Federal Reserve Bank of Minneapolis
Research Department Working Paper

FORECASTING AND CONDITIONAL PROJECTION USING REALISTIC PRIOR DISTRIBUTIONS

By Thomas Doan, Robert Litterman
and Christopher Sims

Working Paper 243
PACS File 3125

August 1983

NOT FOR DISTRIBUTION
WITHOUT AUTHORS' APPROVAL

ABSTRACT

This paper develops a forecasting procedure based on a Bayesian method for estimating vector autoregressions. The procedure is applied to ten macroeconomic variables and is shown to improve out-of-sample forecasts relative to univariate equations. Although cross-variables responses are damped by the prior, considerable interaction among the variables is shown to be captured by the estimates.

We provide unconditional forecasts as of 1982:12 and 1983:3. We also describe how a model such as this can be used to make conditional projections and to analyze policy alternatives. As an example, we analyze a Congressional Budget Office forecast made in 1982:12.

While no automatic causal interpretations arise from models like ours, they provide a detailed characterization of the dynamic statistical interdependence of a set of economic variables, which may help in evaluating causal hypotheses, without containing any such hypotheses themselves.

The views expressed herein are those of the author and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System. The material contained is of a preliminary nature, is circulated to stimulate discussion, and is not to be quoted without permission of the author.

Introduction

We approach the analysis of a group of economic time series as the problem of using a prior joint distribution for the observed values of the series with future values to obtain a posterior distribution for future data conditional on observed data. The methods we suggest are Bayesian in spirit. We do not, however, attempt to make our prior distributions fully reflect our personal a priori knowledge and uncertainty. Instead we aim at a prior distribution which is easily standardized and reproduced by other researchers and reflects aspects of prior distributions which are likely to be similar across many researchers. The posterior distribution produced by our analysis is, of course, just the likelihood function weighted by the prior p.d.f. Our methods can be thought of as a way of reporting the likelihood function to other researchers; it provides a report more useful than the unweighted likelihood function itself for researchers who themselves put little prior probability on regions of the parameter space given low probability by our prior.

We regard conventional methods of developing probability models for econometric time series as unreliable because they do not give probabilistic treatment to the uncertainty arising from researchers' inexact knowledge of the true "model specification." Conventional approaches produce models which can be helpful adjuncts to judgment in producing forecasts, but the implied probability distributions about the forecast which such models generate are almost invariably too optimistic. (The ideas in these first two paragraphs are discussed at more length in Sims (1982) and Litterman (1982).)

Specifying a joint distribution over the hundreds or thousands of interrelated data points available in most applications is a complex task. Any explicit joint probability model we may write down is likely to contain hidden implications which we would reject if we confronted them. Yet there is no joint distribution representing "ignorance" on which we can rely as in some sense conservative. For example, if we take a large-variance joint normal prior on the coefficients of an unrestricted vector autoregressive model for the data as representing ignorance, we are in fact putting high probability on models with very large coefficients. Such models produce erratic, poor forecasts and imply explosive behavior of future data. Most researchers would think it unlikely that such models actually characterize the data, yet use of nonBayesian estimation methods is roughly equivalent to use of the flat priors which put high probability on such models. This is why those making practical use of nonBayesian methods are forced to impose arbitrary or conventional restrictions to simplify their models, eliminating many parameters which it must be admitted are not known to be zero.

Since we and the profession as a whole have little experience with specifying joint distributions for these contexts, in this paper we experiment with a range of prior distributions. The range we consider is indexed by a set of eight parameters. Our view is that a good standard public prior may well be some weighted average of the priors indexed by these parameters. Since the priors with all parameters fixed are much more tractable than would be a weighted integral over the parameters, our hope is that

we will emerge with evidence that for many purposes it will be possible to obtain good results with a single setting of the parameters, without making the extensive explorations which underlie this paper's results. This would occur if over a wide range of reasonable settings for the parameters the model generated similar conditional distributions of the future given past values of the variables in the system.

Another possibility is that, while conditional distributions are sensitive to the parameter setting, the data are fit well only by parameter values in a certain narrow range, and within this range conditional distributions of the future are all similar. This would imply that, though we need to search to find a good parameter vector, we can then generate conditional distributions with a single "good" vector. The inconvenience of having to compute many such conditional distributions and then take weighted averages of the results would be avoided.

While our explorations are in some ways like fitting the parameters of a conventional model--we examine various points in a parameter space and check how well the resulting models fit the data--the motivation and implications of the results are different in important respects. Our ideal conclusion would be that the parameters are "ill-determined"--that the fit is similar across a wide range of parameter settings which all have similar implications.

Of course, there is a question as to what we mean by a particular prior's "fitting" the data well or badly. The Bayesian interpretation is that we have specified our prior incompletely.

The usual Bayesian formulation has a model for the data y specified as a density function $p(y|\theta)$ for y conditional on parameters θ, yielding a joint density for y and θ as the product $p(y|\theta)q(\theta)$, where q is a prior density on the parameters θ. We are introducing an extra layer of parameterization. We specify a model for the data conditional on parameters θ which we call "coefficients". We specify a prior over θ conditional on a second set of parameters π, so that our joint density for the data and the coefficients conditional on π is $p(y|\theta)q(\theta|\pi)$. We leave inexplicit our prior over π, which we need to fully specify the probability distribution of the data. We can in principle integrate $p(y|\theta)\ q(\theta|\pi)$ with respect to θ to obtain the marginal distribution for y given π, which we could call $m(y|\pi)$. If we are not directly interested in θ, we can treat $m(y|\pi)$ as our model for the data. For a fixed set of observed data y, the behavior of $m(y|\pi)$ as a function of π plays the formal role of a likelihood function. As usual in such a context, if our prior density is flat in the region where $m(y|\pi)$ is large, our posterior p.d.f. for π will be proportional to $m(y|\pi)$ and we can think of ourselves as making inferences about the likely values of π. But since here π is interesting mainly for its implications about θ, we do not focus inference on "estimating" π.

Our posterior p.d.f. on θ, for a fully specified prior, would be obtained by forming the marginal joint p.d.f. for θ and y by integrating over π, then applying Bayes' rule. In the case where our prior p.d.f. on π is flat in the relevant region, this leads to a posterior p.d.f. for θ which is a weighted average of

those obtained conditional on π, with the relative weight on π given by $m(y|\pi)$. Thus, when we measure the fit of the model we ought naturally to use the relative size of $m(y|\pi)$. This is formally much like using the likelihood function, and we will occasionally henceforth refer to $m(y|\pi)$ as the likelihood, but it is nonetheless a Bayesian notion, since it is derived by taking the coefficients θ as a priori random.

In fact, we shall see that this Bayesian notion of how well a prior fits the data corresponds to measuring the fit by forecasting performance. That is, with a particular setting of π, we can generate recursively through the sample one-step-ahead forecasts of data at $t + 1$ given data up through t. The measure of fit based on our Bayesian likelihood turns out under our assumptions to be a weighted sum of squares of the one-step-ahead forecast errors. Readers uncomfortable with the Bayesian terminology can think of what we are doing as using π to index forecasting procedures, choosing among procedures by how well they forecast in the sample period. From this perspective, we are taking the large parameter space indexed by θ and reducing it to a smaller one indexed by π. What we are doing is quite different, however, from the conventional "parsimonious parametrization" approach, which would use some subspace of the θ-space, judiciously chosen, as if it were the whole parameter space. Our approach will, for any given choice of π, allow the θ used in forecasting to be more and more strongly data-determined as data accumulates through time, with no subspaces of the θ-space ruled out.

The Forecasting Procedure

The procedures we are about to describe in detail were developed in Litterman (1980, 1981, 1982) and Sims (1980, 1982). Though the procedures are described in general terms, it may help to bear in mind that we will be applying them to a specific set of data. We consider a set of ten variables, measuring output, prices, money, federal government revenues and outlays, stock prices, interest rates, the value of the dollar, the flow of total nonfinancial debt, and the change in business inventories. The data are described fully in the Appendix. Observations begin in 1948:1 and end as of 1983:3. All variables are logged except changes in business inventories and the interest rate; all variables are seasonally adjusted except the interest rate, stock price index, and the trade-weighted dollar; none of which show evidence of a seasonal pattern.

Starting from an unrestricted, time-varying, m'th-order vector autoregressive representation for the n-vector, X,

$$(1) \qquad X_t = A_t(L)\, X_{t-1} + C_t + \varepsilon_t \ ,$$

where $A_t(L)$ is for each t a polynomial of order m in strictly positive powers of the lag operator L, we express our prior separately for each equation as a distribution over the coefficients in A and C. In principle we should also treat the variance of ε_t as uncertain, but instead this is treated as one of the parameters of our prior. Our approach can be thought of as imposing "fuzzy" restrictions on the equation, striking a balance between decreasing variance and increasing bias as the restrictions are tightened

up. What we do thus has antecedents in the literature on shrinkage estimation and its Bayesian interpretation, for example, the works by Hoerl and Kennard (1970), Stein (1974), Shiller (1973) and Leamer (1972, 1978).

The prior is specified as a multivariate normal distribution for the coefficients of the vector autoregression. We refer to changes in the parameters of the prior which lead to smaller (larger) variances of coefficients as tightening (loosening) the prior. The prior means for all coefficients are zero, except for a mean of one at the first lag of the dependent variable in each equation. Thus, in the limit as the prior is tightened around its mean each equation takes the form of a random walk.

$$x_t = x_{t-1} + \varepsilon_t \tag{2}$$

Because most of the variables we use have persistent trends, we always keep the prior for the constant in each equation flat in the relevant region of the parameter space, so the limiting form for each equation is essentially a random walk with drift fit to the data.

$$x_t = x_{t-1} + c + \varepsilon_t \tag{3}$$

While we recognize that a more accurate representation of our prior beliefs would give less weight to systems with explosive roots than is implied by our symmetric distributions around this mean, we doubt that the gain that could be achieved by abandoning the Gaussian form for our prior would be worth the price. In

particular, the likelihood function for data which is not exploding will be quite clear in its rejection of roots significantly outside the unit circle.

We denote by θ_t the parameter vector obtained by stacking up all the coefficients of the vector autoregression. The initial vector θ_0 is given a multivariate normal prior density function with mean $\bar{\theta}$. The covariance matrix of the prior, denoted Σ_0, is generated as a function, F, of a vector of prior parameters, π. Thus, at time 0, we have.

(4) $$\Sigma_0 = F(\pi)$$

(5) $$\theta_0 \sim N(\bar{\theta},\Sigma_0).$$

We postulate change in the coefficients of the autoregression over time according to

(6) $$\theta_t = \pi_8 * \theta_{t-1} + (1-\pi_8) * \bar{\theta} + \mu_t .$$

The parameter π_8 controls the rate of decay toward the prior mean. When it is set to 1, as in a number of our experiments, we are modeling the coefficient variation as a random walk. The random change in the parameter vector, μ_t, is assumed to be drawn from a distribution with zero mean, and covariance matrix proportional to Σ_0.[1/]

[1/] Except that the variance in changes in the constant term is kept equal to the variance of changes in the coefficient on the first own lag, rather than set proportional to the effectively infinite prior variance on the constant term.

The factor of proportionality, π_7, which scales Σ_0 to determine the covariance matrix of μ_t, determines the amount of time variation allowed in the parameter vector.

Having specified the probability model, we apply the Kalman filter equation by equation to obtain recursively posterior modes $\hat{\theta}_t$ for θ_t based on data through t-1. When we have passed through the full sample this way, we end up with a value for the likelihood of the sample and a full-sample estimate of the parameter vector applying at the first post-sample date.

The Kalman filter is easiest to understand for the case where the prior is normal with a fixed covariance matrix and the equation disturbance terms ε_t have known variance. In practice, of course, we do not know the equation disturbance variances a priori. Our procedure is to use σ^2, .9 times the vector of variances of residuals in a univariate autoregressions of order 6, as if it were exactly the vector of variances of equation disturbances for the multivariate system. The results of the filter depend only on the ratios of equation disturbance variances to the elements of Σ_0. We can examine how the likelihood[2/] value behaves as a function of σ_i^2 for equation i, keeping all parameter estimates unchanged. At the highest value of the likelihood, we have the modal estimate of equation disturbance variance for the fixed ratio of σ_i^2 to Σ_0 which generated our estimates, assuming a flat prior on σ_i^2. We will also have an implied rescaling of both our initial σ_i^2 and our initial Σ_0.

[2/] The reader is reminded of our special use for the term likelihood: the marginal p.d.f. for the data conditional on the parameters π.

The value of the likelihood for the rescaled σ_i^2 and Σ_0 at which the posterior p.d.f. is maximized provides a natural measure of fit for individual equations. The likelihood measure is a kind of estimate of the one-step ahead forecast standard error. Scaled to have units of standard errors of one-step ahead prediction errors, it is given by

$$(7) \qquad \left\{T^{-1} \sum_{t=1}^{T} \frac{{}_t\bar{\varepsilon}_{t+1}^2}{(s_{it}^2/s_i^2)}\right\}^{.5} ,$$

where ${}_t\bar{\varepsilon}_{t+1}$ is the one-step ahead forecast error; s_{it}^2 is the theoretical prediction error for equation i at t, s_i^2 is the sample geometric mean of s_{it}^2. It differs from the ordinary root mean square forecast error in that it weights squared forecast errors by the inverses of their theoretical variances. The theoretical variances vary across observations because the component of forecast error variance due to parameter change depends on the values of the independent variables.

We have not seriously explored the potential gains from treating the equations of the system jointly. Least squares equation by equation is fully asymptotically efficient for an unconstrained vector autoregression, because the same variables appear on the right-hand side of each equation. The Bayesian posterior mode is not correctly captured by single equation methods, however, even if priors are normal and independent across equations, unless the prior covariance matrices are proportional to one another and the same multiple of equation disturbance variance in each equation. Furthermore, in a system as large (10

variables) as the one we examine in this paper, there are many (55) free parameters in the disturbance covariance matrix, all of which affect the posterior distribution. It is likely that by imposing an informative prior on the 600 coefficients on lagged variables while using a "flat" prior on the 55 parameters of the covariance matrix we are missing an avenue for improving reliability of these methods.[3/]

The single-equation measures of fit which emerge naturally from the Kalman filter have a multivariate analogue, but it cannot be computed without using a multivariate version of the Kalman filter. We have therefore put primary emphasis on a different class of multivariate measures of fit, the log-determinants of matrices of cross products of k-step ahead out-of-sample forecast errors. The likelihood measure of fit would differ from one based on the determinant of the cross product of one-step ahead forecasts mainly in weighting the errors by the inverses of their conditional variances.

The log-determinants of the matrices of summed cross-products of k-step out-of-sample forecast errors which we rely on as our primary measures of fit are defined by

---

[3/]We could parameterize the model initially in recursive form, with the j'th equation expressing X_{jt} as a linear function of lagged X_t's and current X_{it}'s for $i < j$ and the covariance matrix of equation disturbances specified as diagonal. In such a model single-equation procedures would coincide with multiple-equation procedures because of the diagonality of the disturbance matrix, and most of the free parameters of the covariance matrix of residuals would become coefficients on right-hand-side variables. The difficulty with this approach is that normal prior distributions on coefficients in such a recursive system cannot be chosen to treat variables symmetrically. The potential advantages of including contemporaneous relations among disturbances in the prior distribution are great enough that this approach should be explored, however.

(8) $${}_t\hat{\varepsilon}_{t+k} = ({}_t\hat{X}_{t+k} - X_{t+k}),$$

(9) $$E_k = \sum_{s=1}^{T} ({}_s\hat{\varepsilon}_{s+k}\ {}_s\hat{\varepsilon}'_{s+k})$$

(10) k-step-ahead log-determinant = $\log(|E_k|)$.

Six parameters determine the general form of our function, F. The parameters and their roles are as follows:

parameter	controls:
π_1	relative tightness on own lags
π_2	relative tightness on lags of other variables
π_3	relative tightness on constant term
π_4	differential tightness among other variables
π_5	overall tightness
π_6	tightness on sums of coefficients

Let the ith component of X, x^i have the scalar representation:

(11) $$\begin{aligned} x_t^i = {} & a_{1,1}^i x_{t-1}^1 + a_{1,2}^i x_{t-2}^1 + \dots + a_{1,m}^i x_{t-m}^1 \\ & + a_{2,1}^i x_{t-1}^2 + a_{2,2}^i x_{t-2}^2 + \dots + a_{2,m}^i x_{t-m}^2 \\ & + a_{n,1}^i x_{t-1}^n + a_{n,2}^i x_{t-2}^n + \dots + a_{n,m}^i x_{t-m}^n + c^i + \varepsilon_t \end{aligned}$$

The first five components of π, together with the elements of σ^2 and a set of relative weights, ω_j^i, for $i=1,\dots,n$; $j=1,\dots,n$, define a diagonal matrix of variances for the coefficients. For coefficients of own lags, that is, $a_{i,k}^i$, $k = 1,2,\dots,m$, we assume the variance is given by

$$(12) \qquad \mathrm{Var}\,(a^i_{i,k}) = \frac{\pi_5 \cdot \pi_1}{k \cdot \exp(\pi_4 \cdot \omega^i_i)}$$

For lags of other variables in a given equation, that is, $a^i_{j,k}$, k = 1,2,...,m, and i not equal to j, we assume the variance is given by

$$(13) \qquad \mathrm{Var}\,(a^i_{j,k}) = \frac{\pi_5 \cdot \pi_2 \cdot \sigma^2_i}{k \cdot \exp(\pi_4 \cdot \omega^i_j) \cdot \sigma^2_j} \quad \text{for not equal to } j.$$

For the constant term in each equation we assume the variance is given by

$$(14) \qquad \mathrm{Var}\,(c^i) = \pi_5 \cdot \pi_3 \cdot \pi_2 \cdot \sigma^2_i \,.$$

The σ^2_i scale factors are present to take account of the units of the data in determining the prior tightness for coefficients on different variables.

The relative weights, ω^i_j , are a set of numbers which we specify to reflect our a priori knowledge about the likelihood that lags of variable j will have nonzero coefficients in equation i. The larger is ω^i_j ,the closer to zero we feel that coefficient is likely to be. For most of the variables we have specified ω^i_i equal to 0 and ω^i_j equal to 1 for i not equal to j. For the interest rate and the trade-weighted dollar we specified w^i_i equal to 1 and ω^i_j equal to 2 for i not equal to j. These weights, relative to the others, reflect our belief that these variables are a priori more likely to behave like random walks. Finally, for the stock price index we specified w^i_i equal to 1 and w^i_j equal to 5 to reflect our strong belief that this variable behaves like a random walk.

Given the above tightnesses on individual coefficients, based on π_1 through π_5, we also wished to impose a prior belief that the sums of coefficients on own lags are close to 1, and on lags of other variables are close to 0. This does not affect the mean of our prior. Consider a diagonal block of variances, M, for a vector of coefficients, θ, on lags of variable j in equation i, defined by parameters π_1 through π_5. Let the vector S be defined by

$$(15) \qquad S = \left\{\frac{\pi_6 \cdot \sigma_i}{\sigma_j}\right\} \cdot [1\ 1\ \ldots\ 1] \ .$$

Then following the heuristic logic of Theil's mixed estimation procedure, we can introduce a "dummy observation" of the form

$$(16) \qquad S\theta = v \ ,$$

with the variance of v set to one, by modifying M to take the new form

$$(17) \qquad N = M \left\{\frac{MSS'M}{1+S'MS}\right\}$$

Improving Forecast Accuracy

Before we search over the prior parameters, we generate a set of benchmark univariate, fixed-coefficient, autoregressions. Based on the results in Litterman (1982), which viewed out-of-sample forecast performance as a function of lag length for many of these variables, we chose to include six lags in each equation and a constant term. For this set of equations, and all subsequent specifications, we calculate sets of 1, 3, 6, and 12-step ahead forecast errors for each month from 1951:1 through

1980:12. We compute log-determinant measures of fit as well as standard errors for each variable, and we look at three ten-year sub-periods, as well as the overall fit in order to gauge the consistency of the results. The overall measure of forecast accuracy to which we give primary attention is the full-period log-determinant of the covariance matrix of one-step ahead forecast errors. The univariate results are presented in Table 1.

Table 1

Univariate Forecasting Performance

Variable	Period 1	Period 2	Period 3	Overall
1-Step Horizon				
Standard Errors				
Real GNP	0.965	0.796	0.818	0.863
GNP deflator	0.204	0.136	0.208	0.186
M1	0.292	0.321	0.498	0.381
Stock Price Index	2.895	3.280	3.695	3.306
Treasury bill rate	0.231	0.233	0.871	0.537
Trade weighted dollar	0.295	0.355	1.981	1.174
Flow of total debt	18.863	8.945	7.155	12.741
Change in inventories	4.967	4.981	6.531	5.542
Federal outlays	5.209	3.839	4.383	4.512
Federal receipts	4.300	4.330	3.905	4.183
Log determinant	-69.035	-70.741	-62.819	-64.395
3-Step Horizon				
Standard Errors				
Real GNP	1.492	1.059	1.337	1.309
GNP deflator	0.506	0.368	0.495	0.460
M1	0.651	0.616	1.015	0.782
Stock Price Index	5.901	6.608	7.532	6.714
Treasury bill rate	0.502	0.588	1.849	1.157
Trade weighted dollar	0.562	0.753	4.429	2.614
Flow of total debt	40.328	16.797	14.865	26.642
Change in inventories	5.868	5.507	8.185	6.627
Federal outlays	5.647	4.819	4.532	5.022
Federal receipts	5.279	5.263	4.178	4.934
Log determinant	-58.277	-60.593	-52.840	-53.523
6-Step Horizon				
Standard Errors				
Real GNP	2.549	1.490	2.035	2.071
GNP deflator	0.968	0.673	0.885	0.851
M1	1.151	1.137	1.303	1.208
Stock Price Index	9.598	10.751	11.487	10.641
Treasury bill rate	0.910	0.902	2.410	1.576
Trade weighted dollar	0.882	1.060	6.811	4.012
Flow of total debt	46.127	21.069	18.809	31.227
Change in inventories	6.700	5.866	9.282	7.427
Federal outlays	9.336	6.986	6.224	7.631
Federal receipts	8.073	6.090	4.932	6.496
Log determinant	-50.336	-54.638	-47.249	-46.909

12-Step Horizon

Standard Errors				
Real GNP	4.564	2.749	3.176	3.581
GNP deflator	2.880	1.349	1.722	2.088
M1	2.335	2.041	1.875	2.093
Stock Price Index	17.924	14.725	15.505	16.109
Treasury bill rate	1.184	1.340	2.717	1.878
Trade weighted dollar	2.315	1.275	9.811	5.866
Flow of total debt	64.439	32.025	24.778	43.939
Change in inventories	7.680	6.562	10.340	8.346
Federal outlays	16.970	11.555	10.316	13.265
Federal receipts	19.904	10.006	7.516	13.574
Log determinant	-39.560	-48.346	-42.100	-38.284

The extent of our investigation of different settings of the π vector was constrained by the expense of evaluating the forecast performance for each value. Although our calculations were performed on a Cray-1 computer at the University of Minnesota which is both extremely fast and inexpensive, each evaluation of forecast performance for a given value of π required approximately 60 seconds and cost about $30. About half of the time for a given run was involved in the recursive estimation of the $\hat{\theta}_t$'s, the rest was used in generating the 12-step ahead forecasts each period and doing the accounting necessary to generate forecast accuracy statistics.

We chose to focus primarily on two dimensions of the prior, the overall tightness and the degree of time variation of the parameters. Our previous experience with priors of this form has suggested that the degree of parameterization of an equation is an important determinant of forecast accuracy. Viewing the specification of a forecasting equation as the construction of a signal extraction filter, it is clear that equations with too many free parameters tend to pick up excess noise and to generate poor

out-of-sample forecasts. Equations with too few parameters fail to pick up the signal. The specification of a prior provides a flexible format through which one can confront the tradeoff between increasing signal extraction capabilities and over-fitting the data. By adjusting the tightness of the prior, one can tune the filter along this dimension.

We focus on the forecast performance as a function of the amount of time variation in order to investigate the degree to which results might be improved by relaxing the usual assumption of constant coefficients. We hope not only to increase forecast accuracy, but also to generate a more realistic description of the uncertainty of forecasts, particularly of those at multi-step horizons.

As a first step in this investigation we focused on how much improvement in forecasting would be possible by searching along these two dimensions. Taking as given the parameter values:

$$\pi_1 = .05$$
$$\pi_2 = .001$$
$$\pi_3 = 10^5$$
$$\pi_4 = 2$$

we began by minimizing the one-step ahead log-determinant as a function of π_5 and π_7. An informal search requiring about 50 function evaluations led us to the values $\pi_5 = 1.4$ and $\pi_7 = .23 \cdot 10^{-7}$.

Up to this point we had not yet begun to constrain the sum of coefficients or to allow decay of the parameter esti-

mates. In effect we had, by default, set $\pi_6 = 0$ and $\pi_8 = 1$. Over the range we examined, forecast performance varied little as we changed π_5 and π_7. It was clear, though, that for these parameters we had found values in a neighborhood of no more than a few percent from the point at which our one-step log-determinant measure was minimized.

The amount of parameter variation allowed at this specification is small. The implied standard error of the change in the first own lag, for example, over the entire sample is on the order of .001. Since the prior mean of this parameter is 1, parameter drift might be taken as negligible. This result may seem surprising at first, but it should not be. In a model with 61 coefficients on the right-hand side, any very substantial amount of parameter drift implies large standard errors of one-step ahead forecasts. The fact that simple random walk models forecast economic time series as well as they do over relatively long time spans is inconsistent with large amounts of parameter variability. One way to read our conclusion is that allowing for parameter drift improves forecasts very little and that since doing so is expensive, in many applications it will be reasonable to use fixed-coefficient models.

However, in a model with 61 coefficients on the right-hand side of each equation, even small amounts of variance in parameter changes can contribute a substantial amount to forecast error. Furthermore in multiple-step forecasts, markov parameter drift of the type our model allows builds up very rapidly in the estimated standard errors of forecast. Therefore, it is important

to allow for parameter drift if one wants to obtain more than point forecasts.

One puzzle we found was that the 12-step ahead log-determinant reached a minimum with priors that were both tighter and that allowed less time-variation than the prior which was best at the shortest forecast horizon. Though the differences in fit are small, the pattern of tighter priors leading to relatively better performance at distant horizons motivated our making further experiments with the form of the prior.

Since our conclusion about the amount of time variation seems to us important, we examined the possibility that it is dependent on the particular form in which we allow parameter variation. We performed the following experiment: we compared the forecasting performance of two constant-coefficient specifications, the first of which uses all available observations at each point in time, and the second of which uses only the 120 most recent observations (if that many are available). By the one-step ahead log-determinant measure, the first specification performs better. Thus, dropping observations, even those more than ten years old, causes the log-determinant to rise. Interestingly, the forecasting performance at longer horizons did improve with the old observations dropped. The conclusion that time variation is small relative to sampling error in coefficient estimates seems to be upheld. Since dropping observations gives more weight to the prior, it appears that long-horizon forecasts might be improved by assuming decay of parameters toward their prior means.

An additional restriction which we considered in the hope that it would allow more time variation in parameters was to impose that the sums of coefficients on lags of each variable in each equation do not vary too much. We found that if this restriction was imposed very tightly, then considerably more time-variation in individual coefficients was possible before the forecasting performance worsened. However, none of these specifications performed as well as those without the tight restriction. The best performance along this added dimension was achieved when π_6 was between 5. and 1., that is, with standard deviations around sums of coefficients of between .2 and 1. In choosing π_6 we also considered various values of π_7, but the returns to this search were not large. Of the combinations of values that we tried, the best was $\pi_6 = 1$ and $\pi_7 = 10^{-7}$. At this specification the standard error of parameter change over the full sample is approximately double what it was at the previous best-fitting specification.

A second type of structure we imposed on the time-variation of parameters was to specify that the coefficients slowly decay toward the prior mean. This structure is implemented by choosing values of the decay parameter, π_8, slightly less than 1.

In performing this experiment we reestimated the coefficients with each new observation, but cost considerations prevented us from revising the coefficient estimates at each step in the forecasting recursion. In one sample forecast where we did take account of parameter decay at the .9975 per period rate, we

found that the forecasts changed only by about .1 percent at the 12-step horizon and about 1.5 percent at the 48-step horizon.

Letting $\pi_8 = .999$ in this type of specification was somewhat successful in terms of improving forecast performance, but it did not provide much room for allowing a larger degree of time variation. At this value for π_8, doubling the time-variation parameter, π_7, to $.2 \cdot 10^{-6}$ marginally improved the one-step forecasts, but led to a much larger decrease in accuracy at longer horizons. Increasing the rate of decay to .9975 caused the forecast performance at a one-step horizon to worsen by about the same amount as that at a 12-step horizon improved, with both changes very small. Larger amounts of decay caused decreases in accuracy at all horizons.

Based on these findings, we adopted as our preferred specification the following parameter values:

$$\pi_1 = .05$$
$$\pi_2 = .005$$
$$\pi_3 = 10^5$$
$$\pi_4 = 2$$
$$\pi_5 = 1.4$$
$$\pi_6 = 1.0$$
$$\pi_7 = 10^{-7}$$
$$\pi_8 = .999$$

The forecast accuracy statistics at this specification are given in Table 2.

Table 2

Final Specification Forecast Performance

Variable	Period 1	Period 2	Period 3	Overall
1-Step Horizon				
Standard Errors				
Real GNP	0.925	0.800	0.759	0.831
GNP deflator	0.199	0.131	0.210	0.183
M1	0.269	0.320	0.474	0.365
Stock Price Index	2.832	3.294	3.734	3.308
Treasury bill rate	0.229	0.241	0.847	0.525
Trade weighted dollar	0.319	0.323	1.957	1.160
Flow of total debt	18.087	8.248	6.694	12.110
Change in inventories	4.839	4.963	6.446	5.465
Federal outlays	4.879	3.500	4.485	4.327
Federal receipts	4.267	4.325	4.008	4.202
Log determinant	-69.497	-71.310	-63.114	-64.829
3-Step Horizon				
Standard Errors				
Real GNP	1.355	1.083	1.178	1.211
GNP deflator	0.506	0.329	0.518	0.459
M1	0.580	0.625	0.944	0.735
Stock Price Index	5.841	6.567	7.556	6.692
Treasury bill rate	0.493	0.560	1.746	1.096
Trade weighted dollar	0.670	0.677	4.339	2.565
Flow of total debt	36.154	13.336	13.101	23.499
Change in inventories	5.610	5.402	8.247	6.549
Federal outlays	5.568	4.204	4.416	4.767
Federal receipts	4.931	5.151	4.440	4.850
Log determinant	-59.240	-61.680	-53.379	-54.349
6-Step Horizon				
Standard Errors				
Real GNP	2.172	1.677	1.729	1.872
GNP deflator	0.826	0.567	0.976	0.808
M1	1.073	1.205	1.234	1.173
Stock Price Index	9.709	10.563	11.721	10.696
Treasury bill rate	0.810	0.793	2.273	1.466
Trade weighted dollar	1.135	0.990	6.666	3.946
Flow of total debt	39.045	15.522	14.429	25.649
Change in inventories	6.140	5.899	9.290	7.276
Federal outlays	8.794	5.145	5.539	6.696
Federal receipts	6.900	5.969	5.670	6.202
Log determinant	-52.102	-55.858	-47.723	-48.051

12-Step Horizon

Standard Errors				
Real GNP	3.500	3.235	3.113	3.287
GNP deflator	1.627	1.122	1.992	1.620
M1	2.150	2.339	1.925	2.145
Stock Price Index	17.891	14.460	16.367	16.300
Treasury bill rate	1.033	1.108	2.469	1.672
Trade weighted dollar	1.984	1.187	9.510	5.651
Flow of total debt	53.879	20.877	8.594	35.045
Change in inventories	6.644	6.800	10.318	8.100
Federal outlays	17.870	7.174	8.061	12.052
Federal receipts	14.370	9.269	9.113	11.187
Log determinant	-44.207	-50.265	-41.723	-40.680

In comparing the performance of different systems it is useful to note that, aside from covariance terms, changes in the log-determinant represent a sum of the percent changes in the variance of forecast errors from each equation. Multiplying the change by five (divide by 20 to get standard errors for ten variables and multiply by 100 to get percent) gives a rough estimate of the average percent change in forecast standard errors of the equations. Thus, we observe an average of about 2 percent improvement in the one-step forecast errors in going from univariate to the final specification, about 12 percent at the 12-step horizon.

In searching informally over parameters of our prior we were encouraged by our finding that forecast performance was generally insensitive to the variation in parameters that we looked at. All of the sets of parameter values we looked at had log-determinants closer to our final choice than to the univariate, indicating the lack of sensitivity of forecast performance over the range of priors we investigated.

In order to investigate this sensitivity more carefully, however, we looked at forecast performance over a larger grid of values for the overall tightness and time variation parameters of our prior. The grid was chosen to cover a region several orders of magnitude wide along both dimensions, far outside the range we would consider reasonable.

Our preferred prior overall tightness of 1.4 represents a scaling up of the variances of all coefficient prior distributions by 40 percent from our original specification. For our grid search we chose to look at the values: .014, .14, .7, 1.4, 2.8, and 14. The final value for our time variation parameter was 10^{-7}. We chose a grid along this dimension of 10^{-15}, 10^{-7}, 10^{-6}, and 10^{-5}. The first value represents essentially no parameter variation, while the last specifies an order of magnitude larger than our preferred value.

The overall accuracy of forecasts generated by our vector autoregressions turns out to be a well behaved function of the prior parameters over which we searched. We present the results of the grid search as a series of charts. The overall forecast accuracy is shown from two different views in Charts 1 and 2. Here forecast accuracy is represented by the height of a surface for each point on our grid. The height is given by

$$(18) \qquad 5 \cdot [\log|E_1| - \log|E_1(\pi_5,\pi_7)|]$$

where E_1 is the cross-product matrix of the one-step ahead forecast errors for our preferred specification and $E_1(\pi_5,\pi_7)$ is the cross-product matrix of one-step ahead forecast errors for the point on the grid (π_5,π_7).

Charts 1 and 2

How Forecast Accuracy Varies With Two Dimensions of the Prior

Chart 1. Front View

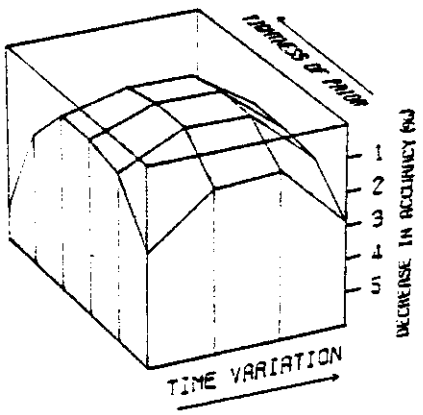
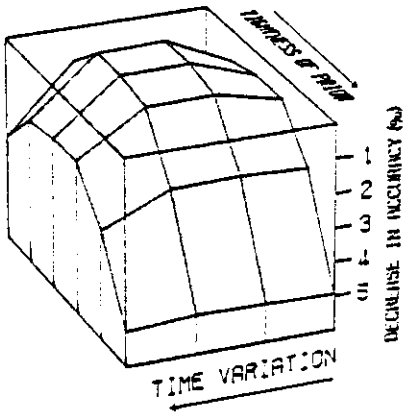


Chart 2. Back View



It can be clearly seen in these charts that the accuracy surface is not sensitive to even order-of-magnitude changes in these parameters of our prior. Because we would give low weight to regions of our grid away from the center, we interpret this result as indicating that if we think of ourselves as having a prior which is a mixture of normal priors indexed by the values of π, we would end up with a posterior much like that for our final chosen specification.

A slightly more detailed picture of the forecast performance over our grid is given in Charts 3 through 8. Here we display the accuracy surfaces for each of our three nonoverlapping sub-periods for the one-step ahead and the 12-step ahead horizons. The consistency of the shape of this surface over the different periods is reassuring. It would appear reasonable to assume that any choice of values for π_5 and π_7 in a wide range around the center of this grid would be likely to remain close to the optimal choice, at least for one-step forecasts.

The results for the 12-step ahead horizon, displayed in Charts 6 through 8 are less consistent over time. In general, though, they reflect the finding that tighter priors with less time variation of parameters appear to perform better at forecasting over longer horizons.

What have we accomplished through this specification search? By some standards, the answer would appear to be not much. After a complex and somewhat expensive (the total computing cost was about $3000) search we find a specification which generates out-of-sample forecast errors which average a few percent

Charts 3–8

Forecast Accuracy Surfaces in Three Nonoverlapping Subperiods

All axes of these charts have the same scales as those of Chart 1.

Charts 3–5. One-Step-Ahead Forecast Horizon

Chart 3. From January 1952 to December 1961

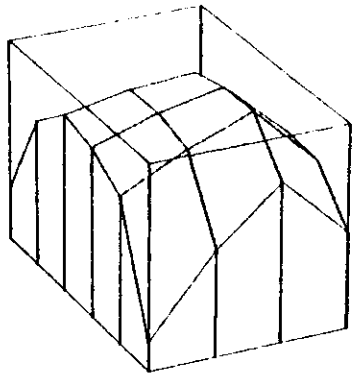


Chart 4. From January 1962 to December 1971

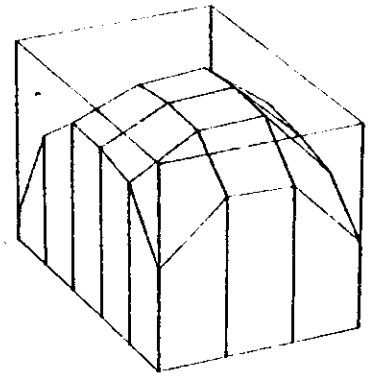
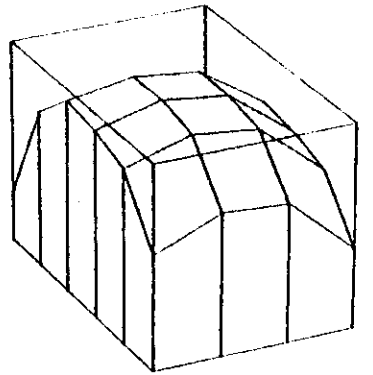


Chart 5. From January 1972 to December 1981



Charts 6–8. Twelve-Step-Ahead Forecast Horizon

Chart 6. From January 1952 to December 1961

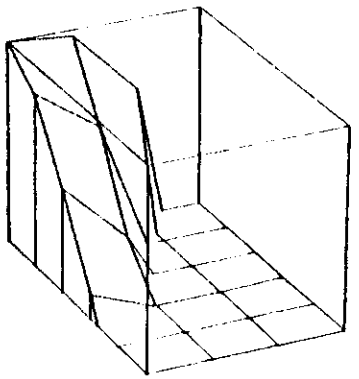


Chart 7. From January 1962 to December 1971

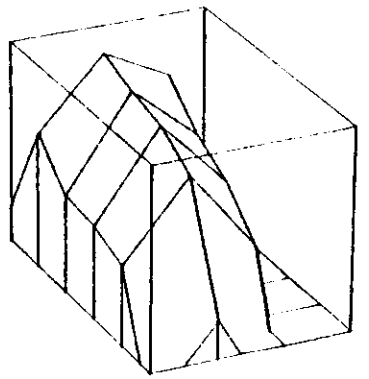
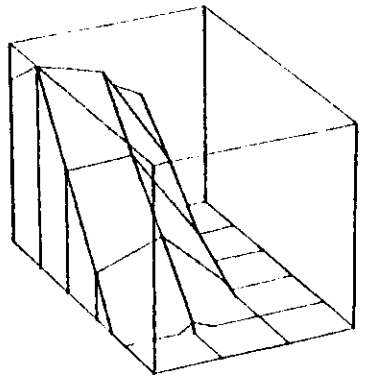


Chart 8. From January 1972 to December 1981



smaller than simple univariate autoregressions. On the other hand, as we pointed out earlier, our search here has been aimed at testing the usefulness of certain ways of specifying a prior. Nearly all the advantage of the multivariate procedures over the univariate procedures in forecasting performance could have been obtained without allowing for parameter drift (a major source of computational expense) and without searching over most of the dimensions we explored. A more difficult question is whether by searching as we have, we have ended up with a reliable probability distribution for future data.

Despite the small absolute gain in forecast accuracy, it is significant that we have documented a consistent gain from the use of a formally explicit multivariate method in a system of this size. This has not been done before, to our knowledge. The difference in accuracy between multivariate and univariate methods which we find is substantial relative to differences in forecast accuracy ordinarily turned up in comparisons across methods, even though it is not large relative to total forecast error. Moreover, if we think of a decomposition of movements in the data into signal and noise, with noise being the dominant component, then a 2 percent increase in forecast accuracy must represent a much larger percentage increase in the amount of signal that is being captured. And with a multivariate probability model which has some claim to accuracy, we can generate conditional distributions of future time paths of a vector of economic variables which capture the most important cross-variable relations.

Forecasts and Conditional Projection

The main purpose of generating a model like ours is to use data at a given date t to make assessments of what is likely to happen after t. We describe here some ideas for making such assessments which are in some ways new, yet which could be applied to any time series model, not just models like that we have constructed.

Obviously one can construct a forecast of the most likely path of the economy. For our model this is just a matter of recursively forecasting one-step ahead with the autoregressive equations, using forecast values as if they were actual data as the date is advanced into the future. The appropriate procedure is to use the most recent estimate of the randomly varying parameters and vary them during the forecasting recursion according to their equation of evolution (6), ignoring the random term in that equation. Of course, when π_8 is 1.0, this amounts to holding the parameters constant. Because the forecasts after the first period data are nonlinear functions of the parameters, they are not unbiased. That is, they do not represent the conditional expectation of future data. One can, at considerable expense, evaluate the conditional expectation by stochastically simulating the model and integrating the posterior distribution of forecasts by Monte Carlo methods. In one experiment using data through 1982:12 we found that the differences in forecasts based on time invariant coefficients, coefficients decaying at the rate .999, and those generated by Monte Carlo integration were quite small relative to the uncertainty in the forecasts.

We present in Charts 9 through 24 two forecasts from the model for 1983 through 1986. The first is based on data through December 1982, the second is based on data through March 1983. The charts in both cases show a forecast of an extremely vigorous recovery, compared to those published forecasts circulating in February 1983. The Congressional Budget Office (CBO), for example, forecast real GNP growth during calendar 1983 of only 4 percent, with inflation at 4.7 percent and the Treasury bill rate at 6.8 percent. As of December, the model forecast real GNP growth at 8.8 percent combined with 5.9 percent inflation and an interest rate of 8.7 percent. Data for the first quarter suggested that the recovery began with less strength than the model anticipated. These observations did not have a significant impact on the forecast growth rates in future quarters, however.

Perhaps the most obvious first step beyond preparing a forecast in using a model to evalutate future prospects is to ask how likely are other possible paths. We can ask, for example, how likely it is that the CBO's projected output and price level growth rates and Treasury bill rates will be realized. In answering these questions, however, we will be taking seriously the cross-variable relationships estimated by the model. Before doing so, it is perhaps useful to investigate those aspects of the model.

The favorable comparison between the forecast performance of our final specification and that of the univariate equations suggests that the cross-variable interactions which are captured by our equations represent predictable responses. More-

Charts 9–24
Forecasts

——— Actuals From January 1981 to March 1983

Forecasts: ----- As of December 1982 (From January 1983 to December 1986)

— — — As of March 1983 (From April 1983 to December 1986)

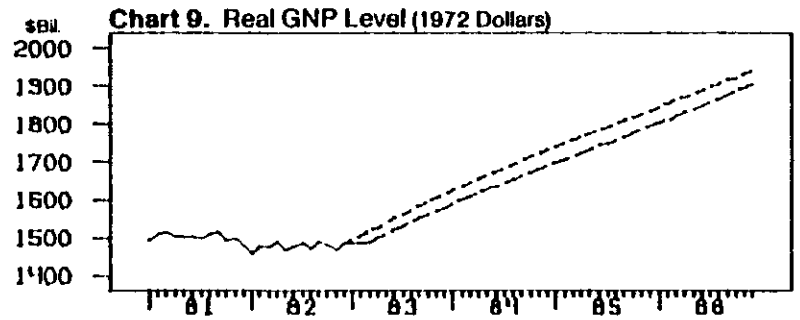


Chart 9. Real GNP Level (1972 Dollars)

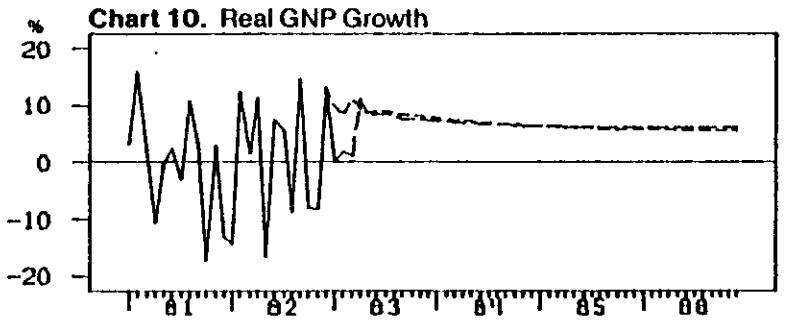


Chart 10. Real GNP Growth

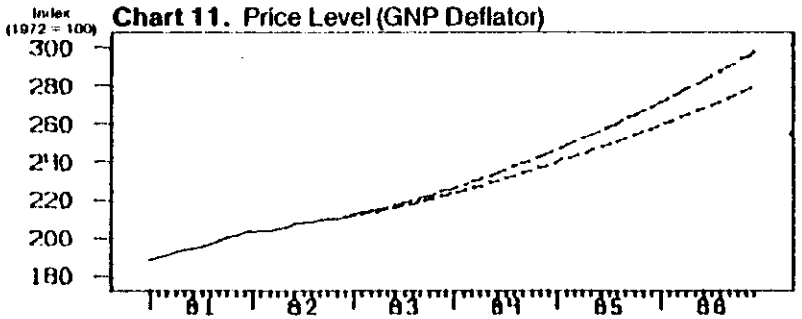


Chart 11. Price Level (GNP Deflator)

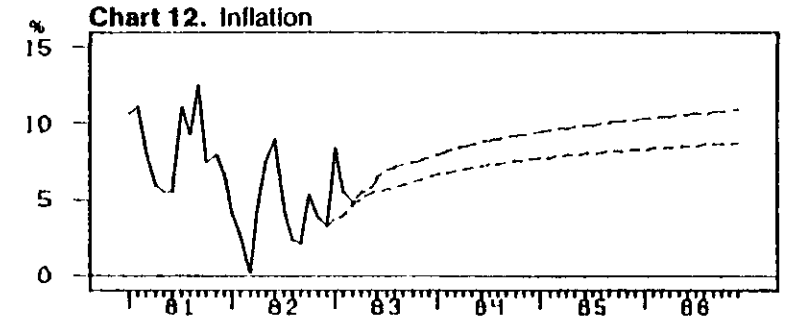


Chart 12. Inflation

Chart 13. Money Supply (M1)

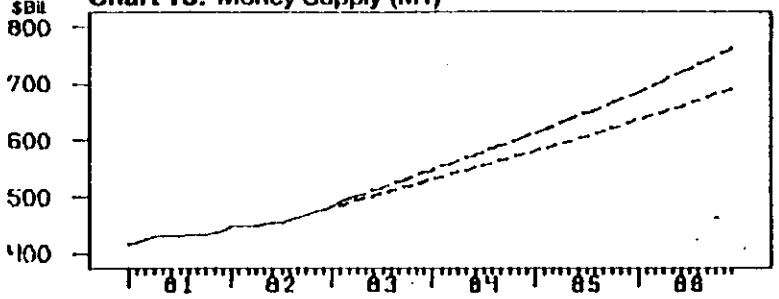


Chart 14. Money Supply Growth

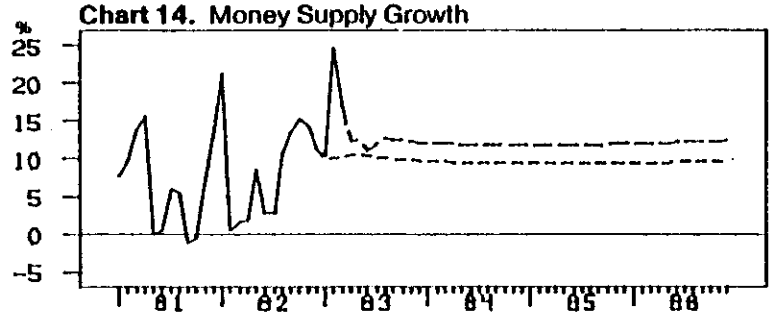


Chart 15. Stock Price Level (Standard & Poor's 500 Index)

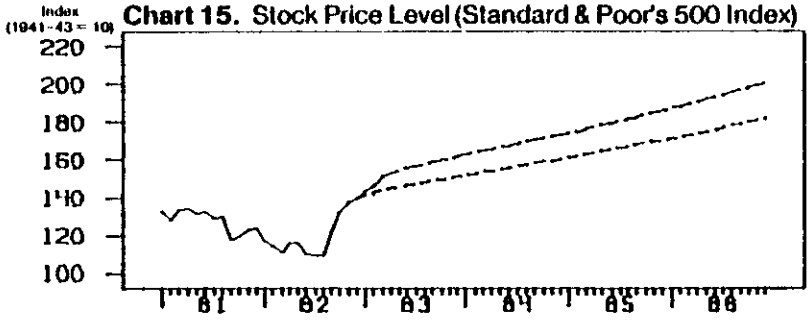


Chart 16. Stock Price Growth

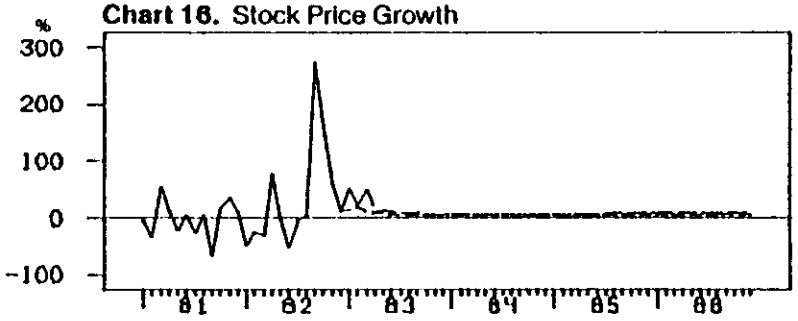


Chart 17. Treasury Bill Rate (3-Month)

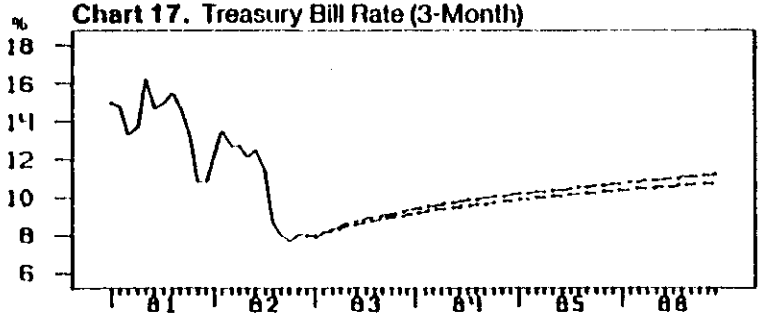
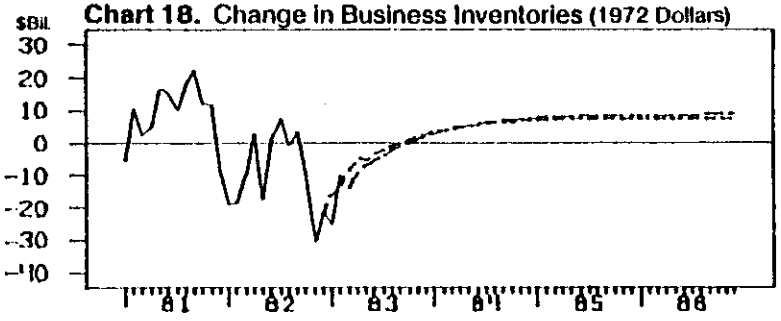


Chart 18. Change in Business Inventories (1972 Dollars)



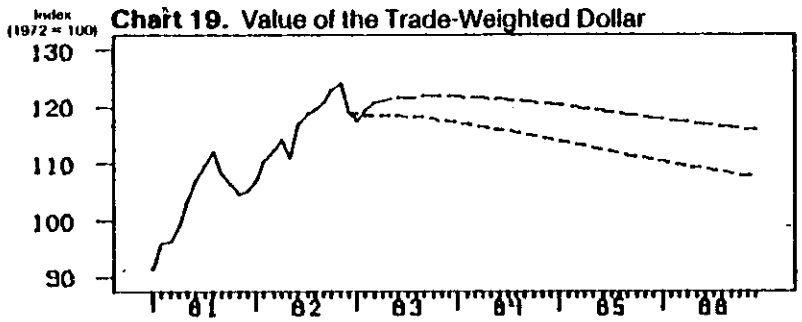


Chart 19. Value of the Trade-Weighted Dollar

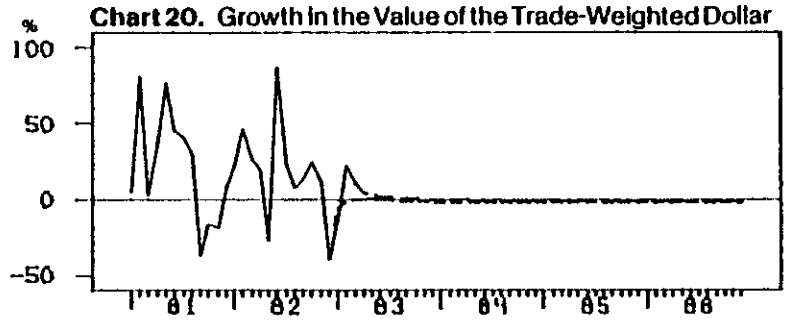


Chart 20. Growth in the Value of the Trade-Weighted Dollar

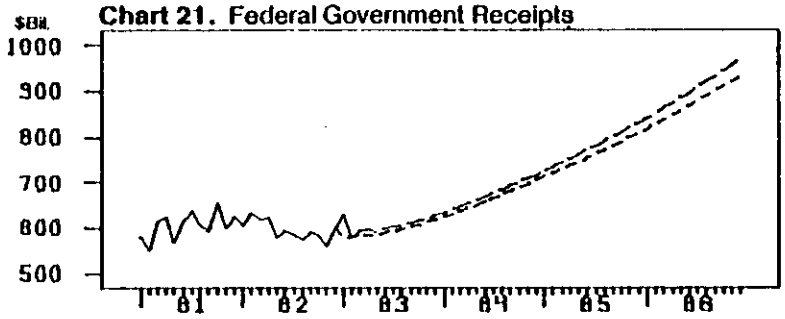


Chart 21. Federal Government Receipts

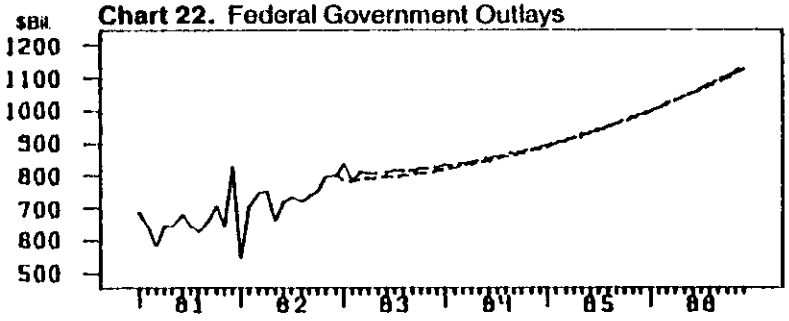


Chart 22. Federal Government Outlays

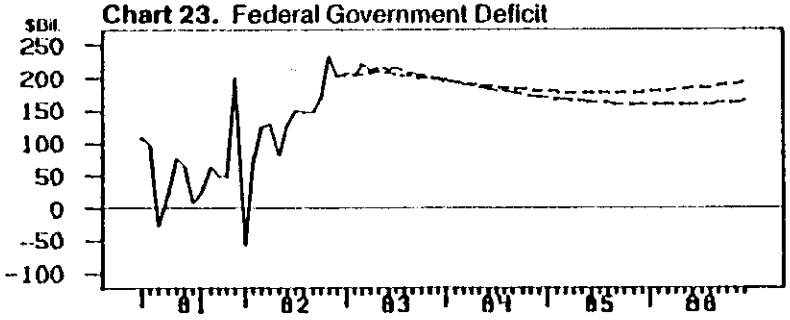


Chart 23. Federal Government Deficit

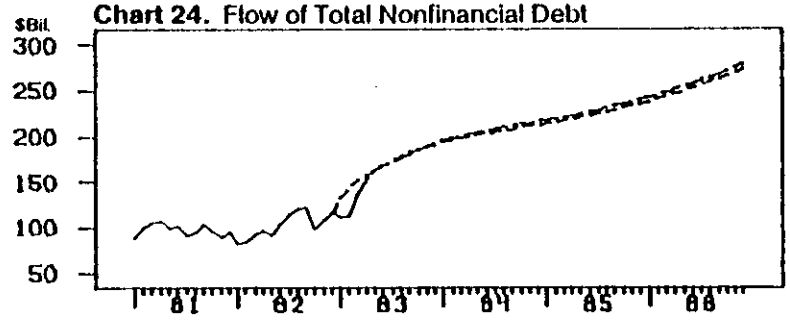


Chart 24. Flow of Total Nonfinancial Debt

over, it turns out that these responses explain a significant proportion of the variation in most of the variables in the model and, with a few exceptions, they remain fairly stable across different sub-periods of the sample.

One measure of the size of the cross-variable interactions is the proportion of the forecast error variance of a variable explained by orthogonalized innovations in the other variables in the system. This measure is based on a decomposition of the variance of the k-step forecast into a sum of components associated with each of a set of orthogonal innovations. See Sims (1980). Although the decomposition depends on the ordering chosen for the orthogonalization, our point here is merely to demonstrate the extent to which interactions amoung variables are captured. We have looked at several orderings, and this aspect of the decomposition is not affected.

For some variables, such as the stock price index, our prior against cross-variable response is so strong that virtually none is allowed. Own innovations explain over 95 percent of stock price forecast errors, even at a 48 month horizon. For other variables in the system, however, the cross-variable responses, shown in Table 3, are significant.

Table 3

Variance Decomposition

Below are the proportions of forecast variance at a 48-month horizon explained by own innovations. The orthogonalization order for this decomposition is the order shown.

M1	29.2
Stock price index	95.1
Treasury bill rate	59.9
Flow of total debt	76.9
GNP deflator	28.4
Change in inventories	76.0
Real GNP	11.7
Federal outlays	79.7
Federal receipts	65.1
Trade Weighted Dollar	54.0

We next display the responses of real GNP to the orthogonalized innovations. These responses also demonstrate the extent to which the model is capable of incorporating multivariate interactions, as well as the extent to which such responses are stable over time. The responses, shown in Charts 25 through 34, were estimated independently over three nonoverlapping subperiods, the same prior being imposed at the beginning of each. Many of the responses are of substantial magnitude relative to the size of the response to own innovations, and for the more significant responses there appear to be strong similarities across the time periods.

Charts 25-34

Responses of Real GNP in Three Nonoverlapping Periods

——— From July 1948 to January 1960

----- From February 1960 to August 1971

— — — From September 1971 to March 1983

Responses are to orthogonalized innovations of the same magnitude for each period. The graphs are displayed in the order that the innovations are orthogonalized. The size of the shock for each variable is normalized to be one standard error of the distribution of the residuals for that variable over the full period. Responses are displayed over a horizon of 48 months in units of percent.

Chart 25. To an Innovation in Stock Prices

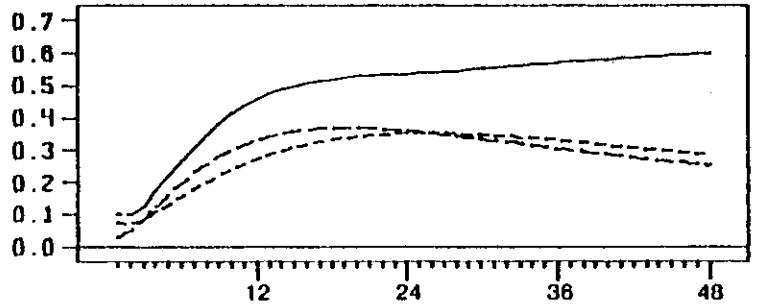


Chart 26. To an Innovation in Treasury Bill Rates

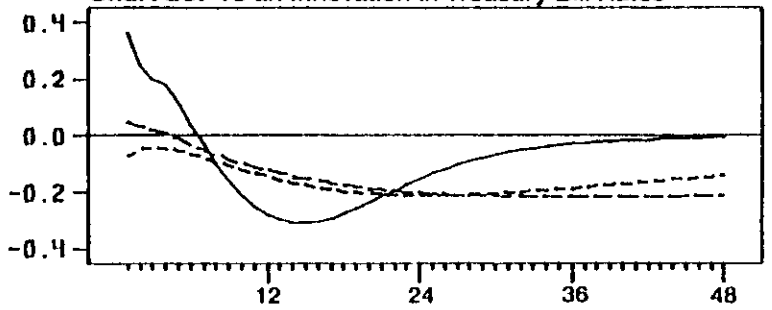


Chart 27. To an Innovation in the Money Supply

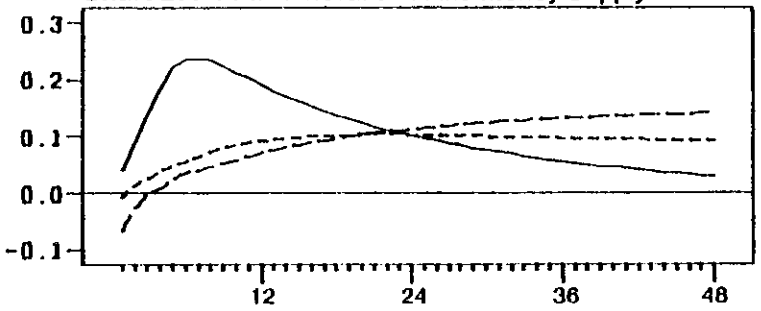


Chart 28. To an Innovation in the Trade-Weighted Dollar

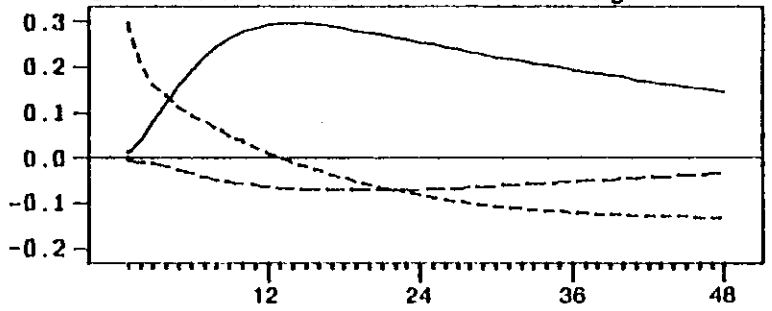


Chart 29. To an Innovation in the Flow of Total Nonfinancial Debt

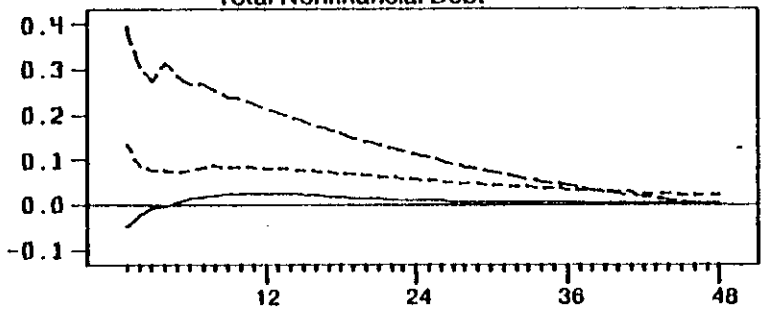


Chart 30. To an Innovation in Federal Government Outlays

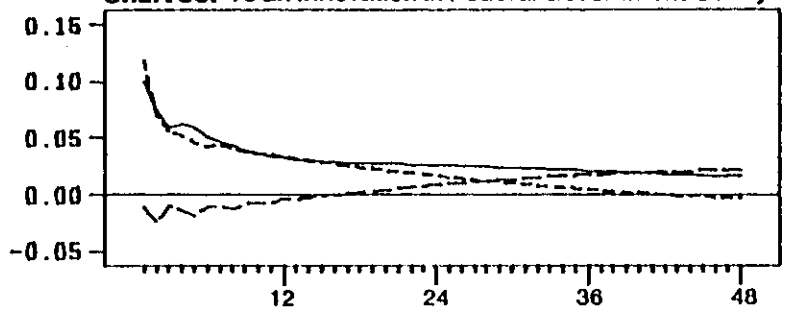


Chart 31. To an Innovation in Federal Government Receipts

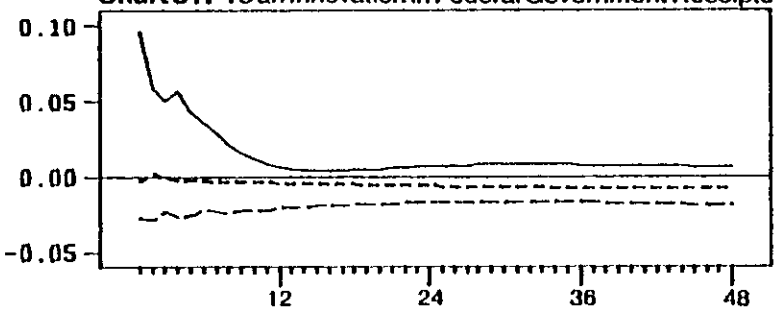


Chart 32. To an Innovation in the Change in Business Inventories

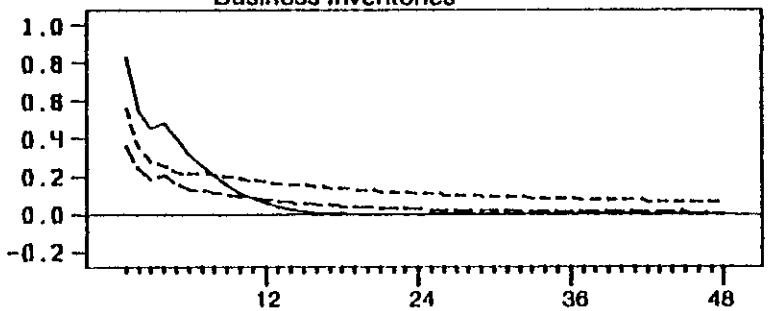


Chart 33. To an Innovation in the Price Level

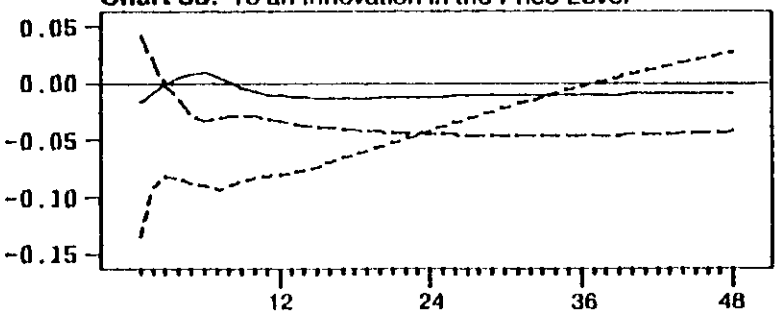
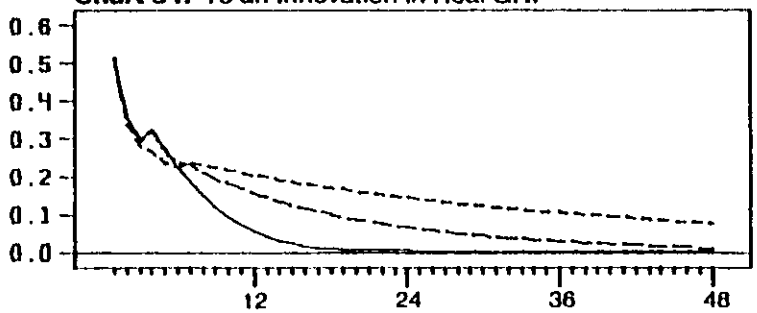


Chart 34. To an Innovation in Real GNP



The responses are scaled to show percent movements in real GNP following orthogonalized innovations in each of the other variables. The size of the shock, which is the same for each period, is normalized to be one standard error of the distribution of innovations over the entire period. The largest responses of real GNP are to innovations in real GNP, the change in business inventories and the stock price index. These responses are all similar across the different sub-periods. The responses of output to interest rates and money innovations are also substantial, and also relatively similar in their dynamic pattern in different sub-periods. The other responses are not particularly consistent over time periods, but for the most part they are not large either.

With regard to the question of consistency across sub-periods, some readers will undoubtedly be more impressed at first glance by the variations in some of the responses than by the similarities of others. Perhaps the most natural metric for measuring the degree of stability of the responses through time, though, is the out-of-sample forecasting accuracy metric which we have already stressed. We know that our specification does well by that measure. What we find encouraging in looking at these response patterns, and the earlier decompositions of forecast variance, is that the prior which led to relatively accurate forecasts is also capable of capturing significant cross-variable interaction, even in these three sub-periods, which each includes only a very limited amount of data.

There is no unambiguously correct way to measure how likely it is that a particular condition on the projected future

path of the economy will be realized. Of course the probability that any set of equality restrictions will be exactly realized is zero. When we ask how likely a projected path is we ordinarily mean to ask how likely it is that the actual path will differ from the model's most likely projection as much as the projected path and "in the same direction". There is no mechanical way to determine, from the path alone, what paths "differing in the same direction" might be. In our example, we might be interested in the probability that real GNP growth will be at least as low as CBO's 4 percent and inflation and the interest rate also at least as low as their projections. But one might instead consider that only the GNP growth rate differences are interesting, so that forecasts differing in the same direction as CBO's are all those with growth rates at least as low. Or one might suppose that the critical thing about the CBO forecast is its lower real interest rate and desire therefore to check the plausibility of its projected gap between inflation rates and interest rates.

If a class of future paths is specified, one can measure the probability of the class directly--by stochastically simulating the model if no computationally cheaper analytic method is available. The method is expensive, however, both in computer time and in its requirement for careful thought about the class of paths to be assessed. Instead, one can mechanically construct a class of paths from specified restrictions. A natural way of doing this is available when the joint density function of future paths is unimodal and has convex level surfaces (like a normal density). We can construct first the most likely path satisfying

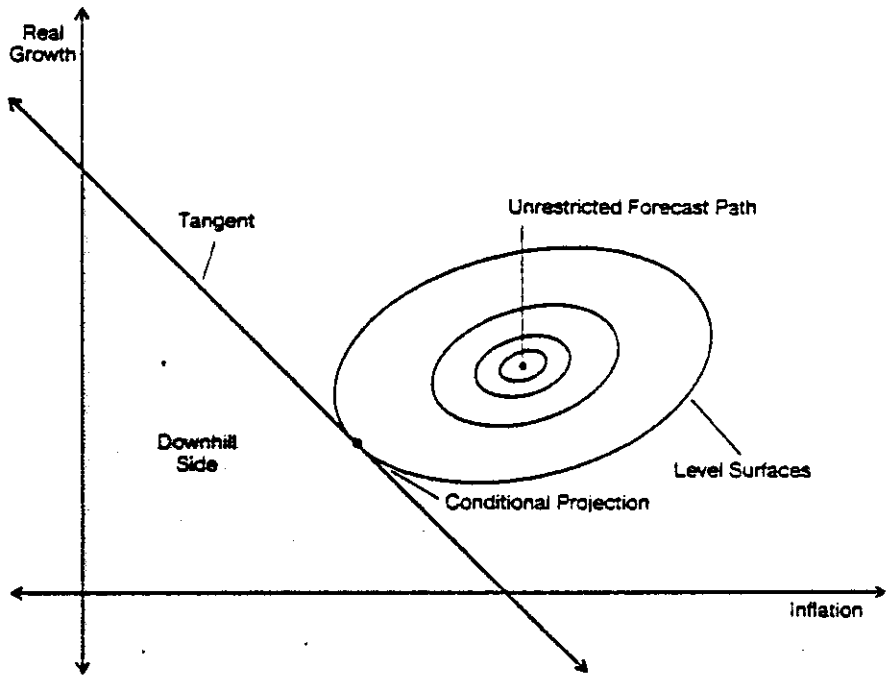
the restrictions, then consider the class of all paths lying on the downhill side of the tangent plane to the level surface at that point in the space of future paths. Chart 35 shows the nature of the set of paths whose probability would be measured in a two-dimensional special case.

For a normal p.d.f., this leads to using the square root of the usual chi-squared statistic as if it were a normal random variable and measuring plausibility by the probability in the upper tail of the normal p.d.f. at the level of this statistic. One might wonder why it is not best, for the case of a normal distribution over future paths, to measure the plausibility of a set of linear restrictions directly by the significance level of its associated chi-squared statistic, using as degrees of freedom the number of restrictions applied. This is, after all, how we would "test" whether the restrictions are "true" using classical methods. Such a procedure treats as the class of paths whose probability is to be measured all paths with lower likelihood than the most likely path satisfying the restrictions. Thus, if the model asserts that real growth will be at 8 percent and inflation at 6 percent and someone claims that instead growth will be at 4 percent and inflation at 9 percent, the claim is in some sense different from the model assertion in one direction--it is more pessimistic. The standard use of the chi-squared statistic would assess the likelihood of the pessimistic forecast by looking at the probability of all paths at least as unlikely, including those which are unlikely because they are much more optimistic than the model. The index we use here instead looks only at paths lying on

Chart 35

Construction of an Implausibility Index

The *implausibility index* is a measure of the probability the model gives to outcomes on the downhill side of a tangent to the forecast's level curve at the point of a conditional projection.



"one side" of the claimed path. This includes some paths with less inflation and much less real growth as well as paths with more inflation and less real growth, so it is not so narrow a class as that of paths with both less real growth and more inflation. However, the tradeoff between inflation and real growth implicit in defining the class of paths "more pessimistic" than that claimed is constructed mechanically from the covariance matrix of paths. This will at best approximate the way we would construct a class of more pessimistic paths if we thought about it carefully. Nonetheless we apply this measure of plausibility here.

To do so, we must first find the model's projection of the most likely future path for the economy subject to the condition that the CBO forecasts for annual average growth rates are satisfied. Such conditional projections may be interesting in their own right as part of a description of the likelihood function, and for other applications we will mention below.

The principle is that the model provides a joint conditional density function for future paths of the process. We simply use that function to find likelihood-maximizing paths subject to certain restrictions on the future paths. These computations cannot in general be carried out recursively forward in time as can the point forecasts, because a constraint on future values of a variable in the system can carry information about the likely current value of all variables. If, for example, I know that money stock will grow slowly between 12 and 18 months from now, and if I know that money stock is negatively correlated with

disturbances in the interest rate 12 to 18 months earlier, then I should think it likely that interest rates will rise soon.

The computations are simplest when the model is stationary and concerned only with second-order properties, so we first describe our procedure within the confines of the prediction problem for covariance stationary processes. The vector stochastic process x_t: $t = \ldots,-2,-1,0,1,2,\ldots$ is assumed to be covariance stationary and linearly regular. The moving average representation (MAR) is

$$(19) \qquad x_t = \sum_{s=0}^{\infty} B_s u_{t-s}$$

where the innovations u_t are uncorrelated both across time and contemporaneously. The MAR is normalized so that $E(u_t u_t') + I$.

A linear constraint upon future values of x is a linear constraint upon future values of the innovations process u. The constraint on x is transformed into the equivalent constraint on u. This has some computational advantages when, as is likely for models of this type, we have already computed the coefficients of the MAR in any case. The least squares estimate of the constrained u's is computed, and the least squares projection of x subject to this constraint is obtained by constructing the path for x implied by the computed innovations.

Let $[y|Q]$ denote the orthogonal projection of the random variable y onto the closed subspace Q in the Hilbert space of finite variance random variables on the underlying probability space. If y is a vector, the projection is done component by component. $H_x(t)$ is the closure of the subspace of finite linear

combinations of x_s for $s < t$. Consider the projection of x_{t+k} on the information set consisting of $H_x(t)$ and

$$(20) \qquad S^* x = \sum_{j=-\infty}^{\infty} S_j x_{t-j} ,$$

where

$$(21) \qquad \sum_{j=-\infty}^{\infty} S_j S_j < \infty .$$

S_j is dimension qxn, where q is the number of constraints and n is the dimension of x_t. The sequence S contains the coefficients on past and future x values in a set of constraints. The projection we are considering can be thought of as the best linear predictor of x_{t+k} given knowledge of x values up to time t and in addition knowledge of the linear combinations of past and future x's whose coefficients are in S. In practice, the S sequence will be zero except for a finite number of terms. Applying the law of recursive projections:

$$(22) \qquad [x_{t+k} | H_x(t) + \mathrm{span}(S^* x)] = [x_{t+k} | H_x(t)] +$$

$$[(x_{t+k} - [x_{t+k} | H_x(t)]) | \mathrm{span}(S^* x - [S^* x | H_x(t)])]$$

and

$$(23) \qquad x_{t+k} - [x_{t+k} | H_x(t)] = \sum_{s=0}^{k-1} B_s u_{t+k-s}$$

Now,

$$(24) \qquad S^* x - [S^* x | H_x(t)] = S^*(x - [x | H_x(t)])$$

$$(25) \qquad x_s - [x_s | H_x(t)] = 0 \text{ for } s < t$$

(26) $$x_s - [x_s | H_x(t)] = \sum_{j-0}^{s-t+1} B_j u_{s-j} \text{ for } s > t ,$$

so

(27) $$S^*(x - [x | H_x(t)]) = \sum_{m=1}^{\infty} S_{-m} \left(\sum_{j-0}^{m-1} B_j u_{t+m-j} \right)$$

(28) $$= \sum_{k=1}^{\infty} \left(\sum_{j=0}^{k-1} S_{j-k} B_j \right) u_{t+k}$$

(29) $$= R^* u = \sum_j R_j u_{t-j}$$

By linearity, the second projection in (22) can be written

(30) $$\sum_{s=0}^{k-1} B_s [u_{t+k-s} | R^* u] .$$

It can be verified, using the orthogonality principle for projections, that the projection $[u_{t+k-j} | R^* u]$ is

(31) $$R_{j-k} \left\{ \sum_{j=-\infty}^{-1} R_j R_j' \right\}^{-1} (R^* u).$$

These are the least squares projections of the future innovations. The projection (22) is thus the sum of the unconstrained forecast plus

(32) $$\sum_{s=0}^{k-1} B_s u_{t+k-s}$$

which can be obtained by simulating the model beginning at t+1, using the u as the innovations. In a particular application, a value for S^*x is usually supplied; the equivalent value for R^*u is the difference between S^*x and the forecast value for S^*x.

To see how this works in a simple case, suppose that data on M1 are available only with a two week delay, while interest rate TB is available on a daily basis. We have a weekly model which we wish to use to forecast, but at t have data on M_s only

for $s < t-1$. Here, "purely for forecasting purposes, we need to make a projection conditional on TB_{t-1} and TB_t.

With the VAR normalized so that B_0 is lower triangular and TB comes above M1 in the ordering of variables, the moving average coefficients needed are the responses of R to orthogonalized shocks in itself at zero and one step and in M1 at one step; call these b_{b0}, b_{b1}, and b_{m1}. With v_t and w_t as the innovations in TB and M1 at t,

$$(33) \qquad \begin{bmatrix} TB_{t-1}-\hat{TB}_{t-1} \\ TB_t-\hat{TB}_t \end{bmatrix} = \begin{bmatrix} b_{b0} & 0 \\ b_{b1} & b_{m1} \end{bmatrix} \begin{bmatrix} v_{t-1} \\ w_{t-1} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ b_{b0} & 0 \end{bmatrix} \begin{bmatrix} v_t \\ w_t \end{bmatrix}$$

The 2x2 matrices in this are, respectively, R_{-1} and R_{-2} in the notation above. The most convenient way to do this computation is to stack the set of innovations. With

$$(34) \qquad \hat{U} = [v_{t-1} w_{t-1} v_t w_t]'$$

$$(35) \qquad R = \begin{bmatrix} b_{b0} & 0 & 0 & 0 \\ b_{b1} & b_{m1} & b_{r0} & 0 \end{bmatrix}$$

$$(36) \qquad r = \begin{bmatrix} TB_{t-1}-\hat{TB}_{t-1} \\ TB_t-\hat{TB}_t \end{bmatrix}$$

the formula for the constrained U vector becomes $\hat{U} = R'(RR')^{-1}r$, which is the solution of the problem: min $U'U$ subject to $RU = r$.

One important variant on this procedure is to add the additional constraint that only certain innovations are allowed to be nonzero. We might want to do this if we had in mind interpre-

tations for certain innovations. For example, if we regarded money and interest rates as "monetary policy variables", we might suppose that innovations in those variables represented changes in policy. Then a forecast conditional on low inflation and on innovations being zero in all variables other than these monetary policy variables would display the most likely way for monetary policy to generate low inflation.[4/]

Holding certain innovations to zero in the conditional projection can be accomplished simply by eliminating the columns in the R matrix which correspond to the variables whose innovations must be zero. For computing constrained paths, the normalization of the MAR used to obtain orthonormal u's has no effect except on the computational burden: if $E(u_t u_t') = \Sigma$, the formula, using the stacking from above, is $\hat{U} = (\Sigma \otimes I)R'[R(\Sigma \otimes I)R']^{-1}$ where a different R' matrix is obtained using the nonorthogonalized MAR. Orthogonalization eliminates the need for the I by incorporating a factorization of Σ into the MAR and thus into the R matrix. However, when innovations for certain variables are constrained to be zero orthogonalization is no longer innocuous, since the definition of a variable's innovations depends on the orthogonalization. For example, the least squares constrained

---

[4/] There are a number of models in the literature which identify innovations in certain variables as generated by policy or which go still further and treat certain policy variables as exogenous, hence Granger causally prior, and as entirely determined by policy. In fact, this kind of assumption is probably the norm in models which are used to generate implications for policy. We regard such assumptions as frequently being interesting speculative hypotheses, but seldom solidly justifiable as "a priori knowledge."

path may prove to be obtained primarily through innovations in the policy variables when one ordering is used, but through innovations in the nonpolicy variables in another.

Though the proofs above were limited to covariance stationary processes, the method will still work, e.g., if x has an invariant autoregressive representation with unstable roots.

Our experience suggested that, though models with time-invariant coefficients generate reasonable forecasts, they have a tendency to generate unreasonably optimistic estimates of the likely size of future forecast errors--even when sampling error in the estimated coefficients (which we ignored above) is allowed for. One of the objectives of our research has been to discover whether our random parameter specification avoids this optimistic tendency.

We will compare four different estimates of the covariance matrix of forecast errors. The first matrix, F, is generated from the usual innovation covariance matrix, Σ, estimated by taking cross products of in-sample residuals based on a fixed-coefficients model. The k-step forecast error covariance matrix is given as

$$(37) \qquad \Sigma_k = \sum_{s=0}^{k-1} B_s \, \Sigma \, B_s'$$

where the B_s's are the coefficients in the MAR associated with the fixed-coefficient model.

Our second estimate of the forecast error covariance matrix is o, the estimate obtained by using a time-varying coefficient model, but taking the end-of-period coefficient estimater as

fixed and using the out-of-sample one-step ahead forecast errors to estimate to covariance matrix of innovations.

Another estimate of the forecast error covariance matrix is obtained by a Monte Carlo simulation of the full random parameter model from the end-of-sample initial conditions. This estimate we will call M.

Finally, another way to assess likely forecast accuracy which is in some sense conservative is to recursively generate forecasts over a range of horizons at each sample point, using data only up to the forecast date in making each set of forecasts. Forming the sample covariance matrix, V, of realized forecast errors at various horizons gives us a direct measure of likely forecast error variances at those horizons. This procedure assumes that the stochastic process for the vector of forecast errors by horizon is jointly stationary, but requires no assumption that the model justifying the forecast procedure is also generating the data.

Our experiments with these four different ways of estimating forecast error covariance matrices gave no clear ranking of the methods. The estimated standard errors of forecasts at 12 and 48 month horizons are shown in Table 4. Each of our difference estimates, F, O, M, and V at times gives both the largest and the smallest estimated standard errors. This result is certainly due in part to the small samples we are using. In our Monte Carlo estimates we used only 200 draws, and for the generation of historical second moments in V we use 240 observations, which represent only 5 nonoverlapping 48-month periods.

Table 4

Standard errors of 12-month forecasts,
estimated various ways

	M1	STOCKS	TBILL	DEBT	PGNP	CBI	RGNP
F	.0136	.1360	1.4113	.2666	.0120	7.9992	.0233
O	.0111	.1130	1.3225	.2822	.0088	8.3999	.0188
V	.0233	.1136	1.2549	.2394	.0202	7.2900	.0293
M	.0129	.0999	1.3039	.2917	.0136	7.8663	.0206

	OUTL	RCPT	TRDOL
F	.0857	.0958	.0457
O	.0890	.1008	.0400
V	.0826	.0927	.0293
M	.0949	.0952	.0412

Standard errors for 48-month forecasts
estimated various ways

	M1	STOCKS	TBILL	DEBT	PGNP	CBI	RGNP
F	.0352	.2422	1.5435	.3010	.0245	8.2767	.0651
O	.0277	.2399	1.5542	.3359	.0240	8.9611	.0612
V	.0834	.1255	1.3493	.1906	.0871	7.1102	.0589
M	.0882	.2191	1.8737	.4500	.0948	11.0070	.1083

	OUTL	RCPT	TRDOL
F	.1048	.1218	.0889
O	.1128	.1354	.0891
V	.1164	.1530	.0826
M	.2054	.2136	.0344

F: Estimating fixed coefficient model, using in-sample residuals to estimate innovation covariance matrix.

O: Using end-of-sample time-varying coefficient estimates as if fixed, treating historical out-of-sample one-step forecast error second moment matrix as if it were the innovation covariance matrix.

V: Historical second moments of out-of-sample forecast errors.

M: Monte-carlo estimates of forecast errors based on time-varying coefficients model started up from end-of-sample initial conditions.

Our original suspicion, that estimates of uncertainty, such as F based on fixed coefficient models, would badly underestimate the average out-of-sample multi-step forecast errors as measured in V, was only occasionally observed. At the 48-step horizon F badly underestimates the size of observed errors only for money and prices. In those two cases the Monte Carlo estimates in M, based on the time-varying specifications, did give estimates much closer to the observed results. More often, however, the estimates in F were larger than the observed forecast variance, and the Monte Carlo estimate in some of those cases gave even larger estimates. It is possible, of course, that the use of V as a standard of comparison is inappropriate. When parameters are varying through time the uncertainty also varies, and at a given time it may be very different from an estimate based on average errors in the past. For the trade-weighted dollar the Monte Carlo estimate suggests much less uncertainty than the others, and it is certainly conceivable that this is correct.

The time-varying parameters specification used in this paper implies a conditionally heteroscedastic nongaussian distribution for the forecast errors. If we form the sample covariance matrix, V, of forecast errors and form conditional projections as minimum mean square error predictions using V, we are therefore contradicting the probability model which justifies our forecast-

ing procedure. However, it is not clear whether that model is more realistic than one which uses V to form conditional projections.

Using V to form conditional projections is only in a sense conservative. It is unlikely to greatly underestimate the magnitude of errors, even at long horizons. But when we estimate the whole of V without applying Bayesian methods we are losing the stability provided by Bayesian "shrinkage" toward a prior mean. In particular when we start comparing conditional projections to form conclusions about how much variables respond to each other, use of V may give an exaggerated view of how strong the interaction among variables in the data is.

A Gaussian covariance-stationary process generates a normal joint distribution for future paths given the past, with some covariance matrix. However, that covariance matrix has a special structure. To take the simplest case, consider the covariance matrix of one- and two-step ahead forecasts for a univariate process. If innovation variance is one, the variance of two-step ahead forecast errors, s_{22}, is $1 + b^2$, that for one-step ahead forecast errors, s_{11}, is 1, and the covariance of one- and two-step ahead forecasts, s_{12}, is b, where b is the coefficient on the first lagged innovation in the MAR. Thus, the square root of $s_{22}-s_{11}$ is s_{12}. But for a process such that minimum variance forecasts are nonlinear functions of the data such a restriction on the covariance matrix of forecast errors is not in general satisfied. For example, suppose $y(t) = e(t) + \mathrm{sgn}(e(t-1))$, where $e(t)$ is i.i.d. uniformly on $(-.5,.5)$ and the function sgn has value 1 if its argument is positive and -1 if its argument is

negative. Clearly we can determine sgn(e(t-1)) from knowledge of y(t-1), so the one-step ahead forecast error variance is the variance of e(t), i.e., 1/12. The variance of the two-step ahead forecast error is 1+(1/12), and the covariance of one- and two-step ahead forecasts is not $(s_{22}-s_{11})^{.5} = 1$ but instead .25.

To find the best linear forecast for a given fixed V not generated by a covariance stationary process therefore requires some modification of our procedure. For this case the difference between the constrained forecast and the unconstrained forecast is $VS'(SVS')^{-1}r$, where the matrix S is taken directly from the constraints on x. A restriction corresponding to the restriction that only certain variables have nonzero innovations can be obtained by examining the meaning of a Choleski factorization of V into LL', L lower triangular. If $EUU' = I$, then if $W = LU$, $EWW' = V$. The Choleski factorization transforms the forecast error W into LU, where each component of U is created as that part of the corresponding element of W which is uncorrelated with the previously defined U's. This is precisely how the orthogonalized innovations decompose the forecast error in the covariance stationary case: the innovation for variable j at step k is the (normalized) part of the forecast error which is orthogonal to the innovations in all variables for steps < k and for variables < j at step k. L describes an analog of the moving average representation: each column gives the response of the system to a unit shock in the corresponding component of U. If $W = VS'(SVS')^{-1}r$, then $U = L^{-1}W$, and if R is defined as SL, then $U = R'(RR')^{-1}r$. Again, by cutting the appropriate columns out of the matrix R, restrictions that certain innovations remain at zero can be implemented.

In this paper we have conditioned on constraints which involve projections 48 periods into the future. Because of the size of the system, a full V or M matrix would be 480 x 480. Rather than attempt to operate with such a huge matrix we have restricted ourselves to looking at the conditional projections for a nonconsecutive sequence of horizons between 1 and 48 steps into the future, with all constraints being put only on those included horizons. That is, instead of forming the covariance matrix for 1 through 48 step ahead forecasts, we form the covariance matrix for 1, 2, 3, 6, 9, 12, 18, 24, 30, 36, and 48 step ahead forecasts. The restrictions we consider on future paths must then be defined in terms of these horizons.

Tables 5-8 show the unconditional forecast produced by the model and forecasts conditional on the CBO's 1983 and 1984 averages for interest rates, inflation and real growth, using three covariance matrices: M, V, and O. Though they were not constrained to match the CBO projections for the deficit, these forecasts agree with it fairly closely. CBO projects 194, 197, 214, and 231 billion dollars for fiscal years 1983-86, and all the model projections are in this same range.

Table 5

Unconditional Forecast

Continuously compounded precentage change at annual rate from previous period on table with period previous to 83-1 taken as 82-12. Except that TBILL is in percent and CBI and DEFICIT are in billions of dollars at annual rates.

Key to variable names

M1	Money Supply	STOCKS	Stock Price Index
TBILL	Treasury Bill Rate	DEBT	Flow of Total Nonfinancial Debt
PGNP	GNP deflator	CBI	Change in Business Inventories
RGNP	Real GNP	OUTL	Federal Outlays
RCPT	Federal Receipts	TRDOL	Trade weighted dollar

DEFICIT Federal Deficit (Monthly figures at annual rates, in $bill.)

YR	MO	M1	STOCKS	TBILL	DEBT	PGNP	CBI
83	1	9.38	14.38	8.01	144.54	4.08	-15.61
83	2	9.46	15.70	8.13	111.67	4.26	-13.49
83	3	9.60	11.86	8.27	59.06	4.97	-7.70
83	6	9.81	7.47	8.59	37.94	5.48	-3.11
83	9	9.36	6.38	8.85	30.21	5.91	.49
83	12	9.11	6.03	9.08	16.31	6.33	3.24
84	6	8.86	6.01	9.44	9.60	6.78	6.04
84	12	8.66	6.08	9.74	6.80	7.20	7.19
85	6	8.57	6.17	9.98	7.84	7.47	7.57
85	12	8.56	6.28	10.20	9.96	7.66	7.63
86	12	8.61	6.45	10.60	12.93	7.85	7.48

		RGNP	OUTL	RCPT	TRDOL	DEFICIT
83	1	9.77	-35.04	-45.36	-.13	201.24
83	2	8.38	.12	-3.96	4.02	203.22
83	3	10.62	7.08	10.32	5.60	202.86
83	6	9.00	2.56	2.72	5.81	203.93
83	9	8.47	1.40	5.64	4.24	198.40
83	12	7.83	4.04	8.40	3.23	193.86
84	6	7.06	5.64	10.30	2.23	184.75
84	12	6.33	7.80	11.44	1.41	179.95
85	6	5.85	9.38	11.88	.97	179.71
85	12	5.53	10.58	12.06	.74	183.87
86	12	5.28	11.75	12.19	.59	203.03

Table 6

Using the Simulated Random Coefficients M Matrix:
Model Forecast Conditional on CBO Average Real Growth,
Inflation, and Bill Rate for 1983 and 1984.
Percentage Growth Rates at Annual Rates Between Listed Dates
(Except TBILL, CBI and DEFICIT)

YR	MO	M1	STOCKS	TBILL	DEBT	PGNP	CBI
83	1	7.56	-11.52	7.89	134.16	4.32	-15.83
83	2	7.20	.24	7.63	91.68	3.84	-14.26
83	3	7.68	6.00	7.43	82.68	4.44	-11.10
83	6	8.72	-15.08	6.96	29.68	5.04	-7.24
83	9	8.48	12.60	6.56	18.00	5.20	-7.50
83	12	10.08	-.64	6.24	10.12	4.36	-7.05
84	6	8.68	2.60	7.06	40.16	4.06	2.44
84	12	7.18	8.94	7.74	19.42	5.14	.12
85	6	6.72	13.96	8.78	-12.52	4.74	3.55
85	12	6.34	9.18	9.76	-5.68	5.46	6.87
86	12	4.61	4.02	11.03	-11.47	5.64	5.34

		RGNP	OUTL	RCPT	TRDOL	DEFICIT
83	1	4.68	-65.40	-63.24	3.24	190.34
83	2	6.60	2.40	42.00	.12	171.62
83	3	4.20	15.60	8.40	6.12	177.43
83	6	6.16	6.00	-7.60	3.04	200.22
83	9	1.92	4.80	-7.20	2.28	220.02
83	12	2.76	12.00	5.20	-2.60	236.64
84	6	6.56	-1.60	4.80	.14	216.09
84	12	2.84	4.00	-2.00	.44	238.32
85	6	5.20	3.80	6.60	2.08	234.46
85	12	3.90	-1.20	12.80	2.80	189.37
86	12	2.37	12.70	9.30	4.51	239.55

Table 7

Using the Empirically Estimated V Matrix:
Model Forecast Conditional on CBO Average Real Growth,
Inflation, and Bill Rate for 1983 and 1984.
Percentage Growth Rates at Annual Rates Between Listed Dates
(Except TBILL, CBI and DEFICIT).

YR	MO	M1	STOCKS	TBILL	DEBT	PGNP	CBI
83	1	4.56	-3.00	7.73	155.76	4.08	-15.22
83	2	5.28	5.40	7.38	113.64	4.08	-13.82
83	3	7.08	9.84	7.04	49.80	4.68	-9.12
83	6	8.20	6.52	6.56	30.84	4.68	-6.93
83	9	7.56	9.72	6.79	24.76	4.92	-5.24
83	12	7.24	6.64	6.81	23.96	4.92	-2.09
84	6	6.66	8.42	7.42	25.04	4.46	3.34
84	12	6.68	12.68	7.38	15.68	4.74	1.78
85	6	5.56	8.50	8.04	11.76	5.54	8.81

85	12	6.32	13.54	8.57	-3.32	6.08	9.17
86	12	5.40	8.84	9.85	-3.95	6.40	7.69

		RGNP	OUTL	RCPT	TRDOL	DEFICIT
83	1	2.28	-34.20	-46.44	.84	202.31
83	2	1.44	-3.60	-12.00	2.64	205.71
83	3	4.80	-1.20	7.20	3.12	201.50
83	6	3.20	1.60	-.40	4.32	205.18
83	9	4.44	-5.60	1.60	1.04	192.06
83	12	5.52	2.00	4.40	.40	189.54
84	6	5.00	6.00	3.80	2.00	201.87
84	12	4.40	4.80	6.20	-3.36	202.51
85	6	6.50	.20	4.20	2.96	190.35
85	12	5.86	2.80	6.60	.26	180.88
86	12	4.44	8.50	11.30	.58	176.97

Table 8

Using the Fixed Coefficients O Matrix:
Model Forecast Conditional on CBO Average Real Growth,
Inflation, and Bill Rate for 1983 and 1984.
Percentage Growth Rates at Annual Rates Between Listed Dates
(Except TBILL, CBI and DEFICIT).

YR	MO	M1	STOCKS	TBILL	DEBT	PGNP	CBI
83	1	6.48	-14.16	7.76	109.20	3.72	-15.71
83	2	6.48	-13.20	7.53	76.08	3.60	-13.97
83	3	7.32	-15.24	7.28	24.48	4.32	-8.84
83	6	8.36	-18.64	6.81	18.64	4.56	-6.78
83	9	8.32	-14.56	6.54	28.32	4.92	-5.16
83	12	8.44	-8.88	6.57	27.36	5.44	-5.04
84	6	7.72	-2.76	7.24	15.46	4.36	.48

84	12	7.78	5.06	7.56	13.62	4.84	-.22
85	6	9.06	8.30	8.39	16.96	7.90	5.48
85	12	8.58	8.48	9.04	5.16	8.12	7.50
86	12	8.43	8.52	9.91	6.44	8.26	8.09

		RGNP	OUTL	RCPT	TRDOL	DEFICIT
83	1	4.92	-39.00	-41.64	.24	196.89
83	2	4.08	-1.20	-3.60	3.48	197.85
83	3	6.36	6.00	8.40	4.32	197.68
83	6	4.16	1.20	0.00	4.48	200.02
83	9	3.88	-.40	1.20	2.32	197.49
83	12	2.84	.80	2.00	1.20	196.13
84	6	5.58	6.00	3.40	.72	209.90
84	12	3.82	6.20	4.40	-1.18	222.01
85	6	8.32	5.20	11.00	-.48	209.48
85	12	6.82	8.00	12.80	-.66	201.77
86	12	5.58	10.70	13.50	-.28	202.89

The "implausibility index" for the fixed coefficients forecast, generated as the root sum of squares of the standardized shocks required to generate the forecast, is 4.4, improbable if treated as a one-tailed normal or t test statistic. For the forecast generated from the V matrix the index is 3.0 and for the M matrix it is 3.3--both smaller than for the fixed coefficients model, but still in the range of implausibility.

All the conditional forecasts show an initial sharp contraction in both outlays and receipts, and all show slower money growth than the unconditional forecast. On the other hand,

the degree to which money growth is reduced is much larger in the V forecast than in either of the other two, and the reduction in stock prices is much greater in the fixed-coefficients model than in the other two. We should note that the results from the simulation-based M matrix differ noticeably between an M matrix based on 200 random draws and one based on 100 random draws, and because the empirical V matrix is also based on a sample of only a few hundred highly dependent observed forecast errors, it too is probably infected by substantial sampling error. Thus, though noticeable differences exist, they may be inherent statistical error rather than fundamental differences in the results based on these different approaches.

To understand why it emerges as implausible, it may help to examine the time sequence of standardized shocks implied by the forecast, as displayed in Table 9 for the empirical V version. Note that there are no standardized shocks after 1984.12 because the constraints involved no dates after that.

Table 9

Standardized Shocks Generating the Table 7 Projection

YR	MO	M1	STOCKS	TBILL	DEBT	PGNP	CBI
83	1	-0.9	-0.4	-0.6	0.1	0.3	-0.0
83	2	-0.4	-0.3	-0.6	0.1	0.3	-0.0
83	3	-0.3	-0.3	-0.5	0.1	0.3	-0.0
83	6	-0.3	-0.3	-0.9	0.3	0.4	0.1
83	9	-0.3	-0.3	-0.6	0.2	0.3	-0.1
83	12	-0.4	-0.2	-0.4	-0.0	0.2	-0.0
84	6	-0.7	-0.2	0.1	-0.1	-0.1	-0.2
84	12	-0.3	-0.1	-0.2	-0.1	0.1	-0.4

		RGNP	OUTL	RCPT	TRDOL
83	1	-0.8	0.0	-0.0	0.2
83	2	-0.7	0.0	-0.0	0.1
83	3	-0.5	0.0	0.0	0.2
83	6	-0.8	-0.1	0.0	0.2
83	9	-0.6	-0.2	0.1	0.2
83	12	-0.4	-0.0	0.1	0.0
84	6	-0.6	0.0	0.0	0.0
84	12	-0.5	0.0	0.0	0.0

Because the model shows a strong connection of M1 innovations and stock price innovations to subsequent output and (to a lesser extent) price movements, both these variables show a sequence of fairly large negative standardized shocks. One possible interpretation of the projection is an "irrational monetarist" one. A less expansionary monetary policy than the model's uncon-

strained forecast leads to correct anticipations of lower future inflation and to lower nominal interest rates. Because of some kind of price rigidity or money illusion (perhaps an inability of wage contracts to lower their rates of increase fast enough) the lower inflation rate leads to persistently lower output.

As one of us (Sims (1983)) has recently argued, though, the practice of identifying policy actions with innovations in policy variables, which underlies much standard manipulation of econometric models for policy analysis as well as some rational expectations macroeconomics, requires justification, which may not be easy to find. One could interpret Tables 7 and 8 as showing the response of the economy to public recognition that capacity utilization is likely to remain low and unemployment high, due to continued slow adaptation of the industrial economies to high energy prices and to the nominal inertia of the wage and price setting mechanism. On this interpretation new information appears first in the financial variables money, the bill rate, and the stock price index because all three (with a partially accommodative monetary policy) react quickly to the public's anticipations of the future. They therefore do not reflect policy decisions and the difference between the CBO and the central model projection cannot be read as displaying the effect of contractionary monetary policy.

One interpretation which is not consistent with the model is the idea that deficits might be critical to the difference between the model's expansionary central forecast and the less vigorous CBO forecast. Differences between the deficit

predictions for these conditional projections and those for the models central forecast are slight. Furthermore in an experiment we do not report in detail we tried imposing a constraint that the deficit be down to 2 percent of GNP by 1984.12. That projection showed expenditures lower and revenues higher, with hardly any other change in the forecast relative to the model's unconstrained forecast. The implausibility index for this forecast ranged from .62 to 1.2 for the three methods, indicating that it is not at all unlikely.

In a more extreme experiment, the deficit was constrained to reach zero at 1984.6 and stay there. This conditional projection, and the shock associated with it, based on the M covariance matrix are shown in Tables 10 and 11. This projection has an implausibility index ranging from 3.6 to 12, with the lowest value coming from this variable-parameters projection. This range is large, but of course all put the constraint in the region of great implausibility.

Table 10

Projection Constrained to Produce Deficit of Zero for 1984.6
and Thereafter Percent Changes from Previous Date
(Except TBILL CBI and DEFICIT), Using Matrix M

YR	MO	M1	STOCKS	TBILL	DEBT	PGNP	CBI
83	1	9.48	13.80	8.05	130.08	4.20	-15.27
83	2	9.48	18.84	8.09	113.52	4.44	-12.49
83	3	10.08	26.40	8.28	93.36	5.04	-7.14
83	6	10.24	16.92	8.56	42.24	5.76	-3.20
83	9	9.92	9.00	8.72	-6.12	6.08	-.12
83	12	9.96	12.64	8.93	18.92	7.52	4.68
84	6	9.22	7.66	9.70	18.20	8.34	8.87
84	12	9.22	4.34	9.85	-9.40	9.06	9.76
85	6	9.62	7.18	9.93	10.60	9.00	8.68
85	12	9.42	5.64	10.34	15.08	9.04	7.50
86	12	9.40	4.87	10.58	10.08	9.31	7.82
		RGNP	OUTL	RCPT	TRDOL	DEFICIT	
83	1	9.96	-33.00	-39.24	-.96	199.62	
83	2	6.96	-7.20	-7.20	3.96	198.42	
83	3	11.88	1.20	7.20	1.92	195.73	
83	6	9.48	3.60	5.20	5.40	195.16	
83	9	9.04	-4.80	12.00	4.80	167.95	
83	12	9.44	-18.00	11.60	2.40	116.14	
84	6	8.04	-4.40	29.80	2.44	0.00	
84	12	6.72	11.60	11.60	2.12	0.00	
85	6	5.08	10.20	10.20	2.34	0.00	
85	12	5.40	12.20	12.20	1.08	0.00	
86	12	5.29	15.60	15.60	1.58	0.00	

Table 11

Standardized Shocks Generating the Table 10 Projection

YR	MO	M1	STOCKS	TBILL	DEBT	PGNP	CBI
83	1	0.0	-0.0	0.1	-0.1	0.0	0.0
83	2	0.0	0.1	-0.2	-0.0	0.1	0.1
83	3	0.1	0.3	0.2	0.1	-0.0	0.1
83	6	0.2	0.3	-0.1	0.1	0.1	-0.0
83	9	0.1	0.1	-0.1	-0.4	-0.0	-0.1
83	12	0.3	0.1	-0.1	-0.2	0.3	0.3
84	6	0.0	0.2	0.3	0.1	0.3	0.0
84	12	-0.0	0.1	-0.0	-0.1	0.2	0.1
85	6	-0.0	0.0	-0.1	-0.1	0.1	-0.0
85	12	0.0	0.0	0.0	-0.1	0.1	0.0
86	12	-0.0	0.0	0.0	-0.0	0.0	-0.0
		RGNP	OUTL	RCPT	TRDOL		
83	1	-0.0	0.1	0.1	-0.1		
83	2	-0.4	-0.0	0.0	-0.0		
83	3	0.0	-0.1	0.0	-0.3		
83	6	0.0	-0.0	0.1	-0.0		
83	9	0.1	-0.1	0.1	0.1		
83	12	-0.0	-0.6	0.3	-0.1		
84	6	-0.1	-0.7	0.9	-0.0		
84	12	0.0	-0.3	0.4	-0.0		
85	6	0.0	-0.1	0.2	0.0		
85	12	0.0	-0.2	0.3	-0.0		
86	12	0.0	-0.0	0.0	0.0		

The constraint of a zero deficit by 1984.6 produces noticeable effects on the projections for other variables, with even more rapid expansion than in the central forecast in the period before 1984.6, followed by a sharp reduction in output growth rate and a rise in interest rates when the deficit takes its sharpest drop. This is consistent with a "Keynesian" inter-

pretation that expansion tends to reduce deficits by raising the tax base faster than it raises government spending plans, at least in the short run, and that when taxes are raised and expenditure reduced, there are subsequent contractionary effects on the economy. The model then can be interpreted as saying that the most likely way to arrive at a zero deficit is to have a lucky expansion in output soon, combined with an unusually large rise in taxes and decline in expenditures later.

It should also be noted that there are several ways to model the effects of a correctly anticipated future reduction in the deficit which imply that it would have current expansionary effects on demand, combined with contractionary effects when it actually occurs. Since there is more than one way to get such a result and none of them are simple, we omit laying out such a theory. We only point out that it is possible to interpret the initial expansion in the projection with small future deficit as directly produced by anticipations of the small future deficit.

The model shows less impact of drastic changes in future deficits than many economists would think likely. Though the modest implausibility index for the drastic deficit reduction of Tables 10 and 11 shows that the model's deficit forecasts have shown substantial error in the historical sample, probably announced and believed changes of such magnitude have not occurred before. In that case the conditional forecast in these Tables would not be a good guide to the likely effects of an announced and believed change. On the other hand, if changes of this magnitude have not been announced and believed before, that is reason

to question whether a believable announcement of this type is possible.

The model does systematically associate the lower growth path of the CBO forecast with a sharp initial reduction in the total size of the federal budget, with expenditures and receipts moving down together. This kind of effect suggests either a substantial short run balanced budget multiplier, or real interactions of federal expenditures with the private sector--phenomena which play a minor role in currently fashionable approaches to macroeconomics.

As a kind of consistency check of these results, we also investigated the posterior distribution directly using the Monte Carlo method to integrate various regions and to evaluate conditional expectations. For example, to judge the plausibility of the CBO forecast in another way, we counted how many of our 200 simulations had real GNP growth lower than the CBO projected in 1983 and 1984. There were only four such simulations, confirming the implausibility of the CBO forecast according to our model. In a similar experiment we found 37 simulations had the price level growing less rapidly than the CBO forecast. There was only one simulation which had both real GNP and price level growth lower than the CBO.

A forecast conditioned on low deficits was formed by averaging the 60 simulations with the lowest deficit forecasts for the period 1984:6 to 1986:12. The average deficit path for this group was negative for the period, smoothly declining from current levels to zero in March 1985, and ending the period with a 100

billion dollar surplus. Consistent with the conditional projections above, this subset of the simulations had lower interest rates, higher stock prices, and more rapid growth of money, prices and output. The deficit forecast here is not forced to zero as in the earlier experiment, and growth in output stays above the overall average until late in 1986.

Conclusion

As is clear from these examples, when models like this one are used for policy analysis they yield no automatic causal interpretations. They provide a detailed characterization of dynamic statistical interdependence of a set of economic variables, which may help in evaluating causal hypotheses, without containing any such hypotheses themselves.

Data Appendix

The data for this study consist of ten series of monthly observations for the period 1948:1 through 1983:3. Some of the series were taken directly from sources given below, others were constructed by interpolating quarterly data. Where data was published in seasonally adjusted form, it was used. In other cases, in which only not seasonally adjusted data was published and where there was evidence of a seasonal pattern, the data was adjusted by us prior to use. Details of the data construction procedures are given below. The data set itself, which is based on data published as of May 1983, is available from Litterman for a nominal charge.

The four series which rely on interpolation are real GNP, the Change in Business Inventories, the GNP deflator, and the Flow of Total Nonfinancial Debt. Real GNP is generated as the sum of nine components, three of which are components of consumption and are available on a monthly basis. The other six components, one of which, the Change in Business Inventories, is included separately, and the GNP deflator are based on interpolation of the quarterly National Income and Product Accounts. The Flow of Total Nonfinancial Debt is an interpolation of a quarterly series included in the Federal Reserve Flow of Funds accounts. The interpolations use related monthly series following the procedures of Chow-Lin [1972] and Litterman [1983].

Seasonal adjustment was required for federal government receipts and expenditures and for several of the monthly series used in the interpolations. Below we describe the interpolation

and seasonal adjustment procedures and the steps used for construction of each of the individual series.

Interpolation

Two interpolation procedures are used; one is the Chow Lin [1972] procedure where errors are assumed to follow a first-order markov process, the other is a variation of Chow-Lin when the error process follows a random walk with a first-order markov driving process. The latter method, denoted RW below, was tried first when interpolation was required. It is based on the assumption that the unobserved monthly series of interest, y_t, is related to a vector of observed monthly series, x_t, by the relation:

$$y_t = x_t \beta + u_t$$

The error process u_t is assumed to follow a random walk:

$$u_t = u_{t-1} + e_t,$$

where e_t is a first-order markov process:

$$e_t = \alpha e_{t-1} + v_t .$$

Litterman [1983] shows how to estimate α, β and the monthly values of y, given quarterly averages of y and the monthly values of x. He finds that relative to other standard approaches, this procedure reduces the interpolation error in several cases where quarterly averages of observed monthly data were considered. In cases where the estimated markov parameter, α, for this procedure was negative, however, the RW procedure did not perform well. For this reason, where we encountered negative estimates of α, we used

the Chow-Lin procedure, denoted CL below. In the CL model the error term, u_t, itself follows a first-order markov process.

Seasonal Adjustment

Where seasonal adjustment was necessary the procedure we followed was a frequency domain method based on Nerlove [1964] and Geweke [1978]. In brief, the steps were as follows:

1. Deterministic constant, trend and monthly seasonals were removed.
2. A short order autoregressive representation with seasonal lags was used to forecast and backcast two years of data.[1/]
3. The series with deterministic part removed and extensions appended is fourier transformed and the spectrum is estimated.
4. The fourier transform of the data is divided at seasonal frequencies by the ratio of the estimated spectrum to an estimate of the nonseasonal spectrum at that frequency. The estimate of the nonseasonal spectrum is obtained as a quadratic curve fit across seasonal frequencies to periodogram ordinates at each end of the seasonal band.

[1/] Early versions of the data set, including thoses used for the out-of-sample forecasting experiments left this step out and padded with zeros rather than forecasts. The seasonally adjusted series generated without this step suffered at the ends of the data from a detectable modulation of the seasonal pattern which led to our adoption of this procedure.

5. The adjusted fourier transform is transformed back to the time domain and constant and trend are added.

Individual Series

Money Supply

Seasonally adjusted monthly values for the money supply, M1, as published by the Federal Reserve Board were used for the period 1959:1 to 1983:3. Values for M1 during the period of 1948:1 through 1958:12 were generated by scaling the old M1 series by the ratio of the new to the old value for 1959:1.

Treasury Bill Rate

This series is monthly averages of yields on 3-month Treasury securities.

Stock Price Index

This series is monthly averages of the Standard and Poor's Index of 500 securities prices.

Flow of Total Nonfinancial Debt

This is an interpolated version of the quarterly Flow of Total Nonfinancial Debt published in the Flow of Funds data set by the Federal Reserve Board. The quarterly series was constructed by summing seasonally adjusted Nonfinancial Sector Credit Market Debt and Foreign Corporate Equities and subtracting Credit Market Funds Raised by Foreigners. These series are labeled F394104005, F263164003, and F264102005, respectively, in the Flow of Funds accounts.

The related monthly series used in the CL interpolation were commercial and industrial loans; the change in consumer credit outstanding; the consumer price index, seasonally adjusted; T-Bills; M1; stocks; and a constant and trend.

Because Flow of Funds data are released with essentially a one-quarter lag, the equation relating monthly variables to the quarterly variable together with the projected residuals was used to extend the data set through the first quarter of 1983, for which no quarterly observation was yet available. Also, the flow of debt series begins in 1952, requiring the use of the equation in a similar manner to extend observations back over the first four years of our sample.

Trade-weighted Value of the U.S. Dollar

The Commerce Department's Index of the Weighted Average Exchange Value of the U.S. Dollar was used for the period in which it is available, 1967:1 through 1983:3. For the earlier period a trade-weighted dollar was constructed following the usual formula and weights, except that it was based on only the exchange rates between the U.S. and Germany, France, and the United Kingdom, rather than on the ten countries in the current index. The constructed series was scaled so that the value for 1967:1 coincides with the current index. Over the period 1967:1 through 1969:12 the actual and the constructed indices were observed to move quite closely, differing at any point by less than .3 percent.

Federal Government Outlays

Federal government budget outlays on a unified basis are available from the Treasury Department monthly, not seasonally adjusted from 1968:2. Annual values are published for the prior years in our sample. The earlier annual data was linearly interpolated using the monthly outlays series on a cash basis, which is available for this period. The entire monthly series was then seasonally adjusted as described above.

Federal Government Receipts

The federal government budget receipts series was constructed using data analogous to that available for outlays, and was also seasonally adjusted as described above.

GNP Deflator

The monthly GNP deflator was based on a RW interpolation using monthly data on the Consumer Price Index, the Producer Price Index, and a constant and trend. The two monthly price indices are published in level form on a not seasonally adjusted basis, and thus were seasonally adjusted as described above prior to use in the interpolation.

Change in Business Inventories

The monthly Change in Business Inventories was generated by summing monthly Nondurable and Durable Changes in Business Inventories series which were each separately interpolated. The Nondurable inventories was based on a CL interpolation. The related monthly series were the Net Change in Inventories on Hand and on Order. Wholesale Inventories on Nondurable Goods, Total

Inventories of Nondurable Goods, Finished Inventories of Nondurable Goods, and a constant, trend and dummies for constant and trend over the period 1948:1 through 1957:12, during which the finished goods inventories are not available.

The Change in Business Inventories of Durable Goods series was generated using a CL interpolation. The related monthly series were the Net Change in Inventories on Hand and on Order and the series for durable goods corresponding to those used in the nondurables interpolation.

Real GNP

In addition to the change in business inventories, five other components of Real GNP were interpolated: Real Business Fixed Investment, Residential Investment, Government Purchases, Exports, and Imports. Real Business Fixed Investment was interpolated using the CL method. Related monthly series included the Index of Industrial Production, the level of Contracts and Orders for Plant and Equipment in 1972 dollars, the Composite Index of Capital Investment Commitments, New Orders for Capital Goods, the Treasury Bill Rate, Commercial and Industrial Loans, and a constant and trend.

The interpolation of Residential Investment used the RW method. Related monthly series were New Private Construction in constant dollars; Total Private Construction Put in Place, which was seasonally adjusted and deflated using the GNP deflator; Expenditures on Private Construction of Residential Buildings, which was deflated using the GNP deflator; and a constant, trend and dummies for periods over which the monthly series were not available.

The interpolation of Government Purchases presented a bit of a problem because we could not find series which would explain its movements. We ended up using the RW interpolation method with a constant and trend.

Exports and Imports were interpolated using the CL method. The related series were Merchandise Trade Exports and Imports, respectively, with constant and trend. Both trade series were deflated using the GNP deflator.

REFERENCES

Chow, B. C. and Lin, A., (1971). Best Linear Unbiased Interpolation, Distribution and Extrapolation of Time Series by Related Series. Review of Economics and Statistics, 53, pp.372-375.

Geweke, J., (1978). "The Temporal and Sectoral Aggregation of Seasonally Adjusted Time Series." Seasonal Analysis of Economic Time Series, (A Zellner Ed.) U.S. Department of Commerce.

Hoerl, A. E. and Kennard, R.W., (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics, 12, pp. 55-67.

Kloek, T. and H. K. Van Dijk, (1978). Bayesian Estimates of Equation System Parameters: An Application of Integration by Monte Carlo, Econometrica, 46, pp. 1-20.

Leamer, E. E., (1972). A Class of Informative Priors and Distributed Lag Analysis. Econometrica, 40, pp. 1059-1081.

Leamer, E. E., (1978). Specification Searches. New York: Wiley

Litterman, R. B., (1980). Techniques for Forecasting With Vector Autoregressions. PH.D. Dissertation, University of Minnesota.

Litterman, R. B., (1981). A Bayesian Procedure for Forecasting With Vector Autoregressions. MIT working paper.

Litterman, R. B., (1982). Specifying Vector Autoregressions for Macroeconomic Forecasting. Federal Reserve Bank of Minneapolis Working Paper.

Litterman, R. B., (1983). A Random-walk, Markov Procedure for Distributing Time Series. Journal of Business and Economic Statistics, 1, pp. 169-173.

Nerlove, M., (1964). Spectral Analysis of Seasonal Adjustment Procedures. Econometrica, 32, pp. 241-286.

Shiller, Robert J., (1973). A Distributed Lag Estimator Derived from Smoothness Priors. Econometrica, 41, pp. 775-788.

Sims, C. A., (1981). Macroeconomics and Reality. Econometrica, 48, pp. 1-48.

Sims, C. A., (1981). An Autoregressive Index Model for the U.S. 1948-75. Large Scale Macro Econometric Models (J. Kementa and J. B. Ramsey, Eds.) North Holland.

Sims, C. A., (1982). Policy Analysis with Econometric Models. Brookings Papers on Economic Activity, 1, pp. 107-154.

Stein, C. M., (1974). Multiple Regressions. Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling (I. Olkin, Ed.) Stanford University Press, Chapter 37.