Federal Reserve Bank of Minneapolis
Research Department Working Paper

FORECASTING WITH BAYESIAN VECTOR AUTOREGRESSIONS--FIVE YEARS OF EXPERIENCE

Robert B. Litterman

Working Paper 274

Revised July 1985

Forecasting with Bayesian Vector Autoregressions--Five 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, economic forecasting with Bayesian vector autoregressions (BVAR), which has proved over the past several years 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 and inexpensive to use, and they 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 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, third, its implementation, and finally the performance record of a small BVAR model that has been used during the past five years.

The Problem of Economic Forecasting

The problem of forecasting is to use past and current information to generate 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 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 there-

fore 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 and the price level or values 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 are 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 propagates 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 propagate 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, and losses 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 of that event, increased industrial concentration was a popular explanation, as were a decline in competition and the failure of the price system. More recent examinations have stressed both real and monetary causes, but come to less than complete agreement. (See, for example, Brunner [1981].) On the one hand, Gordon and Wilcox [1981, p. 77], 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 [1981, p. 152], 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 and 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. However, 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 Data Resources (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 [1983, p. 129]. "Actual investment, in the modern theory,

is viewed as a partial adjustment of the capital stock toward the desired level," Eckstein writes [p. 131]. 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 critical 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. This type of analyst might find a dozen theories of investment and give various weights to them. In a hypothetical calculation of the implied prior distribution for coefficients, the analyst 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. The analyst 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 the relatively small sample, one 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 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 doing this 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 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 believes that Y may be affected by variables x_1 through x_N, which are ordered according to how likely the analyst believes the coefficient on that variable is to be different from zero. In a typical forecasting application, this is likely to be possible. I 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

$$b_j \sim N(0., j^{-2}). \tag{1}$$

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

I 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, I will normalize the x's to be all independent, serially uncorrelated standard Gaussian variates. In each simulation, I 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, I assume the equation error variance is known. I repeat the experiment 3,000 times, and each time I generate artificial data and reestimate models to determine forecasting performance.

I estimate seven models by OLS, models including the most important one, two, three, four, five, and six variables as well as a model in which the number of included variables is chosen by a stepwise procedure which picks the smallest number of variables such that one cannot reject the hypothesis that the excluded variables are all equal to zero at a 5 percent significance level. I compare the mean square error (MSE) 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.

$$\text{MSE} \equiv \sum_{s=1}^{3000} \left[\sum_{j=1}^{6} (b_j - \hat{b}_j)^2 \right] / 3000. \qquad (2)$$

The results for various numbers of observations and equation error variances are given in Table I. Several interesting results are demonstrated in this exercise. First, notice that

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	Population R-Squared	Model	Observations 13		19		31	
4.0	.27	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)
1.0	.60	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)
.05	.97	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)

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 either when 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 specifications 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 much better than the standard approach. I include the results from estimation using the prior

$$b_j \sim N(0.,(j/2.)^{-2}) \tag{3}$$

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 one's prior information about lags k and k + 1. An alternative approach which more closely reflects one's 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.

The Vector Autoregression Representation

An m^{th} order autoregressive representation for the n-vector Y is given by

$$\underset{nx1}{Y(t)} = \underset{nx1}{D(t)} + \sum_{j=1}^{m} \underset{nxn}{B_j}\ \underset{nx1}{Y(t-j)} + \underset{nx1}{\varepsilon(t)} \qquad t = 1,\ldots T \qquad (4)$$

$$E[\varepsilon(t)\varepsilon(s)'] = \begin{cases} \Sigma & \text{if } s = t \\ 0 & \text{otherwise} \end{cases}$$

where D(t) captures the deterministic component of Y(t). In general D(t) is a linear function of an n x d matrix of parameters, C. In the examples which follow D(t) includes a constant term for each component of Y.

The i^{th} equation has the following scalar form:

$$\begin{aligned} Y_i(t) = d^i(t) &+ b^i_{11}Y_1(t-1) + b^i_{21}Y_1(t-2) + \ldots + b^i_{m1}Y_1(t-m) \\ &+ b^i_{12}Y_2(t-1) \qquad\qquad + \ldots + b^i_{m2}Y_2(t-m) \\ &+ b^i_{1n}Y_n(t-1) \qquad\qquad + \ldots + b^i_{mn}Y_n(t-m) \\ &+ \ldots \quad + \varepsilon_i(t) \end{aligned} \qquad (5)$$

where b^i_{jk} above is the k^{th} element of the i^{th} row of B_j in matrix notation, and $d^i(t)$ is the i^{th} element of the deterministic component.

For ease of exposition we also adopt the somewhat misleading notation (since X includes lagged Y's)

$$\underset{Tx1}{Y} = \underset{Txp}{X}\ \underset{px1}{\beta} + \underset{Tx1}{\varepsilon} \tag{6}$$

to refer to this equation. Using this notation the estimator suggested here is

$$\beta^k = (X'X + kR'R)^{-1}(X'Y + kR'r). \tag{7}$$

This estimator combines the data generated by the model in (6), assuming $\varepsilon \sim N(0,\sigma^2 I)$, with the prior information contained in specification

$$\underset{qxp}{R}\ \underset{px1}{\beta} = \underset{qx1}{r} + \underset{qx1}{\nu} \qquad \nu \sim N(0,\lambda^2 I) \tag{8}$$

where $k = \sigma^2/\lambda^2$. Ridge estimators correspond to setting R = I, the identity matrix, and r = 0, the p-dimensional zero vector. Stein type estimators are generated by taking R = X and r = 0. Other estimators of this type which impose smoothness across coefficients in distributed lag models have been suggested by Leamer [1972] and Shiller [1973].

Rather than impose smoothness, the estimator suggested here imposes the information that a reasonable approximation of the behavior of an economic variable is a random walk around an unknown, deterministic component. All equations in the system are given the same form of prior distribution. For the i^{th} equation this distribution is centered around the specification

$$Y_i(t) = Y_i(t-1) + d^i(t) + \varepsilon_i(t). \tag{9}$$

The parameters are all assumed to have means of zero except the coefficient on the first lag of the dependent variable, which is given a prior mean of one. The parameters are assumed to be uncorrelated with each other and to have standard deviations which decrease the further back they are in the lag distributions. In general, the prior distribution is much looser, that is, has larger standard deviations on lag coefficients of the dependent variable than it is on other variables in the system. Generally, without observing the data very little is known about the distribution of the parameters of the deterministic component. In order to represent this ignorance, a noninformative prior is used. The flat prior is not a proper probability distribution, but is justified in the usual manner as an approximation to a proper, but suitably diffuse prior.

The prior which has been described here is not derived from a particular economic theory, and, in this sense, the restrictions it imposes may be referred to as instrumental. The intuition behind its use is its ability to capture more accurately uncertain a priori information than other standard methods of restricting VAR representations. Probably the most objectionable aspects of this prior are its reflection of complete ignorance about the deterministic components and its prior mean, which reflects a nonstationary process. Both of these specifications are likely candidates for modification in particular applications. On the other hand, these parts of the prior are the areas in which the prior is most uncertain anyway, and thus, they are

the areas in which the data will dominate most completely. For this reason forecasting performance should be relatively insensitive to specification of other reasonably loose priors with respect to the constant and the first lag of the dependent variable. It certainly may be true, however, that if one were forecasting growth rates of real GNP, for example, a random walk prior is not appropriate, one might do better by specifying a mean of less than one on the first own lag.

The justification for this prior is simply that through its use we are able to express more realistically our true state of knowledge and uncertainty about the structure of the economy. When there are known relationships among variables, whether derived from economic theory or other considerations, that information should be imposed in the estimation process. We are, however, sympathetic to the many economists who feel that the theory which is typically used to identify the equations of econometric models is not valid. Lucas and Sargent [1979], for example, contend that "probabilistic microeconomic theory almost never implies either exclusion restrictions that were suggested by Keynes or those that are imposed by macroeconomic models." Similarly, Sims [1980] suggests that "claims for identification in these models cannot be taken seriously," and that, "a more systematic approach to imposing restrictions could lead to capture of empirical regularities, which remain hidden to standard procedures, and, hence, lead to improved forecasts and policy projections."

Forecasting with a Vector Autoregression

In this section we describe the application of this method in a forecasting experiment with a particular VAR system. The empirical work reported here was performed in 1979. It led to the specification of a model which has been used on a monthly basis for forecasting in subsequent years. The results of that real-time forecasting experiment are reported in the final section of this paper. The work reported here is taken from Litterman (1980b). More recent surveys of developments in BVAR modeling can be found in Todd [1984], Litterman [1984c] and Doan, Litterman and Sims [1984]. The system includes quarterly observations on the following seven variables: annual growth rates of real GNP, RGNP; annual inflation rates (growth rates of the GNP deflator), INFLA; the unemployment rate, UNEMP; logged levels of the money supply, M1; logged levels of gross private domestic investment, INVEST; the rate on four- to six-month commercial paper, CPRATE; and the change in business inventories, CBI. Observations were obtained from 1948-1 through 1979-3.

Each equation in this seven-variable system includes six lags of each variable and a constant term, a total of forty-three free parameters. In the context of this system it is shown first that the posterior mean estimators can lead to a consistent, large improvement forecast performance relative to unrestricted ordinary least squares estimation.

The prior information we specify treats each equation in the same manner. The matrix R is normalized so that λ is the

standard deviation on the first lag of the dependent variable. Given λ, the standard deviations of further coefficients in the lag distributions are decreased in a harmonic manner. The coefficient on own lag j, $j = 2, \ldots, 6$, is given an independent normal prior distribution with mean zero and standard deviation λ/j. The standard deviations on lags of variables other than the dependent variable are made tighter around zero at all lags by a factor, $\theta = .2$ to reflect the assumption that own lags account for most of the variation of a given variable.

The standard deviations around coefficients on lags of other than the dependent variable are not scale invariant. For example, how tight a standard deviation of .1 is on lags of GNP in an interest rate equation will depend on whether GNP is measured in dollars, or in billions of dollars. Thus, in general, the prior cannot be specified completely without reference to the data.

This scale problem is usually solved in the ridge regression context by transforming the data so that $X'X$ is a correlation matrix. In effect, this scales the implicit prior by the standard deviations of the independent variables. I am led away from this approach because I suspect that the scale of the response of one economic variable to another is more often a function of the relative sizes of unexpected movements in the two variables than of the relative sizes of their overall standard errors. In the results reported here the measure of the size of unexpected movements in variable i is taken to be the estimated

standard error $\hat{\sigma}_i$, of the residuals in an unrestricted univariate autoregression with a constant and six lags.

In summary, letting δ_{ij}^{ℓ} be the standard deviation of the prior distribution for the coefficient on lag ℓ of variable j in equation i, then

$$\delta_{ij}^{\ell} = \begin{cases} \lambda/\ell & \text{if } i = j \\ \dfrac{\theta\lambda\hat{\sigma}_i}{\ell\hat{\sigma}_j} & \text{if } i \neq j \end{cases} \tag{10}$$

Thus, to put the prior for the i^{th} equation in the form of (8) we make R a diagonal matrix with zeros corresponding to deterministic components and elements $[\lambda/\delta_{ij}^{\ell}]$ corresponding to the ℓ^{th} lag of variable j. r is a vector of zeros and a 1 corresponding to the first lag of the dependent variable. R'R is singular here reflecting the improper flat prior on coefficients of the deterministic component. This explains why the prior is expressed as in (5) rather than as a proper probability density for β. As noted above, this procedure is justified as an approximation to a proper, but locally uniform, prior distribution.

A gain in efficiency could be made by estimating all equations together via a seemingly unrelated regression procedure which uses the information contained in the covariances of residuals across equations. I do not attempt such a procedure primarily because of the computational burden; it would require inversion of an $n^2m + nd$ (301 in this case)--order matrix.

If σ^2 and λ^2 were known, the estimator in (7) would have a Bayesian justification as a posterior mean. When σ^2 and λ^2 are not known one is faced with a problem which is usually encountered in the context of shrinkage estimators, that is determining how far to shrink. A Bayesian solution which takes λ as given and a diffuse prior distribution for σ leads to a normal-t posterior density for β which would require an intractable numerical integration in order to calculate the posterior mean. I chose instead an approximation based on the suggestion by Zellner [1971], Section 4.2, of using $\hat{\sigma}$, the estimated standard error of the unrestricted OLS regression in place of σ. I use instead $\hat{\sigma}_i$, the univariate regression standard error, simply because in large VAR systems with few or no degrees of freedom $\hat{\sigma}$ may be an unreliable estimator or may not exist.

The results in Table II demonstrate the improvements in forecasting which were produced by imposing the prior on the seven-variable system above. Using data beginning in 1948-1, the system in its unrestricted form and combined with the above prior with several degrees of tightness (values of λ) is estimated each quarter from 1971-1 to 1975-3. Each period the resulting estimates are used to make forecasts of one to eight steps ahead using the chain rule of forecasting. The chain rule takes estimated one-step-ahead forecasts as the basis for two-step-ahead forecasts and so on. Evidence in Fair [1978] suggest that this method produces good approximations to the posterior mean multistep forecasts suggested in Chow [1973].

Table II

Theil Coefficients 1971-1 to 1975-4

	Forecast Horizon: Quarter Ahead								
	1	2	3	4	5	6	7	8	Average
Real GNP Growth									
No Prior	1.89	1.13	1.12	1.23	.98	1.11	1.18	.98	1.20
$\lambda = .5$	.84	.67	.67	.76	.69	.65	.71	.65	.70
$\lambda = .3$	.79	.64	.66	.75	.69	.67	.72	.66	.70
$\lambda = .2$	.77	.64	.67	.75	.70	.68	.73	.67	.70
$\lambda = .1$	.82	.73	.71	.77	.75	.74	.77	.73	.76
Inflation									
No Prior	1.90	1.17	1.11	.86	.90	1.05	1.08	1.03	1.14
$\lambda = .5$	1.02	.91	.96	.93	.92	.94	.93	.91	.94
$\lambda = .3$	1.01	.91	.96	.92	.91	.92	.91	.88	.93
$\lambda = .2$	1.00	.92	.95	.91	.91	.92	.91	.89	.92
$\lambda = .1$	.98	.92	.94	.92	.93	.94	.94	.92	.94
Unemployment									
No Prior	1.33	1.43	1.60	1.75	1.75	1.66	1.63	1.78	1.62
$\lambda = .5$	.83	.87	.80	.76	.79	.78	.76	.73	.80
$\lambda = .3$	.78	.81	.76	.74	.75	.75	.74	.73	.76
$\lambda = .2$	.77	.79	.74	.74	.75	.76	.76	.76	.76
$\lambda = .1$	.80	.82	.80	.80	.81	.82	.82	.84	.81

	Forecast Horizon: Quarter Ahead								
	1	2	3	4	5	6	7	8	Average
Investment									
No Prior	1.23	1.33	1.48	1.71	1.61	1.39	1.20	1.20	1.39
$\lambda = .5$	1.07	1.09	1.26	1.32	1.20	1.12	1.08	1.06	1.15
$\lambda = .3$	1.06	1.07	1.23	1.30	1.20	1.14	1.12	1.12	1.15
$\lambda = .2$	1.03	1.03	1.17	1.25	1.17	1.13	1.13	1.14	1.13
$\lambda = .1$	.97	.96	1.04	1.12	1.08	1.07	1.08	1.12	1.06
CBI									
No Prior	1.03	.97	.96	10.6	.98	.91	.90	.90	.96
$\lambda = .5$	.93	.85	.91	.95	.86	.80	.80	.82	.87
$\lambda = .3$	.93	.86	.90	.93	.85	.80	.80	.82	.86
$\lambda = .2$	.94	.87	.89	.93	.86	.81	.82	.83	.87
$\lambda = .1$	.96	.91	.91	.95	.89	.85	.86	.87	.90
CPRATE									
No Prior	1.09	1.30	1.32	1.43	1.55	1.57	1.45	1.18	1.36
$\lambda = .5$	.87	.95	.88	.77	.69	.66	.60	.52	.74
$\lambda = .3$	.91	.97	.88	.75	.68	.63	.57	.49	.74
$\lambda = .2$	.93	.95	.86	.73	.66	.61	.55	.48	.72
$\lambda = .1$	.93	.92	.84	.74	.68	.63	.57	.52	.73
M1									
No Prior	.86	1.03	.94	.85	.83	.87	.93	1.00	.91
$\lambda = .5$	.37	.36	.31	.34	.30	.28	.29	.26	.31
$\lambda = .3$	.37	.32	.26	.28	.25	.24	.24	.22	.27
$\lambda = .2$	.37	.30	.25	.26	.23	.23	.24	.22	.26
$\lambda = .1$	.37	.30	.25	.26	.25	.25	.26	.25	.27

Mean square error and Theil coefficients are calculated for each variable at each forecast horizon. The Theil coefficient scales the root mean square error by the root square error of no-change forecasts. This scaling allows comparison to some extent across variables and across horizons. The main result is very clear in Table II. For each of the seven variables, at all horizons, there is an obvious improvement in forecasting as the prior is imposed, relative to the unrestricted model. Values of .5, .3, .2, and .1 were tried for the tightness parameter, λ with the prior. Recall that λ is the standard deviation of the first lag of the dependent variable in each equation. All other standard errors are scaled relative to it. The best overall results were generated with $\lambda = .2$. It is clear that forecasting results are not overly sensitive to changes in this parameter.

The improvement in forecasting demonstrated in Tables II is not particularly surprising. It simply reflects the over-parameterization of the unrestricted system. A more interesting question is how well do the posterior mean estimators forecast relative to other alternative methods. One indication is given by a comparison of these results with the forecast performance of univariate autoregressive equations with constant, six lags and no prior for the same period. Such an equation is, of course, the limiting case of this prior as θ goes to 0, and λ goes to infinity such that $\theta\lambda$ goes to zero. The system with the prior specified above and appropriate λ's almost uniformly outperforms the univariate equations. There is an obvious qualification to these re-

sults however, which is that the optimal λ could not have been known ahead of time. For this reason an additional experiment was performed to compare this prior and the optimal λ with other forecasting methods over the subsequent period 1976-1 to 1979-3. The forecast statistics in the earlier period were compiled as if data for 1976-1 and later were not available in order to avoid biasing the second experiment.

This second experiment was designed to allow comparison not only with ARIMA and univariate autoregression models, but also with the compiled records of two professional forecasters, Data Resources, Inc. (DRI) and Chase Econometric Associates, Inc. (CHASE). The compiled records for these forecasters were taken from the Statistical Abstract published monthly by the Conference Board.

Each of four mechanical forecasts of the quarterly data were updated on a monthly basis using the new or revised information actually available to the professional forecasters at the beginning of the particular month. For example, following the standard convention, the one-step-ahead forecast made in January is a forecast of the fourth quarter data based on the final data for the third quarter. The February one-step-ahead forecast is of first quarter data on the basis of preliminary fourth quarter values. This procedure is followed primarily because it ensures that all of the information used by the models was available to forecasters at the time of their forecast.

The results in Table III show that the posterior mean estimator performed quite favorably in comparison with the other models. It is also clear that during this period no obvious advantage over standard univariate time-series methods was obtained by the professional forecasters' use of structural models, larger information sets and judgemental adjustment.

Forecasting With BVARs

The empirical work reported above 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 prior is the same as shown above except the relative weight parameter, θ, is set at .3. It is now five years later, and I continue to generate forecasts with essentially 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, Data Resources, Wharton EFA, and Chase Econometrics.

Over the past five years I have sent, at no charge, the BVAR forecasts on a regular basis to a list of interested parties consisting primarily of academics. A common response that I have received has been an impression that there is something different or wrong with the BVAR forecasts because they are too "volatile" or "wild," relative to standard forecasts.

Table III

Mean Square Errors of Forecasts* 1976-1 to 1979-4

	Forecast Horizon: Quarters Ahead						
	2	3	4	5	6	7	8
Real GNP Growth							
DRI	2.726	2.801	2.951	3.388	3.566		
	(42)	(39)	(33)	(19)	(12)		
CHASE	3.052	3.391	3.408	3.875	4.224	4.043	3.809
	(42)	(39)	(36)	(33)	(29)	(23)	(14)
ARIMA	2.882	3.071	3.076	3.181	3.209	3.266	3.418
	(42)	(39)	(36)	(33)	(30)	(27)	(24)
UNIVARIATE AR	3.192	3.401	3.405	3.656	3.391	3.331	3.374
	(42)	(39)	(36)	(33)	(30)	(27)	(24)
BAYESIAN VAR	2.841	3.053	2.948	2.959	3.021	2.999	3.281
	(42)	(39)	(36)	(33)	(30)	(27)	(24)
Inflation							
DRI	1.605	1.929	2.277	2.894	2.727		
	(42)	(39)	(33)	(18)	(12)		
CHASE	1.565	2.039	2.412	2.780	3.221	3.602	3.151
	(42)	(39)	(36)	(33)	(28)	(23)	(14)
ARIMA	1.674	1.907	1.755	2.211	2.327	2.773	2.773
	(42)	(39)	(36)	(33)	(30)	(27)	(24)
UNIVARIATE AR	2.289	2.735	3.111	3.526	3.940	4.235	4.539
	(42)	(39)	(36)	(33)	(30)	(27)	(24)
BAYESIAN VAR	1.469	1.624	1.441	1.710	1.640	1.793	1.514
	(42)	(39)	(36)	(33)	(30)	(27)	(24)
Unemployment							
DRI	.341	.449	.485	.494	.430		
	(42)	(39)	(33)	(19)	(12)		
CHASE	.510	.817	1.040	1.201	1.477	1.817	2.301
	(42)	(39)	(36)	(33)	(29)	(23)	(14)
ARIMA	.466	.712	.915	1.073	1.236	1.384	1.491
	(42)	(39)	(36)	(33)	(30)	(27)	(24)
UNIVARIATE AR	.362	.493	.541	.566	.576	.526	.461
	(42)	(39)	(36)	(33)	(30)	(27)	(24)
BAYESIAN VAR	.383	.497	.559	.627	.738	.845	.961
	(42)	(39)	(36)	(33)	(30)	(27)	(24)

*Number of observations given in parentheses.

Such a reaction was perhaps to be expected given, for example, that in my first forecast, published May 1, 1980, the unemployment rate was forecast to rise from the then current rate of 6.1 percent, to above 11 percent by the end of 1982. The DRI, Chase, and Wharton forecasts at that time projected the unemployment rate to peak between 7.5 and 8.2 percent.

There is at least one obvious explanation for the different behavior of the BVAR forecast from other published forecasts. The BVAR forecast is the unadjusted product of a statistical procedure designed to pick a point as close as possible to the future value of the variable in question. Other forecasts are typically sold to clients and are judgementally adjusted, presumably in ways that are designed to maximize the demand for the forecast. It is not at all clear that an unbiased forecast is also a profit-maximizing forecast. For example, faced with the outlook for the unemployment rate described above in May 1980, a profit-maximizing forecaster might have published a forecast with the unemployment rate rising only to 9 percent even though his own model projected unemployment rising to 11 percent. The cost in terms of lost credibility of deviating from the range of other forecasters would have to have been weighed against the questionable benefit of positioning one's forecast further above the range of other forecasters.

For whatever reason, the deviation of the BVAR forecast from the range of the Chase, DRI and Wharton forecasts has been obvious. In Figures 1 through 6, we illustrate this phenomenon in

Figure 1

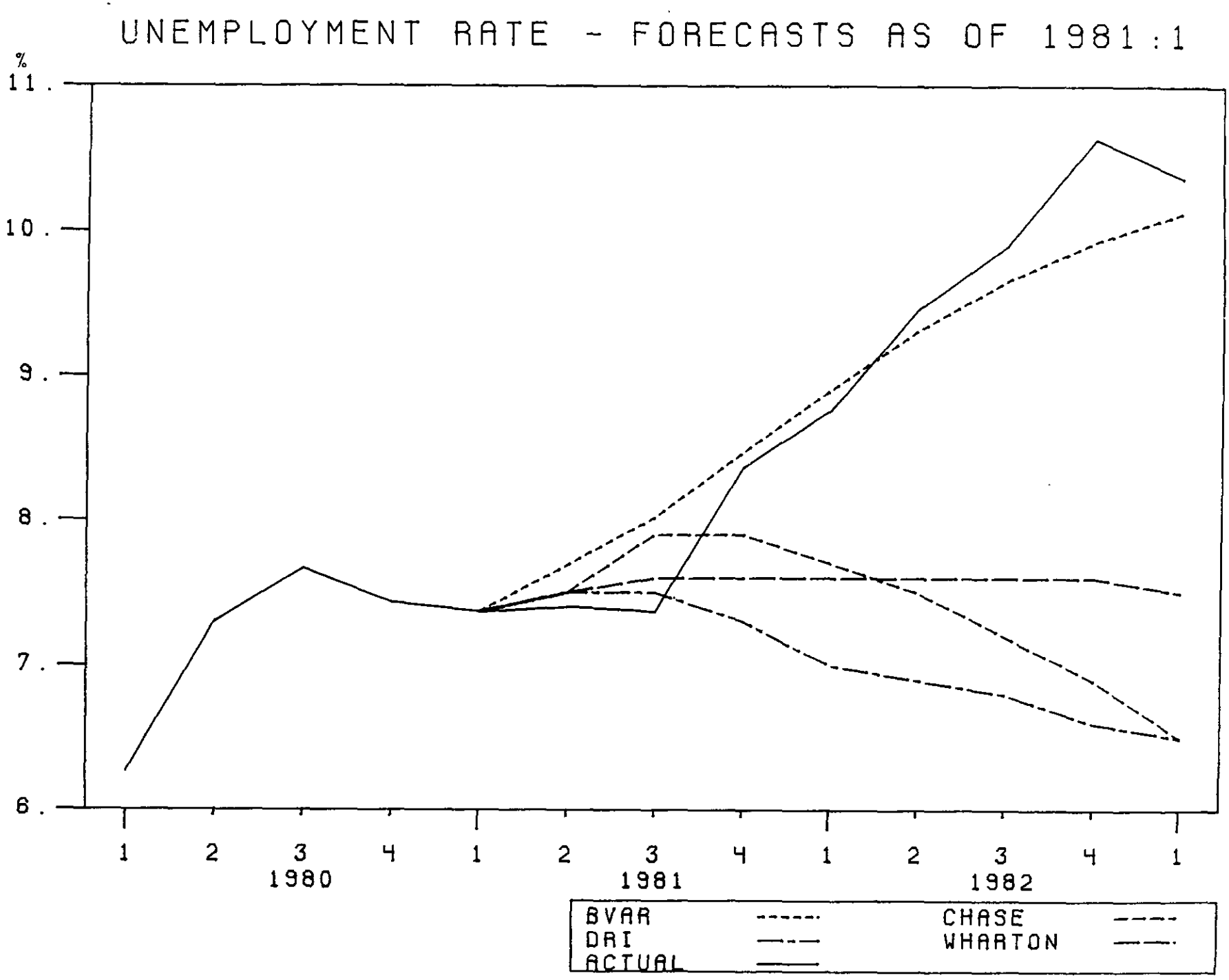

UNEMPLOYMENT RATE - FORECASTS AS OF 1981:1
%
11.
10.
9.
8.
7.
6.
1
2
3
4
1980
1981
1982
BVAR
DRI
ACTUAL
CHASE
WHARTON

Figure 2

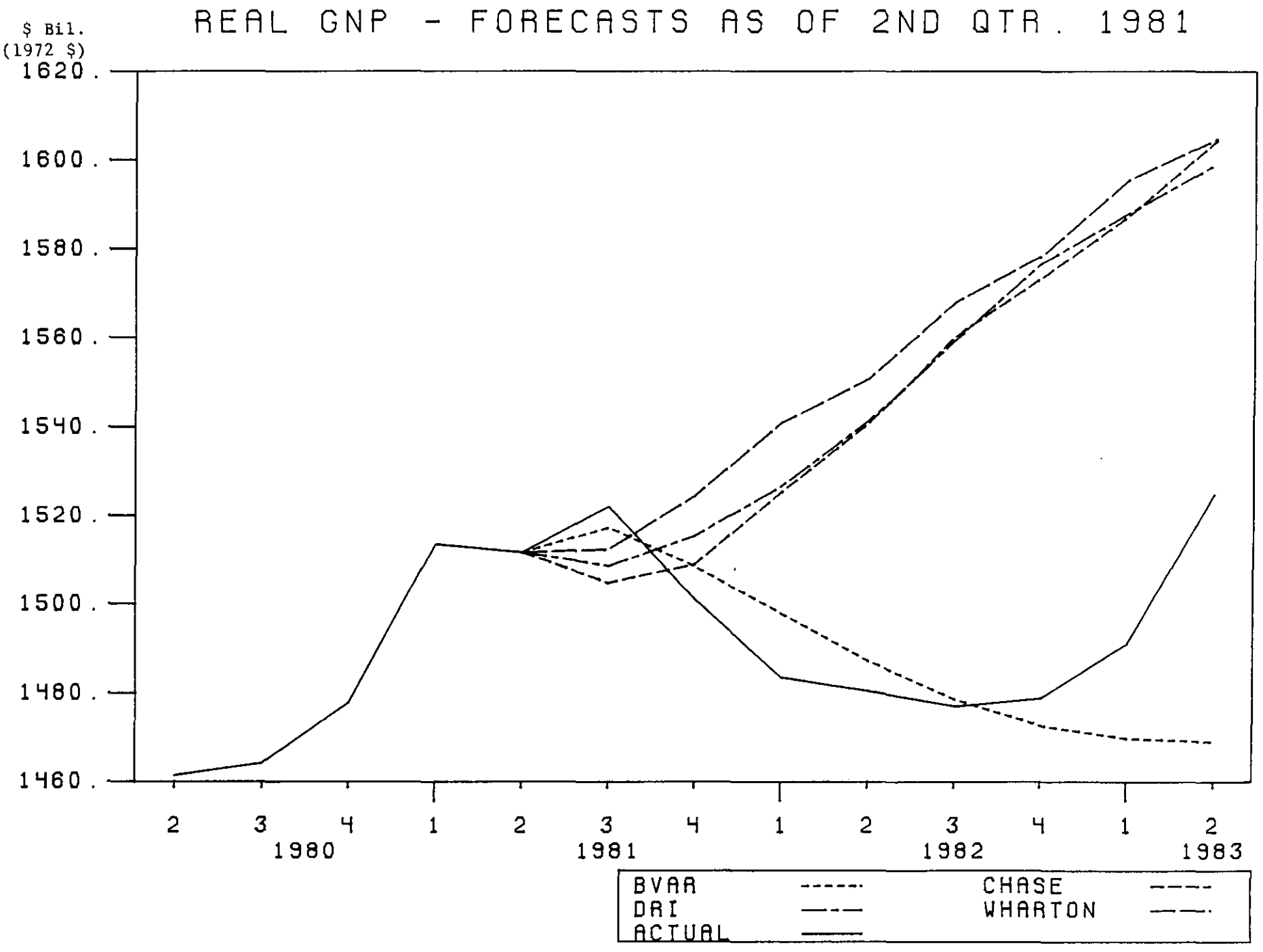

REAL GNP - FORECASTS AS OF 2ND QTR. 1981
$ Bil.
(1972 $)
1620.
1600.
1580.
1560.
1540.
1520.
1500.
1480.
1460.
2
3
4
1
2
3
4
1
2
3
4
1
2
1980
1981
1982
1983
BVAR
DRI
ACTUAL
CHASE
WHARTON

Figure 3

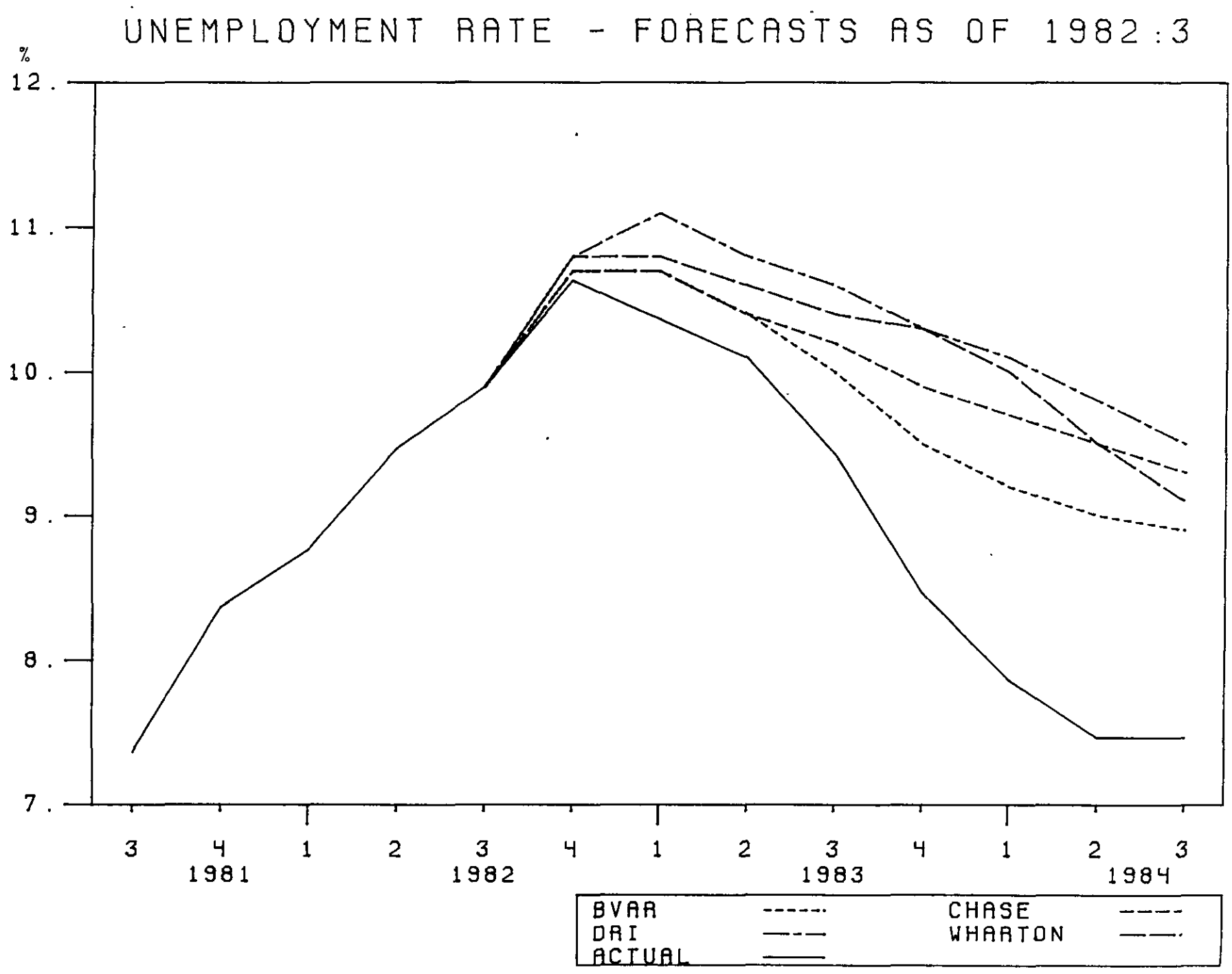

UNEMPLOYMENT RATE - FORECASTS AS OF 1982:3
%
12.
11.
10.
9.
8.
7.
3
4
1
2
3
4
1
2
3
4
1
2
3
1981
1982
1983
1984
BVAR
DRI
ACTUAL
CHASE
WHARTON

Figure 4

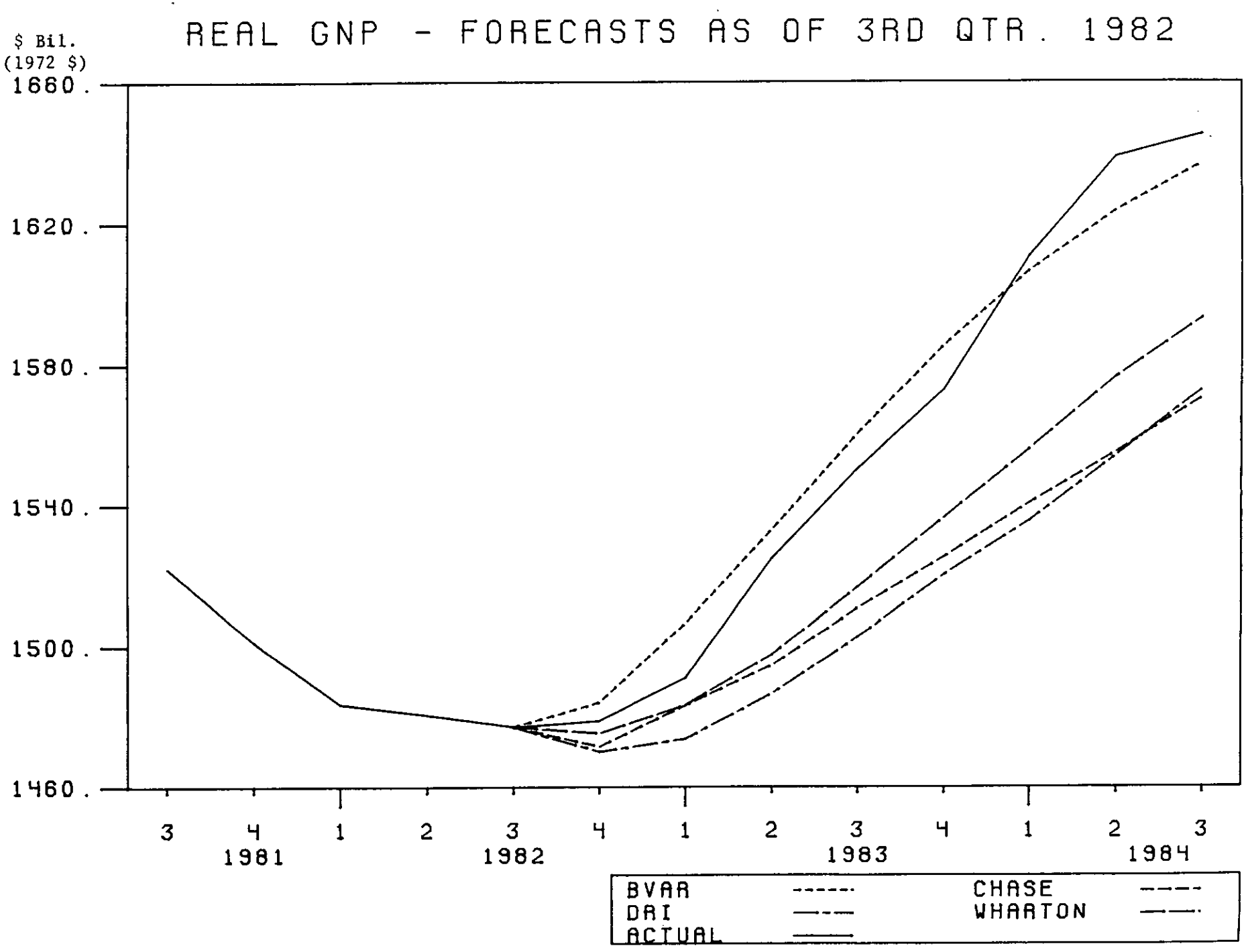

REAL GNP - FORECASTS AS OF 3RD QTR. 1982
$ Bil.
(1972 $)
1660.
1620.
1580.
1540.
1500.
1460.
3
4
1
2
3
4
1
2
3
4
1
2
3
1981
1982
1983
1984
BVAR
DRI
ACTUAL
CHASE
WHARTON

Figure 5

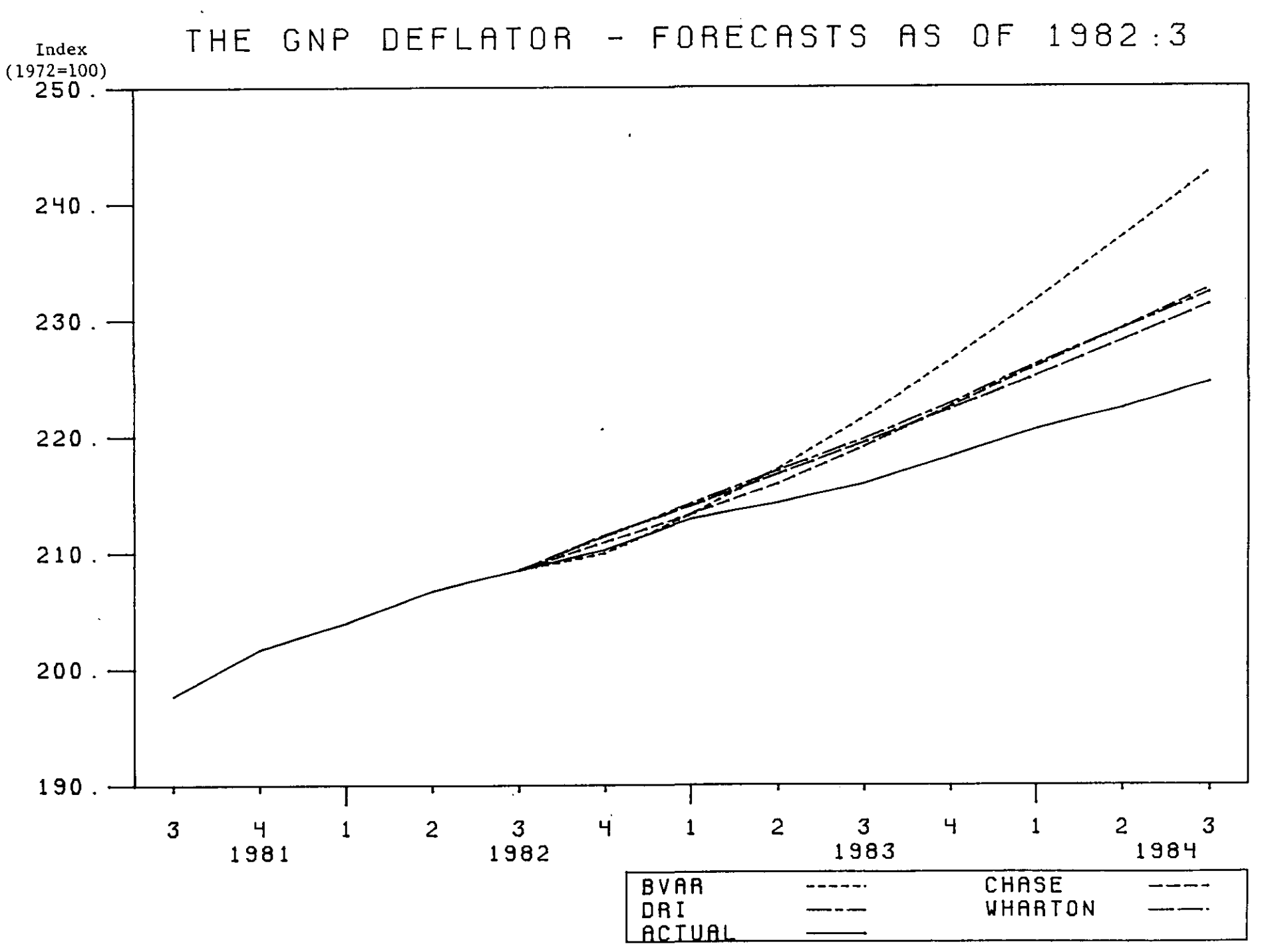
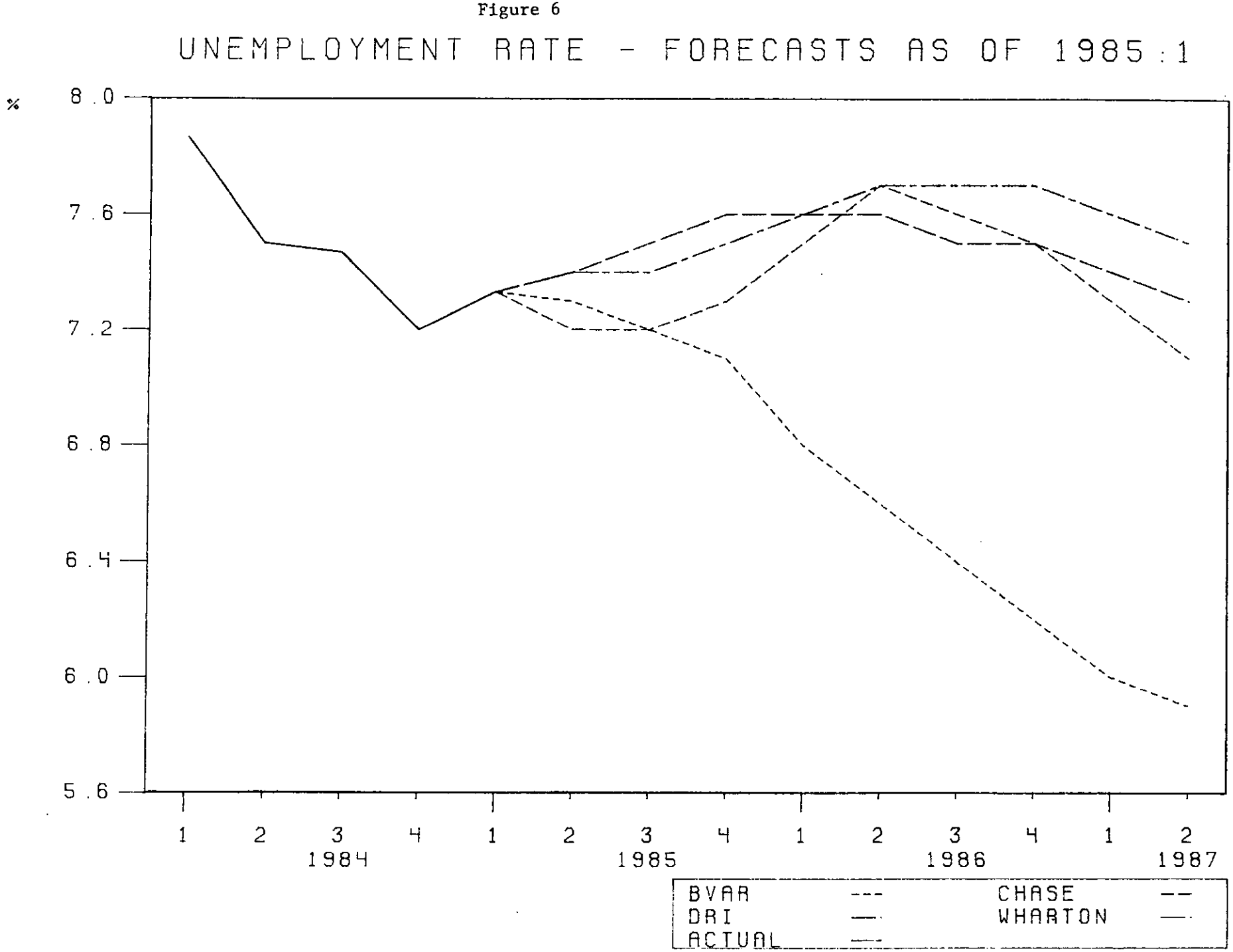

THE GNP DEFLATOR - FORECASTS AS OF 1982:3
Index
(1972=100)
250.
240.
230.
220.
210.
200.
190.
3
4
1
2
3
4
1
2
3
4
1
2
3
1981
1982
1983
1984
BVAR
DRI
ACTUAL
CHASE
WHARTON

Figure 6


UNEMPLOYMENT RATE - FORECASTS AS OF 1985:1
%
8.0
7.6
7.2
6.8
6.4
6.0
5.6
1 2 3 4 1 2 3 4 1 2 3 4 1 2
1984 1985 1986 1987
BVAR
DRI
ACTUAL
CHASE
WHARTON

a few representative forecasts. The deviation of the BVAR forecast from the range of DRI, Chase and Wharton is clear. The reader may also be surprised at how far the actual realized values (the solid line in the figures) are from the range of forecasts. This latter phenomenon illustrates how misleading it can be to follow the common practice of using the range of forecasts as a measure of the range of likely outcomes.

In any case, it should be clear that variance over time in forecasts--or variance with respect to the mean of a distribution of forecasts--is not, in itself, a negative property of a forecasting technique. If the volatility of the forecast represents a correct assessment of the impact of new information, then it is a desirable property. To the extent that a forecasting procedure is too volatile, for example, overly sensitive to new information, that excessive sensitivity will show up as an increased mean square forecast error. We will use this measure of forecast performance to compare the BVAR forecasts with other published forecasts later in this paper.

Another common complaint about BVAR models (and more generally about time series models) is that they never forecast turning points. This criticism is clearly not valid with respect to this BVAR model. Figures 1 through 4 have been chosen specifically to illustrate how turning points in the real economy over the past four years often have been forecast much more accurately by the BVAR model than by the conventional forecasters.

This selective sampling of forecasts cannot provide a basis for judging the relative accuracy of the BVAR technique--that is the subject of the rest of this paper. Nonetheless, lest I leave the wrong impression from this small selection of forecasts, in Figure 5 the outstanding failure of the BVAR model is shown--that is, its projection of accelerating inflation over the past two years. In Figure 6, I show the most recent forecast of the unemployment rate, which again exhibits a substantial difference between the BVAR forecast and the conventional forecasters.

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.

The timing of release of economic forecasts is another important consideration in any forecasting competition. Forecasts are not generally published on 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 were 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 I was 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.

In a recent experiment I found that inclusion of two variables, the value of the trade-weighted dollar and a measure of stock prices, dramatically reduced the out-of-sample forecast errors of the model over the last nine years. In particular, for the one variable which has performed most poorly in the model described here, the GNP deflator, this change in specification reduced the one-year-ahead forecast root-mean-square error by 32 percent. I now include these variables in the model, but it would be unfair to compare the performance of this respecified model with the actual real time performance of others.

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 multistep level-corrected forecasts which can be compared with currently published levels to measure forecast errors. This is the procedure used here. The exact formula is shown below in equation (12).

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, when such input is involved, 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 of 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 forecasts from the BVAR model which I have published for the past five 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 much 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.

Because the model structure has changed recently, however, one cannot expect the forecast performance statistics to apply exactly to the new structure. What one would like to do is to generate procedures which can be expected to give accurate projections of what the performance statistics are likely to be for various model structures. Such a procedure is illustrated below.

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 May 1985. Where observations were not available for one of the forecasters (in a few cases, eight-quarter-ahead forecasts were not published), observations at that horizon and variable 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 the preliminary or the first or second revised NIPA estimates of the previous quarter. In this comparison, which is presented in Table IV, 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 1983 are all included as separate observations 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 used is the familiar root mean square error (RMSE). For the unemployment rate, the RMSE measure of s-quarter-ahead forecast performance is simply

$$\left[\frac{1}{T}\sum_{t=1}^{T}(A_t - {}_sF_t)^2\right]^{1/2} \qquad (11)$$

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

Table IV

BVAR Model Forecast Performance Comparison
Root Mean Square Forecast Errors

Variable	Model	Forecast Horizon in Quarters (number of observations) 1 (60)	2 (58)	3 (55)	4 (52)	5 (49)	6 (46)	7 (43)	8 (38)
Real GNP									
	BVAR	.833	1.095	1.556	1.829	1.785	1.882	2.170	2.957
	Chase	.795	1.367	2.092	2.690	3.124	3.508	3.724	3.576
	DRI	.704	1.208	1.901	2.484	2.940	3.405	3.714	3.648
	Wharton	.696	1.220	1.869	2.472	2.839	3.234	3.546	3.479
GNP Deflator									
	BVAR	.487	1.056	1.874	2.966	4.258	5.571	6.842	8.031
	Chase	.345	.569	.929	1.432	2.080	2.761	3.497	4.211
	DRI	.340	.560	.838	1.328	1.940	2.633	3.421	4.216
	Wharton	.385	.652	1.036	1.555	2.182	2.841	3.595	4.305
Nominal GNP									
	BVAR	.998	1.563	2.577	3.567	4.316	5.130	5.615	6.469
	Chase	.935	1.645	2.575	3.487	4.347	5.301	6.060	6.343
	DRI	.804	1.339	2.198	3.127	4.063	5.137	6.093	6.544
	Wharton	.823	1.565	2.501	3.518	4.449	5.395	6.279	6.637
Unemployment Rate									
	BVAR	.217	.496	.749	.923	1.061	1.110	1.543	1.696
	Chase	.250	.546	.861	1.152	1.441	1.674	1.897	1.938
	DRI	.210	.508	.811	1.161	1.500	1.787	1.997	2.019
	Wharton	.224	.523	.840	1.105	1.359	1.530	1.645	1.674

annualized) in quarter t made r periods earlier, the formula for the RMSE at an s-quarter horizon is

$$\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\left(A_t-\left\{A_{t-s}\cdot\prod_{r=1}^{s}\left(1.+\frac{\overset{*}{F}_{t+r-s}}{100.}\right)\right\}\right)}{A_t}\right]^{1/2} \quad . \qquad (12)$$

Perhaps the most important point to be made in interpreting the results in Table IV is that they are based on a small sample. The number of observations listed under 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 IV because we know they are based on a small, correlated sample, and we have no measures of significance.

Despite this high degree of sampling error in Table IV, 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 range of thousands of dollars per year. The BVAR model can be estimated, and forecasts generated, on a 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 that time series techniques may be useful for very short-term forecasts, but structural models are needed to capture the turning points in business cycles necessary for accurate forecasting at longer horizons. (See, for example, the opinion of L. R. Klein, as quoted in Lupoletti and Webb [undated, p. 7].)

These conclusions would be stronger if we could approximate the distributions of the performance statistics. Unfortunately, there is not much that can be done to model the statistical properties of a short series of judgemental forecasts. For the BVAR forecasts, however, there is an underlying probability model and a reproducible forecasting procedure which can be used to generate a distribution for the measures of expected forecast error variance. Table V shows the actual BVAR performance results (from Table IV) along side a sampling theoretic measure of the mean and standard error of these statistics. These moments are based on simulations of repeated out-of-sample application of the BVAR forecasting technique to artificial data generated from the original estimated probability model. The exact steps involved in this exercise are given in an Appendix.

The standard errors of the simulation RMSE statistics provide at least a rough guide to the uncertainty of these forecast performance measures. In Table V we use the standard error measure to normalize the distance between the BVAR RMSE statistic and that of the best alternative RMSE performance from Table IV for each variable at each horizon. Using this metric we see that

Table V
Monte Carlo Simulation Measures of the Distance Between
the RMSE and the Best Alternative Performance

Variable	Forecast Horizon in Quarters							
	1	2	3	4	5	6	7	8
Real GNP								
BVAR RMSE	.833	1.095	1.556	1.829	1.785	1.882	2.170	2.957
Monte Carlo Simulation Mean	.824	1.271	1.645	1.994	2.360	2.757	3.179	3.593
Monte Carlo Simulation Standard Error	.131	.256	.395	.543	.692	.844	1.023	1.234
Distance (in Standard Errors) from BVAR to Best Alternative	-1.05	.44	.79	1.21	1.52	1.60	1.35	.42
GNP Deflator								
BVAR RMSE	.487	1.056	1.874	2.966	4.258	5.571	6.842	8.031
Monte Carlo Simulation Mean	.459	.875	1.360	1.897	2.461	3.045	3.652	4.256
Monte Carlo Simulation Standard Error	.060	.159	.293	.465	.680	.926	1.195	1.496
Distance (in Standard Errors) from BVAR to Best Alternative	-2.45	-3.10	-3.57	-3.56	-3.40	-3.16	-2.85	-2.55
Nominal GNP								
BVAR RMSE	.998	1.563	2.577	3.67	4.316	5.130	5.615	6.469
Monte Carlo Simulation Mean	.980	1.611	2.209	2.760	3.285	3.818	4.391	4.942
Monte Carlo Simulation Standard Error	.149	.276	.434	.640	.887	1.174	1.436	1.716
Distance (in Standard Errors) from BVAR to Best Alternative	-1.30	-.81	-.87	-.69	-.29	.01	.31	-.07
Unemployment Rate								
BVAR RMSE	.217	.496	.749	.923	1.061	1.110	1.543	1.696
Monte Carlo Simulation Mean	.297	.514	.693	.834	.959	1.088	1.220	1.367
Monte Carlo Simulation Standard Error	.051	.122	.199	.268	.332	.398	.469	.544
Distance (in Standard Errors) from BVAR to Best Alternative	-.13	.10	.31	.68	.90	1.06	.22	-.04

the most significant difference occurs with respect to inflation in which case for all horizons the BVAR model performance is more than two standard errors worse than the best alternative. On the other hand, for real GNP the BVAR model performs more than one standard error better than the best alternative at the four through seven quarter horizons. For nominal GNP these effects offset and the BVAR performance is somewhat worse at shorter horizons and a little better at longer horizons. For the unemployment rate the BVAR performs better than the best alternative for the two through seven quarter horizons, with the magnitude of the difference reaching one standard error at the six step horizon.

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, and therefore to reflect the relative information content of the BVAR forecast. One measure which does reflect that tendency of other forecasts to bunch together is the proportion of times a given forecaster is closest to the actual. By this measure the results clearly favor the BVAR model. Of the 1,604 forecasts considered, the BVAR model was most accurate 34.8 percent of the time. The percent of times each of the other forecasters was most accurate was 16.4, 27.3, and 21.6, for Chase, DRI, and Wharton, respectively.

Postscript

Over the five 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 [1984a]; and Doan, Litterman, Sims [1984].) The Federal Reserve Bank of Minneapolis has developed a larger (46-equation) monthly national forecasting model (Litterman [1984d]); several regional BVAR models have been developed; (Amirizadeh and Todd [1984]) 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 [1984b]).

Appendix

The bootstrap procedure used to estimate standard errors of the forecast performance statistics is as follows:

I. Bayesian vector autoregressive system is estimated over the base period, 1949:3 through 1980:1.

II. Each quarter from 1980:1 through 1985:1, a one-step-ahead forecast is made for each variable, the forecast errors are saved, and the equation estimates are updated using the Kalman filter. The final coefficient estimates are saved for use in generating artificial data.

III. One hundred simulations are performed. In each simulation the following steps are taken:

   A. Artificial data is generated based on the probability structure estimated in steps I and II.

      1. One hundred and twenty-three (the number of observations in the base period) uniform random integers I_i, $i = 1, \ldots, 123$, were drawn from the interval [1,123].

      2. Artificial data is generated using shocks randomly drawn from the base period residual vectors. Initial conditions are taken to be those as of 1949:2. Then each period, t, from 1949:3 through 1980:1 a new observation is obtained as a sum of the forecast based on the estimated coefficients plus the vector of residuals from period $I_{[t-1949:3\ +\ 1]}$.

3. A similar procedure is used to generate artificial data for the forecast period. Here 22 random integers J_i, i = 1, ... 22, are drawn from the interval [1,...,22]. The artificial data through 1980:1 is appended with 22 additional observations obtained as the sum of the forecast at time t plus the vector of forecast errors from the period $J_{[t-1980:1 + 1]}$.

B. A new Bayesian vector autoregressive system is estimated on the artificial data using observations 1949:3 through 1980:1.

C. Each quarter of the forecast period a one-step-ahead forecast is made of the artificial data, the forecast errors are saved, and the coefficient estimates are updated using the Kalman filter.

D. The root-mean-square-error (RMSE) statistics for each variable at each horizon over the 22-quarter forecast period are calculated.

IV. The mean and standard error across simulations of the RMSE statistics are calculated.

This procedure gives a Monte Carlo measure of the uncertainty of the RMSE statistic obtained when the Bayesian forecasting procedure is applied to an 22-quarter sample of artificial data generated with a probability structure estimated from the actual data.

References

Brunner, Karl, ed. 1981. The Great Depression Revisited. Boston: Martinus Nijhoff Publishing.

Doan, Thomas; Robert Litterman; and Christopher Sims. 1984. "Forecasting and Conditional Projection Using Realistic Prior Distributions." Econometric Reviews 3 (1), pp. 1-144.

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

Gordon, Robert J., and James A. Wilcox. 1981. "Monetarist Interpretations of the Great Depression: An Evaluation and Critique." In The Great Depression Revisited, ed. Karl Brunner, pp. 49-107. Boston: Martinus Nijhoff Publishing.

Jorgenson, Dale W. 1963. "Capital Theory and Investment Behavior." American Economic Review 53 (May): 247-59.

Leamer, Edward E. 1972. "A Class of Informative Priors and Distributed Lag Analysis," Econometrica 40, pp. 1059-81.

Litterman, Robert B. 1980. "Techniques for Forecasting Using Vector Autoregressions." Ph.D. diss., University of Minnesota.

_____. 1982. "Optimal Control of the Money Supply." Federal Reserve Bank of Minneapolis Quarterly Review 6 (Fall): 1-9.

_____. 1984a. "Specifying Vector Autoregressions for Macroeconomic Forecasting." Research Department Staff Report 92, Federal Reserve Bank of Minneapolis. Also forthcoming in Bayesian Inference and Decision Techniques With Applica-

tions: Essays in Honor of Bruno de Finetti, ed. P. Goel. Amsterdam: North-Holland.

_____. 1984b. "The Costs of Intermediate Targeting." Research Department Working Paper 254, Federal Reserve Bank of Minneapolis.

_____. 1984c. "Above-Average National Growth in 1985 and 1986." Federal Reserve Bank of Minneapolis Quarterly Review 4 (Fall): 3-7.

_____. 1984d. "Forecasting and Policy Analysis With Bayesian Vector Autoregression Models." Federal Reserve Bank of Minnepaolis Quarterly Review 4 (Fall): 30-41.

_____. 1980b. "A Bayesian Procedure for Forecasting With Vector Autoregression." MIT Working Paper.

Litterman, Robert B., and Thomas M. Supel. 1983. "Using Vector Autoregressions to Measure the Uncertainty in Minnesota's Revenue Forecasts." Federal Reserve Bank of Minneapolis Quarterly Review 7 (Spring): 10-22.

Lucas, Robert E., Jr., and Thomas J. Sargent. 1978. "After Keynesian Macroeconomics," in After the Phillips Curve: Persistence of High Inflation and High Unemployment, Federal Reserve Bank of Boston.

Lupoletti, William M., and Roy H. Webb. Undated. "Defining and Improving the Accuracy of Macroeconomic Forecasts: Contributions From a VAR Model." Federal Reserve Bank of Richmond, Photocopy.

Meltzer, Allan H. 1981. "Comments on 'Monetarist Interpretations of the Great Depression.'" In The Great Depression Revisited, ed. Karl Brunner, pp. 148-64. Boston: Martinus Nijhoff Publishing.

Sims, Christopher A. 1982. "Policy Analysis with Econometric Models." In Brooking Papers on Economic Activity, ed. William C. Brainard and George L. Perry, 1: 107-52. Washington, D.C.: Brookings Institution.

Shiller, Robert J. 1973. "A Distributed Lag Estimator Derived from Smoothness Priors," Econometrica 41, pp. 775-88.

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

Todd, Richard M. 1984. "Improving Economic Forecasting With Bayesian Vector Autoregression." Federal Reserve Bank of Minneapolis Quarter Review 4 (Fall): 18-29.

Zellner, Arnold. 1971. An Introduction of Bayesian Inference in Econometrics, New York: John Wiley and Sons, Inc.