hylleberg, S. R.F. Engle, C.W.J. Granger, and B.S. Yoo, (1990) "Seasonal Integration and Cointegration," Journal of Econometrics, 44, p.215-238.

Ilmakunnas, Pekka, (1990) "Testing the Order of Differencing in Quarterly Data: An Illustration of the Testing Sequence," Oxford Bulletin of Economics and Statistics, vol. 52, no. 2, p.79-88.

Johansen, Soren, (1988) "Statistical Analysis of Cointegration Vectors,"Journal of Economic Dynamics and Control, vol. 12, p.231-254.

and Katarina Juselius, (1990) "Maximum Likelihood Estimation and Inference on Cointegration - With Applications to the Demand for Money," Oxford Bulletin of Economics and Statistics, vol. 52, no. 2, p.169-210.

Kang, Heejoon and Charles R. Nelson, (1984) "Pitfalls in the Use of Time as an Explanatory Variable in Regression," Journal of Business and Economic Statistics, vol. 2, no. 1, p.73-82.

Watson, Mark W., and L. Pastuszek, and E. Cody, (1987) "Forecasting Commercial Electricity Sales," Journal of Forecasting, vol. 6, p.117-136.

White, K.J., Wong, S.D., Whistler, D. and Haun, S.A., SHAZAM User's Reference Manual Version 6.2, McGraw-Hill, 1990.

Table 1
Testing for (Seasonal) Unit Roots vs.
Deterministic Seasonals for Residential Electricity Sales
using the sample period 1978.01-1992.03

Hypothesis Test	KWH	REAL PRICE/KWH	HOUSEHOLDS	UNEMPLOYMENT RATE
Models with Constant and no Trend				
$\beta_1=0$	-0.04	-1.72	-2.46	-1.99
$\beta_1=\mu=0$	6.65	10.73	3.04	2.43
$\beta_2=..=\beta_{12}=0$	11.16	13.00	47.60	21.21
$\beta_1=..=\beta_{12}=0$	10.57	12.74	99.31	21.48
Models with Constant and Trend				
$\beta_1=0$	-3.13	-3.07	-0.32	-2.01
$\beta_1=\delta=\mu=0$	8.06	10.49	2.02	4.03
$\beta_1=\delta=0$	4.98	5.91	3.00	2.08
$\beta_2=..=\beta_{12}=0$	12.28	12.79	45.39	20.92
$\beta_1=..\beta_{12}=0$	12.09	13.46	66.72	21.36

$$\psi(B)y_{8,t} = \beta_1 y_{1,t-1} + \beta_2 y_{2,t-1} + \beta_3 y_{3,t-1} + \beta_4 y_{3,t-2} + \beta_5 y_{4,t-1} + \beta_6 y_{4,t-2} + \beta_7 y_{5,t-1} + \beta_8 y_{5,t-2} + \beta_9 y_{6,t-1} + \beta_{10} y_{6,t-2} + \beta_{11} y_{7,t-1} + \beta_{12} y_{7,t-2} + \mu + \delta t + \epsilon_t$$

The order of the autoregressive component, $\psi(B)$, was found by the minimum of the AIC and the Bayesian Schwarz criterion.

Table 2
LONG RUN MODEL USING MONTHLY DATA --- KWH=f(HOUSEHOLDS(POH),UE)
SMPL 1978.01 1990.12

```
R-SQUARE =    0.3486      R-SQUARE ADJUSTED =    0.3401
VARIANCE OF THE ESTIMATE-SIGMA**2 =    83.659
STANDARD ERROR OF THE ESTIMATE-SIGMA =    9.1465
SUM OF SQUARED ERRORS-SSE=   12800.
MEAN OF DEPENDENT VARIABLE =   65.663
LOG OF THE LIKELIHOOD FUNCTION = -565.126

VARIABLE   ESTIMATED  STANDARD    T-RATIO
  NAME    COEFFICIENT   ERROR     153 DF
POH       0.10730E-02 0.1356E-03   7.914
UE       -0.88990     0.5362      -1.660
CONSTANT  -20.362      13.32      -1.528

DURBIN-WATSON = 1.0586     VON NEUMANN RATIO = 1.0654     RHO =  0.46625
COEFFICIENT OF SKEWNESS =    0.2393 WITH STANDARD DEVIATION OF 0.1943
COEFFICIENT OF EXCESS KURTOSIS =   -1.0117 WITH STANDARD DEVIATION OF 0.3862
```

Table 3
SHORT RUN MODEL 1 OF MONTHLY SALES (LEVELS)
SMPL 1978.01 1990.12

```
 R-SQUARE =     0.9510      R-SQUARE ADJUSTED =    0.9441
VARIANCE OF THE ESTIMATE-SIGMA**2 =    7.1807
STANDARD ERROR OF THE ESTIMATE-SIGMA =    2.6797
SUM OF SQUARED ERRORS-SSE=    962.21
MEAN OF DEPENDENT VARIABLE =    65.669
LOG OF THE LIKELIHOOD FUNCTION = -359.602
```

VARIABLE NAME	ESTIMATED COEFFICIENT	STANDARD ERROR	T-RATIO 134 DF
ESRCPUS	0.79693	0.8352E-01	9.542
ESRCPUS	0.12256	0.8339E-01	1.470
RPRICE	-12.929	3.125	-4.137
RPRICE	12.436	3.141	3.959
ZWHDPUS	0.19960E-01	0.3898E-02	5.120
ZWHDPUS	0.75184E-02	0.4284E-02	1.755
ZWCDPUS	0.71521E-01	0.1373E-01	5.208
ZWCDPUS	0.60464E-02	0.1478E-01	0.4091
M1	-2.2555	1.784	-1.264
M2	-14.882	2.312	-6.436
M3	-11.677	2.576	-4.533
M4	-7.4514	2.707	-2.753
M5	-2.1899	3.238	-0.6762
M6	4.2237	4.230	0.9985
M7	3.3884	5.429	0.6242
M8	-4.2251	5.820	-0.7260
M9	-7.6814	4.970	-1.545
M10	-11.243	3.310	-3.397
M11	-7.3279	1.972	-3.716
CONSTANT	-4.6103	6.999	-0.6587

```
DURBIN-WATSON = 1.8601     VON NEUMANN RATIO = 1.8722     RHO =   0.06759
...DURBIN H STATISTIC CANNOT BE COMPUTED
COEFFICIENT OF SKEWNESS =  -0.3004 WITH STANDARD DEVIATION OF 0.1955
COEFFICIENT OF EXCESS KURTOSIS =   0.3405 WITH STANDARD DEVIATION OF 0.3886

HETEROSKEDASTICITY TESTS
  E**2 ON YHAT:            CHI-SQUARE =       9.575 WITH 1 D.F.
  E**2 ON YHAT**2:         CHI-SQUARE =      10.219 WITH 1 D.F.
  E**2 ON LOG(YHAT**2):    CHI-SQUARE =       8.747 WITH 1 D.F.
  E**2 ON X (B-P-G) TEST:          CHI-SQUARE =     36.428 WITH  19 D.F.
  E**2 ON LAG(E**2)  ARCH TEST:    CHI-SQUARE =      0.208 WITH   1 D.F.
  LOG(E**2) ON X (HARVEY) TEST:    CHI-SQUARE =     20.582 WITH  19 D.F.
  ABS(E) ON X (GLEJSER) TEST:      CHI-SQUARE =     31.519 WITH  19 D.F.

RAMSEY RESET SPECIFICATION TESTS USING POWERS OF YHAT
 RESET(2)=  0.74743      - F WITH DF1=   1 AND DF2= 133
 RESET(3)=  0.37196      - F WITH DF1=   2 AND DF2= 132
 RESET(4)=  0.25040      - F WITH DF1=   3 AND DF2= 131

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	0.0672	0.0806	0.8343	2.1786	1.8601	0.7097
2	-0.0622	0.0806	-0.7725	0.8668	2.1087	1.3220
3	-0.1453	0.0806	-1.8030	1.9282	2.2650	4.6806

Table 4
SHORT RUN MODEL 2 OF MONTHLY SALES - AVERAGE PRICE (LEVELS)
SMPL 1978.03 1990.12

R-SQUARE = 0.9453 R-SQUARE ADJUSTED = 0.9379
VARIANCE OF THE ESTIMATE-SIGMA**2 = 7.9192
STANDARD ERROR OF THE ESTIMATE-SIGMA = 2.8141
SUM OF SQUARED ERRORS-SSE= 1053.2
MEAN OF DEPENDENT VARIABLE = 65.834
LOG OF THE LIKELIHOOD FUNCTION = -362.796

VARIABLE NAME	ESTIMATED COEFFICIENT	STANDARD ERROR	T-RATIO 133 DF
ESRCPUS	0.70738	0.8462E-01	8.359
ESRCPUS	0.21557	0.8429E-01	2.558
ARPRICE	-0.57992	0.6256	-0.9270
ZWHDPUS	0.21095E-01	0.4105E-02	5.139
ZWHDPUS	0.13156E-01	0.4478E-02	2.938
ZWCDPUS	0.67337E-01	0.1439E-01	4.681
ZWCDPUS	0.18045E-01	0.1523E-01	1.185
M1	-3.6057	1.905	-1.892
M2	-19.784	2.185	-9.054
M3	-18.141	2.149	-8.441
M4	-12.446	2.489	-4.999
M5	-6.1402	3.249	-1.890
M6	2.0333	4.440	0.4579
M7	3.5957	5.754	0.6249
M8	-4.2191	6.157	-0.6852
M9	-8.2636	5.268	-1.569
M10	-11.797	3.535	-3.338
M11	-7.0516	2.117	-3.331
CONSTANT	-5.5139	7.383	-0.7468

DURBIN-WATSON = 1.8410 VON NEUMANN RATIO = 1.8532 RHO = 0.07802
...DURBIN H STATISTIC CANNOT BE COMPUTED
COEFFICIENT OF SKEWNESS = -0.1882 WITH STANDARD DEVIATION OF 0.1968
COEFFICIENT OF EXCESS KURTOSIS = 0.0295 WITH STANDARD DEVIATION OF 0.3911

HETEROSKEDASTICITY TESTS
E**2 ON YHAT: CHI-SQUARE = 5.966 WITH 1 D.F.
E**2 ON YHAT**2: CHI-SQUARE = 6.146 WITH 1 D.F.
E**2 ON LOG(YHAT**2): CHI-SQUARE = 5.682 WITH 1 D.F.
E**2 ON X (B-P-G) TEST: CHI-SQUARE = 30.709 WITH 18 D.F.
E**2 ON LAG(E**2) ARCH TEST: CHI-SQUARE = 1.285 WITH 1 D.F.
LOG(E**2) ON X (HARVEY) TEST: CHI-SQUARE = 25.013 WITH 18 D.F.
ABS(E) ON X (GLEJSER) TEST: CHI-SQUARE = 23.645 WITH 18 D.F.

RAMSEY RESET SPECIFICATION TESTS USING POWERS OF YHAT
RESET(2)= 1.4100 - F WITH DF1= 1 AND DF2= 132
RESET(3)= 0.70901 - F WITH DF1= 2 AND DF2= 131
RESET(4)= 0.53694 - F WITH DF1= 3 AND DF2= 130

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	0.0779	0.0811	0.9600	3.2273	1.8410	0.9398
2	-0.0280	0.0811	-0.3454	0.3864	2.0497	1.0623
3	-0.0557	0.0811	-0.6871	0.7278	2.0943	1.5503

Table 5
SHORT RUN MODEL 3 OF MONTHLY SALES (FIRST DIFFERENCES)
SMPL 1978.01 1991.12

R-SQUARE = 0.9362 R-SQUARE ADJUSTED = 0.9283
VARIANCE OF THE ESTIMATE-SIGMA**2 = 6.5068
STANDARD ERROR OF THE ESTIMATE-SIGMA = 2.5508
SUM OF SQUARED ERRORS-SSE= 949.99
MEAN OF DEPENDENT VARIABLE = 0.14071
LOG OF THE LIKELIHOOD FUNCTION = -378.541

VARIABLE NAME	ESTIMATED COEFFICIENT	STANDARD ERROR	T-RATIO 146 DF
DKWH	-0.21742	0.7617E-01	-2.854
DKWH	-0.29202	0.6381E-01	-4.576
DPRI	-11.378	2.963	-3.841
ZWHDPUS	0.20271E-01	0.3593E-02	5.642
ZWHDPUS	0.77360E-02	0.4067E-02	1.902
ZWCDPUS	0.70113E-01	0.1264E-01	5.546
ZWCDPUS	0.20543E-01	0.1412E-01	1.455
M1	1.5737	1.858	0.8472
M2	-8.7991	2.595	-3.391
M3	-6.7576	2.572	-2.627
M4	-7.6785	2.482	-3.094
M5	-1.5306	2.983	-0.5131
M6	4.5289	3.836	1.181
M7	3.8948	4.860	0.8015
M8	-2.9328	5.250	-0.5586
M9	-5.3684	4.596	-1.168
M10	-10.766	3.097	-3.476
M11	-7.4624	1.848	-4.038
CONSTANT	-16.289	3.742	-4.353

DURBIN-WATSON = 1.9675 VON NEUMANN RATIO = 1.9795 RHO = 0.01568
DURBIN H STATISTIC (ASYMPTOTIC NORMAL) = 0.97528
COEFFICIENT OF SKEWNESS = -0.3512 WITH STANDARD DEVIATION OF 0.1890
COEFFICIENT OF EXCESS KURTOSIS = 0.4087 WITH STANDARD DEVIATION OF 0.3758

HETEROSKEDASTICITY TESTS
E**2 ON YHAT: CHI-SQUARE = 1.891 WITH 1 D.F.
E**2 ON YHAT**2: CHI-SQUARE = 9.052 WITH 1 D.F.
E**2 ON LOG(YHAT**2): CHI-SQUARE = 6.121 WITH 1 D.F.
E**2 ON X (B-P-G) TEST: CHI-SQUARE = 32.738 WITH 18 D.F.
E**2 ON LAG(E**2) ARCH TEST: CHI-SQUARE = 0.003 WITH 1 D.F.
LOG(E**2) ON X (HARVEY) TEST: CHI-SQUARE = 21.960 WITH 18 D.F.
ABS(E) ON X (GLEJSER) TEST: CHI-SQUARE = 31.386 WITH 18 D.F.

RAMSEY RESET SPECIFICATION TESTS USING POWERS OF YHAT
RESET(2)= 4.2684 - F WITH DF1= 1 AND DF2= 145
RESET(3)= 2.3668 - F WITH DF1= 2 AND DF2= 144
RESET(4)= 1.5678 - F WITH DF1= 3 AND DF2= 143

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	0.0157	0.0778	0.2013	0.4088	1.9675	0.0413
2	0.0965	0.0778	1.2393	1.9664	1.8032	1.6149
3	-0.3629	0.0778	-4.6620	4.9072	2.6909	24.0202

Table 6
ERROR CORRECTION MODEL 1 - (LEVELS)
SMPL 1978.01 1990.12

```
 R-SQUARE =    0.9761     R-SQUARE ADJUSTED =    0.9725
VARIANCE OF THE ESTIMATE-SIGMA**2 =    3.5331
STANDARD ERROR OF THE ESTIMATE-SIGMA =    1.8797
SUM OF SQUARED ERRORS-SSE=   469.91
MEAN OF DEPENDENT VARIABLE =   65.669
LOG OF THE LIKELIHOOD FUNCTION = -304.416
```

VARIABLE NAME	ESTIMATED COEFFICIENT	STANDARD ERROR	T-RATIO 133 DF
ESRCPUS	1.1250	0.6484E-01	17.35
ESRCPUS	-0.86706E-01	0.6112E-01	-1.419
RPRICE	-2.8418	2.353	-1.208
RPRICE	2.3949	2.362	1.014
ZWHDPUS	0.17271E-01	0.2744E-02	6.294
ZWHDPUS	0.19694E-01	0.3177E-02	6.198
ZWCDPUS	0.59712E-01	0.9685E-02	6.166
ZWCDPUS	0.60861E-01	0.1136E-01	5.358
M1	1.0672	1.283	0.8319
M2	-6.6165	1.767	-3.745
M3	-4.8122	1.898	-2.535
M4	-5.2016	1.908	-2.726
M5	-3.5635	2.275	-1.567
M6	-0.94929	2.999	-0.3165
M7	-0.57520	3.823	-0.1505
M8	-4.1229	4.082	-1.010
M9	-2.1970	3.517	-0.6246
M10	-3.1068	2.422	-1.283
M11	-3.4736	1.421	-2.444
E1	-0.92432	0.7830E-01	-11.80
CONSTANT	-22.971	5.150	-4.460

```
DURBIN-WATSON = 1.6730    VON NEUMANN RATIO = 1.6839    RHO =   0.16017
DURBIN H STATISTIC (ASYMPTOTIC NORMAL) =    3.3477
COEFFICIENT OF SKEWNESS =    0.1188 WITH STANDARD DEVIATION OF 0.1955
COEFFICIENT OF EXCESS KURTOSIS =    0.0858 WITH STANDARD DEVIATION OF 0.3886

HETEROSKEDASTICITY TESTS
  E**2 ON YHAT:           CHI-SQUARE =       8.722 WITH 1 D.F.
  E**2 ON YHAT**2:        CHI-SQUARE =       9.647 WITH 1 D.F.
  E**2 ON LOG(YHAT**2):   CHI-SQUARE =       7.706 WITH 1 D.F.
  E**2 ON X (B-P-G) TEST:          CHI-SQUARE =     34.810 WITH  20 D.F.
  E**2 ON LAG(E**2)  ARCH TEST:    CHI-SQUARE =      0.269 WITH   1 D.F.
  LOG(E**2) ON X (HARVEY) TEST:    CHI-SQUARE =     20.781 WITH  20 D.F.
  ABS(E) ON X (GLEJSER) TEST:      CHI-SQUARE =     30.744 WITH  20 D.F.

RAMSEY RESET SPECIFICATION TESTS USING POWERS OF YHAT
 RESET(2)=    47.136      - F WITH DF1=   1 AND DF2= 132
 RESET(3)=    23.645      - F WITH DF1=   2 AND DF2= 131
 RESET(4)=    15.969      - F WITH DF1=   3 AND DF2= 130

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	0.1588	0.0806	1.9704	3.3505	1.6730	3.9586
2	0.0931	0.0806	1.1551	1.5003	1.7993	5.3280
3	-0.0211	0.0806	-0.2617	0.2728	2.0252	5.3988

Table 7
SIMPLE ERROR CORRECTION MODEL 2 - AVERAGE PRICE
SMPL 1978.03 1990.12

```
 R-SQUARE =    0.9756      R-SQUARE ADJUSTED =    0.9720
VARIANCE OF THE ESTIMATE-SIGMA**2 =    3.5674
STANDARD ERROR OF THE ESTIMATE-SIGMA =    1.8888
SUM OF SQUARED ERRORS-SSE=    470.90
MEAN OF DEPENDENT VARIABLE =    65.834
LOG OF THE LIKELIHOOD FUNCTION = -301.617
```

VARIABLE NAME	ESTIMATED COEFFICIENT	STANDARD ERROR	T-RATIO 132 DF
ESRCPUS	1.1183	0.6527E-01	17.13
ESRCPUS	-0.78654E-01	0.6108E-01	-1.288
ARPRICE	-0.54398	0.4199	-1.296
ZWHDPUS	0.17504E-01	0.2770E-02	6.320
ZWHDPUS	0.21256E-01	0.3072E-02	6.920
ZWCDPUS	0.58457E-01	0.9681E-02	6.038
ZWCDPUS	0.64642E-01	0.1085E-01	5.957
M1	0.80593	1.325	0.6084
M2	-7.3988	1.758	-4.209
M3	-5.8092	1.736	-3.347
M4	-6.0607	1.744	-3.475
M5	-4.2598	2.186	-1.949
M6	-1.3743	2.992	-0.4593
M7	-0.48717	3.875	-0.1257
M8	-3.9309	4.133	-0.9512
M9	-1.9416	3.570	-0.5438
M10	-2.7925	2.475	-1.128
M11	-3.2013	1.452	-2.204
E1	-0.94962	0.7433E-01	-12.78
CONSTANT	-23.091	5.143	-4.490

```
DURBIN-WATSON = 1.6930     VON NEUMANN RATIO = 1.7042     RHO =   0.14963
DURBIN H STATISTIC (ASYMPTOTIC NORMAL) =    3.1074
COEFFICIENT OF SKEWNESS =    0.1601 WITH STANDARD DEVIATION OF 0.1968
COEFFICIENT OF EXCESS KURTOSIS =    0.1854 WITH STANDARD DEVIATION OF 0.3911

HETEROSKEDASTICITY TESTS
  E**2 ON YHAT:           CHI-SQUARE =        6.339 WITH 1 D.F.
  E**2 ON YHAT**2:        CHI-SQUARE =        7.025 WITH 1 D.F.
  E**2 ON LOG(YHAT**2):   CHI-SQUARE =        5.588 WITH 1 D.F.
  E**2 ON X (B-P-G) TEST:         CHI-SQUARE =      27.700 WITH   19 D.F.
  E**2 ON LAG(E**2)  ARCH TEST:   CHI-SQUARE =       0.194 WITH    1 D.F.
  LOG(E**2) ON X (HARVEY) TEST:   CHI-SQUARE =      19.275 WITH   19 D.F.
  ABS(E) ON X (GLEJSER) TEST:     CHI-SQUARE =      28.262 WITH   19 D.F.

RAMSEY RESET SPECIFICATION TESTS USING POWERS OF YHAT
 RESET(2)=    50.025      - F WITH DF1=    1 AND DF2= 131
 RESET(3)=    24.981      - F WITH DF1=    2 AND DF2= 130
 RESET(4)=    16.921      - F WITH DF1=    3 AND DF2= 129

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	0.1485	0.0811	1.8314	3.0599	1.6930	3.4208
2	0.0961	0.0811	1.1846	1.5362	1.7936	4.8616
3	-0.0072	0.0811	-0.0883	0.0919	1.9924	4.8697

Table 8
ERROR CORRECTION MODEL 3 (FIRST DIFFERENCES)
SMPL 1978.02 1991.12

```
 R-SQUARE =    0.9400      R-SQUARE ADJUSTED =    0.9322
VARIANCE OF THE ESTIMATE-SIGMA**2 =    6.1604
STANDARD ERROR OF THE ESTIMATE-SIGMA =    2.4820
SUM OF SQUARED ERRORS-SSE=   893.25
MEAN OF DEPENDENT VARIABLE =   0.14071
LOG OF THE LIKELIHOOD FUNCTION = -373.461
```

VARIABLE NAME	ESTIMATED COEFFICIENT	STANDARD ERROR	T-RATIO 145 DF
DKWH	-0.12780	0.7978E-01	-1.602
DKWH	-0.22376	0.6604E-01	-3.388
DPRI	-10.420	2.900	-3.593
ZWHDPUS	0.20224E-01	0.3496E-02	5.785
ZWHDPUS	0.95936E-02	0.4004E-02	2.396
ZWCDPUS	0.71452E-01	0.1231E-01	5.804
ZWCDPUS	0.30659E-01	0.1413E-01	2.169
M1	1.5196	1.808	0.8407
M2	-8.0729	2.536	-3.183
M3	-5.4827	2.538	-2.161
M4	-6.3355	2.455	-2.580
M5	-1.4773	2.902	-0.5090
M6	3.0585	3.764	0.8126
M7	1.5701	4.790	0.3278
M8	-4.8394	5.147	-0.9402
M9	-5.5245	4.472	-1.235
M10	-9.2496	3.055	-3.028
M11	-6.3833	1.833	-3.483
E1	-0.21166	0.6975E-01	-3.035
CONSTANT	-18.155	3.693	-4.917

```
DURBIN-WATSON = 1.9513    VON NEUMANN RATIO = 1.9632    RHO =   0.02125
...DURBIN H STATISTIC CANNOT BE COMPUTED
COEFFICIENT OF SKEWNESS =   -0.2776 WITH STANDARD DEVIATION OF 0.1890
COEFFICIENT OF EXCESS KURTOSIS =    0.1870 WITH STANDARD DEVIATION OF 0.3758

HETEROSKEDASTICITY TESTS
  E**2 ON YHAT:           CHI-SQUARE =        0.577 WITH 1 D.F.
  E**2 ON YHAT**2:        CHI-SQUARE =        5.514 WITH 1 D.F.
  E**2 ON LOG(YHAT**2):   CHI-SQUARE =        5.657 WITH 1 D.F.
  E**2 ON X (B-P-G) TEST:         CHI-SQUARE =      27.678 WITH  19 D.F.
  E**2 ON LAG(E**2)  ARCH TEST:   CHI-SQUARE =       0.827 WITH   1 D.F.
  LOG(E**2) ON X (HARVEY) TEST:   CHI-SQUARE =      15.120 WITH  19 D.F.
  ABS(E) ON X (GLEJSER) TEST:     CHI-SQUARE =      26.387 WITH  19 D.F.

RAMSEY RESET SPECIFICATION TESTS USING POWERS OF YHAT
 RESET(2)=     7.0529      - F WITH DF1=   1 AND DF2= 144
 RESET(3)=     3.6075      - F WITH DF1=   2 AND DF2= 143
 RESET(4)=     2.4032      - F WITH DF1=   3 AND DF2= 142

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     0.0211      0.0778       0.2714       0.5197      1.9513       0.0750
   2     0.0591      0.0778       0.7585       1.1913      1.8640       0.6645
   3    -0.3685      0.0778      -4.7339       5.0394      2.6854      23.7663
```

TABLE 9

RESIDENTIAL ELECTRICITY SALES, BILLIONS OF KWH
ONE YEAR AHEAD FORECASTS

Year	Actual KWH	Long Run	Short Run 1	Short Run 2	ECM 1	ECM 2	Short Run 3	ECM 3
1986	819.09	802.28	824.48	804.22	828.54	807.37	845.71	852.30
1987	850.41	830.41	844.44	840.61	846.53	841.76	854.29	855.27
1988	892.87	861.02	897.75	899.39	894.20	897.78	901.99	897.23
1989	905.52	893.24	879.21	874.77	878.76	873.69	884.07	883.91
1990	924.02	904.80	902.94	901.75	901.18	899.57	913.83	912.46
1991	957.02	910.09	924.02	931.08	936.16	941.58	921.53	936.41

TWO YEAR AHEAD FORECASTS

	Actual KWH	Long Run	Short Run 1	Short Run 2	ECM 1	ECM 2	Short Run 3	ECM 3
1987	850.41	818.83	839.05	834.66	842.84	837.30	861.65	873.03
1988	892.87	850.61	896.19	896.78	894.06	895.50	927.77	926.95
1989	905.52	878.90	882.49	878.73	882.22	878.89	946.33	941.53
1990	924.02	900.69	878.88	887.97	876.60	885.15	843.61	842.37
1991	957.02	905.42	914.41	927.18	911.00	925.02	901.54	898.71

ONE YEAR AHEAD FORECAST ERRORS

Year	Actual KWH	Long Run	Short Run 1	Short Run 2	ECM 1	ECM 2	Short Run 3	ECM 3
1986	819.09	16.81	-5.39	14.87	-9.45	11.72	-26.62	-33.21
1987	850.41	20.00	5.97	9.80	3.88	8.65	-3.88	-4.86
1988	892.87	31.85	-4.89	-6.52	-1.33	-4.92	-9.12	-4.36
1989	905.52	12.29	26.31	30.75	26.76	31.83	21.46	21.61
1990	924.02	19.22	21.07	22.27	22.84	24.45	10.19	11.56
1991	957.02	46.93	33.00	25.95	20.87	15.45	35.50	20.62
	AVERAGE	24.52	12.68	16.19	10.59	14.53	4.59	1.89
	MEDIAN	19.61	13.52	18.57	12.38	13.59	6.31	7.20
	MAE	24.52	16.11	18.36	14.19	16.17	17.80	16.04
	MAPE	2.72	1.76	2.04	1.56	1.80	1.98	1.82

TWO YEAR AHEAD FORECAST ERRORS

	Actual KWH	Long Run	Short Run 1	Short Run 2	ECM 1	ECM 2	Short Run 3	ECM 3
1987	850.41	31.58	11.36	-15.57	7.57	13.11	-11.24	-22.62
1988	892.87	42.25	-3.32	-46.37	-1.19	-2.63	-34.90	-34.09
1989	905.52	26.63	23.03	14.14	23.30	26.63	-40.81	-36.01
1990	924.02	23.33	45.14	17.56	47.42	38.87	80.41	81.65
1991	957.02	51.61	42.62	-3.16	46.03	32.00	55.48	58.31
	AVERAGE	35.08	23.77	-6.68	24.63	21.60	9.79	9.45
	MEDIAN	31.58	23.03	-15.57	23.30	26.63	-11.24	-22.62
	MAE	35.08	25.09	19.36	25.10	22.65	44.57	46.54
	MAPE	3.86	2.72	2.16	2.71	2.47	4.85	5.08

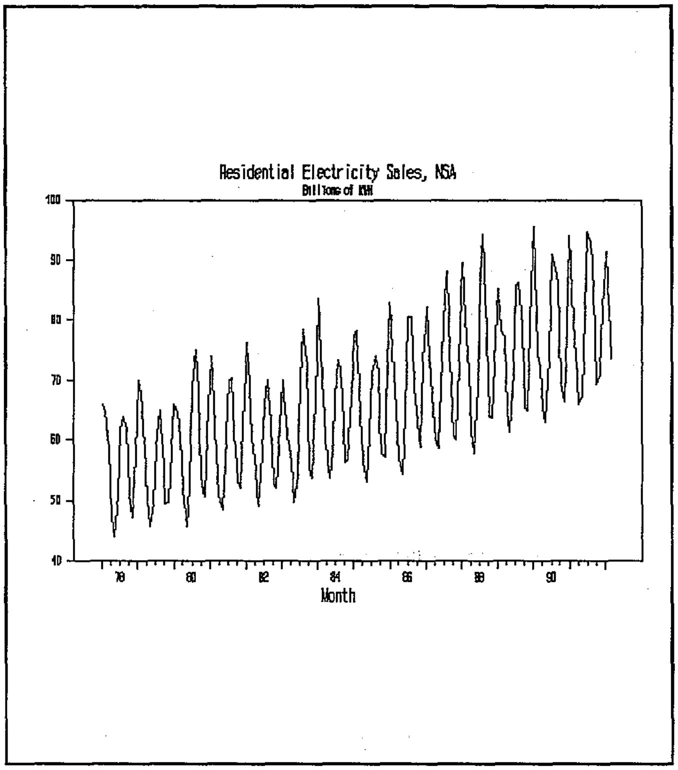


Figure 1

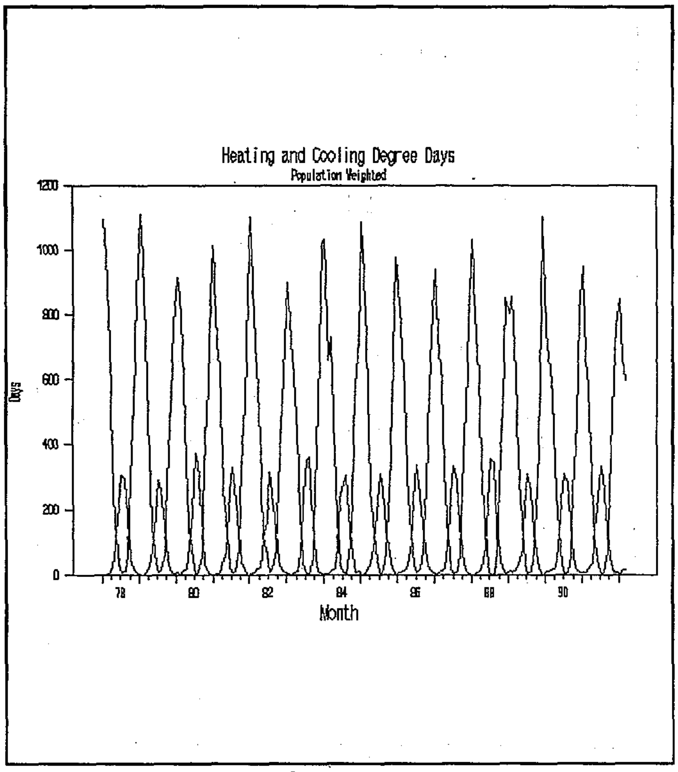


Figure 2

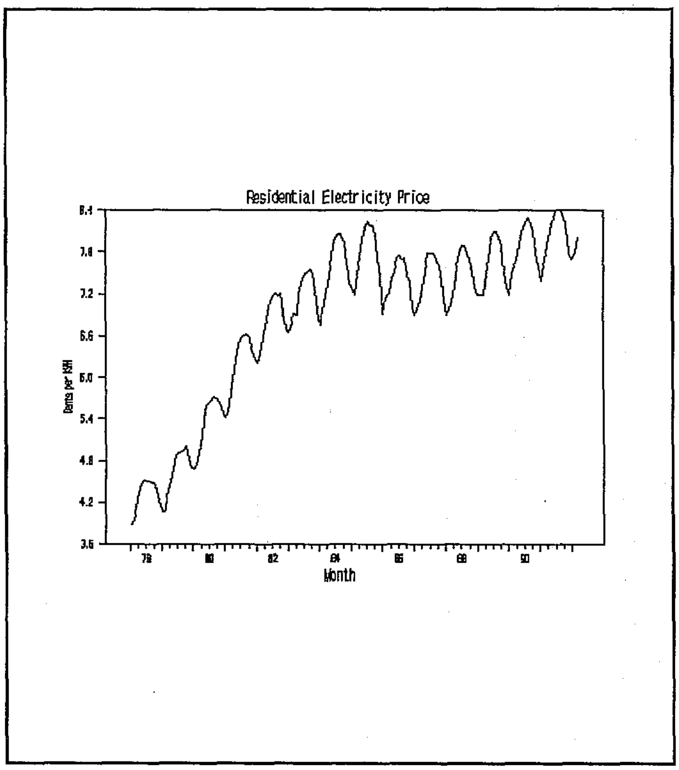


Figure 3

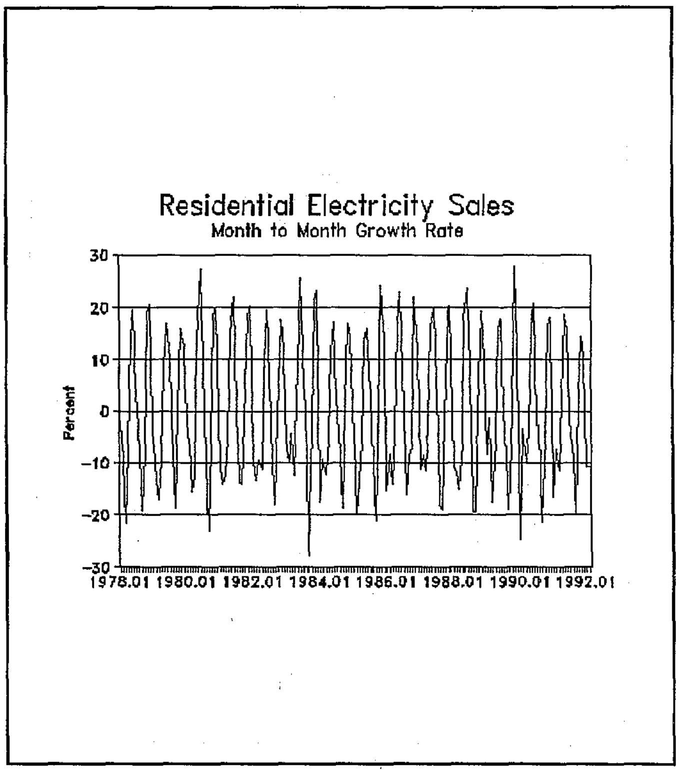


Figure 4

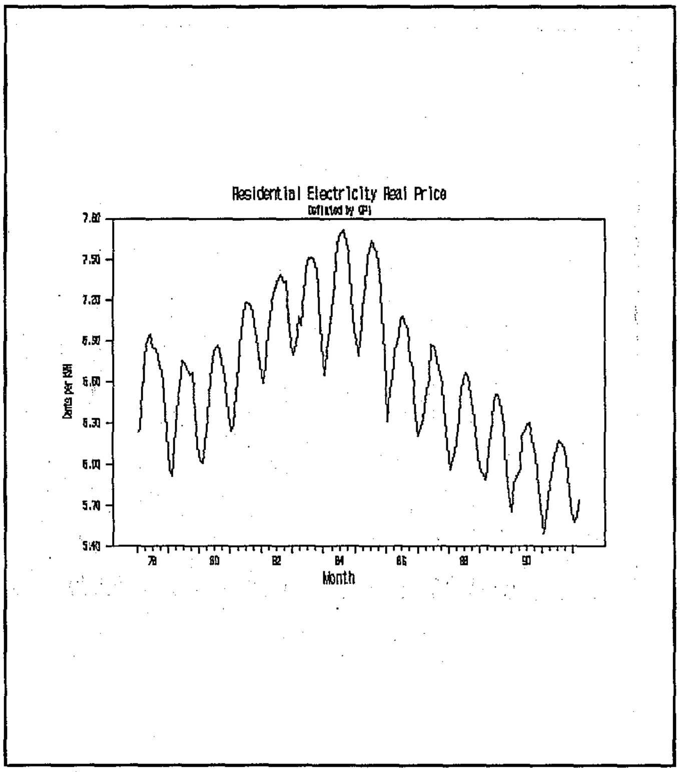


Figure 5

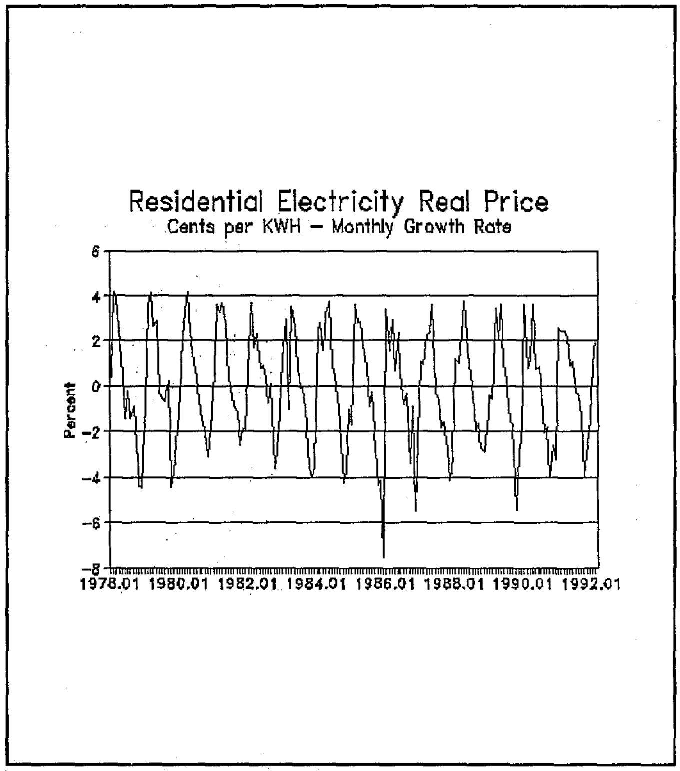


Figure 6

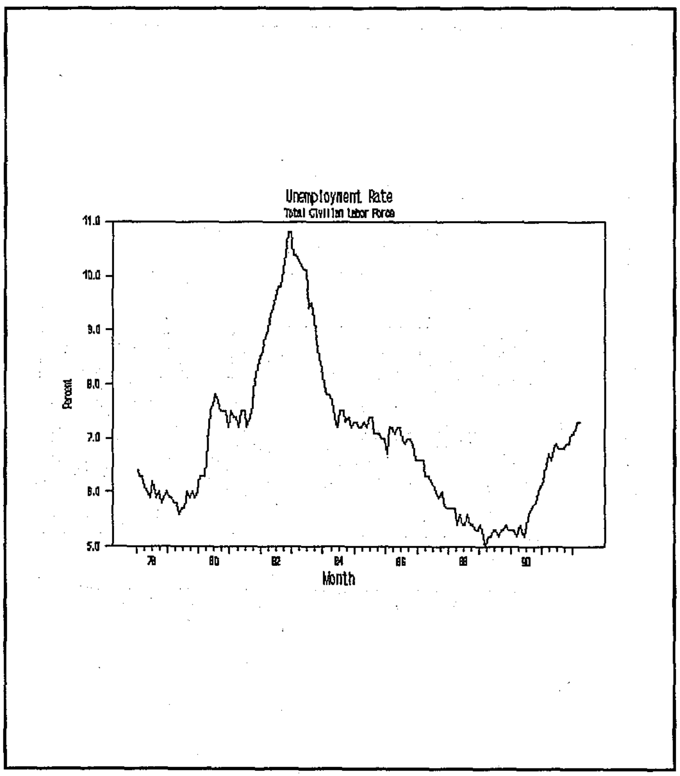


Figure 7

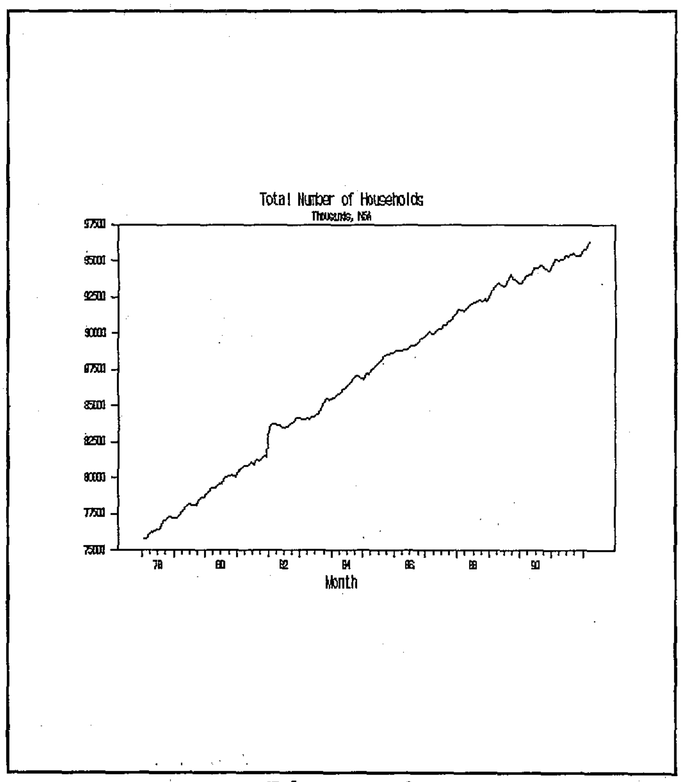


Figure 8

Determinants of Short-term Agricultural Loan Rates at Commercial Banks

Paul A. Sundell, Economic Research Service, U.S. Department of Agriculture

Short-term debt financing from commercial banks is an important component of the cost of capital for many agricultural investment projects. In 1991, farmers spent $13.4 billion on total interest expenses of which $6.2 billion was short-term interest expenses. Commercial banks make more nonreal estate farm loans than any other farm lender. Little econometric work has been done to model short-term farm lending rates. This paper develops econometric models for forecasting interest rates charged by small and large banks on nonreal estate farm loans by small and large banks. The paper is useful in examining interest rate linkages from the macroeconomy to the agricultural sector.

What Determines Bank Lending Rates?

In determining how much to charge for a loan--its loan pricing decision--the bank must cover the basic costs involved in the loan. The loan rate reflects funds costs, default risk, and transactions costs.

(1) Loan rate = f(cost of funds to banks, risk, transactions costs)

Banks vary significantly in terms of their funds costs, default risk characteristics of their loans, and transactions costs. Funds costs include the costs of debt (primarily deposits) to the bank as well as returns to bank equity holders. Default risk involves the costs per dollar of loan from defaults. Transactions costs include such costs as general overhead and transaction costs in loan processing. Small and large banks involved in agricultural lending differ significantly in terms of these cost factors. The next section of the paper examines these cost factors in somewhat greater detail.

Cost of Funds to Banks

Banks raise funds through the debt (primarily deposit) and equity markets. Funds raised through the deposit market may be divided between core deposits (small deposits -under $100,000- generally raised in the bank's service area) and managed liabilities (large deposits over $100,000, bank borrowings, and all foreign deposits). Even in a deregulated interest rate environment, core deposits are less sensitive and respond with a lag to changes money market rates.

Small banks typically maintain higher capital to asset ratios than their large bank counterparts.[1] Since equity is a residual claim earnings, equity is more costly at the margin than debt for banks over most of the range of debt and equity combinations. Smaller banks, because of their smaller size and less diversified loan portfolios, generally have greater uncertainty of achieving positive taxable income, thus the value of their debt and depreciation tax shields are more uncertain. To the extent that the expected tax shield for smaller banks is more uncertain, the required return on small bank equity will be higher than for comparable larger banks.

Another area of difference between large and small banks is the relative importance of average and marginal cost of funds pricing in agricultural loan rate determination. Large banks in general give greater weight to the marginal cost of bank funds in their pricing of agricultural loans than small banks. On the other hand, small banks give relatively greater weight to average cost of funds in the pricing of their agricultural loans. The greater importance of average cost pricing is presented in Table 1. Table 1 shows an American Bankers Association survey of of banks concerning indexes used in agricultural loan rate determination.

The choice of marginal and average cost pricing in agricultural loan pricing is jointly determined by the preferences of banks and agricultural borrowers. In some cases, both borrowers and lenders may benefit from an average cost of funds approach. For borrowers, loan rates determined by the bank's average cost of funds generally are less volatile. For banks, when a change in average interest rate expenses generates an exact change in average interest earnings on their loan portfolio, average cost pricing will help hedge the banks' net worth exposure to interest rate changes.

In markets where there is aggressive competition among banks for loans, loans rates tend to be determined more by the marginal cost of bank funds. Marginal cost pricing allows loan rates to fall more rapidly when interest rates and bank cost of funds are falling, thus enhancing the ability of banks to compete for loans in a falling interest rate environment. Therefore, borrowers, with access to many banks and willingness to accept somewhat greater volatility of marginal cost of funds pricing, may find that long-run borrowing costs are lower with marginal cost pricing.

---

[1]. Stockholders of small banks with undiversified loan portfolios are especially likely to demand a larger expected return on the bank's stock if transaction costs make it costly for stockholders to diversify away the diversifiable unique risk of the bank's equity. Transactions costs generate costs that cause perfectly diversified portfolios to be excessively costly and unique risk to be important in the determination of the expected return on equity. For a greater discussion of this point, the interested reader should refer to Flannery (pp. 458-459).

Risk
The two major sources of individual loan and portfolio risk are default risk and interest rate risk. Default risk is the risk that the stated contractual interest rate will not be paid either in its entirety or on time. The default risk premium is the additional expected return on the loan above the default-free rate (Treasury yields) the bank demands for bearing default risk. The default risk premium will normally flow to holders of the banks equity and non-insured debt. Interest rate risk is the risk that the value of the asset (in this case a loan) will decline if general interest rates rise. Typically, longer term fixed rate loans will bear a term premium to compensate the lender for higher interest rate risk. Individual and portfolio interest rate risk can be reduced through the use of variable rate lending as well as interest rate hedges such as interest rate futures and options.

Loans that are strongly correlated with returns to the bank's loan portfolio will typically be charged a premium. In general, banks that do not hold diversified loan portfolios will be subject to greater portfolio risk. Because undiversified loan portfolios increases the risk of volatile earnings and bankruptcy, bank stockholders will typically demand higher expected returns on the equity of banks with relatively undiversified loan portfolios. Therefore, banks with undiversified loan portfolios are forced to charge higher loan rates to achieve higher long-run returns for their stockholders. Many small agricultural banks typically have fewer opportunities to diversify risk and therefore may be forced to charge somewhat higher loan rates for similar loans than their larger, more diversified banking counterparts.

In addition to setting loan rates based on returns on comparable risk-free assets, default risk, and portfolio risk, banks may use credit rationing to control risk. As real borrowing rates rise above a critical point, the probability of borrower default may rise significantly, causing the loan's expected return to fall with higher loan rates, thus lowering the loan's expected return. As illustrated by Stiglitz and Weis, the risk of borrower default risk may rise significantly with rising real borrowing costs for two reasons:(1) during high real interest rate periods, the best borrowers with the greatest credit flexibility may withdraw from the market and (2) higher real loan rates will likely cause firms to choose more projects with higher expected returns but greater risk. Bank profit maximization often involves the rationing of credit away from high risk borrowers that are willing to agree to higher loan rates but are denied credit because of high and rising likelihood of default. The amount of credit rationing present will be a positive functions of the perceived riskiness of various borrower classes and the real risk free rate. High real interest rates may result in overall reduced loan margins as high risk borrowers are rationed out of the market.

Transactions Costs and Other Factors in Loan Pricing
Transactions costs include such as administrative expenses in loan processing and general overhead. Transactions costs per dollar of loan generally declines as loan size increases. This inverse relationship reflects the significant fixed component of gathering and analyzing credit information and loan processing. Because per dollar transactions costs decline with the size of loan, borrowers are typically charged lower per dollar loan fees as loan size increases.

An additional issue in loan costs are whether the presence of economies of scale and economies of scope drive down the costs of providing banking services, such as loans at large banks. Overall, the empirical evidence coupled with the recent large scale merger activity in banking suggests larger banks may have some cost advantages over their smallest counterparts; however, the issue is far from settled. In addition smaller banks with specialized agricultural lending departments may have advantages in evaluating specific agricultural risk, and providing lending services tailored to agricultural borrowers. Therefore, even if large banks have some general cost advantages in banking, small agricultural banks may have certain technical advantages in farm lending compared to their larger nonagricultural bank counterpart.

Methodology and Data
Having discussed general factors in loan rate determination, this section discusses estimating reduced form equations for agricultural loan rates. The following explanatory variables were used. The large CD rate and the prime rate were used as proxies for marginal and average costs of funds to banks respectively. Some financial economists have argued the national prime rate reflects large banks average costs of funds with a markup (Brady and Goldberg (1982)). Changes in the prime likely overstate changes in the average costs of funds for banks heavily involved in agricultural lending, especially smaller banks that have a large base of consumer deposits that adjust somewhat slowly especially during times of rising interest rates (Wenninger and Mahoney et. al.).[2] To capture the slower adjustment of funds costs especially at small rural banks, the average rate last quarter for consumer deposits maturing between 92 and 182 days and average rate last year for deposits maturing in 1 to 2.5 years were included in the model as well.
Feedbacks from agricultural loan rates to national money market rates are viewed as minimal.

Farm default risk and default risk premiums were made a function of farmer debt to asset ratios as well as interest coverage ratios. Farmer default risk can be expected to increase as the proportion of farm assets acquired through debt

---

[2]. Data for core consumer deposit interest rates are from The Monthly Survey of Selected Deposits published in the Money Stock Release (H.6). Consumer deposit rate data is for small time deposits with maturities of 92 to 182 days and 1 to 2.5 years. The deposit rates are for the last Wednesday of each month.

increases. The farm business debt service ratio is included to monitor current liquidity conditions in farming. [3] Given farming's large investment in illiquid farmland and equipment as well as cyclical changes in farm income, the debt service ratio may provide additional information about overall default risk in farming not solely provided by the debt to asset ratio.

Other variables related to credit rationing and bank risk aversion were included in the model. The bank credit rationing variables included last quarter's real ex post t-bill rate and last quarter's real small and large bank agricultural lending rates as well as last year's bank return on equity. End of the quarter loan to deposit ratios for agricultural banks and large commercial banks were included to examine if farm loan interest margins increased with larger loan to deposit ratios. Increasing loan to deposit ratios, especially for smaller banks, may be symptomatic of overall falling bank liquidity. These additional variables related to credit rationing and loan deposit ratios along with the farm business debt service ratio were found to be insignificant in all the regressions.

Regression Results

The reduced form equations were estimated over the 1984IV-1992I period. Financial deregulation progressed rapidly in the 1980's. In October 1983, all interest rate ceilings on time deposits were removed. To allow for the adjustment of the pricing of small time deposits at commercial banks to the deregulation of small time deposits, the estimation period used is 1984IV-1992I. Explanatory variables that were not significant in any regressions were dropped. The regression results are shown in table 2.

The regressions indicate that the relevant cost of funds variables are, as expected, the most important determinants of short-term agricultural loan rates at large and small banks. For the small bank equation, the econometric evidence clearly indicates the prime is an important relative cost variable. Econometric evidence by Brady and Goldberg (1984) indicates that the prime rate fully responds to an increase in the large CD rate with a 2 or 3 month lag. The exact shape of the lag will depend on factors such as bank loan demand and competitive pressures from bank and nonbank lenders.

Given the relatively larger and more significant coefficient for the prime rate in the small bank regression and the insignificance of the 3-month CD rate in the small bank regression, the results indicate average cost of funds pricing is more important for small banks in pricing short-term agricultural loans. On the other hand, both the 3-month CD rate and the prime rate were significant for the large bank rate equation, suggesting that both marginal and average costs considerations are important in determining lending rates at large banks.

The coefficients for the average rate on the 1.0 to 2.5 year small time deposits were significant for both the large and small bank regression. However, the size and significance of the small time deposit coefficients were much larger in the small bank regressions. The results reflect the greater importance of core deposits and average cost pricing to small banks relative to large banks. The combination of average cost pricing, lagged adjustment of consumer deposits to changes in open market rates, and the typical longer maturity of core deposits when compared to managed liabilities help explain the much slower adjustment of agricultural loan rates at small commercial banks.

The evidence was mixed concerning interest rate term premiums. As suggested earlier, as the percentage of fixed rate loans and the average maturity of loan increases, interest rate risk and overall term premiums are likely to increase. Likewise as the average maturity of short-term farm loans increases, term premiums are likely to increase. The coefficient for the percentage of fixed rate loans was positive and but not significant in the equation for small bank loan rates. The coefficients for the average maturity of agricultural loans at small and large banks were not significant in their respective equations.

The significant negative coefficients on the percentage of loans at fixed rates at large banks may reflect a tendency for very short-term loans to low-risk borrowers to generally be made at fixed rates. Given the relatively fewer agricultural loans made by large banks, a few very large fixed rate short-term farm loans may have a significant impact on the aggregate agricultural loan rate at large banks. Because of the well under one year average maturity of nonreal estate farm loans, term premiums probably are quite small for most short-term farm loans.

The coefficients on the average size of loan variables were negative and significant as expected. The coefficients for the size of loan was significant at the ten percent level for the small bank equation and at least at the five percent level for the large bank equations. The larger coefficient for the small bank equation probably reflects the typically much smaller size of farm loans at small banks. For example, in 1991 the average size non real estate at small banks was $13,900 and $107,000 at large banks.

---

[3]. Farm liquidity is measured by the farm business debt service coverage ratio. This ratio is defined to be net cash farm income and farm interest payments divided by interest and principal payments. Data on the farm business debt service coverage ratio may be found in the National Financial Summary, 1991 (forthcoming) and past issues of the Agricultural Income and Finance Situation and Outlook.

Examination of the residuals for the large bank equation indicated a large positive outlier for the fourth quarter of 1990IV. Examination of the recursive residuals (discussed in the next section) also indicated an outlier for 1990IV. Therefore, a dummy variable was added for the fourth quarter of 1990IV. Discussions with Nicholas Walraven at the Federal Reserve Board indicate that changes in the volume of very large short-term agricultural loans typically priced at or below the prime are a significant source of volatility to the quarter to movements in the large bank interest rate series. In the case of the 1990IV, the data indicated the volume of these large declined. Specifically, the average size of loans over $250,000 made by large banks declined by over $500,000 in 1990IV before increasing by nearly $300,000 in the first quarter of 1991.

Diagnostic Checking

Diagnostic tests were performed on the small and large bank regressions. Well specified models typically yield residuals that are distributed normally and independently with constant variances. Departures from independent and normally distributed residuals with constant variance residuals indicate the residuals contain information that the model builder should exploit in specifying the model. In using ordinary least squares estimation, white noise residuals are necessary for consistent estimators of the variance covariance and hypothesis testing. If dependence between the explanatory variables and the residuals exists, because of omitted variables, simultaneous equation relationships, or errors in functional form, OLS parameter estimates will fail to be consistent.

Diagnostic tests used in this paper include specific tests of residual behavior where a specific alternative hypothesis to well behaved residuals is tested and general tests of mispecification where a specific alternative to random residuals is not tested. Examples of specific tests of residual behavior include the Lagrange Multiplier (LM) tests of joint significance of various degrees of autocorrelation and Engle's autoregressive conditional heteroscedasticity test. Examples of general tests of residual behavior include the White heteroskedasticity test, the general tests of mispecification by Ramsey and the various recursive residual tests by Brown, Durbin, and Evans. A very brief discussion of each of the tests is provided in table 3. Readers not acquainted with the tests will likely benefit from examining the original articles.

As shown by examining the diagnostic tests provided in table 2 and the recursive residual test in table 3, the performance of the final small bank loan model performed extremely well. None of the diagnostic test reported in table 2, could reject the hypothesis of a well specified model with well behaved disturbances at the 5 percent significance level. The recursive residual tests were extremely well behaved with the recursive residuals, cusum, and cusum squared residuals all well within their confidence bands.

The large bank equation also performed well. However, examination of the OLS residuals and the recursive residual indicated an abnormally large residual for 1990IV. As mentioned earlier, large short-term loans priced well below prime have been a source of variability in the large bank series. The recursive residual series for the final large bank equation (equation 5) <u>without</u> the dummy variable presented in figure 6 indicated well behaved recursive residuals with the exception of 1990IV. The inclusion of the dummy variable improved the equations fit and reduced evidence of lower order autocorrelation. The large bank equation with the dummy variable showed no evidence of autocorrelation and passed the variance heteroscedasticity and specification error tests at the five percent significance level.

Conclusion

This paper examined determinants of short-term agricultural loan rates at small and large banks. Empirical support was found for the view that short-term agricultural bank loans are determined in part on an average cost of funds basis while large banks price agricultural loans on a combination of their marginal and average cost of funds basis. In addition, farmer default risk and lender willingness to bear risk unique to agriculture were more important to the determination of small agricultural loan rates.

Diagnostic tests were performed on the small and large bank equations that supported the view that the equations were well specified.

Table 1. Indexes Used In Variable Rate Determination By Agricultural Banks (Percentage of Banks)

	Bank Asset Size (in million dollars)	
	Less than $100	$100 or More
T-Bill rate	9.2	14.6
External prime rate	32.0	45.8
Average cost of bank funds	51.0	35.4
Marginal cost of bank funds	2.6	4.2
Other	23.5	12.5
No. of Banks	153	48

Table 2.
Small Bank and Large Bank Short-Term Non-Real Estate Farm Loan Regressions, 1984IV-1992I.
(t-statistics in parentheses)

	Small Bank		Large Bank		
	(1)	(2)	(3)	(4)	(5)
Constant	4.12	3.82	2.77	2.92	2.44
(6.42)	(8.67)	(2.99)	(3.90)	(7.33)	
Prime rate	0.38	0.48	0.33	0.38	0.47
(first month	(2.53)	(11.77)	(1.59)	(2.28)	4.49
of quarter)					
3-month CD rate	0.07		0.61	0.55	0.49
(first month of quarter)	(0.65)		(4.07)	(4.48)	(5.34)
Average rate on com. bank	0.31	0.28	0.13	0.13	0.09
1 to 2.5 year small time	(4.68)	(5.71)	(1.53)	(1.79)	(2.06)
deposits (over preceding					
4 quarters)					
Ratio of farm debt to	0.08	0.09	-0.03		-0.02
farm assets (market	(3.19)	(5.18)	(-0.78)		(-0.73)
value last year)					
Average Size of loan	-0.03	-0.03	-0.005	-0.005	-0.008
(in thousands)	(-1.78)	(-1.86)	(-2.30)	(-2.40)	(-4.12)
Percentage of loans	0.56	0.52	-1.15	-1.27	-1.35
at a fixed rate	(1.21)	(1.16)	(-2.43)	(-3.33)	(-3.74)
Dummy Var. 90.IV				0.83	0.84
				(3.67)	(3.76)
Adjusted R^2	0.968	0.969	0.978	0.982	0.982
Std. error	0.177	0.174	0.240	0.193	0.191

Diagnostic Tests (significance levels in parentheses)

	(1)	(2)	(3)	(4)	(5)
Durbin-Watson	2.31	2.37	2.61	2.19	2.20
Breusch-Godfrey LM					
lag 1 (significance)	1.18	1.53	4.00	0.66	0.70
	(0.28)	(0.22)	(0.05)	(0.42)	(0.65)
lag 4 (significance)	6.62	7.25	8.96	9.73	9.16
(0.16)	(0.12)	(0.06)	(0.05)	(0.06)	
Jarque-Bera (normality)		1.12			2.50
		(0.57)			(0.29)
ARCH					
(lag 1, LM test)		0.76			2.42
		(0.38)			(0.12)
(lag 4, LM test)		2.24			2.45
		(0.69)			(0.65)
White (LM)		8.05			--.--
		(0.62)			--.--
Ramsey Specification Test		1.48			1.19
(F Test)		(0.24)			(0.33)

Table 3.
Brief Overview of Diagnostic Tests

Breusch-Godfrey generalized autocorrelation test. Tests autocorrelation by regressing OLS residuals on the set of explanatory variables and lagged residuals of general order k. The Lagrange multiplier nR^2 statistic from this regression is distributed as a chi-squared with k degrees of freedom.

Jarque Barque Normality test. If the residuals are normal, the test statistic: $n(((\text{skewness}^2) / 6) + ((\text{kurtosis} - 3)^2 / 24))$ will be distributed chi-squared with 2 degrees of freedom.

Autoregressive Conditional Hetroscedasticity (ARCH). The ARCH model allows the variance of the error term to vary through time by making the variance of this period's error term a function of the variances of previous error terms. The hypothesis of ARCH residuals is tested by regressing the squared OLS residuals on k lagged squared residuals and a constant. Under the null hypothesis of constant variance over time the lagged squared residuals will be jointly insignificant and the test statistic NR^2 will be distributed as a chi-squared random variable with k degrees of freedom. Lags of order 1 and 4 were tested.

White Heteroskedasticity test. In testing for hetroscedasticity, White suggests regressing the OLS squared errors on the explanatory variables, their squares, and cross products. The test statistic nR^2 from the above regression tests the joint significance of the above explanatory variables. The test statistics distributed as a chi-square random variable with the number of degrees of freedom equal to the number of explanatory variables (not counting the constant term). In applying the White heteroscedasticity test to my data, the cross product terms were dropped because of the relatively small sample size. The White test is a general test and the finding of joint significance of the explanatory variables may be caused by specification error as well as heteroscedastic disturbances. The White test could not be used for the dummy variable large bank model because of perfect multicollinearity.

Ramsey test. The Ramsey test is a general test of misspecification. The test consists of adding to the regression as explanatory variables squared and higher powers of the fitted values from the OLS regression. In the case of mispecification, the error term from the OLS regression will in general be composed of the true error term and an additional component resulting from misspecification. In general, the mispecification component of the error term will not be random and the will be related to fitted values of the model. Test results used second, third, and fourth power of the fitted values for both the small and large bank equations. F tests for the joint significance of the powers of the fitted residual terms in included in diagnostic tests.

Recursive Residuals. Recursive residuals are one step ahead forecast errors multiplied by a scaling factor. The scaling factor is the square root of the ratio of the variance of the in-sample error term to the variance of the forecast error term. The recursive residuals are generated by repeatedly reestimating the model to include the previous period's observation. Therefore, for a model with k explanatory variables (not counting the constant term), the first $k+2$ observations are used to estimate the first recursive residual. If the variance of the in-sample error term is normally and independently distributed with constant variance, the recursive residuals will be normally distributed with mean 0 and variance equal to the variance of the in-sample error term. If the recursive residuals do not exhibit a random normal pattern, evidence of parameter instability, heteroskedasticity, or mispecification exists. The paper also reports the cusum and cusum squared recursive residual tests. The cusum test reports the cumulative sum of the recursive residuals at each point in time divided by the standard error of the regression. If parameters and variance of the error term is stable over time and the model is correctly specified, the expected value of all the recursive residuals should be zero. Therefore, the expected value of the sum of the recursive residuals should also be equal to zero. Prolonged deviations outside the confidence intervals above and or below zero suggests model misspecification. The cusum squared test is the ratio of the cumulative squared sum of recursive residuals through time t divided by the cumulative sum of all the recursive residuals for the entire sample period. As with the Cusum test, movements outside the confidence intervals indicates parameter instability, non constant error term variance or specification error from omitted relevant explanatory variables or functional form.

REFERENCES

American Bankers Association. 1989 Farm Credit Situation Survey Washington D.C. 1989.

Barrickman John. "Loan Pricing Guidelines," Journal of Agricultural Lending, Vol. 2, Issue 1 (Spring 1988) pp. 2-6.

Barry Peter J. and Jeffrey D. Calvert. "Loan Pricing and Profitability by Agricultural Banks," Agricultural Finance Review, Vol. 43, 1983 pp. 21-29.

Brady, Thomas F. The Role of the Prime Rate in the Pricing of Business Loans. Board of Governors, Staff Paper No. 146, Washington D.C. November 1985.

Brown R. L., J. Durbin, and J. M. Evans. "Techniques for Testing the Constancy of Regression Relationships" (with discussion), Journal of the Royal Statistical Society, Series B, Vol. 37, 1975, pp.149-163.

Buck, Walter H. "A Risk Management Approach to Pricing Loans and Leases," Journal of Commercial Bank Lending, April 1979, pp. 2-15.

Brunner Allan D., John V. Duca, and Mary M. McLauglin. "Recent Developments Affecting the Profitability and Practices of Commercial Banks," in Federal Reserve Bulletin, July 1991, pp.505-527.

Engle, R. "Autoregressive Conditional Heteroscedasticity, with Estimates of the Variance of United Kingdom Inflations, Econometrica, Vol. 50, pp. 987-1008.

Flannery, Mark J. "A Portfolio View of Loan Selection and Pricing," in Handbook for Banking Strategy, R.C. Aspinwall and R.A. Eisenbeis eds., Wiley Publishing Co, New York, pp. 457-472.

Godfrey, L. G. "Testing Against General Autoregressive and Moving Average Error Models when the Regressors Include Lagged Dependent Variables," Econometrica, Vol. 46, 1978, pp. 1293-1302.

Goldberg, Michael A. "The Pricing of the Prime Rate," Journal of Banking and Financing, Vol. 6 (1982) pp. 277-296.

Goldberg, Michael A. "The Sensitivity of the Prime Rate to Money Market Conditions," Journal of Financial Research, Vol. 3, No. 4 (Winter 1984) pp. 269-280.

Jackson, Michael A. "What Borrowers Look for in an Agricultural Lender," Journal of Agricultural Lending, Vol. 1, Issue 4 (Winter 1988) pp. 22-25.

Mahoney Patrick I., Alice P. White, Paul F. Obrien, and Mary M. Mclaughlin. Responses to Deregulation: Retail Deposit Pricing from 1983 through 1985, Staff Study No. 151, Board of Governors of the Federal Reserve System, 1987.

Ramsey J. B. "Tests for Specification Error in Classical Linear Least Squares Regression Analysis," Journal of the Royal Statistical Society, Series B, Vol. 31, 1969, pp. 350-371.

Sinkey, Joseph F. Jr. Commercial Bank Financial Management In the Financial Services Industry, Third Edition 3 (especially chapter 19) MacMillan Publishers 1989.

Stiglitz, J. E. and A. Weis. "Credit Rationing in Markets with Imperfect Information," American Economic Review, June 1981, pp.373-409.

United States Department of Agriculture. National Financial Summary, 1990, Nov. 1991.

Walraven, Nicholas A. and John Rosine. Agriculture Financial Databook, Board of Governors, Washington D.C. (various issues).

Wenninger, John. "The Responsiveness of Interest Rate Spreads and Deposit Market Flows to Changes in Market Interest Rates," Quarterly Review, Federal Reserve Bank of New York, Fall 1986, pp. 1-10.

White Halbert. "A Heteroskedasticity-Consistent Estimator and a Direct Test for Heteroskedasticity," Econometrica, Vol. 48, pp. 817-838.

Figure 1. Small Bank Recursive Residuals

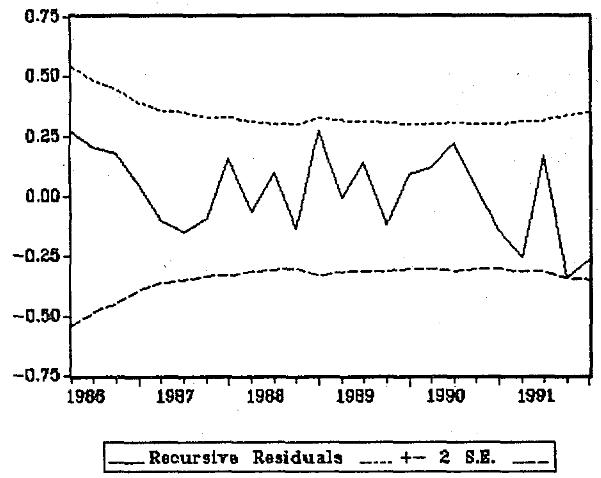


Figure 2. Small Bank Cusum of Recursive Residuals

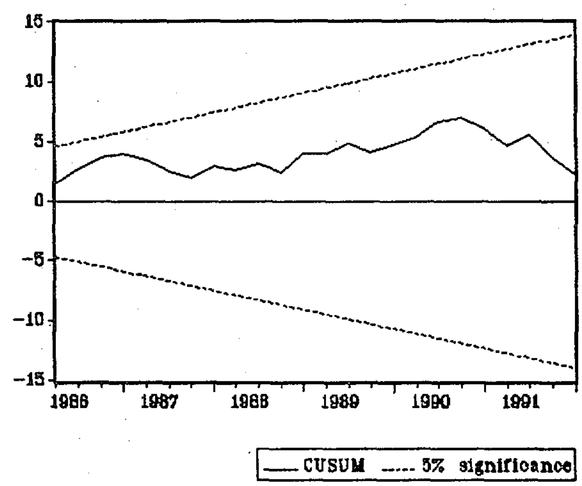


Figure 3. Small Bank Cusum Squared Recursive Residuals

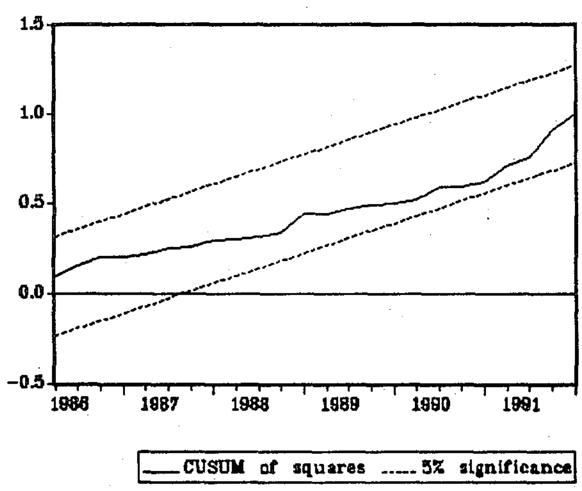


Figure 4. Large Bank Recursive Residuals

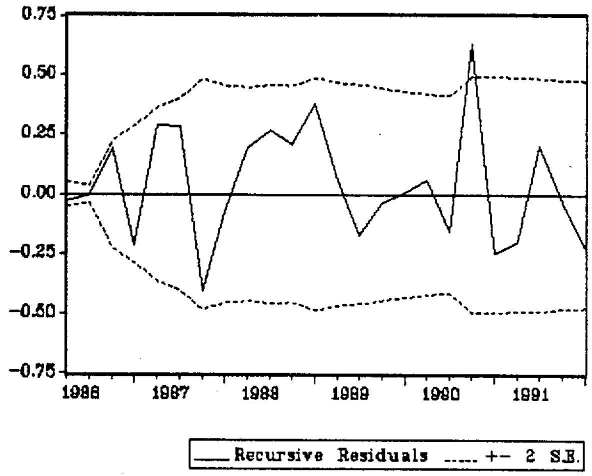


Figure 5. Large Bank Cusum of Recursive Residuals

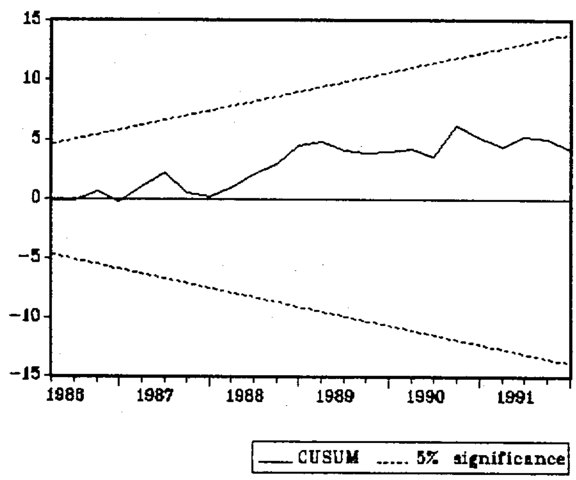
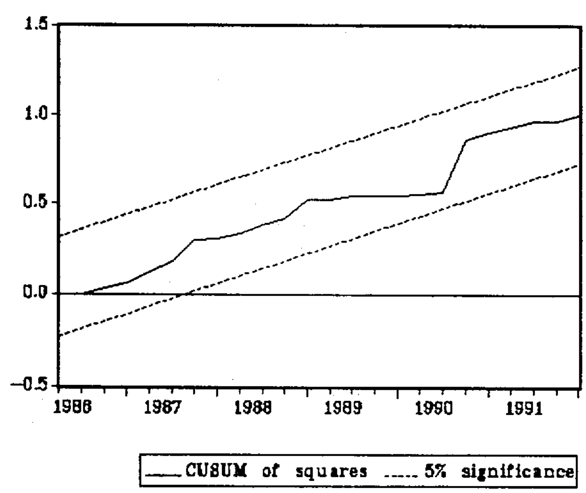


Figure 6. Large Bank Cusum Squared Recursive Residuals



The T-Test in Model Selection: An Empirical Study

David H. Richardson[1,2], Great Falls Analytics

1. Regression Model Selection

It is not likely that any particular linear regression specification will correspond exactly to the object of study. The initial specification is based on our subject matter knowledge together with experience and, at the outset, we may have a great deal to learn. Often

> "there is uncertainty as to the appropriate model to be used. As a consequence, investigators begin with an initial set of specifications and then modify their models by testing ..."
> [Bock *et. al.*, 1973, p. 109]

The need for a method of choosing among competing models by testing is, in part, due to the great quantity of data that is available.

> The data banks of the National Bureau of Economic Research contain time-series data on 2000 macroeconomic variables. Even if observations were available since the birth of Christ, the degrees of freedom in a model explaining gross national product in terms of all of these variables would not turn positive for another two decades.
> Leamer, 1983, p. 286

> We are overwhelmed by the sheer quantity of data and hence it is not surprising that
> Among the various possible sources of false statistical models, the omission or relevant explanatory variables or the inclusion of extraneous explanatory variables, are the most likely and pervasive.
> Judge *et. al.*, p. 854

However, economic theory can often indicate *a priori* the regressors that should be included and the signs of some of the coefficients. Use of this prior information results in better estimators in the classical as well as the Bayesian approach as shown, e.g., in Theil [1971, pp. 42-45] and Judge *et. al.* [1985, pp. 857-59]. "Hence, mechanical reliance on goodness of fit should be avoided as much as possible." [Amemiya, 1980, p. 331] Nevertheless, once the prior economic information has been incorporated, it is often necessary to choose a single regression specification and it would be helpful to be able to do so on the basis of a single statistic.
R^2, perhaps the first choice, has an obvious weakness since it can always be increased by increasing the number of regressors. Theil [1961, p. 213] proposed R-bar squared, R^2 adjusted for degrees of freedom: the higher R-bar squared, the better the specification. More recently, several statistics have been proposed on the basis of information theory, the most popular being Mallows' C_p, Akaike's AIC, and Amemiya's PC. Another possibility is to include all regressors, i.e., not select at all, as suggested by a strict interpretation of the maximum likelihood principle or by Bayesian inference with a noninformative, i.e., Jeffreys, prior. [See, e.g., Schwarz 1978, p. 461, Box and Tiao, 1973, pp. 25-60, Chow, 1981, pp. 30-33.] It must be remembered, however, that these decision rules have been proposed, to some extent, in an *ad hoc* manner, none of them are admissible, and their "sampling properties are virtually unknown." [Judge *et. al.*, 1985, p. 888] The purpose of this paper is to study empirically some of the sampling properties of these methods on the basis of the accuracy of out-of-sample forecasts (OOSF's).

The comparison of OOSF's is an activity for which there is a long-standing tradition in the forecasting literature. Mayer [1975] searched the literature for studies in which at least three models were estimated and for which, either in the original article or subsequently, OOSF results were presented. The best fitting models, by R-bar squared, within sample, and the best OOS, by either the standard error of forecast (SEF) or mean absolute error, were selected and compared. It was found that of the 13 cases in which there were three models, 8 times, 62%, the best fitting model was also best OOS. Of the 23 cases in which there were four models, only for 8, 35%, was the best fitting model also best OOS. Similarly, of the 35 cases in which there were five models, only for 12, 37% was the best fitting model also best OOS. Therefore "if one is interested in hypotheses that are valid beyond the sample period, goodness of fit statistics are a very poor guide." [Mayer, 1975, p. 882]

The information base utilized is the data and analysis generated by the 1990-91 Hedonic Regression Project [HRP] of the Consumer Prices Branch, Bureau of Labor Statistics, U. S. Dept. of Labor. The HRP involved the specification and estimation of 63 regression equations using the expert system SAMUEL. Since the specification procedure was the same over all of the models there was an opportunity to study the specification procedure itself. A report on the HRP is found in Section 2. The decision rules implied by R-bar squared, C_p, AIC and PC can be expressed as the minimum t-value or t-cutoff required to keep a regressor, as documented in Section 3 and Appendix A. The data were divided into two halves, and, using SAMUEL, a model was specified using each half. Each half was then forecast from the other and the SEF was computed for each of the decision rules considered, for both halves of the data. The decision rules were then compared on the basis of weighted and normalized mean SEF. A report of these results is found in Section 4. No difference in forecasting ability was found among models selected on the basis of these criteria, a result that is broadly consistent with that of Mayer [1975]. The question of the possible bias of conventional estimates of the standard error of estimate, SEE, is addressed in Section 5. The conclusions are presented in Section 6.

2. The Hedonic Regression Project

The Consumer Price Index (CPI) is calculated by pricing a market basket of consumer products each month, and comparing these prices to the previous month's prices. The prices of the products are gathered by BLS "shoppers" and every effort is made to continue to price the identical product. On occasion, however, products are discontinued and a decision must be made whether to substitute a comparable product or to terminate the series and start a new one based on the new product. Since price increases often occur with model changes, terminating one series and introducing another tends to bias the measured inflation rate downward, and hence it is important to substitute a comparable product if possible.

Since the products are not identical, part of the observed price difference is a result of changes in the attributes of the product. The purpose of the HRP was to establish a procedure which would save as many of the quotes as possible by estimating the effect of these changes on the price. To do this 63 regression equations were specified and estimated on the basis of data provided by the 19 BLS managers and analysts upon whom the success of the project depended. Two models were estimated for each product, one based on price as the dependent variable and one based on the log of the price. The independent variables were the attributes of the product. The regressions for all of the products were specified using the expert system SAMUEL developed by the author. Since the specification decisions were made automatically, the overall procedure was the same for all of the models. This mechanical repetition provided an opportunity to study the rules themselves.

Before any regressions were run there was a discussion between the analyst and the author to quantify the prior information. The prior included, *inter alia*, the enumeration of an initial set of independent variables and the t-cutoff. A prior sign, i.e., a constraint that a coefficient be positive or negative, was optionally specified for each of the independent variables. These signs were **not** intended to anticipate the data: the sole criterion for the prior sign was whether the opposite sign could possibly make sense with respect to the subject matter. For example, a longer as opposed to a shorter automobile battery warranty cannot, *per se*, be a disadvantage from the point of view of the consumer. Therefore, the sign on the coefficient of the length of the warranty was required to be positive, i.e., it cannot decrease the price. No consideration was given to any possible explanation for the opposite sign on the basis of the correlation of the warranty length with other variables in the model.

SAMUEL eliminates variables sequentially until a specification that meets all of the relevant criteria has been determined. While SAMUEL looks for the best model on the basis of the prior and the data, it is not feasible to run all possible regressions. Rather, SAMUEL determines a model that satisfies all of the explicit criteria and fits the data as well as possible on the basis of the given sequence of decisions. The decision-making process itself has been optimized to provide as good a model as possible on one pass. Incorrect signs are considered more harmful than low t-statistics and hence the signs are dealt with first. The most deleterious incorrect sign is deleted at each step until the signs on all of the coefficients are consistent with the prior. Once all of the signs are correct, variables with low t-statistics are deleted until all of the absolute t-statistics are larger than t-cutoff. As might be expected, the estimated coefficients tend to vary most in the first few steps, and then gradually converge to their final values. The results then go back to the analyst. If anomalies were discovered, further runs were made until the analyst, SAMUEL, and the author were all satisfied.

As an example of the results of the HRP, the specified and estimated model for automobile batteries is reported in Exhibit 2.1. The output includes a standard regression printout followed by a report on the normalized residuals, the absolute residuals divided by the SEE, larger than 2. There is then a report on large partial correlations among the independent variables, if any, and, if the dependent variable is logged, a computation of R^2 on the basis of the original data. In this case the log of the price, LPRICE, was regressed against several variables including the length of the warranty, WARRANT, and dummies for the trade of the old battery, TRADEIN, and terminals on the side rather than the top of the battery, SITERM. The price of the battery increases by about .55% per month of the warranty, decreases by about 2.32% for a trade-in, and increases by about 6.97% for side terminals. Therefore, if the warranty were extended from 48 months to 60 months we would expect the price to rise by about 6.60%. If the price increased more than this we would measure inflation, while if the increase were less, we would measure deflation.

3. The Theoretical Basis of the t-cutoffs

We must choose between nested regression models on the basis of the criteria mentioned in Section 1. As shown in Appendix A, the selection problem reduces to the determination of the level of significance or the cutoff value of an F test. Since SAMUEL makes specification decisions sequentially, one variable at a time, the F-statistic has one degree of freedom and is equivalent to the square of a t-statistic. Classical hypothesis testing has lead researchers to choose the t-cutoff at the 95% confidence level, $|t| = 2$, approximately. On the basis of a strict interpretation of the maximum likelihood principle or Bayesian Inference with a Jeffreys prior on the regressors we would include the largest possible model, i.e., there would be no selection, and $|t| = 0$. As is well known R-bar squared is maximized by deleting all variables with $|t| < 1$.

Mallows, not quite in the spirit of Information Theory, suggests not minimizing C_p, but rather choosing a specification in which C_p is "small." As shown in Judge *et. al.*, [1985, pp. 867,868] this rule leads to the t-cutoff $|t|=1$. Extending this development and minimizing C_p, the t-cutoff is found to be the square root of 2, i.e., $|t|$ =sqrt2. With respect to PC, Amemiya [1980, p. 348] finds that the unrestricted model is superior if $F_{PC} > 2T/(T+NRHS-1)$ where T is the number of observations and NRHS the number of regressors in the unrestricted model. If T increases without bound with NRHS constant, this converges to $F_{PC}>2$ or $|t|$ >sqrt2. In terms of this study the small difference between F_{PC} and sqrt2 is inconsequential, and the t-cutoff for PC was set equal to sqrt2. Amemiya [1980, pp. 340-44] showed that Akaike's AIC is equivalent either to PC or C_p minimized, depending on how the variance of the observations is estimated. Hence maximizing information by all three approaches leads to the t-cutoff $|t|$ =sqrt2, approximately. The criteria proposed by Shibata [1981] and Breiman and Freedman [1983] are asymptotically equivalent to C_p, PC, and AIC. Therefore the four t-cutoffs, 0, 1, sqrt2, and 2, were used to generate the forecasting models studied.

A complete analysis was carried out for the data in two parts, the odd and the even observations. The first step was to use SAMUEL to specify a base model, in which all the signs were correct. Four models were then specified and estimated, one for each of the four t-cutoffs respectively. The dependent variable was either price or its log, whichever fit the data better, when the model was specified and estimated for the HRP. The odds were then forecast from the evens, and *vice versa*, and the SEF was computed for each of the four t-cutoffs, for both halves of the data. An attempt was made to analyze all of the products studied in the HRP. One, miscellaneous fruits, was eliminated because there was too much data to deal with conveniently; and five were eliminated because, in the original specification, NRHS was larger than half the sample size and a regression based on half the data could not be run. A summary of the models specified and estimated is found in Appendix B.

4. The Results

The SEF was computed for all of the specified and estimated models as a function of t-cutoff. A regression of WSEF, SEF weighted by the number of forecasts and multiplied by 100, on dummies for the four different t-cutoffs generated the results reported in Exhibit 4.1. The coefficients in Exhibit 4.1 are proportional to the weighted average of the SEF for each of the four t-cutoffs. Normalizing by dividing by the coefficient for the 0.0 t-cutoff, 170.8558, yields the relative SEF, RSEF, for each t-cutoff as reported in Exhibit 4.2.

For all practical purposes, there is no difference among the SEF's in Exhibits 4.1 and 4.2. A regression of WSEF on a constant together with dummies for $t=1$, t=sqrt2, and $t=2$ is presented in Exhibit 4.3. There is no statistically significant difference between any of them and the constant, while the negative R-bar squared indicates that the three dummies, as a group, are not significant at the usual levels. This is broadly consistent with the results of Mayer [1975]. Therefore, we have the (perhaps) surprising conclusion that the coefficients of the t-cutoffs are not significantly different from each other at the usual levels, and the t-cutoff has, for all practical purposes, no effect on the forecasting accuracy of the resulting equation.

One reason for this is that any procedure which omits variables is not Bayes and hence is inadmissible. [Cohen, 1965] In fact under quadratic loss the Bayes estimator of the vector of regressions coefficients is just the posterior mean, as noted by Leamer [1979, p. 508]. Ordinarily the posterior mean will have no zero elements and hence Leamer suggests that while inference must be based on the posterior mean, interesting summaries of the results may be presented in terms of restricted coefficient vectors. Leamer also notes that the information criteria are all couched in terms of the true parameters and the need to estimate some of the parameters completely eliminates the potential gain. The Bayes estimator, however, is not superior but merely equal to the information based estimators in the HRP sample. The reason for the lack of superiority may be related to the (often) very large number of variables and the inclusion of variables for which a reasonable prior is concentrated near the origin.

5. Bias in the SEE

If the specified model is identical to the one which generated the data, the SEE will be an unbiased estimator of the SEF. Since deleting variables when $|t|<1$ reduces the SEE but not necessarily the SEF, we may expect the specification process to introduce a downward bias in the SEE as an estimator of the SEF. Conversely, if $|t|>1$, an upward bias will be introduced. Therefore, it is of interest to study log(bias) = log(SEF/SEE), the log of the resulting bias, as a function of the other characteristics of the model. With a little manual specification effort, as reported in Appendix C, the model reported in Exhibit 5.1 was specified and estimated. As shown, the bias increases quite rapidly with t-cutoff, from $e^{2.70}=14.8$ at $|t|=0$ to $e^{3.55}=34.9$ at $|t|=2$, a ratio of 2.36:1.

The bias decreases substantially with DF, however. Summing the effect of the three DF coefficients, we find that log(bias) decreases from -2.27 when DF = 10 to -6.08 when DF = 600, a ratio of .022:1. The effect of increasing DF is reported in detail in Exhibit 5.2. Increasing NRHS, however, increases log(bias) very rapidly. Summing the two NRHS coefficients we find that log(bias) increases from .04 at NRHS = 1 to 5.93 at NRHS = 100, a ratio of 361.8:1, as reported in Exhibit 5.3.

While in the case of a known, correctly specified, model SEE is unbiased for SEF, this is clearly not true in the case of models specified on the basis of the data. Here there is substantial bias, a bias which increases rapidly with the t-cutoff and the number of regressors. Since, however, increasing the t-cutoff usually reduces the number of regressors, it is not clear which effect will dominate in any particular case. Increasing the sample size unambiguously reduces the bias.

These results broadly confirm those of Makridakis and Winkler [1989, pp. 331-332]: "Emphatically, however, post-sample errors do not behave as theory suggests." Using the data generated by the M-competition for the eight best (in terms of accuracy) forecasting methods one period ahead they found that the range of the bias ratio (not logged) was "1.27 to 2.03 for yearly data, 1.34 to 6.35 for quarterly data, and 1.04 to 1.65 for monthly data." [Makridakis and Winkler, 1989, p. 336] The reasons suggested for this are structural change in the model and overfitting. Since the HRP data is based on a sample at a single point in time it may be hypothesized that overfitting, here in the sense of including just too many variables, is the cause of most of the observed inflation in the bias. This hypothesis is troubling, however, since the degrees of freedom correction is designed to deal with just this problem.

6. Conclusion

Several different model specification methods have been evaluated on the basis of their OOSF's using the data and analysis generated by the 1990-91 HRP of the Consumer Prices Branch, BLS, which involved the specification and estimation of 63 regression equations using the expert system SAMUEL. Since the specification procedure was the same over all of the models the rules themselves could be studied. The model specification methods considered were the selection of significant regressors, Theil's R-bar squared, Mallows' C_p, Akaike's AIC, Amemiya's PC, and a strict interpretation of the maximum likelihood principle. It was determined that the decision rules of these methods could be expressed as selection on the basis of minimum t-cutoffs, i.e., the values 0, 1, sqrt2, and 2. The data were divided into two halves and, using SAMUEL, a model was specified and estimated on the basis of each half. Each half was then forecast from the other and the standard error of forecast was computed for each of the decision rules and for both halves of the data. The decision rules were then compared on the basis of the weighted, normalized, mean standard error of forecast. Surprisingly, it was found that there was no difference in forecasting ability among the models based on the different methods and their respective t-cutoffs.

The practice of "data mining," running large numbers of regressions in an attempt to get an equation with significant coefficients, has generated serious concern. Equations thus derived are the basis of research reports and are used in forecasting. When published, these reports have the effect of filling the literature with Type I errors, positive results that arise by chance, a disconcerting situation. [Lovell, 1983] With respect to models used for forecasting, however, on the basis of our results, the data miners have been wasting their time: modeling by selecting the variables on the basis of t-tests results in forecasts that are neither better nor worse than those generated by models specified without such selection.

The data were also used to comment on the possible bias of conventional estimates of the standard error of estimate. While in the case of a known model the standard error of estimate is unbiased for the standard error of forecast, this is not true in the case of models specified on the basis of the data. Here there is substantial bias, a bias which increases rapidly with the t-cutoff and the number of regressors. Since, however, a higher t-cutoff results in fewer regressors, it is not clear which effect will predominate in any particular case. Increasing the sample size unambiguously reduces the bias.

Footnotes

1. Great Falls Analytics, 1629 Columbia Rd., NW, Suite 704, Washington DC. 20009.

2. This work was partially supported by the Bureau of Labor Statistics, U.S. Dept. of Labor, Consumer Proces Branch, Walter Lane, Chief, under purchase order B9J04797. The Project Manager was David Barry of the Consumer Prices Branch, BLS. The Consumer Prices Branch analysts, on whose work the project was totally dependent, were Lisa Alexander, Royce Buzzell, Joseph Chalena, Karen Flynn-Huff, Charles Fortuna, Carrick Herbert, Christine Jackson, Yvette Jackson, Jeffrey Kellar, Timothy LaFleur, Pau; Liegey, Sylvia Norman, Michael Reese, Tod Reese, Bryan Richardson, Jon Schomaker and Alexander Talmor. All of this assistance is gratefully acknowledged.

References

Akaike, H., 1974, "A New Look at the Statistical Identification Model." IEEE: Trans. Auto. Control, 19, 716-23.
-----, 1981, "Likelihood of a Model and Information Criteria." J. of Econometrics, 16, 3-14.

Amemiya, T., 1980, "Selection of Regressors." International Economic Review, 21, No. 2, 331-54.

Bock, M. E., T. A. Yancey and G. G. Judge, 1973, "The Statistical Consequences of Preliminary Test Estimators in Regression." J. of the American Statistical Assn., 68, 109-16.

Box, G. E. P. and G. C. Tiao, 1973, Bayesian Inference in Statistical Analysis. Reading, Mass.: Addison Wesley Publishing Co.

References cont.

Breiman, L. and D. Freedman, 1983, "How Many Variables should be Entered in a Regression Equation?" J. of the American Statistical Assn., **78**, 131-36.

Chow, G., 1981, "A comparison of Information and Posterior Probability Criteria for Model Selection." J. of Econometrics, **16**, 21-33.

Cohen, A. , 1965, "Estimates of Linear Combinations of the Parameters in the Mean Vector of a Multivariate Distribution." Annals of Mathematical Statistics, **36**, 78-87.

Judge, G. G., Griffiths, W. E., Hill, R. C., Lutkepohl, H., and Lee, T. C., 1985, The Theory and Practice of Econometrics. New York: John Wiley & Sons.

Kennedy, W. J. and T. A. Bancroft, 1971, "Model Building for Prediction in Regression Based on Repeated Significance Tests." Annals of Mathematical Statistics, **42**, 1273-84.

Leamer, E. E., 1979, "Information Criteria for Choice of Regression Models." Econometrica, **47**, 507-510.
-----, 1983, "Model Choice and Specification Analysis" in Griliches, Z. and M. D. Intriligator, eds., Handbook of Econometrics, **1**, p. 286. Amsterdam: North-Holland Publishing Co.

Lovell, M. C., 1983, "Data Mining." The Review of Economics and Statistics, **LXV**, 1-12.

Mallows, C. L., 1973, "Some Comments on C_p." Technometrics, **15**, 661-76.

Makridakis, S. and R. L. Winkler, 1989, "Sampling Distributions of Post-Sample Forecasting Errors." Applied Statistics, 38, 331-342.

Mayer, T., 1975, "Selecting Hypotheses by Goodness of Fit." The Economic Journal, **85**, 877-83.

Schwarz, G., 1978, "Estimating the Dimension of a Model." Annals of Statistics, **6**, 461-464.

Shibata, R., 1981, "An Optimal Selection of Regression Variables." Biometrika, **68**, 45-54.

Theil, H., 1961, Economic Forecasts and Policy. Amsterdam: North-Holland Publishing Co.
-----, 1971, Principles of Econometrics. New York: John Wiley & Sons.

Appendix A. The Variable Selection Criteria

Assume that there are two models, indexed by $i=0,1$, with $K+i$ independent variables, respectively. Each of the selection criteria of Section 3 can be expressed in terms of an F(1,T-K-1) variable, the square of a t(T-K-1) variable. The correctly specified regression model is

$$y = XB_i + u$$

where

y is the T-vector of observations on the dependent variable;
u is the T-vector of independent, identically distributed (iid) disturbances;
B_i is the true $K+i$-vector of coefficients;
X_i is the $T\times(K+i)$ matrix of the independent variables;
$b_i = (X_i'X_i)^{-1}X_i'y$ is the $K+i$-vector of least squares coefficients;

V(u) is the variance of u; and
$s_i^2 = y'[I-(X_i'X_i)^{-1}X]y/(T-K-i)$ is the least squares estimate of V(u).

The problem is to determine whether to include variable $K+1$ in the model, i.e., whether $i=0$ or $i=1$ is optimal.

It is well known that if we choose i such that the standard error of estimate is minimized we include variables if $t \geq 1$ but not otherwise. To include the $K+1^{th}$ regressor

$$s_0^2 - s_1^2 = [(T-K)s_0^2 - (T-K-1)s_1^2] - s_1^2 \geq 0$$

$$(T-K)s_0^2 - (T-K-1)s_1^2 \geq s_1^2$$

$$[(T-K)s_0^2 - (T-K-1)s_1^2]/s_1^2 = t^2 \geq 1$$

and $i=1$ is preferred if $|t| \geq 1$.

Mallows' C_p minimizes the risk function

$$R(b_i) = E(b_i-B_i)'X_i'X_i(b_i-B_i)$$

for $i = 0,1$. This results in

$$C_p = (T-K)s^2 + (2K-T)V(u)$$

Estimating V(u) by s_1^2 we have

$$C_{p0} = (T-K)s_0^2 + (2K-T)s_1^2$$

and

$$C_{p1} = (T-K-1)s_1^2 + (2K+2-T)s_1^2 = (K+1)s_1^2$$

Minimizing C_p over i

$$C_{p0} - C_{p1} = (T-K)s_0^2 + (K-1-T)s_1^2 \geq 0$$

$$(T-K)s_0^2 - (T-K-1)s_1^2 \geq 2s_1^2$$

$$[(T-K)s_0^2 - (T-K-1)s_1^2]/s_1^2 = t^2 \geq 2$$

and $i=1$ is preferred if $|t| \geq \text{sqrt}(2)$.

Amemiya's PC is intended to minimize the expected mean square error of Forecast (EMSEF).

$$EMSEF = E(y_{T+1}-y^*_{T+1})^2 = s_i^2[1 + x_{i,T+1}(X_i'X_i)^{-1}x'_{i,T+1}]$$

over i = 0,1. Here x'_{T+1} is the is the K+i-vector of the first OOS observation on the independent variables, y_{T+1} is the corresponding value for the dependent, and y^*_{T+1} is its forecast.
Taking the expectation yields

$$PC_1 = (T+K+1)s_1^2/T$$

and

$$PC_0 = (T+K)s_0^2/T$$

If i=1

$$PC_0 - PC_1 = [(T+K)s_0^2 - (T+K+1)s_1^2]/T \geq 0$$

$$(T-K)s_0^2 - (T-K)(T+K+1)s_1^2/(T+K) \geq 0$$

$$(T-K)s_0^2 - (T-K-1)s_1^2 \geq 2Ts_1^2/(T+K)$$

$$t^2 \geq 2T/(T+K)$$

If T increases without bound with K constant, this converges to $t^2>2$ or $|t|>$sqrt2. In terms of this study the small difference between F_{PC} and sqrt2 is of no consequence, and, for PC, the t-cutoff was set at sqrt2.

Appendix B. THE FORECASTING EXPERIMENT

The results of the forecasting experiment are reported in Exhibit B.1. There are three data lines for each product followed by a blank line. The first line in each group is a heading which begins with the ELI number, the code by which the components of the CPI are organized. This is followed by the common name of the product and the forecast sample size for the odds and the evens respectively. The following two lines report NRHS, t-cutoff, and SEF. To conserve space the odd and even results have been averaged.

Appendix C. The Specification of Log(Bias)

Two regressions were run to specify the model for log(bias) reported in Exhibit 5.1. The first specification included dummies for the four t-cutoffs together with the linear, square, and log transformations of DF and NRHS. The least satisfactory variable in Exhibit C.1 is NRHS since $|t| = .72$. Deleting it results in the regression shown in Exhibit 5.1, the specified model for log(bias).

Exhibit 2.1
Automobile Batteries

Variable	Coefficient	Std. Error	t-statistic
Constant	3.573	.0594	60.2
Tradein	-.0232	.0069	-3.35
Warrant	.0055	.0009	5.80
Siterm	.0697	.0316	2.20
Brand1	-.1634	.0444	-3.68
Brand2	.1950	.0579	3.37
Brand3	.3067	.0606	5.06
Brand4	.4287	.0944	4.54
Brand5	.1531	.0661	2.32
Small City	-.1196	.0481	-2.49
Retail1	.1724	.0501	3.44
Retail2	.2746	.0462	5.94
East	.1017	.0385	2.64
Midwest	.1470	.0356	4.13

Obs. = 125 SEE = .157 R^2 = .662 Adjusted R^2 = .623

Large Residual/SEE

Obs.	Outlet	Quote	Normalized Residual
11	.647E+06	2.0000	2.40
70	.108E+07	4.0000	-2.94

There are 0 outliers larger than 3 in absolute value.

Cross-Correlations Larger than .60

Corr	Var1	Var2
-.86094	Warrant	Constant

In the Original Metric, R^2 = .688

Exhibit 4.1
WSEF on Dummies for the t-cutoffs

t-cutoff	Coefficient	Stand. Error	t-statistic
.0000	170.86	10.7	16.0
1.000	170.38	10.7	16.0
1.414	170.37	10.7	16.0
2.000	170.13	10.7	16.0

Obs. = 416 SEE = 108.6 R^2 = .000 Adjusted R^2 = -.007

Exhibit 4.2
Relative SEF's for the t-cutoffs

t-cutoff	RSEF
.0000	1.00
1.000	.997
1.414	.997
2.000	.996

Exhibit 4.3
WSEF on Dummies for t-cutoffs with Constant

Variable	Coefficient	Stand. Error	t-statistic
Constant	170.	10.7	16.0
t-cutoff=1.000	.485	15.1	.032
t-cutoff=1.414	.009	15.1	.001
t-cutoff=2.000	-.239	15.1	-.016

Obs. = 416 SEE = 108.6 R^2 = .000 Adjusted R^2 = -.007

Exhibit 5.1
The Determinants of Log(Bias)

Variable	Coefficient	Stand. Error	t-statistic
t-cutoff=.0000	2.70	.5507	4.90
t-cutoff=1.000	3.30	.5494	6.00
t-cutoff=1.414	3.46	.5409	6.40
t-cutoff=2.000	3.55	.5310	6.69
NRHS Squared	.000410	.0000922	4.45
Log of NRHS	.397	.09617	4.13
Deg. Freedom	.00979	.00272	3.59
Deg. Free. Squared	-.0000149	.0000045	-3.29
Log of Deg. Free.	-1.03	.151	-6.80

Obs. = 208   SEE = .397 R^2 = .638 Adjusted R^2 = .624

Exhibit 5.2
Effect of Degrees of Freedom on Log(Bias)

DF	Log(Bias)	DF	Log(Bias)	DF	Log(Bias)
10	-2.27	90	-3.87	300	-4.27
20	-2.89	100	-3.90	350	-4.43
30	-3.22	120	-3.96	400	-4.63
40	-3.42	140	-4.00	450	-4.90
50	-3.57	160	-4.03	500	-5.23
60	-3.68	180	-4.06	550	-5.62
70	-3.76	200	-4.09	600	-6.08
80	-3.82	250	-4.16		

Exhibit 5.3
The Effect of the Number of Regressors on Log(Bias)

NRHS	Log(Bias)	NRHS	Log(Bias)	NRHS	Log(Bias)
1	.041	14	1.13	40	2.12
2	.277	15	1.17	45	2.34
3	.440	16	1.21	50	2.58
4	.557	17	1.24	55	2.83
5	.649	18	1.28	60	3.10
6	.726	19	1.37	65	3.39
7	.793	20	1.35	70	3.70
8	.852	22	1.43	75	4.02
9	.905	24	1.50	80	4.37
10	.955	26	1.57	85	4.73
11	1.00	28	1.64	90	5.11
12	1.05	30	1.72	95	5.51
13	1.09	35	1.91	100	5.93

Exhibit B.1.
Basic Forecasting Results

ELI NRHS	PRODUCT SEF	ODD OBS t-cutoff	EVEN OBS NRHS	SEF	t-cutoff
03011	GROUND BEEF		245	245	
19.5	.530	.0000	17	.535	1.000
13.5	.530	1.414	11	.531	2.000
03041	OTHER ROAST		135	134	
11.5	1.20	.0000	8	1.21	1.000
7.5	1.21	1.414	5.5	1.23	2.000
03042	STEAK	371	371		
24	.170	.0000	19.5	.170	1.000
18	.170	1.414	17.5	.170	2.000
03043	RIBS	176	176		
34	.760	.0000	19	.752	1.000
14.5	.764	1.414	12.5	.751	2.000
03043	VEAL	43	42		
18	.505	.0000	12	.506	1.000
11	.469	1.414	8	.404	2.000
04011	BACON	249	249		
18	.627	.0000	14	.631	1.000
12	.633	1.414	9	.649	2.000
04021	PORK CHOPS		237	236	
22.5	.580	.0000	17	.582	1.000
14.5	.587	1.414	11.5	.584	2.000
04032	HAM	43	42		
14	1.76	.0000	8.5	1.65	1.000
4.5	1.65	1.414	4	1.67	2.000
04042	SAUSAGE	211	211		
23	.589	.0000	13.5	.592	1.000
11	.590	1.414	9	.593	2.000
05011	HOT DOGS	102	101		
24	.217	.0000	16.5	.222	1.000
14	.223	1.414	12.5	.229	2.000
05012	BOLOGNA SALAMI		131	130	
20	.302	.0000	12.5	.300	1.000
8.5	.309	1.414	5	.314	2.000
05013	OTHR LNCHMEAT		243	242	
24.5	.877	.0000	16	.880	1.000
14.5	.887	1.414	10	.904	2.000
08011	EGGS	259	259		
22	.229	.0000	16.5	.229	1.000
15.5	.229	1.414	13.5	.234	2.000
10021	CHEESE	266	265		
51	.596	.0000	28.5	.600	1.000
22.5	.592	1.414	17	.599	2.000

ELI NRHS	PRODUCT SEF	ODD OBS t-cutoff	EVEN OBS NRHS	SEF	t-cutoff
10041	ICE CREAM	259	259		
34	.324	.0000	25.5	.318	1.000
20	.314	1.414	16.5	.310	2.000
11011	APPLES	345	345		
18.5	.351	.0000	12.5	.349	1.000
9	.344	1.414	6	.337	2.000
11031	ORANGES	384	383		
18	.668	.0000	13	.668	1.000
7.5	.670	1.414	6.5	.672	2.000
12011	POTATOES	262	262		
24.5	.359	.0000	18	.357	1.000
15.5	.358	1.414	11	.353	2.000
12021	LETTUCE	246	245		
20	.237	.0000	13	.240	1.000
11	.241	1.414	7.5	.238	2.000
12031	TOMATOES	262	262		
20.5	.264	.0000	11.5	.265	1.000
9.5	.269	1.414	6	.267	2.000
13013	FRUIT JUICE	282	282		
35	.616	.0000	21	.612	1.000
15	.609	1.414	8.5	.599	2.000
13031	CANNED FRUIT	172	172		
53	.330	.0000	26.5	.332	1.000
20	.338	1.414	18	.347	2.000
16014	PEANUT BUTTER	96	96		
17.5	.495	.0000	11.5	.499	1.000
8.5	.484	1.414	7.5	.491	2.000
17012	SOFT DRINKS		75	74	
28	.729	.0000	14.5	.750	1.000
10.5	.753	1.414	9	.734	2.000
18022	FROZEN DISH		85	84	
28	.320	.0000	16.5	.341	1.000
13.5	.339	1.414	9	.328	2.000
18031	SNACK FOODS		216	215	
38.5	.541	.0000	26	.538	1.000
21	.533	1.414	10.5	.509	2.000
20011	BEER AT_HOME	231	230		
30	2.49	.0000	19.5	2.48	1.000
15	2.47	1.414	13	2.48	2.000
20021	WHISKEY HOME	125	125		
20	.383	.0000	13	.386	1.000
9.5	.384	1.414	6.5	.374	2.000
20022	OTHER LIQUOR	108	108		
23	5.14	.0000	14.5	5.19	1.000
8.5	4.95	1.414	7.5	4.97	2.000

ELI NRHS	PRODUCT SEF	ODD OBS t-cutoff	EVEN OBS NRHS	SEF	t-cutoff
20031	WINE AT_HOME	242	242		
54.5	7.91	.0000	17.5	7.80	1.000
10	7.71	1.414	7.5	7.63	2.000
20051	BEER NOTHOME		59	59	
17	.972	.0000	11	.938	1.000
10.5	.953	1.414	8	.929	2.000
21031	STDNT HOUSING	95	95		
35.5	688.	.0000	17.5	693.	1.000
15.5	711.	1.414	8.5	720.	2.000
29021	SOFAS	49	48		
36.5	794.	.0000	21.5	794.	1.000
18.5	859.	1.414	11	902.	2.000
31011	TELEVISION	134	133		
39	.233	.0000	22	.237	1.000
15.5	.244	1.414	12.5	.247	2.000
31033	RECORDED MUSIC	37	36		
19	4.68	.0000	9	4.80	1.000
8	4.74	1.414	3.5	4.96	2.000
36041	MEN'S SHIRTS	399	398		
26	.378	.0000	19.5	.377	1.000
17	.375	1.414	12.5	.379	2.000
37013	BOY'S SHIRTS	127	127		
19.5	.323	.0000	12	.322	1.000
10	.329	1.414	9.5	.329	2.000
37016	BOY'S OUTRWEAR	88	88		
19	.431	.0000	11.5	.430	1.000
7	.436	1.414	6.5	.433	2.000
38011	WMN'S OUTRWEAR	423		423	
34	.495	.0000	30	.497	1.000
28.5	.506	1.414	23.5	.520	2.000
45031	MOTOR CYCLES		94	94	
37	.242	.0000	23.5	.231	1.000
15	.234	1.414	9.5	.230	2.000
46011	USED CARS	166	166		
59	834.	.0000	50.5	825.	1.000
48.5	823.	1.414	47	830.	2.000
47012	LEADED GAS		54	54	
24.5	.0843	.0000	13	.0737	1.000
10.5	.0718	1.414	5	.0594	2.000
47014	REGULAR GAS		280	280	
34.5	.0886	.0000	16.5	.0886	1.000
13.5	.0886	1.414	10.5	.0886	2.000
47015	MIDGRADE GAS		62	62	
27.5	.119	.0000	15	.112	1.000
12.5	.109	1.414	6.5	.111	2.000

ELI NRHS	PRODUCT SEF	ODD OBS t-cutoff	EVEN OBS NRHS	SEF	t-cutoff
47016	PREMIUM GAS		278	278	
33.5	.111	.0000	17	.110	1.000
13	.110	1.414	10	.110	2.000
48021	AUTO BATTERIES	63	63		
29.5	15.1	.0000	19.5	14.5	1.000
13	13.5	1.414	10	13.9	2.000
52051	RENTAL CARS		42	42	
18	116.	.0000	8.5	97.8	1.000
7	97.8	1.414	4	98.8	2.000
53011	AIRLINE FARES	158	157		
31.5	157.	.0000	21.5	153.	1.000
16.5	155.	1.414	12.5	157.	2.000
53023	OCEAN CRUISES	56	56		
22.5	2034.	.0000	9.5	1967.	1.000
7	1920.	1.414	6	1904.	2.000
56031	EYE GLASSES		50	49	
36.5	69.1	.0000	23.5	66.6	1.000
17.5	70.1	1.414	12.5	70.3	2.000
67041	TRADE SCHOOLS	38	38		
33.5	4680.	.0000	28	4826.	1.000
24.5	3999.	1.414	19.5	3045.	2.000
68031	FUNERAL SVCS		42	41	
20	.412	.0000	11	.382	1.000
8.5	.382	1.414	5	.354	2.000

Exhibit C.1
The Determinants of Log(Bias)

Variable	Coefficient	Stand. Err.	t-stat.
t-cutoff = .0000	2.5226	.601173	4.20
t-cutoff = 1.000	3.1084	.608540	5.11
t-cutoff = 1.414	3.2711	.603403	5.42
t-cutoff = 2.000	3.3637	.592445	5.68
NRHS	-.0251	.034766	-.72
NRHS Squared	.00069	.000402	1.72
Log of NRHS	.59939	.296412	2.02
Deg. Freedom	.00982	.002726	3.60
Deg. Free. Squared	-.00001	.000004	-3.26
Log of Deg. Free.	-1.0383	.151928	-6.83

Obs. = 208 SEE = .397 R^2 = .639 Adjusted R^2 = .623

The United Nation's Millennium Project

Jerry Glenn, The United Nations University
The United Nations University, an autonomous organ of the UN, is examining the feasibility of organizing futures research to continuously up-date and improve humanity's thinking about the future and make it available for public education and feedback. In collaboration with the Smithsonian Institution, The Futures Group, and the EPA, the UNU is linking futurists, scholars, and institutions around the world to create an international information system of forecasts, key questions, lessons from history, and potential futures research agendas. Called the "Millennium Project," it will also evaluate futures research methodology and the potential for setting standards, along with integrating forecasts to describe potential futures, and propose policy choices.

Differing Forecasting Styles Between Economists and Futurists

Robert L. Olson and Jonathan Peck, Institute for Alternate Futures
Forecasting in Federal government agencies is usually based on the assumptions and forecasting styles of economics. Most people involved in futures research, however, believe that the model of economic forecasting is too confining. While economists forecast from past data and extrapolate trends, futurists construct alternative scenarios that explore broader possibilities of change in technology, the economy and society. While economists generally adopt a short-term perspective, futurists take a long-term perspective on inter-generational costs, benefits and trade-offs. Additionally, futurists favor a more participatory model in which people work together to clarify "Preferred Futures" and the strategies to achieve them. The largest improvements in Federal government forecasting will come from adopting a broader and more powerful array of forecasting tools, treating economic methods as one useful approach among many.

Future Studies and Sustainable Development

Donald R. Lesh and Diane G. Lowrie, Global Tomorrow Coalition
Concerns for the future only become meaningful to senior decision makers--and thus begin to drive bureaucratic, organizational, corporate, educational, and civil structures--when present realities suggest unacceptable levels of threat, or that such threats may arise. Otherwise, future studies, projections, scenarios, and models are viewed as abstractions only distantly related to current decision-making. Research centers, individual analysts, nongovernmental organizations, and communication media can play a direct role in the linkages between future studies and current decision making, through (a) alerting leaders and the public to potential future threats; (b) building understanding of the scientific basis for concern; (c) suggesting policy courses and actions to mitigate or avoid future threats; (d) encouraging public demand for timely response. As an adjunct to the decision making process, the Global Tomorrow Coalition employs a variety of tools and forums in efforts to improve the quality of foresight and leadership toward the goal of a more sustainable development.

Error Measures for Generalizing About Forecasting Methods: Empirical Comparisons

J. Scott Armstrong, The Wharton School
Fred Collopy, Case Western Reserve University
This study evaluated measures for making comparisons of errors across time series. We analyzed 90 annual and 101 quarterly economic time series. We judged error measures on reliability, construct validity, sensitivity to small changes, protection against outliers, and their relation to decision making. The results lead us to recommend the Geometric Mean of the Relative Absolute Error (GMRAE) when the task involves calibrating a model for a set of time series. The GMRE compares the absolute error of a given method to that from the random walk forecast. For selecting the most accurate methods, we recommend the Median RAE (MdRAE) when few series are available and the Median Absolute Percentage Error (MdAPE) otherwise. The Root Mean Square Error (RMSE) is not reliable, and is therefore inappropriate for comparing accuracy across series.

Evaluating the 1980-90 Occupational Projections

Neal Rosenthal, Division of Occupational Outlook, Bureau of Labor Statistics
The discussion will focus on the lessons learned from the evaluation of the 1980-90 BLS occupational projections. Information will also be presented on the problems encountered in conducting the evaluation. Date on 1980 and 1990 actual employment and projected 1990 employment will be available to the audience.

☆U.S. GOVERNMENT PRINTING OFFICE:1993 -341-937/82632