Federal Reserve Bank of Minneapolis
Research Department Working Paper

FORECASTING WITH BAYESIAN VECTOR AUTOREGRESSIONS--
FOURS YEAR OF EXPERIENCE

Robert B. Litterman

Working Paper 259
PACS File 3125

August 1984

NOT FOR DISTRIBUTION
WITHOUT AUTHOR'S APPROVAL

Forecasting with Bayesian Vector Autoregressions--
Fours Years of Experience

Introduction

Forecasting the economy is a risky, often humbling task. Unfortunately, it is a job that many statisticians, economists, and others are required to engage in. This paper describes a technique which can make this task easier, and appears likely to produce improved results.

The technique, economic forecasting with Bayesian vector autoregressions (BVAR), over the past several years has proved to be an attractive alternative in many situations to the use of traditional econometric models or to other time series techniques. The BVAR models are relatively simple to use, inexpensive, and generate forecasts which have been as accurate, on average, as several of the most expensive forecasts currently available.

Moreover, relative to the widely used macroeconometric models, the BVAR approach has a distinct advantage in two respects. First, and most important, it does not require judgemental adjustment. Thus, it is a scientific method which can be evaluated on its own, without reference to the forecaster running the model. Second, it generates not only a forecast, but a complete, multivariate probability distribution for future outcomes of the economy which appears to be much more realistic than those generated by other competing approaches.

I will consider first the problem of economic forecasting, then the justification for the Bayesian approach, and finally, the performance record of a small BVAR model that has been used during the past four years.

The Problem of Economic Forecasting

The problem of forecasting is to use past and current information to generate from it a probability distribution for future events. Generally speaking, this is one of the basic problems of statistical analysis and there are many well-known statistical procedures which have been developed and used successfully to forecast in a variety of contexts.

Some particular difficulties arise, however, in forecasting economic data. First, there is only a limited amount of data, and what is available is often severely contaminated with measurement error. Second, many complicated relationships which are only poorly understood and probably evolving over time interact to generate the data. Finally, it is generally impossible to perform randomized experiments to test hypotheses about those economic structures. In this adverse environment, most of the standard statistical approaches do not work well.

The fact that aggregate economic quantities are usually measured with considerable error is well known. Conceptual problems, seasonal adjustment, changes in the mix of goods and services, and the nonreporting of cash and barter transactions are just of a few of the sources of this noise.

The sense in which there is only a limited amount of data is perhaps not so obvious. After all, the total quantity of economic data which is processed and available on computer data bases today is enormous. The paucity of useful data arises because of the pervasive interdependencies in the economy, and therefore in economic data. When we talk of forecasting the economy, we usually are referring to the problem of predicting either values of economic aggregates such as GNP or the price level, or of variables which are closely related to such aggregates. Most forecasts are short to medium term, and much of the variation in these aggregate variables at these horizons seems to be generated by an underlying phenomenon, "the business cycle." The sense in which data is scarce is that the entities that we are really trying to measure and forecast are business cycles, and the number of observations of business cycles relevant for use in forecasting today's economy is relatively small. Moreover, the structure of the economy appears to be evolving through time and government policies are constantly changing so the relevance of older observations is always called into question. Thus, despite the existence of larger and larger data bases, the small sample-size problem is likely to be with us for the foreseeable future.

Although explanations abound, very little is known with certainty about what causes and propogates business cycles. Theories point to a variety of sources of economic shocks and mechanisms for generating serial correlations in economic data. I believe that a realistic representation of the current state of economic theory requires a tremendous degree of uncertainty about the structure of the economy. If this is true, then a Bayesian procedure that can more accurately represent that uncertainty can produce a significant improvement over conventional techniques in our ability to generate a realistic probability distribution for future economic events.

The first point in this argument is the assumption that there is a high degree of uncertainty in our understanding of the structures which cause and propogate fluctuations in economic variables. Consider the list one could develop of the possible mechanisms causing business cycles. It would have to include a variety of both real and monetary factors. The real shocks would include, for example, crop failures and other weather-related events, wars, changes in fiscal policies, and fluctuations in international trade. The monetary shocks would include fluctuations in the money stock, changes in the international monetary system, and financial system shocks such as bank failures, speculative bubbles in asset prices, or a loss of confidence in the payments mechanism. Newer equilibrium business cycle theories focus on the effects of incomplete information, wage contracts, and responses to unanticipated changes in nominal quantities.

In recent years there has been a renewed interest in, but little agreement about, the causes of the Great Depression. At the time, increased industrial concentration was a popular explanation, as well as a decline in competition, and the failure of the price system. More recent examinations (e.g., Meltzer [1981]) have stressed both real and monetary causes, but come to less than complete agreement. Gordon and Wilcox [1981], for example, stress as causes the overproduction of capital due to "overbuilding of residential housing in the mid-1920s and the effect on consumer spending of the overshooting of the stock market during its 1928-29 speculative bubble" followed by declining population growth and its effect on residential housing.

Meltzer, on the other hand, cites "higher tariffs under Hawley Smoot, and retaliation abroad." He also mentions attempts to maintain the gold standard as well as anticipations of higher labor costs and lower after-tax returns to capital, changes in budget policy, interest rates, and stock prices.

The point of this discussion is that there are a multitude of economic theories of the business cycle, most of which focus on one part of a complex, multifaceted problem. Most economists would admit that each theory has some validity, though there is wide disagreement over the relative importance of the different approaches. It may be unnecessary to belabor this point perhaps the profusion of economic theories is obvious. On the other hand, a naive investigation into the workings of the current genre of large macroeconometric models might lead one to a completely opposite conclusion. Each of the behavioral equations in these models is typically based on a specific economic theory, and the theories in different models are often very similar. If one were to study only the equations in these models, one might conclude that there is a good deal of consensus on the economic structures involved.

Consider, for example, the investment equations in the DRI model. These equations are based on "the modern econometric theory of business fixed investment, developed by Dale Jorgenson [1963]," according to the description in Eckstein [1982]. "Actual investment, in the modern theory, is viewed as a partial adjustment of the capital stock toward the desired level," he writes. The desired level is then expressed as a function of expected output, the production technology, and factor prices. The model includes an equation with investment explained by the lagged stock of capital, the expected utilization rate, and distributed lags on a measure of the rental price of capital, on the ratio of interest payments to cash flow of nonfinancial corporations, and on real output.

Even if one accepts the Jorgenson theory as a reasonable approach to explaining investment, the empirical implementation described above does not adequately represent the true uncertainty about the determinants of investment. In the theory, expected output plays a criticl role in generating investment. Thus, any information which affects the future course of the economy will affect investment. Yet, in the DRI equation all such effects are delivered through a proxy term which is simply a fixed distributed lag on output. The empirical implementation of the theory requires many restrictions (here the exclusion from the expectation formulation of direct influence from variables that affect the course of future output) which are not particularly motivated by the theory itself.

A Bayesian who might try to derive from the Jorgenson theory a prior probability distribution for coefficients on variables in the model would presumably generate priors that were more informative for the coefficients on those variables directly incorporated in theory, and flatter about those that might enter through their effect on future output. Yet in the implementation described above, the implied priors have just the reverse property. Variables picked out by the theory, there lagged capital stock and factor prices, are included with flat priors on the coefficients, and other variables about which the theory says little, here all the excluded variables, are given coefficients with very informative priors--they are all set to zero.

Moreover, a thorough Bayesian would probably not be satisfied to give probability only to the Jorgenson theory. He might find a dozen theories of investment and give various weights to them. In a hypothetical calculation of the implied prior distribution for coefficients, he would likely find a wide range of variables which one or more of the theories picks out as likely to affect investment, and the effects would come through a wide variety of channels. He would thus find prior distributions for coefficients on many variables which looked similarly imprecise.

In the non-Bayesian approach to equation specification, the standard practice, aptly illustrated above, is to include only a few explanatory variables suggested by a given theory, and to exclude the rest. This practice is based on a practical recognition by the econometrician that given, his relatively small sample, he can ask only so much from the data. The problem with this approach, from the perspective of the Bayesian who considers several theories plausible, is that the non-Bayesian begins with very similar prior information for a variety of variables and he is forced in each case to make a decision to include or exclude the variable. For the Bayesian either choice is an extreme; the choice to include represents that nothing is known about the coefficient, the choice to exclude represents that the coefficient is known to be zero.

The Problem of Dimensionality

The standard approach to specifying equations recognizes that given a limited number of observations one must be very parsimonious about adding explanatory variables. Each additional coefficient must be estimated from the data, and while it will always improve the fit in sample (though not always when adjustment is made for degrees of freedom), in the forecasts generated by the equation there will be a tradeoff between decreased bias and increased variance. In a Bayesian specification framework, this tradeoff disappears in that a mean square error loss function is minimized by including all relevant variables along with prior information which accurately reflects what is known about the likely values of their coefficients. Of course there are practical limits to the extent to which variables can be included, but the limitations are due to computational feasibility, rather than being due to the lack of degrees of freedom.

One way to think about this problem is to view the forecasting equation as a filter which must pick out from the din of economic noise a weak signal which reveals the likely future course of the variable of interest. The standard approach takes the position that the best one can do is to rely on economic theory to suggest at most a few places to look for useful information. The search for information becomes narrowly focused. The alternative BVAR approach is based on a view that useful information about the future is likely to be spread across a wide spectrum of economic data. If this is the case, a forecasting equation which captures and appropriately weights information from a wide range of sources is likely to work better than one with a narrow focus. The appropriate weights are the coefficient estimates which combine information in the prior with evidence from the data.

We can illustrate the advantage of the Bayesian approach in a simple experiment designed to simulate the problem of modeling in an environment where the structure is uncertain. Suppose the analyst is interested in forecasting the variable Y, and he believes that Y may be affected by variables x_1, through x_N, which are ordered by the analyst according to how likely he believes the coefficient on that variable is to be different from zero. In a typical forecasting application this is likely to be possible. We will represent the analyst's prior as a set of independent distributions, with the coefficients, b_j, on variable x_j taken to be distributed

$$b_j \sim N(0.,j^{-2}).$$

In the usual specification procedure the analysts would likely either pick a few of the x's which he believed to be the most important, or he might order them and use a stepwise pretesting procedure to identify those variables to include in his "final" specification.

We compare the forecast errors made by either of those types of approaches with the results of specifying the Bayesian prior and using the posterior mean estimate as the basis for forecasting. In this simulation we will normalize the x's to be all independent, serially uncorrelated standard Gaussian variates. In each simulation, we generate data on Y by picking random x's and random coefficients from the normal distributions specified in the prior. For the purpose of simplifying the calculations we assume the equation error variance is known. We repeat the experiment 3,000 times, where in each case we generate artificial data and reestimate models to determine forecasting performance.

We estimate seven models by OLS, models including the most important one, two, three, four, five, and six variables, as well as a model where the number of included variables is chosen by a stepwise procedure which picks the smallest number such that one cannot reject the hypothesis that the excluded variables are all equal to zero at a 5 percent significance level. We compare the mean square error of coefficient estimates (where coefficients on excluded variables are taken to have estimates of zero) by these methods with the mean square error of the Bayesian posterior mean estimates.

The results for various numbers of observations and equation error variances are given in Table 1. Several interesting results are demonstrated in this exercise. First, notice that the usual concern about parsimony is well founded. Excluding variables whose coefficients are likely to be close to zero is better than including them in the standard approach when either the error variance is large, so that the R-squared (proportion of variance explained by the regression) is small, or when the number of observations is relatively small. Notice also that the use of a stepwise testing approach does not offer much room for improvement over a shrewd choice of a fixed set of variables to include. Finally, notice that the Bayesian approach offers a very significant advantage over any of the other specification whenever the number of observations relative to the R-squared is such that exclusionary restrictions might be desirable.

Admittedly, this experiment gives an unrealistic advantage to the Bayesian approach in that the coefficients are drawn from exactly the distribution which is included in the prior used for estimation. However, even when the prior variance is off by a factor of four, it generally works better than the standard approach. We include the results from simulating estimation using the prior

$$b_j \sim N(0., (j/2)^{-2}$$

as the line Wrg-Bayes in the table.

A similar problem arises in choosing a lag length in a time series approach. Dozens of formulas have been suggested for picking the appropriate lag length to satisfy this or that criterion in a variety of contexts. What such formulas ignore is that the reason one wants to choose a lag length in the first place is because one has prior information that more recent values of the variable in question have more information than more distant values. Truncation at a lag length, k, generates an estimate which reflects inappropriately that there is a clear break in ones prior information about lags k and k + 1. An alternative approach which more closely reflects ones actual prior information is to include as long a lag as is computationally feasible, with a prior distribution on the coefficients reflecting the fact that coefficients on longer lags are more likely to be close to zero. Of course this requires one to specify how quickly one's prior tightens around zero, but any such specifications within a wide range should be more appropriate than the prior implicit in either truncation at a given k or truncation based on a function of the evidence in the data.

The BVAR approach does not include any coefficients on moving average terms, as is usual practice in the ARIMA time series estimation approach. The use of moving average terms is designed to lead to parsimoniously parameterized representations which can generate long, and potentially infinite dimensional, autoregressive representations. The disadvantages of including moving average terms are well known; identification of the order of moving average and autoregressive lag lengths is difficult and estimation requires a nonlinear procedure. In multivariate contexts these problems are usually severe; whether they can be overcome in this context is perhaps an open question. To my knowledge there is no evidence available, such as I will present below for a BVAR model, to suggest that multivariate ARIMA models can consistently perform at least as well as the standard econometric models in real-time, out-of-sample economic forecasting.

Experience with BVARs

It is often difficult to evaluate a statistical method until it is used to solve applied problems. In order to illustrate the usefulness of this technique, I will briefly discuss its development and use at the Federal Reserve Bank in Minneapolis.

A major role of the research department at a regional Federal Reserve Bank, is to prepare a macroeconomic forecast for the Bank's president before each Federal Open Market Committee meeting. The BVAR approach was developed about six years ago in response to a perceived need to supplement, and possibly replace the use of a standard structural model in this context.

There were two major problems with the structural model as it was then used. First the model was very expensive to run, costing in the tens of thousands of dollars per year. Second, the model did not produce forecasts which were considered "reasonable" when allowed to run without judgemental input. The input which was required was a set of specified paths for many of the important variables in the model. In neither respect was the Bank's model different from most other structural models in use then or now.

The original idea to use a vector autoregression for forecasting at the Bank was suggested by Professor Thomas Sargent, an advisor to the Bank. His suggestion was motivated by concurrent research of Christopher Sims (later published as Sims [1980]) into the use of VARs in analyzing economic data.

My own involvement, which began as a research assistant at the Bank, was to write a program to estimate and forecast with VARs. Once this was done and we had specified what seemed to be a reasonable set of variables and lag lengths, we decided to test how well such a model would have performed over the previous several years making forecasts each quarter based only on previous observations. When we made this calculation we were surprised and disappointed to find that our preferred specification would have produced forecast errors consistently much larger than those of simple univariate time series models. In fact, we found that on average we would have done worse than almost any simple extrapolative method.

We tried a variety of changes in the specification such as replacing variables and down-weighting older observations, none of which made much difference. What we had not appreciated was the degree to which we were suffering from overparameterization, trying to fit too many parameters with too few observations. Before we gave up the VAR approach altogether, however, Sims suggested we try using a Bayesian approach to alleviate the problem.

The first time we implemented what we thought might be a reasonable Bayesian specification we found a substantial improvement in forecasting performance. The prior we used was based not so much on any particular economic theory as on the notion that most economic variables appear to approximate random walks, possibly with trends. What we did was to specify independent normal prior distributions for all of the parameters of the VAR with means of zero (except for the first lag of the dependent variable in each equation, which as given a prior mean of one) and standard deviations that were functions of a small set of hyperparameters of the prior. One of the hyperparameters, for example, scaled the standard deviations of prior distributions for coefficients on lags of all variables other than the dependent variable in each equation. By making this parameter zero (and making the standard deviations on own lags large), we could essentially duplicate the univariate benchmarks we had been using. What we then found with a little experimentation was that for a wide range of hyperparameter settings the forecast performance improved to a point much better than the benchmarks or any of the other models we had tried.

This work became the basis for my thesis at the University of Minnesota and led to a number of subsequent papers by myself and others at the Bank. (For example, Anderson [1979]; Doan, Litterman, Sims [1984]; Litterman [1980a], [1980b], [1984]; Sargent [1979]). It also led to my specifying a simple six-variable, six-lag quarterly model which I began to use to forecast with on a regular basis each month beginning in May 1980. The variables in that model are real GNP, the GNP price deflator, real business fixed investment, the 3-month treasury bill rate, the unemployment rate, and the money supply. The specification of that model is described in detail in Litterman [1980b]. It is now four years later and I continue to generate forecasts with the same model once a month. In the remainder of this paper, I will compare the forecasts generated by that BVAR model with those of three of the best known commercial forecasting services.

Measuring Forecast Performance

Before presenting the comparison, it will be useful to review some of the difficulties in interpreting evidence in forecast performance comparisons. In making this comparison I am, in effect, setting up a form of after-the-fact competition in which the rules and object of the competition were not specified ahead of time to the players. In this situation, there is an obvious potential risk that by selective reporting of results one could give a misleading picture of the results. This is especially true since different models are designed for different purposes, are specified at different levels of aggregation, and are used to forecast over various horizons.

Fortunately, there is a widespread agreement that the variables and horizons considered here are indeed those of primary interest. For many years the Statistical Abstract, a publication of the New York based Conference Board, has included each month a set of one- through eight-quarter-ahead forecasts of a number of commercial forecasting firms for four variables of primary economic interest, real GNP, nominal GNP, the unemployment rate, and the GNP price deflator. This publication is the source of data and the basis for the forecast comparison I make here.[1/]

The timing of release of economic forecasts is another important consideration in any forecasting competition. Forecasts are not generally published on exactly the same date, so they will to some extent be based on slightly different information sets. Forecasts of macroeconomic variables are generally dated according to the latest available National Income and Product Accounts (NIPA) data which was available at the time of release, and I follow that convention.

Notice that in the forecast comparison made here all participants were operating in "real time," making forecasts each month over a period of four years. Thus, we need not worry about how to interpret "out-of-sample" forecasts which are made after the fact. The all too common reporting of results from so-called "forecasting experiments" in which actual values are used for exogenous variables, those not included in the model, are subject to obvious criticism. Less obvious, but still problematical, are out-of-sample experiments in which a given specification is estimated using data only up to a certain date in order to make a forecast as of that date. Such simulations are certainly useful in some contexts; results from such an experiment, for example, were the reason we were led to use a Bayesian procedure. But for the most part, such comparisons cannot be used to rank models because it is very difficult to know how important after-the-fact information was in generating the specifications which were used in such an experiment. Today, for example, most conventional econometric models have highly developed energy sectors which in out-of-sample experiments are quite useful in forecasting the economic data of the seventies. Of course, no one was using those models at the time, and we can only guess today at what structures will be needed to forecast the economy in the future.

Another issue which arises, is how to define the target that everyone is trying to forecast. The answer is obvious for series such as an interest rate, which does not get revised, but not so obvious for historical economic data which are constantly revised. Scheduled revisions take place in NIPA data for at least three years, and benchmark revisions may make the historical data look quite different from the data observed at the time forecasts were made. Since these revisions generally affect levels and short-run growth rates rather than growth over several quarters, one approximate solution to this problem is to use the forecasted growth rates, applied to currently published base levels, to generate multi-step level-corrected "forecasts" which can be compared with currently published levels to measure forecast errors. This is the procedure used here.

Finally, one has to ask what it is that is being judged. Those who have not attempted to use large econometric models are probably unaware of the importance of the judgemental input, sometimes referred to as "tender loving care," which is applied by the forecaster. There is abundant evidence that the standard econometric models cannot be used mechanically to generate forecasts that compare in accuracy with those that are produced with judgemental input. This judgemental input is unfortunate, however, because it makes such forecasts nonreproducible and essentially takes them out of the realm of scientific study. My own guess is that forecast performance is much more related to the individual producing the forecast than to the model being used. In any case in order to judge a model, as opposed to the person running the model, one would like to have at least both the unadjusted and the adjusted forecasts for comparison. This information is unavailable, however, since unadjusted forecasts from these models are never published. In these circumstances it becomes very difficult to know how to interpret the forecast performance from a given commercial model. One might expect the performance to change, for example, when personnel at the firm change.

I think an important distinction can be drawn between forecasts from such models and the forecasts from the BVAR model which I have published for the past four years because the latter are purely mechanically produced forecasts without judgemental adjustment. Furthermore, they have been generated by a model whose specification has not changed over that period of time. They thus represent reproducible data, the statistical properties of which could be expected to remain stable if the model were to be used in the future.[2/]

A Forecast Performance Comparison

The forecast performance comparison is based on the monthly forecasts of the BVAR model, the Data Resources model, the Wharton model, and the Chase Econometrics Model. The first forecast was made in May 1980 and the last in February 1984. Where observations were not available for one of the forecasters (in a few cases eight-quarter-ahead forecasts were not published), observations for all forecasters were dropped from the sample. Because forecasts are made monthly of quarterly data, there are three forecasts for each observation of a given variable at a given horizon. These are sometimes referred to as early, middle, and late quarter forecasts depending on whether they are based on preliminary, or first or second revised NIPA estimate of the previous quarter. In this comparison, which is presented in Table II, I aggregate the results for these three months into a single category. Thus, for example. forecasts of data for the first quarter of 1984 made in January, February, and March of 1983 are all included in the five-quarter-ahead category (note that the one-quarter-ahead forecast refers to a forecast of the current quarter).

The measure of forecast accuracy, which is used, is the familiar root mean square error (RMSE). For the unemployment rate the RMSE measure of s-quarter-ahead forecast performance is simply

$$\left[\frac{1}{T}\sum_{t=1}^{T}(A_t - {}_sF_t)^2\right]^{1/2}$$

Where A_t is the actual value at time t, and ${}_sF_t$ is the forecast made s quarters earlier.

For the variables real GNP, the GNP deflator, and nominal GNP, errors are expressed as percentages of the level of the actual value. Due to the above mentioned correction for historical revisions, the formula for these variables appears somewhat complicated. Letting A_t be the actual value of the level of the variable at time t, and ${}_r\dot{F}_t$ be the forecasted percent growth (not annualized) in quarter t made r periods earlier, the formula for the RMSE at an s-quarter horizon is

$$\left\{\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\left(A_t - \{A_{t-s}\cdot\prod_{r=1}^{s}(1. + \frac{{}_r\dot{F}_{t+s-r}}{100.})\}\right)}{A_t}\right)^2\right\}^{1/2}$$

Perhaps the most important point to be made in interpreting the results in Table II, is that they are based on a small sample. The number of observations listed under the each horizon is small to begin with, and the errors in each category, particularly at long horizons, will be highly correlated. It is difficult to judge the results in Table II because we know they are based on a small, correlated sample, and we have no measures of significance. There is, of course, not much that can be done for judgemental forecasts.3/

For the BVAR model, however, there is an underlying probability model which can be used to generate measures of expected forecast error variance. These measures suggest that the above statictics do suffer from considerable sampling variability. The five-step forecast variance for real GNP, for example, is expected to be 2.710 percent, rather than the measured RMSE of 1.854 percent. On the other hand, the five-step forecast variance of the GNP deflator is expected to be 1.453 percent rather than the measured 3.670 percent. (It is intended that a Monte Carlo experiment will be performed to generate standard errors for the BVAR statistics in Table II.)

Despite the high degree of sampling error inherent in Table II, a few results are clear. It is demonstrated here that a time series forecasting procedure operating in real time, without judgemental adjustment, can produce forecasts which are at least competitive with the best forecasts commercially available. This is not a small achievement. The commercial forecasts are sold for prices in the tens of thousands of dollars per years. The BVAR model can be estimated, and forecasts generated, on an IBM personal computer in approximately three minutes.

A second result of interest is that the BVAR model appears to do relatively better at longer horizons. My interpretation of this tendency is that it reflects the significant advantage that the judgemental forecasts had in forecasting the current quarter during the first two years of the forecasting period. In any case, it clearly calls into question a common perception (e.g., Klein [1982]) that time series techniques may be useful for very short-term forecasts, but that structural models are needed to capture the turning points in business cycles necessary for accurate forecasting at longer horizons.

Although the RMSE is probably the best overall measure of forecast accuracy, it fails to reflect the degree to which the judgementally adjusted forecasts of the commercial firms tend to bunch together relative to the BVAR model. One measure which does reflect that tendency is the proportion of times a given forecaster is closest to the actual. These results clearly favor the BVAR model. Of the 1,128 forecasts considered, the BVAR model was most accurate 39.1 percent of the time. The percent of times each of the other forecasters was most accurate was 18.7, 19.1, and 23.1, for Chase, DRI, and Wharton, respectively.

Postscript

Over the four years since the model described above was specified, the state of the art of using BVARS has advanced considerably. In particular, models with time-varying parameters and much more sophisticated prior distributions have been developed. (See, for example, Sims [1982]; Litterman [1982]; and Doan, Litterman, Sims [1984]). As mentioned in footnote 2, the Federal Reserve Bank of Minneapolis has developed a larger (56 variable) monthly national forecasting model; several regional BVAR models have been developed; and the BVAR technique has also been used in applications to forecast state revenues (Litterman and Supel [1983]), to control the money supply (Litterman [1982]) and to measure the costs of intermediate targeting by the Federal Reserve System (Litterman [1984]).

Footnotes

[1]The history of commercial forecasts for the other variables included in the BVAR model is not generally available, although the records for real investment and the treasury bill rate have been kept by Stephen McNees, who has informed me privately that he will soon publish a more complete and independent evaluation on the BVAR forecasts as well as those of several of the commercial forecasting services.

[2]Being purely mechanical is not an easy task. In making forecasts, any quarterly model is at a tremendous disadvantage if it does not use recent information on monthly variables in forecasting the current quarter. My procedures has been to use an auxiliary set of equations with monthly data to generate the forecast of the current quarter. These current quarter values are then used by the quarterly model to forecast future quarterly values. The auxiliary equations have been updated during the four years. For the first several years, only three monthly variables in the model were used to forecast current quarter data for all six variables. Thus, for example, no current quarter monthly price data was used in forecasting the current quarter GNP deflator. This weakness of the model led to systematically worse one-quarter-ahead errors in the BVAR model than in the judgemental forecasts against which it is being compared. For the past two years a much larger and more sophisticated monthly BVAR model developed at the Federal Reserve Bank of Minneapolis has been used to generate the current quarter values. This new model, which is based on publicly available data and which also does not use

judgemental adjustment, appears to produce current quarter forecast which are as accurate as those produced with judgemental input.

3/ The commercial firms do not attempt to quantify uncertainty by publishing standard errors of forecasts. Instead, they typically publish alternative forecast scenarios which are given subjective probabilities. It is not clear how one could use such scenarios to quantify uncertainty, however.

References

Anderson, Paul. 1979. "Vector Autoregression, A new Tool for Regional Forecasting," Summer Quarterly Review, Federal Reserve Bank of Minneapolis.

Brunner, Karl. 1981. "The Great Depression Revisited," Martins Nijhoff Publishing, Boston.

Doan, Thomas, Robert B. Litterman and Christopher A. Sims. 1984. "Forecasting and Conditional Projection Using Realistic Prior Distributions," forthcoming in Econometric Reviews.

Eckstein, Otto. 1983. "The DRI Model of the U. S. Economy," McGraw-Hill, New York.

Gordon, Robert J., and James A. Wilcox. 1981. "Monetarist Interpretations of the Great Depression: An Evaluation and Critique," in 'The Great Depression Revisited,' K. Brunner ed., Martins Nijhoff Publishing, Boston.

Jorgenson, Dale W. 1963. "Capital Theory and Investment Behavior," American Economic Review, proceedings pp. 247-259.

Litterman, Robert B. 1980a. "Techniques for Forecasting with Vector Autoregressions," Doctoral thesis, University of Minnesota.

__________. 1980b. "A Bayesian Procedure for Forecasting with Vector Autoregressions," MIT working paper. (Available from the Research Department, Federal Reserve Bank of Minneapolis).

__________. 1982. "Specifying Vector Autoregressions for Macroeconomic Forecasting, Federal Reserve Bank of Minneapolis Staff Report 92, forthcoming in Studies in Bayesian Econometrics and Statistics, Volume III, in honor of Bruno de Finetti, (P. Goel, Ed., North Holland).

Meltzer, Allan H. 1981. "Comments on "Monetarist Interpretations of the Great Depression" in 'The Great Depression Revisited,' K. Brunner ed., Martins Nijhoff Publishing, Boston.

Sargent, Thomas J. 1979. Vector Autoregression . . ." in Summer Quarterly Review, Federal Reserve Bank of Minneapolis.

Sims, Christopher A. 1980. "Macroeconomics and Reality," Econometrica 48, pp. 1-48.

Table I

Simulation Comparison of Bayesian
With Standard Specification Approaches

Mean Square Error of Estimated Coefficients
(percentage increase relative to Bayesian estimates)

Equation Error Variance = 4.
Population R-squared = .27

Model	Observations 13		19		31	
OLS Variable 1	.902	(46)	.772	(53)	.656	(73)
OLS Vars (1,2)	1.092	(78)	.777	(54)	.555	(46)
OLS Vars (1-3)	1.532	(149)	.954	(90)	.597	(58)
OLS Vars (1-4)	2.059	(235)	1.221	(142)	.699	(84)
OLS Vars (1-5)	2.842	(362)	1.567	(211)	.850	(124)
OLS Vars (1-6)	4.227	(587)	1.934	(284)	1.023	(170)
OLS Stepwise	1.873	(204)	1.085	(116)	.693	(83)
Bayesian Vars (1-6)	.615		.503		.379	
Wrg-Bayes Vars (1-6)	.809	(32)	.673	(34)	.518	(37)

Equation Error Variance = 1.0
Population R-squared = .60

Model	Observations 13		19		31	
OLS Variable 1	.629	(102)	.585	(152)	.546	(255)
OLS Vars (1,2)	.483	(55)	.398	(72)	.330	(109)
OLS Vars (1-3)	.508	(63)	.357	(54)	.259	(64)
OLS Vars (1-4)	.584	(88)	.370	(60)	.234	(49)
OLS Vars (1-5)	.742	(138)	.417	(80)	.238	(51)
OLS Vars (1-6)	1.059	(240)	.480	(107)	.258	(63)
OLS Stepwise	.657	(111)	.421	(81)	.267	(69)
Bayesian Vars (1-6)	.311		.232		.158	
Wrg-Bayes Vars (1-6)	.421	(35)	.320	(38)	.220	(39)

Equation Error Variance = .05
Population R-squared = .97

Model	Observations 13		19		31	
OLS Variable 1	.546	(1507)	.530	(2870)	.516	(5145)
OLS Vars (1,2)	.296	(771)	.277	(1451)	.260	(2543)
OLS Vars (1-3)	.184	(442)	.166	(833)	.150	(1424)
OLS Vars (1-4)	.117	(244)	.101	(464)	.085	(760)
OLS Vars (1-5)	.067	(97)	.054	(206)	.041	(319)
OLS Vars (1-6)	.042	(24)	.019	(7)	.010	(4)
OLS Stepwise	.055	(62)	.023	(31)	.012	(19)
Bayesian Vars (1-6)	.034		.018		.010	
Wrg-Bayes Vars (1-6)	.047	(39)	.023	(30)	.012	(19)

Table II

BVAR Model Forecast Performance Comparison
Root Mean Squared Forecast Errors

Real GNP

Forecast Horizon in Quarters (number of observations)

Model	1 (46)	2 (43)	3 (40)	4 (37)	5 (34)	6 (31)	7 (28)	8 (23)
BVAR								
Actual	.878	1.192	1.677	1.969	1.854	1.974	2.274	3.021
Mean	1.037	1.567	1.993	2.359	2.701	3.025	3.367	3.706
Std. Dev.	.173	.351	.536	.717	.904	1.113	1.334	1.571
Chase	.812	1.500	2.271	2.868	3.316	3.738	3.935	3.803
DRI	.727	1.273	2.024	2.639	3.104	3.596	3.914	3.888
Wharton	.661	1.246	1.978	2.628	3.089	3.566	3.990	4.015

GNP deflator

Forecast Horizon in Quarters (number of observations)

Model	1 (46)	2 (43)	3 (40)	4 (37)	5 (34)	6 (31)	7 (28)	8 (23)
BVAR								
Actual	.539	1.143	1.917	2.829	3.670	4.644	5.674	6.364
Mean	.528	.865	1.196	1.525	1.841	2.139	2.422	2.691
Std. Dev.	.067	.178	.317	.472	.647	.827	1.011	1.202
Chase	.340	.614	.952	1.450	2.100	2.834	3.574	4.246
DRI	.357	.592	.846	1.368	2.062	2.879	3.755	4.610
Wharton	.471	.727	1.094	1.587	2.215	2.982	3.810	4.519

Unemployment Rate

Forecast Horizon in Quarters (number of observations)

Model	1 (46)	2 (43)	3 (40)	4 (37)	5 (34)	6 (31)	7 (28)	8 (23)
BVAR								
Actual	.246	.568	.845	1.039	1.175	1.168	1.670	1.670
Mean	.327	.551	.730	.864	.961	1.036	1.103	1.162
Std. Dev.	.057	.131	.212	.284	.347	.403	.454	.499
Chase	.278	.622	.969	1.307	1.585	1.798	2.080	2.198
DRI	.228	.583	.916	1.336	1.718	2.027	2.282	2.391
Wharton	.248	.596	.931	1.198	1.505	1.689	1.817	1.932

Nominal GNP

	Forecast Horizon in Quarters (number of observations)							
	1 (46)	2 (43)	3 (40)	4 (37)	5 (34)	6 (31)	7 (28)	8 (23)
Model								
BVAR								
Actual	1.045	1.685	2.837	3.841	4.385	4.861	5.037	4.752
Mean	1.198	1.854	2.406	2.886	3.312	3.704	4.091	4.484
Std. Dev.	.179	.377	.606	.845	1.094	1.356	1.634	1.916
Chase	.958	1.779	2.820	3.838	4.951	6.181	7.255	7.883
DRI	.848	1.465	2.428	3.481	4.603	5.967	7.300	8.181
Wharton	.834	1.690	2.742	3.869	5.044	6.235	7.477	8.162