Techniques of Forecasting
Using Vector Autoregressions

Robert B. Litterman

Revised November 1979

Working Paper #: 115

PACS File #: 2700

 This research has been supported in part by the Social Science Research Council.

I would like to thank Thomas Sargent and Christopher Sims for many valuable suggestions in the preparation of this paper.

CHAPTER I
VECTOR AUTOREGRESSIONS

Introduction

Large, simultaneous equation econometric models are routinely used today for forecasting and policy analysis in both business and government. Nevertheless, there has existed for many years a persistent current of economic thought which questions the validity of using these models to project economic variables. Many economists[1/] feel that the equations in typical econometric models are formulated primarily by means of spurious economic theory. This paper presents an alternative methodology for projecting economic time series. The method is basically statistical in nature. The model of the economy which underlies this work is that of a linear stochastic difference equation, more commonly referred to in this context as a vector autoregression (VAR).

Each element of a vector of economic variables of interest is regressed on its own lagged values and the lagged values of every other variable in the system. The particular techniques described here use Bayesian priors to apply a consistent set of restrictions for the purpose of minimizing the mean square error of forecasts.

The vector autoregression specification is very general. It is capable of modeling arbitrarily well any covariance stationary stochastic process. The main weakness of this specification, and the reason it has not been used extensively in the past for economic forecasting, is that the number of free parameters increases quadratically with the number of variables in a system, and for even moderately-sized systems the model becomes highly overparameterized relative to the number of available observations.

Given quarterly postwar data, for example, a model with five lags and approximately 20 variables would have close to zero degrees of freedom. Estimation of such a model without restrictions would give a near perfect fit of the data, but the mean square error of out-of-sample forecasts would be very large.

Several common types of macroeconomic models may be viewed as vector autoregressions with particular classes of restrictions. The reduced forms of traditional simultaneous equation econometric models are special cases of vector autoregressions. Such models are estimated and identified through the imposition of huge numbers of exclusionary restrictions on structural equations and the restrictions implied by the categorization of variables into exogenous and endogenous.

Another common forecasting method, the autoregressive-integrated-moving-average (ARIMA) models popularized by Box and Jenkins [1970], generate stochastic processes which, under the usual invertibility assumption, have autoregressive representations. While such models have theoretical virtues and are capable of capturing long lag distributions with parsimonious parameterizations, they have several drawbacks. The addition of a moving average, which differentiates ARIMA models from vector autoregressions, causes a loss of linearity which makes estimation, statistical inference, interpretation, and prediction more difficult even in the univariate case, and the difficulties increase dramatically with multivariate ARIMA models.[2/] The univariate ARIMA models which are commonly used may be viewed as multivariate models which have severe cross-equation exclusionary restrictions.

The equilibrium solutions of rational expectations models[3/] are another special case of vector autoregressions. Here the assumption of optimizing behavior of agents in the economy generally leads to a set of complicated, cross-equation restrictions.

One characteristic the simultaneous equation, ARIMA, and rational expectations models have in common is that the restrictions are imposed with complete certainty. In contrast, the procedure suggested here will add information in the form of a probability distribution.

It is primarily the recent criticisms of typical large simultaneous equations models by Lucas, Sargent, and Sims which has motivated this work. Nevertheless, such criticism is not entirely new. Liu [1960], for example, suggested that the structures of such models are identified only through the "omission of relevant variables," and that consequently "least-squares-reduced-form equations are likely to be the best forecasting equations."

A recent summary of the limitations of large overidentified models is contained in Sims [forthcoming]. Sims concludes that, "claims for identification in these models cannot be taken seriously," and that, "a more systematic approach to imposing restrictions could lead to capture of empirical regularities which remain hidden to the standard procedures and hence lead to improved forecasts and policy projections."

The main thrust of Lucas' [1976] criticism of the large econometric models, that their structures do not remain invariant with respect to contemplated policy changes and that therefore they are not useful for conditional forecasting, applies equally forcefully to vector autoregression specifications. The solution to the problem of constructing models which are policy invariant seems to be along the lines sketched by Hansen and Sargent [forthcoming]. They summarize the techniques required for the application of restrictions derived from rational expectations hypotheses.

On the other hand, for the problem of unconditional forecasting, one implication of rational expectations theory is that the imposition of exclusionary restrictions on behavioral equations so common in large-scale modeling is not justified. The usual method of imposing restrictions seems to be to assume that no variables enter a particular equation other than those for which there is a particular economic theory to justify their inclusion. When expectations, which in general are conditioned on the past values of all variables in the system, enter decision functions, then the opposite assumption would seem to be more appropriate, that in general it is likely that movements of all variables affect the behavior of all other variables.

The organization of this thesis is as follows: The first chapter gives a brief survey of the theory and estimation techniques for vector autoregressions. In Chapter II the solution to the problem of overparameterization through the addition of instrumental information in the form of Bayesian priors is described. Chapter III considers the problem of making and evaluating forecasts. Various uses of the VAR forecasting method are described in Chapter IV, and the final chapter extends the model to the case of time-varying parameters.

Representation Theory

The statistical model underlying the vector autoregression procedure is a linear dynamic system with an (nx1) vector of outputs, Y, which is generated by a stochastic difference equation. The vector Y can include variables thought to be endogenous or exogenous, and it might include, for example, GNP, investment, interest rates, prices, money supply, and employment. Each variable is treated as a linear function of its own lagged values and lagged values of each of the other variables plus a random disturbance.

We begin by modeling nondeterministic, zero mean covariance stationary processes. Later we will relax the assumption of stationarity. A natural mathematical setting for the discussion of stationary time series is the Hilbert space with norm

$$||Y||^2 = E[Y'Y] \tag{1}$$

where E is the expectation operator.

The linear projection operator, $P[\cdot|S]$, on the complete linear space S is defined by: $||P[Y|S]-Y|| = \min_{X\varepsilon S} ||X-Y||$. $P[Y|X_1,X_2,\ldots,X_K]$ refers to the projection of Y onto the space X^n, the n-dimensional cartesian product space, where X is the space spanned by the nk components of X_1, X_2, ..., X_K.

It is possible to show that linear projections of Y(t) on the spaces spanned by the sets $\{Y(t-1)\}$, $\{Y(t-1),Y(t-2)\}$, ... converge to a random variable, say, $\hat{Y}(t)$.[4/] Letting $\hat{Y}^m(t) = P[Y(t)|Y(t-1),\ldots,Y(t-m)]$ then for any $\delta > 0$ there exists an $N(\delta)$ such that $||\hat{Y}(t)-\hat{Y}^m(t)||^2 = E[\sum_{i=1}^{n} (\hat{Y}_i(t)-\hat{Y}_i^m(t))^2] < \delta$ for all $m > N(\delta)$. Define the Y process innovation as $\varepsilon(t) = Y(t) - \hat{Y}(t)$. Since $\hat{Y}(t)$ is the projection of Y(t) on $\{Y(t-1),Y(t-2),\ldots\}$, $\varepsilon(t)$ is orthogonal to the space spanned by past Y's, which also includes all past ε's. Thus $P[\varepsilon(t)|Y(t-1), Y(t-2),\ldots,\varepsilon(t-1),\varepsilon(t-2),\ldots] = 0$.

Thus, the projection based on a finite past can approximate the projection based on the infinite past arbitrarily well. Furthermore, since $\hat{Y}(t)$ minimizes $||Y(t)-Z||$ over Z on the space spanned by past Y's, these projections are optimal linear predictions in the sense of minimizing the error variance. Now assume Y is generated by an m^{th}-order stochastic difference equation of the form

$$\underset{nx1}{Y(t)} = \underset{nx1}{D(t)} + \underset{nxn}{B_1}\ \underset{nx1}{Y(t-1)} + \ldots + \underset{nxn}{B_m}\ \underset{nx1}{Y(t-m)} + \underset{nx1}{\varepsilon(t)} \tag{2}$$

where D(t), the deterministic component of Y(t), typically might include a polynomial in t and seasonal dummies. Under mild regularity conditions the B_j's are uniquely determined by the population orthogonality conditions

$$E[\varepsilon'(t)Y(t-j)'] = 0 \quad j=1,2,\ldots,m. \tag{3}$$

The above representation can be written more compactly using lag operator notation as

$$(I-B(L))Y(t) = D(t) + \varepsilon(t) \tag{4}$$

with $B(L) = \sum_{j=1}^{m} B_j L^i$, and L is the lag operator defined by $L^j(Y(t)) = Y(t-j)$. The nondeterministic part of Y is given by

$$Z(t) = Y(t) - (I-B(L))^{-1}D(t) \tag{5}$$

which has the moving average representation

$$Z(t) = [I-B(L)]^{-1}\varepsilon(t) = M(L)\varepsilon(t) = \sum_{j=0}^{\infty} M_j\varepsilon(t-j) \tag{6}$$

where each M is an nxn matrix and $M_0 = I$. The M's and B's are related by the matrix Fourier transform relation

$$(I-B_1e^{-iw} - \ldots - B_me^{-iwm})^{-1} = \sum_{j=0}^{\infty} M_je^{-iwj}. \tag{7}$$

The autoregressive representation generates a broad class of stochastic processes. According to a theorem of Wold, any stationary stochastic process can be represented as the sum of a deterministic component and a nondeterministic component which is representable as a moving average. Thus, all stationary stochastic processes for which the moving average component is invertible can be represented as an infinite-order autoregression and approximated arbitrarily well by finite-order autoregressions. In addition, a wide class of nonstationary series may be represented by an autoregressive model if the first m values of an m^{th}-order process are taken as predetermined or given a probability distribution.

The optimal linear projection property of the vector autoregressive representation for covariance stationary stochastic processes is an important motivation for the choice of autoregressive estimators in the analysis which follows. The assumption of stationarity, however, is not needed and is in some ways inappropriate in this investigation.

One purpose of the stationarity assumption is that it provides a basis for the limiting behavior of time series estimators as the number of observations grows. Asymptotic theory is usually relied upon largely because it is the only theory which is available to describe the properties of these estimators.

In this thesis, however, the focus is an attempt to find complex multivariate interactions in small samples of data. In applications of the forecasting procedure to be developed here there are many parameters, the number of degrees of freedom is small, and large sample theory simply does not apply. In its place a Bayesian justification will be developed.

Another purpose for assuming stationarity is to restrict the parameter space. In the univariate, first-order process

$$Y(t) = \lambda Y(t-1) + \varepsilon(t), \qquad (8)$$

for example, stationarity is equivalent to the restriction $|\lambda| < 1$. More generally, in the m^{th}-order vector autoregression (1), stationarity of the non-deterministic part is equivalent to the restriction that the roots of the determinantal equation

$$|Iz^m - B_1 z^{m-1} - \ldots - B_m| = 0 \tag{9}$$

be less than one in absolute value.

For many economic variables such as the levels of employment, output, prices, and productivity there is no reason to assume stationarity. Of course, generally the assumption of stationarity is applied to variables which have been first transformed by, for example, first differencing, or taking logs and removing time trends. However, if the motivation for a transformation is simply to induce stationarity, then the subsequent imposition of stationarity is no longer a meaningful restriction on the data. In any case, as a restriction on the parameter space, stationarity does not seem to deserve universal application any more than any other particular restriction derived from economic theory.

Autoregressive representations have been analyzed extensively in the literature on time series. Recent expositions include Anderson [1971], Fuller [1976], and Hannan [1970]. The form of the autoregression model in (1), with a deterministic component and roots possibly outside the unit circle, is suggested, in a univariate context, in Chapter 8 of Whittle [1963].

Estimation Theory

In a vector autoregressive system with n variables there are n separate equations, each of which has the same explanatory variables. In a system with m lags of each variable and deterministic component D(t),[5/] a function of the (nxd) matrix of parameters C, the i^{th} equation has the following scalar form:

$$\begin{aligned} Y_i(t) = d^i(t) &+ b^i_{11}Y_1(t-1) + b^i_{21}Y_1(t-2) + \ldots + b^i_{m1}Y_1(t-m) \\ &+ b^i_{12}Y_2(t-1) \qquad\qquad\qquad + \ldots + b^i_{m2}Y_2(t-m) \\ &+ b^i_{1n}Y_n(t-1) \qquad\qquad\qquad + \ldots + b^i_{mn}Y_n(t-m) \\ &+ \varepsilon_i(t) \end{aligned} \tag{10}$$

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

We now derive the conditional likelihood function. Suppose we have observations on Y(t), t = -m+1, -m+2, ..., 0, 1, ..., T generated by equation (1). Let $\varepsilon(t) = u(t)$ where u(t) is distributed as multivariate normal, $N(0,\underset{nxn}{\sum_u})$, independent in time. The log likelihood for u(t) is

$$L(u(t)) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\log|\textstyle\sum_u| - \frac{1}{2}u(t)'\textstyle\sum_u^{-1}u(t) \tag{11}$$

and the joint log likelihood is

$$L(u(t), t=1,\ldots,T) = -\frac{Tn}{2}\log 2\pi - \frac{T}{2}\log|\textstyle\sum_u| - \frac{1}{2}\sum_{t=1}^{T} u'(t)\textstyle\sum_u^{-1}u(t). \tag{12}$$

When Y(-m+1), ..., Y(0) are taken as fixed, (1) defines a 1-1 transformation of Y(1), ..., Y(T) into u(1), ..., u(T) with unit Jacobian. Thus, we can substitute [(I-B(L))Y(t)-D(t)] for u(t) and write the log likelihood for Y(1), ..., Y(T) given Y(-m+1), ..., Y(0) as

$$L(y(t),t=1,\ldots,T|\Sigma_u,B(L),C) = -\frac{Tn}{2}\log 2\pi - \frac{T}{2}\log|\Sigma_u| \quad (13)$$
$$-\frac{1}{2}\sum_{t=1}^{T}[(I-B(L))Y(t)-D(t)]'\Sigma_u^{-1}[(I-B(L))Y(t)-D(t)].$$

Σ_u will be positive definite at the maximum, so that a condition for L to be maximized with respect to Σ_u is that $\frac{\partial L}{\partial \Sigma_u^{-1}} = 0$.

Let

$$\underset{n\times 1}{u}(t,B(L),C) = \hat{u}(t) = [(I-B(L))Y(t)-D(t)], \quad (14)$$

then a first-order condition for the maximization of L is given by

$$\frac{\partial L}{\partial \Sigma_u^{-1}} = \frac{T}{2}\Sigma_u + [\frac{\partial}{\partial \sigma^{ij}}(-\frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{n}\hat{u}_i(t)\sigma^{ij}\hat{u}_j(t))] = 0 \quad (15)$$

where $\Sigma_u^{-1} = [\sigma^{ij}]$. The square brackets indicate matrices whose i, j^{th} element is given inside.

Let $\underset{n\times n}{S} = \frac{1}{T}\sum_{t=1}^{T}\hat{u}(t)\hat{u}(t)'$, then the above condition implies $\Sigma_u = S$, which is true for any values of B(L) and C. We can then form the concentrated log likelihood function L* by substituting S for Σ_u to get

$$L^*(Y(t),t=1,\ldots,T|B(L),C) = -\frac{Tn}{2}\log 2\pi - \frac{T}{2}\log|S| \quad (16)$$
$$-\frac{1}{2}\sum_{t=1}^{T}\hat{u}(t)'S^{-1}\hat{u}(t)$$
$$= -\frac{Tn}{2}(\log 2\pi - 1) - \frac{T}{n}\log|S|.$$

In general we minimize log $|S|$ with respect to B(L) and C in order to maximize the likelihood function. It is a standard result that when the right-hand-side variables are the same in all equations, as they are in the unrestricted VAR which we have in (1), minimization of log $|S|$ is solved by minimizing the sum of squared residuals in each equation separately.[6] Thus, OLS estimates equation by equation are maximum likelihood estimates conditioned on the initial observations. More generally, a variety of alternative assumptions on the properties of the Y's and ε's will insure consistency and asymptotic normality of the least squares estimates.[7/]

Under the assumption of stationarity one might want to use the unconditional likelihood approach which includes the observations on Y(t), t = -m+1, ..., 0 in the likelihood function. The asymptotic distribution of maximum likelihood estimator is the same in either case. When only small samples are available, one might expect the unconditional likelihood approach, which includes more information, to generate better results.

One case in which the two approaches will be very different is when the observations appear to be generated by nonstationary processes. In particular, using conditional maximum likelihood it is possible to estimate $\hat{B}(L)$, which would generate a nonstationary process, whereas this is not possible with unconditional maximum likelihood. In the latter case, estimates which approach regions of nonstationarity imply a large variance of the observed process which causes the likelihood function to explode downward. If one suspects that stationarity may not be a valid assumption, as we often do here, unconditional maximum likelihood does not make sense, even in small samples.

Another case in which the two methods will differ is when the initial observations are in a region which would be given low probability by the stationary distribution of the process implied by the conditional maximum likelihood estimator. Unconditional maximum likelihood may also be more biased than conditional, even when the initial observation is likely, if it is nonetheless fixed. These points are illustrated in monte carlo experiments which follow. The experiments generate the distributions of the unconditional and conditional maximum likelihood estimates in a univariate, first-order process under different initial conditions and true parameter values.

In each experiment 30,000 sets of 10 observations for a scalar variable $Y(t)$, $t=1, 2, \ldots, 10$ are generated according to

$$Y(t) = \rho Y(t-1) + \varepsilon(t) \qquad \varepsilon(t) \sim N(0,1).^{8/} \tag{17}$$

In the first three experiments $Y(0)$ is fixed. In experiment 1, $\rho = .999$ and $Y(0) = 0$; in experiment 2, $\rho = .9$ and $Y(0) = 0$; and in experiment 3, $\rho = .9$ and $Y(0) = 6.88$. In the last two experiments $Y(0)$ is drawn randomly from the stationary distribution of the Y process implied by ρ, that is, $N(0,(1-\rho^2)^{-1})$. In experiment 4, $\rho = .999$; and in experiment 5, $\rho = .9$.

In each of the experiments the unconditional and conditional maximum likelihood estimates of ρ, $\hat{\rho}_u$, and $\hat{\rho}_c$, respectively, were obtained for each set of observations on Y. Following Anderson [1971], Section 6.11, let

$$P_0 = \sum_{t=1}^{9} Y^2(t) \qquad P_0' = Y^2_{(0)} + Y^2_{(10)} \qquad P_1 = \sum_{t=1}^{10} Y(t)Y(t-1). \tag{18}$$

Then

$$\hat{\rho}_c = P_1/(P_0+Y^2(0)) \tag{19}$$

and $\hat{\rho}_u$ is the root less than 1 in absolute value which solves

$$-\frac{10}{11}\rho_u^3 + \frac{9}{11}P_1\rho_u^2 + (\frac{12}{11}P_1 + \frac{1}{11}P_0^1)\rho_u - P_1 = 0. \tag{20}$$

For each experiment we generate the means of the estimators given the N = 30,000 drawings,

$$\bar{\hat{\rho}}_u = \frac{1}{N}\sum_{i=1}^{N}\hat{\rho}_u(i) \text{ and } \bar{\hat{\rho}}_c = \frac{1}{N}\sum_{i=1}^{N}\hat{\rho}_c(i) \tag{21}$$

where $\hat{\rho}_u(i)$ and $\hat{\rho}_c(i)$ are the estimates in the i^{th} drawing, and mean square errors of the distributions of these estimates

$$MSE_u = \frac{1}{N}\sum_{i=1}^{N}(\hat{\rho}_u(i)-\rho)^2 \qquad MSE_c = \frac{1}{N}\sum_{i=1}^{N}(\hat{\rho}_c(i)-\rho)^2. \tag{22}$$

In addition, for each drawing we calculate the unconditional and conditional maximum likelihood estimates of the error variances, $\hat{\sigma}_u^2$ and $\hat{\sigma}_c^2$ using

$$\hat{\sigma}_u^2 = \frac{P_0'-2P_1\hat{\rho}_u+(1+\rho_u^2)P_0}{11} \tag{23}$$

and

$$\hat{\sigma}_c^2 = \frac{P_0'-2P_1\hat{\rho}_c+(1+\hat{\rho}_c^2)P_0+(\hat{\rho}_c^2-1)Y^2(0)}{10} \tag{24}$$

and their means.

$$\bar{\hat{\sigma}}_u = \frac{1}{N}\sum_{i=1}^{N}\hat{\sigma}_u(i) \text{ and } \bar{\hat{\sigma}}_c = \frac{1}{N}\sum_{i=1}^{N}\hat{\sigma}_c(i). \tag{25}$$

The results of these experiments are given in Table 1. The distributions of $\hat{\rho}_u$ and $\hat{\rho}_c$ are plotted in graphs 1 to 5.[9/] The first three experiments illustrate the fact that even when the true autoregressive process which generates the observations is stationary, if the initial observations are fixed rather than drawn from the implied distribution of the process, unconditional maximum likelihood estimation in small samples may give very biased results. This bias may be particularly strong when the fixed initial observations are not likely, given the stationary distribution of the process. In experiment 3, for example, Y(0) = 6.88 is an observation three standard deviations from the mean of 0. The distribution of unconditional maximum likelihood estimates given this initial observation is badly biased upward due to the small likelihood of such an observation given a $\rho = .9$.

On the other hand, the maximum likelihood estimator may be biased relative to the conditional likelihood estimator even when the fixed initial observation is quite likely, as in experiments 1 and 2 where it is equal to 0., the mean of the process.

TABLE 1

EXPERIMENT 1: $\rho = .999$, $Y(0) = 0$

$\bar{\hat{\rho}}_u = .765$	$\bar{\hat{\sigma}}_u = .87$	$MSE_u = .104$
$\bar{\hat{\rho}}_c = .863$	$\bar{\hat{\sigma}}_c = .89$	$MSE_c = .088$

EXPERIMENT 2: $\rho = .9$, $Y(0) = 0$

$\bar{\hat{\rho}}_u = .689$	$\bar{\hat{\sigma}}_u = .85$	$MSE_u = .099$
$\bar{\hat{\rho}}_c = .767$	$\bar{\hat{\sigma}}_c = .90$	$MSE_c = .089$

EXPERIMENT 3: $\rho = .9$, $Y(0) = 6.88$

$\bar{\hat{\rho}}_u = .975$	$\bar{\hat{\sigma}}_u = 1.23$	$MSE_u = .0074$
$\bar{\hat{\rho}}_c = .873$	$\bar{\hat{\sigma}}_c = .887$	$MSE_c = .0061$

EXPERIMENT 4: $\rho = .999$, $Y(0) \sim N(0,(1-\rho^2)^{-1}(\approx 500.2))$

$\bar{\hat{\rho}}_u = .978$	$\bar{\hat{\sigma}}_u = .978$	$MSE_u = .0056$
$\bar{\hat{\rho}}_c = .979$	$\bar{\hat{\sigma}}_c = .886$	$MSE_c = .0073$

EXPERIMENT 5: $\rho = .9$, $Y(0) \sim N(0,(1-\rho^2)^{-1}(\approx 5.26))$

$\bar{\hat{\rho}}_u = .798$	$\bar{\hat{\sigma}}_u = .927$	$MSE_u = .047$
$\bar{\hat{\rho}}_c = .797$	$\bar{\hat{\sigma}}_c = .894$	$MSE_c = .056$

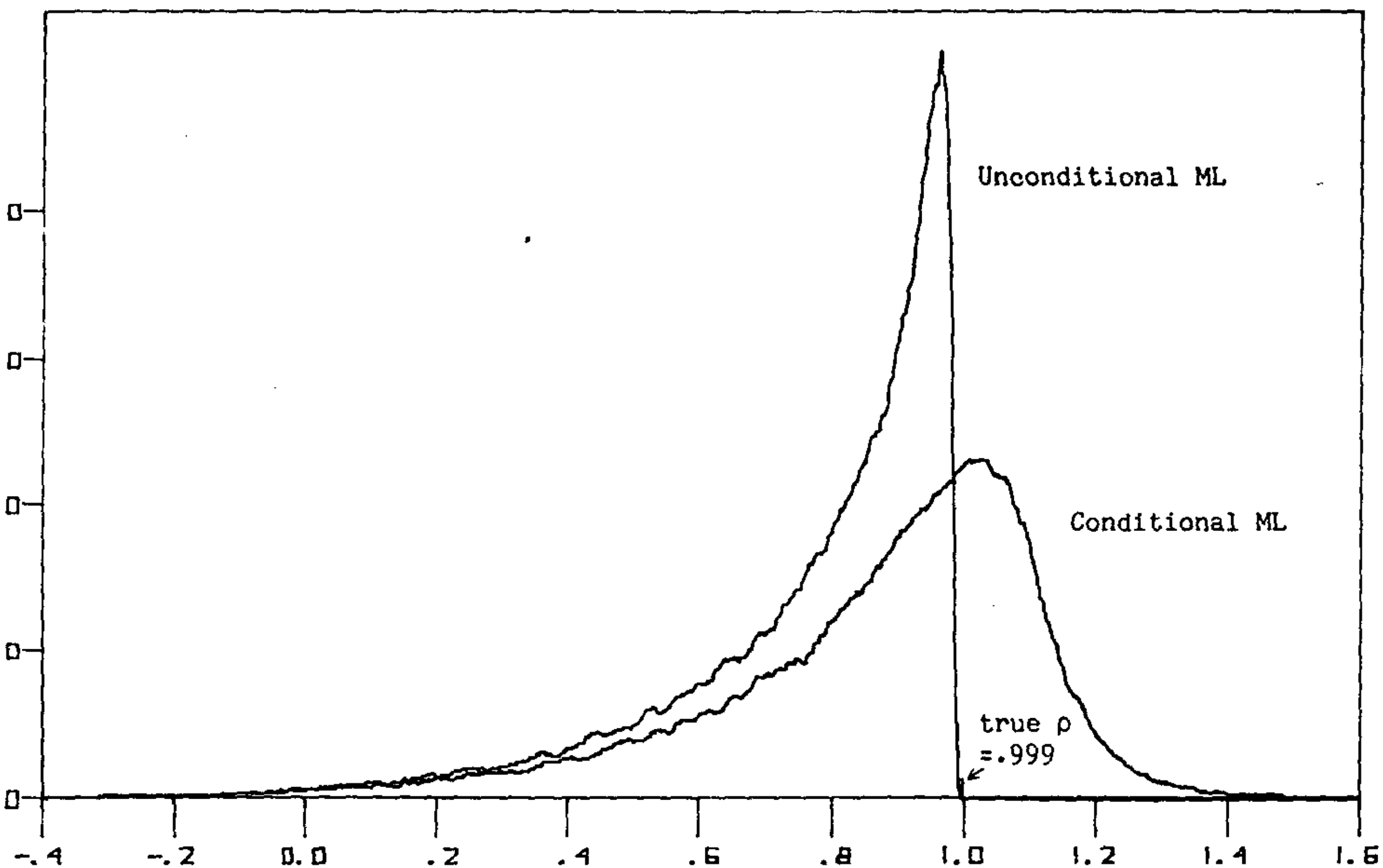
When the initial observations are drawn from the stationary distribution of the process, as in experiments 4 and 5, the use of conditional maximum likelihood is not as efficient as unconditional estimation. Nevertheless, in these univariate, first-order models, conditional maximum likelihood does not appear to suffer appreciably from the problem of increased bias relative to the unconditional estimation.

These univariate results demonstrate some of the reasons for caution when considering the use of unconditional maximum likelihood, but they cannot illustrate all of the issues raised when the two methods are compared in a higher-order or multivariate context. In these experiments the only way the initial observation can be atypical is in terms of its absolute size. More generally, for example, a particular ratio of observations may be unlikely.

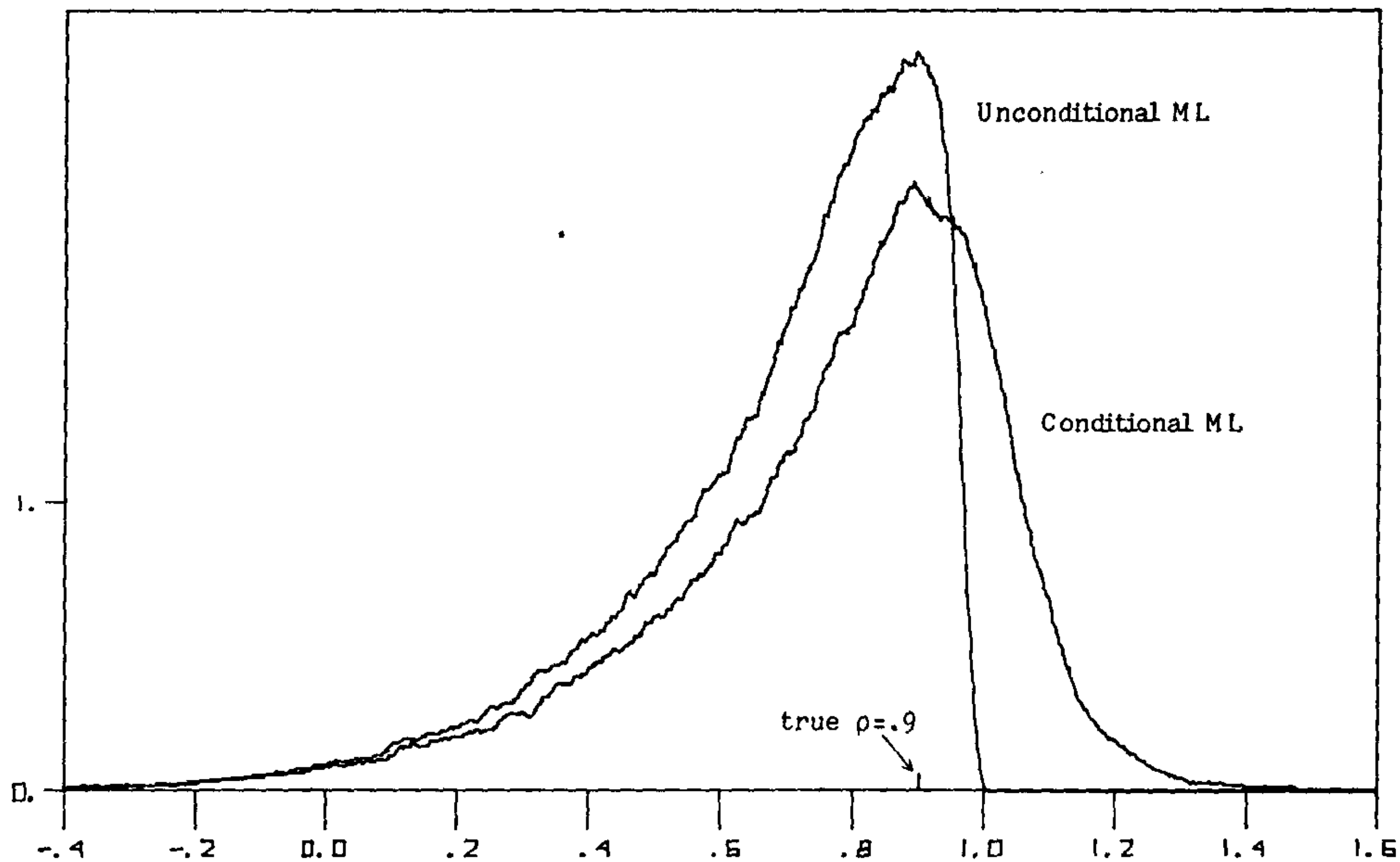
In practice, economists do not seem to have strong beliefs that the usual initial observations in economic time series, typically in the late 1940s and early 1950s, are necessarily drawn from a stationary distribution corresponding to the process which has generated data since that time. The fact that economic time series applications often ignore earlier data when it is available seems to indicate a suspicion that there may have been a significant structural change in the economy during the preceeding years. This suspicion justifies conditioning on the initial observations, since it implies a large uncertainty concerning their distribution and, in particular, whether they were generated by the process which led to the subsequent observations.

One might wish to specify a prior distribution for the initial observations representing one's uncertainty more precisely than either the conditional or unconditional maximum likelihood approaches allow. In practice, a Bayesian approach of this kind will require nonlinear procedures which would greatly increase the expense of estimation. Because of these considerations, the conditional likelihood function has been used throughout this investigation.

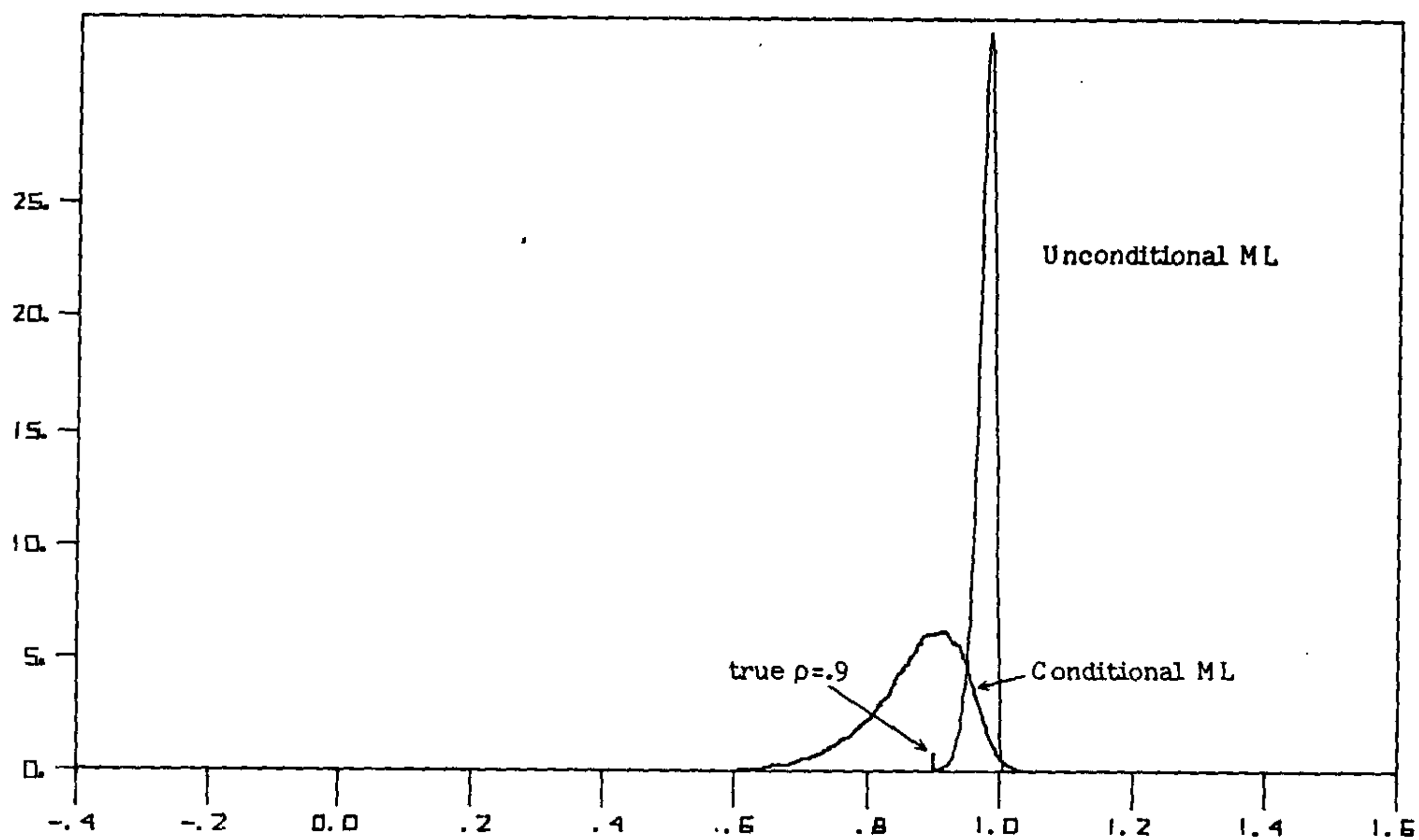
Experiment 1



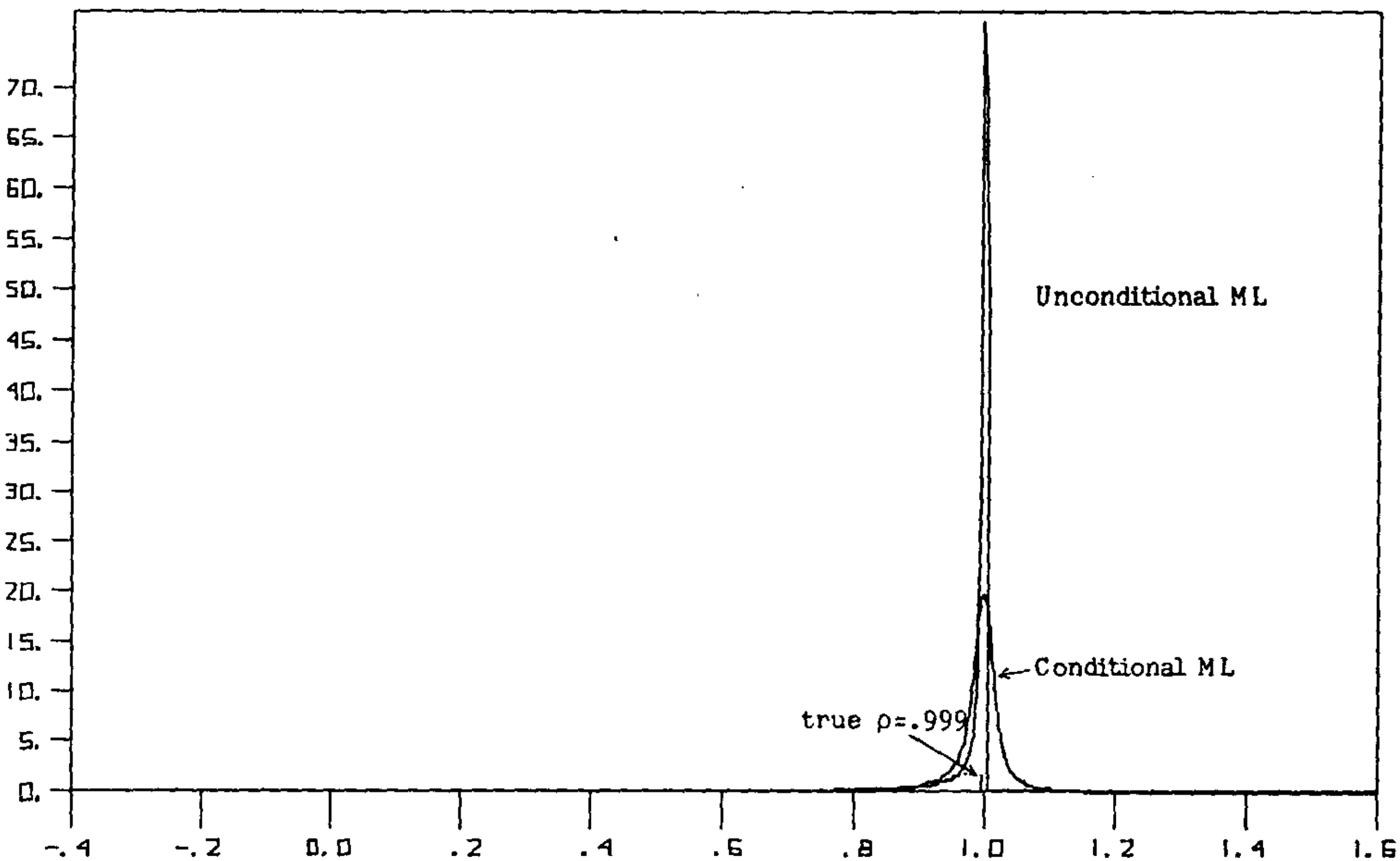
Experiment 2



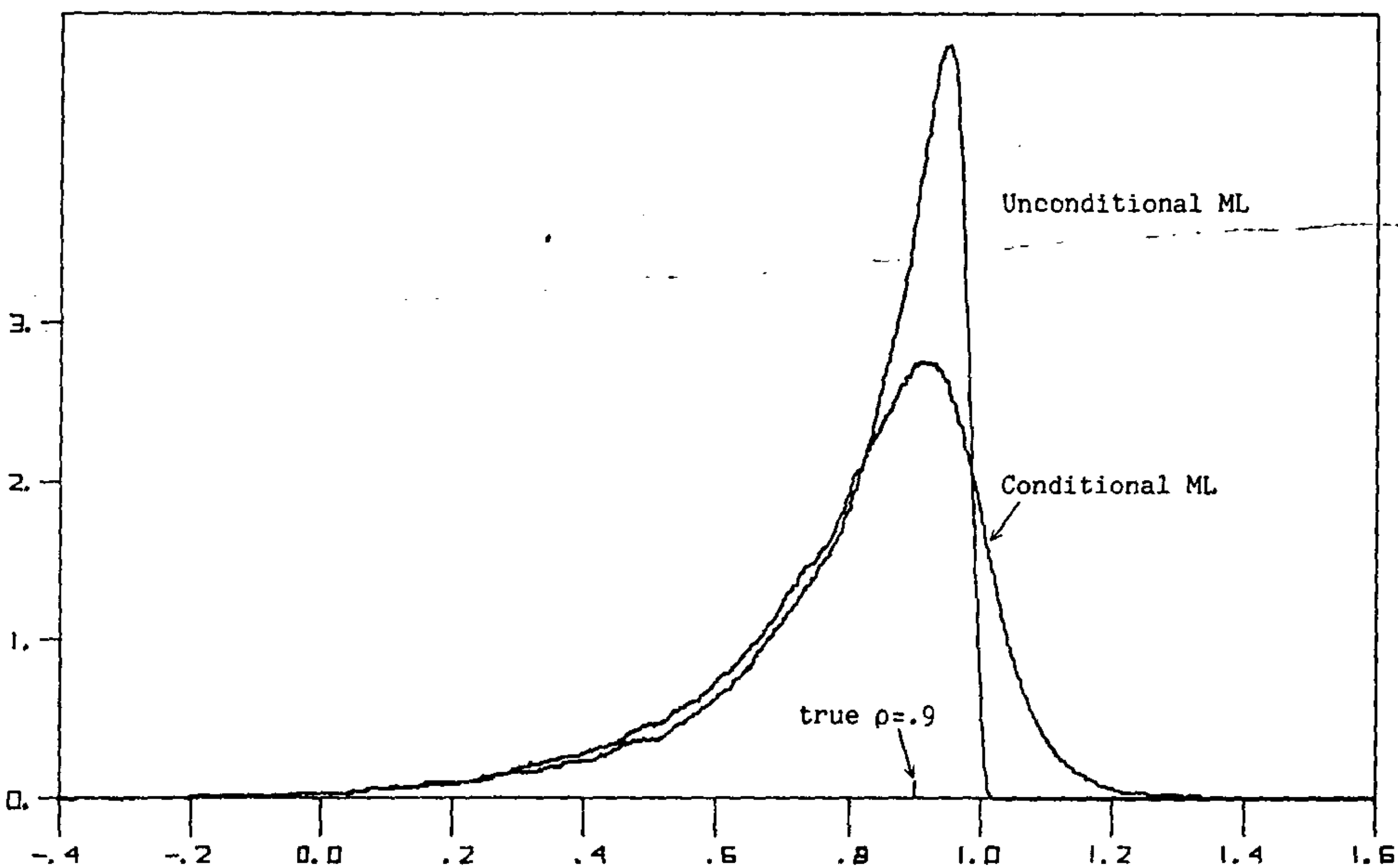
Experiment 3



Experiment 4



Experiment 5



CHAPTER II
RESTRICTIONS IN THE FORM OF PRIORS

Biased Estimation

Autoregressive specifications often lead to multicollinearity problems and large sampling errors in estimation. This is particularly true in typical vector specifications which have relatively few degrees of freedom.

Several procedures (Chamberlain and Leamer [1976]) have been devised to overcome this type of problem. Included in this category are, for example, ridge regression (Hoerl and Kennard [1970]) and Stein rule estimators (Stein [1974]). These procedures have been justified in the literature on the grounds that they can generate biased estimators which have smaller mean square errors than OLS estimates. The use of ridge-type estimators in the univariate autoregressive context has been suggested by Swamy and Rappaport [1975] and is discussed in the context of distributed lags by Maddala [1977].

Each of these methods may be given a Bayesian interpretation, which amounts to specifying the prior distribution for which the biased estimator is the posterior mean. In essence these procedures imply a combination of data evidence and information supplied by the investigator. The Bayesian interpretation is useful in that it makes explicit the exact content of the information being added to the data.

In the normal linear model

$$Y = X\beta + e \qquad e \sim N(0,\sigma^2), \tag{26}$$

the standard ridge estimators, for example, are given by

$$b^k = (X'X+kI)^{-1}X'Y \text{ with } k > 0. \tag{27}$$

These estimators are posterior means corresponding to prior distributions given by $\beta \sim N(0,\lambda^2 I)$ with $k = \sigma^2/\lambda^2$. Each coefficient in the prior is distributed independently and normally with mean zero and variance inversely proportional to k.[10/] Here, and throughout this section, σ^2 and λ^2 are treated as known.

Ridge estimation developed as a procedure for overcoming large sampling errors associated with multicollinear data. An investigator using ridge regression attempts to accomplish this, at least implicitly, by adding to the data the a priori information that the larger the coefficients are in absolute value, the more unreasonable they are. The ridge estimator is sometimes referred to as a "shrinkage" estimator because it shrinks coefficient estimates toward zero. Following Leamer [1978] we can illustrate the effect of ridge regression in two dimensions in terms of a graph of the "contract curve" which balances data evidence against a priori information. Relative to any point not on this curve, there is a point on the curve which is more compatible with both the data and the prior distribution. The ridge class of estimators traces out this curve as k varies, beginning at the OLS estimate for k = 0 and moving to 0 as k goes to infinity.

The Stein class of estimators is given by

$$b^k = (X'X+kX'X)^{-1}X'Y \text{ with } k > 0. \tag{28}$$

The implicit prior is of the form $\beta \sim N(0,\lambda^2(X'X)^{-1})$. Stein estimators are also shrinkage estimators and differ from ridge estimators only in the metric through which the shrinkage is applied.

Following Maddala [1977] we can define a generalized ridge estimator

$$\beta^k_{GR} = (X'X+k\Delta^{-1})^{-1}(X'Y+k\Delta^{-1}\theta) \tag{29}$$

which corresponds to a prior distribution on β of $N(\theta,\lambda^2\Delta)$ with $k = \sigma^2/\lambda^2$. The variance of this estimator is given by $\sigma^2(X'X+k\Delta^{-1})^{-1}$.

One example of a generalized ridge estimator for distributed lags is given by Leamer [1972]. He considers a geometrically decaying response pattern in the model

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \cdots \tag{30}$$

with mean vector defined by

$$E[\alpha] = \alpha_0 \qquad E[\beta_i] = mr^i \qquad 0 < r < 1. \tag{31}$$

The variance matrix is generated by the principle of proportionality

$$\mathrm{Cov}[\beta_i \beta_j] = \lambda^2 w^{|i-j|} r^{i+j-2} \qquad 0 \le w \le 1 \tag{32}$$

and $\mathrm{Var}[\alpha] = \lambda^2 a$.

Thus, for a fourth-order lag

$$\theta = \begin{bmatrix} \alpha \\ m \\ mr \\ mr^2 \\ mr^3 \end{bmatrix} \qquad \Delta = \begin{bmatrix} a & 0 & 0 & 0 & 0 \\ 0 & 1 & wr & w^2r^2 & w^3r^3 \\ 0 & wr & r^2 & wr^3 & w^2r^4 \\ 0 & w^2r^2 & wr^3 & r^4 & wr^5 \\ 0 & w^3r^3 & w^2r^4 & wr^5 & r^6 \end{bmatrix}. \tag{33}$$

This prior implies, among other things, that the coefficients on more distant lags have relatively tight marginal distributions around zero. The prior distributions we will develop later in this section for the vector autoregressive model will also have this property.

Shiller's [1973] smoothness prior for distributed lags is another example of a generalized ridge estimator. The smoothness prior for the m^{th}-order lag (without constant) is given by

$$R_d\beta = w \text{ with } w \sim N(0,\sigma_w^2 I) \tag{34}$$

where the (m-d-1xm) matrix R_d is a matrix of d+1 differences; i.e., for d = 2

$$R_d = \begin{bmatrix} 1 & -3 & +3 & -1 & 0 & & \cdots & & 0 \\ 0 & 1 & -3 & +3 & -1 & 0 & & \cdots & 0 \\ 0 & 0 & 0 & \cdots & & 1 & -3 & +3 & -1 \end{bmatrix}. \tag{35}$$

Because R'R is not of full rank, Shiller's prior distribution on w does not translate into a proper prior distribution for β. In order to represent prior information in this situation we write the generalized ridge estimator as

$$\hat{\beta}^k_{GR} = (X'X+k(R'R))^{-1}(X'Y+kR'r) \tag{36}$$

corresponding to implicit prior information given by

$$R\beta = r + v \qquad v \sim N(0,\lambda^2 I) \tag{37}$$

with $k = \sigma^2/\lambda^2$. The prior distribution for β in this case is improper. It is justified as an approximation to a proper prior which combines the information in (37) with a proper, but locally uniform, prior distribution for linear combinations of β orthogonal to the space spanned by $R\beta$. That is, (37) is obtained as the limit, as $\tau \to \infty$, of the proper prior distribution for β determined by specifying a (d+1xm) matrix S, such that

$$Q\beta = \begin{bmatrix} r \\ 0 \end{bmatrix} + \begin{bmatrix} v \\ w \end{bmatrix} \qquad \begin{matrix} v \sim N(0,\lambda^2 I) \\ w \sim N(0,\tau^2 I) \end{matrix} \tag{37a}$$

where $Q = \begin{bmatrix} R \\ S \end{bmatrix}$ $RS' = 0$ and rank $[Q] = m$. When R'R is invertible, then (36) is identical to (29) with

$$\theta = (R'R)^{-1}R'r \text{ and } \Delta = (R'R)^{-1}. \tag{38}$$

As Sims [1974] suggests, "the whole notion that lag distributions in econometrics ought to be smooth is . . . at best weakly supported by theory or evidence." Rather than imposing smoothness, the information about coefficient values which we will combine with the data is derived from the assumption that a reasonable approximation of the behavior of an economic variable is a random walk around an unknown, deterministic component. Thus, the prior distribution for the parameters of the i^{th} equation in the vector autoregression is centered around the specification

$$Y_i(t) = Y_i(t-1) + d_i(t) + \varepsilon_i(t). \tag{39}$$

The parameters are all assumed to have means of zero except the coefficient on the first lag of the dependent variable, which is given a prior mean of one. The parameters are assumed to be uncorrelated with each other and to have standard deviations which decrease the further back they are in the lag distributions. In general, the prior distribution on lag coefficients of the dependent variable is much looser, that is, has larger standard deviations, than it is on other variables in the system.[11/]

The specification of knowledge about the deterministic component in an autoregressive model is not generally independent of the specification for other parameters. For example, in the univariate, first-order model

$$Y(t) = \alpha + \beta Y(t-1) + \varepsilon(t) \tag{40}$$

if we know that Y is a stationary process with mean M_y, then taking expectations and solving, we have $M_y = \alpha/(1-\beta)$. This nonlinear relationship implies that independent normal prior distributions on M_y and β cannot be transformed into a multivariate normal prior distribution on α and β. Of course, even if it were known that $M_y = \alpha = 0$, it would not be possible to represent the stationarity assumption in a normal prior distribution for β.[12/]

On the other hand, when we do not assume stationarity, M_y need not exist, and without observing the data very little will generally be known about the distribution of the parameters of the deterministic component. In order to represent this ignorance a noninformative prior, that is, one which gives equal weight to all possible parameter values, is used. This flat prior is not a proper probability distribution, but it is the limit of proper distributions with increasing standard deviations and, thus, generates results which are arbitrarily close to those which would be generated with a given set of data by a proper, but suitably diffuse, prior. For example, with a noninformative prior distribution on α in (40) and a normal prior distribution with mean of 1 and variance λ^2 for β, the improper prior distribution for α and β may be written

$$p(\alpha,\beta|\lambda^2) \propto \exp\left(-\frac{(\beta-1)^2}{2\lambda^2}\right). \tag{41}$$

This prior distribution is the limit if we consider the class of proper prior distributions given by

$$\begin{bmatrix}\alpha\\ \beta\end{bmatrix} \sim N_\tau(\theta,\lambda^2\Delta) \text{ with } \theta = \begin{bmatrix}0\\ 1\end{bmatrix},\ \Delta = \begin{bmatrix}\tau^2 & 0\\ 0 & 1\end{bmatrix}, \tag{42}$$

and let τ go to infinity.

More generally, in the model

$$\underset{T\times 1}{Y} = \underset{T\times K_1}{Z}\ \underset{K_1\times 1}{\gamma} + \underset{T\times K_2}{X}\ \underset{K_2\times 1}{\beta} + \underset{T\times 1}{u} \qquad u \sim N(0_T,\sigma^2 I_T) \tag{43}$$

where we may view $Z\gamma$ as the deterministic component, it can be shown that the least squares estimate of β is given by

$$\hat{\beta} = (X'M_2X)^{-1}X'M_2Y \tag{44}$$

with $M_2 = (I_T - Z(Z'Z)^{-1}Z')$. Since M_2 is idempotent, this implies that the least squares estimate of β in (43) may be found by first subtracting out the deterministic components of X and Y, that is, by forming

$$\tilde{X} = M_2X = X - Z(Z'Z)^{-1}Z'X \tag{45}$$

and

$$\tilde{Y} = M_2Y = Y - Z(Z'Z)^{-1}Z'Y \tag{46}$$

the residuals of regressions of X and Y on Z, and then regressing $\tilde{Y}$ on $\tilde{X}$.

If we consider a class of priors on γ and β which are of the form

$$p(\gamma,\beta) = p_v(\gamma)\cdot p_1(\beta) \tag{47}$$

where $p_v(\gamma)$ is multivariate normal with mean zero and variance $v^2I_{K_1}$, we can, following Leamer [1978], derive a similar result. The marginal posterior density function of β is given by

$$\begin{aligned} f(\beta|Y) &\propto \int_\gamma L(Y|\beta,\gamma)P(\gamma,\beta)d\gamma \\ &= \int_\gamma L(Y|\beta,\gamma)p_v(\gamma)d\gamma p_1(\beta) \\ &= L(Y|\beta)p_1(\beta) \end{aligned} \tag{48}$$

where we use $L(Y|\beta)$ to represent the marginal likelihood of Y generated by the process

$$Y = X\beta + \varepsilon \qquad \varepsilon = u + Z\gamma \tag{49}$$

so that given the prior, p_v, on γ, ε is normal with mean zero and covariance matrix $(\sigma^2I_T+v^2ZZ')$.

Thus,

$$L(Y|\beta) \propto |(\sigma^2I_T+v^2ZZ')|^{-1} \exp\{-\frac{1}{2\sigma^2}(Y-X\beta)(\sigma^2I_T-v^2ZZ')^{-1}(Y-X\beta)\}. \tag{50}$$

For large v $|(\sigma^2 I_T + v^2 ZZ')|$ behaves as $(\sigma^2)^{(T-K_1)}|Z'Z|v^{K_1}$ and

$$(\sigma^2 I_T + v^2 ZZ')^{-1} = \sigma^{-2} I_T - \sigma^{-2} Z[Z'Z + \frac{\sigma^2}{v^2} I_K]Z' \tag{51}$$

approaches the singular matrix $\sigma^{-2}M_2$. Thus, the limiting likelihood function is

$$L(Y|\beta) \propto |Z'Z|^{-1}\sigma^{-(T-K_1)} \exp\{-\frac{1}{2\sigma^2}(Y-X\beta)M_2(Y-X\beta)\}. \tag{52}$$

With a prior on γ of the form given in (47), as v gets large the posterior density of β approaches that obtained by first subtracting out the deterministic components of Y and X.

The prior distribution which has been described here is not derived from a particular economic theory, and in this sense, the information it represents may be referred to as instrumental. The intuition behind its use is two-fold. Consider a univariate autoregression. The univariate model requires specification of a lag length, m. In choosing a larger m, the trade-off is between the possibility of increasing the explanatory power of the data through inclusion of further lags and the possibility of decreasing the accuracy of parameter estimates because of the larger estimation problem. However, this approach, of specifying a model by the choice of a finite lag length, m, has the unfortunate property that it embodies the a priori assumption that nothing is known about the values of the first m coefficients in the lag distribution, whereas it is known with certainty that all coefficients on lags greater than m are zero. By specifying a prior distribution with decreasing standard deviations on coefficients of larger lags, we are, in general, able to estimate more lags and at the same time are making a more consistent assumption about the distribution of relevant information in the past values of the variable. A similar type of prior information is used in Leamer's [1972] work on distributed lags.

The second point is that when one attempts to model the effects of other variables in an equation, the standard procedure is to make extensive use of exclusionary restrictions. Coefficients are added only on variables which a particular economic theory suggests should have an effect. It is not so much the particular theories which are used in this way that are objectionable, as the manner in which they are used. There is rarely an attempt to justify the absence of variables on the basis of an economic theory, despite the fact that a zero restriction implies the existence of very certain prior information. On the other hand, theory is often cited to justify the inclusion of particular variables, but then in the estimation process the investigator acts as if it is the coefficients on just those variables about which he is completely ignorant. When information, whether derived from a particular theory or otherwise, is added in terms of zero restrictions, the choice is extreme, to include or to exclude; there is no middle ground. For this reason, the use of exclusionary restrictions does not allow the realistic specification of a priori knowledge. Our specification, in general, includes coefficients on all variables in the system at several lags, but also includes a prior which suggests, with varying degrees of uncertainty, that those coefficients are close to zero. We are thus able to allow for a fuller range of interaction among variables and at the same time are able to specify with greater flexibility how likely we believe it to be that the interaction does exist. In sum, the justification for this prior is simply that through its use we are able to express more realistically our state of knowledge and uncertainty about the structure of the economy, and that in doing so we are more likely to find the regularities in the data which will lead to better forecasts.

Probably the two most objectionable aspects of this prior are its reflection of complete ignorance about the deterministic components and its prior mean, which reflects a nonstationary process. Both of these specifications are likely candidates for modification in particular applications. On the other hand, these are the areas in which the prior distribution is most uncertain anyway, and, thus, they are the areas in which the data will most likely dominate. For this reason, forecasting performance will be insensitive to specification of other reasonably loose priors on the deterministic components and means of other than the one suggested here. It certainly may be true that for the unemployment rate, or interest rates, for example, a random-walk prior is not appropriate, and one might do better by specifying a mean of less than one on the first own lag.

When there are known relationships among variables, whether derived from economic theory or other considerations, that information should be imposed in the estimation process. As mentioned in the introduction, there are many economists who feel that the theory which is typically used to identify the equations of econometric models is not valid. With most economic systems, and particularly when rational expectations mechanisms and dynamic optimization problems are involved, interactions among variables will be complex and the form of lag distributions often will be unknown. A primary purpose of this investigation is to determine if, despite the absence of strong a priori beliefs on restrictions derived from economic theory, one could improve estimates and, in particular, forecasting performance, by including instrumental information of the type described above.

Bayesian Ridge Estimators

In the preceeding section we introduced a general class of ridge estimators corresponding to Bayesian priors. We also described the intuition behind a particular prior distribution to be used here with vector autoregressions. In this section we fill in some details of the exact specification of the prior distribution and develop the formal Bayesian justification for the estimators used in this study.

To begin with we suggest a parameterization of the prior distribution which treats all equations symmetrically. First we specify λ, a constant standard deviation on the first lag of the dependent variable in each equation. Then the standard deviations of further coefficients in the lag distributions can be decreased in a harmonic manner, according to a parameter γ_1. Standard deviations on other variables in the system can be made tighter than own lags according to a parameter γ_2.

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

This scale problem is usually solved in the standard ridge regression context by, in effect, scaling the implicit prior by the standard deviations of the independent variables. This is accomplished by using transformed data so that X'X is a correlation matrix. We are led away from the standard approach because of a difference in our prior information and that implicit in ridge regression. We suspect that the size of the response of one economic variable to another is more often a function of the relative sizes of the unexpected movements of the two variables than of their overall standard deviations.

In results reported here the measure of the unexpected movements is taken to be the residuals in unrestricted OLS vector autoregressions. For large systems with few (or no) degrees of freedom, residuals from univariate regressions could be used.

A specific example may illustrate the type of problem that can arise when scaling according to standard deviations of the variables in a VAR context. Consider again a system which includes GNP and an interest rate. Suppose that in the data both real GNP and the inflation rate fluctuate around positive means. Now consider alternative regression equations for nominal versus real GNP. We expect, a priori, the coefficients on interest rates in both these equations to be approximately the same size.

Nominal GNP, however, will exhibit a positive trend which over any significant interval will cause the ratio of standard errors of nominal GNP to the interest rate to be much greater than the ratio of standard errors of their residuals in these regressions. Real GNP has no trend, so the ratio of its standard error to that of the interest rate will be much closer to the ratio of their residuals, which will also be about the same as the ratio of residuals in the system with nominal GNP.

Scaling the standard deviations of our prior distribution by the relative sizes of residuals in OLS regressions would work for either equation. Scaling by the relative standard deviations of the variables themselves, however, would lead to a lower prior on interest rates in the nominal GNP equation which would misrepresent our prior opinions.

To summarize, our specification of a symmetric prior depends on three parameters, λ, γ_1, and γ_2, and the following rule. Let δ_{ij}^{ℓ} be the standard deviation of the coefficient on lag ℓ of variable j in equation i. Also, $\hat{\sigma}_i$ is the standard error of the residuals in the i^{th} equation in the unrestricted OLS estimation of the system. Then

$$\delta_{ij}^{\ell} = \begin{cases} \dfrac{\lambda}{\ell^{\gamma_1}} & \text{if } i = j \\ \dfrac{\lambda\gamma_2\hat{\sigma}_i}{\ell^{\gamma_1}\hat{\sigma}_j} & \text{if } i \neq j \end{cases} \quad . \tag{53}$$

Thus, to put the prior in the form

$$R\beta = r + v \qquad v \sim N(0,\lambda^2 I) \tag{54}$$

we make R_i a diagonal matrix with zeros corresponding to deterministic components and elements $[\lambda/\delta_{ij}^{\ell}]$ corresponding to the ℓ^{th} lag of variable j. r_i is a vector of zeros and a one corresponding to the first lag of the dependent variable.

Notice that we have been treating the prior distribution for each equation separately. We will indeed estimate each equation separately. This treatment is justified when there is no prior because the explanatory variables are the same in each equation. When we add prior information which is different for different equations in the system, there could be a gain in efficiency by estimating all equations together via a seemingly unrelated regression procedure which uses the information contained in cross-equation covariances.[13/] We do not attempt such a procedure primarily because of the computational burden; it requires inversion of an n^2m+nd-order matrix, where n is the number of variables, m the number of lags, and d the number of parameters in the deterministic component of each equation. The departures from full efficiency will depend on how far from diagonal is the covariance matrix of residuals from different equations and the relative strength of asymmetric prior information to data evidence.

In the previous section it was claimed that the estimator in (36) corresponded to the prior distribution given in (37) when σ^2 and λ^2 are known.

To see this, we begin with the prior (37), which can be written

$$P(\beta_i) \propto \lambda^{-p} \exp\{-\frac{1}{2\lambda^2}(R_i\beta_i - r_i)'(R_i\beta_i - r_i)\} \tag{56}$$

where $p = \text{rank}\,(R_i'R_i)$.

In this section we treat each equation separately and let $\underset{T\times 1}{Y_i'} = (Y_i(1),\ldots,Y_i(T))$ and similarly let $\underset{T\times k}{X}\ \underset{k\times 1}{\beta_i} + \underset{T\times 1}{\varepsilon_i}$ represent the stacked right-hand side of (10), where $k = nm + d$ is the number of explanatory variables.

Now, the likelihood function can be written

$$L(Y_i|\beta_i\sigma) \propto \sigma^{-n} \exp\{-\frac{1}{2\sigma^2}(Y_i - X\beta_i)'(Y_i - X\beta_i)\}. \tag{57}$$

Combining (56) and (57) according to Bayes law, and dropping the i subscript, we obtain the posterior density function:

$$P(\beta|Y,\sigma) \propto \lambda^{-p}\sigma^{-n} \exp\{-\frac{1}{2\sigma^2}M\} \tag{58}$$

where

$$M = [(\bar{R}\beta-\bar{r})'(\bar{R}\beta-\bar{r})+(Y-X\beta)'(Y-X\beta)] \tag{59}$$

and

$$\bar{R} = \frac{\sigma}{\lambda}R \qquad \bar{r} = \frac{\sigma}{\lambda}r. \tag{60}$$

Let

$$\tilde{X}' = (X'\bar{R}') \qquad \tilde{Y}' = (Y'\bar{r}') \tag{61}$$

$$\hat{\beta}_m = (X'X+\bar{R}'\bar{R})^{-1}(X'Y+\bar{R}'\bar{r}) = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{Y} \tag{62}$$

and

$$\tilde{U}_m = (\tilde{Y}-\tilde{X}\hat{\beta}_m). \tag{63}$$

It can easily be shown that

$$M = [\hat{U}_m'\hat{U}_m+(\beta-\hat{\beta}_m)'(\tilde{X}'\tilde{X})(\beta-\hat{\beta}_m)]. \tag{64}$$

It is apparent that the posterior density function for β is a k-dimensional multivariate normal with mean $\hat{\beta}_m$ and covariance matrix $\sigma^2(\tilde{X}'\tilde{X})^{-1}$. $\hat{\beta}_m$ is equal to $\hat{\beta}^k_{GR}$ given in (36) with $k = \sigma^2/\lambda^2$.

Of course, we will rarely know σ^2, and we begin again with (56), add a diffuse prior on σ, $P(\sigma) \propto \frac{1}{\sigma}$, as suggested in Zellner [1971], and taking λ as given we have the prior density function

$$P(\beta,\sigma) \propto \sigma^{-1}\lambda^{-p} \exp\{-\frac{1}{2\lambda^2}(R\beta-r)'(R\beta-r)\}. \tag{65}$$

Proceeding as above we derive the joint posterior density

$$P(\beta,\sigma|Y) \propto \frac{1}{\sigma^{n+1}} \exp\{-\frac{1}{2\sigma^2}[\hat{U}_m'\hat{U}_m+(\beta-\hat{\beta}_m)\tilde{X}'\tilde{X}(\beta-\hat{\beta}_m)]\} \tag{66}$$

$$= \frac{\exp\{-\frac{1}{2\sigma^2}\hat{U}_m'\hat{U}_m\}}{\sigma^{n-p+1}|\tilde{X}'\tilde{X}|^{1/2}} \frac{|\tilde{X}'\tilde{X}|^{1/2}}{\sigma^p} \exp\{-\frac{1}{2\sigma^2}(\beta-\hat{\beta}_m)\tilde{X}'\tilde{X}(\beta-\hat{\beta}_m)\} \tag{67}$$

$$= \frac{\exp\{-\frac{1}{2\sigma^2}\hat{U}_m'\hat{U}_m\}}{\sigma^{n-p+1}|\tilde{X}'\tilde{X}|^{1/2}} \cdot N(\hat{\beta}_m,\sigma^2(\tilde{X}'\tilde{X})^{-1}). \tag{68}$$

The conditional distribution for β given σ is normal. Thus, to find the mode of the posterior density function we maximize

$$P(\hat{\beta}_m(\sigma),\sigma) \propto \frac{1}{\sigma^{n+1}} \exp\{-\frac{1}{2\sigma^2}\hat{U}_m'\hat{U}_m'\} \tag{69}$$

which is derived by substituting $\hat{\beta}_m$ for β in (66) above.

For many purposes we might want to minimize a quadratic loss function in β. In this case we calculate the posterior mean of β.

$$E(\beta) = \int_\beta \beta P(\beta)d\beta \tag{70}$$

where $P(\beta)$ is the marginal posterior density of β found by integrating out σ. This function, a normal -t,[14/] is difficult to evaluate directly, and the above numerical integration is not feasible.

We note that by changing the order of integration we can show that this posterior mean is a weighted average of the $\hat{\beta}_m$'s treated as a function of σ. The weights are given by the marginal posterior density of σ

$$P(\sigma) \propto \frac{\exp\{-\frac{1}{2\sigma^2}\hat{U}_m'\hat{U}_m\}}{\sigma^{n-p-1}|\tilde{X}'\tilde{X}|^{1/2}}. \tag{71}$$

We have

$$\begin{aligned} E(\beta) &= \int_\beta \beta[\int_\sigma P(\beta,\sigma)d\sigma]\,d\beta \\ &= \int_\sigma\int_\beta \beta P(\beta|\sigma)P(\sigma)d\beta d\sigma \\ &= \int_\sigma P(\sigma)[\int_\beta \beta P(\beta|\sigma)d\beta]\,d\sigma \\ &= \int_\sigma P(\sigma)\hat{\beta}_m(\sigma)d\sigma. \end{aligned} \tag{72}$$

Computation of $\hat{\beta}_m(\sigma)$ and $P(\sigma)$ are too time consuming to allow evaluation of this integral by standard numerical integration techniques. I have attempted to approximate this expectation by taking a weighted average of $\hat{\beta}_m$'s with weights proportional to the corresponding $P(\sigma)$'s. These approximations are expensive, however, and generally are similar to the mean of the normal approximation of this posterior distribution suggested by Zellner [1971], Section 4.2. That normal approximation has mean $\hat{\beta}_m(\hat{\sigma})$ and variance matrix $\hat{\sigma}^2(\tilde{X}'\tilde{X})^{-1}$ where $\hat{\sigma}$, the estimated standard error of the OLS regression, is substituted for σ in the definition of $\tilde{X}$.

Unless otherwise noted, this estimator, $\hat{\beta}_m(\hat{\sigma})$, is the estimator which is used in this study. It is also the mixed estimator developed by Theil [1963] when the information in (37) is treated as a second sample, independent of the original data.

CHAPTER III

FORECASTING WITH VECTOR AUTOREGRESSIONS

The Bayesian justification of the estimators of the preceeding chapter, because it remains valid for any prior distribution, does not reflect on whether such estimators have value in forecasting economic time series. The latter question is an empirical one and is the subject matter of this chapter. The formula for prediction with vector autoregressions is given first. This is followed by a description of the statistics which are used to compare forecast performance. The third section provides a digression on the validity of generating performance comparisons on the basis of seasonally adjusted data. The fourth section proposes an improvement on a forecast evaluation procedure suggested by Fair [forthcoming]. Finally, a method is suggested for specifying priors in larger systems in which symmetric treatment of variables seems inappropriate. The forecasting performance of a system of this type is compared with the compiled performances of several forecasters along with univariate autoregressions and ARIMA models.

Projections

Projections using the coefficient estimates of the autoregressive representation are made according to Wold's "chain rule of forecasting."

Let $P_t Y(t+k)$ be the projection of $Y(t+k)$ made on the basis of information available at time t. For sake of exposition consider a simplified version of (1) in which $Y_t = \beta_0 D(t) + \sum_{\ell=1}^{L} \beta_\ell Y(t-\ell) + \varepsilon(t)$. Substituting estimates $\hat{\beta}_i$ for β_i and taking expectations at time t of Y(t+1) we have

$$P_t Y(t+1) = \hat{\beta}_0 D(t+1) + \sum_{\ell=1}^{L} \hat{\beta}_\ell Y(t+1-\ell) \tag{73}$$

which is well defined, since everything on the right-hand side is available at time t.

To find $P_t Y(t+k)$ we project both sides of

$$Y(t+k) = \beta_0 D(t+k) + \sum_{\ell=1}^{L} \beta_\ell Y(t+k-\ell) + \varepsilon(t+k) \tag{74}$$

on information available at time t and substitute $\hat{\beta}_i$ for β_i to get

$$P_t Y(t+k) = \hat{\beta}_0 D(t+k) + \sum_{\ell=1}^{k-1} \hat{\beta}_\ell P_t Y(t+k-\ell) + \sum_{\ell=k}^{L} \hat{\beta}_\ell Y(t+k-\ell). \tag{75}$$

This formula determines a recursive projection procedure, Wold's chain rule, which generates forecasts indefinitely far into the future.

Another procedure for generating a k-step forecast, $P_t Y(t+k)$, is to estimate the regression equation

$$Y(t+k) = A_0 D(t+k) + \sum_{\ell=1}^{L} A_\ell Y(t-\ell), \; t=1, \; T. \tag{76}$$

In this equation Y(t+k) is projected directly on information available at time t. One uses $P_t Y(t+k) = \hat{A}_0 D(t+k) + \sum_{\ell=1}^{L} \hat{A}_\ell Y(T-\ell)$ as the forecast.

If the true coefficients in both projection formulas were known, the forecasts would be the same. This second procedure, however, does not generate recursive projections and therefore is not useful for generating sample paths.

Performance Statistics

All statistical forecasting models of economic series involve a high degree of oversimplification and approximation. Thus, it is useful to have an empirical procedure to check how reasonable are the forecasts generated by a model and how they compare with those of other models.

A natural procedure to test a forecasting model is to generate a sequence of coefficient estimates and their forecasts and produce sample performance statistics such as the bias, mean absolute error, and mean square error of those forecasts.

The procedure for generating the forecast performance statistics reported here is the following. First, a set of data with observations t=1, ..., T is divided into an estimation period, t=1, ..., T_0 followed by a projection period, t=T_0+1, ..., T. Next a model is specified. The model may be a vector autoregression which requires specification of the variables, lag length, and possibly a prior distribution on the coefficients, or it may be any other well-defined forecasting procedure such as an ARIMA model with specified orders.[15/]

The coefficients of the model are estimated using the data in the estimation period. A sequence of forecasts of the next m values of Y is generated by using the chain rule. The coefficients are then reestimated using an additional observation, that is, adding Y(T_0+1) to the sample, and a new set of m forecasts is generated. This procedure of reestimating coefficients and forecasting is repeated for each observation in the projection period. The computations required for this technique are not intractable even with large vector autoregressive models, including those derived as approximations to posterior means, because updating the coefficient estimates each period does not require reapplication of OLS to larger and larger data sets. The equivalent sequence of linear least squares estimates may be generated by use of the Kalman filter, a recursive algorithm described later in this paper.

The forecast errors generated during the projection period are used to form performance statistics for the model. A list of definitions of some commonly used sample statistics is given below.

The statistics are defined in terms of the k-step-ahead forecast made at time t, $P_t(t+k)$; and the actual value at time t, A(t). Let T_k be the set of t's in the projection period for which the k-step-ahead forecast errors are known,[16/] L_k be the number of elements in T_k, and N be the frequency of the data per year, i.e., four for quarterly, twelve for monthly.

Mean Error k-Steps Ahead
$$\equiv [\sum_{t\epsilon T_k} P_t(t+k)-A(t+k)]/L_k \qquad (77)$$

Mean Percent Error k-Steps Ahead
$$\equiv [\sum_{t\epsilon T_k} \frac{[P_t(t+k)-A(t+k)]\cdot 100}{A(t+k)}]/L_k \qquad (78)$$

Mean Percent Growth Error (in annual terms) k-Steps Ahead
$$\equiv [\sum_{t\epsilon T_k} \frac{[P_t(t+k)-A(t+k)]\cdot 100\cdot N}{A(t)\cdot k}]/L_k \qquad (79)$$

Mean Absolute Error k-Steps Ahead
$$\equiv [\sum_{t\epsilon T_k} |P_t(t+k)-A(t+k)|]/L_k \qquad (80)$$

Mean Absolute Percent Error k-Steps Ahead
$$\equiv [\sum_{t\epsilon T_k} |\frac{[P_t(t+k)-A(t+k)]\cdot 100}{A(t+k)}]|/L_k \qquad (81)$$

Mean Absolute Percent Growth Error (in annual terms) k-Steps Ahead
$$\equiv [\sum_{t\epsilon T_k} |\frac{[P_t(t+k)-A(t+k)]\cdot 100\cdot N}{A(t)\cdot k}|]/L_k \qquad (82)$$

Root Mean Square Error k-Steps Ahead
$$\equiv [(\sum_{t\epsilon T_k} [P_t(t+k)-A(t+k)]^2)/L_k]^{1/2} \qquad (83)$$

Theil's U Statistic k-Steps Ahead
$$\equiv [\frac{\sum_{t\epsilon T_k} [P_t(t+k)-A(t+k)]^2}{\sum_{t\epsilon T_k} [A(t)-A(t+k)]^2}]^{1/2} \qquad (84)$$

For illustrative purposes an example of the performance statistics for an eight-lag vector autoregression model estimated using OLS is shown here. Quarterly observations of real GNP, M1, and the GNP price deflator were used beginning in 1954-1.[17/] The projection period is 1970-1 through 1978-1. The units of the variables are billions of 1972 dollars, billions of current dollars, and 1972 = 100, respectively.

	Steps Ahead	L_k	Real GNP	Money	Prices
Mean Error	1	32	.682862	-.282718	.066854
	2	31	3.558445	-.489344	.005614
	3	30	6.654699	-.754810	-.267618
	4	29	8.300401	-1.038095	-.431948
Mean Percent Error	1	32	.060892	-.125927	.057011
	2	31	.391168	-.252625	.055021
	3	30	.742653	-.415313	-.056337
	4	29	.919827	-.591788	-.093193
Mean PCT Growth Error	1	32	.203119	-.517532	.219291
	2	31	.735785	-.532468	.087560
	3	30	.919112	-.598251	-.121751
	4	29	.791110	-.655932	-.163796
Mean ABS Error	1	32	8.655970	1.154115	.668666
	2	31	14.156463	2.501070	1.417094
	3	30	24.076290	4.245739	2.145536
	4	29	37.092176	6.787852	3.275013
Mean ABS Percent Error	1	32	1.108620	.467390	.426903
	2	31	1.780782	1.020501	.898786
	3	30	3.012092	1.703680	1.369456
	4	29	4.625275	2.677226	2.067312
Mean ABS PCT Growth Error	1	32	4.439494	1.898055	1.736071
	2	31	3.583295	2.105232	1.859973
	3	30	4.055892	2.381728	1.922407
	4	29	4.687835	2.851695	2.212934
Root Mean Square Error	1	32	10.965852	1.469385	.828557
	2	31	17.737345	3.075096	1.657500
	3	30	27.868638	5.208654	2.523325
	4	29	43.352891	7.679900	3.727461
Theils U	1	32	.921579	.381233	.295304
	2	31	.824734	.403743	.295751
	3	30	.915069	.452651	.301145
	4	29	1.106923	.496577	.340096

One use of these performance statistics is to compare the forecasting ability of vector autoregressions with other popular models. Figure 2 shows such a comparison on the basis of Theil U statistics. The particular autoregressive specification shown was not the result of an extensive search for the best variables, lag lengths, prior parameters, and so on. It was, in fact, the first specification attempted for this comparison. A more complete comparison of McNee's results with those of a 15-variable VAR is given later in this chapter.

The forecast performance statistics suggested in this section, however, are most useful when they form the basis for comparing different models on the same data set. Comparison with real-time forecasters, such as is made in Figure 2, must be viewed with caution because real-time forecasts are necessarily based on a different information set than that which forms the basis of the data available in current economic time series.

The actual numbers in most economic time series change over time for two reasons which may tend to improve their predictability. The first reason is that revisions in the data are regularly made as more information becomes available. The second reason is that seasonal adjustment procedures use future as well as past values of not seasonally adjusted data, and therefore change as more information becomes available. This later problem is shown in the following section to be not very significant in one application.

Forecasters operating in real time, on the other hand, may gain a significant short-run advantage by using not only quarterly data, but also monthly and weekly data which may be available. Also, because they are made judgmentally, these forecasts take into account a much wider information set including short-range factors such as weather and strikes. It seems likely that these differences become less important for forecast horizons of several or more quarters.

Figure 2

Theil U statistics provide a standard for comparisons of forecasting performance. Here are shown the performance statistics compiled by McNees (1975) for six forecasters over the period 1970:2 – 1975:3. Also shown are the corresponding performance statistics for a vector autoregressive (VAR) system. Variables included were GNP, the implicit price deflator, business fixed investment, and a commercial paper rate. The system contained eight lags and a prior with parameters $\gamma_1 = .8$, $\gamma_2 = 1.0$, and $\lambda = .1$.

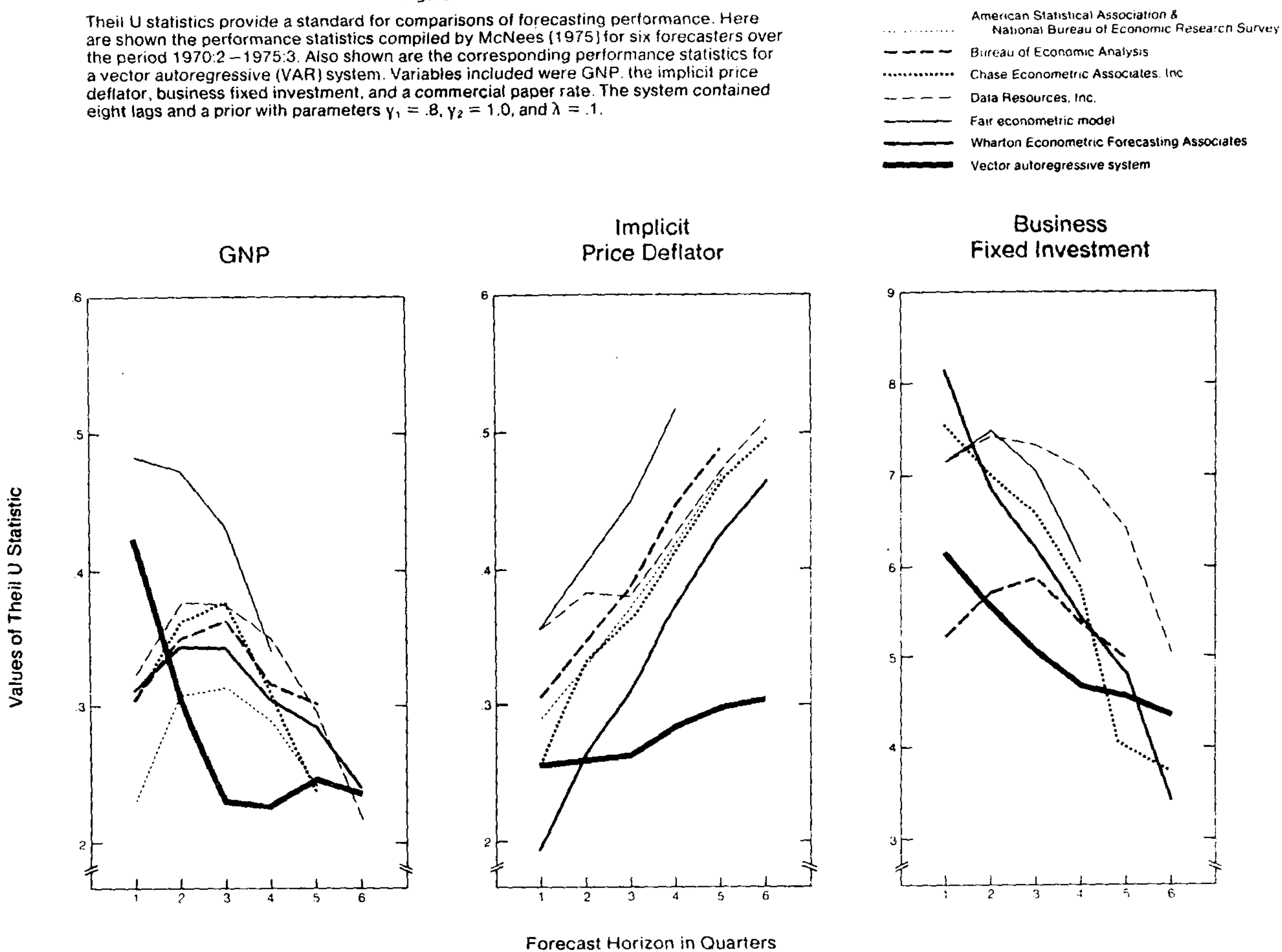


Figure 3

Theil U statistics for a three-variable eight-lag system illustrate the typical pattern of forecasting performance as tighter priors (smaller values of λ) are placed on the autoregressive specification. Forecasts generally improve (smaller Theil U statistics are better) and then worsen as prior tightness increases. Earlier projection periods, which are based on fewer observations, and longer forecast horizons generally require tighter priors. The system shown here has a prior with $\gamma_1 = .8$, $\gamma_2 = 1.0$, and λ as indicated.

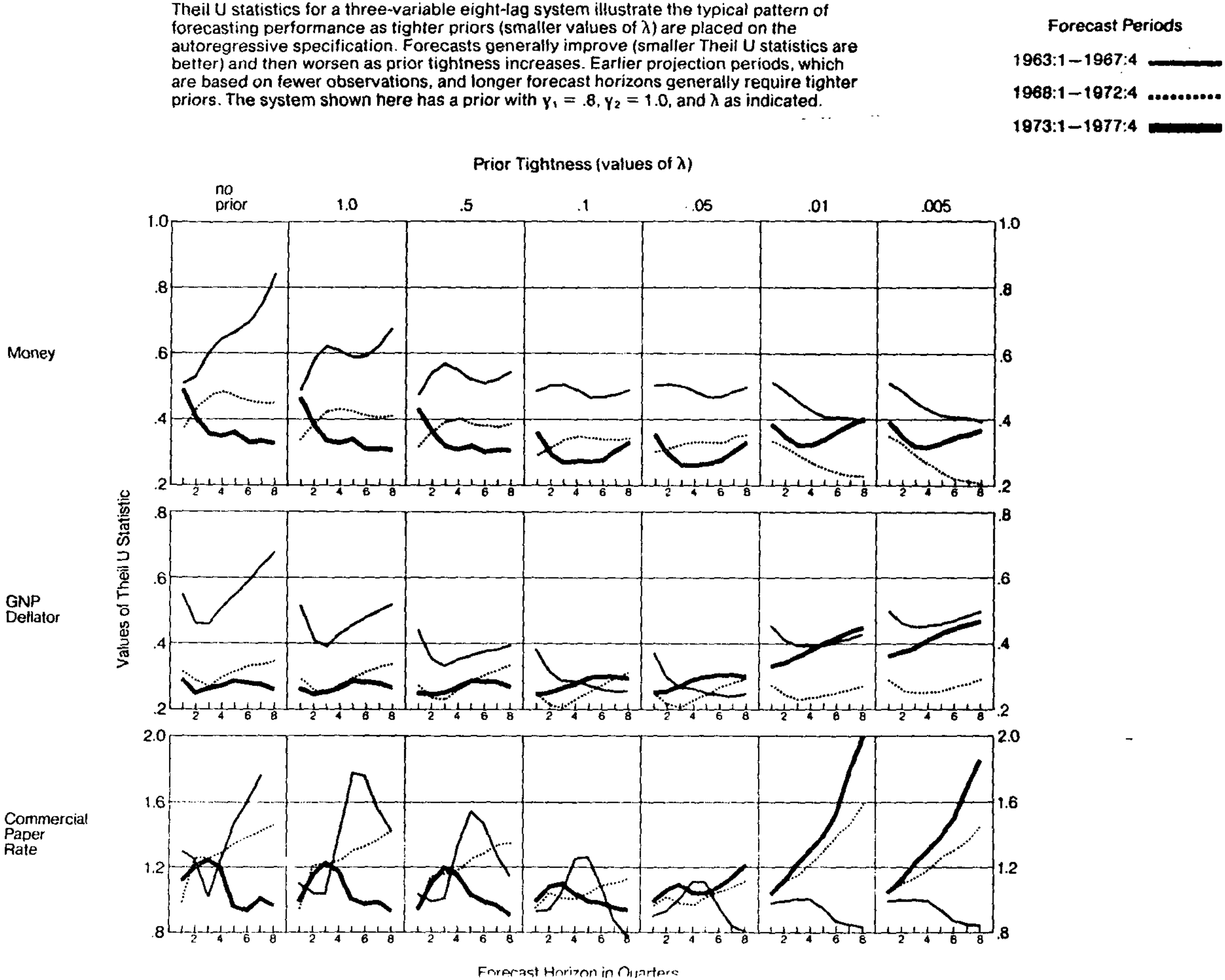
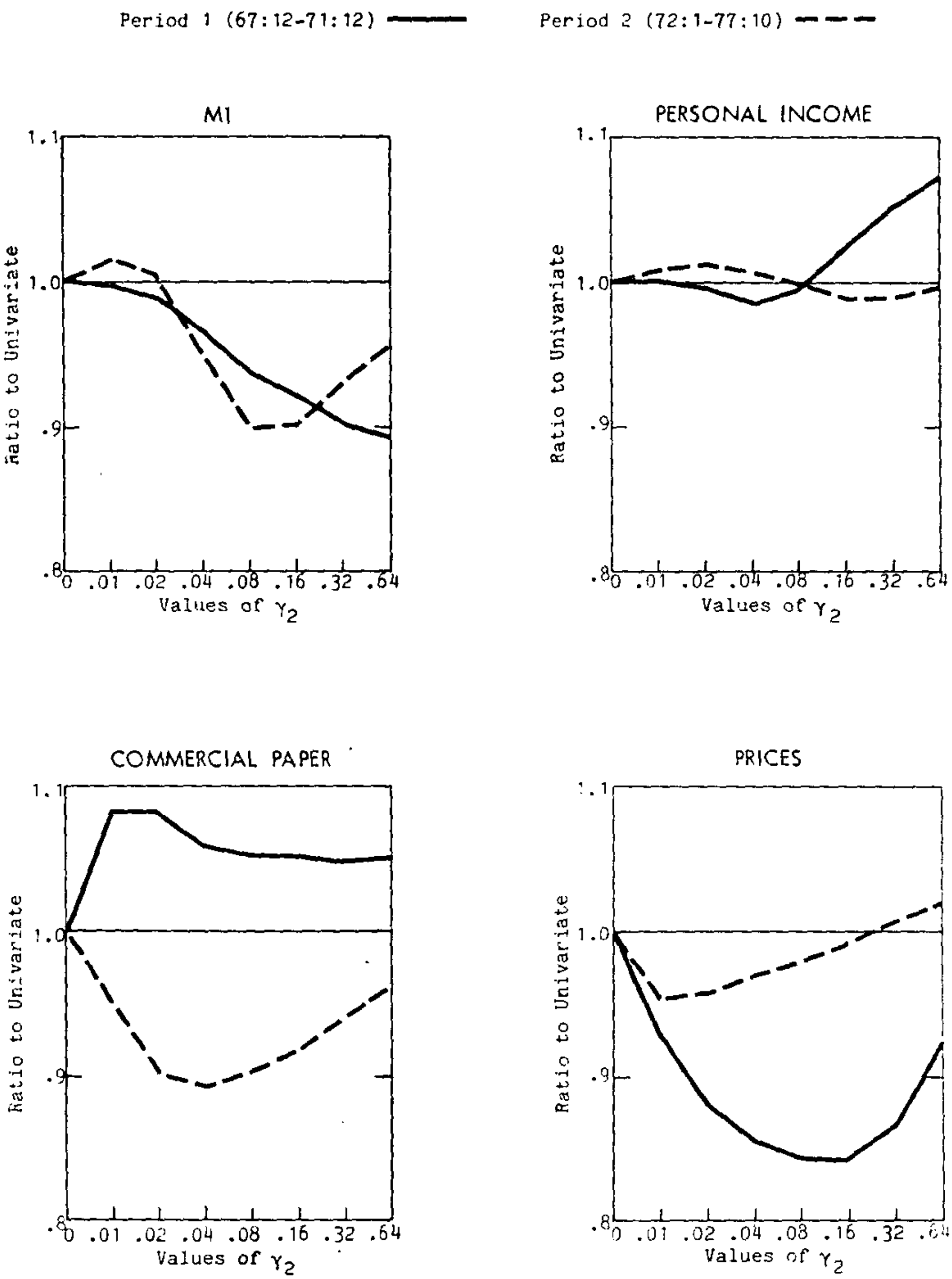


Figure 4

Changes in forecast performance as a function of movement from a univariate to multivariate specifications are shown on these graphs generated by a monthly model which has the variables M1, prices, personal income, and the prime rate on 4- to 6-month commercial paper. Each equation includes 6 lags of each variable and a constant. Shown are the ratio to univariate of mean square errors of 4-month horizon forecasts from specifications varying prior parameter γ_2. The values of the other prior parameters are λ = .2 and γ_1 = .5, reflecting relatively diffuse information about own-lag coefficients.



Forecast performance statistics are relatively insensitive to changes in the parameters of the prior over a wide range of values, as can be seen in Figure 3. Here a comparison of Theil U statistics is made over the entire range of λ, the standard error in the prior of the coefficient on the first lag of the dependent variable. This parameter, which represents the ratio of uncertainty in the prior to uncertainty in the data, controls the tightness of the prior. The graphs shown in Figure 3 demonstrate consistent improvement in forecasting performance with the imposition of the prior for the three variables in the system over three different projection periods. The largest gains are generated for the earliest projection period, 1963-1 through 1967-4, and for the forecasts with the longest horizon.

The result that unrestricted OLS systems can be readily improved upon by the imposition of a prior in multivariate autoregressive specifications is of interest, particularly in reasonably sized systems such as was used to generate Figure 3. In general, however, it is a rather weak result, since one could always add variables to a system until the unrestricted model would be so highly overparameterized that it would perform very poorly. A perhaps more interesting comparison is between the multivariate systems and a univariate autoregression. Such a comparison is made in Figure 4, in which the percentage change of root mean square errors of four-month horizon forecasts are graphed as the specification moves from univariate to multivariate. The movement is accomplished by adjusting the parameter γ_2 in the prior which controls the relative tightness on other versus own lags.

While the overall results are mixed, at least for prices consistent improvement in forecasts is apparent as other variables enter the system. Such a result indicates the existence of a causal relation in the Granger "predictability" sense running from at least one of the other variables to prices.

The Problem of Seasonal Adjustment[18/]

Economic forecasts are almost always made in terms of seasonally adjusted data. A problem arises, however, when forecast performance statistics generated according to the method of the previous section using seasonally adjusted data are compared to compiled statistics on forecast errors generated by forecasters operating to real time. Seasonally adjusted economic time series are generated by a two-sided filtering of raw (not seasonally adjusted) data. They are thus based on more information than is available when each value of the seasonally adjusted series first becomes available. The current observation necessarily depends on a seasonal adjustment procedure which uses, at most, current and past values of the raw data.

The purpose of this section is to show that the advantage obtained by using a seasonally adjusted data series is not significant. This is accomplished for a particular example with a strong seasonal by demonstrating a one-sided seasonal adjustment procedure which, when applied to raw data, leads to forecasts of the two-sided seasonally adjusted series with only slightly larger mean square error than those generated by using the past values of the series itself.

The results in this section were motivated by an attempt to compare the forecast errors of two-month growth rates of M1 generated by a vector autoregression with the compiled errors of the Federal Reserve Board.[19/] The standard procedure for obtaining a "final" seasonally adjusted M1 series requires the application of a two-sided filter to the raw data. Letting the two-sided seasonally adjusted data be $M1_S$, the unadjusted data be $M1_N$, then a two-sided filter is represented by

$$M1_S(t) = F[M1_N(t-k), M1_N(t-k+1), \ldots, M1_N(t), \ldots, M1_N(t+k)]. \qquad (85)$$

Thus, future, as well as past, values of the unadjusted $M1_N$ series are required for some length, k, depending on the filter.

The forecast performance statistics in this paper are based on currently available seasonally adjusted data. A forecaster operating in real time, however, faces a different, more difficult forecasting problem. At the time of the forecast, t, the forecaster does not know the final seasonally adjusted numbers for the last k periods, and he must first estimate $M1_S(t-s)$, s=0, 1, ..., k, based on $M1_N(t-s)$, s=0, 1, ..., and then project future values of $M1_S$. I refer to the estimation of $M1_S(t-s)$, s=0, 1, ..., k, on the basis of $M1_N(t-s)$, s=0, 1, ..., as one-sided seasonal adjustment, and let $M1_S^t(t-s)$, s=0, 1, ..., be the estimates. Notice that for s > k, $M1_S^t(t-s) = M1_S(t-s)$. Several methods have been suggested to accomplish this one-sided seasonal adjustment. I use the procedure suggested by Geweke [1978] which has the property of minimizing expected subsequent revision in the seasonal factors for wide sense stationary series whose autocovariance function is known.

In theory one could face the problem of generating comparisons with real-time forecasters by starting with unadjusted data and incorporating a one-sided seasonal adjustment procedure explicitly into the estimation process at each point in time during the projection period. I have not done this because of the large computing expense which would be involved. Not only would there be the cost of seasonal adjustment each period, but more importantly, because the data series themselves change, use of the Kalman filter updating algorithm would no longer be possible and a set of matrix inversions would, thus, have to be performed each period to form the desired projections.

Rather than use this costly procedure throughout, I have followed it here in a univariate system as an experiment in order to estimate the magnitude of the difference between using two-sided and one-sided seasonally adjusted data.

The two-sided seasonal adjustment procedure used here is a simplified version of the multiplicative, ratio-to-moving average method used in the U.S. Department of Commerce Census X-11 program as described in Shiskin, Young, and Musgrave [1967]. Given the entire $M1_N$ series, the seasonally adjusted series, $M1_s$, is formed as follows:

The first step is the calculation of $M1_c$, a centered 23-term symmetric moving average of $M1_N$.

$$M1_c(t) = \sum_{s=-11}^{11} a(s) M1_N(t-s). \tag{86}$$

The a's have the weights, .148, .138, .122, .097, .068, .039, .013, -.005, -.015, -.016, -.011, -.004, for s=0, ± 1, ± 2, ..., ± 11, respectively.

The second step is the calculation of S-I (seasonal-irregular) ratios, formed by dividing $M1_N$ by $M1_c$.

$$S\text{-}I(t) = \frac{M1_N(t)}{M1_c(t)}. \tag{87}$$

The third step is the calculation of seasonal factors, S(t), as a moving average of the S-I ratios individually for each month.

$$S(t) = \sum_{k=-3}^{+3} b(k) \cdot S\text{-}I(t-(k \cdot 12)). \tag{88}$$

The b's have the weights .200, .200, .133, .067 for K=0, ± 1, ± 2, ± 3, respectively.

Finally, the seasonally adjusted series, $M1_s$, is equal to $M1_N$ divided by the seasonal factors.

$$M1_s(t) = M1_N(t)/S(t). \tag{89}$$

It can be seen that this procedure requires values of the unadjusted series 47 steps ahead and previous to each period in order to calculate the seasonally adjusted value. Geweke notes that if we take the procedure in (86)-(89) as given, then the problem of minimizing subsequent revision in the seasonal factors of recent data is solved by optimally forecasting the not seasonally adjusted data.

First, it should be noted that this mechanical procedure, when applied to not seasonally adjusted M1 data, produces a series which closely approximates the published seasonally adjusted M1 series, M1-SA. A comparison of the two during the projection period 1972-2 through 1977-11, along with the unadjusted data, is given in Figure 5. The series labeled "Filtered M1-NSA" is generated using the above method and is based on unadjusted data available through 1978-11 and projections of $M1_N$ beyond that date. Shown are deviations from constant and trend of the logarithms of each of the three series.

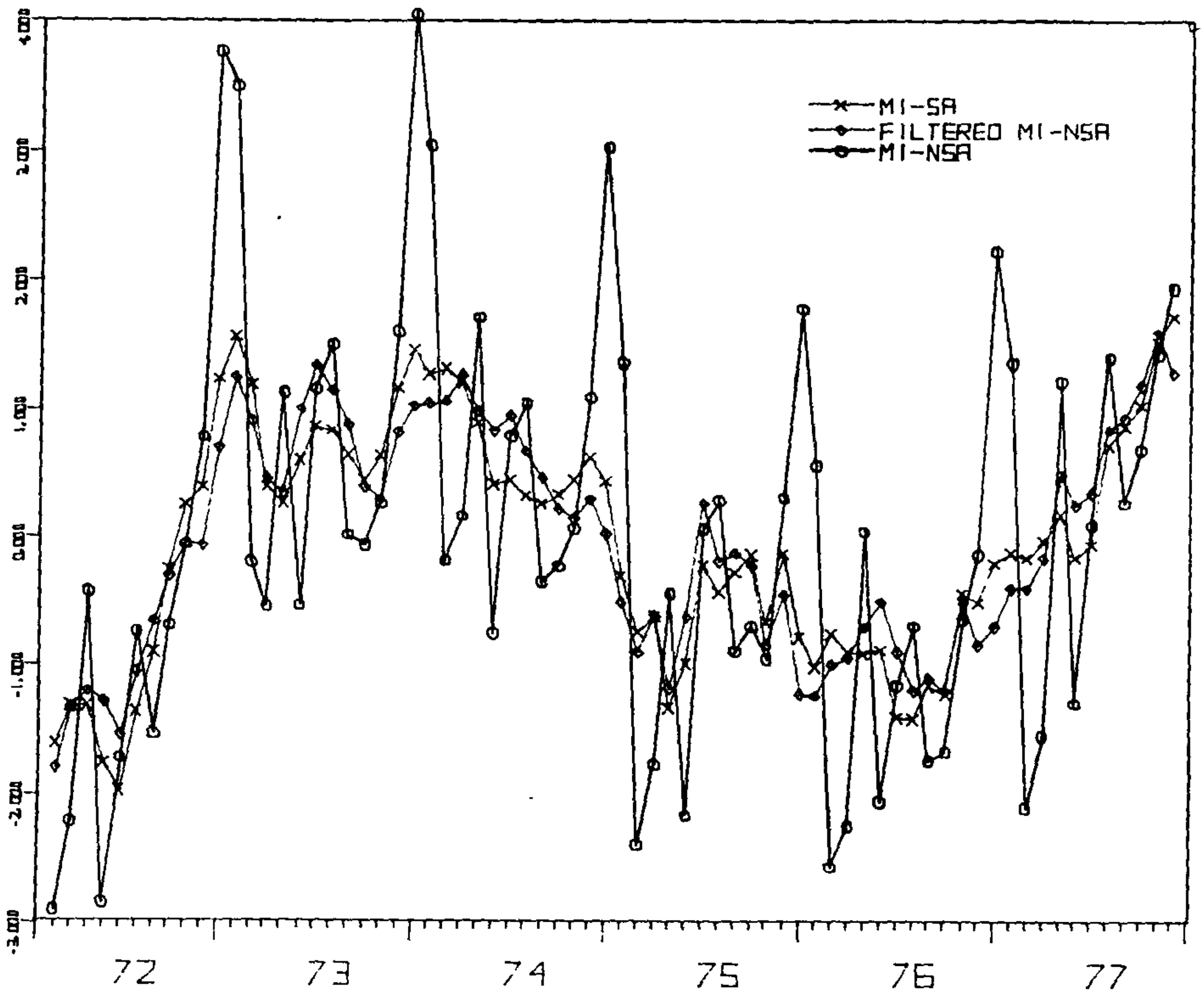
Given the two-sided seasonal adjustment procedure defined above, a one-sided adjustment procedure is defined by a specification of a method for projecting not seasonally adjusted data. The method followed here is suggested by Geweke. The regression of $M1_N$ on its first 7 and 12 through 20 lags is formed, and $M1_N$ is projected ahead using the chain rule. With these values for $M1_N$ the above two-sided adjustment procedure is applied to generate $M1_s^t(t-s)$, s=0, 1, ...

The above techniques allow one to generate two series of growth rates of seasonally adjusted M1 as follows:

1-S NSA(t) = $600[M1_s^t(t)-M1_s^t(t-2)]/M1_s^t(t-2)$, that is, the two-month growth rates of seasonally adjusted M1 which become available month by month through the use of the one-sided seasonal adjustment procedure; and

Figure 5

This graph compares the seasonally adjusted M_1 series generated by the two-sided seasonal filter defined in the text (Filtered M1-NSA) with the published seasonally adjusted series (M1-SA) and not seasonally adjusted series (M1-NSA). All three are shown as deviations from constant and trend of logarithms of the data.



2-S NSA(t) = $600[M1_s(t)-M1_s(t-2)]/M1_s(t-2)$, the two-month growth rates of $M1_s$, the series generated by the two-sided seasonal adjustment procedure applied to the entire not seasonally adjusted M1 series.

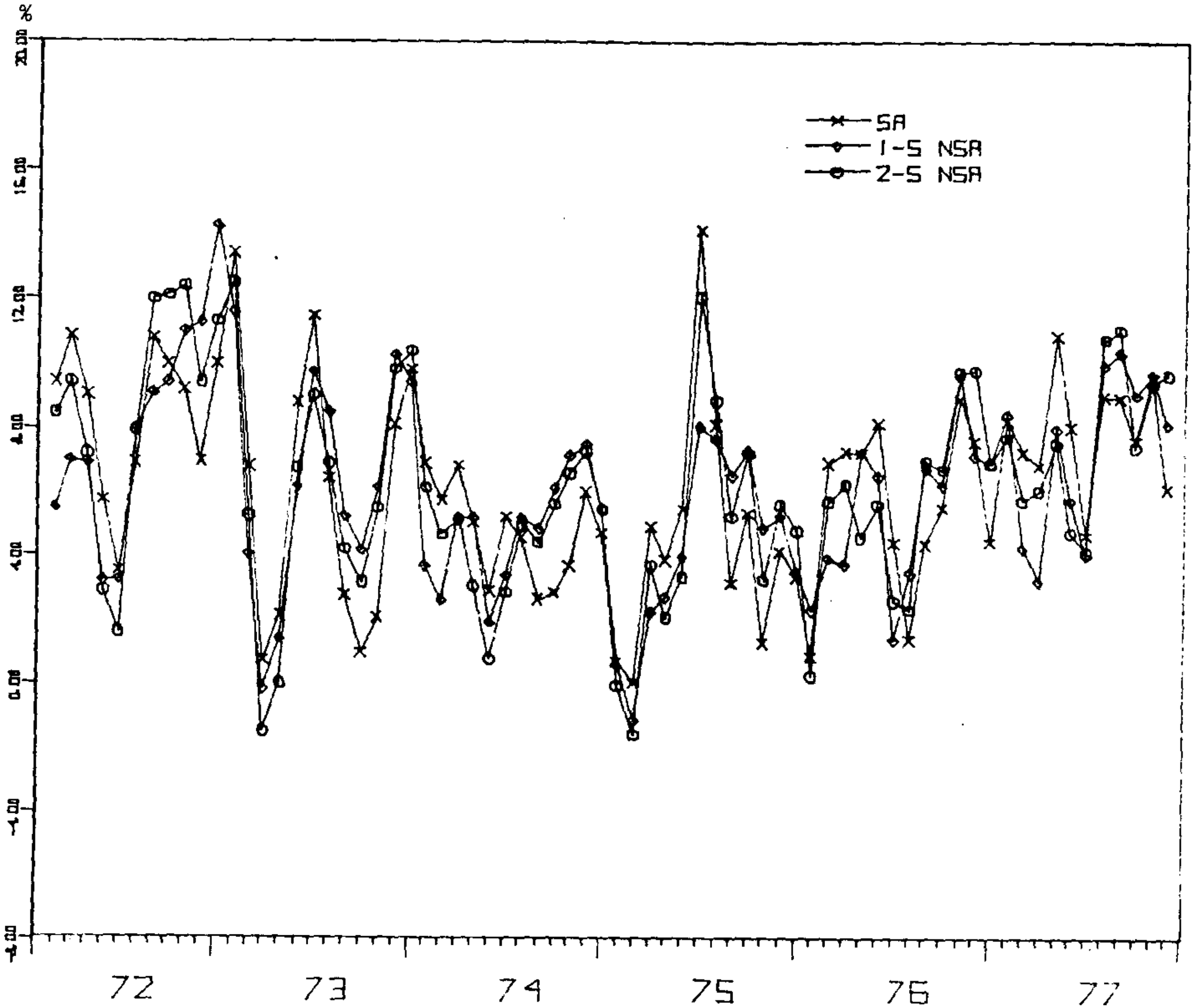
In Figure 6 these two series are plotted along with "SA," the growth rates implied by the final published seasonally adjusted M1 data. The series "1-S NSA" closely approximates both "2-S NSA" and "SA."

Now define "real-time forecasting" as the technique of forecasting future values of $M1_s$ based on $M1_s^t$, that is, of reestimating. $M1_s^t(t-s)$, s=0, 1, ..., on the basis of $M1_N(t-s)$, s=0, 1, ..., each period and using those estimates to project $M1_s(t+1)$, $M1_s(t+2)$, ... I have calculated error statistics comparing the forecasts of growth rates of $M1_s$ generated by this method with those generated by projecting $M1_s$ on itself.

These experiments show that there is only a rather small advantage gained by using the final seasonally adjusted numbers throughout rather than the real-time forecasting method. The one-step projections by a univariate sixth-order autoregression of two-month growth rates of "SA" generate a root mean square error of 2.28 on the projection period. The corresponding error statistic using $M1_s$ to forecast growth rates of $M1_s$ is 2.34. The root mean square error of forecasts using the real-time forecasting method to forecast growth rates of $M1_s$ is 2.43. Thus, a procedure of real-time forecasting exists, which, in the univariate case, generates errors only slightly larger than the method of projecting final seasonally adjusted data.

Figure 6

A comparison of the growth rates of three seasonally adjusted M_1 series shows that applying the one-sided and two-sided adjustment procedures defined in the text to not seasonally adjusted data generates series with growth rates very close to those of the published seasonally adjusted data.



Forecast Evaluation

Fair [1978b] has proposed a new method of evaluating the forecasting performances of different econometric models. This section extends that method by considering several alternative ways of estimating the total uncertainty of a model. It also applies the method to evaluate the effects of applying a Bayesian prior in the estimation of a vector autoregression.

Most studies[20/] have relied upon the calculation of the root mean square error, RMSE, of ex post forecasts, or a scaled version known as the Theil U coefficient. Fair's method uses the ex post forecast errors, but adjusts the RMSE to account for changes in the variance of forecasts over time. Fair [1978c] has presented the results of applying his method to four macroeconomic models including Sims' [1977] six-variable, unconstrained vector autoregression. Here an examination is made of application of the method to the same VAR model and one with the addition of a Bayesian prior distribution.

Since the details of Fair's method are rather involved and fully described in [1978b], they will be only briefly discussed here. The method partitions forecast uncertainty into four sources: (1) error terms, (2) coefficient estimates, (3) exogenous variable forecasts, and (4) misspecification[21/] of the model. Estimates of uncertainty due to error terms and coefficient estimates are calculated at a given time by stochastic simulation. In the VAR models there is no uncertainty from exogenous variables.

Uncertainty from misspecification of the model is estimated by a comparison of the variances computed by stochastic simulation with estimated variances computed from post-sample forecast errors. Fair's procedure relies on an assumption of constancy for the degree of misspecification of the model. Several alternative assumptions will be considered here.

There are two other minor modifications which were made in Fair's procedure for this study.[22/] The first is that the estimator of the covariance matrix of residuals has been corrected for degrees of freedom. This correction is particularly important in highly parameterized models such as the unconstrained VAR, and as will be shown, failure to include it biases Fair's results towards a higher degree of misspecification for these models.

Also, the one-quarter lag which Fair places between his estimation period and the projection period has been dropped. In this respect Fair's procedure mimics the behavior of large econometric modelers who, because of cost considerations, do not reestimate their models on the basis of preliminary data. There would seem to be no reason not to reestimate a linear model on the basis of preliminary data.

The purpose of generating stochastic simulations is to estimate the first two moments of the distributions of forecasts. Note that the mean of a forecast distribution is in general not given by the common procedure of plugging in zero errors as is true of linear projections. Even though the VAR models generate linear one-step-ahead forecasts, the k-step forecast for $k \geq 2$ is a nonlinear function of the parameters. On the other hand, the results in Table 2 of this section support the conclusion of Fair and others that the bias in forecasts generated by the zero error method is relatively small.

The first step in Fair's procedure is to generate error simulations to estimate the mean and variance of forecasts holding the estimated coefficients constant and picking normally distributed random errors with a covariance matrix equal to the sample covariance matrix of the residuals. The standard deviations of these distributions are the error term uncertainties. There is a different error term uncertainty calculated each period for each variable and forecast horizon. The second step is to generate coefficient simulations in which, in addition to random errors, coefficients are picked randomly according to the posterior distribution of the coefficients. The coefficient estimation uncertainties are defined as the differences between the standard deviations of the coefficient simulations and the standard deviations of the error simulations.

If the specification of the model, in this case, for example, a six-variable, four-lag VAR, were the true underlying stochastic mechanism which generated the observations, then the variance of the combined error term and coefficient simulations would be the variance of forecast errors generated by the model. Fair thus suggests that the average difference between the forecast error squared and this variance be called the misspecification of the model. He suggests that for variables with trend rather than averaging directly, the averaging should be made with misspecification expressed as a percent of the variable's level squared.

Given the estimated average misspecification, Fair calculates the total uncertainty of a given forecast; that is, the uncertainty due to errors, coefficient estimates, and misspecification. It is the square root of the sum of coefficient simulation variance and the average misspecification and is expressed either in units of the variable or as a percent of its level depending on whether the variable has a trend.

The six variables in Sims' VAR are Money Supply, Real GNP, the GNP Price Deflator, Wage Rates, the Import Price Deflator, and Unemployment Rate. Each regression equation includes a constant, trend, seasonal dummies, and four lags of each variable in the system. The model is estimated using OLS separately on each of the six equations. All variables except unemployment rate are estimated in logs, but in the process of stochastic simulation, exponentiation is performed so that forecasts, standard errors, and misspecification are all expressed in terms of the original levels. The five variables other than unemployment rate contain trends, so their misspecifications are calculated as percents of their levels.

The quarterly, seasonally adjusted data was taken from the Fair model data bank and is described in Fair [1978a]. The period of estimation begins in 1954-1 for the dependent variables and ends in 1978-1. Forecasts and estimates of misspecification are generated for each of the 35 periods from 1969-1 to 1977-3, using stochastic simulations with 50 drawings of error terms and coefficients. The estimator of total uncertainty is based on a set of forecast simulations made using data through 1978-1. Five hundred drawings of error vectors are made for this final stochastic simulation.

The specification of the second model is exactly the same as the first, except that each equation is estimated with the addition of a prior distribution. In terms of the parameters defined earlier, this is a prior with parameters $\gamma_1 = 1$, $\gamma_2 = .5$, and $\lambda = .1$.

Calculations for the model without a prior required 644 seconds on the CDC Cyber 172 computer at the University of Minnesota, while the model with a prior required 1024 seconds. The additional time was due almost entirely to the necessity of factoring six different coefficient covariance matrices each period rather than the one which is the same for all equations in the no-prior model.

The two modifications which were made in Fair's procedure, correcting for degrees of freedom and dropping the one-quarter estimation lag, made very significant changes in the results. Table 2 compares Fair's results with, first, a duplication of his procedure (slight differences may be attributed to the randomness introduced in stochastic simulations), then to a specification with those above-mentioned two changes, and finally to the model with the prior. The correction for degrees of freedom has the effect of increasing the estimated variance of errors in the model. Thus, the error term uncertainty increases.

TABLE 2
Estimates of Uncertainty

Model I(a) = Results in Fair [3].

Model I(b) = Same model with different simulations.

Model II = Model with corrected degrees of freedom and no estimation lag.

Model III = Model estimated with prior.

a = Uncertainty due to error terms.

b = Uncertainty due to error term and coefficient estimates.

c = Total uncertainty.

d = Estimated degree of misspecification (c-b).

Forecast Horizon	1	2	3	4
Real GNP:				
Model I(a)				
a	.640	.910	1.070	1.290
b	.880	1.320	1.640	2.130
c	1.300	2.290	3.040	4.040
d	.420	.970	1.400	1.910
Model I(b)				
a	.648	.903	1.034	1.288
b	.863	1.320	1.604	2.120
c	1.299	2.292	3.129	4.221
d	.436	.973	1.525	2.100
Model II				
a	.730	1.066	1.231	1.475
b	1.091	1.577	2.036	2.584
c	1.275	2.061	2.557	3.213
d	.183	.483	.520	.629
Model III				
a	.787	1.169	1.408	1.636
b	.992	1.494	2.010	2.570
c	.985	1.563	2.322	3.251
d	-.007	.069	.312	.680
GNP Deflator:				
Model I(a)				
a	.220	.300	.420	.560
b	.280	.420	.630	.880
c	.500	.770	1.340	2.070
d	.220	.350	.710	1.190
Model I(b)				
a	.216	.283	.393	.524
b	.276	.419	.614	.870
c	.504	.791	1.356	2.106
d	.228	.371	.742	1.235

Model II				
a	.255	.355	.504	.677
b	.356	.518	.745	1.029
c	.435	.671	1.057	1.673
d	.079	.154	.312	.644
Model III				
a	.287	.411	.535	.654
b	.325	.528	.707	.927
c	.357	.751	1.197	1.786
d	.032	.223	.491	.859
Unemployment Rate:				
Model I(a)				
a	.210	.350	.430	.490
b	.280	.480	.590	.680
c	.440	.850	1.190	1.300
d	.160	.370	.600	.620
Model I(b)				
a	.213	.351	.420	.485
b	.282	.489	.596	.671
c	.464	.897	1.260	1.410
d	.182	.407	.660	.741
Model II				
a	.252	.430	.539	.607
b	.363	.609	.768	.852
c	.423	.805	1.049	1.219
d	.060	.195	.281	.367
Model III				
a	.302	.465	.597	.661
b	.357	.573	.754	.842
c	.356	.751	1.123	1.445
d	-.002	.178	.368	.602
Money Supply:				
Model I(a)				
a	.690	.860	1.020	1.170
b	.920	1.190	1.560	1.920
c	1.370	1.530	2.290	3.290
d	.450	.340	.730	1.370
Model I(b)				
a	.734	.841	1.054	1.191
b	.921	1.206	1.512	1.912
c	1.344	1.541	2.210	3.159
d	.422	.334	.697	1.247
Model II				
a	.801	1.023	1.170	1.482
b	1.080	1.452	1.939	2.387
c	1.424	1.575	1.616	2.079
d	.344	.123	-.323	-.307
Model III				
a	.898	1.246	1.494	1.817
b	1.090	1.524	2.077	2.512
c	1.350	1.564	2.056	2.305
d	.260	.040	-.021	-.207

Wage Rate:				
Model I(a)				
a	.470	.620	.710	.770
b	.630	.900	1.080	1.230
c	1.230	2.200	3.020	3.710
d	.600	1.300	1.940	2.480
Model I(b)				
a	.491	.651	.750	.776
b	.636	.919	1.059	1.187
c	1.274	2.212	3.085	3.758
d	.637	1.293	2.026	2.570
Model II				
a	.565	.743	.855	.947
b	.758	1.071	1.329	1.571
c	.896	1.508	2.207	2.781
d	.138	.436	.879	1.210
Model III				
a	.651	.843	.969	1.068
b	.766	1.040	1.289	1.493
c	.605	.718	.743	.509
d	-.161	-.322	-.545	-.984
Import Price Deflator:				
Model I(a)	not available			
Model I(b)				
a	1.124	2.073	3.105	4.028
b	1.524	3.011	4.859	6.659
c	3.770	8.470	13.799	19.327
d	2.246	5.458	8.941	12.669
Model II				
a	1.406	2.471	3.752	4.857
b	1.924	3.526	5.861	8.136
c	3.283	7.112	12.401	17.667
d	1.359	3.586	6.540	9.531
Model III				
a	1.646	2.514	3.464	4.292
b	1.995	3.382	4.926	6.423
c	3.207	6.976	11.816	16.924
d	1.211	3.593	6.889	10.501

Similarly, the estimated variance of coefficient estimates also rises and causes the coefficient uncertainty to increase. Since the post-sample forecast errors are approximately the same in both cases, the estimated misspecification decreases substantially when the correction for degrees of freedom is made.

The addition of the prior causes error uncertainty to increase, but decreases the increment due to coefficient estimation. Their sum, uncertainty in the coefficient simulations, is generally about the same for both models, although it is usually slightly smaller in the model with the prior.

Table 3 shows a comparison of the RMSEs for the two unrestricted VARs and the VAR with a prior. The dropping of the estimation lag improved forecasting performance slightly, but an additional larger improvement is gained by the addition of the prior. The forecasts considered here are both the means of the coefficient simulations and the zero error RMSEs. Little, if any, gain in forecast accuracy is obtained by the stochastic simulation process. With only 50 draws per period, the increased variance seems to dominate the decreased bias in most cases.

In the following graphs the misspecification over time for each of the six variables is displayed for both specifications. The series are scaled and plotted with respect to the time at which the forecasts were made.

As a first approximation, the assumption of constancy of the mean over time does not seem unreasonable, although there does appear to be a small positive trend for unemployment and import prices. It is not so much the mean of these series as their variance which appears to change during this period. The increased variance during the 1973-1976 period is apparent in all six variables, and is particularly striking in import prices. To a large degree the increased variance of the misspecification is matched by an increase in the variance of the coefficient simulations. For example, the average standard error of a four-step

TABLE 3

Root Mean-Squared Errors of Ex Post Forecasts

Model I = Fair's VAR.

Model II = Degrees of freedom and estimation lag-corrected VAR.

Model III = VAR with prior.

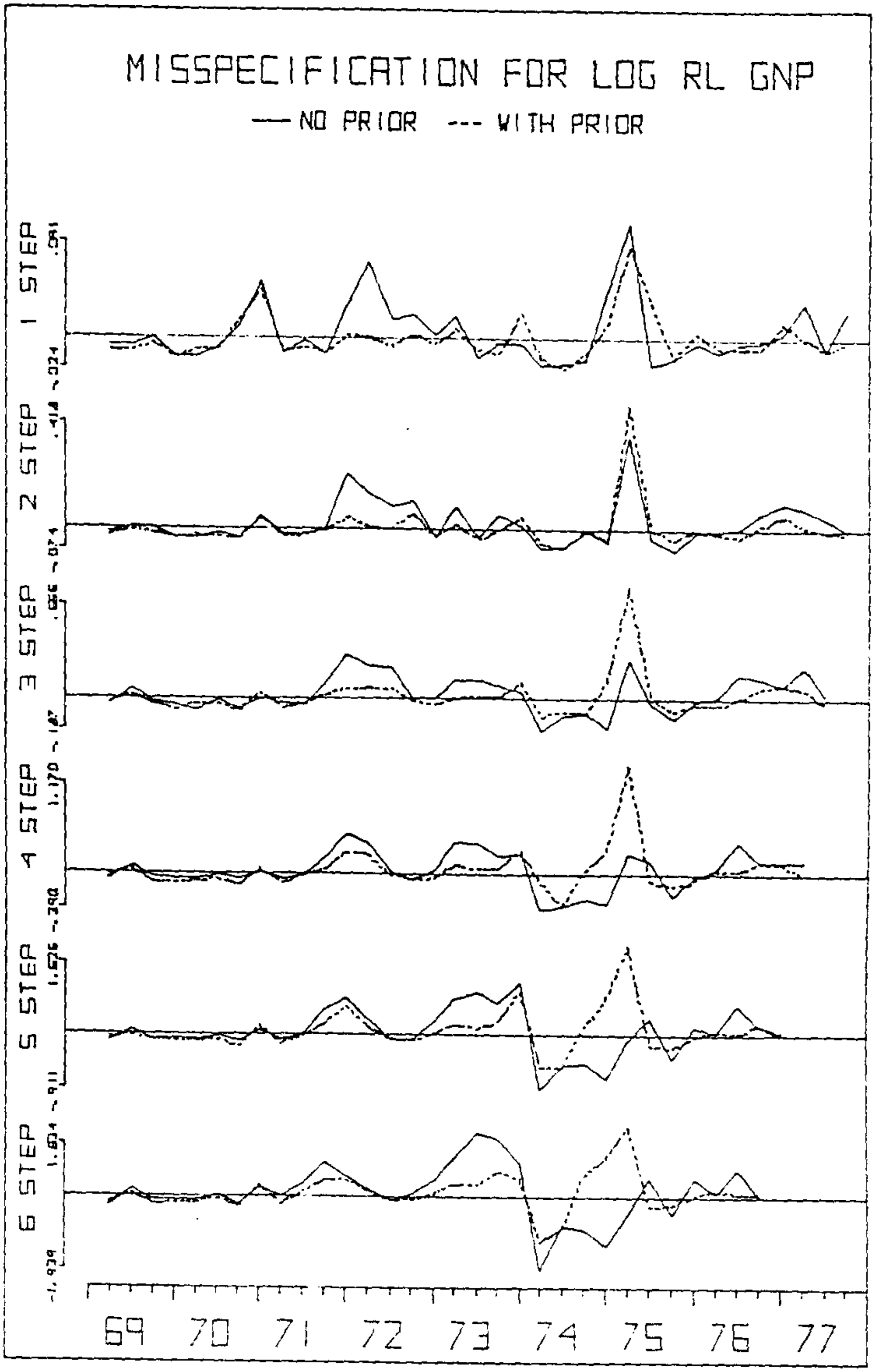
(a) = Forecasts from coefficient simulations.

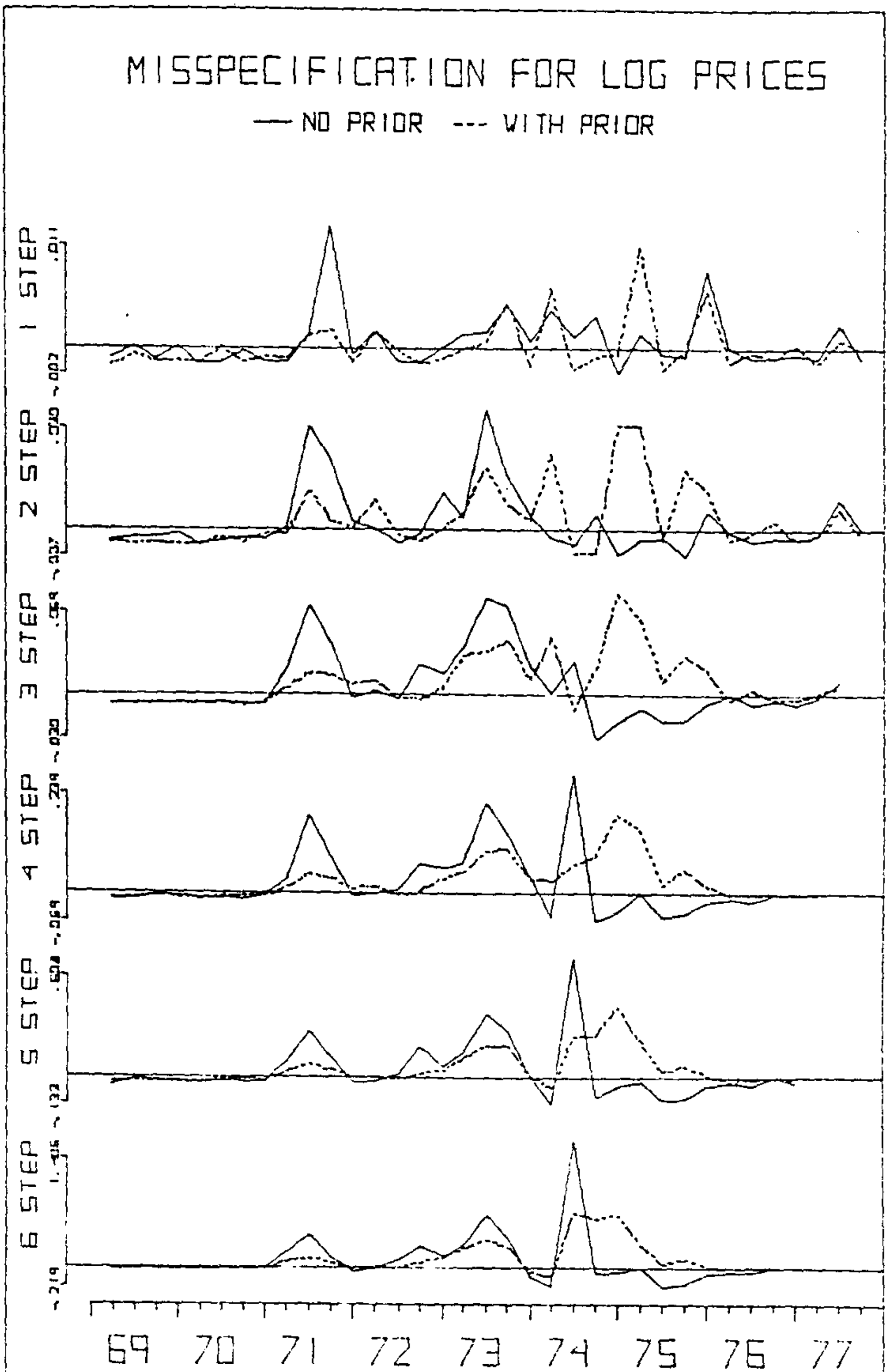
(b) = Forecasts with zero errors.

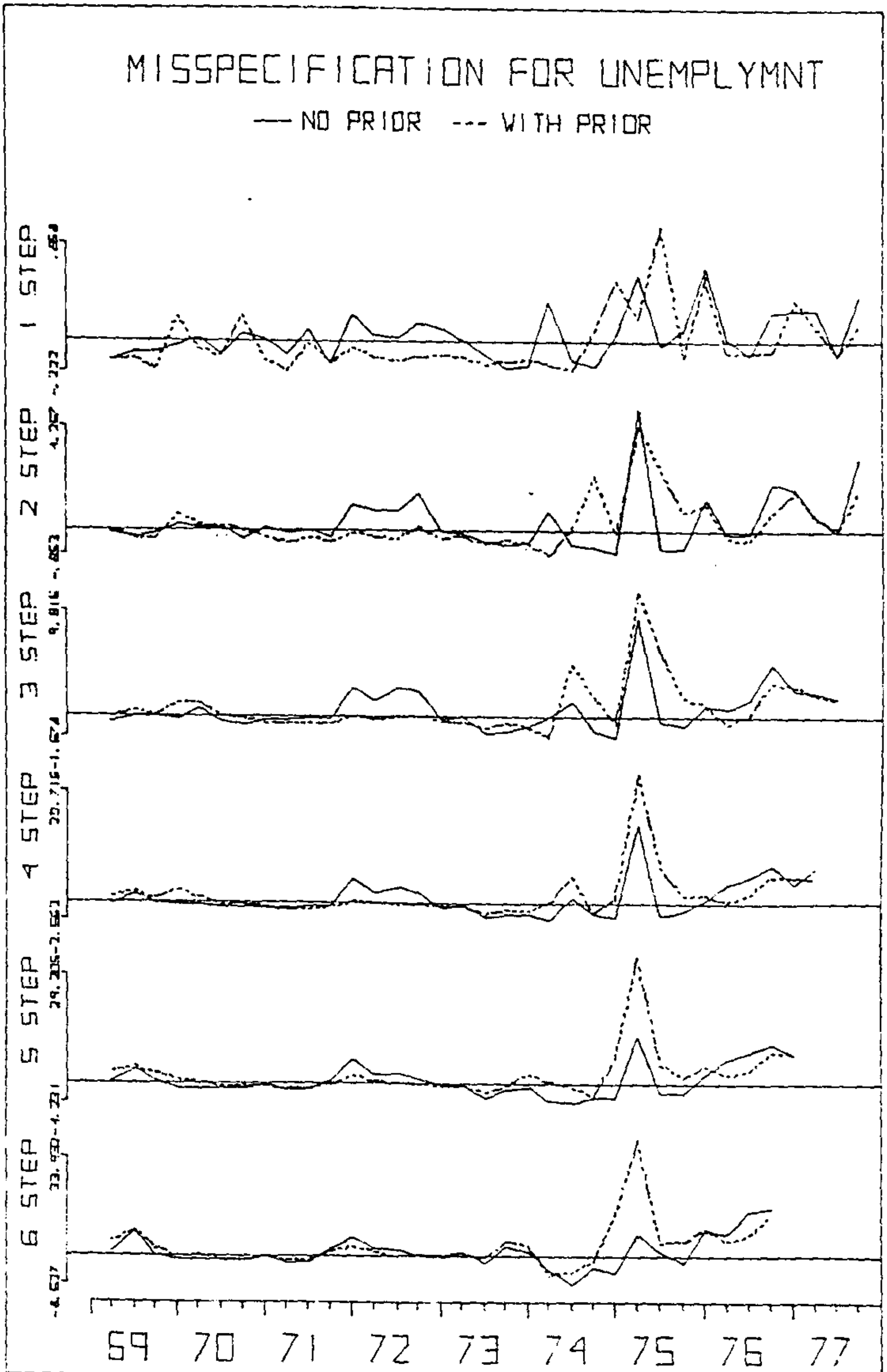
Forecast Horizon		1	2	3	4
Real GNP:					
Model I	a)	1.406	2.494	3.545	4.954
	b)	1.391	2.466	3.511	4.903
Model II	a)	1.329	2.272	3.061	4.159
	b)	1.269	2.201	2.988	4.105
Model III	a)	1.091	1.869	2.778	3.831
	b)	1.069	1.835	2.656	3.705
GNP Deflator:					
Model I	a)	.548	.899	1.521	2.393
	b)	.546	.886	1.498	2.336
Model II	a)	.464	.788	1.276	2.086
	b)	.463	.778	1.235	1.975
Model III	a)	.408	.814	1.308	1.947
	b)	.406	.793	1.270	1.883
Unemployment Rate:					
Model I	a)	.484	.953	1.355	1.547
	b)	.476	.944	1.339	1.522
Model II	a)	.435	.858	1.161	1.407
	b)	.432	.850	1.148	1.395
Model III	a)	.395	.809	1.189	1.561
	b)	.385	.801	1.179	1.547
Money Supply:					
Model I	a)	1.438	1.933	3.072	4.522
	b)	1.472	1.968	3.113	4.548
Model II	a)	1.506	2.018	2.749	3.960
	b)	1.533	2.072	2.741	3.984
Model III	a)	1.319	1.776	2.481	3.252
	b)	1.345	1.744	2.425	3.087
Wage Rate:					
Model I	a)	1.375	2.468	3.601	4.544
	b)	1.349	2.436	3.487	4.399
Model II	a)	1.017	1.790	2.752	3.619
	b)	.994	1.785	2.714	3.611
Model III	a)	.698	1.038	1.415	1.734
	b)	.673	.958	1.336	1.661

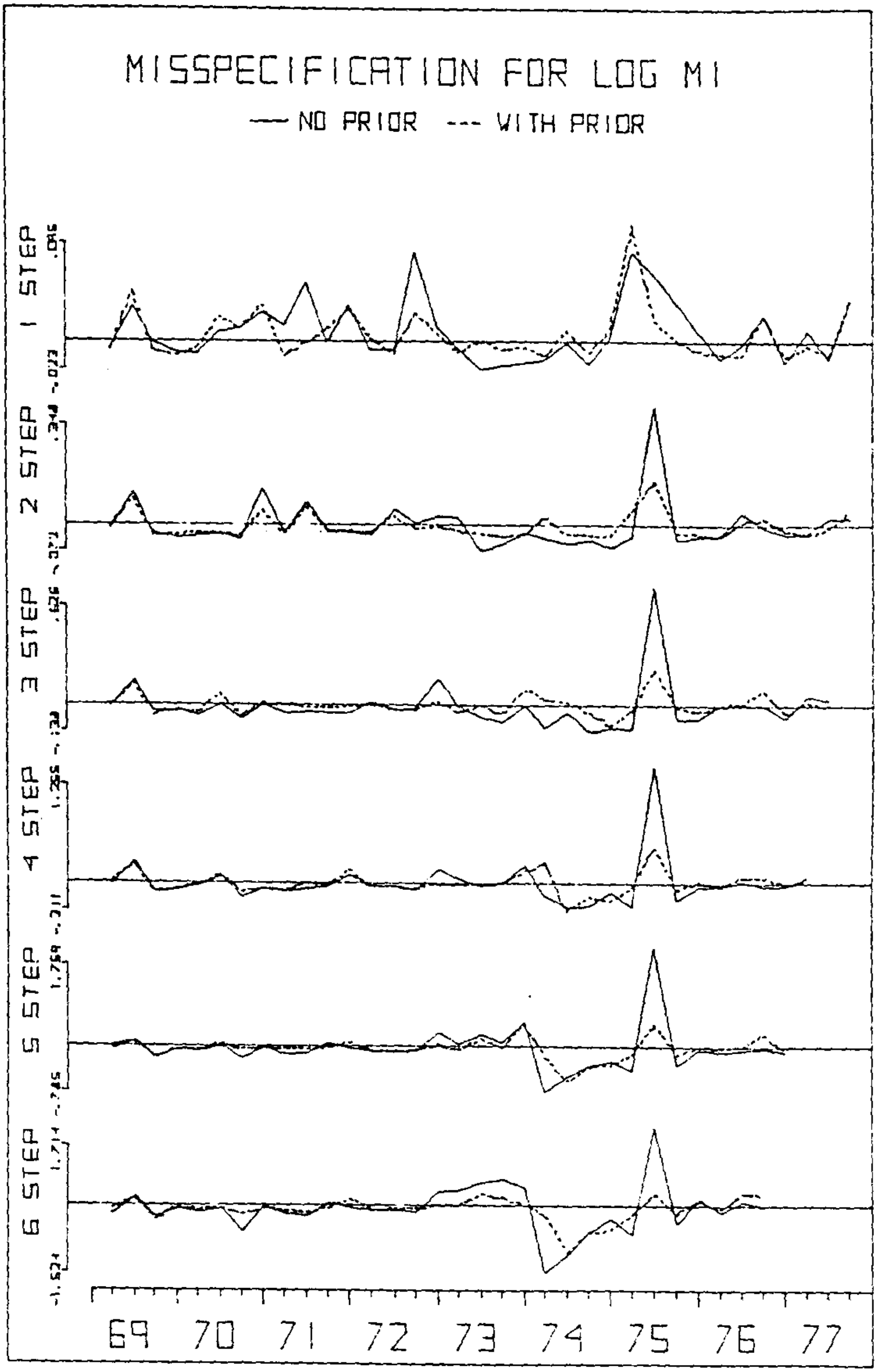
Import Price Deflator:

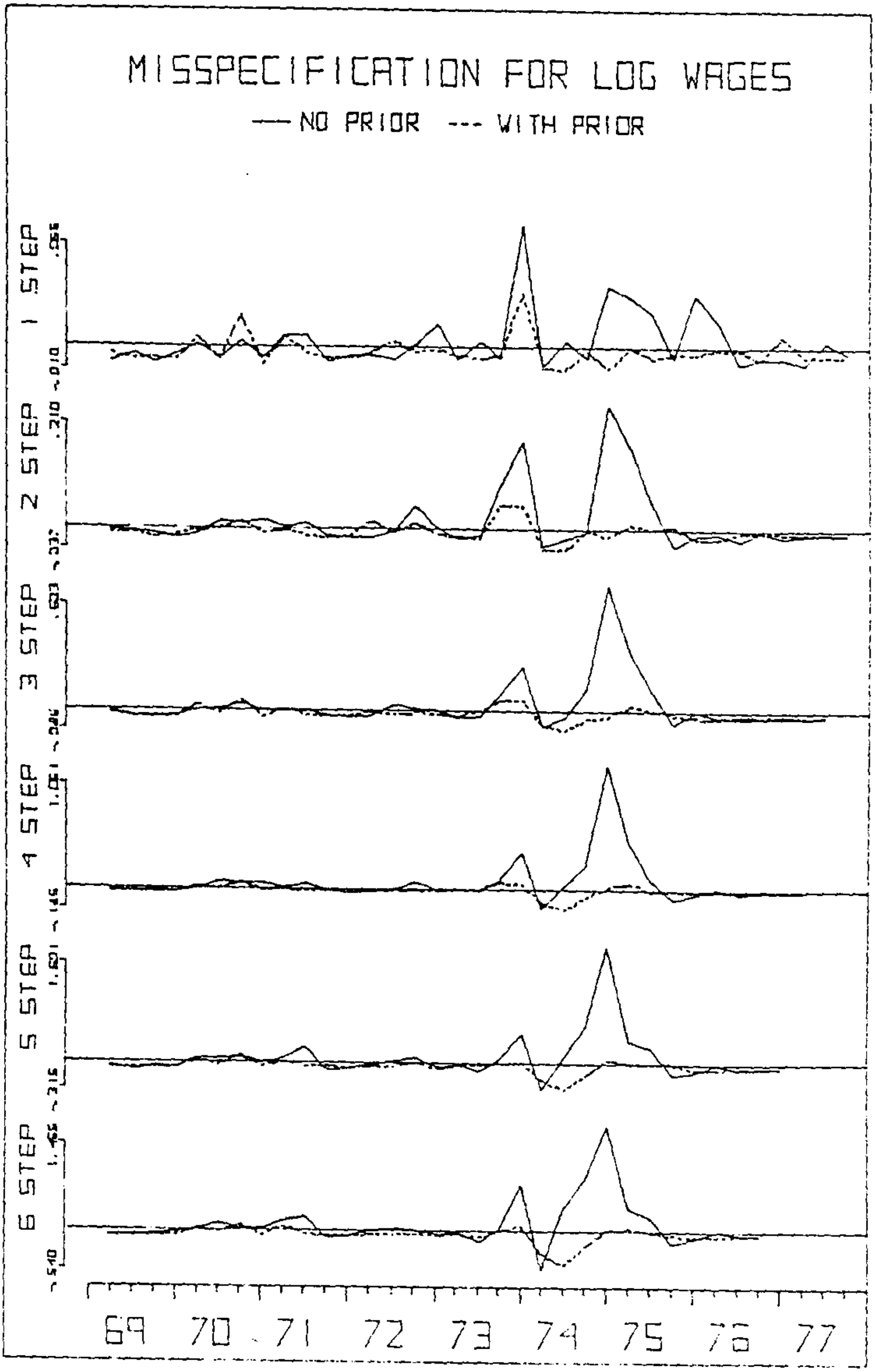
Model I	a)	3.859	9.442	18.039	31.867
	b)	3.922	9.400	17.646	30.469
Model II	a)	3.134	7.627	14.974	26.507
	b)	3.081	7.337	14.199	24.431
Model III	a)	3.037	7.558	14.850	26.285
	b)	2.988	7.329	14.448	25.459

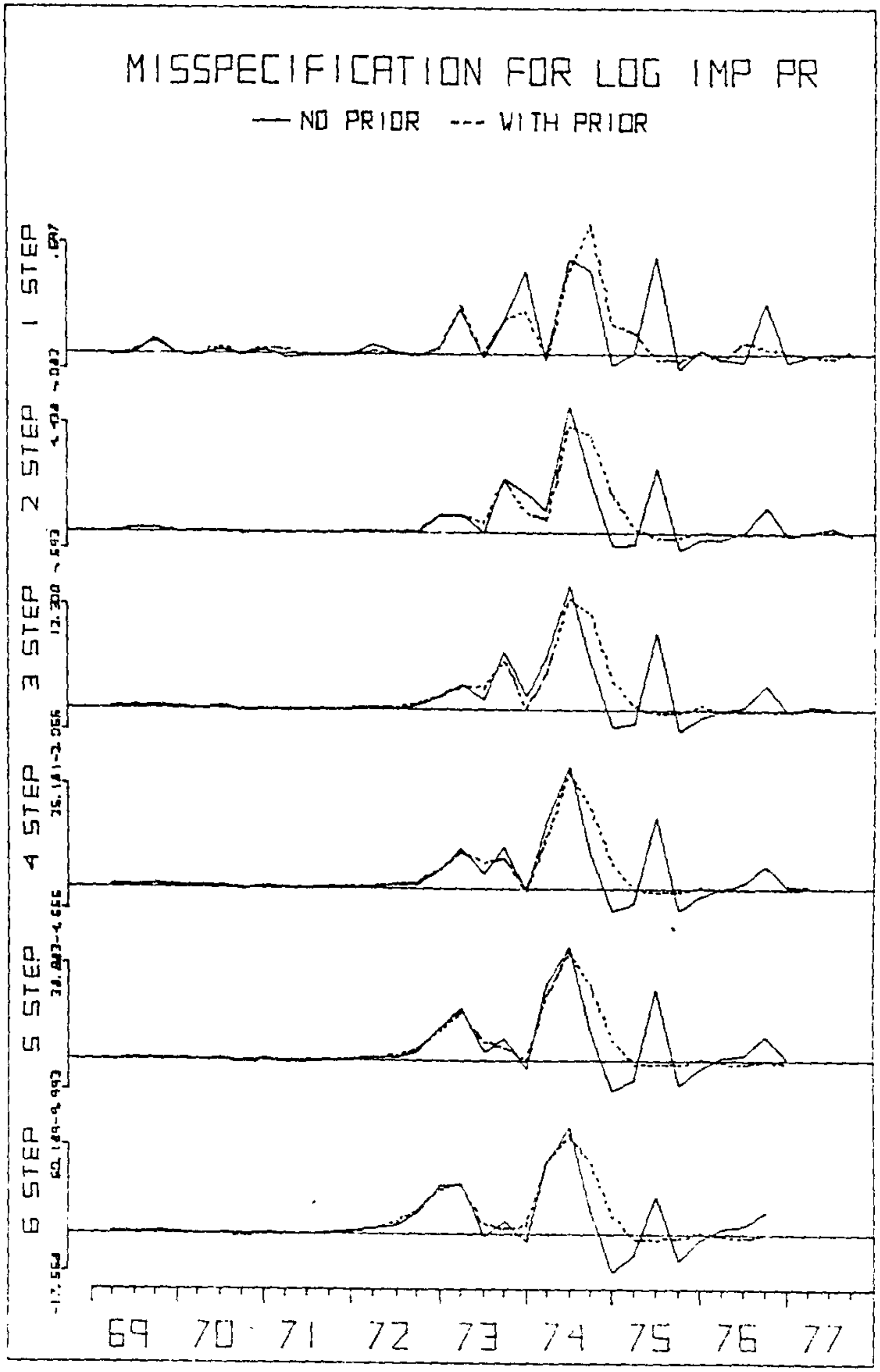

MISSPECIFICATION FOR LOG RL GNP
—— NO PRIOR
--- WITH PRIOR
1 STEP
2 STEP
3 STEP
4 STEP
5 STEP
6 STEP
69
70
71
72
73
74
75
76
77


MISSPECIFICATION FOR LOG PRICES
— NO PRIOR --- WITH PRIOR
1 STEP
2 STEP
3 STEP
4 STEP
5 STEP
6 STEP
69
70
71
72
73
74
75
76
77


MISSPECIFICATION FOR UNEMPLYMNT
— NO PRIOR --- WITH PRIOR
1 STEP
2 STEP
3 STEP
4 STEP
5 STEP
6 STEP
69
70
71
72
73
74
75
76
77


MISSPECIFICATION FOR LOG MI
—— NO PRIOR --- WITH PRIOR
1 STEP
2 STEP
3 STEP
4 STEP
5 STEP
6 STEP
69
70
71
72
73
74
75
76
77


MISSPECIFICATION FOR LOG WAGES
— NO PRIOR --- WITH PRIOR
1 STEP
.056
-.010
2 STEP
.210
3 STEP
4 STEP
-.145
5 STEP
-.315
6 STEP
-.540
69 70 71 72 73 74 75 76 77



MISSPECIFICATION FOR LOG IMP PR
— NO PRIOR --- WITH PRIOR
1 STEP
2 STEP
3 STEP
4 STEP
5 STEP
6 STEP
69
70
71
72
73
74
75
76
77

forecast of import prices made in 1971 was 3.65. By 1973 it was 11.01, and in 1974 it was 47.44.

Fair's measure of total uncertainty may be used to rank the expected forecasting performance in the current period for different models. In this sense it may provide a substitute for the usual practice of ranking models according to their RMSE. Note that if the current stochastic simulation variance is the same as the average over the ex post forecast period, then regardless of its level the total uncertainty equals the RMSE. In this sense, the total uncertainty is basically an RMSE which has been adjusted for the difference between the average model variance over the forecast period and the current model variance.

There is a weakness in this measure of average misspecification. Let e^2_{tk} be the k-step forecast error squared at time t, and $\hat{\sigma}^2_{tk}$ be the estimated coefficient simulation variance. Then $d_{tk} = e^2_{tk} - \hat{\sigma}^2_{tk}$ is the misspecification at time t, and its simple average, $\bar{d}_k$, is Fair's estimate of model misspecification. This estimate seems to be, at best, inefficient. The underlying model assumes that e_{tk} is drawn from a distribution with zero mean and variance $\hat{\sigma}^2_{tk} + \bar{d}_k$. Fair's estimate of $\bar{d}_k$ assumes that each d_{tk} contains equal information about the value of $\bar{d}_k$, which implies that the variance of the d_{tk}'s does not vary with $\hat{\sigma}_{tk}$. This is possible, but would require that for large $\hat{\sigma}_{tk}$'s the e_{tk}'s be drawn from a distribution concentrated at plus and minus $\hat{\sigma}_{tk}$.

Consider the assumption that the e_{tk}'s are distributed normally with mean zero and variance $\sigma^2_{tk} = f(\hat{\sigma}^2_{tk})$ for some specified function f. Then the e^2_{tk}'s are distributed according to a chi-squared distribution, and one can estimate the parameters of f by maximum likelihood procedures.[23/]

Fair's assumption that $\bar{d}_k$ is constant implies that as the uncertainty of a forecast increases, the relative amount of misspecification decreases. In

TABLE 4
Estimates of the Functional Form for Total Uncertainty

Forms are: 1) $\hat{\sigma}_t^2 + \bar{d}$ 2) $\hat{\sigma}^2 + \tilde{d}$ 3) $\bar{\beta}\hat{\sigma}_t^2$ 4) $\hat{\alpha} + \hat{\beta}\hat{\sigma}_t^2$

(Standard Errors in Parenthesis)

	(Step)	$\bar{d}$	$\tilde{d}$		$\bar{\beta}$		$\hat{\alpha}$		$\hat{\beta}$	
Money Supply:										
No Prior	1	.086	.099	(.052)	1.807	(.043)	.175	(.160)	.377	(1.137)
	2	.037	.137	(.090)	1.402	(.044)	.395	(.243)	-.008	(.632)
	3	-.114	.073	(.140)	.978	(.050)	.335	(.261)	.431	(.358)
	4	-.137	-.004	(.199)	.904	(.066)	.181	(.348)	.731	(.356)
Prior	1	.063	.070	(.041)	1.649	(.038)	.138	(.141)	.336	(1.241)
	2	.012	.042	(.066)	1.093	(.038)	.145	(.145)	.531	(.520)
	3	-.008	-.022	(.107)	.950	(.052)	-.003	(.287)	.957	(.610)
	4	-.099	-.031	(.178)	.925	(.065)	.108	(.419)	.798	(.513)
Real GNP:										
No Prior	1	.043	.065	(.043)	1.515	(.036)	.280	(.144)	-.706	(.860)
	2	.176	.267	(.128)	1.904	(.065)	.604	(.322)	-.178	(.767)
	3	.239	.488	(.232)	1.931	(.083)	.984	(.450)	-.024	(.521)
	4	.365	.804	(.396)	1.939	(.110)	1.509	(.794)	.162	(.560)
Prior	1	-.001	.016	(.029)	1.083	(.026)	.245	(.060)	-.993	(.246)
	2	.021	.106	(.093)	1.256	(.047)	.628	(.201)	-.749	(.378)
	3	.135	.259	(.192)	1.391	(.071)	.788	(.496)	.035	(.680)
	4	.396	.571	(.356)	1.597	(.103)	1.109	(.916)	.389	(.820)
GNP Deflator:										
No Prior	1	.006	.005	(.004)	1.397	(.012)	.000	(.013)	1.379	(1.015)
	2	.018	.024	(.013)	1.721	(.020)	.039	(.027)	.510	(.656)
	3	.056	.076	(.036)	2.057	(.036)	.111	(.072)	.503	(.734)
	4	.174	.196	(.090)	2.230	(.059)	.218	(.161)	.863	(.806)
Prior	1	.002	.001	(.004)	1.132	(.013)	-.005	(.008)	1.527	(.731)
	2	.028	.024	(.014)	1.743	(.024)	-.003	(.037)	1.847	(1.203)
	3	.093	.070	(.035)	2.120	(.041)	-.006	(.062)	2.213	(1.111)
	4	.233	.161	(.075)	2.460	(.066)	-.030	(.099)	2.741	(1.164)
Unemployment Rate:										
No Prior	1	.047	.056	(.040)	1.448	(.107)	.092	(.084)	.564	(.643)
	2	.277	.339	(.164)	1.918	(.238)	.511	(.301)	.477	(.655)
	3	.511	.646	(.307)	1.985	(.334)	.998	(.560)	.411	(.658)
	4	.760	1.286	(.507)	2.569	(.435)	2.132	(.740)	-.125	(.413)
Prior	1	-.001	.025	(.038)	1.097	(.091)	.286	(.193)	-.819	(1.111)
	2	.235	.216	(.144)	1.580	(.229)	.087	(.455)	1.350	(1.225)
	3	.691	.665	(.319)	2.055	(.381)	.529	(.911)	1.217	(1.397)
	4	1.378	1.397	(.561)	2.670	(.538)	1.529	(1.143)	.844	(1.149)
Wage Rate:										
No Prior	1	.023	.014	(.019)	1.225	(.024)	.005	(.035)	1.158	(.569)
	2	.113	.019	(.042)	1.273	(.043)	-.073	(.084)	1.737	(.704)
	3	.311	.017	(.072)	1.305	(.067)	-.150	(.123)	1.806	(.630)
	4	.527	.032	(.112)	1.329	(.085)	-.293	(.242)	2.025	(.827)

Prior	1	-.022	-.017	(.010)	.671	(.013)	.008	(.032)	.554	(.504)
	2	-.057	-.040	(.022)	.666	(.019)	.025	(.079)	.493	(.546)
	3	-.111	-.103	(.023)	.600	(.027)	-.126	(.032)	1.066	(.266)
	4	-.197	-.179	(.023)	.589	(.032)	-.203	(.072)	1.108	(.317)
Import Price Deflator:										
No Prior	1	.708	.568	(.206)	3.753	(.117)	-.003	(.251)	3.768	(1.735)
	2	3.814	3.060	(1.056)	5.233	(.248)	.362	(.796)	4.464	(1.862)
	3	11.943	8.851	(3.152)	6.202	(.403)	1.659	(2.137)	4.559	(2.019)
	4	24.592	16.790	(6.103)	7.574	(.566)	8.077	(4.887)	3.165	(1.605)
Prior	1	.630	.499	(.180)	3.660	(.108)	.082	(.245)	3.167	(1.584)
	2	3.723	2.778	(.927)	5.638	(.242)	.571	(.729)	4.313	(1.765)
	3	11.534	7.577	(2.596)	7.377	(.387)	2.115	(1.627)	4.603	(1.865)
	4	24.517	14.957	(5.119)	10.041	(.614)	7.366	(3.655)	3.851	(1.782)

TABLE 5
Four Measures of Total Uncertainty in 1978-1

	(Step)	$\hat{\sigma}_t^2 + \bar{d}$	$\hat{\sigma}_t^2 + \tilde{d}$	$\bar{\beta}\hat{\sigma}_t^2$	$\hat{\alpha} + \hat{\beta}\hat{\sigma}_t^2$
Real GNP:					
No Prior	1	1.275	1.359	1.343	1.400
	2	2.061	2.272	2.176	2.368
	3	2.557	3.004	2.829	3.121
	4	3.213	3.837	3.599	4.022
Prior	1	.985	1.068	1.032	1.214
	2	1.563	1.814	1.675	2.146
	3	2.322	2.575	2.371	2.832
	4	3.251	3.511	3.249	3.698
GNP Price Deflator:					
No Prior	1	.435	.422	.421	.421
	2	.672	.719	.679	.731
	3	1.057	1.148	1.069	1.182
	4	1.674	1.739	1.538	1.759
Prior	1	.357	.344	.345	.328
	2	.752	.721	.697	.693
	3	1.197	1.097	1.029	1.022
	4	1.786	1.570	1.453	1.432
Unemployment Rate:					
No Prior	1	.424	.434	.437	.424
	2	.806	.843	.844	.830
	3	1.049	1.112	1.082	1.114
	4	1.219	1.418	1.365	1.429
Prior	1	.356	.392	.374	.425
	2	.751	.737	.720	.728
	3	1.123	1.111	1.082	1.106
	4	1.445	1.452	1.377	1.459
Money Supply:					
No Prior	1	1.423	1.469	1.451	1.478
	2	1.575	1.864	1.719	1.984
	3	1.616	2.120	1.918	2.230
	4	2.079	2.378	2.270	2.444
Prior	1	1.350	1.374	1.400	1.334
	2	1.564	1.657	1.594	1.639
	3	2.056	2.023	2.025	2.024
	4	2.305	2.449	2.417	2.476
Wage Rate:					
No Prior	1	.897	.844	.839	.844
	2	1.508	1.158	1.209	1.125
	3	2.207	1.391	1.518	1.298
	4	2.782	1.671	1.811	1.437

Prior	1	.605	.641	.627	.634
	2	.718	.820	.849	.886
	3	.743	.797	.999	.712
	4	.509	.664	1.146	.661
Import Price Deflator:					
No Prior	1	3.283	3.063	3.727	3.731
	2	7.112	6.560	8.067	7.690
	3	12.401	11.084	14.595	13.159
	4	17.667	15.300	22.391	17.038
Prior	1	3.207	2.995	3.817	3.666
	2	6.976	6.263	8.031	7.419
	3	11.815	10.002	13.380	11.526
	4	16.924	13.814	20.354	15.250

the context of a more general functional form, $\sigma_{tk} = f(\hat{\sigma}^2_{tk}) = \alpha + \beta\hat{\sigma}^2_{tk}$, Fair has assumed that $\beta = 1$. The asymptotic variances computed from the maximum likelihood procedure are usually too large to reject this hypothesis with strong confidence, particularly since only 35 observations are used in the estimation.

Nevertheless, since the measure of total uncertainty hinges on the estimate of misspecification, one would feel more confident using the total uncertainty measure if it were not highly sensitive to changes in the exact form of the specification of the measure for misspecification of the model.

In Tables 4 and 5 four functional forms for total uncertainty are compared for the two models. The first measure uses $\bar{d}_k + \hat{\sigma}^2_{tk}$ with $\bar{d}_k$ computed by Fair's method. The second uses $\tilde{d}_k + \hat{\sigma}^2_{tk}$, with $\tilde{d}_k$ computed using the more efficient maximum likelihood estimation. The third measure is $\bar{\beta}_k\hat{\sigma}^2_{tk}$, where $\bar{\beta}_k$ is a maximum likelihood estimate of β in the above specification for f with α constrained to be zero. Finally, total uncertainty is estimated by $\hat{\alpha}_k + \hat{\beta}_k\hat{\sigma}^2_{tk}$ with $\hat{\alpha}$ and $\hat{\beta}$ estimated by maximum likelihood.

The overall results do not seem to be highly sensitive to which measure is used. Considering the one-, two-, three-, and four-step forecasts for each of the six variables, there are 24 comparisons of total uncertainty. Of these the model with a prior has less total uncertainty in 16 cases using the first measure, 21 cases using the second, 19 using the third, and 21 using the fourth. These results may be compared with the RMSE measure in which the model with prior was lower in 18 or 20 of the 24 cases based on zero errors and coefficient simulation forecasts, respectively.

One advantage the total uncertainty measures have over RMSE is that they change from period to period to reflect changes in the uncertainty of the model with respect to current conditions. A comparison of the same two models one period earlier, that is, in 1977-4, for the same four measures of total uncertainty gave the model with prior less uncertainty in 15, 17, 14, and 19 cases, respectively.

The increased advantage of the model with prior in 1978-1 suggests that conditions may have entered an area of the state space for which recent history gave relatively less information.

Appendix
Maximum Likelihood Estimator of Total Uncertainty

Let the k-step forecast error at time t be e_{tk}. Assume that e_{tk} is distributed normally with mean zero and variance σ^2_{tk} which is a function of $\hat{\sigma}^2_{tk}$, the estimated coefficient simulation variance. Then e^2_{tk} has the chi-squared function density

$$e^2_{tk} \sim (2\pi\sigma^2_{tk})^{-1/2}\exp(\frac{-e^2_{tk}}{2\sigma^2_{tk}})e^{-1}_{tk}. \tag{90}$$

Suppose $\hat{\sigma}^2_{tk} = \hat{\sigma}^2_{tk} + \tilde{d}$. Given T observations on e^2_{tk} and $\hat{\sigma}^2_{tk}$, the log likelihood function is

$$\text{Log } L(e^2_{tk},\tilde{d},\hat{\sigma}^2_{tk},t=1,T) = \tag{91}$$

$$-\sum_{t=1}^{T}[\frac{1}{2}(\log 2\pi+\log(\tilde{d}+\hat{\sigma}^2_{tk}))+\frac{e^2_{tk}}{2(\tilde{d}+\hat{\sigma}^2_{tk})}+\log e_{tk}]$$

and

$$\frac{\partial \text{ Log } L}{\partial \tilde{d}} = \sum_{t=1}^{T}[\frac{1}{2(\tilde{d}+\hat{\sigma}^2_{tk})}(\frac{e^2_{tk}}{(\tilde{d}+\hat{\sigma}^2_{tk})}-1)] \tag{92}$$

and

$$\frac{\partial^2 \text{ Log } L}{\partial \tilde{d}^2} = -\sum_{t=1}^{T}[\frac{1}{(\tilde{d}+\hat{\sigma}^2_{tk})^2}(\frac{e^2_{tk}}{(\tilde{d}+\hat{\sigma}^2_{tk})}-\frac{1}{2})]. \tag{93}$$

The estimate of $\tilde{d}$ is obtained by maximizing the log likelihood function. Initial estimates of $\tilde{d}$ were obtained by regressing

$$(\frac{e^2_{tk}}{\hat{\sigma}_{tk}} - \hat{\sigma}_{tk}) \tag{94}$$

on $\frac{1}{\hat{\sigma}_{tk}}$. Maximization was then obtained by iterating with a Davidson-Fletcher-Powell algorithm. The asymptotic variance of the maximum likelihood estimate is given by

$$-\left(\frac{\partial^2 \text{ Log } L}{\partial \tilde{d}^2}\right)^{-1}. \tag{95}$$

Similar procedures were followed for the specifications $\sigma_{tk}^2 = \beta\hat{\sigma}_{tk}^2$ and $\sigma_{tk}^2 = \alpha + \beta\hat{\sigma}_{tk}^2$.

Larger Models

The results to this point have described how prior restrictions on relatively small VAR systems can improve forecasting performances over that of unrestricted models. These techniques also allow estimation and forecasting to proceed in VAR systems larger than it is possible to estimate with unrestricted OLS.

For example, consider expansion of a system with a given number of lags by the addition of new variables. Eventually, the number of coefficients will exceed the number of observations and OLS estimation would be impossible. One specification which would allow estimation to proceed is a prior which uniformly restricts other variables in each equation. The limiting case of this prior is a set of univariate autoregressions.

Starting from the limiting case, it is possible to compare forecasting performances as the prior tightness on other variables in each equation is decreased. While for most variables in an experiment of this kind there is a region of improvement followed by worsening, this is not always the case. One exception, for example, is net exports, for which in the mid-seventies a univariate specification dominates ones with any interaction allowed with other domestic macroeconomic variables.

The results of an experiment of this kind are shown below. The VAR system includes 11 variables and 5 lags. The projection period is 1972-1 through 1978-4. The priors in the second and third specifications, in effect, leave own-lag coefficients unconstrained (standard errors are 100.), while lag coefficients of other variables are given standard errors about zero of .001 and .01, respectively. The fourth specification imposes restrictions on own lags, starting with a standard error of .2 about a mean of 1. for the first lag; and standard errors beginning at .01 for the first lag of other variables. The standard errors become smaller for coefficients with larger lags.

As the size of a VAR system increases, it becomes increasingly implausible that the procedure of treating all variables symmetrically is optimal. When there are 10 or 15 variables in the system, for example, there would seem to be a strong case which could be made that forecasting performance might be improved by the imposition of some structure derived from economic theory.

One could conceive of specifying a prior similar to those considered here in which for an N-variable VAR, an additional N^2 parameters specified relative weights for every variable in every equation. A more tractable approach would seem to be to specify a function which generates those N^2 parameters on the basis of, for example, an ordering of the variables.

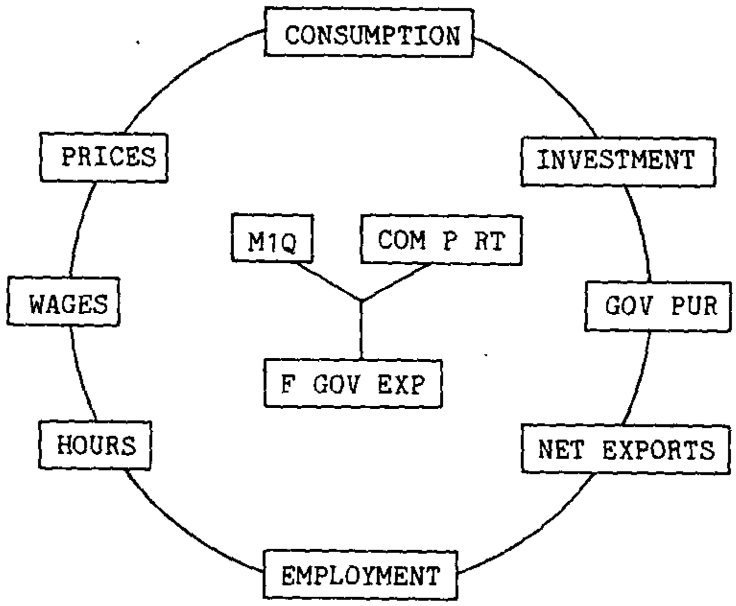
One such function may be represented schematically as a star inside a circle. Variables on the star are those which are assumed to have strong, but equal, impacts on all other variables in the system. Other variables are arranged in a circular ordering with related variables placed close together in the ordering. Relative tightness of the prior on coefficients in an equation are then made a function of the relative positions of the variables.

For example, with N variables in a system, let K of them be specified to be on the star, i.e., in the set S and N-K arranged around the outside, i.e., in the set C, with a given circular ordering. An additional parameter, γ_3, is specified which reduces the standard error of the prior for the coefficients of variable j in equation i by dividing them by δ_{ij}, where

$$\delta_{ij} = \begin{cases} 1 & \text{if } i\varepsilon S,\ j\varepsilon S \\ (N-K)^{\gamma_3} & \text{if } i\varepsilon S,\ j\varepsilon C \\ 2^{\gamma_3} & \text{if } i\varepsilon C,\ j\varepsilon S \\ L_{ij}^{\ \gamma_3} & \text{if } i\varepsilon C,\ j\varepsilon C \end{cases} \tag{96}$$

and L_{ij} is the distance around the circle between variables i and j.

Shown below is a schematic drawing of a circle-star prior for the above-mentioned 11-variable VAR.



In experiments with such priors I have generated some, but usually not large, reduction over symmetrical priors.

The Theil U statistics compiled by McNees [1975] for several large econometric models provide a useful benchmark for comparison with the VAR. For this purpose I constructed a 15-variable VAR which included 10 variables from McNees' study along with 5 other variables (M1, Commercial Paper Rate, Import Price Index, Wages, and Crude Materials Prices). The Theil U statistics for this system are plotted along with the NcNees statistics in the following graphs. The circle-star prior in this case had parameters $\gamma_1 = 1.$, $\gamma_2 = .5$, $\lambda = .1$, $\gamma_3 = 1.$ The variables on the star were M1, Commercial Paper Rate, and Personal Consumption of Nondurables and Services. The circular ordering was Real GNP, Unemployment Rate, Wages, Implicit Price Deflator, Business Fixed Investment, Change in Business Inventories, Personal Consumption of Durable Goods, Federal Government Purchases, Residential Structures, Crude Materials Prices, Import Prices, and Net Exports.[24/]

Theil U Statistics

Univarite Autoregressions

STEPS AHEAD	COM P RT	M1Q	F GOV EXP	CON-SUMPTION	INVEST-MENT	GOV PUR	NET EXPORTS	EMPLOY-MENT	HOURS	WAGES	PRICES	AVE
1	.891	.318	.691	.606	1.022	.987	1.038	.597	.857	.266	.288	.687
2	1.034	.289	.594	.543	1.044	1.072	1.020	.624	.922	.216	.301	.696
3	1.029	.267	.528	.497	1.095	1.098	.991	.640	1.002	.204	.320	.697
4	.975	.282	.498	.513	1.169	1.041	.954	.675	1.041	.193	.358	.701
5	.954	.297	.489	.527	1.209	.984	.921	.697	1.059	.188	.399	.702
6	.962	.306	.487	.521	1.237	.924	.917	.696	1.054	.187	.442	.703
7	.963	.324	.479	.514	1.260	.919	.887	.695	1.079	.191	.482	.708
8	.940	.336	.464	.509	1.309	.882	.878	.699	1.119	.196	.521	.714
AVERAGE	.968	.302	.529	.529	1.168	.989	.951	.665	1.017	.205	.389	.701

Prior: $\gamma_1 = 0$, $\gamma_2 = .00001$, $\lambda = 100$.

STEPS AHEAD	COM P RT	M1Q	F GOV EXP	CON-SUMPTION	INVEST-MENT	GOV PUR	NET EXPORTS	EMPLOY-MENT	HOURS	WAGES	PRICES	AVE
1	.828	.317	.688	.611	1.005	1.014	1.064	.591	.800	.266	.287	.679
2	.924	.288	.589	.553	1.010	1.129	1.094	.616	.831	.216	.300	.686
3	.884	.266	.521	.510	1.040	1.198	1.120	.629	.865	.205	.320	.687
4	.833	.279	.492	.531	1.095	1.186	1.146	.662	.853	.195	.357	.693
5	.818	.294	.485	.552	1.110	1.132	1.162	.686	.829	.189	.397	.696
6	.849	.303	.483	.553	1.111	1.056	1.179	.691	.797	.188	.440	.695
7	.884	.320	.472	.555	1.107	1.047	1.182	.696	.791	.193	.482	.703
8	.892	.331	.455	.560	1.148	1.034	1.198	.713	.805	.198	.522	.714
AVERAGE	.864	.300	.523	.553	1.078	1.099	1.143	.661	.821	.206	.388	.694

Prior: $\gamma_1 = 0$, $\gamma_2 = .0001$, $\lambda = 100$.

STEPS AHEAD	COM P RT	M1Q	F GOV EXP	CON-SUMPTION	INVEST-MENT	GOV PUR	NET EXPORTS	EMPLOY-MENT	HOURS	WAGES	PRICES	AVE
1	.827	.313	.654	.621	.945	1.035	1.085	.556	.779	.259	.274	.668
2	.895	.279	.541	.564	.905	1.179	1.149	.556	.804	.208	.280	.669
3	.827	.251	.477	.526	.915	1.294	1.214	.542	.857	.193	.296	.672
4	.724	.261	.446	.563	.950	1.378	1.284	.551	.874	.182	.321	.685
5	.623	.274	.436	.601	.966	1.470	1.343	.566	.901	.173	.350	.700
6	.563	.281	.423	.621	.965	1.568	1.385	.568	.932	.172	.379	.714
7	.548	.300	.394	.649	.967	1.720	1.425	.572	.991	.176	.406	.740
8	.517	.311	.356	.675	1.020	1.954	1.481	.610	1.108	.181	.427	.785
AVERAGE	.690	.284	.466	.602	.954	1.450	1.296	.565	.906	.193	.342	.704

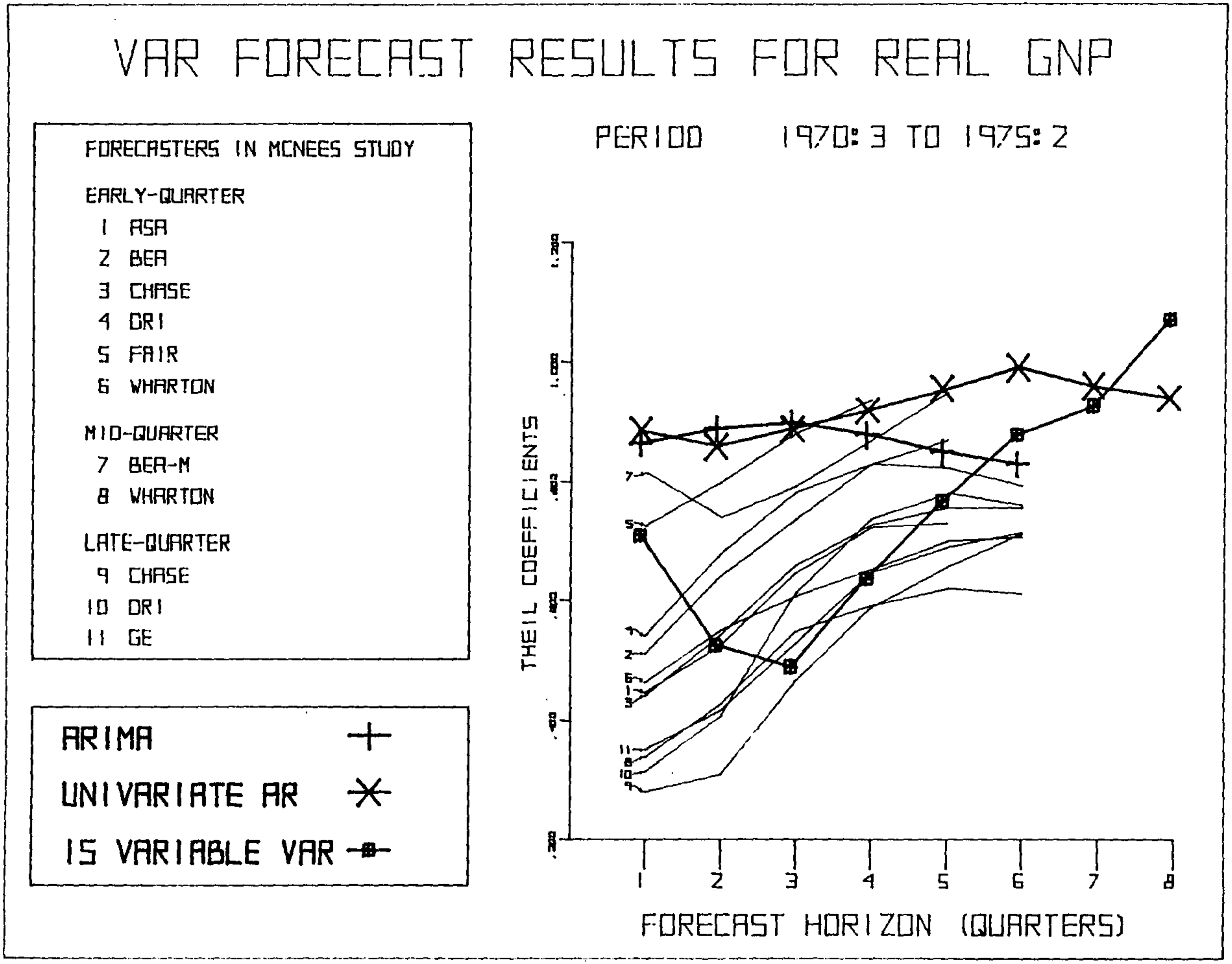
Prior: $\gamma_1 = 1.$, $\gamma_2 = .05$, $\lambda = .2$.

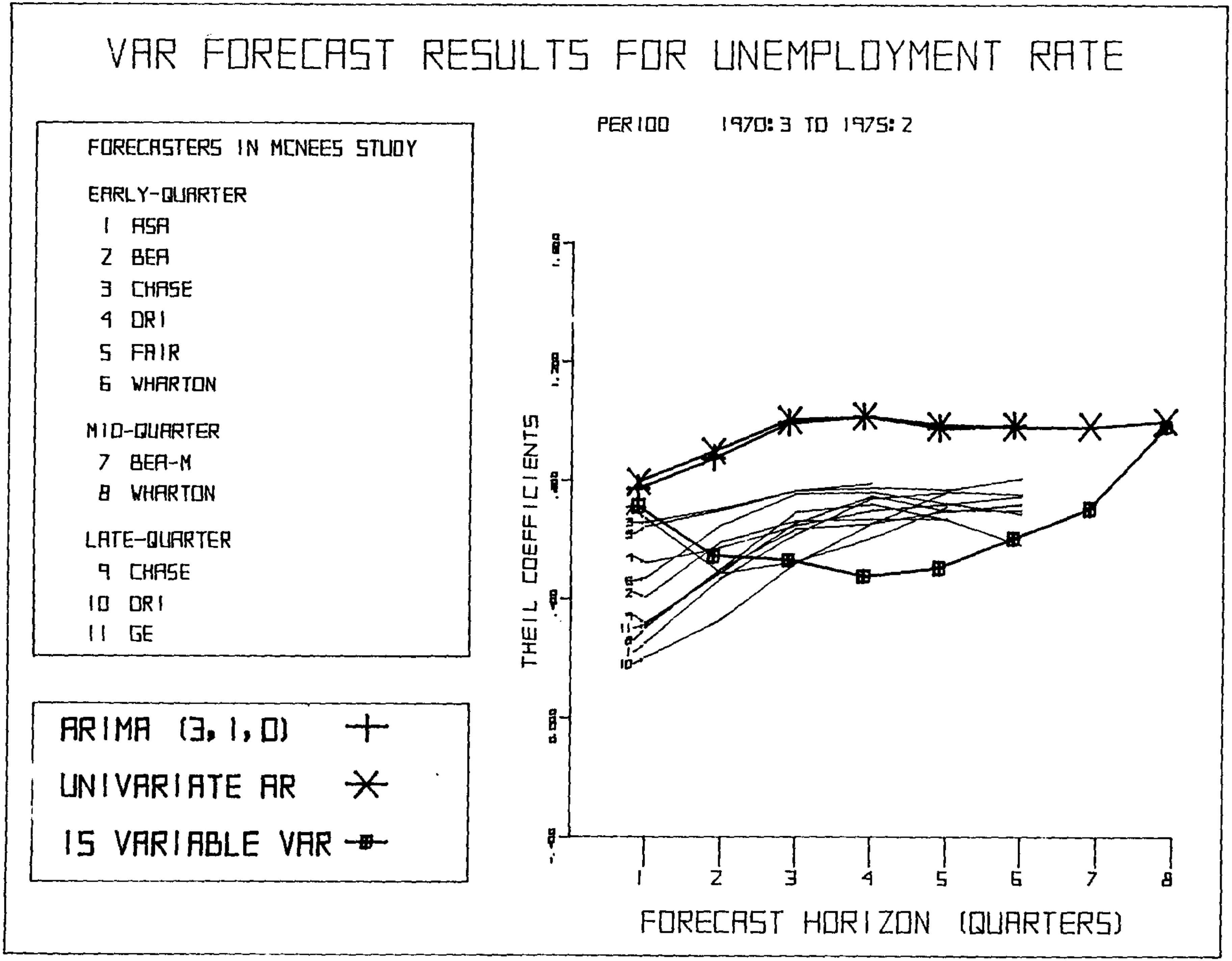
STEPS AHEAD	COM P RT	M1Q	F GOV EXP	CON-SUMPTION	INVEST-MENT	GOV PUR	NET EXPORTS	EMPLOY-MENT	HOURS	WAGES	PRICES	AVE
1	.863	.296	.623	.615	.913	1.034	1.053	.559	.779	.252	.272	.660
2	.912	.259	.520	.567	.910	1.142	1.108	.569	.796	.206	.283	.661
3	.837	.237	.456	.532	.924	1.201	1.141	.573	.833	.194	.305	.657
4	.709	.250	.428	.571	.965	1.197	1.171	.597	.841	.184	.333	.659
5	.670	.262	.419	.619	.999	1.175	1.199	.621	.854	.176	.364	.669
6	.672	.274	.410	.650	1.014	1.133	1.231	.634	.873	.175	.396	.678
7	.637	.295	.394	.686	1.036	1.136	1.261	.652	.918	.180	.427	.693
8	.606	.309	.373	.723	1.105	1.168	1.309	.700	1.012	.186	.452	.722
AVERAGE	.738	.273	.453	.620	.983	1.148	1.184	.613	.863	.194	.354	.675

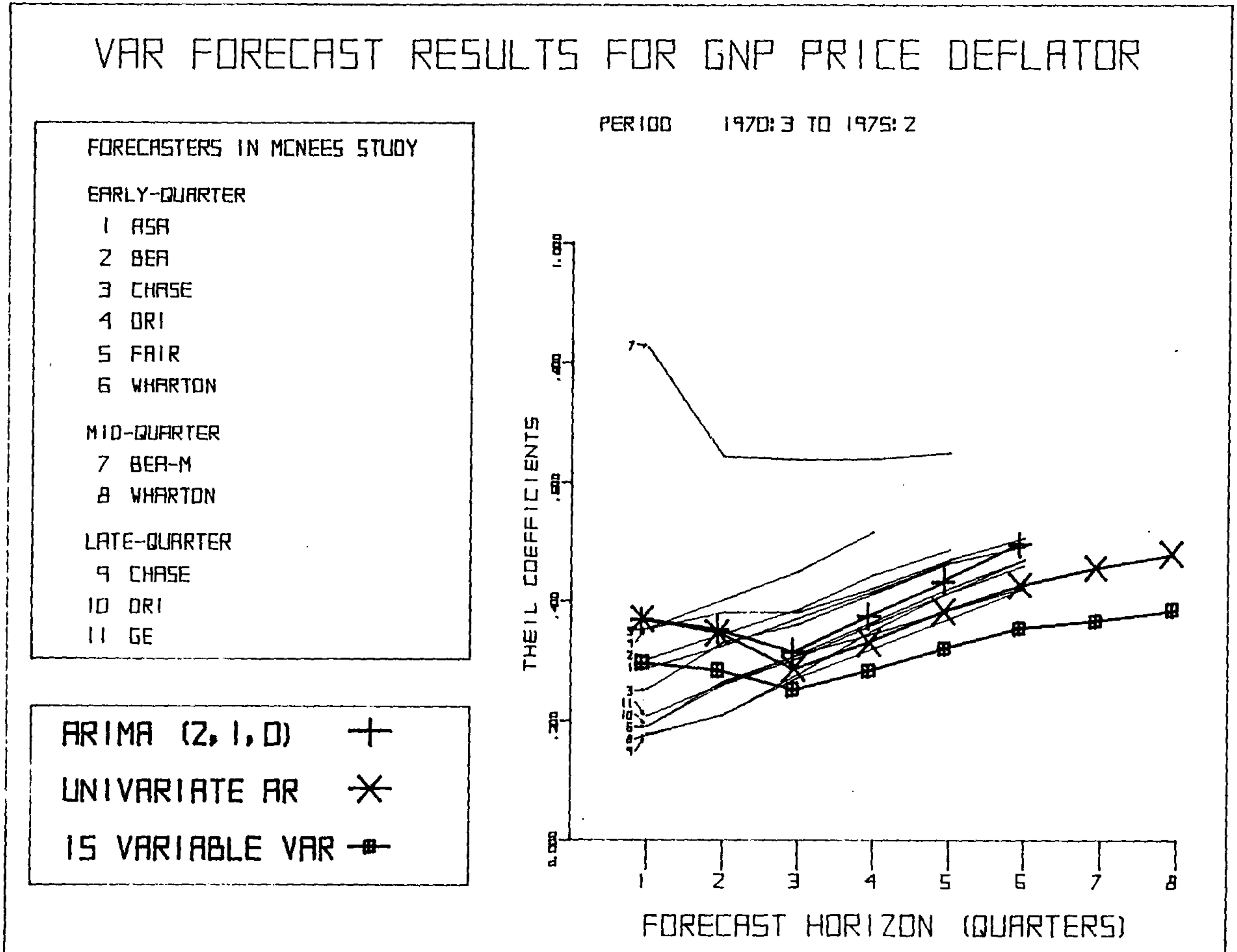
Theil U Statistics for Circle-Star Prior

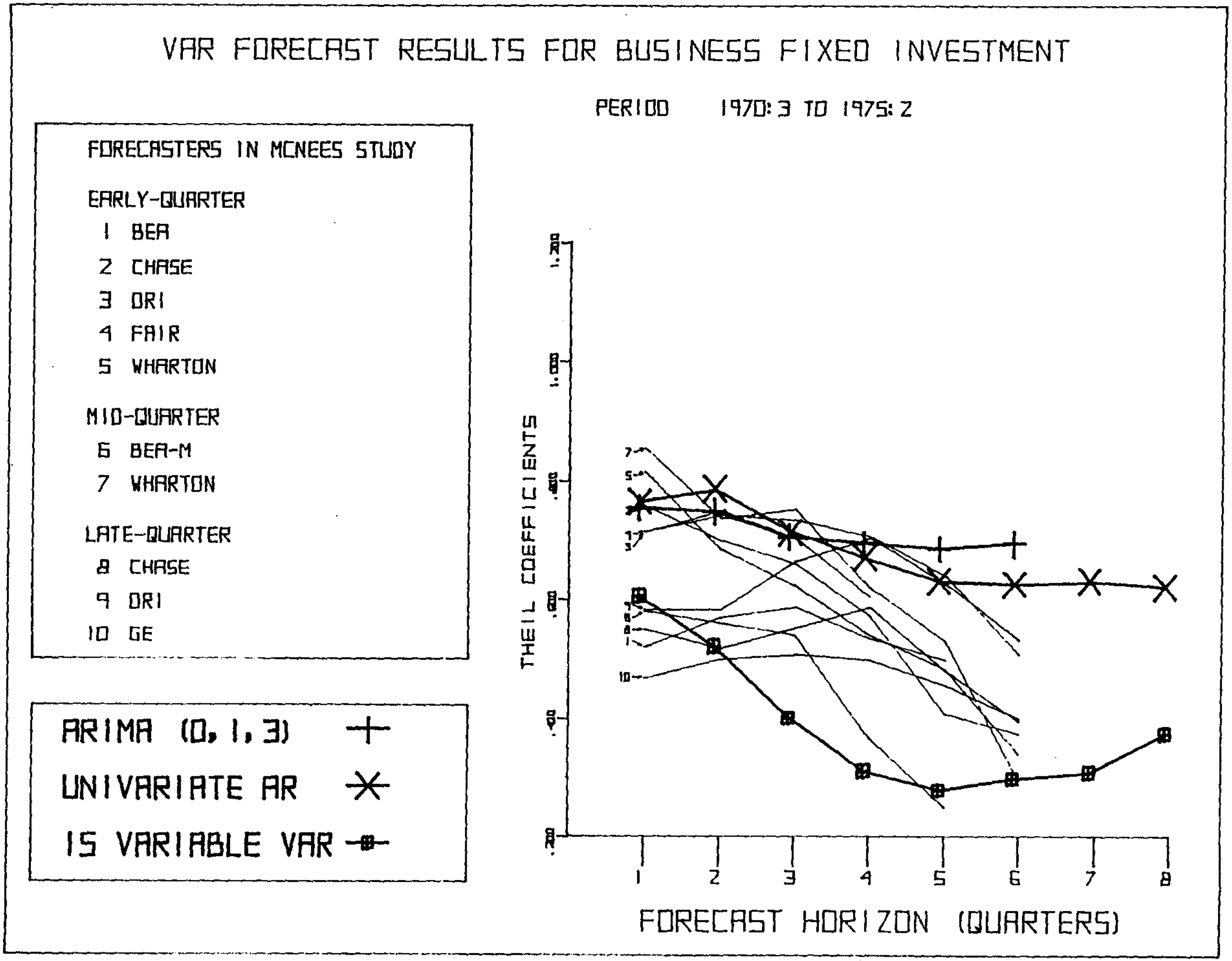
Prior: $\gamma_1 = .5$, $\gamma_2 = .1$, $\lambda = .2$, $\gamma_3 = 1.$

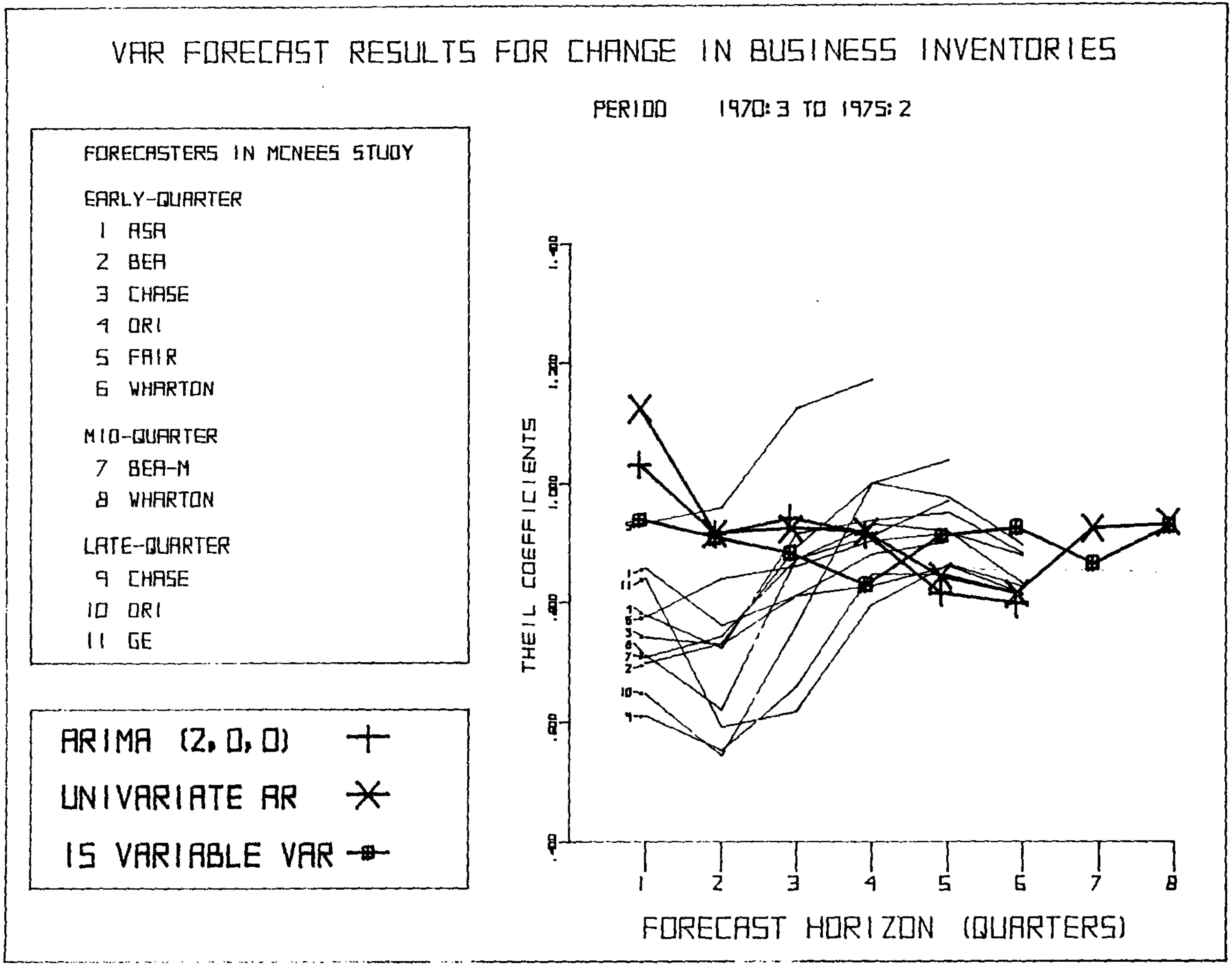
STEPS AHEAD	COM P RT	M1Q	F GOV EXP	CON-SUMPTION	INVEST-MENT	GOV PUR	NET EXPORTS	EMPLOY-MENT	HOURS	WAGES	PRICES	AVE
1	.797	.297	.651	.616	.921	1.040	1.051	.583	.777	.252	.275	.660
2	.843	.263	.555	.565	.903	1.159	1.100	.601	.789	.204	.285	.660
3	.774	.245	.485	.530	.908	1.221	1.140	.609	.814	.189	.306	.656
4	.660	.258	.451	.570	.931	1.209	1.177	.638	.801	.177	.335	.655
5	.601	.269	.438	.614	.939	1.182	1.210	.662	.791	.168	.368	.658
6	.597	.277	.422	.643	.925	1.125	1.240	.674	.787	.166	.402	.660
7	.598	.293	.390	.678	.919	1.122	1.264	.689	.812	.169	.435	.670
8	.592	.301	.347	.713	.961	1.124	1.302	.733	.879	.176	.461	.690
AVERAGE	.683	.275	.467	.616	.926	1.148	1.185	.649	.806	.188	.358	.664

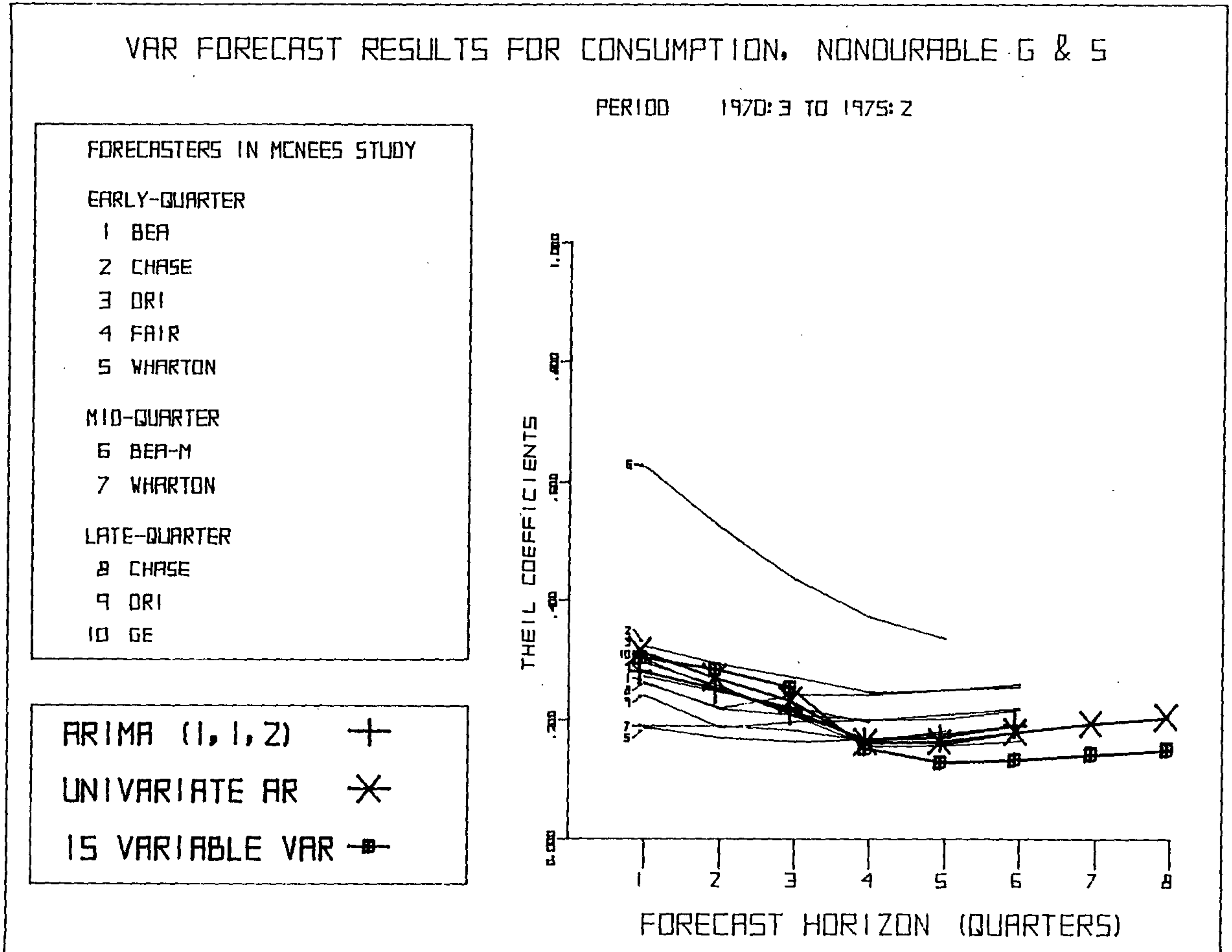

VAR FORECAST RESULTS FOR REAL GNP
FORECASTERS IN MCNEES STUDY
EARLY-QUARTER
1 ASA
2 BEA
3 CHASE
4 DRI
5 FAIR
6 WHARTON
MID-QUARTER
7 BEA-M
8 WHARTON
LATE-QUARTER
9 CHASE
10 DRI
11 GE
ARIMA
UNIVARIATE AR
15 VARIABLE VAR
PERIOD 1970:3 TO 1975:2
THEIL COEFFICIENTS
1 2 3 4 5 6 7 8
FORECAST HORIZON (QUARTERS)


VAR FORECAST RESULTS FOR UNEMPLOYMENT RATE
PERIOD 1970:3 TO 1975:2
FORECASTERS IN MCNEES STUDY
EARLY-QUARTER
1 ASA
2 BEA
3 CHASE
4 DRI
5 FAIR
6 WHARTON
MID-QUARTER
7 BEA-M
8 WHARTON
LATE-QUARTER
9 CHASE
10 DRI
11 GE
ARIMA (3,1,0)
UNIVARIATE AR
15 VARIABLE VAR
THEIL COEFFICIENTS
1 2 3 4 5 6 7 8
FORECAST HORIZON (QUARTERS)

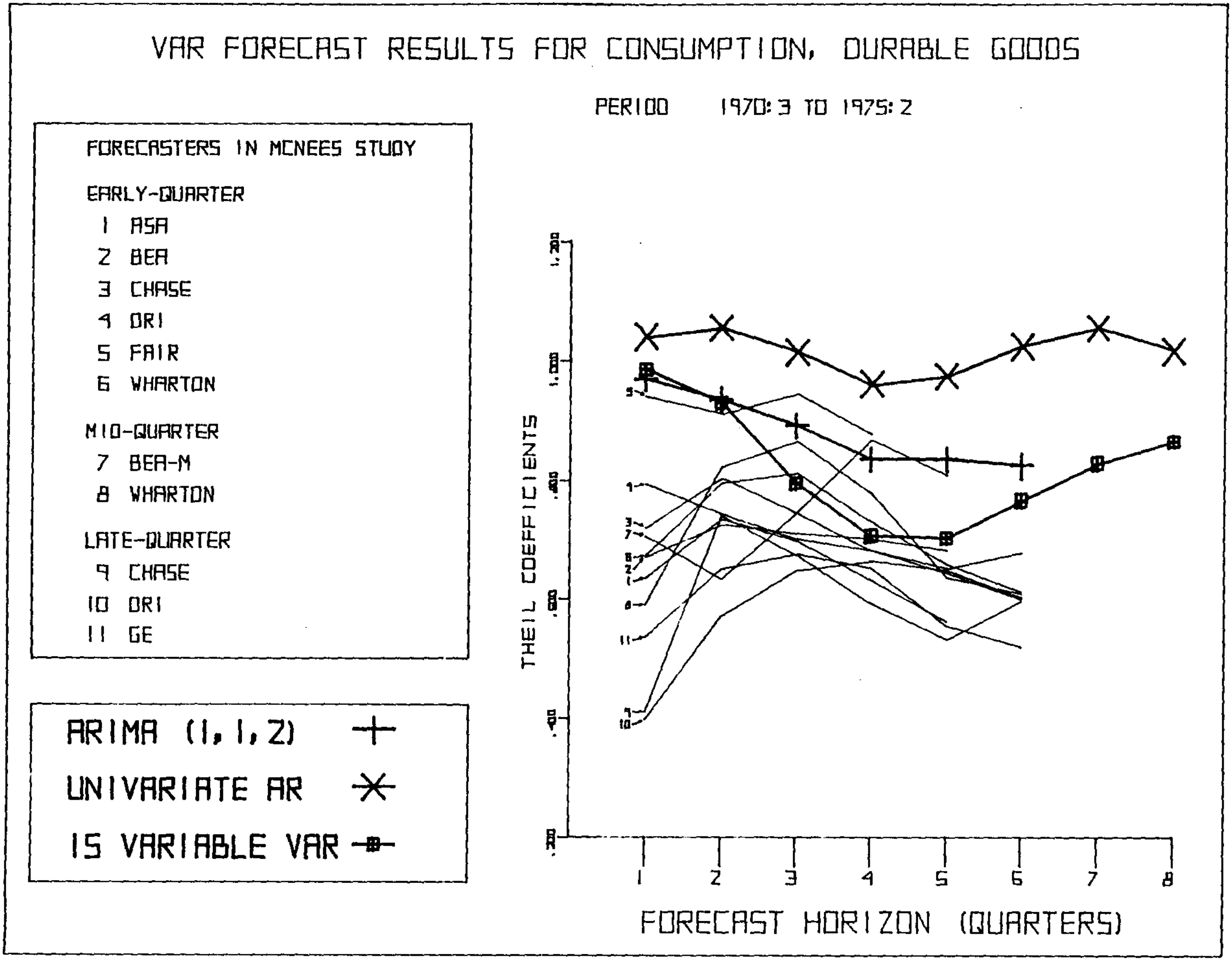

VAR FORECAST RESULTS FOR GNP PRICE DEFLATOR
FORECASTERS IN MCNEES STUDY
EARLY-QUARTER
1 ASA
2 BEA
3 CHASE
4 DRI
5 FAIR
6 WHARTON
MID-QUARTER
7 BEA-M
8 WHARTON
LATE-QUARTER
9 CHASE
10 DRI
11 GE
ARIMA (2,1,0)
UNIVARIATE AR
15 VARIABLE VAR
PERIOD 1970:3 TO 1975:2
THEIL COEFFICIENTS
1
2
3
4
5
6
7
8
FORECAST HORIZON (QUARTERS)

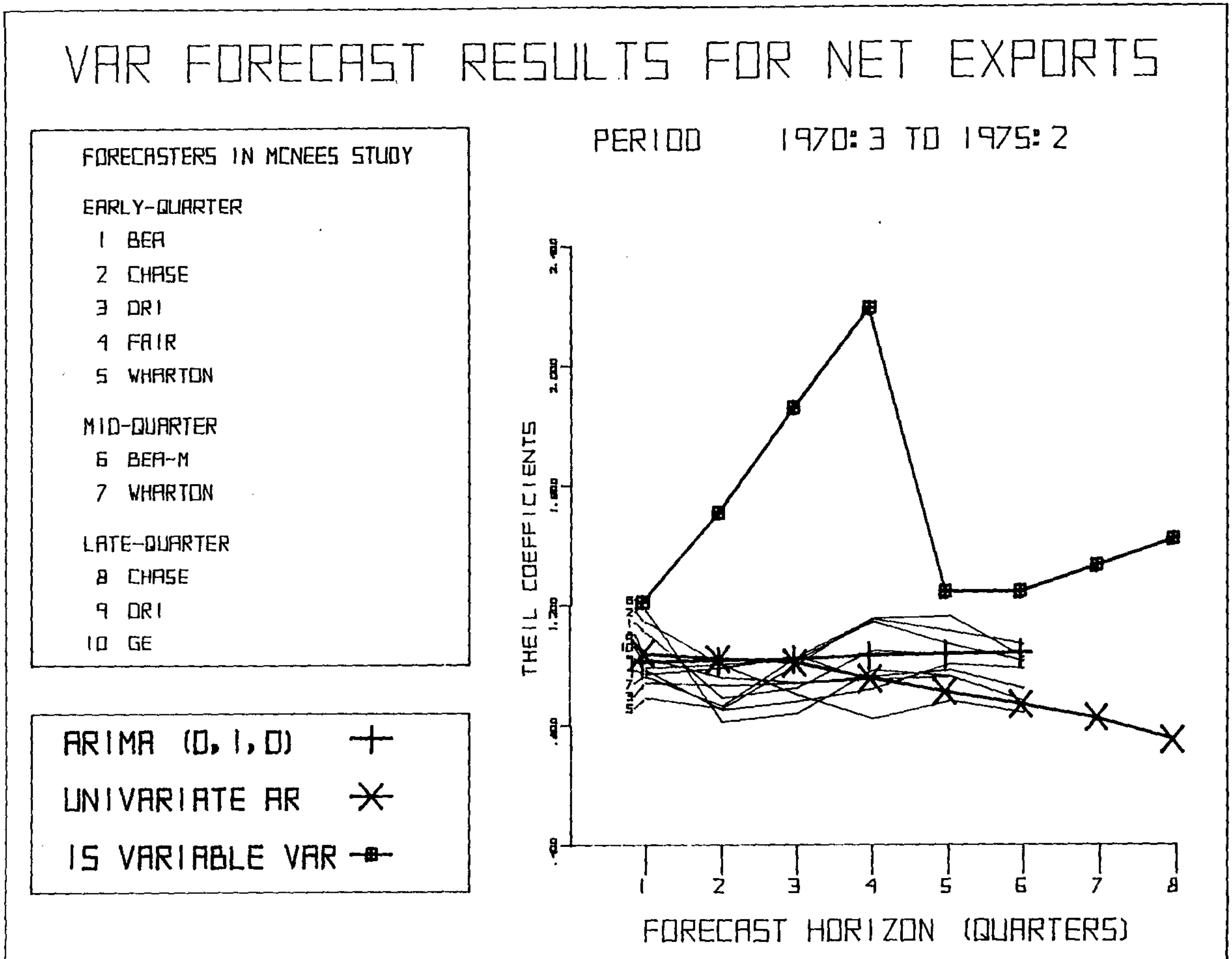

VAR FORECAST RESULTS FOR BUSINESS FIXED INVESTMENT
PERIOD 1970:3 TO 1975:2
FORECASTERS IN MCNEES STUDY
EARLY-QUARTER
1 BEA
2 CHASE
3 DRI
4 FAIR
5 WHARTON
MID-QUARTER
6 BEA-M
7 WHARTON
LATE-QUARTER
8 CHASE
9 DRI
10 GE
ARIMA (0,1,3)
UNIVARIATE AR
15 VARIABLE VAR
THEIL COEFFICIENTS
1.200
1.000
1
2
3
4
5
6
7
8
FORECAST HORIZON (QUARTERS)


VAR FORECAST RESULTS FOR CHANGE IN BUSINESS INVENTORIES
PERIOD 1970:3 TO 1975:2
FORECASTERS IN MCNEES STUDY
EARLY-QUARTER
1 ASA
2 BEA
3 CHASE
4 DRI
5 FAIR
6 WHARTON
MID-QUARTER
7 BEA-M
8 WHARTON
LATE-QUARTER
9 CHASE
10 DRI
11 GE
ARIMA (2, 0, 0)
UNIVARIATE AR
15 VARIABLE VAR
THEIL COEFFICIENTS
1
2
3
4
5
6
7
8
FORECAST HORIZON (QUARTERS)

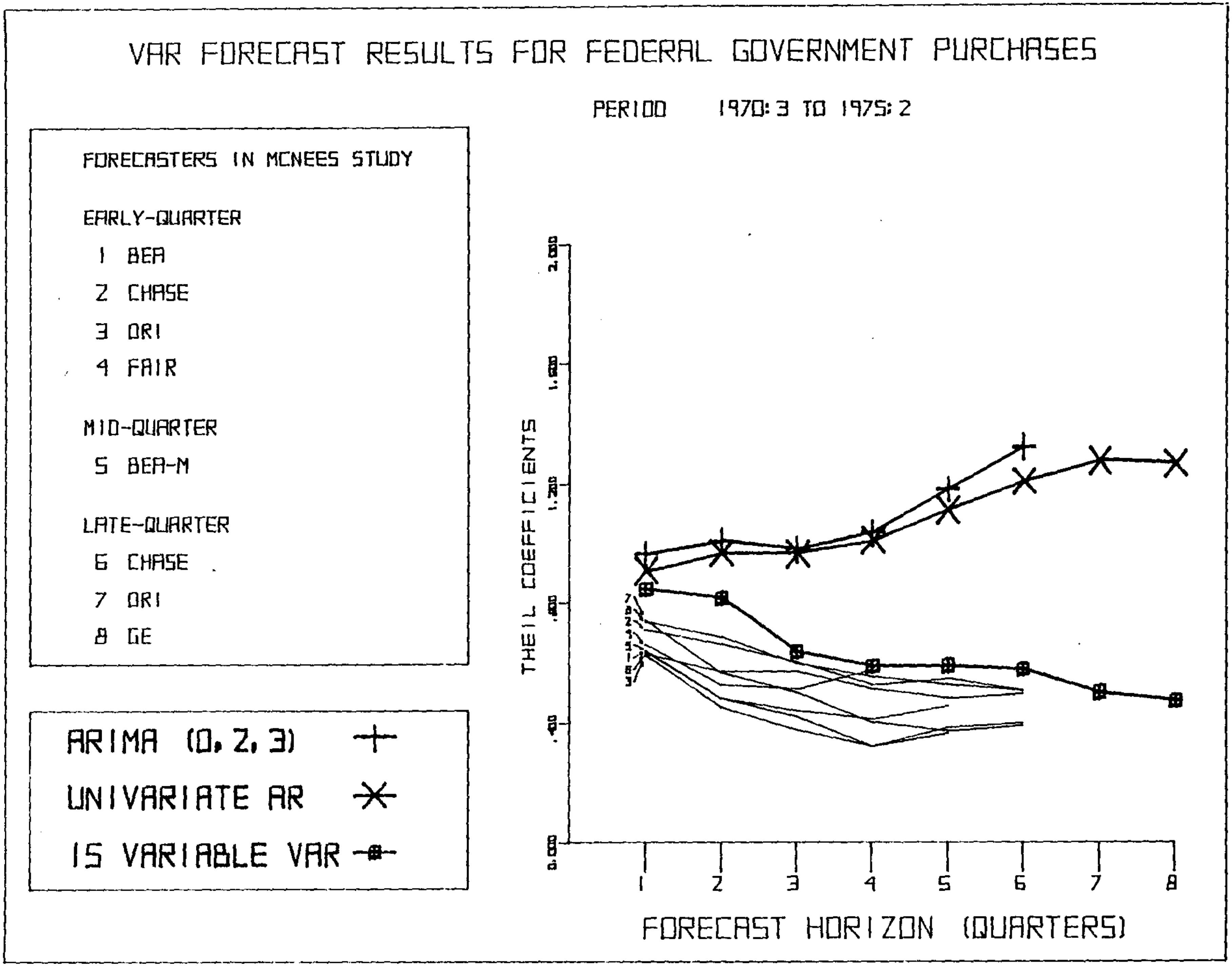


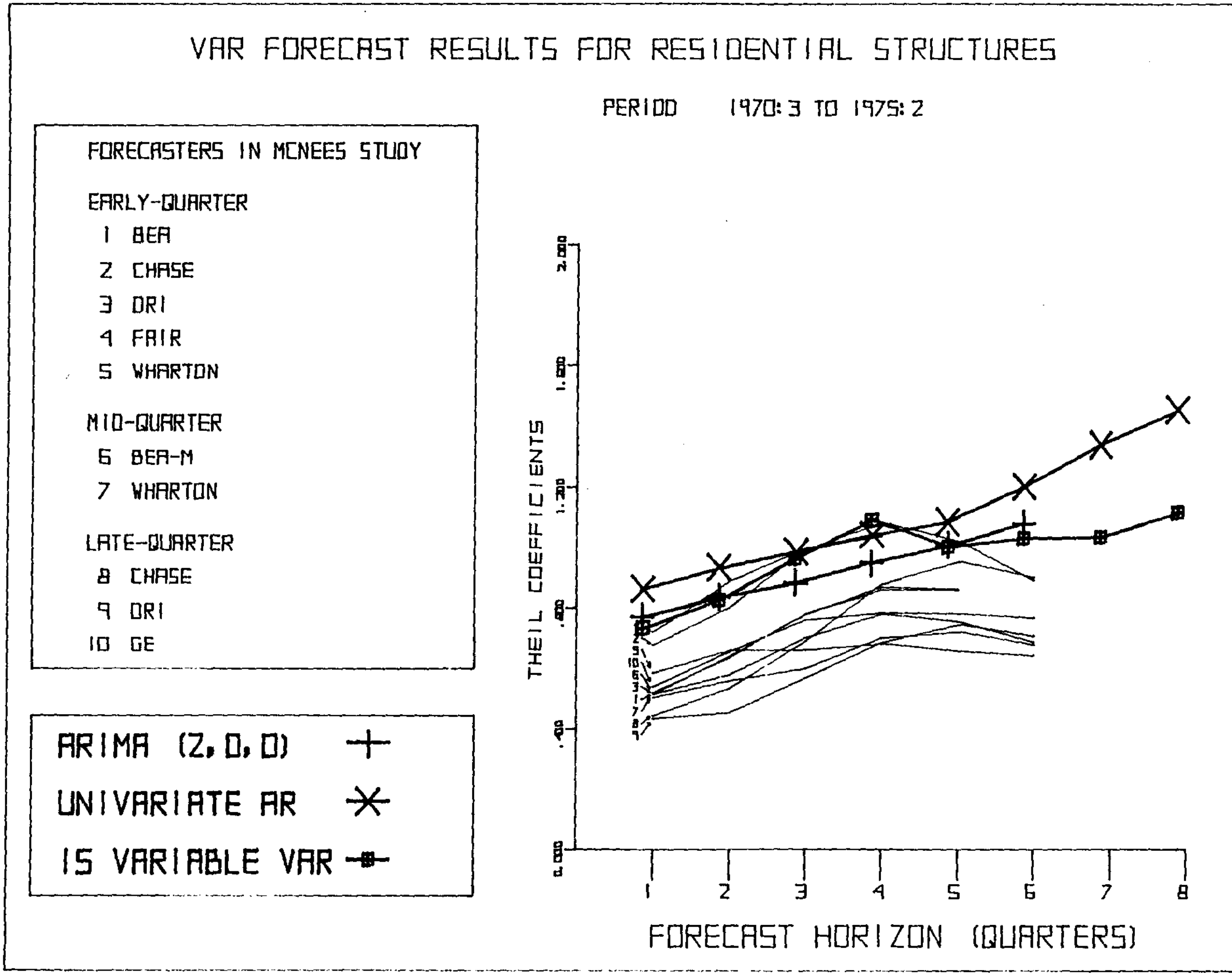
VAR FORECAST RESULTS FOR CONSUMPTION, NONDURABLE G & S
PERIOD 1970:3 TO 1975:2
FORECASTERS IN MCNEES STUDY
EARLY-QUARTER
1 BEA
2 CHASE
3 DRI
4 FAIR
5 WHARTON
MID-QUARTER
6 BEA-M
7 WHARTON
LATE-QUARTER
8 CHASE
9 DRI
10 GE
ARIMA (1,1,2)
UNIVARIATE AR
15 VARIABLE VAR
THEIL COEFFICIENTS
FORECAST HORIZON (QUARTERS)
1
2
3
4
5
6
7
8


VAR FORECAST RESULTS FOR CONSUMPTION, DURABLE GOODS
PERIOD 1970:3 TO 1975:2
FORECASTERS IN MCNEES STUDY
EARLY-QUARTER
1 ASA
2 BEA
3 CHASE
4 DRI
5 FAIR
6 WHARTON
MID-QUARTER
7 BEA-M
8 WHARTON
LATE-QUARTER
9 CHASE
10 DRI
11 GE
ARIMA (1,1,2)
UNIVARIATE AR
15 VARIABLE VAR
THEIL COEFFICIENTS
1 2 3 4 5 6 7 8
FORECAST HORIZON (QUARTERS)



VAR FORECAST RESULTS FOR NET EXPORTS
PERIOD 1970:3 TO 1975:2
FORECASTERS IN MCNEES STUDY
EARLY-QUARTER
1 BEA
2 CHASE
3 DRI
4 FAIR
5 WHARTON
MID-QUARTER
6 BEA-M
7 WHARTON
LATE-QUARTER
8 CHASE
9 DRI
10 GE
ARIMA (0,1,0)
UNIVARIATE AR
IS VARIABLE VAR
THEIL COEFFICIENTS
FORECAST HORIZON (QUARTERS)
1
2
3
4
5
6
7
8


VAR FORECAST RESULTS FOR FEDERAL GOVERNMENT PURCHASES
PERIOD 1970:3 TO 1975:2
FORECASTERS IN MCNEES STUDY
EARLY-QUARTER
1 BEA
2 CHASE
3 DRI
4 FAIR
MID-QUARTER
5 BEA-M
LATE-QUARTER
6 CHASE
7 DRI
8 GE
ARIMA (0, 2, 3)
UNIVARIATE AR
15 VARIABLE VAR
THEIL COEFFICIENTS
1 2 3 4 5 6 7 8
FORECAST HORIZON (QUARTERS)


VAR FORECAST RESULTS FOR RESIDENTIAL STRUCTURES
PERIOD 1970:3 TO 1975:2
FORECASTERS IN MCNEES STUDY
EARLY-QUARTER
1 BEA
2 CHASE
3 DRI
4 FAIR
5 WHARTON
MID-QUARTER
6 BEA-M
7 WHARTON
LATE-QUARTER
8 CHASE
9 DRI
10 GE
ARIMA (2,0,0)
UNIVARIATE AR
15 VARIABLE VAR
THEIL COEFFICIENTS
FORECAST HORIZON (QUARTERS)
1
2
3
4
5
6
7
8

CHAPTER IV
USES OF VECTOR AUTOREGRESSIONS

This chapter considers several of the types of analyses which are possible given an estimated vector autoregressive representation. The first section illustrates how stochastic simulation of an estimated model can generate the probability of any event, such as the onset of a recession, which can be defined in terms of the properties of a sample path. The next section defines the impulse response function, the dynamic response of the system to an innovation in any one of its components. The third section demonstrates how forecast variance can be partitioned among a set of orthogonal error processes defined by a particular ordering of the variables in the system. Finally, an example is given in which several of the techniques developed in this paper are applied.

Sample Paths and Histograms

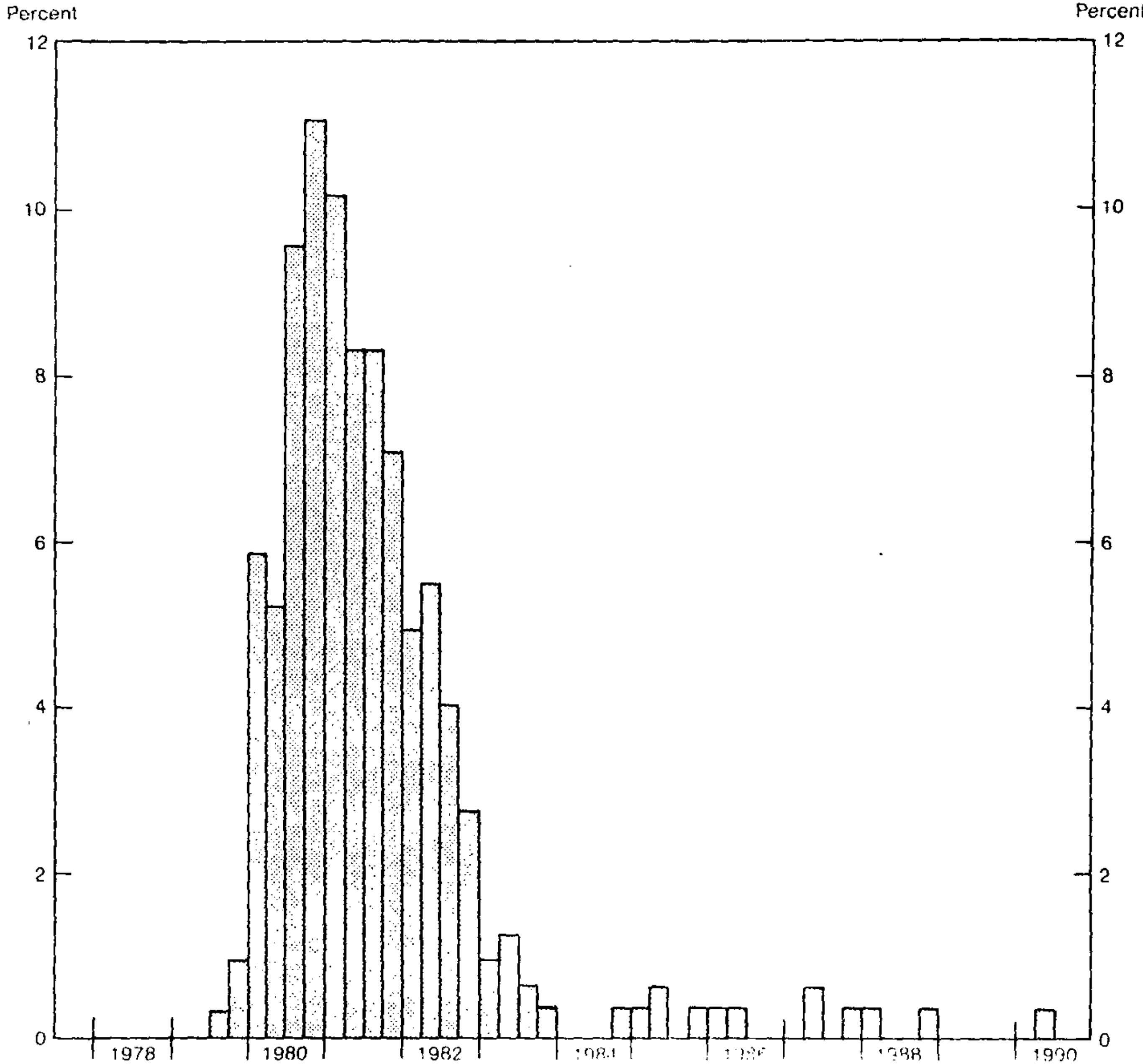
Policymakers often want answers to questions such as "What is the likelihood of a recession within the next year?" Unfortunately, such questions do not lend themselves to convenient analytic solutions, and standard forecasts will generally be of little help. For example, even though it is quite conceivable that a projection of GNP would increase monotonically, such a forecast is not necessarily incompatible with projecting that a recession that year is quite likely to occur.

The explanation of this paradox is that the forecast is an expected value or mean of the distribution projected by the model. A constant forecast, therefore, does not imply that the value of the series will remain constant, but rather that its mean remains constant. The future path of the variables which is actually realized is expected to resemble not the projection itself, but a random sample from the distribution which underlies the projection. Such a sample will generally have many random movements up and down.

One implication of the above discussion is that even given a projection of GNP, it is still difficult to forecast the likelihood of a recession. For example, if we define a recession as three consecutive quarters of decreases in real GNP, then the probability of such an occurrence is a complicated function of the probability distribution of future sample paths. Wecher [1979] has suggested a method for answering such questions. The technique is to build a monte carlo simulator into the prediction program. The estimation procedure generates not only estimated coefficients of the system, but also a distribution from which random shocks are drawn. Wecher's suggestion is to take the current state of the system and repeatedly generate sample paths by feeding in random shocks drawn from the distribution of residuals which has been observed in the past. Then statistics can be generated concerning the relative frequency of any events about which policymakers are concerned. Thus, for example, we can record the percentage of sample paths which include a recession within the next four periods, or generate a histogram of the probable next occurrence of a drop in interest rates. This method is quite general and can be used to answer probabalistic questions about any function of future sample paths. An example is shown in Figure 7.

Figure 7

Histograms, such as the one shown below, are generated by the prediction program using the monte carlo method. This example presents the likelihood of the first occurrence of a downturn (defined as the third quarter of three consecutive decreases) in real GNP. This particular histogram is based on an autoregressive system with GNP, money growth, inflation, the three-month Treasury bill rate, and the unemployment rate.



Impulse Response Matrices

Once an autoregressive representation has been estimated, we can derive the response of the system to an innovation in any of its components. This impulse response function is simply the moving-average representation of the system given in (2). It is a sequence of matrices, M_0, M_1, M_2, M_3, ..., with $M_0 = I$, such that

$$Y(t) = M_0\varepsilon(t) + M_1\varepsilon(t-1) + \dots = M(L)\varepsilon(t) \qquad (97)$$

where the ε's are the innovations in the Y process. The response of the i^{th} variable to a unit innovation in the j^{th} variable K periods earlier is given by the ij^{th} element of M_K.

The relationship between M(L) and B(L) is given in (7). Given the autoregressive representation of a system, B(L), there exists a simple algorithm for generating the sequence of M_i's. If Y(1), ..., Y(P) are set to zero vector in the p^{th}-order autoregressive system, and the innovations are given by $\varepsilon'(0) = [0,0,\dots,1,0,\dots,0]$, where the 1 is in the j^{th} column, and $\varepsilon(1) = \varepsilon(2) = \dots = 0$, then the output vector

$$Y(P+k) = B(L)Y(P+k-1) + \varepsilon(k) \qquad k=0,1,\dots \qquad (98)$$

gives the j^{th} column of the K^{th} impulse response matrix.

Unit innovations may be difficult to interpret, for example, when the standard error, σ_i, of ε_i is very small. For this reason, one may wish to exploit the linearity of the system and generate a scaled version of the impulse response function which gives the response to innovations one standard error in size. In this case one generates the sequence of matrices $\tilde{M}_0$, $\tilde{M}_1$, $\tilde{M}_2$, $\tilde{M}_3$, ... with $\tilde{M}_0 = \text{diag}[\sigma_i]$ and $\tilde{M}_i = M_i\tilde{M}_0$.

Some caution must be exercised in interpreting the impulse response matrices. First it must be recognized that one does not in general expect innovations to different variables to occur independently. The correlation matrix of the residuals will indicate which innovations are likely to occur together. Because this model is linear, the response to a combination of innovations is just the sum of their separate contributions.

Another caution concerns the implicit assumption that the structure of the system does not change. This requires careful consideration when impulse response characteristics may enter into policymaking. For example, if monetary policy generates a money supply process which can be represented as a linear function of lagged variables in a particular system, then inclusion of a money variable in the specification will lead to an estimate of the reduced-form money supply rule. Such a system should correctly forecast the response to a random innovation in the money supply, given that the money supply rule remains unchanged.

One cannot, however, successfully attempt to set some variables such as GNP at a certain level by implementing a new rule defined by the procedure of estimating a VAR system and giving a shock to the money supply of a particular size based on the projected impulse response. The act of setting the money supply according to the impulse response characteristics of the system is itself a change in the money supply rule. The structure of the system is changed. Because expectations change across regimes, the response to a random innovation in the money supply will not be the same as the response to a money supply of the same magnitude which has been set according to a new rule.

Orthogonal Decomposition of Variance

It is sometimes of interest to partition the variance of the forecast error into the proportions attributable to innovations in each variable in the system.

If the contemporaneous innovations in different variables were orthogonal, the decomposition of the variance of forecasts would be straightforward. Thus, define the (nx1) vectors $L_i = (0,0,\ldots,0,1,0,\ldots,0)$, where the one is in the i^{th} position. Then one can write the K-step forecast of variable i as

$$E_t Y_i(t+K) = Y_i(t+K) - L_i\varepsilon(t+K) - L_i M_1\varepsilon(t+K-1) - \ldots - L_i M_{K-1}\varepsilon(t+1) \tag{99}$$

where the M_i's are the moving-average matrices and $M_0 = I$ is assumed.

The variance of the K-step forecast of the i^{th} variable

$$E\{[Y_i(t+K)-E_t Y_i(t+K)]^2\} \tag{100}$$

would be

$$[\sigma_i^2+\sum_{k=1}^{K-1}\sum_j M_k^2(ij)\sigma_j^2] \tag{101}$$

since all cross products would be zero by assumption. Then the percentage of variance in the K-step-ahead forecast of variable i due to innovations in variable J would be given by

$$\frac{100\sum_{k=0}^{K-1} M_k^2(iJ)\sigma_J^2}{\sum_{k=0}^{K-1}\sum_j M_k^2(ij)\sigma_j^2}. \tag{102}$$

In general, however, cross covariances are not zero and some normalization must be made before the forecast variance can be partitioned. The method suggested by Sims [forthcoming] is to orthogonalize the errors according to a given ordering of the variables. Different orderings will lead to different decompositions.

Suppose a three-variable system has the error covariance matrix

$$E(\varepsilon_t \varepsilon_t') = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} = \Omega \tag{103}$$

and we take the ordering as 1, 2, 3.

Then we define a new set of errors, u_{it}, i=1, 2, 3 which span the same space as the ε_{it}'s as follows:

Let $u_{1t} = \varepsilon_{1t}$.

Let u_{2t} be that part of ε_{2t} which is orthogonal to ε_{1t}. That is, let u_{2t} be defined by

$$\varepsilon_{2t} = P_1 \varepsilon_{1t} + u_{2t} \tag{104}$$

$$E[u_{2t} \cdot \varepsilon_{1t}] = 0$$

then

$$P_1 = \frac{E[\varepsilon_{2t} \cdot \varepsilon_{1t}]}{E[\varepsilon_{1t}^2]} = \frac{\sigma_{21}}{\sigma_{11}}. \tag{105}$$

Let u_{3t} be that part of ε_{3t} which is orthogonal to ε_{2t} and ε_{1t}. That is, let u_{3t} be defined by

$$\varepsilon_{3t} = P_2 \varepsilon_{1t} + P_3 \varepsilon_{2t} + u_{3t} \tag{106}$$

$$E[u_{3t} \varepsilon_{it}] = 0, \ i=1, 2.$$

Then

$$\begin{bmatrix} P_2 \\ P_3 \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{13} \\ \sigma_{23} \end{bmatrix}. \tag{107}$$

In this three-variate system we have

$$\begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -P_1 & 1 & 0 \\ -P_2 & -P_3 & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{bmatrix}. \tag{108}$$

In general we will have defined

$$u_t = G\varepsilon_t \tag{109}$$

where G depends on the particular ordering of the variables. Next we form $E[u_t u_t'] = E[G\varepsilon_t \varepsilon_t' G'] = G\Omega G' = \phi$. By construction ϕ is diagonal. Its diagonal elements give the variances of the orthogonal components of the residuals. Rearranging (99), we have

$$Y_{t+k} - E_t Y_{t+k} = \varepsilon_{t+k} + M_1 \varepsilon_{t+k-1} + \dots + M_{k-1}\varepsilon_{t+1}. \tag{110}$$

Substituting for the ε's leads to

$$Y_{t+k} - E_t Y_{t+k} = IG^{-1}u_{t+k} + M_1 G^{-1} u_{t+k-1} + \dots + M_{k-1} G^{-1} u_{t+1}. \tag{111}$$

The variance matrix of the K-step forecast is given by

$$E[(Y_{t+K}-E_t Y_{t+K})(Y_{t+K}-E_t Y_{t+K})'] \tag{112}$$

$$= G^{-1}\phi G^{-1'} + M_1 G^{-1}\phi G^{-1'} M_1' + \dots + M_{K-1} G^{-1}\phi G^{-1'} M_{K-1}'$$

$$= \Omega + M_1 \Omega M_1' + \dots + M_{K-1}\Omega M_{K-1}'$$

which does not depend on the orthogonalization order. Define $H_k = M_k G^{-1}$ and let $h_k(ij)$ be the ij^{th} element of H_k. Then notice that the K-step forecast variance of the i^{th} variable is given by

$$\sum_{k=0}^{K-1} \sum_j h_k(ij)^2 \phi_{jj}. \tag{113}$$

Thus, the percentage of K-step forecast variance in variable i accounted for by the component of innovations in variable J orthogonal to innovations in variables 1, ... J-1, is given by

$$\frac{100 \sum_{k=0}^{K-1} h_k(iJ)^2 \phi_{JJ}}{\sum_{k=0}^{K-1} \sum_j h_k(ij)^2 \phi_{jj}}. \tag{114}$$

An example of the orthogonal decomposition of variance for the system described on page 38 is shown here.

Orthogonalization Order Is:

Real GNP Money Prices

Variance of Orthogonal Components:

22.46848 .38800 .14088

	K-Step Forecast Variance	Percent caused by shocks in: Real GNP	Money	Prices
Real GNP:				
K= 1.	22.4684798	100.0000	0	0
K= 2.	55.7682212	93.9750	3.0165	3.0085
K= 3.	105.7702653	84.2433	8.1983	7.5585
K= 4.	186.9537991	74.1574	13.4568	12.3858
K= 5.	277.9306256	67.1475	15.9818	16.8707
K= 6.	374.3044668	60.9871	15.8518	23.1611
K= 7.	483.3193992	57.5381	14.3844	28.0776
K= 8.	574.0624279	56.3136	12.9117	30.7748
K= 20.	884.7682757	63.2161	10.1328	26.6511
K= 35.	1241.4433123	59.3997	11.9453	28.6550
K=100.	3655.9592969	43.3026	31.3803	25.3171
Money:				
K= 1.	.4175643	7.0797	92.9203	0
K= 2.	1.7504409	16.0334	82.0186	1.9481
K= 3.	3.9124286	17.9789	76.2685	5.7526
K= 4.	6.5453000	17.9734	72.6081	9.4185
K= 5.	9.1612671	19.4133	68.9463	11.6405
K= 6.	11.6525446	19.4017	67.3673	13.2309
K= 7.	14.2594038	18.9300	66.9913	14.0788
K= 8.	16.9587293	20.0646	65.5996	14.3358
K= 20.	62.1298744	27.4154	52.4878	20.0968
K= 35.	175.6838151	32.4665	44.4267	23.1068
K=100.	1698.8725055	52.7956	28.9040	18.3004

Prices				
K= 1.	.1647344	10.8893	3.5931	85.5176
K= 2.	.5349815	6.5199	1.2124	92.2677
K= 3.	1.0591736	5.3365	.7174	93.9461
K= 4.	1.7272635	5.3761	1.1559	93.4679
K= 5.	2.3723499	4.2243	1.8510	93.9247
K= 6.	2.9587831	3.7101	2.7537	93.5362
K= 7.	3.3583191	3.4035	3.6166	92.9799
K= 8.	3.5823575	3.4264	5.1647	91.4089
K= 20.	12.5767273	13.9636	46.3619	39.6745
K= 35.	50.2696125	21.0629	49.3592	29.5779
K=100.	784.3314624	47.4652	30.4551	22.0797

An Example

In this section I will briefly illustrate some possible uses of the vector autoregression tools I have described by applying them to a question of current interest, the relationship between money growth and inflation. Two common assertions concerning these variables, see, for example, Meltzer [1978], are that in the long run inflation depends on the difference between the growth rate of money and the growth rate of GNP, and in the short run inflation may vary due to changes in the demand for money and other factors.

I have examined quarterly, seasonally adjusted, United States data for growth rates of money, prices, output, demand deposits and government expenditures, and interest rates. Observations from 1950-1 to 1978-1.

In considering whether money growth causes inflation, a useful procedure is to test for a Granger [1969] causal ordering. In this case we test the null hypothesis that the lags of money growth are all zero in a regression of inflation on constant, trend, and past rates of inflation and money growth. We also test equations with GNP growth and the difference between money growth and GNP growth included as explanatory variables.

Granger Causal Ordering Test Results

Dependent variable is inflation.
Independent variables: I = inflation; M = money growth; G = GNP growth.

	Equation 1		Equation 2		Equation 3		
	I	M	I	M-G	I	M	G
8 Lags	.0000	.1438	.0000	.3488	.0000	.1681	.6996
12 Lags	.0000	.3117	.0000	.3309	.0000	.1546	.4757

Table 6--Shown are the marginal significant levels for Granger tests in three regressions of inflation on lagged values of inflation and other independent variables.

The results of the Granger tests are inconclusive. The marginal significance level of .1438, for example, means that the explanatory power of money growth in this regression would be attained or surpassed 14 percent of the time by chance in repeated samples even if the true coefficients on lags of money growth were zero. This level, although suggestive, is too large to cause rejection of the null hypothesis that money growth does not cause inflation. The use of forecast performance statistics allows us to test directly the ability of money growth to help forecast inflation over different horizons. We can compare the performance, as measured by the root mean square error of forecasts generated by regressions which do and do not include money growth. In these examples the forecasts were made over the period 1968-1 through 1978-1.

Root Mean Square Errors of Predictions
for Inflation Without and With Money Growth

Quarters Ahead		1	2	3	4	5	6	7	8
8	Without	1.75	2.09	2.41	2.87	3.05	3.16	3.23	3.23
Lags	With	1.75	1.94	2.18	2.62	2.76	2.79	2.82	2.81
12	Without	1.71	2.05	2.34	2.74	2.97	3.15	3.25	3.25
Lags	With	2.02	2.24	2.44	2.85	2.94	2.97	2.97	2.92

Table 7--Forecast errors generated by OLS regressions of inflation on lags of inflation with and without lags of money growth.

As shown above, the addition of money growth clearly helps to predict inflation in the long run.

The addition of prior distributions not only leads to better forecasting performance, but allows the money growth lags to enter the regression with varying amounts of weight. Results in Table 8 show that the best long-run forecasts of inflation are made by giving even more weight to past rates of money growth than to past rates of inflation.

RMS Errors of Predictions for Inflation
With Varying Weight on Money Growth

Quarters Ahead		1	2	3	4	5	6	7	8	12
Values	.00	1.54	1.93	2.24	2.63	2.82	2.99	3.05	3.06	2.94
of γ_2	.25	1.79	2.09	2.36	2.68	2.83	2.96	3.03	3.02	2.93
	.50	1.77	2.06	2.30	2.60	2.74	2.83	2.89	2.87	2.84
	.70	1.78	2.05	2.26	2.56	2.67	2.75	2.78	2.77	2.78
	1.00	1.79	2.05	2.25	2.54	2.63	2.69	2.72	2.71	2.75
	2.00	1.87	2.10	2.27	2.54	2.60	2.63	2.65	2.65	2.73
	5.00	1.97	2.17	2.30	2.58	2.62	2.63	2.65	2.65	2.72

Table 8--Shown are prediction error statistics generated from regressions of inflation on twelve lags of inflation and money growth with white-noise-form priors with $\lambda = .1$, $\gamma_1 = 0$, and γ_2 as shown.[25/] Larger values of γ_2 give more weight to lags of variables other than the dependent variable, in this case more weight to money growth. $\gamma_2 = 0$ excludes money growth, $\gamma_2 = 1$ gives equal weights to both variables, $\gamma_2 = \infty$ leaves other variables unconstrained.

Meltzer's conjecture that it is the difference between money growth and growth of output, rather than money growth alone, which causes inflation is not supported by the data. Both Granger tests (Table 6) and forecast errors (Table 7) indicate that inflation is better explained by money growth alone than by the difference between growth of money and output. On the other hand, if we allow growth rates of money and output to enter the regression separately, then the combination does slightly improve forecasts several periods ahead.

RMS Errors of Predictions for Inflation
With Money and GNP Growth Rates

Regression	1	2	3	4	5	6	7	8	12
I on I, M	1.79	2.05	2.25	2.54	2.63	2.69	2.72	2.71	2.75
I on I, M-G	1.84	2.14	2.41	2.74	2.90	3.02	3.09	3.05	2.84
I on I, M, G	1.84	2.09	2.26	2.53	2.62	2.65	2.68	2.66	2.65

Table 9--These prediction errors were generated from regressions with twelve lags on each independent variable and white-noise-form priors with $\lambda = .1$, $\gamma_1 = 0$, $\gamma_2 = 1$.

The conjecture that the best short-run forecasts of inflation may have to take account of other variables is reflected in the data. By including as independent variables the additional series commercial paper rate, and growth of government expenditures and demand deposits, gains are made (see Table 10) in short-term forecasts.

RMS Errors of Predictions for Inflation
With Additional Independent Variables

C = Commercial paper rate
D = Demand deposit growth
E = Government expenditure growth

Regression on Constant and Trend	1	2	3	4	5	6	7	8
Add Lags of I	2.41	2.49	2.59	2.69	2.74	2.80	2.84	2.83
Add Lags of M, G	1.52	1.90	2.18	2.51	2.68	2.84	2.88	2.90
Add Lags of M, G	1.51	1.84	2.13	2.49	2.67	2.84	2.90	2.91
Add Lags of C, D, E	1.49	1.71	1.92	2.25	2.43	2.64	2.72	2.77

Table 10--These prediction errors were generated from regressions of inflation on expanding sets of independent variables. A random-walk-form prior with parameters $\gamma_1 = .8$, $\gamma_2 = 1.$, $\lambda = .1$ was imposed where applicable.

The dynamic effects of innovations to the different series are given by the impulse response functions shown in Table 11. Suppose a typical innovation (one standard deviation in size) in the money growth series of 1.86 percent occurs. This is to say that in the period of the innovation, money growth is 1.86 percent higher than forecast on the basis of past data. The positive serial correlation of money growth causes increases of 1.29, .67, .32, ... percent to occur in subsequent quarters. The effect (see Figure 8) on inflation one quarter ahead is an increase of .01 percent. The effect two quarters ahead is an increase of .07 percent. These effects continue to grow for the first year up to a rate of increase for inflation of about one-quarter percent per quarter. These increases continue at about the same rate for another two years. Meanwhile GNP growth is also stimulated at first by the positive innovation in money growth, but it then declines and becomes negative. The total effect after four years is basically an increase in money and prices, while the level of GNP is about the same as where it would have been without the money growth innovation.

Impulse Response Function

Response to a One-Standard-Deviation
Innovation in GNP Growth

Periods Ahead	GNP Growth	Money Growth	Inflation
1	4.1038	.0000	.0000
2	2.5815	-.0908	-.0790
3	1.1393	-.3121	-.2297
4	.2508	-.3710	-.2227
5	-.5532	-.4059	-.3088
6	-.9452	-.4313	-.2567
7	-.5628	-.3218	-.2072
8	.1705	.0705	-.1867
9	.2507	.3723	-.1064
10	.2465	.4158	-.0646
11	.2562	.3136	-.0149
12	.1436	.1530	.0204
13	.0226	.0193	.0227
14	.0105	-.0218	.0095
15	.0367	-.0007	.0144
16	-.0316	.0057	.0356
17	-.1024	-.0081	.0534
18	-.1287	-.0161	.0671
19	-.1339	-.0145	.0690
20	-.1082	-.0077	.0598

Response to a One-Standard-Deviation
Innovation in Money Growth

Periods Ahead	GNP Growth	Money Growth	Inflation
1	.0000	1.8625	.0000
2	.1124	1.2890	.1227
3	.1570	.6761	.0757
4	.5637	.3232	.2657
5	.4248	.1712	.2964
6	.0192	.1993	.1543
7	-.1561	.2021	.1942
8	-.1329	.0034	.2794
9	-.4683	-.1589	.2889
10	-.5105	-.1779	.2755
11	-.3367	-.0782	.2350
12	-.2159	.0125	.1775
13	-.1044	.0236	.1454
14	.0176	-.0069	.1144
15	.0641	-.0224	.0705
16	.0504	-.0316	.0259
17	.0668	-.0401	.0150
18	.0888	-.0443	.0405
19	.0778	-.0477	.0516
20	.0548	-.0438	-.0570

Response to a One-Standard-Deviation
Innovation in Inflation

Periods Ahead	GNP Growth	Money Growth	Inflation
1	-.0000	.0000	1.6566
2	-.4022	-.1876	1.1546
3	-.4362	-.2769	.9168
4	-.3228	-.3152	.8568
5	-.2269	-.3408	.5851
6	-.3853	-.3273	.3181
7	-.1855	-.2211	.1590
8	.0869	-.1462	.0385
9	.1910	-.0857	.0890
10	.3224	-.0527	.1563
11	.4064	-.0262	.1968
12	.3632	-.0068	.2258
13	.2593	-.0102	.2249
14	.1986	-.0164	.2082
15	.1322	-.0077	.1785
16	.0629	.0073	.1413
17	.0242	.0273	.1028
18	.0031	.0477	.0645
19	-.0180	.0575	.0289
20	-.0335	.0556	.0004

Table 11--Shown are the impulse response functions for a trivariate eight-lag system estimated with a random-walk-form prior with $\lambda = .1$, $\gamma_1 = 0.$, and $\gamma_2 = 1$.

CHAPTER V
SEQUENTIAL ESTIMATION OF COEFFICIENTS

The Kalman Filter

The procedures and statistical theory described so far have all assumed that the economy which generates the data is one which can be approximated reasonably well by a model with constant coefficients. The possibility that structural changes do occur over time suggests that a statistical model which takes this into account may forecast better than one which does not.

A natural way to generalize the vector autoregressive specification so as to include time-varying coefficients is through the use of the Kalman filter.[26/]

The Kalman filter is an algorithm for making recursive linear least squares projections. Given a current coefficient estimate, the Kalman filter generates a new estimate on the basis of new data.

The Kalman filter will generate sequential ordinary least squares estimates as a special case when coefficients are assumed to remain constant, and it will generate mixed estimates when prior restrictions are imposed. However, it also will generate time-varying coefficient models in which coefficients vary according to any general ARMA structure.

The model underlying the Kalman filter is a two-equation linear dynamic system. The first equation generates time-varying coefficients according to a specified structure.

$$\underset{k\times 1}{\beta_{t+1}} = \underset{k\times k}{A_t}\ \underset{k\times 1}{\beta_t} + \underset{k\times 1}{w_t} \qquad t=0, \ldots, N-1 \qquad (115)$$

where A_t is known for all t and w_t is a random vector.

The second equation is the vector autoregressive structure, and may be written

$$\underset{}{Y_t} = \underset{1xk}{X_t}\ \underset{kx1}{\beta_t} + \underset{1x1}{v_t} \qquad t=1, \ldots, N. \tag{116}$$

Assume β_0, w_t, v_s are independent for all t and s. Assume $E(w_t) = 0$, $E(v_t) = 0$ for all t, $E(\beta_0) = \bar{\beta}_0$. Define $E\{(\beta_0-\bar{\beta}_0)'(\beta_0-\bar{\beta}_0)\} = \underset{kxk}{B}$.

$$E[w_t w_t'] = M_t \qquad E[v_t^2] = N_t. \tag{117}$$

Assume $\bar{\beta}_0$ and B are given, along with M_t and N_t, for all t.

The problem is to find for each t linear least square estimates of β_t and β_{t+1} given values of Y_0, Y_1, ..., Y_t, and X_0, ..., X_t. Let

$$y_t' = [Y_0' Y_1' \ldots Y_t']. \tag{118}$$

Denote the linear least squares projections of β_{t+1} and β_t given y_t by $\hat{\beta}_{t+1|t}$ and $\hat{\beta}_{t|t}$, respectively.

The attractive feature of the Kalman filter algorithm which solves this problem is that the estimate $\hat{\beta}_{t+1|t}$ is obtained by means of a simple equation which involves the previous estimate, $\hat{\beta}_{t|t-1}$ and the new information, Y_t and X_t, but does not involve the past data, Y_{t-1}, Y_{t-2}, ..., Y_0 or X_{t-1}, ..., X_0.

Suppose we have the estimate $\hat{\beta}_{t-1|t-1}$ together with $\hat{\Sigma}_{t-1|t-1}$, an estimate of

$$\Sigma_{t-1|t-1} = E\{(\beta_{t-1}-\hat{\beta}_{t-1|t-1})(\beta_{t-1}-\hat{\beta}_{t-1|t-1})'\}. \tag{119}$$

At time t we receive the new data Y_t and X_t which we assumed is generated by the equation $Y_t = X_t\beta_t + v_t$. The Kalman filter algorithm is as follows:

Step 1: Form

$$\hat{\Sigma}_{t|t-1} = A_{t-1}\hat{\Sigma}_{t-1|t-1}A'_{t-1} + M_{t-1} \tag{120}$$

an estimate of

$$\Sigma_{t|t-1} = E\{(\beta_t-\hat{\beta}_{t|t-1})(\beta_t-\hat{\beta}_{t|t-1})'\}. \tag{121}$$

Step 2: Form

$$\hat{\Sigma}_{t|t} = \hat{\Sigma}_{t|t-1} - \hat{\Sigma}_{t|t-1}X'_t(X_t\hat{\Sigma}_{t|t-1}X'_t+N_t)^{-1}X_t\hat{\Sigma}_{t|t-1}. \tag{122}$$

Step 3: Form

$$\hat{\beta}_{t|t} = A_{t-1}\hat{\beta}_{t-1|t-1} + \hat{\Sigma}_{t|t}X'_tN_t^{-1}(Y_t-X_tA_{t-1}\hat{\beta}_{t-1|t-1}). \tag{123}$$

For the purposes of updating OLS or posterior mean coefficient estimates, one sets $A_t \equiv I$. The constant coefficient specification takes $M_t = 0$. The Kalman filter updating can be begun at any point in time; for example, at the beginning of a projection period, by taking the current OLS or posterior mean and covariance matrix for $\hat{\beta}_{t|t}$ and $\hat{\Sigma}_{t|t}$, respectively.

Time Varying Parameter Specifications

The above-described use of the Kalman filter to update parameter estimates throughout a projection period in the constant coefficients specification suggests a natural extension of the VAR methodology in a time-varying, stochastic coefficients model. In this section I report a set of results showing forecast performances with such models. The experiments I have done are limited in scope and the results are mixed. In contrast to the similar experiments reported earlier with the imposition of priors, I have found little room for consistent improvement in forecast performance through the use of time-varying coefficients models, and often considerable deterioration.

A basic weakness of the Kalman filter approach with stochastic coefficients taken here is the assumption that innovations to coefficients are uncorrelated with errors in the regression equation. This assumption implies that the movements in parameters cannot represent structural changes or changes in policy which are responses to shocks to the state of the economy. Such phenomenon can be handled by more general versions of the Kalman filter.[27/] This consideration suggests one interpretation of the weak nature of the results presented here could be as support of Rosenberg's [1973] contention that "stochastic parameter regression should be employed as a supplement to analysis of systematic variation, rather than as an alternative."

The first problem to be faced in the use of the Kalman filter time-varying coefficients model is the specification of the transition matrices, A_t, and the coefficient error variance matrices, M_t. Three different specifications are considered here. The choices were made to reflect ignorance about the source and form of parameter variation. The three specifications are referred to as follows: (A) the Adaptive Regression Model, in which the coefficient on constant follows a random walk and other parameters do not vary; (B) the Random-Walk

Parameters, in which all coefficients follow a random walk; and (C) the Shrink to Mean Estimator, in which the coefficient vector subject to a white-noise disturbance decays toward an unknown mean.

For each of the models a variety of choices for σ_w^2, the scale of the error variance in the Kalman filter coefficients equation, were considered. Experiments were performed on univariate systems. Each equation included six lags and a constant. The data used in this section consisted of monthly observations on M1, Personal Income, Prices, and the rate on four- to six-month prime Commercial Paper, with observations from 1953-1 to 1979-4. An estimation period of 1953-7 to 1959-12 was used to generate an initial OLS parameter vector β_0 and its covariance matrix $\sum_{0|0}$. In each of the experiments these values were entered into the Kalman filter algorithm, and new parameter estimates and covariance matrices were generated for each month from 1953-7 to 1979-4. The use of OLS estimates as initial conditions is simply a device to generate reasonable starting value on the basis of data preceeding the projection period. One could consider more involved attempts to capture initial conditions, such as beginning with OLS estimates in 1959-12 and filtering backwards in time. Since the statistics in this investigation depend only on estimates during the projection period (after 1959-12), the results should not be sensitive to reasonable alternative initial conditions for the coefficients in 1953-7. Projection statistics were generated for the one-, two-, three-, and four-step-ahead forecast errors for the period beginning in 1960-1 and extending through the end of the data. Thus, for example, there were in each experiment 231 one-step forecast errors. The projection period was further broken down into the 120 observations in the 1960s and the 111 in the 1970s.

Experiment A: The Adaptive Regression Model

In this specification the coefficient variation equation generates a random-walk constant. Let α_t be the coefficient on the constant in the regression and partition β_t as $\beta_t = \begin{bmatrix}\alpha_t \\ \beta^*_t\end{bmatrix}$. The Adaptive Regression specification is as follows:

$$\begin{bmatrix}\alpha_t \\ \beta^*_t\end{bmatrix} = \begin{bmatrix}\alpha_{t-1} \\ \beta^*_{t-1}\end{bmatrix} + \begin{bmatrix}w_t \\ 0\end{bmatrix} \quad E(w_t) = 0 \qquad E(w_t w_s) = \delta_{st}\sigma_w^2. \tag{124}$$

This model, which is very similar to one investigated by Cooley and Prescott [1973], can be viewed as a mechanical version of the common practice of treating the constant as an "add factor" in the equations of macroeconomic models. In such models the coefficient on the constant is typically judgmentally adjusted whenever a pattern of positive or negative errors is perceived.

Note that with the Adaptive Regression specification forecasts are obtained exactly as in the nonstochastic parameter case, by use of the chain rule. Since $E_t\beta(t+k) = E_t\beta(t) + E_t(\sum_{s=1}^{K} w(t+s)) = E_t\beta(t)$, the estimate, $\hat{\beta}(t)$, of $\beta(t)$ is substituted for $\beta(t+s)$, s=0, 1, ... in the chain rule of forecasting. As noted earlier, since multistep forecasts are nonlinear functions of the parameters, the chain rule does not give unbiased estimates of the mean of the forecast distribution. In the case considered here with stochastic parameters, even the assumption that the estimated coefficients are the true ones at that point in time does not eliminate the problem. Nevertheless, the chain rule has been used since unbiased forecasts could be generated only by the stochastic procedures described in Chapter 3 and at considerable expense.

The value of σ_w^2 in this experiment was set at 0., .0001, .001, .01, .1, and .2 times the estimated variance of the OLS coefficient on constant. Root mean square error statistics were generated for the two projection periods and are presented in Table 1. Shown in Table 1 are results from Period 1, 1960-1

Table 1. Adaptive Regression Model

PERIOD 1

VARIANCE

VARIABLE	HORIZON	OLS	0	.0001	.0010	.0100	.1000	.2000
M1	1-STEP	.475388	.475293 .999800	.476240 1.001792	.477096 1.003593	.484119 1.018366	.512004 1.077023	.526233 1.106955
M1	4-STEP	1.334770	1.346807 1.009015	1.351902 1.012835	1.358984 1.018141	1.448808 1.085431	1.750653 1.311577	1.904931 1.427160
PER INCOME	1-STEP	2.062538	2.080264 1.008594	2.081592 1.009236	2.087704 1.012201	2.117040 1.026425	2.213136 1.073016	2.275799 1.103397
PER INCOME	4-STEP	3.823710	3.787996 .990660	3.796614 .992914	3.832180 1.002215	3.990626 1.043653	4.410497 1.153460	4.674494 1.222502
COM P RATE	1-STEP	.174018	.173793 .998707	.173704 .998196	.173118 .994828	.171352 .984680	.171486 .985450	.172767 .992811
COM P RATE	4-STEP	.531484	.539064 1.014262	.537791 1.011867	.529392 .996064	.504007 .948301	.495602 .932487	.505366 .950856
PRICES	1-STEP	.169577	.169922 1.002034	.168137 .991508	.167841 .989763	.175745 1.036373	.197057 1.162050	.205504 1.211862
PRICES	4-STEP	.416253	.421782 1.013285	.399677 .960176	.387092 .929944	.447340 1.074683	.635040 1.525611	.707864 1.700562

Table 1. Adaptive Regression Model

PERIOD 2

			VARIANCE					
VARIABLE	HORIZON	OLS	0	.0001	.0010	.0100	.1000	.2000
M1	1-STEP	1.214176	1.217411	1.218804	1.231647	1.252760	1.322571	1.361882
			1.002664	1.003812	1.014554	1.031776	1.069275	1.121651
M1	4-STEP	2.907251	2.919995	2.929056	3.023351	3.140514	3.673649	4.034826
			1.004384	1.007495	1.039935	1.080235	1.263616	1.387849
PER INCOME	1-STEP	5.945937	5.937621	5.945734	5.977800	6.054767	6.274730	6.376992
			.998601	.999966	1.005559	1.018307	1.055297	1.072496
PER INCOME	4-STEP	11.199724	11.223150	11.263075	11.521066	12.064861	13.702449	14.540051
			1.002094	1.007442	1.028692	1.077246	1.223463	1.298251
COM P RATE	1-STEP	.451026	.446916	.446537	.444656	.445316	.448912	.450075
			.990887	.990047	.986320	.987340	.995313	.997867
COM P RATE	4-STEP	1.397791	1.377105	1.373044	1.355781	1.367330	1.409299	1.418764
			.985201	.982296	.969945	.978206	1.008233	1.015004
PRICES	1-STEP	.390431	.388586	.392495	.392701	.387457	.403359	.414566
			.995274	1.005286	1.005814	.992385	1.033112	1.061873
PRICES	4-STEP	1.213591	1.206071	1.244968	1.251327	1.207778	1.357124	1.458650
			.993504	1.025855	1.031094	.995210	1.118271	1.201929

through 1969-12, and Period 2, 1970-1 through 1979-4. For comparative purposes, the OLS RMSE statistics are given along with the Adaptive Regression Model results. Notice that the OLS results are not identical with the constant coefficient (zero variance) specification because in the later case the OLS estimate is taken as the known initial condition and filtered over the estimation period. If it had been taken as the initial condition at the beginning of the projection period, it would be identical to OLS. Underneath the RMSE statistics for the Adaptive Regression Model are given the ratio of these numbers to those of the OLS model.

The message of these results seems to be that there are no consistently significant improvements to be made in forecasting by adopting a time-varying constant. The most improvement was shown in the 1960's projection period by the interest rate variable and prices. Both variables showed a potential decrease in RMSE of about 7 percent for the best specification of coefficient error variance. However, had either of these specifications been chosen in 1970 for use in the following decade, the results would have turned out worse than OLS. In fact, for the later period the largest improvement for any variable was only 3 percent better than OLS.

Experiment B. The Random-Walk Parameters

The second specification of the coefficient variation equation is simply:

$$\beta_t = \beta_{t-1} + w_t \qquad E(w_t) = 0 \qquad E(w_t w_s') = \delta_{st}\sigma_w^2 M. \qquad (125)$$

Each coefficient follows a random walk. The constant M matrix is taken to be equal to $\Sigma_{0|0}$, the estimated variance matrix of the estimation period OLS coefficient vector. This specification is a special case of the third one (when there is no shrinkage), and the results appear as the last column of Table 2. In

Table 2. Shrink to Mean Estimator

PERIOD 1 1-STEP FORECAST

VALUES OF DECAY FACTOR, D

VARIABLE	VARIANCE	0	.10	.30	.50	.70	.90	1.00
M1	0	.999500	.999817	.999899	1.000093	1.000501	1.001064	.999800
M1	.0010	.999792	.999796	.999846	.999987	1.000271	1.000831	1.003513
M1	.0100	.999729	.999613	.999407	.999142	.998662	1.001111	1.028587
M1	.1000	.999327	.998134	.996285	.995183	.997005	1.020398	1.110125
PER INCOME	0	1.008594	1.008749	1.008951	1.009028	1.009170	1.010882	1.008594
PER INCOME	.0010	1.008589	1.008750	1.008965	1.009043	1.009159	1.010741	1.006530
PER INCOME	.0100	1.008539	1.008764	1.009093	1.009207	1.009171	1.010264	.999146
PER INCOME	.1000	1.007986	1.008902	1.010769	1.012429	1.014429	1.019915	1.023010
COM P RATE	0	.998707	.998724	.998759	.998776	.998770	.999931	.998707
COM P RATE	.0010	.998701	.998719	.998764	.998787	.998805	.999931	.994638
COM P RATE	.0100	.998673	.998701	.998787	.998902	.999069	.999977	.990168
COM P RATE	.1000	.998454	.998535	.998960	.999874	1.001460	1.002827	1.029118
PRICES	0	1.002034	1.001852	1.001492	1.001168	1.001136	1.003550	1.002034
PRICES	.0010	1.002064	1.001861	1.001527	1.001215	1.001191	1.003261	.993838
PRICES	.0100	1.002341	1.002164	1.001846	1.001633	1.001687	1.001504	.981589
PRICES	.1000	1.005066	1.004824	1.004777	1.005325	1.006333	1.004087	.996945

Table 2. Shrink to Mean Estimator

PERIOD 1 4-STEP FORECAST

VALUES OF DECAY FACTOR, D

VARIABLE	VARIANCE	0	.10	.30	.50	.70	.90	1.00
M1	0	1.009018	1.009020	1.009127	1.009556	1.010743	1.012271	1.009018
M1	.0010	1.008972	1.008964	1.009027	1.009330	1.010120	1.010819	1.013932
M1	.0100	1.008559	1.008464	1.008169	1.007478	1.005618	1.007204	1.042642
M1	.1000	1.005053	1.004402	1.002346	.998711	.997564	1.047668	1.141916
PER INCOME	0	.990660	.990934	.991600	.992600	.994744	1.002054	.990660
PER INCOME	.0010	.990622	.990892	.991548	.992531	.994679	1.002617	.999031
PER INCOME	.0100	.990280	.990526	.991099	.991948	.994200	1.007099	1.032188
PER INCOME	.1000	.987557	.987698	.988028	.989027	.995685	1.036159	1.102687
COM P RATE	0	1.014262	1.014309	1.014412	1.014512	1.014746	1.019557	1.014262
COM P RATE	.0010	1.014256	1.014307	1.014424	1.014542	1.014806	1.019504	.999118
COM P RATE	.0100	1.014207	1.014292	1.014512	1.014813	1.015336	1.019210	.973864
COM P RATE	.1000	1.013778	1.014177	1.015370	1.017391	1.020625	1.023566	1.031079
PRICES	0	1.013283	1.012757	1.011608	1.010633	1.010964	1.021525	1.013283
PRICES	.0010	1.013427	1.012889	1.011793	1.010873	1.011284	1.020396	.973994
PRICES	.0100	1.014710	1.014251	1.013427	1.013002	1.013931	1.012951	.902547
PRICES	.1000	1.026892	1.027118	1.028600	1.031981	1.036218	1.007399	.888707

Table 2. Shrink to Mean Estimator

PERIOD 2 1-STEP FORECAST

VALUES OF DECAY FACTOR, D

VARIABLE	VARIANCE	0	.10	.30	.50	.70	.90	1.00
M1	0	1.002664	1.002669	1.002681	1.002691	1.002687	1.002651	1.002664
M1	.0010	1.003085	1.002614	1.002346	1.001975	1.002201	1.007515	1.025303
M1	.0100	1.008057	1.005376	1.001456	.999842	1.004251	1.028227	1.066521
M1	.1000	1.045109	1.028017	1.005011	.999076	1.016329	1.071780	1.129698
PER INCOME	0	.998601	.998630	.998710	.998816	.998930	.998810	.998601
PER INCOME	.0010	.998638	.998617	.998660	.998915	.999520	1.001599	1.018456
PER INCOME	.0100	.999439	.998737	.998244	.999235	1.004154	1.025075	1.086621
PER INCOME	.1000	1.013213	1.004835	.996509	.999711	1.023704	1.104685	1.267614
COM P RATE	0	.990887	.990894	.990907	.990916	.990890	.990812	.990887
COM P RATE	.0010	.990812	.990825	.990861	.990901	.990956	.991422	1.005538
COM P RATE	.0100	.990151	.990244	.990451	.990765	.991548	.996446	1.066103
COM P RATE	.1000	.984888	.985473	.987047	.989821	.996971	1.034373	1.225850
PRICES	0	.995274	.995323	.995426	.995525	.995597	.995712	.995274
PRICES	.0010	.995794	.995546	.995041	.994411	.993430	.990157	1.001713
PRICES	.0100	1.000251	.997531	.992163	.986279	.979558	.973693	1.010929
PRICES	.1000	1.023451	1.009774	.985470	.966317	.957818	.976604	1.043150

Table 2. Shrink to Mean Estimator

PERIOD 2 4-STEP FORECAST

VALUES OF DECAY FACTOR, D

VARIABLE	VARIANCE	0	.10	.30	.50	.70	.90	1.00
M1	0	1.004384	1.004404	1.004411	1.004372	1.004326	1.004871	1.004384
M1	.0010	1.004247	1.004193	1.004043	1.003886	1.004552	1.016016	1.044777
M1	.0100	1.004777	1.004354	1.004191	1.006634	1.019996	1.075403	1.085317
M1	.1000	1.022551	1.021270	1.023987	1.041986	1.097393	1.229835	1.115730
PER INCOME	0	1.002094	1.002166	1.002410	1.002759	1.003053	1.001834	1.002094
PER INCOME	.0010	1.002023	1.002103	1.002416	1.003055	1.004379	1.007792	1.046330
PER INCOME	.0100	1.001542	1.001609	1.002301	1.005223	1.017676	1.061979	1.156411
PER INCOME	.1000	1.001588	1.001312	1.003889	1.019409	1.083589	1.299620	1.381923
COM P RATE	0	.985201	.985193	.985175	.985141	.984984	.964184	.985201
COM P RATE	.0010	.985116	.985110	.985091	.985056	.984971	.965375	1.013808
COM P RATE	.0100	.984370	.984377	.984347	.984325	.984879	.994410	1.109892
COM P RATE	.1000	.978176	.978291	.978295	.978776	.985423	1.045186	1.250085
PRICES	0	.993804	.993832	.993921	.994030	.994031	.993945	.993804
PRICES	.0010	.993429	.993390	.993275	.993097	.992726	.987913	1.029743
PRICES	.0100	.990646	.990048	.988252	.985320	.980373	.972026	1.042742
PRICES	.1000	.978596	.975652	.967867	.957772	.952150	1.004078	1.056189

this table only the ratio to OLS RMSE's appear, since the base period is the same.

The results of this experiment are similar to those of the first. A possibly significant improvement in forecasting of prices is shown in the 1960s with a reduction of 11 percent in the OLS RMSE. Just as in the previous experiment, however, the improvement fails to remain in the later period.

Experiment C. Shrink to Mean Estimators

This specification involves a common adaptation of the Kalman filter algorithm. A set of state variables is appended to the parameter vector. Let β_t be the current parameters in the regression equation $\underset{1x1}{Y_t} = \underset{1xp}{X_t} \underset{px1}{\beta_t} + \underset{1x1}{V_t}$. $\bar{\beta}$ is an unknown mean toward which β_t shrinks over time, subject to a white-noise disturbance. The rate of decay is controlled by a parameter, d. The coefficient equation is written as:

$$\begin{bmatrix} \beta_t \\ \bar{\beta} \end{bmatrix} = \begin{bmatrix} d \cdot I_p & (1-d) \cdot I_p \\ 0_p & I_p \end{bmatrix} \begin{bmatrix} \beta_{t-1} \\ \bar{\beta} \end{bmatrix} + \begin{bmatrix} w_t \\ 0 \end{bmatrix} \begin{matrix} E(w_t) = 0 \\ E(w_t w_s') = \delta_{st} \sigma_w^2 M. \end{matrix} \tag{126}$$

The covariance matrix of this augmented parameter vector may be partitioned as:

$$\underset{2px2p}{\Sigma_{t|t}} = \begin{bmatrix} A_{t|t} & B'_{t|t} \\ B_{t|t} & C_{t|t} \end{bmatrix}. \tag{127}$$

I take $B_{0|0} = 0_p$, and $A_{0|0} = C_{0|0} = M = \Sigma_{0|0}$, the (pxp) OLS estimate. Making these substitutions into the Kalman filter algorithm, one obtains the following:

Step 1: Form

$$\hat{\Sigma}_{t|t-1} = \begin{bmatrix} \hat{A}_{t|t-1} & \hat{B}'_{t|t-1} \\ \hat{B}_{t|t-1} & \hat{C}_{t|t-1} \end{bmatrix} = \tag{128}$$

$$\begin{bmatrix} d^2\hat{A}_{t-1|t-1}+(1-d)d(\hat{B}_{t-1|t-1}+\hat{B}'_{t-1|t-1})+(1-d)^2\hat{C}_{t-1|t-1}+\sigma_w^2 M & d\hat{B}'_{t-1|t-1}+(1-d)\hat{C}_{t-1|t-1} \\ d\hat{B}_{t-1|t-1}+(1-d)\hat{C}_{t-1|t-1} & \hat{C}_{t-1|t-1} \end{bmatrix}$$

Step 2: Form

$$\hat{\Sigma}_{t|t} = \begin{bmatrix} \hat{A}_{t|t} & \hat{B}'_{t|t} \\ \hat{B}_{t|t} & \hat{C}_{t|t} \end{bmatrix} = \tag{129}$$

$$\hat{\Sigma}_{t|t-1} - (X_t\hat{A}_{t|t-1}X'_t+\sigma_v^2)^{-1} \begin{bmatrix} \hat{A}_{t|t-1}X'_tX_t\hat{A}_{t|t-1} & \hat{A}_{t|t-1}X'_tX_t\hat{B}'_{t|t-1} \\ \hat{B}_{t|t-1}X'_tX_t\hat{A}_{t|t-1} & \hat{B}_{t|t-1}X'_tX_t\hat{B}'_{t|t-1} \end{bmatrix}$$

Step 3: Form

$$\hat{\beta}_t = d\cdot\hat{\beta}_{t-1} + (1-d)\overline{\beta}_{t-1} + \hat{A}_{t|t}X'_t\cdot(Y_t-X_t(d\hat{\beta}_{t-1}+(1-d)\overline{\beta}_{t-1}))/\sigma_v^2 \tag{130}$$

$$\hat{\overline{\beta}}_t = \hat{\overline{\beta}}_{t-1} + \hat{B}_{t|t}X'_t\cdot(Y_t-X_t(d\hat{\beta}_{t-1}+(1-d)\overline{\beta}_{t-1}))/\sigma_v^2. \tag{131}$$

In making forecasts with this model one cannot simply substitute $\hat{\beta}_t$ into the chain rule. $E(\beta_{t+K}) = d^K\beta_t + (1-d^K)\overline{\beta}$, thus one substitutes $d^K\hat{\beta}_t + (1-d^K)\hat{\overline{\beta}}$ for β_{t+K}.

The results shown in Table 2 suggest that the Shrink to Mean Estimator does not generally produce improvements in forecasting. Neither do the results exhibit the type of consistent pattern which might suggest directions in which improvements might be obtained.

In the 1960's period a very small improvement in M1, Personal Income, and the Commercial Paper Rate for different forecast horizons is matched with larger increases in forecast errors for these variables in other horizons. For

Prices, the only variable for which there was significant improvement in the Random Walk Parameter specification, all values of d less than one which were considered led to worse than OLS forecasts at all horizons.

In the 1970's period more significant reduction in the RMSEs are generated for Prices and the Commercial Paper Rate. There are also, however, significant increases in errors of M1 and Personal Income.

In conclusion, while the experiment results were not entirely negative, the overall absence of a consistent positive pattern across variables and across time periods, along with the small sizes of the improvements where they do appear, suggests that the Shrink to Mean Estimator is probably not a fruitful path for forecasting purposes.*

---

*In several multivariate experiments results were obtained which matched closely the univariate results reported here.

Footnotes

1/ See, for example, Lucas and Sargent [1978].

2/ The procedures described by Granger and Newbold [1977] illustrate these problems.

3/ An example is given in Sargent [1978].

4/ The proof for the univariate case is in Anderson [1971] on page 419.

5/ D(t) is an (nx1) vector function for which we estimate a (nxd) matrix of parameters, C. For example, we may have a constant and trend, in which case d = 2, $C = (\underset{nx1}{C_0}, \underset{nx1}{C_1})$, and $D(t) = C_0 + C_1 t$.

6/ This is shown for the case of fixed regressors in Anderson [1958], Chapter 8.

7/ I will not discuss these possibilities further. As discussed earlier, asymptotic results are not particularly relevant to this investigation. For a discussion of asymptotic results in the stationary case see Anderson [1971], Anderson and Taylor [1976], and Ljung [1976]. For the nonstationary case see Fuller, Hasza, and Goebel [1979], and Sims [1978].

8/ The program "Normal" on the University of Minnesota computer system was used to generate the random errors.

9/ The graphs were generated by using a moving average to smooth histograms of the estimates with interval size of .002. The smoothing process causes the distribution of the unconditional maximum likelihood estimator to slightly spill over past 1.0, although there are no observations in that region. The scales on the graphs vary so that the area under the curves equals 1.

10/ Ridge estimation is usually performed after the data have been demeaned and scaled so that X'X has unit diagonal elements. Swamy and Rappaport [1975] recommend against this scaling in the context of univariate autoregression because "this practice results in a transformation of the original parameters." However, when the data vary in size, as they usually will in a multivariate regression, the uniform standard deviations of the prior distribution corresponding to a standard ridge regression do not seem appropriate. The solution to this problem to be developed here will differ slightly from the ridge approach. Rather than scaling the data, we scale the size of the standard deviations associated with coefficients on different variables in the prior distribution, and whereas in ridge regression the scaling is according to the standard errors of the independent variables, we scale according to the standard error of each variable's residuals. The procedure is described in more detail in the next section.

11/ Perhaps the most difficult task in a Bayesian analysis is the construction of a prior distribution which accurately captures the a priori knowledge of the investigator. While the prior which is described here necessarily represents only my own opinions, it was developed with the aid of many helpful suggestions from Christopher Sims.

[12]Zellner [1971] has suggested the beta density function $p(\beta) \propto (1-\beta^2)^{-1/2}$ when we know $|\beta| < 1$.

[13]Seemingly Unrelated Regression is discussed in a Bayesian context in Zellner [1971], section 8.5; and Leamer [1978], section 8.3.

[14]Zellner [1971] describes this distribution on page 99.

[15]Following the notation of Granger and Newbold [1977], an ARIMA model specified as $Y_t \sim \text{ARIMA}(p,d,q)$ refers to a process generated by $A(L)(1-L)^d Y(t) = B(L)\varepsilon(t)$ where $\varepsilon(t)$ is a zero-mean white noise, A(L) and B(L) are polynomials in the lag operator of orders p and q, respectively, and d is an integer.

[16]When actual values of Y(t) are unavailable for some values of $t \leq T + m$, then k-step-ahead forecast errors for some values of k will not be known at the end of the projection period.

[17]These data, and all other data referred to in this paper, unless otherwise noted, are the seasonally adjusted values published by the U.S. Department of Commerce, Bureau of Economic Analysis, with revisions available as of the end of the projection period.

[18]The results in this section were generated using the Regression Analysis of Time Series, RATS, computer program written by Thomas A. Doan at the Federal Reserve Bank of Minneapolis.

[19]Although the problem in this comparison caused by seasonal adjustment, as shown here, is not large, two other problems were encountered which prevented a meaningful comparison. The first problem was that a significant proportion of the variance of the sequence of current observations on M1 is eliminated by subsequent revision of the data. The time series of not seasonally adjusted data available at the end of the projection period is smoother than the sequence of current observations. For this reason, a meaningful comparison of forecast errors would have required construction of different data sets at each period in time. The other problem was that Federal Reserve forecasts of M1 growth are made on a weekly basis, none of the weekly forecasts are based on the same information as is contained in a monthly data set, and the standard error of forecasts decreases considerably with each additional week of information.

[20]See, for example, McNees [1975].

[21]Misspecification is used here to refer to the residual variance of forecasts not accounted for by the first three sources. It is not intended to be a measure of misspecification in the standard sense, that is, the degree to which the specification of the model differs from the true underlying structure. If the estimated model is not misspecified in the standard sense, then this measure has an expected value of zero. However, if the model is misspecified, this measure may have a nonpositive expected value.

[22]Fair plans to incorporate the correction for degrees of freedom into a forthcoming version of this paper.

23/ See the appendix to this section for details of the estimation procedure.

24/ Helpful suggestions in the construction of the priors mentioned in this section were given by Thomas Doan and Preston Miller. The ARIMA specifications used for comparison in the following graphs were constructed by Doan.

25/ A white-noise-form prior refers to a prior distribution which has means of zero for all coefficients. A random-walk-form prior refers to the prior distribution in which all means are zero except the mean of the coefficient on the first lag of the dependent variable, which has a mean of one.

26/ Since their introduction (Kalman [1960]), Kalman filters have been used extensively in engineering and control application.

27/ One possible generalization is to specify nonzero covariances of v_t and w_t. Another possibility is to include other driving variables in the state transition equation (117), leading to a systematic parameter variation model. Specification of these types of models require considerable knowledge about the structure of the system and are beyond the scope of this paper.

References

1. Anderson, T. W. (1958) An Introduction to Multivariate Statistical Analysis, New York: John Wiley and Sons, Inc.

2. ________. (1971) The Statistical Analysis of Time Series, New York: John Wiley and Sons, Inc.

3. ________, and John B. Taylor. (1976) "Conditions for Strong Consistency of Least Squares Estimates in Linear Models," Technical Report No. 213, Institute for Mathematical Studies in the Social Sciences, Stanford University.

4. Box, G. E. P., and G. M. Jenkins. (1970) Time Series Analysis, Forecasting and Control, San Francisco: Holden Day.

5. Chamberlain, Gary, and Edward E. Leamer. (1976) "Matrix Weighted Averages and Posterior Bounds," Journal of the Royal Statistical Society, Series B, 73-84.

6. Chow, Gregory C. (1973) "Multiperiod Predictions from Stochastic Difference Equations by Bayesian Methods," Econometrica, 41, 109-118.

7. Cooley, J. F., and E. C. Prescott. (1973) "Varying Parameter Regression: A Theory and Some Applications," Annals of Economic and Social Measurement, 463-473.

8. Fair, Ray C. (1978a) "The Fair Model as of April 15, 1978," mimeo.

9. ________. (1978b) "Estimating the Expected Prediction Accuracy of Econometric Models," Cowles Foundation Discussion Paper No. 480. Forthcoming in International Economic Review.

10. ________. (1978c) "An Analysis of the Accuracy of Four Macroeconometric Models," Cowles Foundation Discussion Paper No. 492. Forthcoming in Journal of Political Economy.

11. Fuller, Wayne A. (1976) Introduction to Statistical Time Series, New York: John Wiley and Sons, Inc.

12. ________, David P. Hasza, and J. Jeffery Goebel. (1979) "Estimation of the Parameters of Stochastic Difference Equations," Iowa State University Statistical Laboratory Report.

13. Geweke, John. (1978) "An Efficient Method for Revising Seasonally Adjusted Time Series," Discussion Paper No. 78-22, Social Science Institute, University of Wisconsin.

14. Granger, Clive W. J. (1969) "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," Econometrica, 37, No. 3, 424-438.

15. ________, and P. Newbold. (1977) Forecasting Economic Time Series, New York: Academic Press.

16. Goldberger, Arthur S. (1964) Econometric Theory, New York: John Wiley and Sons, Inc.

17. Hansen, Lars P., and Thomas J. Sargent. (1979) "Formulating and Estimating Dynamic Linear Rational Expectations Models: I," manuscript. Forthcoming in Journal of Economic Dynamics and Control.

18. Hoerl, Arthur E., and Robert W. Kennard. (1970) "Ridge Regression: Biased Estimation for Nonorthogonal Problems," Technometrics, 12, No. 1, 55-67.

19. Kalman, R. E. (1960) "A New Approach to Linear Filtering and Prediction Problems," Journal of Basic Engineering, 82, 34-45.

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

21. __________. (1978) Specification Searches, Ad Hoc Inference with Nonexperimental Data, New York: John Wiley and Sons, Inc.

22. Liu, Ta-Chung. (1960) "Underidentification, Structural Estimation, and Forecasting," Econometrica, 28, No. 4, 855-865.

23. Ljung, Lennart. (1976) "Consistency of the Least-Squares Identification Method," IEEE Transactions on Automatic Control, 779-781.

24. Lucas, Robert E., Jr. (1976) "Econometric Policy Evaluation: A Critique," in The Phillips Curve and Labor Markets, edited by K. Brunner and A. H. Meltzer, Carnegie-Rochester Conference Series on Public Policy, 19-46, Amsterdam: North Holland.

25. __________, and Thomas J. Sargent. (1978) "After Keynesian Macroeconomics," in After the Phillips Curve: Persistence of High Inflation and High Unemployment, Federal Reserve Bank of Boston.

26. Maddala, G. S. (1977) Econometrics, New York: McGraw Hill.

27. McNees, Stephen K. (1975) "An Evaluation of Economic Forecasts, New England Economic Review, 3-39.

28. Meltzer, Allan H. (1978) "Money Growth and Inflation," The Wall Street Journal, (May 17) 18.

29. Nelson, Charles R. (1972) "The Prediction Performance of the FRB-MIT-PENN Model of the U.S. Economy," The American Economic Review, 62, No. 5, 902-917.

30. Rosenberg, Barr. (1973) "A Survey of Stochastic Parameter Regression," Annals of Economic and Social Measurement, 381-397.

31. Sargent, Thomas J. (1978) "Estimation of Dynamic Labor Demand Schedules Under Rational Expectations," Journal of Political Economy, 86, No. 6, 1009-1044.

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

33. Shiskin, Julius, Allan H. Young, and John C. Musgrave. (1967) "The X-11 Variant of the Census Method II Seasonal Adjustment Program," U.S. Department of Commerce, Bureau of the Census, Technical Paper No. 15.

34. Sims, Christopher A. (1974) "Distributed Lags," Chapter 5 of Frontiers of Quantitative Economics, Volume II, edited by M. P. Intriligator and P. A. Kendrick, 289-332.

35. ________. (1977) "Macroeconomics and Reality," University of Minnesota, Center for Economic Research, Discussion Paper No. 77-91. Forthcoming in Econometrica.

36. ________. (1978) "Least-Squares Estimation of Autoregressions With Some Unit Roots," University of Minnesota, Center for Economic Research, Discussion Paper No. 78-95.

37. Stein, Charles M. (1974) "Multiple Regression," Chapter 37 in I. Olkin (ed.) Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling, Stanford University Press.

38. Swamy, P. A. V. B., and Paul Rappoport. (1975) "Relative Efficiencies of Some Simple Bayes Estimators of Coefficients in Dynamic Models-I," Journal of Econometrics, 3, 273-296.

39. Theil, H. (1963) "On the Use of Incomplete Prior Information in Regression Analysis," Journal of the American Statistical Association, 58, 401-414.

40. Wecher, William. (1979) "Predicting the Turning Points of a Time Series," Journal of Business, 52, No. 1, 35-50.

41. Wold, H. (1954) A Study in the Analysis of Stationary Time Series, (second edition), Stockhold: Almquist and Witsell.

42. Zellner, Arnold. (1971) An Introduction to Bayesian Inference in Econometrics, New York: John Wiley and Sons, Inc.