Discussion Paper 14

Institute for Empirical Macroeconomics
Federal Reserve Bank of Minneapolis
250 Marquette Avenue
Minneapolis, Minnesota 55480

May 1989

A NINE VARIABLE PROBABILISTIC MACROECONOMIC FORECASTING MODEL

Christopher A. Sims
Institute for Empirical Macroeconomics and
University of Minnesota

ABSTRACT

This model extends one originally constructed by Robert Litterman in 1980 and used continuously since then to prepare quarterly forecasts. The current version is 3 variables larger than Litterman's original model, and it now allows time variation in coefficients, predictable time variation in forecast error variance, and non-normality in disturbances. Despite this elaboration the model in a sense has just 12 parameters free to fit the behavior of 9 variables in 9 equations. The paper reports the model structure and summarizes some aspects of its recent forecasting performance.

This material is based on work supported by the National Science Foundation under Grant No. SES-8722451. The Government has certain rights to this material.

A NINE VARIABLE PROBABILISTIC MACROECONOMIC FORECASTING MODEL

by Christopher A. Sims

Revised May 1989

Robert Litterman, beginning around 1980, began forecasting aggregate macroeconomic variables using a small Bayesian vector autoregressive (BVAR) model. The model originally used 6 variables -- Treasury bill rate, M1, GNP deflator, real GNP, real business fixed investment, and unemployment. It performed remarkably well relative to forecasts prepared by commercial forecasting organizations using much more elaborate models. (See Litterman [1986].) In particular, as documented by McNees [1986], it performed better than commercial models for real GNP and unemployment, but worse for the price level. It was easy to see from graphs or tables of the forecasts that the model was extrapolating inflation at a long run average rate, despite many quarters in a row of same-signed errors for forecasts made this way. Attempting to rectify this, Litterman added three additional variables to the model -- trade-weighted value of the dollar, Standard and Poors 500 stock price index, and a commodity price index.

With the model in this form, I took over preparing forecasts with it, starting in the fall of 1986. Litterman had noted a tendency for improvements in the retrospective forecast performance[1] of the BVAR model for inflation to be accompanied by deterioration in its performance for real variables. He had chosen his additional variables aiming to minimize the real-variable

[1] I.e., in the simulated forecast performance of the model when it is repeatedly re-estimated from data up to t and used to forecast data at t+s, s=1,...,k, while t ranges over the historical sample period.

deterioration while improving price forecasts. My own analysis suggested, however, that this attempt was not entirely successful. Furthermore, as I took over the model it had been making a sequence of same-signed errors in forecasting real GNP, which, while not as serious as the earlier sequence of inflation errors, were disturbingly similar in pattern. I decided therefore to complicate the specification of the model in several ways, aiming to find a probability model which would track the shifts in trend inflation rates and productivity growth rates while still performing about as well for real variables as Litterman's original simple 6 variable model.

The resulting model differs from Litterman's in several respects.

- It allows for conditional heteroskedasticity (time varying variances of disturbance terms).

- It allows for nonnormality of disturbances. Specifically, it allows disturbances to be mixtures of two normal random variables.

- It takes account of the connection of the constant term to the means of the explanatory variables using a "dummy initial observation," described below.

- It uses the discrete-time process generated by time averaging of a continuous time random walk as a prior mean, rather than using a discrete time random walk.

- Probably mainly as a result of the first three changes, it fits best with a great deal more implied time variation of parameters than Litterman found optimal with his model.

Likelihood is dramatically higher for this version of the model than for its predecessor. Simulated one through eight step ahead forecasts from the sample period are about as good or a bit better than with the previous model for real variables, much better for price variables, and slightly worse for interest rates.

In what follows the probability model is described, the methods used to fit it to data are laid out, the characteristics of the fitted model are summarized, and the model's recent forecasting record is displayed.

1. Description of the model

The data are a time series of kx1 vectors X(t), determined by a state vector $\beta(t;i,j,s)$ and an equation disturbance u(t;i) according to

$$1)\quad X_i(t) = \sum_{j=1}^{k} \sum_{s=1}^{m} X_j(t-s)\beta(t;i,j,s) + \beta(t;i,j+1,1) + u(t;i)$$

We treat the β's and u's as stochastic processes which generate a distribution, conditional on initial X's, for the other observed X's. In principle, inference on all equations of the system should proceed jointly, as randomness in one equation could be correlated with randomness in other equations.[2] However, because it is computationally convenient and because some tentative experiments have indicated little advantage from full system estimation, we proceed equation-by-equation. What we discuss below, therefore, though we call it the "likelihood" is usually

---

[2]In this model the algebra of the "seemingly unrelated regressions" of econometric textbooks applies. Thus even if the randomness is related across equations, if the same X's appear on the right hand side of each equation and the prior has the same form in each equation, then analysis of the whole system reduces to equation-by-equation analysis. However, the prior we consider is not symmetric across equations.

the component of the likelihood corresponding to one equation under the assumption of independence across equations.

A. Form of the distribution of disturbances

Conditional on prior information, data observable through date t-1 <u>and</u> on $\beta(t-1;i,j,s)$, the vector $(\beta(t;i,\cdot,\cdot),u(t;i))$ is taken to be a mixture of two jointly normally distributed random variables, both with mean $(\beta^*(t-1;i,\cdot,\cdot),0)$ and with variance matrices $V(t;i)$ and $\pi_{12}^2V(t;i)$ respectively, i.e. the vector has p.d.f.[3]

$$2)\quad p(\beta(t),u(t;i)|t-1) = \pi_{11}\phi([\beta^*(t-1),0],V(t;i)) + (1-\pi_{11})\phi([\beta^*(t-1),0],\pi_{12}^2V(t;i)),$$

where $\phi(a,b)$ is the p.d.f. of a normal vector with mean a and variance matrix b.

Conditional on data and prior information observable through t-1 <u>alone</u> $\beta^*(t-1)$ is taken to be normally distributed with covariance matrix $W(t-1)$ and mean $B(t-1)$, i.e. it is taken to have p.d.f.

$$3)\quad q(\beta^*(t-1)) = \phi(B(t-1),W(t-1)).$$

If π_{11} were 0 or 1, equations (1)-(3) would justify applying the Kalman Filter to an observation on $X_i(t)$ to obtain a posterior distribution for β and u. With other values of π_{11}, since the conditional distribution of $X_i(t)$ is nonnormal, the Kalman Filter cannot be applied directly. However, the posterior distribution is still easily obtained by two applications of the Kalman Filter. One applies it once conditional on the $V(t;i)$ covariance

---

[3]Here and below we will use the abbreviated notation a(t) for $a(t;i,\cdot,\cdot)$ where there can be no ambiguity.

matrix, then again conditional on the $\pi_{12}^2 V(t;i)$ covariance matrix. The posterior p.d.f. on $(\beta(t),u(t;i))$ is then a weighted sum of the two resulting normal posterior p.d.f's, with the weights given by the relative likelihoods of the observed $X_i(t)$ under the two normal prior distributions.

The posterior distribution on $\beta(t)$ generated by this procedure is, of course, itself a mixture of normals, not a normal distribution. If $\beta(t+1)$ were related to $\beta(t)$ by a linear equation with normal disturbances, the prior distribution on $\beta(t+1)$ would itself be nonnormal and the Kalman filter would not be applicable at t+1. We assume that $\beta^*(t-1)$ is a function of $\beta(t-1)$ such that it has a normal distribution with the same mean and variance as has $\beta(t-1;,\cdot,\cdot)$, despite the nonnormality of the latter.

If we could represent this change in distribution by supposing that some sort of random noise were added to $\beta(t-1)$, it would be natural to think of this as simply non-normal stochastic time variation in a. However, the nature of the change in distribution precludes its being characterized this way. The assumption is in fact unnatural, justifiable only as a convenient approximation. Note, though, that because our uncertainty about $\beta(t-1)$ cumulates the effects of disturbances at many dates, our posterior for a is likely to be much closer to normality than is the distribution for $\beta(t)-\beta^*(t-1)$. Treating the distribution of the former as approximately normal while carefully accounting for nonnormality in the latter is therefore justifiable as an approximation.[4]

Note that we are in effect assuming that our posterior mean for $\beta(t-1)$ at t-1 is the same as our prior mean for $\beta(t)$. This makes the $E[\beta(t)|t]$ sequence a martingale. There would be no computational or conceptual difficulty with allowing a more general linear dependence of the prior mean for $\beta(t)$ on $\beta(t-1)$, and indeed Litterman and I have both, in this and other models, experimented with specifications where

$$E[\beta(t)|t-1] = \theta E[\beta(t-1)|t-1],$$

with θ a scalar. The best choice of θ has always been close to one, however, so that with sample sizes of the length actually available there has seemed little advantage to freeing i to differ from one.

---

[4]Actually, if the prior at t=0 is normal, the prior at t=1 would be a mixture of two normals, so that by conditioning on each normal component of the prior, Kalman filtering twice for each, we could obtain a new posterior which was a mixture of four normals, etc. However with the number of normal components involved proliferating exponentially, this exact approach would be computationally intractable. A better approximate approach might be to continually keep track of the k most likely of the 2^t branches of the tree of normal components of the mixed posterior distributions, with k set at, say, 4 or 16. Or instead one could at each t convert the posterior for $\beta(t-k;i,\cdot,\cdot)$ conditional on data through t-k to the normal distribution with corresponding mean and variance, treating the disturbances from t-k+1 to t exactly. The procedure actually used is this procedure with k=1, but a k of 2 or 3 would be feasible, at least as an experiment to check the sensitivity of results. One hesitates to work too hard at this, since the mixture of normals assumption itself is an arbitrary convenience. A matrix t distribution would be more plausible, implying a continuous mixture of normals in place of a mixture of just two normals.

B. Initial prior mean

In the model discussed here m, the lag length, is 5 -- slightly over a year, since we are using quarterly data. The vector $B(0;i,j,\cdot)$, the initial prior mean on $\beta(1;i,j,\cdot)$, is set to zero for i not equal to j. The vector $B(0;i,i,\cdot)$ is given by

$$1.2679 \quad -.3397 \quad .0910 \quad -.0244 \quad .00654 \quad .$$

These numbers satisfy $B(0;i,i,s)=(1+\alpha)(-\alpha)^{s}$, which defines (if s is allowed to run to infinity, instead of being truncated at 5) the autoregressive coefficients for an ARIMA(0,1,1) process with moving average parameter $\alpha=2-\sqrt{3}$. It can be shown that this is the form of a unit-averaged Wiener process. Thus the prior mean makes all elements of X behave like unit-averaged Weiner processes with no lagged cross-relations among components of X.[5]

C. The initial Litterman prior

The prior covariance matrix is built up by a sequence of modifications of an initial prior. The initial prior makes each scalar component of the $\beta^{*}(t;i,\cdot,\cdot)$ vector independent of all the others (i.e. it makes the covariance matrix W(0) diagonal) and sets the variance according to

---

[5]Note that in previous published work prior means for BVAR models have generally made the components of X discrete time random walks. The unit-averaged Weiner process prior (at least where the data have in fact all been collected as unit averages) is a notably more accurate naive standard, however. Observe that the Theil U's obtained by using the correct AR in place of a discrete random walk AR for a process which is actually a unit-averaged Weiner processes would be, at forecast horizons 1 through 4, .933, .9732, .9832, .9878.

$$4) \quad \sqrt{\mathrm{Var}(\beta^*(0;i,j,s))} = \frac{\sigma(i)}{\sigma(j)} \pi_1 \pi_2 \delta(i,j) \exp(-\pi_3 \log(s)) \quad , \quad j=1,\ldots,k+1;$$

Here $\sigma(i)$ is a parameter measuring the scale of fluctuations in variable i, taken in practice as the residual standard error from a univariate 5th order VAR fit to the entire sample for $i \le k$. For $j=k+1$ there is only an $s=1$ term, as the corresponding a is the "constant term" (here actually not a constant, but time varying.) For this term $\sigma(j+1)=1/\pi_4$, another unknown parameter. The function $\delta(i,j)$ is the Kronecker delta, 1 for $i=j$ and zero otherwise. Here as elsewhere in this paper the parameters π_i are "unknown constants." In principle, we should specify a prior over them to complete a Bayesian framework for inference. However, because doing so would be inconvenient and we expect our prior on them would be uninformative (i.e., we don't know much about them a priori) we integrate over these parameters informally.

D. The dummy intial observation

The range of differences in observed dynamic behavior for economic time series is fairly large, and indeed a reasonable prior specification for the standard error of $\beta^*(0;i,i,1)$ is about .16. But then this component of uncertainty about β^* alone accounts for an implied standard error of forecast for $X_i(1)$ amounting to 16% of the initial level of X_i. Since the random components in the other elements of $\beta^*(0)$ are all independent of this one, they all serve only to increase the implied forecast errors. We are not in fact this uncertain about the accuracy of naive random-walk forecasts (which is what our initial forecasts, based on prior means for β^*, will be). We are unsure of whether our prior means are exactly right, coefficient by coefficient, but we find it much more likely that the best forecasting model will be one which implies that naive no-change forecasts will be fairly good than that it will be one which implies that great improvements on a no-change forecast are possible. If coefficients deviate from their prior means, we expect other coefficients will deviate in an offsetting way, so that naive no-change forecasts will still be fairly accurate.

To capture this aspect of prior beliefs, we need to introduce appropriate off-diagonal elements into W(t;i), while leaving the diagonal elements relatively undisturbed. One easy way to do this is to introduce a "dummy observation" in which the prior is modified by feeding it into a Kalman filter which takes as observed data for X(t-s), s=1,...,m the actual m intial values of X from the sample and for $X_i(t)$ not the actual $X_i(m+1)$ but instead the model's own forecast, based on the prior mean for β^*, of $X_i(m)$. The data in this dummy observation are weighted by a parameter π_5, which can be expected to be best taken to be near one if the variances of the u(t;i) disturbances have been specified as near to the variance of forecasts from a naive random walk model. Because the Kalman filter finds that with these artificial data the prior mean generates perfect forecasts, the Kalman filter makes the posterior mean the prior mean. Only the variance matrix of the prior mean is changed. The change is of rank one and in practice turns out to have only modest effects on diagonal elements of W.

In most previous published work with BVAR's, there has been a "sum of coefficients" modification to the prior. That modification can be characterized as a sequence of Kalman filtering operations indexed by j=1,...,k, in each of which X(t-s) is set to 0 for s=1,...,m except for $X_j(t-s)$, s=1,...,m, all of which are set to 1, while $X_i(t)$, the dependent variable, is set to 1 if and only if j=i. Because most economic time series are smooth, X(t-s) and

X(t-s-1) have similar values. Thus the dummy initial observation used here is approximately a linear combination of the dummy observations used in imposing the sum of coefficients modifications. In practice, the dummy initial observation seems to reduce or eliminate the usefulness of sum of coefficients dummy observations. This point is substantively important, because heavily weighted sum of coefficients dummy observations push the model toward a limiting form written entirely in terms of differences, which eliminates all long-run relations across variables.

This dummy initial observation idea was discovered in the process of adapting BVAR methodology to a context where the number of series available for a model increases at several dates scattered through an historical sample. A natural approach to such a situation is to begin with a prior for a model with all the variables which will eventually be available, padding the data for variables which are initially unavailable with zeroes. Applying the Kalman filter to the padded data is equivalent to applying it to a smaller model. The prior means and variance matrix of the coefficients on unavailable variables are left unaltered by the Kalman Filter when the data for them is set at zero. However, at the time when data on a series does become available, the prior shows an exaggerated version of the problem described in the text. The new data multiply large prior variances to imply large forecast errors, and the uncertainty about coefficients on the newly entering variable shows no correlation with uncertainty about coefficients on the variables already in the model. We know in fact that the small model estimated up to this point is a good forecasting model and the availability of data on a new variable has not made its forecast accuracy worse. To make the prior reflect this knowledge, a dummy observation, in which the prior mean coefficients at t are presented to the Kalman filter as making perfect forecasts for t+1, is appropriate. The prior mean coefficients are all zero for the newly introduced variable, so the dummy observation expresses confidence in the small model estimated without the new variable. Covariances between coefficients are introduced so that deviations from zero in coefficients on the new variable imply likely offsetting changes in other coefficients to leave forecasts from the previously estimated small model fairly accurate.

E. Relative tightness on durable goods prices

There is an a priori basis for expecting that prices of durable goods frequently traded in open markets will follow stochastic processes well approximated as Weiner processes over short time spans. Thus our prior mean is inherently more attractive for such variables than, e.g., for GNP or unemployment. We therefore introduce into (4) an additional multiplicative factor,

5) $\pi_8 \mathrm{IDGP}(j)$,

where IDGP(j) is zero for variables which are not durable goods prices and one for variables which are. The latter are taken to be the value of the dollar, stock prices, and commodity prices. A case could be made for including three month Treasury bill rates and/or M1 in this list, but they were left out as not actually being prices of durable goods.

F. Covariance matrix of disturbances

The upper left diagonal component of the matrix V(1;i) corresponding to all the β's is taken to be $\pi_6^2 W(0;i)$, i.e. just a scaled version of the initial prior covariance matrix. However, the scale of this matrix is allowed to adapt over time to the observed squared errors in the model. The idea here is very close to that of the ARCH models pioneered by Engle [1982], but differs in that instead of variances being adapted to the sizes of past unobservable true disturbances (u and $\beta-\beta^*$ in our notation), they are adapted to the sizes of past actual forecast errors, i.e. in our notation to sizes of

$$v(t;i) = X_i(t) - \sum_{j=1}^{k} \sum_{s=1}^{m} X_j(t-s)B(t-1;i,j,s) + B(t;i,j+1,1) .$$

The specification adopted here has the advantage that it makes the variance of disturbances at t+1 known at t, allowing a single pass of the Kalman filter through the data to evaluate the sample likelihood function.

More specifically the scales of the $V(t;i)$ matrices are adapted to the recent history of forecast errors in all equations of the system according to the following scheme. Let $v^*(t;i;0)$ be $v(t;i)$ divided by the model's implied variance for $v(t;i)$ conditional on the true disturbance matrix being $V(t;i)$, while $v^*(t;i;1)$ is $v(t;i)$ divided by the model's implied variance for $v(t;i)$ conditional on the true disturbance matrix being $\pi_{12}^2 V(t;i)$. Then let

$$v^{**}(t;i)^2 = p_0 v^*(t;i;0)^2 + p_1 v^*(t;i;1)^2,$$

where p_0 is the posterior probability, given data at t, of the smaller variance normal component of the mixed distribution for the disturbance at t, and $p_1=1-p_0$ is the posterior probability of the other component. If the model is correct, v^{**} should average out to about 1. Let

$$6) \quad \tau^*(t) = \pi_9 + (1-\pi_9) \sum_{i=1}^{k} v^{**}(t;i)^2/k \quad .$$

Let

$$7) \quad \tau(t;i) = \pi_7 + (1-\pi_7)v^{**}(t;i)^2 + \pi_{10}(\tau^*(t)-1) + \delta \quad .$$

Then we take

8) $V(t+1;i) = \tau(t;i)V(t;i)$.

If δ=0 in this specification, each $v(t;i)^2$ is a martingale, but since these terms are necessarily positive, they form martingales bounded below. Thus with δ=0 the model implies v(t;i) converges almost surely to a constant. While this implication is perhaps no more unreasonable than the implications of martingale behavior for β itself (which we have imposed), experimentation with δ nonzero seems warranted. The current version of the model takes δ=.01, which slightly improves fit over δ=0, but there has been no systematic exploration of the likelihood surface in δ as there has been for the π vector.

2. Model Fitting

What we have described above is a 12-parameter probability model for the 9 quarterly observed time series in the model. A classically oriented statistician can ignore the Bayesian jargon in the model description, treat the β's as well as the u's as unobserved random disturbances, and intepret the π's as the model parameters. From this perspective, our estimation procedure is simply maximum likelihood. (Though, as mentioned above, since we add up individual equation log likelihoods to form the system likelihood used as the fit criterion, we are in effect assuming independence of all random disturbances across equations, a potentially unrealistic assumption.)

My own view is that maximum likelihood is justifiable only as an approximation to a Bayesian procedure or as a device for summarizing a likelihood function. The most important single aspect of a likelihood function, at least if it has a well defined peak, is its maximum. Nonetheless, we must bear in mind that the peak might not be well defined, or that the shape of the likelihood may otherwise turn out to differ from the usual Gaussian shape. In practice, this means that if likelihood turns out to be insensitive to some dimension of variation in β, we ought to verify that the implications of the model which are important to us -- forecasts and policy analysis -- are also insensitive to this dimension of variation. If not, results from several parameter settings should be studied.

The derivation and interpretation of the likelihood function for this type of model have been described in Doan, Litterman and Sims [1984]. The mechanics of likelihood maximization have been handled with a nonstandard hillclimbing routine, described in Sims [1986a]. Because each function maximization is relatively expensive (involving a pass through the data with two Kalman filter applications at each sample point), it seemed important to use global information about the shape of the likelihood in deciding on each function evaluation. The program used, BAYESMTH, fits a surface to the observed likelihood values to generate a guess for the location of the function's peak. It is applied iteratively, with on the order of 50 function evaluations used to obtain very rough convergence.[6] An advantage of the Bayesian hillclimbing routine is that at any iteration it can be used to generate a best guess at the shape of the likelihood, which is more important for inference than the precise location of the peak.

[6]Iteration is ordinarily halted when, say, 10 or 12 successive function evaluations produce changes in log likelihood of less than .5.

3. Characteristics of the fitted model

Table 1 below shows the π vector which achieved the highest level of the likelihood function. Observe that with π_{12}= 3.7 and π_{11}=.31, the mixed distribution is strongly nonnormal.

Table 2 shows the simulated forecasting performance of the model and of a smaller model with no time variation (but otherwise the same form of prior distribution.) The point would be clearer if the table showed results for an optimized 9-variable model with no time variation, but the table's results are probably representative -- they suggest that time variation is very important in improving forecast accuracy for the GNP deflator, but not for other variables in the 6 variable model.

Table 1

Likelihood-Maximizing π Vector

π subscript	π value	Description of π
1	.17	Overall tightness
2	.19	Relative tightness on other variables
3	1.08	Exponent for increase in tightness with lag
4	2.97	Standard error of constant term relative to $\sigma(1)$
5	1.16	Weight on initial dummy observation
6	.10	Ratio of initial standard error of time variation to initial prior standard error
7,9,10	.84 .61 .9	Determinants of time variation in disturbance variances. See equations (6-8) in text.
8	.37	Relative tightness of prior on durable goods price equations
11	.31	Probability that disturbance is drawn from normal component with larger standard error
12	3.65	Standard deviation of more diffuse of the two normal components of the disturbance distribution, as a multiple of the standard deviation of the less diffuse

Table 2

Theil U Statistics, 1949:3-1987:2

Current 9 variable model result listed above result for 6 variable model with no time variation in each pair of rows.

	Quarters ahead			
Variable	1	2	4	8
Treasury	.9636	1.0379	.9641	.9950
Bill Rate	.9467	.9723	.9567	.8576
M1	.4661	.4232	.3767	.3761
	.4807	.4353	.3968	.4060
GNP	.3892	.3219	.2850	.2592
Deflator	.4471	.4182	.4436	.4664
Real GNP	.7618	.6984	.6968	.6481
	.7523	.7034	.7022	.6857
Business	.8650	.8791	.9356	.9548
Fixed Investment	.9040	.9382	.9698	.9305
Unemployment	.7956	.8554	.9212	.9775
	.8163	.8680	.9568	1.0477
Trade-Weighted Value of Dollar	.9207	.9640	1.0274	1.1715
S&P 500 Stock Price Index	.8775	.9016	.9201	.9915
Commodity Price Index	.7471	.8036	.8727	.8758

Charts 1-9 show the model's forecasting performance since the first time it was used with the Table 1 π vector, in August 1987. This was of course a difficult period to forecast, because of the October 1987 stock market crash. We will not examine the record exhaustively here, but we can point out a few interesting aspects of it.

Revisions in GNP accounts have made it difficult to evaluate forecasting performance. Forecasts of real GNP growth for 1988 were too low by 2 percent or so, and the change in the forecast (between the D87 and 887 lines) generated by the October crash information moved the forecasts in the wrong direction. On the other hand, GNP growth for 87:3, which was apparently predicted almost perfectly at first, was substantially revised upward later. It would be interesting to check how much of the poor year-ahead 1988 forecasting performance for GNP is accounted for by data revision. A similar difficulty arises in interpreting the forecast record for real business fixed investment because of the substantial downward revision in its 88:1 value. Difficulties in deciding how to deflate the computer component of investment may be a source of this instability in the GNP accounts data.

Commodity price inflation forecasts reacted badly to the 87:4 shock. The jump in inflation in that quarter was extrapolated in the forecasts, yet it proved to be a temporary phenomenon.

Unemployment forecasts were good. Before the 88:4 shock, the model was predicting rapid and continuing decline in unemployment for the following three years. At the time, there were probably no other forecasters predicting nearly as sharp a drop in unemployment. The 87:4 shock accurately dampened the optimism of this forecast, while maintaining the prediction of a declining unemployment rate. It is interesting to note the contrast between the results for real GNP and unemployment. One might have expected that the inaccuracy of the GNP forecasts would have corresponded to similar inaccuracy in the unemployment forecasts. That the unemployment forecasts were good, and revised in the right direction by the 87:4 data, may lend credence to the idea that index number problems in the GNP data are part of the explanation for the poor model performance on GNP forecasts.

Charts 10-13 show some experiments to determine the degree of nonlinearity in the model. In each chart the solid line labeled "85" is the forecast actually prepared in March 1989, extrapolating the 1989:1 treasury bill rate as 8.5%. The other three lines represent forecasts in which the bill rate is set at 6%, 10%, and 11%. Note that 6% and 11% are above and below 8.5% by equal amounts, while 10% is more than halfway from 8.5% to 11%.

The alternative forecasts were prepared in each case by using current quarter data on interest rate, money stock, stock prices, and value of the dollar in the following way. A forecast based on data through 1988:4 was prepared. Data for 1989:1 was set to the forecast values, except for the four variables with current quarter data available. The model coefficient estimates were updated using the mixed forecast and actual values for 1989:1 as if it were actual data. Forecasts were then generated for subsequent quarters with coefficients held fixed at their newly estimated values.

Since the interest rate enters the model untransformed, if the model were linear, the responses to each of these deviations from the 8.5% interest rate level would have the same form, differing only in scale, with the scale factors -1 (for 6%), 1 (for 11%) and .6 (for 10%). Since we re-estimate the model before forecasting, the effects of the variations in 1989:1 interest rate would be somewhat nonlinear in any case, but without the substantial time variation allowed in the model coefficients the nonlinearity would be small. Note that at longer time horizons the forecast interest rate path for the 11% shock is considerably more than twice as far from the base forecast line than is the 10% path. The same nonlinearity is evident at long time horizons in each of the other graphs.

Observe also that the responses to an interest rate shock by GNP deflator inflation and by commodity price inflation are opposite in sign. It is difficult to see how this result fits conventional Keynesian, monetarist, or real business cycle theories.

4. Conclusion

This model represents a further step in a research program attempting to bring into the realm of explicit probabilistic theory more of our uncertainty about the way the economy works. The model has been used for forecasts with the same parameters (the π vector) since mid-1987 and has performed reasonably well. It is a part of a sequence of models which have been used for forecasting since 1980, all of which have made forecasts without any add factors or ad hoc adjustments in response to current data over the entire period of record.

The form of the model has some implications for developments in macroeconomic theory which aim at explaining observed data. The model has substantial time variation in its coefficients, which is essential to generating good forecasts for some variables. Theories which imply stationary linear models for observed data will therofore inevitably fall short. Rational expectations theorists, who have taken the lead in developing explicitly stochastic models, have not yet generated econometrically usable structural models capable of fitting a world of stochastically drifting parameters.

Recently a number of authors (e.g. Bernanke [1986], Blanchard [1986], and myself [1986]) have explored the use of convenient schemes for interpreting stationary VAR models. It is either discouraging or challenging, depending on your point of view, to note that just as tools for convenient identifaction of stationary VAR models begin to be widely used, evidence emerges that stationary VAR models are inadequate. The problem of generating convenient identification schemes for the nonlinear, nonnormal model laid out in this paper appears quite difficult.

Notes to Charts 1-8

Each chart plots four forecasts, together with the "actual" data available at the time of the forecast for the preceding two (for rates of change) or three (for levels) quarters. The forecast plots are labeled with the months in which they were prepared: 8/87, 12/87, 6/88, and 3/89. In each case some monthly data were available on some of the model's variables for the quarter in which the forecast was prepared, but other variables, including GNP accounts, were available only for the preceding quarter. The plots show a vertical line at the quarter through which all data were available at the time of preparation of each forecast. Data revisions and the use of incomplete current data to estimate current-quarter values of some variables cause the plots in some instances to fail to lie on top of one another when it appears from the dating that they should.

Notes to Charts 10-13

Each chart plots the publicly distributed March 1989 forecast (solid line) and three other forecasts based on the same data, except for variations in the assumed 1989:1 interest rate. The interest rates assumed are 6%, 8.5% (the public forecast's value), 10%, and 11%. The method for incorporating the current information into the forecast is described in the text. The vertical line is at 1988:4, the last date for which current data for all variables was available at the time of forecast.

Data Description

Treasury Bill Rate: Three month treasury bill rate, discount on new issues

M1: Pre-1959 data on M1, spliced together with more recent data on M1B. The splicing is done simply by scaling the earlier data to match the level of M1B at the date of switch.

PGNP: GNP deflator.

GNP82: GNP, 1982 prices.

BFI82: Business Fixed Investment in 1982 prices from the GNP accounts.

UNEMP: Unemployment rate, total labor force excluding military overseas.

DOLLAR: Federal Reserve Board trade-weighted index of the value of the U.S. dollar.

STOCKS: The Standard and Poors 500 Stock index.

PR28: Index of producer prices for 28 sensitive materials.
Source: Bureau of Labor Statistics Bureau of Economic Analysis

REFERENCES

Bernanke, B. [1986]. "Alternative Explanations of the Money-Income Correlation", in Carnegie-Rochester Policy Series on Public Policy, Amsterdam: North-Holland.

Blanchard, Olivier and Mark Watson [1986]. "Are All Business Cycles Alike?" in The American Business Cycle, R.J. Gordon, ed. (Chicago: U. of Chicago Press).

Engle, Robert F. [1982]. "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation", Econometrica v. 50, p.987-1008.

Doan, T., R. Litterman and C. Sims [1984]. "Forecasting and Policy Analysis Using Realistic Prior Distributions", Econometric Reviews v.3, no.1, p.1-100.

Litterman, Robert B. [1986. "Forecasting with Bayesian Vector Autoregressions -- Five Years of Experience", J. of Business and Economic Statistics v.4, no.1, p.25-38.

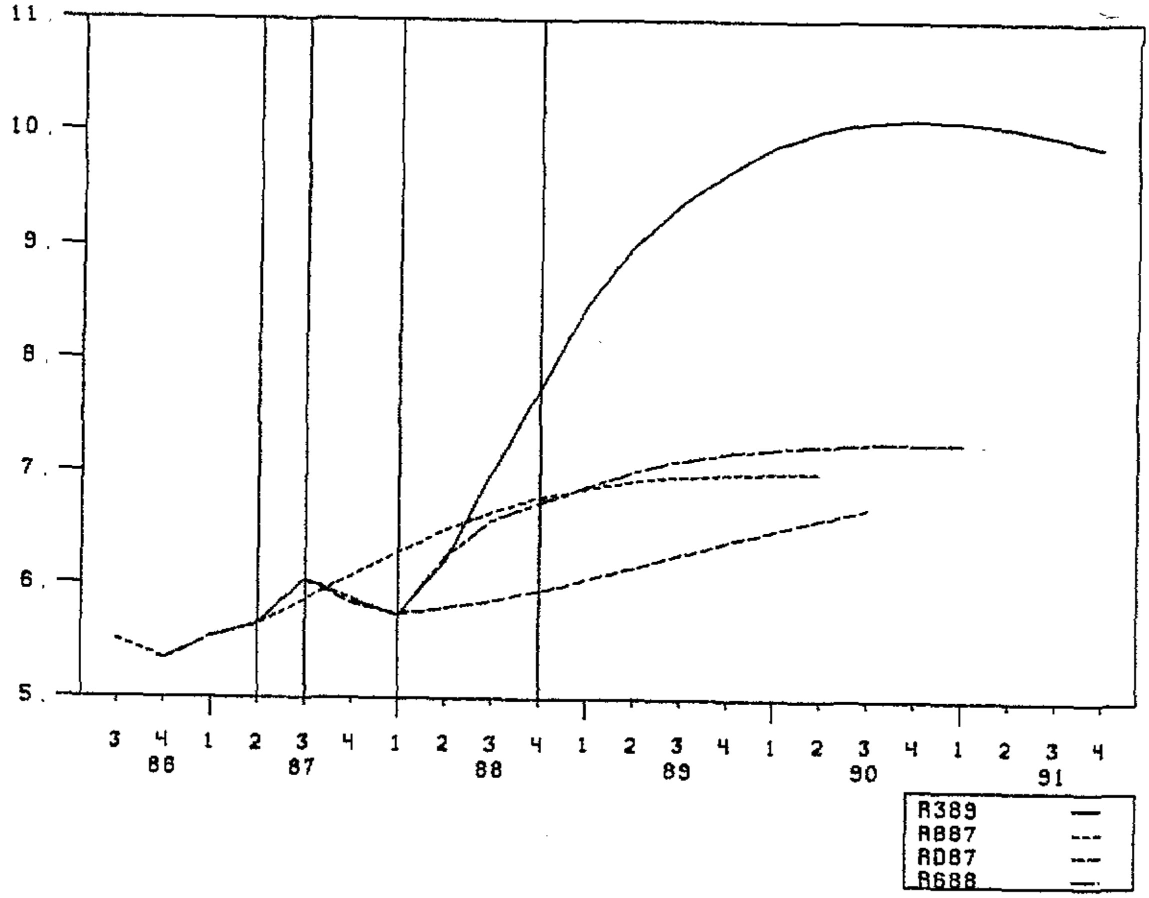
McNees, Stephen K. [1986]. "Forecasting Accuracy of Alternative Techniques: A Comparison of U.S. Macroeconomic Forecasts", J. of Business and Economic Statistics, v. 4, no. 1, p.5-24.

Sims, Christopher [1986a]. "BAYESMTH: A Program for Multivariate Bayesian Interpolation", University of Minnesota Center for Economic Research Discussion Paper no. 234.

________________ [1986b]. "Are Forecasting Models Usable for Policy Analysis", Quarterly Review of the Federal Reserve Bank of Minneapolis, Winter, p.2-16.

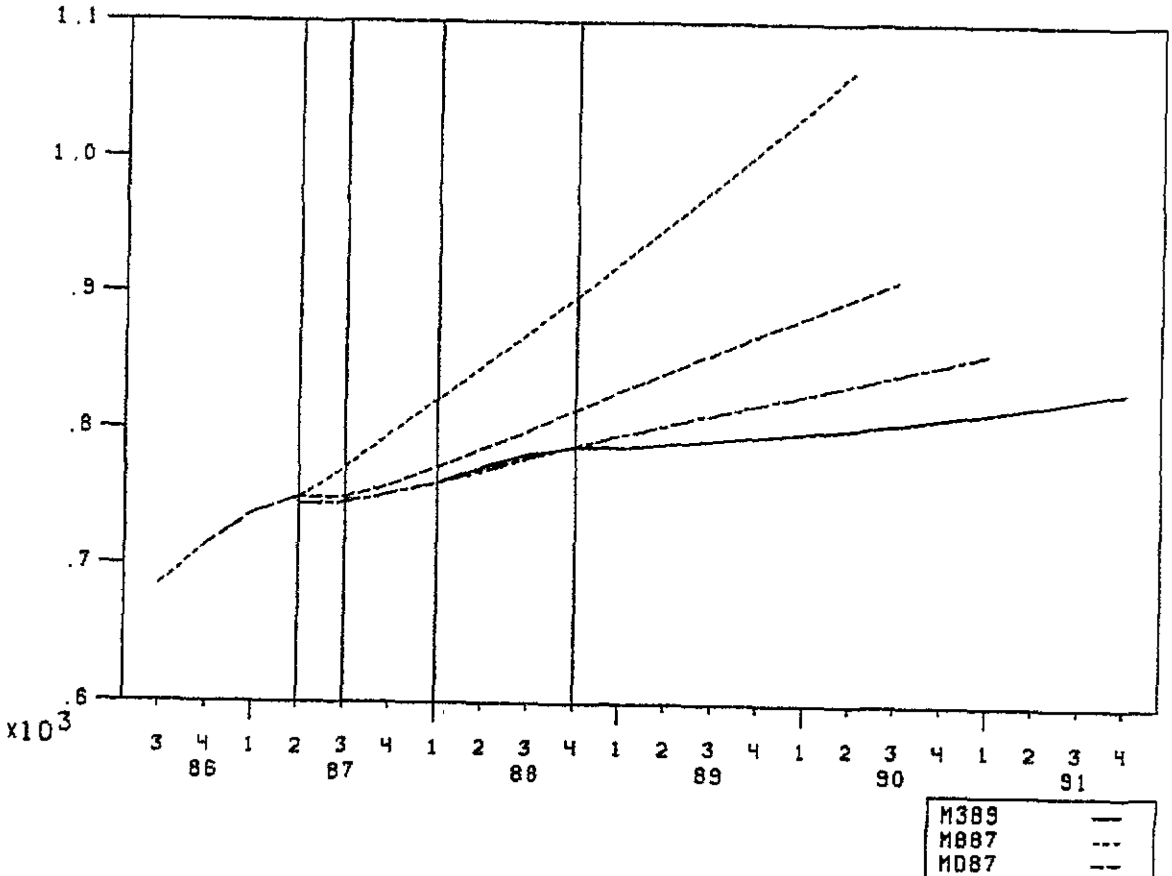
1

TREASURY BILL RATE FORECASTS



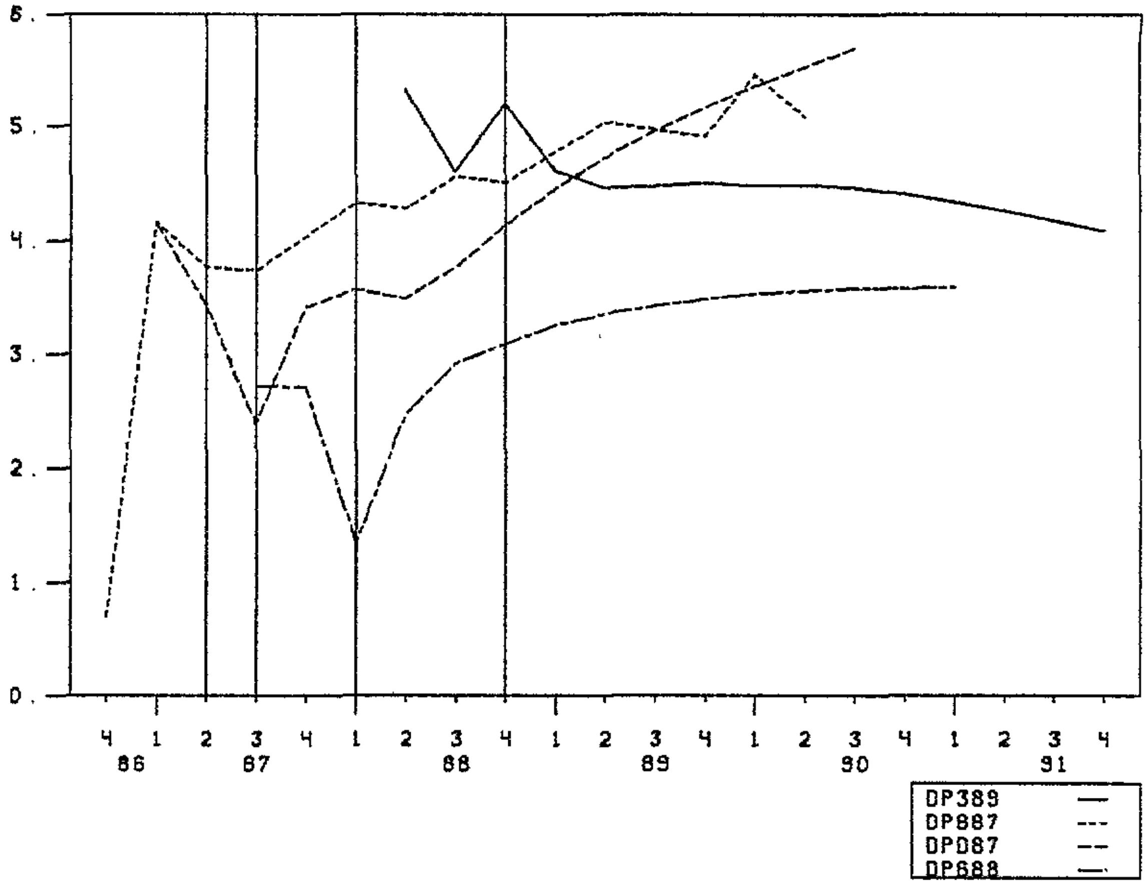
2

MONEY STOCK FORECASTS



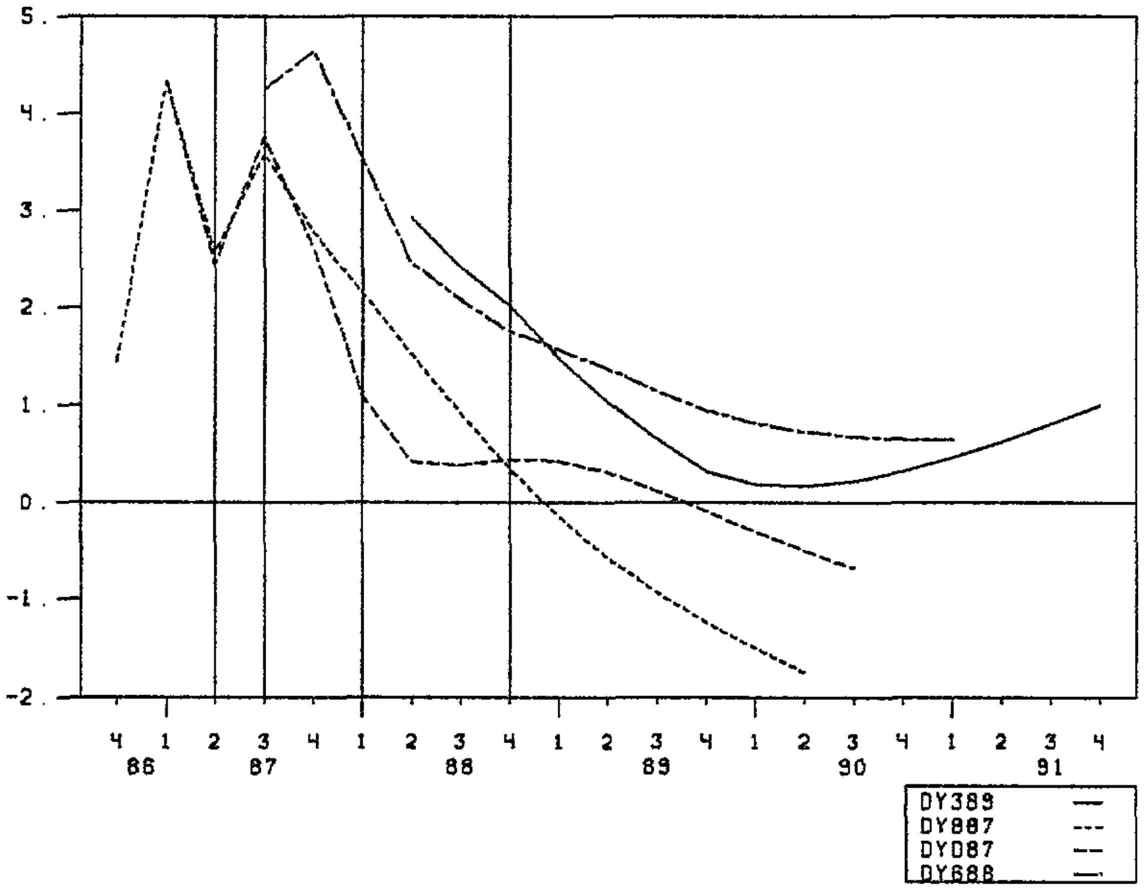
3

INFLATION FORECASTS

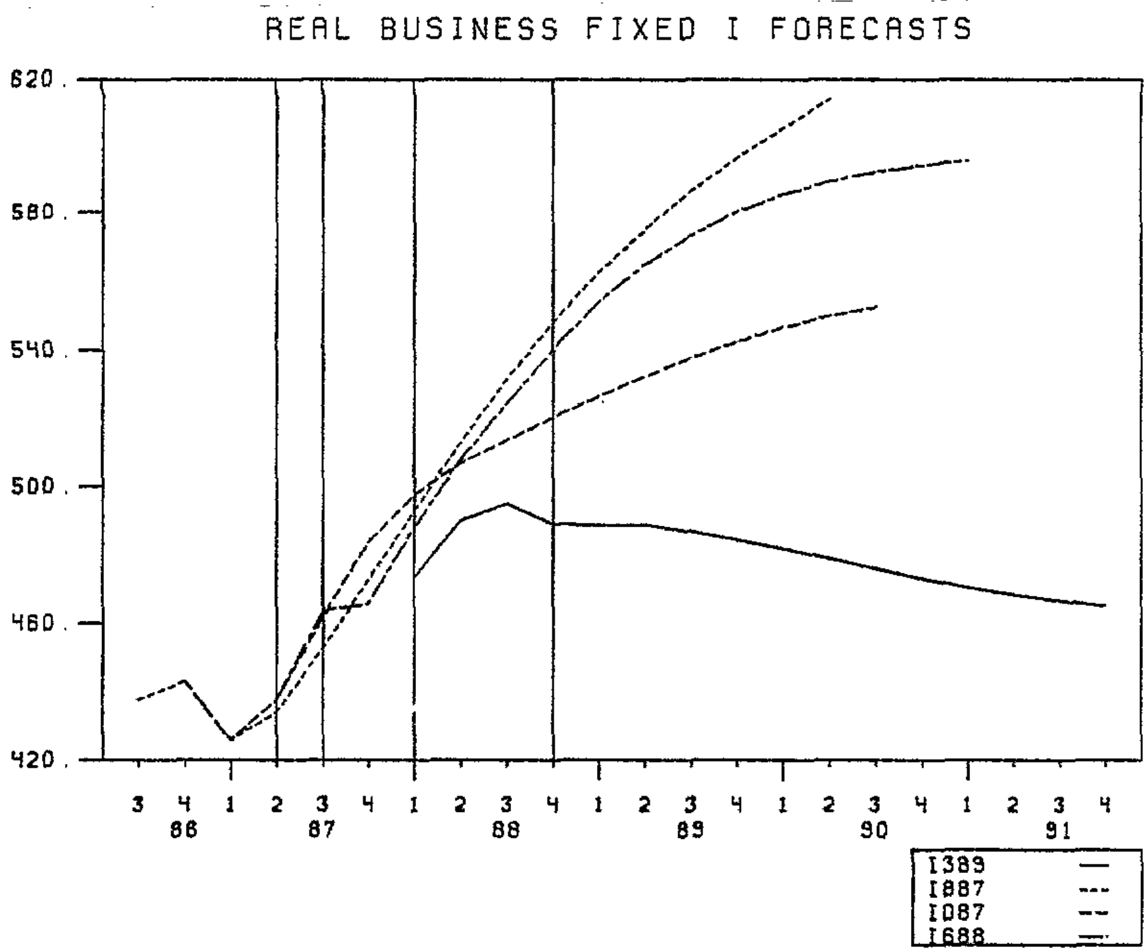


4

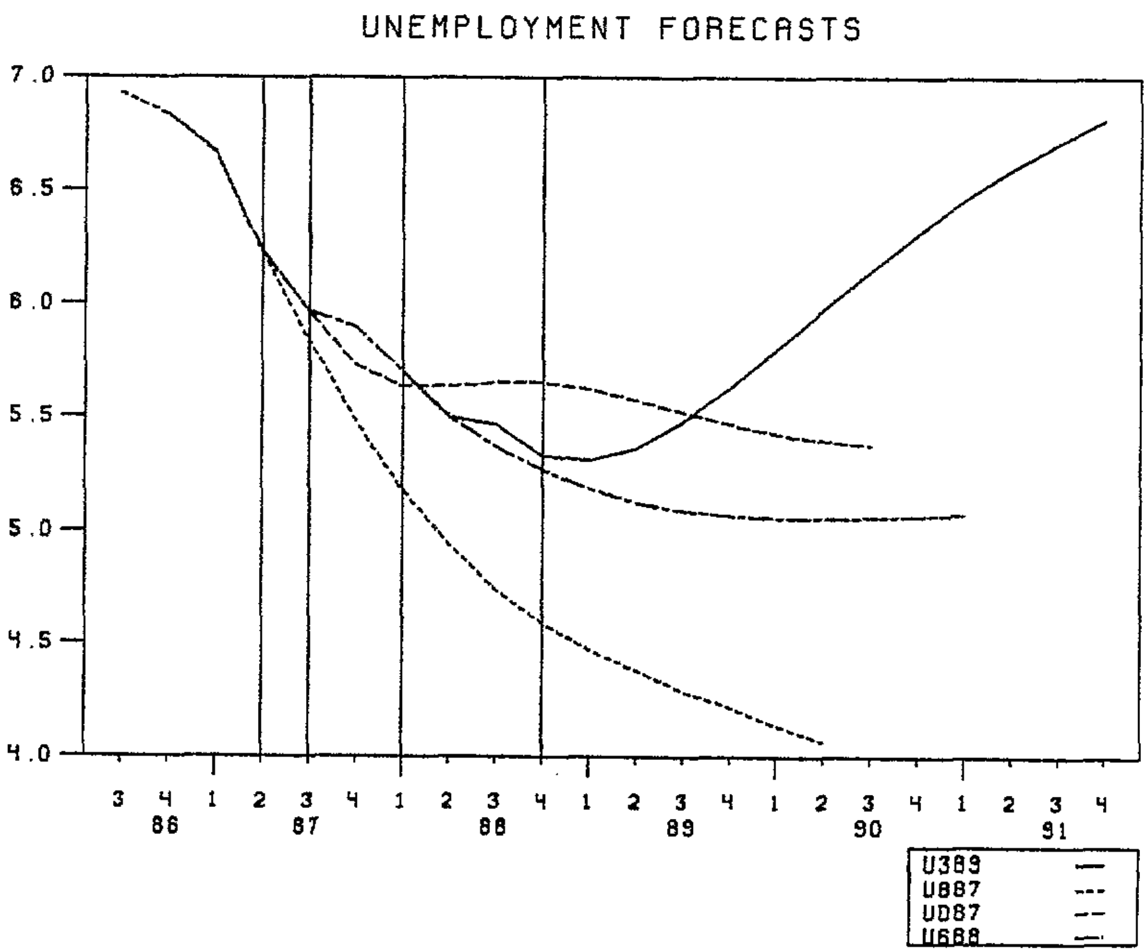
REAL GNP GROWTH FORECASTS



5

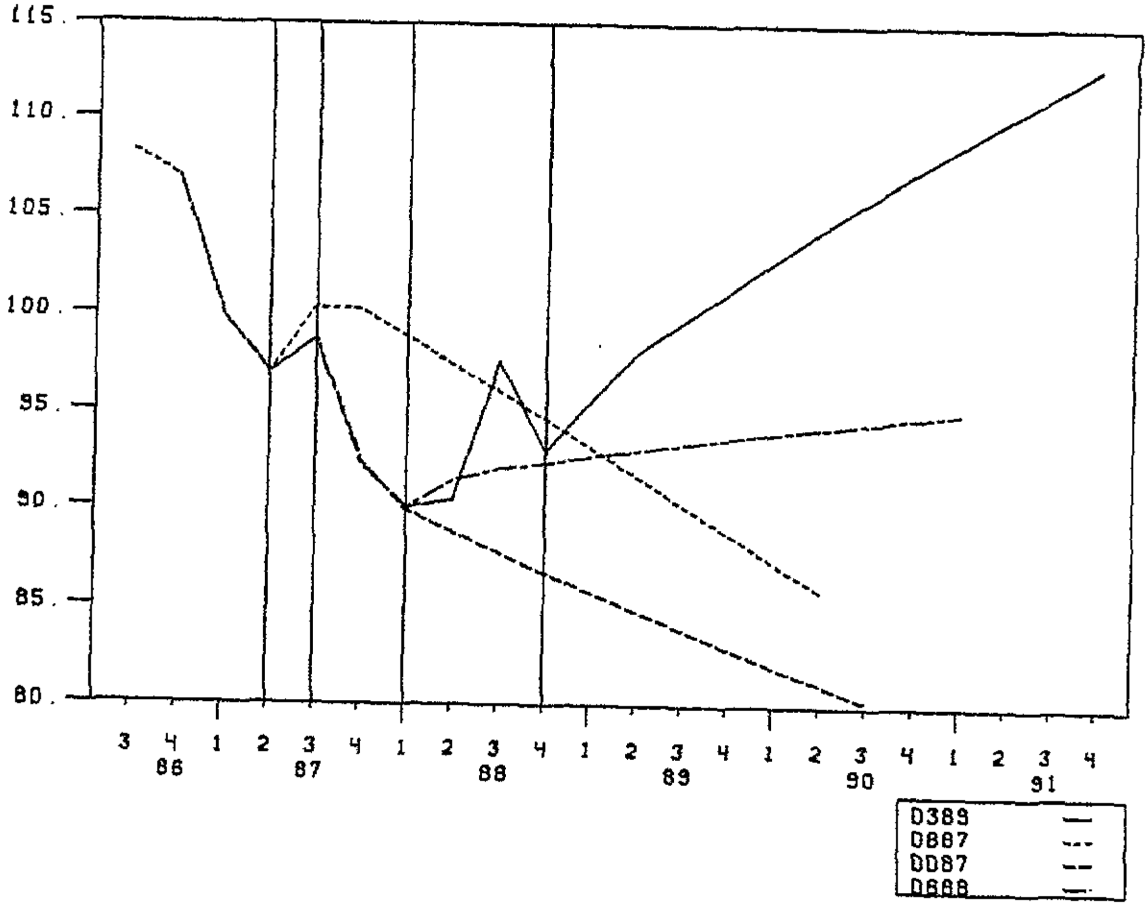

REAL BUSINESS FIXED I FORECASTS
620.
580.
540.
500.
460.
420.
3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4
86 87 88 89 90 91
I389
I887
ID87
I688

6


UNEMPLOYMENT FORECASTS
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4
86 87 88 89 90 91
U389
U887
UD87
U688

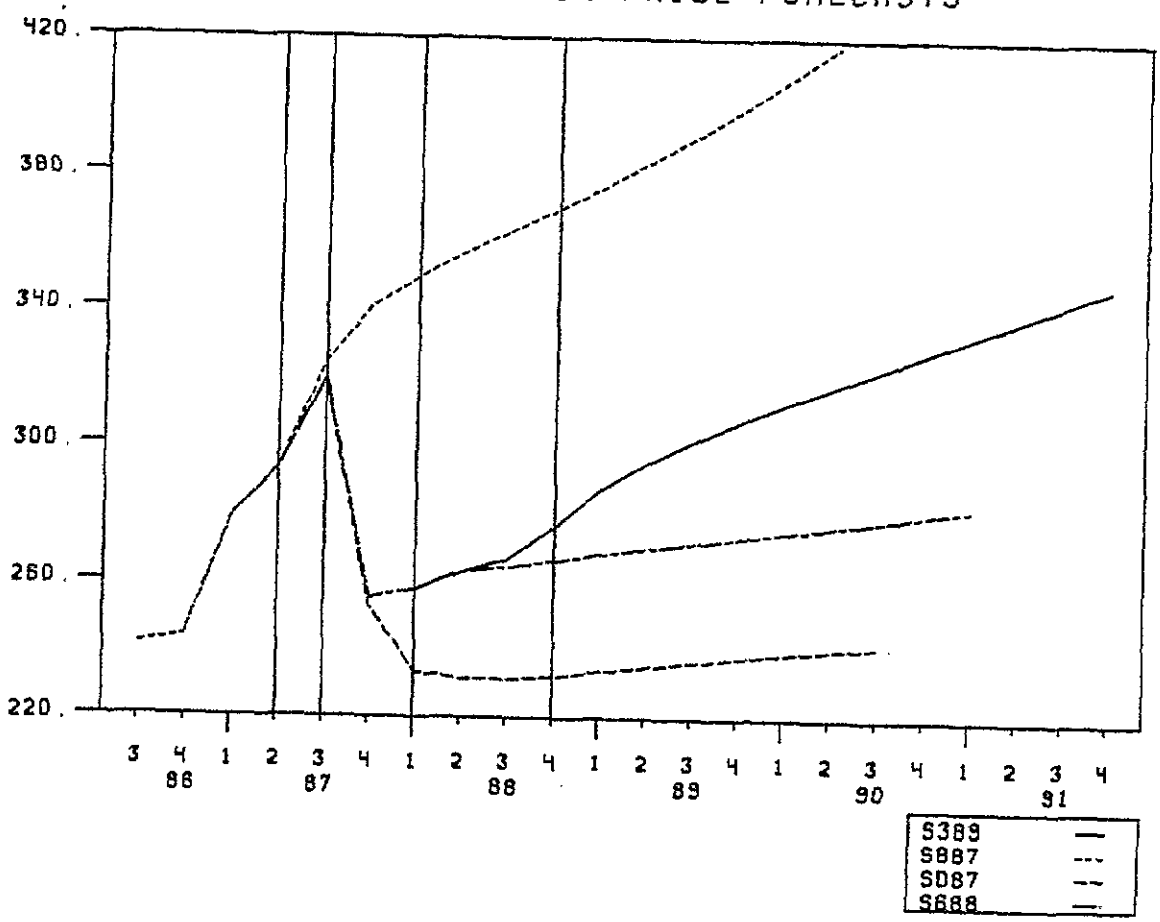
7

INDEX OF DOLLAR"S VALUE FORECASTS



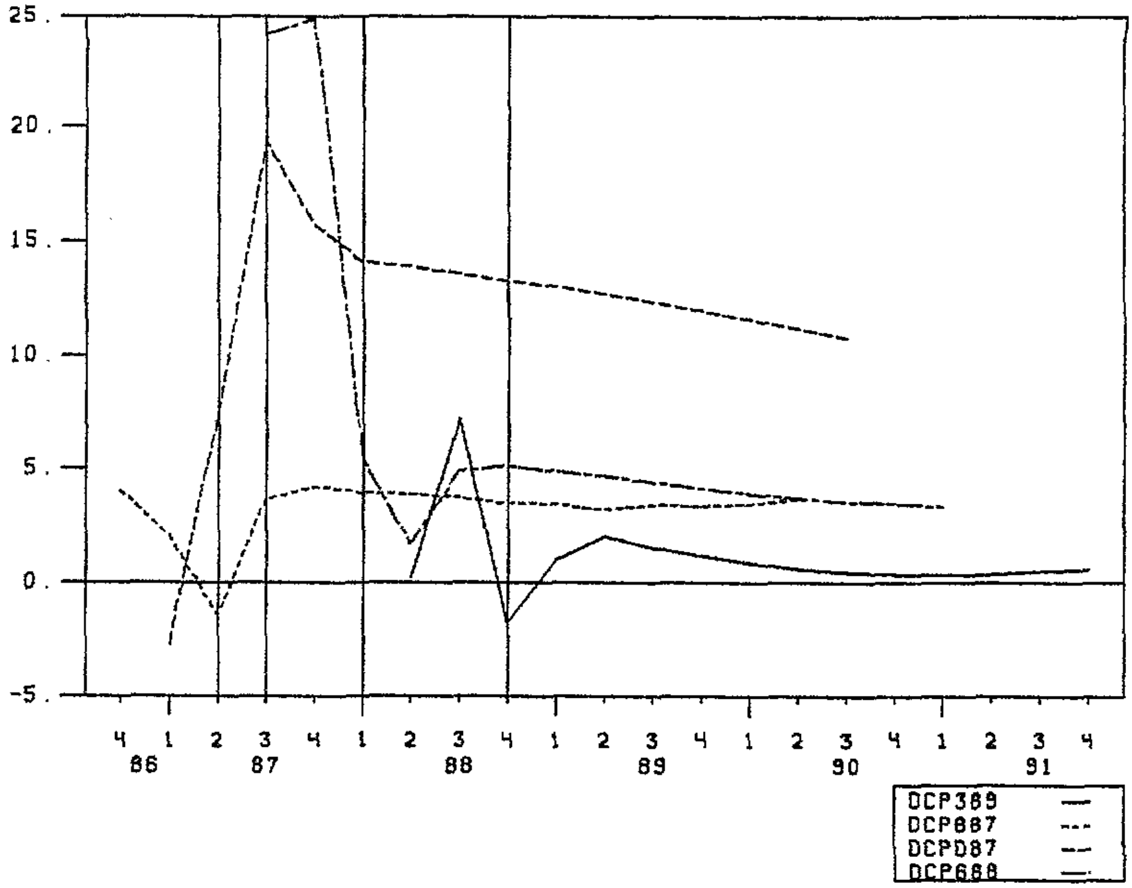
8

S&P 500 STOCK PRICE FORECASTS



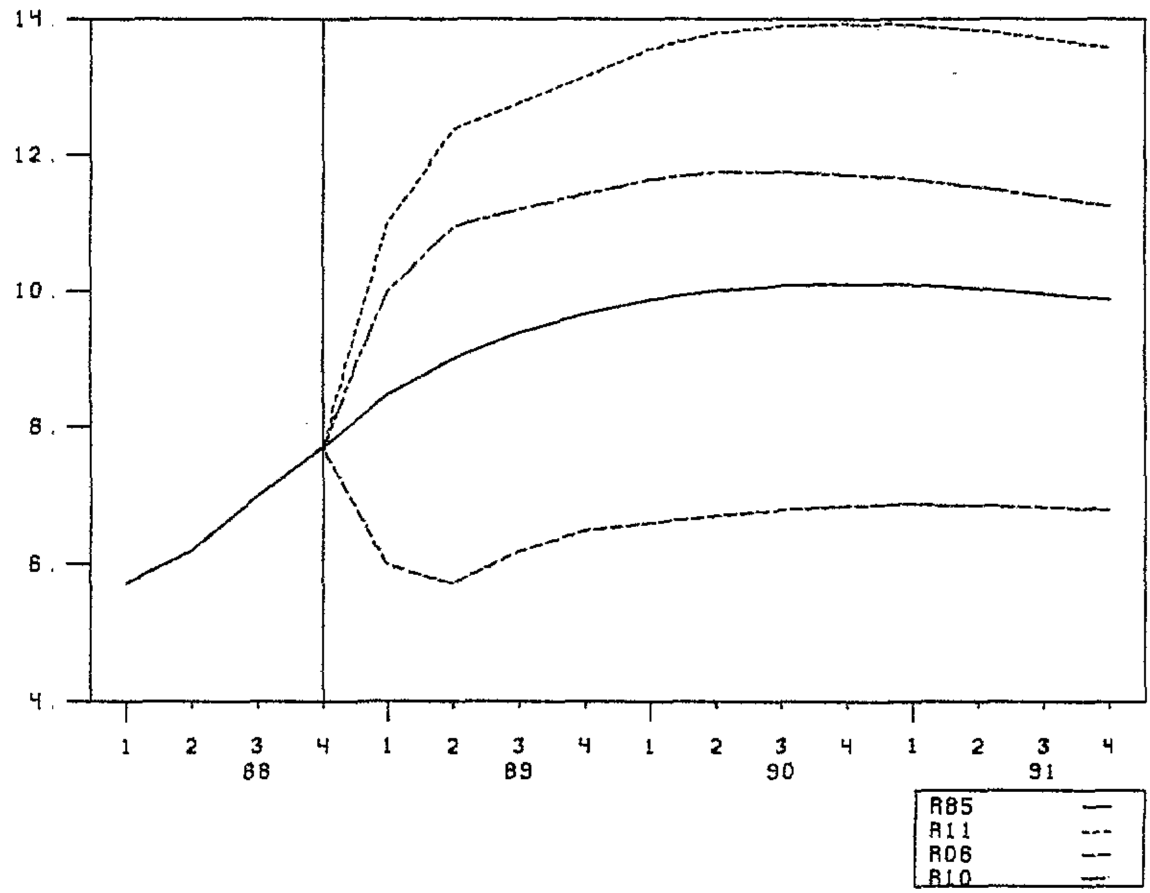
9

COMMOD. PRICE INFL. FORECASTS



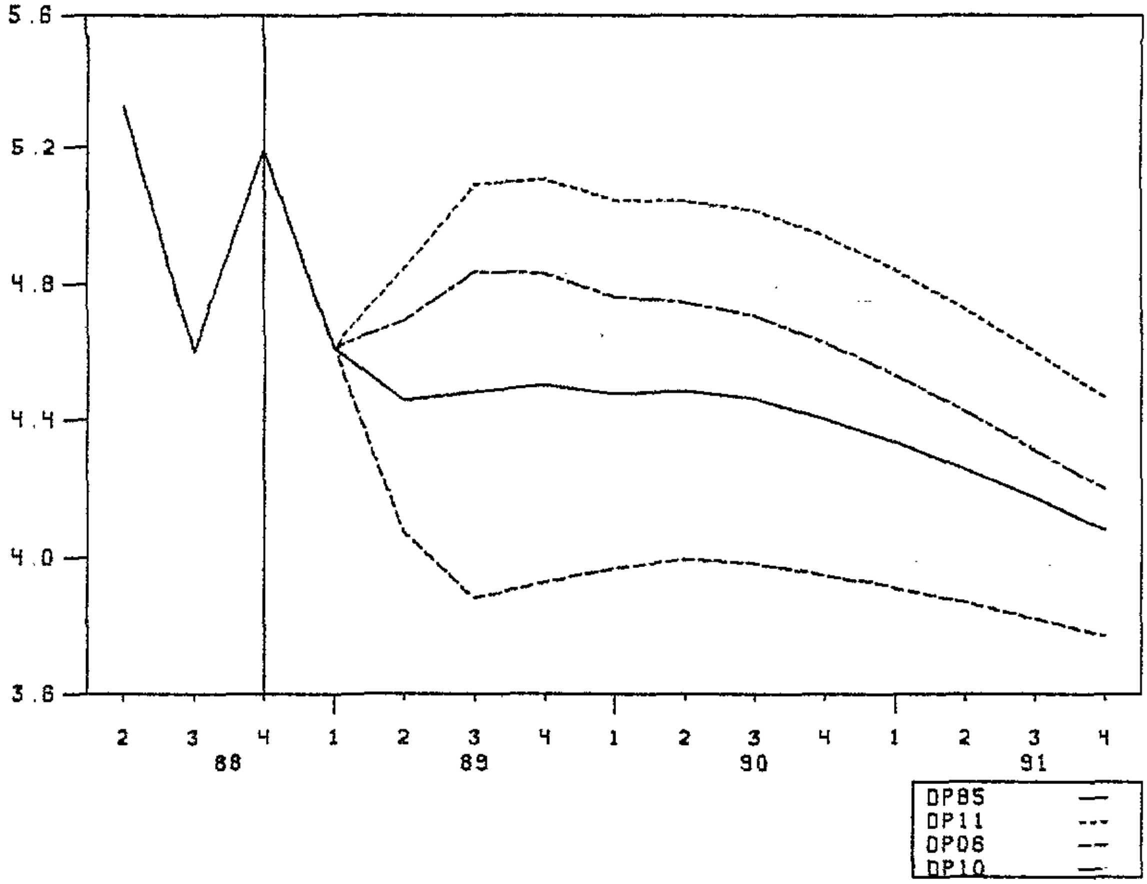
10

R SHOCK EFFECTS: T-BILL RATE



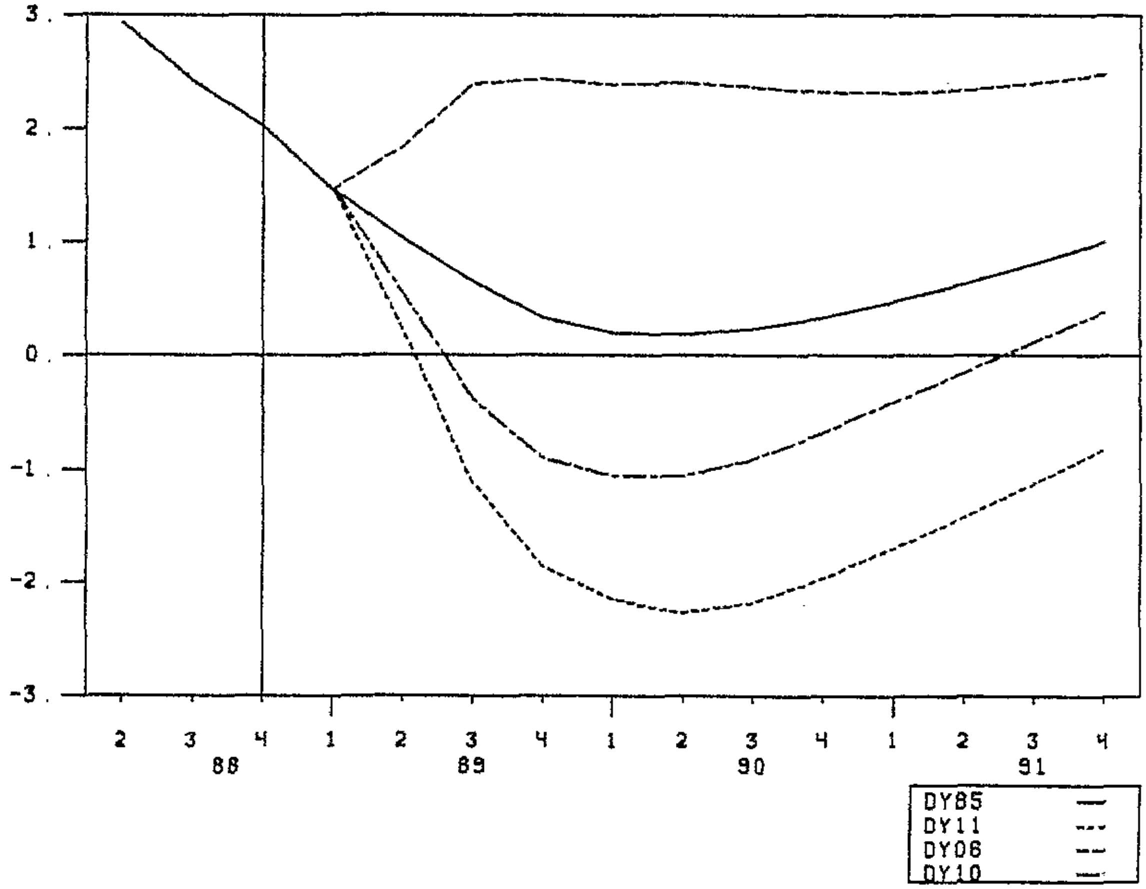
11

R SHOCK EFFECTS: INFLATION



12

R SHOCK EFFECTS: GNP GROWTH



13

A SHOCK EFFECTS: COMMODITY INFLATION