COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale University
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 1025

BAYES METHODS FOR TRENDING MULTIPLE TIME SERIES WITH AN EMPIRICAL APPLICATION TO THE U.S. ECONOMY

by

Peter C.B. Phillips

August 1992

BAYES METHODS FOR TRENDING MULTIPLE TIME SERIES WITH AN EMPIRICAL APPLICATION TO THE US ECONOMY*

by

Peter C. B. Phillips**

Cowles Foundation for Research in Economics
Yale University
and
University of Auckland

0. ABSTRACT

Multiple time series models with stochastic regressors are considered and primary attention is given to vector autoregressions (VAR's) with trending mechanisms that may be stochastic, deterministic or both. In a Bayesian framework, the data density in such a system implies the existence of a time series "Bayes model" and"Bayes measure" of the data. These are predictive models and measures for the next period observation given the historical trajectory to the present. Issues of model selection, hypothesis testing and forecast evaluation are all studied within the context of these models and the measures are used to develop selection criteria, test statistics and encompassing tests within the compass of the same statistical methodology. Of particular interest in applications are lag order and trend degree, causal effects, the presence and number of unit roots in the system, and for integrated series the presence of cointegration and the rank of the cointegration space, which can be interpreted as an order selection problem. In data where there is evidence of mildly explosive behavior we also wish to allow for the presence of co-motion among variables even though they are individually not modelled as integrated series. The paper develops a statistical framework for addressing these features of trending multiple time series and reports an extended empirical application of the methodology to a model of the US economy that sets out to explain the behavior of and to forecast interest rates, unemployment, money stock, prices and income. The performance of a data-based, evolving "Bayes model" of these series is evaluated against some rival fixed format VAR's, VAR's with Minnesota priors (BVARM's) and univariate models. The empirical results show that fixed format VAR's and BVARM's all perform poorly in forecasting exercises in comparison with evolving "Bayes models" that explicitly adapt in form as new data becomes available.

JEL Classification No. 211

Keywords: Bayes model, Bayes measure; Causality; Cointegration; Co-motion; Deterministic trend; Forecast-encompass; One-period ahead forecasts; Order selection; PIC criterion; PICF criterion; RUMPY model; Unit root.

First draft: July, 1992
This version: August, 1992

---

*Invited paper to be presented at the ASA Meetings in Boston, August, 1992.

**All the computations and graphics that are reported in this paper were performed by the author on a 486-33 PC using programs written in GAUSS-386i (Version 2.2). John McDermott kindly ran regressions in RATS to produce the BVARM forecasts. The author thanks the NSF for research support under Grant No. SES 9122142, and Glena Ames for her skill and effort in keyboarding the manuscript.

1. INTRODUCTION

Scientific interest in Bayes methods has grown sharply over recent years and has affected many different fields of statistics. Lindley (1990) and Breslow (1990) provide two recent reviews of ongoing Bayesian research, the first in statistics generally and the second in the field of biostatistics. Both papers document the growing use of Bayesian methodology. In econometrics, Bayesian methods have also become more popular and are now much more frequent in empirical economic research than they were a decade ago. Zellner (1985) traces developments in Bayesian econometrics to the mid 1980's and there has been substantial growth in the range of applications since then, especially to the field of time series and unit root econometrics. A recent special issue of the *Journal of Econometrics* edited by Poirier (1991) illustrates the wide range of empirical research in economics and finance that is now being conducted using Bayesian statistical methods.

The treatment of trending economic time series has received special attention lately – see Sims (1988), Sims-Uhlig (1991), Phillips (1991a, b) and the discussion therein. Much of the attention has focused on issues connected with the construction of "noninformative" priors for the coefficients in simple time series models like autoregressions. As the discussion of the Phillips (1991a) paper and a recent analysis by Berger-Yang (1992) makes clear, the development of noninformative priors even in the simplest example of the AR(1) model is far from straightforward. These problems become more severe in models with many lags and in multivariate models. When dealing with trending multiple time series we must accept that the number of stochastic trends (or unit autoregressive roots) in the system is unknown and is a matter to be determined empirically. If we intend also to allow for the presence of deterministic trends and for possible cointegration among the series that do have unit roots then the problem of specifying any prior, let alone a noninformative prior that works well in the sense of giving reasonable results in a wide range of applications seems a formidable undertaking.

Some success in forecasting exercises has been achieved with Bayesian vector autoregressions (VAR's) that automatically implement priors for the coefficients that are centered on a vector random walk with degrees of tightness in the prior that can be set *a priori* by the researcher. These procedures are discussed by Doan *et al.* (1984), Litterman (1986) and Todd (1990) and are available as a standard option in RATS (1990). Use of the Minnesota priors, as they are termed by Todd, is convenient in applications but ignores the possibility supported both by data and by economic theory that there is important co-trending behavior amongst many economic time series. Moreover, by shrinking the estimated coefficients towards a vector random walk, the Minnesota priors subtly alter the role of the fitted intercept in a VAR to that of a drift and this can have major consequences in forecasting exercises. According to the evidence presented by Litterman and Todd such Bayesian VAR's perform well in practical macroeconomic forecasting. However, the analysis of a variety of forecasting models and methods by Fair-Shiller (1990) indicates that, although VAR forecasts (both with and without Minnesota priors) do well, they are generally outperformed by those of the structural econometric model of Fair (1976) in terms both of accuracy (using a root mean squared error-RMSE-criterion) and of the information contained in the forecasts (which is assessed by a regression of the actual values on the predicted values from the rival models). In any event, Bayesian VAR's with Minnesota priors now establish a strong benchmark against which any competitor model and methodology should be compared.

This paper offers a new approach to modelling and forecasting multivariate time series. We concentrate on linear models with stochastic regressors and most of our attention and our empirical application is devoted to VAR models with possible trends, in which we allow for restrictions on the coefficients that include unit root, cointegrating and noncausality restrictions. The approach we adopt leads to an evolving sequence of "Bayes models" for the data. These "Bayes models" are predictive models for the present observation given the past history of the data. Associated with these models are "Bayes measures" or data distributions that can be used as the basis for comparing rival models. The Bayes measure may also be taken conditional on the realized trajectory of the time series to the present data point, and under this measure the systematic part of the model is an optimal data-dependent predictor that is analogous to the Wiener-Kolmogorov predictor. When the process is initialized with flat priors on the coefficients, the optimal Bayes model predictor is identical to the maximum likelihood estimator (MLE) of the traditional Wiener-Kolmogorov predictor and this correspondence provides a classical interpretation to our approach. In an *evolving* sequence of "Bayes models" we permit model characteristics to be determined on a period by period basis. This includes lag length, trend degree, number of unit roots, causal effects and co-motion among the variables. Co-motion is equivalent to cointegration when the individual series are integrated (i.e. have unit roots) but allows for individual series nonstationarity of the mildly explosive kind also, which is often relevant in our data-dependent "Bayes models." The result of this analysis is an evolving sequence of "Bayes models" where the dimension of the parameter space is time and data dependent as are the parameters themselves. In a certain sense, the outcome is a sequence of best "Bayes models" for the data in a certain class of models and potential restrictions, because the model configuration each period has the highest posterior odds under the respective Bayes measures of the competing models.

The approach just outlined leads naturally to a Bayesian methodology of data-oriented forecasting models wherein Bayesian updating procedures are used to revise the forecasting model period by period. Dynamic linear state space models that are updated by Kalman filter methods belong to the same class and have been reviewed recently by West-Harrison (1989). The approach offered here is related to that work but also embodies a methodology of model selection, hypothesis testing and forecast encompassing that is based on the Bayes measures associated with the various competing models.

The paper proceeds as follows. In the next section we outline the class of linear multiple time series models that interest us and the specializations that attract our main attention in applications. Section 3 develops "Bayes models" and their associated measures for this class. The properties of these models and measures are explored and this extends the analysis of univariate models given in an earlier paper by Phillips-Ploberger (1992). Issues of model choice, hypothesis testing and encompassing are explored in Section 4. As in Phillips-Ploberger (1992), we use the Radon-Nikodym (RN) derivative of the respective "Bayes model" measures as the criterion for model choice and put forward a forecast version of this criterion that is useful for evaluating the forecasts from rival models. In order selection problems attention is given to both lag order and trend degree in what we characterize as "VAR(p) + Tr(r)" models, i.e. VAR models with p lags and an r'th order deterministic trend (with $r \geq -1$, and $r = -1$ corresponding to the "no intercept" case). This section also develops ideas of evolving "Bayes models" and shows how these may be compared with certain fixed format models. Section 5 reports an extended empirical application of this methodology to a "RUMPY" model of the US economy with quarterly data over the period 1959:1-1992:1 on interest rates (R), unemployment (U), money stock (M), prices (P) and real GDP (Y). Both univariate and multivariate "Bayes models" are considered. We conduct within sample and forecast-based tests of money-income and unemployment-money noncausality, unit roots, and co-motion among the variables. Forecast encompassing tests of a sequence of evolving "Bayes models" against fixed format models are performed for individual series as well as the series taken together. Finally, we report some summary comparisons of the forecasting performance of multiple time series evolving "Bayes models" against that of a Bayesian VAR with Minnesota priors (BVARM). For all series, forecasts over the period 1984:1-1992:1 from the evolving "Bayes model" are superior to those from the BVARM and for some series (R, M, U and P) the improvement is substantial.

2. LINEAR MULTIPLE TIME SERIES MODELS WITH TREND AND CO-MOTION

Our starting point is the multivariate linear regression model

$$y_t = \Pi x_t + \varepsilon_t\ , \quad (t = 1, 2, \ldots) \tag{1}$$

whose dependent variable y_t and error ε_t are m-vector valued stochastic processes on a probability space $(\Omega, \mathcal{F}, P)$. Accompanying y_t is a filtration $\mathcal{F}_t \subset \mathcal{F} (t = 0, 1, 2, \ldots)$ to which both y_t and ε_t are adapted. The error vector ε_t in (1) is a martingale difference sequence with respect to $\mathcal{F}_t$ and will be assumed to have constant conditional covariance matrix $\Sigma = E(\varepsilon_t \varepsilon_t' | \mathcal{F}_{t-1})$. The regressors x_t $(k \times 1)$ in (1) are defined on the same probability space and are assumed to have the property that x_t is $\mathcal{F}_{t-1}$-measurable. Π is a matrix $(m \times k)$ of coefficients which we assume can be parameterized in the form

$$\mathrm{vec}(\Pi) = S\alpha + s\ , \tag{2}$$

where S is an $mk \times q$ matrix of known constants whose rank is q and s is a known vector. In (2) α is taken as a q-vector of basic parameters. The formulation allows the same parameters to occur in more than one equation of (1). Further, when $s \neq 0$ certain elements of Π may be preassigned to specific nonzero values such as unity. It will sometimes be useful to us to write the dependence in (2) explicitly in terms of the coefficient matrix Π in (1) as $\Pi = \Pi(\alpha)$, leading to the model

(1′) $$y_t = \Pi(\alpha)x_t + \varepsilon_t\,,\ \ (t = 1, 2, \ldots)$$

A common instance of (1) is the vector autoregression (with $k = p$ lags) given by

(3) $$y_t = \Sigma_{i=1}^{p}\Pi_i y_{t-i} + \varepsilon_t\,,$$

and the same model augmented with a deterministic trend of degree r

(4) $$y_t = \Sigma_{i=1}^{p}\Pi_i y_{t-i} + \Sigma_{j=-1}^{r} c_j t^j + \varepsilon_t\,,\quad c_{-1} = 0\,.$$

Throughout this paper we will refer to the model (4) as a "VAR(p) + Tr(r)" system. When $r = -1$ the model has no fitted intercept and reduces to (3).

Often it is useful to write (4) in difference form as

(5) $$\Delta y_t = Hy_{t-1} + \Sigma_{i=1}^{p-1}\Phi_i\Delta y_{t-i} + \Sigma_{j=-1}^{r} c_j t^j + \varepsilon_t\,.$$

where $H = \Sigma_{k=1}^{p}\Pi_k - I$ and $\Phi_i = -\Sigma_{k=i+1}^{p}\Pi_k$. We call the matrix $K = I + H$ the long run autoregressive coefficient matrix because when there is no trend (i.e. when $r = -1$) the value of K determines the form that the spectral density of the autoregressive process y_t takes at the zero frequency. The format of (5) is convenient when we wish to allow (or test) for unit roots and cointegration in specific equations. For instance, when $H = 0$ there are n unit roots in (5), and when

$$H = \begin{array}{cc} \begin{array}{cc} m_1 & m_2 \end{array} & \\ \begin{bmatrix} B_1 & B_2 \\ 0 & 0 \end{bmatrix} & \begin{array}{c} m_1 \\ m_2 \end{array} \end{array}, \quad \text{and} \quad y_t = \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} \begin{array}{c} m_1 \\ m_2 \end{array},$$

there are m_2 ($= m - m_1$) unit roots in the final m_2 equations of (5) and the linear relation $B_1 y_{1t} + B_2 y_{2t}$ is cointegrating in the first m_1 equations of (5). More generally, we may have

(6) $$H = \begin{bmatrix} B_1 & B_2 \\ 0 & H_{22} \end{bmatrix}\,,\quad H_{22} = \mathrm{diag}(h_{n_2+1}, \ldots, h_n)$$

and in our empirical applications we want to allow for the possibility that, for some i, $h_{ii} > 0$ at least over certain subperiods of data. In such cases h_{ii} is usually very small, so that y_{it} is well modelled as a mildly explosive process over the relevant data. As we will see, such models are supported by the data for several macroeconomic time series in contemporary periods. Notice that stationary relationships among such series are still possible and in (2) we continue to allow for the possible presence of co-movement in such series. We do not use the term cointegration in this case because y_{2t} is not $I(1)$ and the individual series are not integrated processes. Instead, we will describe such processes as y_{1t} and y_{2t} as *co-motional*.

Reduced rank regression models of the type considered by Johansen (1988, 1991) and Ahn-Reinsel (1988, 1990) also come within the framework of (5). In these models the nonlinear parameterization $H = \Gamma J'$ is employed, where Γ $(m \times r)$ and J' $(r \times m)$ are factor loading and cointegrating matrices, respectively. The methods we develop below can be employed in the context of these reduced rank regression models, but in view of the nonlinear specification of the coefficient matrix H the justification for them is asymptotic. We will not use these parameterizations specifically in our empirical application below but do plan to include them in subsequent work and test them against some of the empirical formulations considered here. Notice that specifications such as (6) which allow for possibly explosive processes are not nested within the reduced rank regression format. This is because the coefficient matrix H in (6) is not necessarily of reduced rank (e.g. $B_1 = I_{m_1}$, $H_{22} > 0$) yet it accommodates a stationary relationship between y_{1t} and y_{2t} while allowing for nonstationary (in this case explosive) behavior in both y_{1t} and y_{2t}. Within the framework of (6) we now have the possibility that the co-motional series y_{1t} and y_{2t} actually become cointegrated during certain subperiods of data when the individual series are well modelled as integrated processes (i.e. $H_{22} = 0$). As emphasized in the Introduction, in our methodology specific models are supported on a period by period basis (conditional on past data), yet certain characteristics of the series (like co-motion or cointegration) may persist over many periods. Our approach is therefore to allow the model itself to evolve over time within a certain general class of potentially useful empirical specifications.

More general than the VAR(p) + Tr(r) model (5) is the vector ARMA(p, q) + Tr(r) system

$$(7) \qquad \Delta y_t = H y_{t-1} + \Sigma_{i=1}^{p-1} \Phi_i \Delta y_{t-i} + \Sigma_{j=1}^{q} \Psi_j \varepsilon_{t-j} + \Sigma_{k=-1}^{r} c_k r^k + \varepsilon_t \, .$$

In such models not all the regressors are observable and to fit these models within our framework we need recursive methods such as extended least squares -- see Hannan-Kavalieris (1984), Hannan-Deistler (1988) and Chen-Guo (1991). Phillips-Ploberger (1992) develop Bayes methods for treating scalar models of the form (7) by means of recursions of the type explored in the work of Hannan-

Rissanen (1982, 1983). Their methods can be extended to vector models of the class given by (7). However, we will confine ourselves in this paper to VAR(p) + Tr(r) systems in order to keep the calculations readily manageable within present PC computing facilities – specifically, under one hour of computing time on a 486-33 PC.

Our extension of the VAR(p) + Tr(r) system that fits within the above computational time frame is a model of this form with additional predetermined (i.e. $\mathscr{F}_{t-1}$-measurable) regressors other than lagged dependent variables or their differences. We may write such a model as

(8) $$\Delta y_t = Hy_{t-1} + \Sigma_{i=1}^{p-1} \Phi_i \Delta y_{t-i} + Fz_t + \Sigma_{j=-1}^{r} c_j t^j + \varepsilon_j ,$$

where z_t is $\mathscr{F}_{t-1}$-measurable. Models in this class we characterize as "VARZ(p) + Tr(r)" systems. By suitable definition of the filtration $\mathscr{F}_t$, they include the autoregressive leading indicator (ARLI) models of Garcia *et al.* (1987), Zellner-Hong (1989) and Zellner-Min (1992), in which case the vector z_t includes certain predetermined variables that are designated as leading indicators for y_t. Another instance of a model in the class (8) that is of potential empirical interest arises when z_t includes nonlinear transformations of certain elements of the vector y_{t-1}. Frequently in practice, the elements of y_t are certain nonlinear transformations of the raw data series. These transformations are often taken to achieve variance stabilizing objectives or approximate Gaussianity in the data. It is not always the case that linear cointegrating relations will apply to these transformed series. In such cases, when certain elements of y_{t-1} need to be reverse-transformed or subjected to a further transformation before they accord with a specified cointegrating relation they may be included in the vector z_t precisely in the form they are required. The cointegrating or co-motional relation (in the case of possibly explosive processes) is then embodied in an appropriate subvector of the component "$Hy_{t-1} + Fz_t$" of the model (8). For instance, if H has the form (6) then we may have

$$Hy_{t-1} + Fz_t = \begin{bmatrix} B_1 & B_2 \\ 0 & H_{22} \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} F_1 \\ 0 \end{bmatrix} z_t ,$$

and the linear relation $B_1 y_{1t-1} + B_2 y_{2t-1} + F_1 z_t$ is the cointegrating or co-motional component.

3. BAYES MODELS AND MEASURES

Let ε_t = iid $N(0, \Sigma)$ in (1) and let us assume for the time being that the error variance matrix Σ is known. Then the inference problem presented by (1) is linear in parameters and Gaussian. This facilitates an exact development of the Bayesian theory. Our approach to dealing with nuisance parameters like Σ in practice will be classical, i.e. we will replace Σ in our formulae with what would be a consistent estimator of Σ if the true data generating mechanism falls within the class of models given by (1) with a homogeneous error variance matrix. We recognize that not all Bayesians will be comfortable with the "mixture of paradigms" that this approach apparently involves. Nevertheless, it has the advantage of avoiding the specification of high dimensional nuisance parameter prior distributions and this offers potentially greater advantages in cases where transient dynamics are dealt with semiparametrically. (In the latter case we might, for example, use nonparametric spectral density matrix estimates rather than specify internal dynamics directly through a model such as a VAR. Many of the ideas presented here also apply in such semiparametric models. They will be pursued in later work.)

It is convenient to write model (1) with the restrictions (2) imposed as follows:

$$y_t = (I \otimes x_t')(S\alpha + s) + \varepsilon_t = W_t\alpha + (I \otimes x_t')s + \varepsilon_t$$

or

$$z_t = W_t\alpha + \varepsilon_t \tag{9}$$

where $z_t = y_t - (I \otimes x_t')s$. The maximum likelihood estimator (MLE) of α in (9) is

$$\begin{aligned} \tilde{\alpha}_n &= \left(\Sigma_1^n W_t'\Sigma^{-1}W_t\right)^{-1}(\Sigma_1^n W_t'\Sigma^{-1}z_t) \\ &= \left[S'(\Sigma^{-1} \otimes X_n'X_n)S\right]^{-1}[S'(\Sigma^{-1} \otimes X_n')\mathrm{vec}(Y_n') - S'(\Sigma^{-1} \otimes X_n'X_n)s] \end{aligned} \tag{10}$$

where $Y_n' = [y_1, \ldots, y_n]$ and $X_n' = [x_1, \ldots, x_n]$. We shall use the notation

$$A_n = A_n(\Sigma) = \Sigma_1^n W_t'\Sigma^{-1}W_t = S'(\Sigma^{-1} \otimes X_n'X_n)S \ . \tag{11}$$

It is also convenient to write (9) in the canonical form

$$\underline{z}_t = \underline{W}_t\alpha + \underline{\varepsilon}_t \tag{9'}$$

where $\underline{z}_t = \Sigma^{-1/2}z_t$, $\underline{W}_t = \Sigma^{-1/2}W_t$ and $\underline{\varepsilon}_t = \Sigma^{-1/2}\varepsilon_t \equiv$ iid $N(0, I_m)$.

Let P_n^α be the probability measure of Y_n. When $\alpha = 0$ we will denote this measure by P_n and use it as a reference measure in some of the analysis that follows. Conditional on $\mathcal{F}_0$ and α the density of Y_n with respect to Lebesgue measure (ν) is

$$\text{(12)}\quad \begin{aligned} \mathrm{pdf}(Y_n|\mathcal{F}_0, \alpha) = dP_n^\alpha/d\nu &= (2\pi)^{-nm/2}|\Sigma|^{-n/2}\mathrm{etr}\left\{-(1/2)\Sigma^{-1}(Y_n' - AX_n')(Y_n' - AX_n')'\right\} \\ &= (2\pi)^{-nm/2}|\Sigma|^{-n/2}\exp\left\{-(1/2)\Sigma_1^n(\underline{z}_t - \underline{W}_t\alpha)'(\underline{z}_t - \underline{W}_t\alpha)\right\} . \end{aligned}$$

Since $\tilde{\alpha}_n$ is the least squares estimator of α in (9′), the likelihood ratio process $L_n(\alpha) = dP_n^\alpha/dP_n$ has the simple exponential form

$$\text{(13)}\quad L_n(\alpha) = \exp\{(1/2)[\tilde{\alpha}_n'A_n\tilde{\alpha}_n - (\tilde{\alpha}_n - \alpha)'A_n(\tilde{\alpha}_n - \alpha)]\} .$$

When we combine (13) with a prior density $\pi(\alpha)$ for α we obtain the following joint density of (α, Y_n) conditional on $\mathcal{F}_0$:

$$\text{(14)}\quad \begin{aligned} \mathrm{pdf}(\alpha, Y_n) = \pi(\alpha)\mathrm{pdf}(Y_n|\mathcal{F}_0, \alpha) &= \left[(2\pi)^{-(nm-q)/2}|\Sigma|^{-n/2}|A|^{-1/2}\mathrm{etr}\left\{-(1/2)\tilde{E}_n'\tilde{E}_n\right\}\right] \\ &\cdot\left[\pi(\alpha)(2\pi)^{-q/2}|A|^{1/2}\exp\{-(1/2)(\tilde{\alpha}_n - \alpha)'A_n(\tilde{\alpha}_n - \alpha)\}\right] \end{aligned}$$

where $\tilde{E}_n' = [\tilde{\underline{\varepsilon}}_1, \ldots, \tilde{\underline{\varepsilon}}_n]$ and $\tilde{\underline{\varepsilon}}_t = \underline{z}_t - \underline{W}_t\tilde{\alpha}_n$. For $\pi(\alpha) = \pi_0 =$ constant this joint density gives rise to the marginal posterior density of α

$$\text{(15)}\quad \Pi_n(\alpha) = (2\pi)^{-q/2}|A_n|^{1/2}\exp\{-(1/2)(\tilde{\alpha}_n - \alpha)'A_n(\tilde{\alpha}_n - \alpha)\} \equiv N(\tilde{\alpha}_n, A_n^{-1}) .$$

Integrating α out of (14) we obtain the Bayesian data density for Y_n

$$\text{(16)}\quad \mathrm{pdf}(Y_n|\mathcal{F}_0) = \pi_0(2\pi)^{-(nm-q)/2}|\Sigma|^{-n/2}|A_n|^{-1/2}\mathrm{etr}\left\{-(1/2)\tilde{E}_n'\tilde{E}_n\right\} .$$

We now let Q_n be the measure whose density with respect to ν is (16). This measure is σ-finite, rather than a proper probability measure because the prior $\pi(\alpha) = \pi_0$ is improper. Similar calculations can, of course, be made for various proper prior densities $\pi(\alpha)$ in place of π_0 and one of these will be discussed below. The likelihood ratio of Q_n with respect to the reference measure P_n is given by the ratio

$$\text{(17)}\quad \frac{dQ_n}{dP_n} = \frac{dQ_n}{d\nu}/\frac{dP_n}{d\nu} = \pi_0(2\pi)^{q/2}|A|^{-1/2}\exp\{(1/2)\tilde{\alpha}_n'A_n\tilde{\alpha}_n\} .$$

At this point the analysis is formally identical to that of the univariate model considered in detail in Phillips-Ploberger (1992). The following results are therefore an immediate consequence of the analysis in that earlier paper.

3.1. THEOREM *Under the model* (1)-(2) *with the uniform prior* $\pi(\alpha) = \pi_0 = (2\pi)^{-q/2}$ *the Bayesian data density of* Y_n *conditional on* $\mathcal{F}_0$ *and* Σ *is given by*

(18) $$\frac{dQ_n}{dP_n} = |A_n(\Sigma)|^{-1/2}\exp\{(1/2)\tilde{\alpha}_n' A_n(\Sigma)\tilde{\alpha}_n\} .$$

The process dQ_n/dP_n *is a local* P_n*-martingale.* □

3.2. REMARKS

(i) The process dQ_n/dP_n is not itself integrable but it does have finite conditional expectation and satisfies the martingale property

$$E\left(\frac{dQ_n}{dP_n}|\mathcal{F}_{n-1}\right) = \frac{dQ_{n-1}}{dP_{n-1}} \quad \text{a.s. } (P_n) .$$

(ii) The uniform prior $\pi(\alpha) = \pi_0$ is a scale factor in (17). When $\pi_0 = (2\pi)^{-q/2}$, the density ratio dQ_n/dP_n is given by (18). There is an explanation for this choice of scale factor. Consider the scalar model (1) ($m = 1$) with $\Sigma = \sigma^2$, $s = 0$ and S a selection matrix of zeros and ones of full column rank q. Then, under P_n, $y_t = \varepsilon_t \equiv$ iid $N(0, \sigma^2)$. If x_t involves lagged dependent variables, then $X_n'X_n/n \to_{a.s.} \sigma^2 I_k$ under P_n, where $k = \dim(x_t)$. If we now standardize the ratio (18) so that non-trivial asymptotics obtain we have

$$2\ \ln[n^{-q/2}(dQ_n/dP_n)] = \tilde{\alpha}_n' A_n \tilde{\alpha}_n - \ln(n^q|A_n|) \to_d \chi_q^2 ,$$

since $n^{-q}|A_n| = |n^{-1}S'X_n'X_nS/\sigma^2| \to_{a.s.} |S'S| = 1$. Thus, under P_n twice the logarithm of the likelihood ratio (18) is χ_q^2 when scaled by $n^{q/2}$ to avoid degenerate asymptotics. Therefore, use of the prior $\pi_0 = (2\pi)^{-q/2}$ leads to a Bayes likelihood ratio test (to use the description in Phillips-Ploberger (1991)) that is asymptotically equivalent to the classical Wald test (based on $\tilde{\alpha}_n' A_n \tilde{\alpha}_n$) of the hypothesis $H_0 : \alpha = 0$ in (1). Use of the prior $\pi_0 = (2\pi)^{-q/2}$, therefore, makes Bayesian inference that is based on the posterior odds ratio dQ_n/dP_n asymptotically equivalent to classical inference based on the Wald, LM or LR tests of H_0. The same result applies in the multivariate model (1) when $s = 0$ and $S = I_{mk}$.

(iii) As indicated in (ii) above, the ratio dQ_n/dP_n in (18) is a likelihood ratio. As a Bayesian quantity it is simply the posterior odds in favor of the more complex model (1) over the same model with $\alpha = 0$. The odds ratio is computed using the Radon-Nikodym derivative of the measure Q_n with respect to the base measure P_n. In the univariate case, Phillips-Ploberger (1992) explores the use of this odds ratio as a tool for model selection and hypothesis testing. That paper also derives the model of the data for which Q_n is the associated measure. While Q_n is σ-finite (with an improper density with respect to ν) the conditional density given $\mathcal{F}_{n-1}$ (i.e. information in the given sample trajectory up to $n-1$) is proper provided there is sufficient data in $\mathcal{F}_{n-1}$ to determine $\tilde{\alpha}_n$, i.e. $S'(\Sigma \otimes X_n'X_n)S > 0$. This will be so if $nm > q$, when $\mathrm{rank}(X_n'X_n) = \min(k, n)$. As soon as the minimal data requirements for the definition of $\tilde{\alpha}_n$ (and, hence, dQ_n/dP_n) are met, the conditional density of Q_n given $\mathcal{F}_{n-1}$ can be derived and with it the model of the data that corresponds to the evolving Bayes measure Q_n. The calculations are analogous to those given in Phillips-Ploberger (1992), so we state only the result here.

3.3. THEOREM

(a) *The Bayesian conditional density of y_t given $\mathcal{F}_{t-1}$ is*

(19) $$dQ_t/dQ_{t-1} = \mathrm{pdf}(y_t|\mathcal{F}_{t-1}) = (2\pi)^{-m/2}|F_t|^{-1/2}\exp\{-(1/2)v_t'F_t^{-1}v_t\}\ ,\ t = [q/m]+1, [q/m]+2, \ldots$$

where $[q/m]$ is the smallest integer $\geq q/m$.

(b) *The "Bayes model" corresponding to this data density is*

(20) $$y_t = \Pi(\tilde{\alpha}_{t-1})x_t + v_t\ ,\quad \text{where}\quad v_t|_{\mathcal{F}_{t-1}} \equiv N(0, F_t)\ ,$$

that is

(21) $$\begin{aligned} E(v_t|\mathcal{F}_{t-1}) &= 0 \\ E(v_tv_t'|\mathcal{F}_{t-1}) &= F_t = \Sigma + (I \otimes x_t')SA_{t-1}^{-1}S'(I_m \otimes x_t) \\ &= \Sigma + W_tA_{t-1}^{-1}W_t'\ , \end{aligned}$$

and $\tilde{\alpha}_{t-1}$ is the MLE *of α in* (1)-(2) *based on information in $\mathcal{F}_{t-1}$.* □

3.4. REMARKS

(i) The conditional density (19) is forward looking in the sense that it is the density of y_t given $\mathcal{F}_{t-1}$ and it is defined as soon as there are enough observations in the trajectory to estimate the q-vector α by $\tilde{\alpha}_{t-1}$. Likewise, the "Bayes model" (20) is a predictive model in which, taking the conditional expectation with respect to Q_t measure, $E_{Q_t}(y_t|\mathcal{F}_{t-1}) = \Pi(\tilde{\alpha}_{t-1})x_t$ is the best Bayes estimate of the location of y_t given the information in $\mathcal{F}_{t-1}$. Observe that this location estimate is identical to the MLE of the best predictor $E_{P_t^\alpha}(y_t|\mathcal{F}_{t-1}) = \Pi(\alpha)x_t$, where the conditional expectation is taken with respect to P_t^α measure and α is the "true value" of the parameter in (1'). This MLE, viz. $\Pi(\tilde{\alpha}_{t-1})x_t$, is precisely the predictor we would use in classical inference. Thus, from the perspective of the model that is used to make predictions, there is no difference between the Bayesian and classical approaches. On the other hand, when a proper (or "informative") prior for α is employed then the posterior mean or Bayes estimate $\tilde{\alpha}_{t-1}^\pi$ of α will depend on the prior and will generally differ from $\tilde{\alpha}_{t-1}$. In such a case the "Bayes model" (20) will also change. When the prior $\pi(\alpha) = N(\bar{\alpha}, V_\alpha)$ is conjugate it is easy to show that the resulting "Bayes model" has the form

$$y_t = \Pi(\hat{\alpha}_{t-1}^\pi)x_t + v_t^\pi\ , \quad \text{where} \quad v_t^\pi|_{\mathcal{F}_{t-1}} = N(0,\ F_t^\pi) \tag{20'}$$

and

$$F_t^\pi = \Sigma + (I_m \otimes x_t')S\ \mathrm{var}(\hat{\alpha}_{t-1}^\pi)S'(I_m \otimes x_t)$$

where $\mathrm{var}(\hat{\alpha}_{t-1}^\pi)$ is the posterior variance of $\hat{\alpha}_{t-1}^\pi$. A simple calculation shows that the posterior mean and variance formulae needed for the above equations are:

$$\hat{\alpha}_n^\pi = \left[V_\alpha^{-1} + A_n\right]^{-1}[V_\alpha^{-1}\bar{\alpha} + A_n\tilde{\alpha}_n]\ ,$$

which is a matrix weighted average of the prior mean $\bar{\alpha}$ and the MLE $\tilde{\alpha}_n$; and

$$\mathrm{var}(\hat{\alpha}_n^\pi) = \left[V_\alpha^{-1} + A_n\right]^{-1}A_n\left[V_a^{-1} + A_n\right]^{-1}\ .$$

When $V_\alpha^{-1} = 0$, $\pi(\alpha)$ is uniform and (20') reduces to the "Bayes model" (20) given in Theorem 3.3.

(ii) As remarked in (i) the conditional mean of y_t under Q_t given $\mathcal{F}_{t-1}$ is the same as the classical ML estimate of the best predictor of y_t. The conditional variance matrix F_t in (21) is also identical to the classical formula for the variance matrix of the prediction error $y_t - A(\tilde{\alpha}_{t-1})x_t$. When x_t is fixed and non-random, the prediction error variance matrix is precisely F_t as given in (21). When x_t is a stochastic regressor with predetermined variables, then F_t in (21) is the asymptotic variance matrix of the forecast error. In either case, the formula for F_t corresponds to what would be used in classical inference.

(iii) "Bayes models" of the type (20) belong to a general class of data-oriented forecasting models wherein Bayesian updating procedures are used to revise the model on a period by period basis. Dynamic linear models that are updated by Kalman filter methods belong to the same category -- see Harrison-Stevens (1976) for an early development of this technique and West-Harrison (1989) for a recent review. All of these methods are recursive in the sense that the previous estimate or forecast is updated using the latest available observation. The present approach offers an advantage over traditional implementations of the Kalman filter in forecasting because associated with a forecasting model like (20) is the underlying Bayes measure Q_t. This measure can be used to compare alternative models, to test hypotheses about the model and to evaluate forecasts. Phillips-Ploberger (1992) and Phillips (1992a, b) develop a methodology of model selection, hypothesis testing and forecast encompassing for univariate models that is based on this approach. The ideas in those papers are extended here to multivariate systems. □

3.4. THEOREM

(a) *The* MLE $\tilde{\alpha}_t$ *is a local Q_t-martingale and satisfies*

$$E_{Q_t}(\tilde{\alpha}_t | \mathcal{F}_{t-1}) = \tilde{\alpha}_{t-1} , \quad \text{a.s.}$$

(b) *Under the Bayes measure Q_t, the distribution of $\tilde{\alpha}_t$ given $\mathcal{F}_{t-1}$ is*

$$\tilde{\alpha}|_{\mathcal{F}_{t-1}} = N(\tilde{\alpha}_{t-1}, A_{t-1}^{-1} - A_t^{-1}) .$$

(c) *The posterior distribution of α, viz.* $\Pi_t = N(\tilde{\alpha}_t, A_t^{-1})$, *is a local Q_t-martingale and satisfies*

$$E_{Q_t}[N(\tilde{\alpha}_t, A_t^{-1}) | \mathcal{F}_{t-1}] = N(\tilde{\alpha}_{t-1}, A_{t-1}^{-1}) , \quad \text{a.s.} \quad \square$$

3.5. REMARKS

(i) Theorem 3.4 shows that the MLE $\tilde{\alpha}_t$ and its posterior distribution, when it is treated as a Bayes estimate, have martingale-like properties under the Bayes measure Q_t. These martingale properties formalize the Bayesian updating process making it clear that they hold under the new measure Q_t.

(ii) When we carry out Bayesian inference on the model (1′) under a uniform prior, the probability measure of the data transforms according to the RN derivative (18), which is a consequence of the use of Bayes rule. Under the new measure the model is the predictive "Bayes model" (20). The optimal predictor of y_t in this model and under the new measure Q_t is simply the conditional mean given by $E_{Q_t}(y_t|\mathcal{F}_{t-1}) = \Pi(\tilde{\alpha}_{t-1})x_t$. Thus, the conventional Wiener-Kolmogorov prediction theory holds in this new (predictive model and predictive measure) context. Applying the Wiener-Kolmogorov prediction theory to the original model (1′) we have the optimal predictor $E_{P_t^\alpha}(y_t|\mathcal{F}_{t-1}) = \Pi(\alpha)x_t$. The two applications of this traditional theory of prediction give identical results when we replace the "true value" α in $A(\alpha)x_t$ by the best (regular) estimate of α based on information in $\mathcal{F}_{t-1}$, i.e. the MLE $\tilde{\alpha}_{t-1}$. □

4. MODEL CHOICE, TESTING AND ENCOMPASSING

4(a) Order Selection

Every "Bayes model" like (20) has associated with it its own σ-finite measure Q_t. Different models of the data may then be compared in terms of their respective "Bayes model" measures. This principle, put forward in Phillips-Ploberger (1992), is applicable to problems of order selection in multivariate models of the form (20).

We designate two competing "Bayes models" of the data using argument notation to signify the number of regressors as follows: the first model has k regressors

$$H(Q_n^k) : y_{n+1} = \tilde{\Pi}_n(k)x_{n+1}(k) + v_{n+1}(k)\ , \quad v_{n+1}(k)|_{\mathcal{F}_n} \equiv N(0,\ F_{n+1}(k))\ ;$$

and the second model has $K \geq k$ regressors

$$H(Q_n^K) : y_{n+1} = \tilde{\Pi}_n(K)x_{n+1}(K) + v_{n+1}(K) \ , \quad v_{n+1}(K)|_{\mathcal{F}_n} \equiv N(0, F_{n+1}(K)) \ .$$

Each of these models and the associated coefficient estimators has the properties described in Theorems 3.3 and 3.4 under their respective measures. Since the issue at first is one of order selection and concerns only the number of regressors we leave aside for the moment matters of coefficient restrictions like (2).

Pairwise evaluations of models like $H(Q_n^k)$ and $H(Q_n^K)$ can be conducted using the likelihood ratio of the measures associated with them. In this case, we have

$$\begin{aligned} dQ_n^k/dQ_n^K(\Sigma) = {} & |\Sigma^{-1} \otimes X_n(k)'X_n(k)|^{-1/2}\mathrm{etr}[(1/2)\Sigma^{-1}\tilde{\Pi}_n(k)X_n(k)'X_n(k)\tilde{\Pi}_n(k)] \\ & /|\Sigma^{-1} \otimes X_n(K)'X_n(K)|^{-1/2}\mathrm{etr}[(1/2)\Sigma^{-1}\tilde{\Pi}_n(K)X_n(K)'X_n(K)\tilde{\Pi}_n(K)] \end{aligned} \tag{22}$$

which measures support in the data for the more restrictive model $H(Q_n^k)$ against that of the more complex model $H(Q_n^K)$. Given that Σ is usually unknown we propose replacing it in (22) by $\tilde{\Sigma}_K$, the least squares estimate of Σ from the model with K regressors $H(Q_n^K)$. This leads to the order estimator

$$\hat{k} = \mathrm{argmin}_k \ \mathrm{PIC}_k \ ,$$

where

$$\mathrm{PIC}_k = dQ_n^K/dQ_n^k(\tilde{\Sigma}_K) \ . \tag{23}$$

By choosing $\hat{k}$ we are maximizing $1/\mathrm{PIC}_k$ over $k \leq K$ and thereby the model $H(Q_n^k)$ that is *a posteriori* most likely according to its data density given the available data. This criterion was explored in Phillips-Ploberger (1992) where some favorable simulation results for univariate autoregressive order selection is reported. In the univariate case, an alternative form of this criterion was given by Wei (1992) and called the Fisher information criterion (FIC).

4.1. THEOREM

(a) *The order estimator* $\hat{k}$ *minimizes*

$$\mathrm{tr}(\tilde{\Sigma}_K^{-1}SS_k) + \ln|\tilde{\Sigma}_K^{-1} \otimes A_n(k)| \tag{24}$$

where $SS_k = Y_n'Y_n - \tilde{\Pi}_n(k)A_n(k)\tilde{\Pi}_n(k)'$.

(b) *The criterion* (23) *is asymptotically equivalent to*

(25) $$\ln|\tilde{\Sigma}_k| + (1/n)\ln|\tilde{\Sigma}_K^{-1} \otimes A_n(k)| \ ;$$

and, for stationary VAR *systems (with $k = pm$ where p = number of lags),* (23) *is asymptotically equivalent to the* BIC *criterion*

(26) $$\mathrm{BIC}_p = \ln|\tilde{\Sigma}_{pm}| + m^2 p \ \ln(n)/n \ . \quad \square$$

4.2. REMARKS

(i) An alternative version of the PIC criterion is based on the ratio of the conditional data densities $Q_n(\cdot|\mathcal{F}_K)$ of the respective models over the same subsample of data, viz. $n > K$. Conditional on $\mathcal{F}_K$ we have

$$dQ_n^k/dQ_n^K(\Sigma) = \left(\Pi_{K+1}^n |F_t(K)|/|F_t(k)|\right)^{1/2} \exp\left\{(1/2)\Sigma_{K+1}^n [v_t(K)'F_t(K)^{-1}v_t(K) - v_t(k)'F_t(k)^{-1}v_t(k)]\right\} .$$

As before, we estimate Σ by $\tilde{\Sigma}_K$ giving the following alternative form of the PIC criterion

(27) $$\begin{aligned}\mathrm{PICF}_k &= dQ_n^k/dQ_n^K(\tilde{\Sigma}_K)|_{\mathcal{F}_K} \\ &= \left(\Pi_{K+1}^n |\tilde{F}_t(K)|/|\tilde{F}_t(k)|\right)^{1/2} \exp\{(1/2)\Sigma_{K+1}^n [v_t(K)'\tilde{F}_t(K)^{-1}v_t(K) - v_t(k)'\tilde{F}_t(k)^{-1}v_t(k)]\}\end{aligned}$$

where

$$\begin{aligned}\tilde{F}_t(k) &= \tilde{\Sigma}_K + W_t(k)\tilde{A}_{t-1}^{-1}(k)W_t(k)' \ , \\ \tilde{A}_t(k) &= A_t(\tilde{\Sigma}_K, k) = (\tilde{\Sigma}_K^{-1} \otimes X_t(k)'X_t(k)) \ .\end{aligned}$$

with a similar definition for $\tilde{F}_t(K)$. In $X_t(k)$, $W_t(k)$ and $A_t(\Sigma, k)$ we use the additional argument k to signify the number of regressors in that model (as in $x_t(k)$).

(ii) The PICF criterion (27) has an interesting interpretation as an encompassing test statistic. If $\mathrm{PICF}_k > 1$, the density for the model with k regressors exceeds the density of the model with K regressors when both are evaluated at the sample data. This can be restated by saying that the model with k regressors encompasses the model with K regressors in terms of their respective probability densities. This is called *distributional encompassing* over the sample $t \in [K+1, n]$ in Phillips (1992b).

(iii) We observe that in the present case

$$\hat{F}_t(k) = \tilde{\Sigma}_K + (I \otimes x_t')(\tilde{\Sigma}_K \otimes (X_{t-1}'X_{t-1})^{-1})(I \otimes x_t)$$
$$= \tilde{\Sigma}_K[1 + x_t'(X_{t-1}'X_{t-1})^{-1}x_t]$$

which is the usual formula for the forecast error covariance matrix in an unrestricted multivariate model. In this form it is clear that $\hat{F}_t(k)$ is invariant to linear transformations of the regressors. The same invariance therefore holds for the criterion PICF in (27). This is not true for PIC as given in (22) and (23).

4(b) Testing Unit Roots, Causality and Co-Motion

We will be concerned initially with hypotheses about the coefficient matrix Π in (1) that can be parameterized in the form $\Pi = \Pi(\alpha)$ as specified in (2), i.e.

$$H_0 : \mathrm{vec}(\Pi) = S\alpha + s\ , \quad \text{or} \quad H_0 : S_\perp'\mathrm{vec}(\Pi) = s_\perp$$

where $S_\perp$, whose dimension is $mk \times q_\perp$ with $q_\perp = mk - q$, is the orthogonal complement of S and $s_\perp = S_\perp's$.

A "Bayes model" test of H_0 is constructed by taking the RN derivative of the respective measures Q_n^π and $Q_n^\alpha = Q_n^{\Pi(\alpha)}$ of the two models

$$H(Q_n^\pi) : y_n = \tilde{\Pi}_{n-1}x_n + v_n^\pi\ ,$$

and

$$H(Q_n^\alpha) : y_n = \Pi(\tilde{\alpha}_{n-1})x_n + v_n^\alpha\ .$$

The likelihood ratio of the measures is

$$(28)\quad dQ_n^\pi/dQ_n^\alpha(\Sigma) = |A_n(\Sigma)|^{1/2}|\Sigma^{-1} \otimes X_n'X_n|^{-1/2}\mathrm{etr}[(1/2)\Sigma^{-1}\tilde{\Pi}_n X_n'X_n\tilde{\Pi}_n']\exp\{-(1/2)\tilde{\alpha}_n'A_n(\Sigma)\tilde{\alpha}_n\}$$

and the posterior odds test criterion is

(C1) *Accept $H(Q_n^\alpha)$ in favor of $H(Q_n^\pi)$ if $dQ_n^\pi/dQ_n^\alpha(\hat{\Sigma}) < 1$.*

Here $\hat{\Sigma}$ is the MLE of Σ under the general model (1).

An alternative form of the test statistic is given in the following and can be derived by straightforward regression algebra.

4.2. LEMMA

$$(29)\quad dQ_n^{\pi}/dQ_n^{\alpha}(\bar{\Sigma}) = |S_{\perp}'\{\bar{\Sigma}\otimes(X_n'X_n)^{-1}\}S_{\perp}|^{-1/2} \exp\Big\{(1/2)[S_{\perp}'\mathrm{vec}(\bar{\Pi})-s_{\perp}]'[S_{\perp}'\{\bar{\Sigma}\otimes(X_n'X_n)^{-1}\}^{-1}S_{\perp}][S_{\perp}'\mathrm{vec}(\bar{\Pi})-s_{\perp}]\Big\}\ . \quad \square$$

4.3. REMARKS

(i) In (29) the argument of the exponential is one half times the Wald statistic for testing the restrictions H_0. When H_0 is true this quantity is $O_p(1)$ as $n \to \infty$. However, if the persistent excitation condition $\lambda_{\min}(X_n'X_n) \to \infty$ a.s. is satisfied, the factor $|S_{\perp}'\{\bar{\Sigma} \otimes (X_n'X_n)^{-1}\}S_{\perp}| \to 0$ a.s. and (29) converges in probability to zero. On the other hand, when H_0 is false it is easily seen that (29) diverges. Hence the rule (C1) leads to a correct decision as $n \to \infty$ and the probability of both type I and type II errors goes to zero.

(ii) The statistic (29) can be used to test noncausality (where certain submatrices of Π have only zero elements), unit autoregressive roots (where certain elements of Π take on the value unity – e.g. the diagonal elements of $K = I+H$ in (5)) and co-motion (where in the system (5) the matrix H takes on the form specified in (6), i.e. nonzero off diagonal elements in any row of H prescribe the coefficients of co-motional series). Each of these hypotheses falls in the framework of H_0 and can be tested by means of the statistic (29) using the criterion (C1). We may wish to test these hypotheses individually, conditionally or jointly. The nature of the statistic makes it possible to compare one hypothesis with another as the next remark explains.

(iii) The statistic (29) assesses the evidence in the data in favor of the "Bayes model" $H(Q_n^{\alpha})$ versus that of $H(Q_n^{\pi})$, the unrestricted model. (The closer (29) is to zero the more we favor $H(Q_n^{\alpha})$.) When we have a group of competing hypotheses we can assess their odds against $H(Q_n^{\pi})$ and select that which the data most favors, i.e. select the "Bayes model" with the highest odds in its favor. In effect, if $H(Q_n^{\alpha_1})$ and $H(Q_n^{\alpha_2})$ are two competing explanations of the data then we simply compute

$$(30)\qquad \frac{dQ_n^{\alpha_1}}{dQ_n^{\alpha_2}}(\bar{\Sigma}) = \frac{dQ_n^{\alpha_1}}{dQ_n^{\pi}} \Big/ \frac{dQ_n^{\alpha_2}}{dQ_n^{\pi}}$$

and choose in favor of $H(Q_n^{\alpha_1})$ if this ratio exceeds unity, i.e.

(C2) $$\textit{Accept } H(Q_n^{\alpha_1}) \textit{ in favor of } H(Q_n^{\alpha_2}) \textit{ if } dQ_n^{\alpha_1}/dQ_n^{\alpha_2}(\bar{\Sigma}) > 1 .$$

(iv) In multivariate models like (1) the number of potential competing hypotheses is very large, even for moderate values of m and k. There is, therefore, some potential for the Bayesian equivalent of data mining in the use of the criteria (29) and (30), i.e. searching among large numbers of models for the one most favored by the data. This problem was discussed by Jeffreys (1961, p. 249 and pp. 253-254) in the context of using posterior odds to select from a large number of alternatives, and recently by Zellner-Min (1992) in the context of choosing regressors in time series models of the autoregressive leading indicator type. Jeffreys suggested a way of dealing with this "selection problem" by adjusting the prior odds to take into account the total number of alternatives. The result of such adjustment is that when testing a large number (say #) of (irrelevant) alternatives the prior odds should typically be set much higher than unity -- Jeffreys calculation based on independent alternatives leads to a prior odds of setting of 0.7#. When there are fewer alternatives that are of real interest and some theory to guide us in determining these alternatives, the need for such adjustments seems less urgent. In our empirical application later in this paper we employ recursive calculations of the odds ratio criterion so that the evidence in the trajectory in favor or against the alternatives is seen to evolve over time. The course of the recursive plot of the criterion then provides accumulated sample evidence which shows whether the odds are increasing or decreasing with the accrual of more data. This approach, we believe, helps to mitigate the effects of the "selection problem." Nevertheless, the role of the prior odds in determining the outcome of the test when there are many alternatives should always be borne in mind.

(v) Hypothesis testing using (29) is entirely analogous to the earlier problem of order selection using the PIC criterion (23) -- in the latter case the hypothesis is simply that of exclusion restrictions. In both cases the test statistic is based on the likelihood ratio of the respective Bayes measures of the competing models. This analogy, in which order selection and explicit hypothesis testing come within the same statistical paradigm, is especially interesting in the case of reduced rank regression models of the VAR type. In such models, like (5) above with a coefficient matrix H of possibly reduced rank, decisions need to be made on the VAR lag order (p) and the rank (r) of the coefficient matrix H. It is important to recognize in such models that the decision on r is an order selection problem just as that concerning p. These decisions are usually made sequentially first by determination of p and subsequently by determination of r -- as in the classical likelihood ratio procedures of Johansen (1988, 1991) and Ahn-Reinsel (1988, 1990). When the "Bayes model" methods discussed here are employed in this context the decision on the order (p, r) can be made simultaneously by choosing the order pair of the "Bayes model" whose odds are *a posteriori* the highest. The criterion in this case is very similar to (29) but needs to allow for the nonlinearity of the reduced rank hypothesis. The theory is in this case asymptotic. It will be reported and applied in later work.

(vi) Just as in the case of the problem of order selection there is also a predictive version of the criterion (29) that is based explicitly on the densities of the Bayes measures Q_n^π and Q_n^α. We condition on the same startup period (i.e. $n \geq k$) in order to initialize the MLE's $\tilde{\Pi}_n$ and $\tilde{\alpha}_n$ and hence the "Bayes model" measures. We then compare the respective densities of these measures over the same subsample of data $t \in [k+1, n]$. The criterion is analogous to (27) and has the form

$$dQ_n^\pi / dQ_n^\alpha(\tilde{\Sigma})\big|_{\mathscr{F}_k} = \left(\Pi_{k+1}^n |\tilde{F}_{t\alpha}| / |\tilde{F}_{t\pi}|\right)^{1/2} \exp\left\{(1/2)\Sigma_{k+1}^n [v_t^{\alpha\prime} \tilde{F}_{t\alpha}^{-1} v_t^\alpha - v_t^{\pi\prime} \tilde{F}_{t\pi}^{-1} v_t^\pi]\right\} \tag{31}$$

where

$$\tilde{F}_{t\alpha} = \tilde{\Sigma} + W_{t\alpha} \tilde{A}_{t-1\alpha}^{-1} W_{t\alpha}'$$

$$\tilde{A}_{t\alpha} = A_{t\alpha}(\tilde{\Sigma}) = \Sigma_{s=1}^t W_{s\alpha}' \tilde{\Sigma}^{-1} W_{s\alpha}$$

and

$$W_{t\alpha} = (I \otimes x_t') S$$

with corresponding definitions for the case of π-subscripted variables, except that $W_{t\pi} = I \otimes x_t'$. As with (29) we can employ the estimate $\tilde{\Sigma}$ from the unrestricted model $H(Q_n^\pi)$ in (31) and the variable definitions above. If (31) exceeds unity we conclude that the "Bayes model" $H(Q_n^\alpha)$ encompasses the unrestricted "Bayes model" $H(Q_n^\pi)$ in terms of their respective data densities over the sample $t \in [k+1, n]$. In other words, the within sample predictions of the model $H(Q_n^\alpha)$ outperform those of $H(Q_n^\pi)$ after controlling for the effect of the relevant restrictions in $H(Q_n^\alpha)$. (The latter is achieved through the presence of the first factor on the right side of (31) which embodies the penalties associated with the use of each model.)

4(c) Forecast Encompassing Tests

As pointed out earlier in Remarks 4.2 (i) and (ii), when it is written in its predictive density form the likelihood ratio criterion for competing "Bayes models" is a distributional encompassing test. The explicit form of the criterion given by PICF in (27) compares two models of different orders (k and K regressors, respectively) on the basis of their within sample predictive ability. Likewise (31) compares the restricted "Bayes model" $H(Q_n^{\alpha})$ with the unrestricted "Bayes model" $H(Q_n^{\pi})$ in terms of their respective predictive distributions.

This principle is easily extended to evaluate out of sample forecasts from rival models. Phillips (1992a, b) employed this idea in univariate models and we extend that approach here to multivariate models. Suppose, for instance, we wish to compare the models $H(Q_n^{\alpha})$ and $H(Q_n^{\pi})$ in terms of their one-period ahead forecast performance over the period $t = n+1, \ldots, N$. The forecast encompassing test statistic is

$$(32) \qquad dQ_N^{\pi}/dQ_N^{\alpha}(\tilde{\Sigma})|_{\mathcal{F}_n} = \prod_{t=n+1}^{N} (|\hat{F}_{t\alpha}|/|\hat{F}_{t\pi}|)^{1/2}\exp\left\{(1/2)[v_t^{\alpha\prime}\hat{F}_{t\alpha}^{-1}v_t^{\alpha} - v_t^{\pi\prime}\hat{F}_{t\pi}^{-1}v_t^{\pi}]\right\} ,$$

where

$$\hat{F}_{t\alpha} = \tilde{\Sigma}_t + W_{t\alpha}\hat{A}_{t-1\alpha}^{-1}W_{t\alpha}' ,$$
$$\hat{A}_{t\alpha} = A_{t\alpha}(\tilde{\Sigma}_t) = \Sigma_{s=1}^{t} W_{s\alpha}'\tilde{\Sigma}_t^{-1}W_{s\alpha} ,$$

and $\tilde{\Sigma}_t$ is the MLE of Σ from the unrestricted model $H(Q_n^{\pi})$ based on data up to the latest observation (t). Notice that in (32), as distinct from (31), we allow $\tilde{\Sigma}_t$ (and hence the forecast error variance matrices $\hat{F}_{t\alpha}$ and $\hat{F}_{t\pi}$) to evolve recursively over the forecast period. We conclude that $H(Q_n^{\alpha})$ encompasses $H(Q^{\pi})$ in terms of forecast performance over the period $t \in [n+1, N]$ when $dQ_n^{\pi}/dQ_n^{\alpha}(\tilde{\Sigma})|_{\mathcal{F}_n} < 1$.

If we wish to compare competing restricted models like $H(Q_n^{\alpha_1})$ and $H(Q_n^{\alpha_2})$ on the basis of forecasts over $[n+1, N]$ we can do so by relating the performance of these models to that of $H(Q_n^{\pi})$ as in (30), i.e. by multiplying the RN derivatives. We get

$$(33) \qquad dQ_N^{\alpha_1}/dQ_N^{\alpha_2}(\tilde{\Sigma})|_{\mathcal{F}_n} = [dQ_n^{\alpha_1}/dQ_n^{\pi}(\tilde{\Sigma})|_{\mathcal{F}_n}][dQ_n^{\pi}/dQ_n^{\alpha_2}(\tilde{\Sigma})|_{\mathcal{F}_n}] ,$$

where in each case we use the recursive estimate $\bar{\Sigma}_t$ from $H(Q_t^{\tau})$ in the computation of the forecast error variance matrix. When

$$dQ_N^{\alpha_1}/dQ_N^{\alpha_2}(\bar{\Sigma})|_{\mathscr{F}_n} > 1 \tag{34}$$

we conclude that the forecast performance of the two models favors $H(Q_N^{\alpha_1})$. The ratio (34) records the posterior odds in favor of $H(Q_N^{\alpha_1})$ from this forecast evaluation.

4(d) Evolving "Bayes Models" and Structural Change

"Bayes models" like $H(Q_t^{\alpha})$ are updated period by period as new observations become available. This means that the coefficients (i.e. $\bar{\alpha}_{t-1}$) and the forecast error variance matrix (i.e. $\hat{F}_{t\alpha}$) are revised each period for $t \in [n+1, N]$. We may also allow the model $H(Q_t^{\alpha})$ itself to evolve in the sense that the pattern of restrictions $H_0 : \text{vec}(\Pi) = S\alpha + s$ that are supported by the data may change over the period $[n+1, N]$. In particular, we may wish to ascertain whether the data continues to support over the full period the same model characteristics in the following categories:

(a) lag length and trend degree;

(b) number of unit roots;

(c) noncausality restrictions;

(d) co-motion of variables in the system.

In an evolving "Bayes model" we permit the characteristics (a)-(d) to be determined on a period by period basis. The outcome is a "Bayes model" sequence like $H(Q_t^{\alpha})$ but one in which the dimension of α is itself time dependent. We call this evolving "Bayes model" a sequence of best "Bayes models" because the model configuration in each period has the highest posterior odds according to (30) against rival models in the same class. We write this evolving sequence of best "Bayes models" as

$$H(Q_t^B) : y_t = \Pi(\bar{\alpha}_{t-1}, \hat{q}_{t-1})x_t + v_t^{\alpha}(\hat{q}_{t-1}) \tag{35}$$

with

$$E(v_t^{\alpha}(\hat{q}_{t-1})|\mathscr{F}_{t-1}) = 0 \ ,$$
$$E(v_t^{\alpha}(\hat{q}_{t-1})v_t^{\alpha}(\hat{q}_{t-1})'|\mathscr{F}_{t-1}) = F_{t\hat{q}_{t-1}} = \Sigma + W_{t\hat{q}_{t-1}}A_{t-1\hat{q}_{t-1}}^{-1}W_{t\hat{q}_{t-1}}'$$

and where the additional suffix $\hat{q}_{t-1}$ signifies that a "Bayes model" with $\dim(\alpha) = \hat{q}_{t-1}$ is selected using the criterion (30) after the $t-1$'th observation.

Having determined $H(Q_t^B)$ it is interesting to consider how the model configuration evolves over a subperiod of data such as $t \in [n+1, N]$. In our empirical application of these ideas in the next section we show how this can be done quite conveniently and economically by suitably designed graphics. When there are changes in the model configuration (such as an increase in the number of unit roots or a breakdown in a cointegrating or co-motional relationship) this can be interpreted as a form of structural change. The model sequence $H(Q_t^B)$ is therefore adaptive in the sense that the model dimension adapts to the needs of the data within the general framework of the model class (1)-(2) and the class of restrictions implied by the model characteristics (a)-(d) above.

It is of interest to compare the sequence $H(Q_t^B)$ with a fixed format "Bayes model" sequence, $H(Q_t^F)$ say, where the dimension of the model (F) is fixed. The comparison between these models can be made on the basis of their respective predictive densities over a forecast horizon such as $t \in [n+1, N]$. It will, for instance, often be interesting to use a fixed format VAR model, such as a VAR(4) + Tr(1) model, as a base model for comparison purposes. The forecast encompassing statistic for comparing these models is

$$(36) \qquad dQ_N^B/dQ_N^F(\tilde{\Sigma})\big|_{\mathcal{F}_n} = \prod_{t=n+1}^{N} \left(|F_{t\hat{q}_{t-1}}|/|F_{tF}|\right)^{1/2} \exp\left\{(1/2)[v_{t\hat{q}_{t-1}}^{\alpha'} \hat{F}_{t\hat{q}_{t-1}}^{-1} v_{t\hat{q}_{t-1}}^{\alpha} - v_t^{F'} \hat{F}_{tF}^{-1} v_t^F]\right\}$$

where the suffice "F" indicates that fixed format "Bayes model" formulae apply.

On the basis of their respective forecasting performance over $t \in [n+1, N]$, the "Bayes model" sequence $\{H(Q_t^B)\}_{n+1}^N$ is preferred to the fixed format model sequence $\{H(Q_t^F)\}_{n+1}^N$ if

$$dQ_N^B/dQ_N^F(\tilde{\Sigma})\big|_{\mathcal{F}_n} > 1 \; .$$

We would then conclude that forecasts from $H(Q_t^B)$ encompasses those of the fixed model $H(Q_t^F)$ over $t \in [n+1, N]$. As with the criteria (32) and (33), we employ the recursive estimates $\tilde{\Sigma}_t$ in (36). This error covariance matrix estimator can be obtained from the model $H(Q_t^B)$, since this estimator is consistent for Σ if (1) is the true model (given that a consistent model selection principle is being used in determining $H(Q_t^B)$), or it can be obtained form the general model $H(Q_t^T)$ in the class as in the case of

(32) and (33). We use estimates of Σ that are based on the model $H(Q_r^*)$ in our empirical application reported below.

5. AN EMPIRICAL APPLICATION TO A RUMPY MODEL OF THE US ECONOMY

5(a) US Macroeconomic Data

The data used are quarterly US macroeconomic time series taken from the Citibase data set over the period 1959:1-1992:1. Table 1 provides details of the five series that we use in this study and the variable notation that we employ. All variables except for the interest rate are seasonally adjusted and all series except for the interest rate are measured as natural logarithms. The interest rate variable is taken in reciprocals of levels, i.e. as 1/"level" where "level" is measured in % p.a. terms. The reciprocal transformation (i.e. $x \rightarrow 1/x$) is variance stabilizing and was found by the author in earlier work (1992a, b) to reduce the volatility in interest rates series that often occurs at higher interest rate levels. The series are graphed together in Figure 0 and the graphs of the T_Bill and 1/T_Bill series in this Figure clearly show the variance stabilizing property of the reciprocal transformation. Note that the methods of Section 4 continue to apply to models that include levels rather than reciprocals of levels in some equations. This can be achieved through the use of a "VARZ(p) + Tr(r)" model of the type (8) discussed earlier. This would be relevant if T_Bill levels (rather than reciprocals) were cointegrating or co-motional with other variables in the model. For the present application we did not, however, find it necessary to employ this extension.

To assist in evaluating the support in the data for various hypotheses and to conduct some forecast encompassing exercises we split the overall sample into two, giving: a sample period 1959:1-1983:4, and a forecast period 1984:1-1992:1. The division was made so that the sample period contained 100 observations, leaving 33 observations in the forecast period. No other sample split was tried. But clearly the results reported below would involve some changes contingent on the timing of this split. As it stands the forecast period of 33 observations is considered to be long enough to provide a fairly rigorous evaluation of the models and methods.

TABLE 1: Series Notation and Description

#	Variable	Description	Citibase	Sample Period	Forecast Period
1	T_Bill	3-month Treasury Bill avetion average discount (% pa; reciprocals)	FYGN3	1959:1-1983:4	1984:1-1992:1
2	M1	M-1: sum of currency, travellers cheques, demand deposits and other checkable deposits ($b, sa; logs)	FM1	1959:1-1983:4	1984:1-1992:1
3	GDP	Real GDP (1987$b, sa; logs)	GDPD	1959:1-1983:4	1984:1-1992:1
4	Un	Unemployment rate: all workers 16 years and over (%, sa; logs)	LHUR	1959:1-1983:4	1984:1-1992:1
5	PGDP	Implicit price deflator of GDP (1987 = 100, sa; logs)	GDPD	1959:1-1983:4	1984:1-1992:1

5(b) Univariate "Bayes Models" for the Series

Each series was considered in turn and univariate evolving "Bayes models" in the AR(p) + Tr(r) class were constructed using the PIC model selection criterion. The algorithm involved order selection of the autoregressive lag order (p) and trend degree (r) using (23) and subsequent use of the PIC criterion as in (29) and (C1) in order to determine whether the best "Bayes model" involved a unit autoregressive root. The procedure involves these steps:

STEP 1. Set maximum orders for the AR(p) and Tr(r) components. Here we used max_p = 6 and max_r = 1.

STEP 2. Use PIC as in (23 with an AR(max_p) + Tr(max_r) reference model to select the AR order ($\hat{p}$).

STEP 3. Select the trend degree ($\hat{r}$) using PIC as in (23) with the reference model AR($\hat{p}$) + Tr(max_r).

STEP 4. If $\hat{p} \geq 1$, compare the "Bayes model" AR($\hat{p}$) + Tr($\hat{r}$) with a "Bayes model" of the same order having a unit root in the autoregressive component. This is accomplished directly by using (29) and criterion (C2).

These steps are performed on a period by period basis over the forecast period 1984:1-1992:1. Steps 2 and 3 can, in fact, be combined in a single step by considering an array of regressions of the

AR(p) + Tr(r) form against the AR(max_p) + Tr(max_r) base model. This is, of course, more costly in computation time. Since this alternative led to only minor differences in outcome in a few periods and to almost identical forecasting results the simpler procedure involving the two steps is very useful, especially as an economy in computation in the vector case considered later.

The algorithm was applied to the five RUMPY series described above. In addition, one-period ahead forecasts over 1984:1-1992:1 were generated from the best "Bayes model" sequence $\{H(Q_t^B)\}_{t=1983:4}^{1991:4}$ and a sequence of fixed format models of the form AR(4) + Tr(1) over the same period. Figures 1-5 show the forecast performance of each of these model sequences. For each series there are six figures that display the following: the series and relevant forecast period (Figure (a)); the prediction errors from the rival models (Figure (b)); the characteristics (viz., lag and trend orders and whether or not there is a unit root) of the evolving "Bayes model" (Figure (c)); a recursive plot of the forecast encompassing test of the evolving model versus the fixed format model (Figure (d)); a recursive plot of the value of the long run autoregressive coefficient (Figure (e)); recursive 95% Bayes confidence bands around the forecasts from the evolving "Bayes model" plotted against the actual value (Figure (f)). In addition to these Figures, Table 2 tabulates some of the numerical details of these forecasting exercises and records the evolving form of the best "Bayes model."

TABLE 2: Univariate Forecasting Exercises 1984:1-1992:1

Series	Forecast RMSE's		Number of model changes	Best "Bayes model" 1992:1	Parameter count ratio "Bayes model"/ fixed model	Long run autoregressive coefficient 1992:1	Posterior odds in favor of unit root 1992:1	Forecast encompassing test dQ^B/dQ^F in 1992:1
	"Bayes model"	Fixed model						
T_Bill^{-1}	0.0115	0.0156	3	AR(3)$^{-1}$	3/6	0.9923	78.74016	106.4536
T_Bill*	0.5984	0.6841						
T_Bill†	0.5317	0.6483	2	AR(4)$^{-1}$	4/6	0.9915	68.4931	34.7545
M1	0.0110	0.0105	6	AR(2)$^{-1}$+T(1)	3/6	0.9738	1.3685	0.0213
GDP	0.0048	0.0047	4	AR(3)$^{-1}$	2/6	1.0005	3.7907	1.4832
Un	0.0228	0.0246	9	AR(3)$^{-1}$	2/6	0.9998	454.5454	3.5327
PGDP	0.0025	0.0027	4	AR(3)$^{-1}$	2/6	1.0004	769.2307	7.1438

Legend: * forecasts for the T_Bill series in levels were obtained from the models in reciprocals.
† forecasts for the T_Bill series in levels were obtained from models using levels.
AR(p)$^{-1}$ = AR(p) model with a unit autoregressive root.

The main results are as follows:

(i) All of the series are found to have a unit root at the end of the forecasting period 1992:1. The posterior odds in favor of the presence of a unit root range from 769:1 in the case of PGDP to 1.3685:1 in the case of M1. Unemployment and prices have unit roots throughout the period 1984:1-1992:1. For most of this period real GDP is found to have a mildly explosive long run autoregressive coefficient in preference to a unit root. Only in the last few periods, starting at the model for the 1990:4 quarter does the "Bayes model" convert to having a unit root. The model for real GDP changes to a unit root model during the 1990-1991 recession, resets back to a mildly explosive nonstationary model in the first part of 1991 and quickly changes again to a unit root model by the end of 1991. The "double dip" recession phenomena is captured by this double change in the model to a unit root model. As Figure 3(c) makes clear, when the model for real GDP changes to a unit root model the selected lag length increases from an AR(2) to an AR(3). Figure 3(e) shows the declining value of the long run AR coefficient from 1990:1 onwards and Figures 3(b) and (f) show that the biggest forecast error occurs at the onset of the 1990-1991 recession. Observe that with the model change in the evolving "Bayes model," the forecast error is smaller for the 1990:4 quarter than that of the fixed format AR(4) + Tr(1) model.

(ii) The better forecast performance of the evolving "Bayes model" for GDP over the recession period 1990-1991 leads to the sharp rise in the value of the forecast encompassing statistic dQ^B/dQ^F shown in Figure 3(d). By the end of the forecast period $dQ^B/dQ^F = 1.4832$, and thus at this point the odds in favor of the evolving "Bayes model" over the fixed format model are 1.48:1 on the basis of their forecast performance. Note that this is the case even though the RMSE over the entire period is slightly higher for $H(Q_n^B)$ than it is for $H(Q_n^F)$, i.e. 0.0048 against 0.0047.

(iii) The best "Bayes model" sequence encompasses the forecasts of the AR(4) + Tr(1) model also for the T_Bill, Un and PGDP series. In each of these cases the dominance is uniform throughout the period 1984:1-1992:1. It is especially strong for the T_Bill series, where $dQ^B/dQ^F = 106.4536$ and for PGDP, where $dQ^B/dQ^F = 7.1438$. The only series for which $H(Q^B)$ does not forecast-encompass $H(Q^F)$ is M1. For the sub-period 1986:1-1987:1 we have $dQ^B/dQ^F > 1$ for this series but otherwise the statistic is consistently below unity, indicating that there are features of the series that are useful for forecasting that are contained in the AR(4) + Tr(1) model but not in $H(Q_n^B)$. As is clear from Figure 2(c), $H(Q_n^B)$ undergoes several major changes in form, from a unit root model with trend (an $\text{AR}(1)^{-1}$ + Tr(1)) to a mildly explosive model with no trend (an AR(2)) and back to a unit root model with trend at the end of the period. This continual (and apparently fruitless) searching for a suitable model indicates that there may be no suitable model within the AR(p) + Tr(r) class. It is therefore of special interest to see how well this series is modelled when embedded in a multiple time series setting.

(iv) All of the evolving "Bayes models" are more parsimonious than the AR(4) + Tr(1) model. For three series (GDP, Un and PGDP) only a two parameter model is selected, compared with the six parameters in the AR(4) + Tr(1). Note that the presence of a unit root in the evolving "Bayes model" also serves to reduce the parameter count.

(v) In the graphs, Figure (b) shows the forecast errors from each model over the forecast period and Figure (f) shows 95% confidence bands around the forecast together with the actual series. The root mean squared error (RMSE) of forecast for each model is given in Table 2 and on Figure (b) for each series. For three series (T_Bill, Un, PGDP) the RMSE of $H(Q^B)$ is substantially better (by around 10% or more), for one series (GDP) the RMSE is about the same, and for M1 the RMSE of $H(Q^B)$ is about 5% worse than that of $H(Q^F)$. In the case of the T_Bill series the improvement in the forecasts from $H(Q^B)$ over those of $H(Q^F)$ is substantial, amounting to a 26% reduction in RMSE.

(vi) Interestingly, the unemployment series is found to have a unit root and the finding is robust over the period 1984-1992 as indicated in Figure 4(c). (Note that the model "changes" for this series amount only to a variation in the selected lag length between an AR(2) and an AR(3).) The unit root finding corresponds to that for data on unemployment from other countries (e.g. Phillips (1992b) for Australian data). It does not correspond, however, to results from the analysis of very long historical time series for the USA -- notably Nelson-Plosser (1982), and recently, using data-based methods related to those of the current paper, Phillips-Ploberger (1992) and Phillips (1992a). Thus, when taken in this context, the issue of whether a series is best modelled by a process with a unit root or can be viewed in terms of the series length and, hence, its initialization. The present finding for the series over the shorter time frame can be reconciled with that over the longer historical period by arguing that over the shorter time frame the unit root model is locally the better and more economical choice of model. This is certainly borne out here by its forecasting performance in relation to a model (viz. $H(Q^F)$ that admits a level stationary interpretation.

(vii) As discussed in Section 4(a) above, the T_Bill model is constructed in reciprocals of levels to help smooth out the volatility of the series over the sample. Forecasts for both reciprocals of levels (shown in Figure 1) and levels but generated from the model in reciprocals (shown in Figure 1') are given for this series. In addition we show results in Table 2 and Figure 1" for evolving "Bayes models" and fixed format models for this series in levels. Note that the evolving "Bayes model" in reciprocals is the more parsimonious (3 parameter versus 4) and has forecasting performance (in terms of RMSE) that dominate the fixed model in both cases (i.e. in both reciprocals and levels). Note also that according to the distribution theory of the evolving "Bayes model" the forecasts for the T_Bill series obtained using the model in reciprocals are median unbiased predictors but not optimal predictors in the conditional mean sense because the conditional expectation (i.e. $E_Q(\cdot|\mathcal{F}_{t-1})$) is not finite. This suggests that the RMSE is not a very satisfactory criterion for comparing the levels T_Bill forecasts when the model generating the forecasts is taken in reciprocals.

5(c) Multivariate "Bayes Models" for the Series

For multiple time series models the problem of model choice is substantially more complex than that for univariate series. This is not only a matter of there being an enormous number of potential choices, even for a five equation system such as that considered here. There is, in addition, the "selection problem" discussed in Remark 4.3 (iii) above. Thus, although it is possible to compare a very large number of alternative specifications using the criteria given in (29) and (30) and then to select that model with the highest odds against a general reference model (like a VAR(4) + Tr(1)), it is not at all clear that the use of a prior odds setting of unity is reasonable in the decision criterion (C2) when this large number of discriminating choices is being made. Even the computational demands become considerable by 486-33 PC standards when large numbers of such discriminating choices need to be made period by period in the construction of an evolving multivariate "Bayes model." It is important to recall at this point that the model characteristics to be investigated involve not only lag length and trend degree but also noncausality restrictions, numbers of unit roots, and comotional behavior among the levels of the series.

The class of multivariate models considered in the present application were of the form given by (5), i.e. VAR(p) + Tr(r) systems, which we write here as

(37) $$y_t = Ky_{t-1} + \Sigma_{i=1}^{p-1} \Phi_i \Delta y_{t-i} + \Sigma_{j=-1}^{r} c_j t^j + \varepsilon_t$$

with corresponding "Bayes models" of the form

(38) $$y_t = \Pi(\tilde{\alpha}_{t-1}) x_t + v_t^{\alpha}$$

as in (20), where the restrictions on the coefficient matrix Π are parameterized according to (2). To deal with some of the dimensionality problems presented by the system (37) we implement the following algorithm:

Step 1. Set initial maximum orders for the AR(p) and Tr(r) components in (37). Here we used max_p = 6 and max_r = 1.

Step 2. Use PIC as in (23) with a VAR(max_p) + Tr(max_r) reference model to select the VAR lag order $\hat{p}$ and trend degree $\hat{r}$. This gives a VAR($\hat{p}$) + Tr($\hat{r}$) base model [1].

Step 3. Compare ($\hat{p}$, $\hat{r}$) with the corresponding pairs ($\hat{p}_i$, $\hat{r}_i$) obtained for individual equations using the univariate series algorithm in Section 5(b). Construct a new base model with orders ($\hat{p}$, $\hat{r}$) determined by $\hat{p} = \max(\hat{p}, \max_i(\hat{p}_i))$ and $\hat{r} = \max(\hat{r}, \max_i(\hat{r}_i))$. But in the new base model the coefficient matrices Φ_j in (37) are diagonal for $j > \hat{p}$ and have non-zero diagonal elements $[\Phi_j]_{ii}$ only when $j \leq \hat{p}_i$. Similarly, when $j > \hat{r}$, $[c_j]_i$ is non-zero only when $j \leq \hat{r}_i$. This is the base model [2].

Step 4. Assuming that $\hat{p} > 1$, compare the base model [2] with "Bayes models" of the same order but with the levels coefficient matrix K taking the following forms and select the model with the highest posterior odds:

(39) $$K = \text{diag}(k_{11}, \ldots, k_{mm}) ,$$

(40) $$K = \text{diag}(k_{11}, \ldots, k_{mm}) \text{ and } k_{ii} = 1 ,$$

(41)
$$K = \begin{bmatrix} k_{11} & 0 & 0 & 0 & 0 \\ k_{21} & k_{22} & k_{23} & 0 & k_{24} \\ 0 & 0 & k_{33} & 0 & 0 \\ 0 & 0 & 0 & k_{44} & 0 \\ 0 & 0 & 0 & 0 & k_{55} \end{bmatrix},$$

(42)
$$K = \begin{bmatrix} 1 & 0 & & & \\ k_{21} & k_{22} & k_{23} & 0 & k_{24} \\ 0 & 0 & k_{33} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

STEP 5. Conduct non-causality tests of the following hypotheses

NC(i) Un does not cause M1,

NC(ii) M1 does not cause GDP.

These tests are accomplished by comparing "Bayes models" with and without the restrictions

(43)
$$[\Phi_j]_{24} = 0 \,, \quad [\Phi_j]_{32} = 0 \,, \quad \text{for all } \; j \le p$$

in terms of (29) and the criterion (C2). □

The above algorithm is designed to deal with dimensionality difficulties by utilizing information from univariate "Bayes models" in the specification of the lag order and trend degree. In addition, only three general classes of restrictions are considered. These relate to:

M(i) the presence of unit roots in individual equations of the model;

M(ii) the present of co-motional relationships in individual equations that are of interest from the point of view of relevant economic theory; and

M(iii) non-causality restrictions that may be suggested by theory.

In our empirical application we specify the coefficient matrix K in (41) and (42) to allow for a co-motional relationship between M1, GDP, PGDP and the T_Bill in the M1 equation. This relationship is tested against the unrestricted K, a diagonal K, a diagonal K with some unit roots and finally K as given in (42) which accommodates a co-motional relationship in the M1 equation and unit roots in the T_Bill, Un and PGDP series (as indicated in the univariate analyses of these series).

Additional specifications are possible, but (39)-(42) capture what seem to be the most relevant prior specifications of the levels coefficient matrix K, allowing for unit roots, mildly explosive or stable roots and potential co-motion among the variables M1, GDP, PGDP and the T−Bill. Notice that (41) and (42) place the long-run money demand relationship between M1, GDP, PGDP and the T_Bill in the M1 equation. This accords with the notion that in the long run the money stock does not depart too much from the demand for money that is determined by the level of real income, prices and the level of interest rates. In a more complex model we may wish to allow liquidity preference to rely on both a short run interest rate (as a form of own-return) and a long rate (as a cost of holding money). We may also wish to augment the interest rate equation to accommodate a response to any excess demand for money holdings by allowing levels of M1, GDP, and PGDP to appear in the interest rate equation. Extensions along these and other lines will be explored in later work with a larger model.

Step 5 in the algorithm incorporates tests of two important non-causality hypotheses. The first of these is the analysis of a potential policy effect of the unemployment rate on the money stock. To the extent that variables like real income influence interest rates and the demand for money holdings we would certainly expect some indirect effects (e.g. through real GDP) of the rate of unemployment on the money stock. The hypothesis NC(i) in Step 5 examines whether there is, in addition, any direct effect of the unemployment rate on M1, arising (for instance) from responses of the monetary authority to changes in the level of unemployment.

The second non-causality hypothesis NC(ii) in Step 5 involves the effects of M1 on real GDP. Causal relations between money and income have been much discussed in the empirical econometric literature [see Sims (1972), Feige-Pearce (1979), Hsiao (1979), Osborn (1984) and Boudjellaba *et al.* (1992), *inter alia*]. Some of this work, notably Hsiao (1979), has also given attention to issues of model selection that involve lag order choice in the VAR's in which the causality tests are to be conducted. In this respect, our methodology is more closely related to that of Hsaio's study than that of other work in this area. Recent econometric theory on classical hypothesis testing of non-causality by Sims-Stock-Watson (1991) and Toda-Phillips (1991) indicates that there are major difficulties (of non-standard limit theory and nuisance parameters in the asymptotics) in mounting asymptotically valid non-causality tests in possibly nonstationary VAR's in levels and in reduced rank regression ECM's. The Bayesian approach used here avoids those difficulties and will be pursued more systematically in later work. As far as our present application goes, it is straightforward to test NC(i) and NC(ii) by employing a "Bayes model" likelihood ratio test of the restrictions (43) using the RN derivative (39) of the respective "Bayes model" measures and the decision criterion (C2). Note that NC(i) and NC(ii) can be tested individually, jointly or even conditionally. For example, NC(ii) can be tested given the outcome of NC(i). This is, in fact, the way these tests will be conducted here.

The above algorithm was implemented on a period by period basis over the forecast horizon 1984:1-1992:1. Test criteria based on (29) were computed recursively over this period for each of the main model characteristics described by M(i), M(ii) and M(iii) above. One period ahead forecasts were generated from the following "Bayes models":

B(i) "Bayes model^{nc}" = "Bayes model" of the form (38) with no co-motional relationship and with unit roots incorporated as in the coefficient matrix (40) when supported by the data using the PIC criterion (29).

B(ii) "Bayes model$^{coint.}$" = "Bayes model" again of the general form of (38) with unit roots incorporated as determined by the data and the co-motional relationship in the M1 equation imposed for each period according to the specification of the K matrix given in (42).

B(iii) "Bayes model^{c+nc}" = "Bayes model" as in B(ii) but the co-motional relationship for M1 is incorporated only when it is supported by the data as determined by the PIC criterion (29).

B(iv) "Bayes model$^{caus.}$" = "Bayes model" with causal effects of Un on M1 and M1 on GDP included.

B(v) "Bayes model$^{diag.}$" = "Bayes model" with diagonal coefficient matrices and a general error covariance matrix.

In addition, forecasts were generated from several fixed format "Bayes models" for comparative purposes. These were

F(i) VAR(2) + Tr(−1): this is, in fact, the base model [1] determined by empirically in Step 2 of the algorithm. The lag order and trend degree were unchanged throughout the period 1984:1-1992:1.

F(ii) VAR(4) + Tr(0).

F(iii) VAR(4) + Tr(−1).

F(iv) VAR(4) + Tr(1).

Models F(ii) and F(iii) serve as good reference models for evaluating forecast performance because the lag order $p = 4$ and trend degrees $r = 0, 1$ are popular choices in empirical work with quarterly macroeconomic time series using unrestricted VAR's.

Finally, we will report forecast results obtained from the standard settings in the RATS package for forecasting with VAR's using Minnesota priors. We classify these models as follows

R(i) VAR(4) + Tr(0); with standard RATS (1989) package Minnesota priors.

R(ii) VAR(4) + Tr(1), with standard RATS (1989) package Minnesota priors.

The standard setting for the Minnesota priors is determined automatically (i.e. by default) in the RATS package. It involves a flat prior on the trend coefficients and independent normal priors for the autoregressive coefficients that are centered on $K = I$, $\Phi_j = 0$ in (37), i.e. a vector random walk.

The results for B(i), the "Bayes model^{nc}" are given in Figures 6-10. We graph the results for each series separately. Moreover, for each series we give the prediction errors (in Figure (a)), 95% Bayes confidence bands for the forecast (Figure (b)) and recursive plots of the unit root test for this equation (Figure (c)) and the value of the long run AR coefficient (Figure (d)). This combination of figures makes it easy to compare the results with those for the univariate series given earlier in Figures 1-6 and those for F(ii), the fixed format "VAR(4) + Tr(0)," which are shown in Figure (a) for each series.

The results for B(ii), the "Bayes model$^{coint.}$," are given in Figures 13-17, again for each series. In these figures we use F(i), the "VAR(2) + Tr(−1)," as the fixed format reference model. Figures 18-22 graph the prediction errors for B(iii), the "Bayes model^{c+nc}"; Figures 18-22 show the prediction errors for B(iv), the "Bayes model^{c+nc}"; and Figures 23-27 graph the prediction errors for B(v), the "Bayes model$^{diag.}$," against those of R(i), the Bayesian VAR(4) + Tr(0) with Minnesota priors.

Figures 11(a) and 11(b) give recursive plots for the causality tests (43) of Un on M1 and M1 on GDP. Figure 11(c) plots the test for co-motion among M1, GDP, PGDP and the T_Bill, and Figure 11(d) gives recursive plots of the estimated coefficients of GDP, PGDP and T_Bill variables in this relation.

5(c) Discussion

Evaluating the results of a multivariate forecasting exercise is far from straightforward and becomes more complex the greater is the number of different models under consideration. Here we have three main "Bayes models," viz. B(i)-(iii), three fixed format reference models, viz. F(i)-(iii), and two VAR models with Minnesota priors, viz. R(i)-(ii). Figures 6-10 and 13-22 give a detailed picture of the forecast performance of models B(i)-(iii) in relation to the two fixed models F(i) and F(ii). Since the results from F(iii) were very close to those of F(ii) we do not include them in the graphical displays. The results can be summarized as follows:

(i) As with the univariate analyses, unit roots (i.e. $k_{ii} = 1$ in (40) and (41)) were supported over the entire period 1984:1-1992:1 for the T_Bill, Un and PGDP equations -- see Figures 6(c), 9(c) and 10(c). For GDP the presence of a unit root was convincingly rejected by the criterion (29) and (C2) in favor of a mildly explosive long run AR coefficient -- see Figure 8(c), but note the steadily declining value of the recursive long run AR coefficient graphed in Figure 8(d). Again, this corresponds closely with the univariate results except for the latter model's switch to a unit root for GDP during the 1990-1991 recession. For the M1 equation the outcome is rather different from the univariate case. As seen in Figure 7(c), the presence of a unit root in the M1 equation is rejected until the 1989:2 quarter, when it is accepted for two quarters. After this the unit root is marginally rejected until the last few quarters. This outcome corroborates the evidence from the univariate analysis of some uncertainty over the specification of the M1 equation, especially for the latter part of the forecast period. Looking at the data for M1 in Figure 7(b) it is apparent that there was a substantial slowdown and fluctuation in M1 growth over 1987:1-1989:4. Both the univariate and the multivariate models sustain some large forecast errors in this period. For instance, in 1989:1 actual M1 is outside the 95% confidence band for the forecast, as is apparent in Figure 7(b). Comparing the prediction errors shown in Figure 2(a) with those in Figure 7(a), it is clear that there are substantial gains in forecast accuracy in moving from the univariate to the multivariate model. The RMSE of the "Bayes model^{nc}" forecasts of M1 is 0.0091, compared with a RMSE of 0.0110 for the univariate "Bayes model" forecast of M1. Use of the multivariate model leads to a substantial reduction (20%) in the RMSE of forecast and therefore suggests that across-variable dependence is important for modelling this series.

(ii) Figure 11(c) shows the results for the test of co-motion or cointegration in the M1 equation. This particular test is based on a direct comparison of the "Bayes model" with a level coefficient matrix K as given in (42) and with an intercept in (37) of the form

$$c_0' = [0, c_{02}, 0, 0, 0]$$

(thereby allowing a non-zero intercept in the M1 equation), and the corresponding "Bayes model" with no co-motion, i.e. with $k_{21} = k_{23} = k_{24} = c_{02} = 0$. The recursive plot of the logarithm of the test statistic (29) is shown in Figure 11(c). Figure 11(d) shows recursive plots of the corresponding fitted values of the co-motional or cointegrating coefficients. The results are quite striking. The coefficients undergo dramatic swings in the short sub-period 1986:3-1987:4, e.g. from large positive values to large negative values and back again to positive values in the case of the real GNP and T_Bill coefficients. From the graph of the series given in Figure 7(b) it is apparent that this is the period when M1 growth sustains the slowdown mentioned earlier. Evidence in support of co-motion in the M1 equation (as given by the dQ_t/dQ_t^0 ratio shown in Figure 11(c)) is quite strong until this period, when it starts to fall quickly. By the 1987:4 quarter, the test leads to a rejection of co-motion in the M1 equation. Following the 1987:4 quarter, the estimated coefficients appear to stabilize again and this is maintained until the final quarter 1992:1. The test statistic continues to reject co-motion until the 1991:1 quarter after which $dQ_t/dQ_t^0 > 1$ until the end of the period. These results indicate that there is certainly some instability in an M1 equation with levels of the other variables (GDP, PGDP and T_Bill) as regressors. Moreover, if these variables are relevant in determining M1 then it would seem useful to have some mechanism for switching their effects off when the regression becomes unstable. This is precisely what the evolving "Bayes model" $H(Q_t^B)$ is designed to achieve.

We will examine the effects of the switching mechanism in this case later on when we consider the forecast performance of models B(ii) and B(iii).

(iii) Forecasts for the T_Bill from the "Bayes model^{nc}" (i.e. B(i)) are broadly the same in terms of RMSE as those of the best univariate "Bayes model" – compare Figures 1(b) and 6(a). However, as is clear in Figure 6(a), the fixed VAR(4) + Tr(0) model has substantially larger forecast errors in many periods and a RMSE (0.0163) that is 38% bigger than that of the "Bayes model^{nc}" (viz. 0.0118). Another variable for which the "Bayes model^{nc}" generates much better forecasts than the fixed format VAR models is GDP. Here the RMSE for the VAR(4) + Tr(0) model (0.0085) is 51% larger than that of the "Bayes model^{nc}" (0.0056). This is a very substantial difference and the improvement in forecasting performance of the "Bayes model^{nc}" is clearly visible in Figute 8(a). A similar comment applies to the unemployment rate forecasts. Here again the "Bayes model^{nc}" forecasts substantially improve on those of the fixed models VAR(4) + Tr(0) and VAR(2) + Tr(−1)

(iv) Figures 11(a) and 11(b) give the outcomes of the non-causality hypothesis tests NC(i) and NC(ii). The tests are mounted by first implementing the "Bayes model" test of NC(i) followed by the corresponding test of NC(ii) given the outcome of the test of NC(i). In the case of NC(i): no causal effects of Un on M1, we see that the odds are decidedly in favor of non-causality and this is maintained throughout the period 1984:1-1992:1. With regard to NC(ii): no causal effects of M1 on GDP, we see that the data does not favor NC(ii) until the 1986:3 quarter, from which point the odds are clearly in favor of non-causality from M1 to GDP. A joint test of hypotheses NC(i) and NC(ii) taken together led to acceptance of non-causality of Un on M1 and M1 on GDP throughout the entire period. This hypothesis was therefore incorporated into the evolving "Bayes model^{nc}."

(v) Figures 13-17 show the prediction errors for the five series that are obtained from model B(ii): the "Bayes model$^{coint.}$." In this model a co-motional relation in the M1 equation is included for every quarter of the forecasting period, i.e. the results of the recursive test of co-motion shown in Figure 11(c) are *not* incorporated in the construction of the sequence of "Bayes models." The forecast performance of this model is very similar to that of B(i): the "Bayes model^{nc}." Relative to B(i) there is a slight deterioration in the forecasts of the M1 and GDP series, a slight improvement in the forecasts of Un, and a definite improvement in the forecasts of PGDP (the RMSE falls 20% from

0.0032 for this series for B(i) to 0.0025 in B(ii)). Thus, the main effect of imposing co-motion is on the PGDP series, not M1 or GDP. Looking at Figure 17 (for the B(ii) model) in comparison with Figure 10(a) (for B(i)) it is clear that the improvement in forecast accuracy occurs over the whole period, with the exception of the last 2 or 3 observations. We will assess whether the overall improvement in forecasting performance from moving to B(ii) from B(i) is significant in Section 5(d) below.

(vi) Figures 18-22 give the prediction errors for the series from model B(iii): the "Bayes model$^{\text{coint.+non-coint.}}$." In this model we employ a switching mechanism that is triggered by the test for co-motion or cointegration reported in Figure 11(c). This mechanism determines whether the evolving "Bayes model" incorporates co-motion in the M1 equation or not. As is apparent from the recursive plot in Figure 11(c), the switch from a "cointegrated model" to a "non-cointegrated model" occurs several times over the course of the forecast period. The forecast results from B(iii) are, however, quite similar to those of B(ii). The GDP and Un series predictions are slightly better by overall RMSE for B(iii) than they were for B(i). Forecasts for the T_Bill series have the same RMSE, while those for PGDP and M1 lead to a slight increase in the RMSE. Thus, real variable forecasts improve a little by implementing the switching mechanism but the monetary variable forecasts deteriorate slightly in comparison with B(ii) where the co-motional relation in the M1 equation is sustained throughout the period.

5(d) Forecast Encompassing

Forecast encompassing tests of the type discussed in Section 4(c) were conducted to assess whether on the basis of actual forecast performance the realized posterior odds favored one model or another. The models compared were the following

F(i): VAR(2) + Tr(−1)

F(iv): VAR(4) + Tr(1)

B(i): "Bayes model$^{\text{nc}}$"

B(ii): "Bayes model$^{\text{coint.}}$," i.e. Bayes with cointegration or co-motion included

B(iv): "Bayes model$^{\text{caus.}}$," i.e. "Bayes model" with causal effects included

The results are shown in Figures 12(a)-(d) in a series of recursive plots of the forecast encompassing test statistics given in (32), (33) and (36).

Figures 12(a)-(b) plot the logarithm of the statistic dQ^B/dQ^F given in (36) comparing the forecasts of the model B(i) against those of the two fixed models F(iv) and F(i). The evidence from these tests overwhelmingly favors B(i), leading us to conclude that, on the basis of their respective forecast performance, the odds in support of the model B(i) over either F(iv) or F(i) are more than e^{10} : 1. Clearly model F(i) which is the base model [1] that is determined in Step 2 of the algorithm by choosing lag length and trend degree in the VAR, dominates F(iv). Thus, model choice even in terms only of lag length and trend degree leads to improved forecasting performance. But the evidence in favor of model B(i) over F(i) shown in Figure 12(b) shows that substantial gains in forecasting accuracy are still possible after the choice of lag length and trend degree.

Figure 12(c) gives a recursive plot of the forecast-encompassing test of B(i) against B(iv). This test compares two "Bayes models" and is constructed using (33). The more general model in this case is B(iv) because it allows for potential causal effects from Un to M1 and M1 to GDP to be included. As is apparent from the figure, $dQ^{B^{nc}}/dQ^{B^c} > 1$ throughout the forecast period, although the margin is not great early on in the period. Thus, the evidence from forecasts obtained with these two models favors B(i), i.e. the model with no causal effects from Un to M1 and M1 to GDP. This outcome corroborates the earlier findings from the direct tests of the non-causality restrictions that the data supports these restrictions (refer to Figures 11(a) and 11(b) and the discussion in Section 5(c), Remark (iv)).

Finally, Figure 12(d) plots the forecast encompassing test of model B(i) against B(ii). The recursive plot of the ratio $dQ^B/dQ^{B^{coint}}$ (again, constructed using formula (33) for the two models) shows that the forecast performance favors B(ii), i.e. "Bayes model$^{coint.}$," early in the forecast period. But, from quarter 1985(ii) onwards, $dQ^B/dQ^{B^{coint}} < 1$ and the odds are subsequently in favor of the model B(i) on the basis of the forecasting performance of the two models. By the end of the period the odds clearly support model B(i). Interestingly B(i) dominates B(ii) on the basis of this criterion even though the model B(ii) leads to improved predictive accuracy over B(i) for some series (viz. PGDP and Un) as discussed in Section 5(c), Remark (v). This is because model B(ii) has more parameters than B(i) and the encompassing statistic $dQ^{B}/dQ_{B^{coint}}$ evaluates the forecast errors produced from the two models in terms of their respective forecast error covariance matrices, which in turn depend on the number of estimated parameters. Thus, although B(ii) does improve predictive accuracy for some series (but *not* all), the improvement needs to be significant relative to the number of additional parameters that are employed in the model. According to the encompassing test, this is not so by the end of the forecast period and, indeed, we conclude that forecasts from B(i) encompass those of model B(ii). A similar conclusion is reached for model B(iii): The "Bayes model$^{coint.+non\text{-}coint.}$," although we do not report the graphics for this case.

5(e) Summary Results

Table 3 summarizes details of the forecasting accuracy of all of the main models that were considered as serious competitors. These include: the univariate "Bayes models" (from Section 5(b)); the VAR "Bayes models" B(i), B(ii), B(iii); the fixed format models F(i), F(ii), F(iv); and the two Bayesian VAR models with standard (i.e. RATS-package implemented) Minnesota priors R(i) and R(ii).

To assist in making a global comparison across all five series we computed the following total relative RMSE criterion defined as follows:

$$TRE = \prod_{i=1}^{5} RMSE_i / RMSE_i[VAR(4) + Tr(1)] \ ,$$

where

$RMSE_i$ = root mean squared error of forecast for series i,

$RMSE_i[VAR(4) + Tr(1)]$ = root mean squared error of forecast for series i from the base model "VAR(4) + Tr(1)".

The statistic TRE is calculated for each of the rival models and uses the RMSE's from the fixed format "VAR(4) + Tr(1)" as a base model for comparison. In effect, the base model RMSE's provide the units for comparing the RMSE's of the five different series on which the TRE statistic depends. Comparisons across models using the TRE criterion are then straightforward. We prefer models with smaller TRE values -- the closer TRE is to zero the more that model's forecasts dominate those of the base model (for which TRE = 1).

The results from Table 3 are quite clear. The three VAR "Bayes models" *and* the univariate "Bayes models" are all preferred to the fixed format VAR models and the BVAR's with Minnesota priors (BVARM's). As is apparent from the table, the BVARM's do improve on the forecasting performance of fixed format models with the same number of parameters. For instance, TRE[for VAR(4) + Tr(0)] = 0.7663 whereas TRE[for BVARM(4) + Tr(0)] = 0.4409. This accords with earlier findings by Litterman (1986) and others in support of the BVARM method. However, Table 3 also makes it clear that much greater gains in forecasting accuracy can be obtained using the "Bayes model" methodology. All of the "Bayes models" give striking improvements in forecast accuracy over BVARM's. For example, model B(ii): the "Bayes model$^{coint.}$" with cointegration imposed has TRE("Bayes model$^{coint.}$") = 0.1985 whereas that of model R(i): "BVARM(4) + Tr(0)" is TRE["BVARM(4) + Tr(0)"] = 0.4409, which is more than twice as much.

TABLE 3: Multivariate Forecasting Exercises: RMSE's over 1984:1-1992:1

Series	Bayes model^{nc}	Bayes model^{c}	Bayes model^{nc+c}	Bayes model$^{diag.}$	Univariate Bayes model	VAR(2) + T(−1)	VAR(4) + T(0)	VAR(4) + T(1)	BVAR(4) + Tr(0) with Minnesota priors	BVAR(4) + Tr(1) with Minnesota priors
T_Bill	0.0118	0.0120	0.0120	0.0115	0.0115	0.0172	0.0163	0.0163	0.0164	0.0168
M1	0.0091	0.0092	0.0095	0.0098	0.0110	0.0096	0.0088	0.0091	0.0104	0.0106
GDP	0.0056	0.0057	0.0056	0.0052	0.0048	0.0065	0.0085	0.0085	0.0055	0.0063
Un	0.0227	0.0226	0.0225	0.0237	0.0228	0.0241	0.0341	0.0355	0.0263	0.0272
PGDP	0.0032	0.0025	0.0028	0.0024	0.0025	0.0034	0.0033	0.0040	0.0032	0.0033
TRE	0.2439	0.1985	0.2151	0.1862	0.1933	0.4912	0.7663	1.0000	0.4409	0.5625

Legend: TRE = Total relative error = $\Pi_{i=1}^{5}$ $RMSE_i/RMSE_i$(VAR(4) + Tr(1)).
Bayes model^{nc} = evolving "Bayes model" with no co-motion.
Bayes model^{c} = evolving "Bayes model" with co-motion in M1 equation.
Bayes model^{nc+c} = evolving "Bayes model" with model switch to co-motion when favored by the data.

Interestingly, the univariate "Bayes model" (i.e. the "Bayes model" constructed for the individual series) does very well in terms of TRE in these summary comparisons. Looking at the RMSE's for the individual series, it is apparent that the gains over B(ii), for example, come in forecasting the T_Bill and GDP series. Note that the univariate "Bayes model" for M1 does much worse than B(ii), having a RMSE = 0.0110 against that of 0.0092 for B(ii). However, this is not enough to offset the gains from the improved forecasts of the T_Bill and GDP series. Overall, TRE[univariate "Bayes models"] = 0.1933 whereas TRE[B(ii)] = 0.1985. The difference is small, but it indicates that there are likely to be some potential improvements in our algorithm for choosing the best VAR "Bayes model." This is also borne out by the results for B(v), the "Bayes model$^{\text{diag.}}$", which has the smallest TRE (0.1862) relative to the base model. Use of the general error covariance structure in B(v) improves the forecasts of M1 and PGDP over those from the univariate model, gives identical RMSE results for the T_Bill series but leads to a slight deterioration in the GDP and Un series forecasts relative to the univariate "Bayes models." On the other hand, as is apparent from Figures 23-27 B(v), the "Bayes model$^{\text{diag.}}$," outperforms R(i), the Bayes model VAR(4) + Tr(0) with Minnesota priors, for each of the series. The improvement in forecast performance is particularly evident in the case of the T_Bill and PGDP series.

6. CONCLUSIONS

The methodology of this paper involves the use of data-oriented forecasting models that are revised period by period using a Bayesian updating process. The model revisions apply to the parameters and to the dimension of the parameter space, so that the selected "Bayes models" have characteristics that are temporally dependent and may therefore evolve with the data over time. In building Bayesian models of this type for trending multiple time series we take into account a variety of characteristics including lag length and trend degree, the number of unit roots, causality effects and potential co-motion among the variables in the system. Model selection, hypothesis testing and even forecast evaluations are all conducted by the same principle, which utilizes the likelihood ratio of the "Bayes measures," conditional on the trajectory that has already occurred, of the rival "Bayes models." From this perspective, our approach has both a Bayesian and a classical interpretation.

The algorithm that implements the methodology in the multivariate case allows the use of results from univariate "Bayes models" (for instance, on lag length or trend degree), and economic theory can be used to guide the configuration of model characteristics that are to be considered each period in competing models of the data. The class of models considered in our empirical application of these ideas is the "VAR(p) + Tr(r)" system. Since we allow for nonstationary series, including both integrated and possibly mildly explosive processes, the VAR's are permitted to have "structural" components (i.e. relations between the nonstationary levels of some of the series) as well as nonstationary (unit root or mildly explosive) and transient components.

An earlier paper (1992a) by the author put forward a precept for empirical work called the Exxon Valdez principle to the effect that fixed format time series models have difficulty adapting quickly to new data (just as big tankers cannot stop or turn corners in a hurry). The principle is especially relevant in time series models with deterministic and stochastic trends. Deterministic trends are powerful elements in determining the course of time series predictions, just as their associated regression coefficients have faster rates of convergence than that of stationary regressors. However, when deterministic trends are inappropriately included as regressors there is a far greater potential of poor forecasting performance. This was illustrated in the exercises conducted in the earlier paper (1992a). Similar effects also operate in multivariate models but here the role of stochastic trends also becomes important. For in a multivariate model there is also the possibility of inappropriately including the level of a nonstationary variable as one of the regressors in the model. In an unrestricted VAR, every equation contains the levels of all other variables and the risk of sustaining poor forecasts through the inappropriate inclusion of regressors with stochastic trends is thereby greatly increased. The problem arises also in Bayesian VAR's with Minnesota priors, for although all the estimated across-variable effects are shifted towards a prior mean of zero in this case, some of the effects of trending (and potentially spurious) regressors remain. Moreover, since the default Minnesota priors are flat on the deterministic trend components, the influence of these regressors is not diminished by the prior. In fact, it is actually enhanced because the Minnesota priors pull the estimated model towards a vector random walk and in such a model an intercept and linear trend serve the role of a linear drift and a *higher order* (quadratic) trend, respectively. These arguments suggest that Bayesian VAR's with Minnesota priors are exposed to the risk of poor forecasting performance from the mechanical treatment of trending regressors (and even intercepts) in much the same way as fixed format VAR models.

The forecasting exercises undertaken in this paper bring empirical evidence to bear in support of the above arguments. Fixed format VAR models and Bayesian VAR models with Minnesota priors all perform poorly in the forecasting exercises here in comparison with "Bayes models" that explicitly incorporate model selection principles. Our model selection procedures provide no evidence for deterministic trends in our Bayesian VAR models and, further, do not support the presence of intercepts in those regressions either. These are strong conclusions, especially when taken in the context of present applied econometric practice with VAR's. Nevertheless, the forecasting results with our "Bayes models" provide some hard evidence that these model choices are correct, in the explicit sense that they lead to improved forecasting performance. Outside the context of this forecasting exercise with RUMPY data, it seems fair to conclude that the presence of intercepts and trends should be subject to more careful scrutiny than seems heretofore to have been the case. One of the features of the model selection methods given here is that they can serve a role in helping to determine whether such regressors should be included in time series regression models. We hope to report further empirical evidence on these issues with larger models and with data from other countries at a later date.

7. APPENDIX

7.1. PROOF OF THEOREM 3.4

Applying recursive least squares to the model in transformed form (9′) we have

$$\begin{aligned}
\tilde{\alpha}_t &= \tilde{\alpha}_{t-1} + \left(\Sigma_1^t \underline{W}_s' \underline{W}_s\right)^{-1} \underline{W}_t'(\underline{z}_t - \underline{W}_t \tilde{\alpha}_{t-1}) \\
&= \tilde{\alpha}_{t-1} + A_t^{-1} W_t' \Sigma^{-1}(y_t - A(\tilde{\alpha}_{t-1})x_t) \\
&= \tilde{\alpha}_{t-1} + A_t^{-1} W_t' \Sigma^{-1} v_t \ .
\end{aligned}$$

Since $E_{Q_t}(v_t|\mathcal{F}_{t-1}) = 0$, (a) follows directly. We also deduce from this expression the conditional density

$$\tilde{\alpha}_t|_{\mathcal{F}_{t-1}} \equiv N(\tilde{\alpha}_{t-1}, A_t^{-1} W_t' \Sigma^{-1} F_t \Sigma^{-1} W_t A_t^{-1}) \ ,$$

and noting that

$$\begin{aligned}
A_{t-1}^{-1} &= \left(A_t - W_t' \Sigma^{-1} W_t\right)^{-1} \\
&= A_t^{-1} + A_t^{-1} W_t' \left(\Sigma - W_t A_t^{-1} W_t'\right)^{-1} W_t A_t^{-1} \\
&= A_t^{-1} + A_t^{-1} W_t' \left[\Sigma^{-1} + \Sigma^{-1} W_t \left\{A_t - W_t \Sigma^{-1} W_t'\right\}^{-1} W_t' \Sigma^{-1}\right] W_t A_t^{-1} \\
&= A_t^{-1} + A_t^{-1} W_t' \Sigma^{-1} [\Sigma + W_t A_{t-1}^{-1} W_t'] \Sigma^{-1} W_t A_t^{-1} \\
&= A_t^{-1} + A_t^{-1} W_t' \Sigma^{-1} F_t \Sigma^{-1} W_t A_t^{-1} \ ,
\end{aligned}$$

we obtain

$$\tilde{\alpha}_t|_{\mathcal{F}_{t-1}} \equiv N(\tilde{\alpha}_{t-1}, A_{t-1}^{-1} - A_t^{-1}) \ ,$$

as required for (b). Part (c) follows as in Theorem 2.5(c) of Phillips-Ploberger (1992). □

7.2. PROOF OF THEOREM 4.1

Since $SS_k - SS_K = \tilde{\Pi}_n(K) A_n(K) \tilde{\Pi}_n(K)' - \tilde{\Pi}_n(k) A_n(k) \tilde{\Pi}_n(k)'$ we have

$$\text{PIC}_k = \text{etr}[(1/2)\tilde{\Sigma}_K^{-1}(SS_k - SS_K)] |\tilde{\Sigma}_K^{-1} \otimes A_n(k)|^{1/2} / |\tilde{\Sigma}_K^{-1} \otimes A_n(K)|^{1/2} \ .$$

Minimizing PIC_k is therefore equivalent to minimizing

$$\mathrm{tr}(\bar{\Sigma}_K^{-1} SS_k) + \ln|\bar{\Sigma}_K^{-1} \otimes A_n(k)|$$

giving (24). To a first order approximation we have

$$\begin{aligned}\ln|\hat{\Sigma}_k| &\sim \ln|\hat{\Sigma}_K| + \mathrm{tr}|\hat{\Sigma}_K^{-1}(\hat{\Sigma}_k - \hat{\Sigma}_K)] \\ &= \ln|\hat{\Sigma}_K| + (1/n)\mathrm{tr}[\bar{\Sigma}_K^{-1}(SS_k - SS_K)] \ .\end{aligned}$$

Hence, minimizing (24) with respect to k is asymptotically equivalent to minimizing

$$\ln|\hat{\Sigma}_k| + (1/n)\ln|\bar{\Sigma}_K^{-1} \otimes A_n(k)|$$

giving (25). When the system is a stationary VAR, $(1/n)A_n(k) = (1/n)X_n(k)'X_n(k) = O_p(1)$ so that

$$\ln|\bar{\Sigma}_K^{-1} \otimes A_n(k)| = \ln(n^{mk}|\bar{\Sigma}_K^{-1} \otimes n^{-1}A_n(k)|) = mk\ \ln(n) + O_p(1)$$

leading to (26) upon the replacement $k = mp$ where p = number of lags in the VAR. □

7.3. PROOF OF LEMMA 4.2

We start by writing the model (1) in the form

$$y_t = (I_m \otimes x_t')\mathrm{vec}(\Pi) + \varepsilon_t$$

and transform the regressors with the orthogonal matrix $C = [S,\ S_\perp]$. Note that if S is not orthonormal we can, in order to achieve this, replace it with the matrix $S(S'S)^{-1/2}$, without loss of generality as far as the restrictions (2) are concerned. If we simultaneously transform the model by premultiplication with $\Sigma^{-1/2}$ and subsequently stack observations we have

$$\begin{aligned}y &= ZCC'\mathrm{vec}(\Pi) + \eta\ , \quad \eta \equiv N(0,\ I_{nm}) \\ &= Z_1S'\mathrm{vec}(\Pi) + Z_2S_\perp'\mathrm{vec}(\Pi) + \eta\end{aligned}$$

where $Z' = [\Sigma^{-1/2} \otimes x_1,\ \ldots,\ \Sigma^{-1/2} \otimes x_n]$, $Z_1 = ZS$ and $Z_2 = ZS_\perp$. Let ss_H and ss be the residual sum of squares from estimating this model subject to the restrictions $H_0 : S_\perp'\ \mathrm{vec}(\Pi) = s_\perp$ and without these restrictions, respectively. Then using (28) and (16) we find

$$dQ_n^\pi/dQ_n^\alpha(s) = |A_n(\Sigma)|^{1/2}|\Sigma^{-1} \otimes X_n'X_n|^{-1/2}\exp\{(1/2)(ss_H - ss)\}\ .$$

But

$$ss_H - ss = (S_\perp' \operatorname{vec}(\tilde{\Pi}_n) - s_\perp)'(Z_2'\bar{P}_1 Z_2)(S_\perp' \operatorname{vec}(\tilde{\Pi}_n) - s_\perp)$$

where $\bar{P}_1 = I_n - Z_1(Z_1'Z_1)^{-1}Z_1$, and

$$\begin{aligned}(Z_2'\bar{P}_1 Z_2)^{-1} &= [(C'Z'ZC)^{-1}]_{22} = [C'(Z'Z)^{-1}C]_{22} \\ &= S_\perp'(Z'Z)^{-1}S_\perp = S_\perp'\{\Sigma \otimes (X_n'X_n)^{-1}\}S_\perp .\end{aligned}$$

Hence

$$ss_H - ss = (S_\perp' \operatorname{vec}(\tilde{\Pi}_n) - s_\perp)'\left[S_\perp'\{\Sigma \otimes (X_n'X_n)^{-1}\}S_\perp\right]^{-1}(S_\perp' \operatorname{vec}(\tilde{\Pi}_n) - s_\perp) .$$

Next $|\Sigma^{-1} \otimes X_n'X_n| = |Z'Z|$ and $|A_n(\Sigma)| = |Z_1'Z_1|$, so that

$$|A_n(\Sigma)|^{1/2}|\Sigma^{-1} \otimes X_n'X_n|^{-1/2} = (|Z'Z|/|Z_1 Z_1|)^{-1/2} = |Z_2'\bar{P}_1 Z_2|^{-1/2} = |S_\perp'\{\Sigma \otimes (X_n'X_n)^{-1}\}S_\perp|^{-1/2} .$$

It follows that

$$\begin{aligned}dQ_n^\pi/dQ_n^\alpha(\Sigma) &= |S_\perp'\{\Sigma \otimes (X_n'X_n)^{-1}\}S_\perp|^{1/2} \\ &\quad \cdot \exp\left\{(1/2)(S_\perp' \operatorname{vec}(\tilde{\Pi}_n) - s_\perp)'\left[S_\perp'\{\Sigma \otimes (X_n'X_n)^{-1}\}S_\perp\right]^{-1}(S_\perp' \operatorname{vec}(\tilde{\Pi}_n) - s_\perp)\right\}\end{aligned}$$

and (29) follows directly.

REFERENCES

Ahn, S. K. and G. C. Reinsel (1988). "Nested reduced rank autoregressive models for multiple time series," *Journal of the American Statistical Association*, 83, 849-856.

Ahn, S. K. and G. C. Reinsel (1990). "Estimation for partially nonstationary multivariate autoregressive models," *Journal of the American Statistical Association*, 85, 813-823.

Berger, J. and R. Yang (1992). "Noninformative priors and Bayesian testing for the AR(1) model," Purdue University, mimeographed.

Boudjellaba, H., J-M. Dufour and R. Roy (1992a). "Testing causality between two vectors in multivariate ARMA models," *Journal of the American Statistical Association* (forthcoming).

Boudjellaba, H., J-M. Dufour and R. Roy (1992b). "Simplified conditions for non-causality between vectors in multivariate ARMA models," University of Montreal, mimeographed.

Breslow, N. (1990). "Biostatics and Bayes" (with discussion) *Statistical Science*, 5, 269-283.

Chen, H. F. and L. Guo (1991). *Identification and Stochastic Adaptive Control*. Boston: Birkhauser.

Doan, T. A. (1990). *RATS User's Manual: Version 3.10*. Minneapolis, MN: VAR Econometrics.

Doan, T., R.B. Litterman and C. Sims (1984). "Forecasting and conditional projections using realistic prior distributions," *Econometrics Reviews*, 3, 1-100.

Fair, R. C. (1976). *A Model of Macroeconomic Activity, Vol. 2, The Empirical Model*. Cambridge: Ballinger.

Fair, R. C. (1986). "Evaluating the predictive accuracy of models," in: Z. Griliches and M. D. Intriligator (eds.), *Handbook of Econometrics*, Vol. 3. Amsterdam: North-Holland.

Fair, R. C. and R. J. Shiller (1990). "Comparing information in forecasts from econometric models," *American Economic Review*, 80, 375-389.

Feige, E. L. and D. K. Pearce (1979). "The causal relationship between money and income: Some caveats for time series analysis," *Review of Economics and Statistics*, 61, 521-533.

Garcia-Ferrer, A., R. A. Highfield, F. Palm and A. Zellner (1987). "Macroeconomic forecasting using pooled international data," *Journal of Business and Economic Statistics*, 5, 53-67.

Hannan, E. J. and M. Deistler (1988). *The Statistical Theory of Linear Systems*. New York: John Wiley & Sons.

Hannan, E. J. and L. Kavalieris (1984). "Multivariate linear time series models," *Advances in Applied Probability* 16, 492-561.

Hannan, E. J. and J. Rissanen (1982). "Recursive estimation of ARMA order," *Biometrika*, 69, 273-280 [Corrigenda, *Biometrika*, 1983, 70].

Harrison, P. J. and C. F. Stevens (1976). "Bayesian forecasting (with dicussion)," *Journal of the Royal Statistical Society*, Series B, 38, 205-247.

Hsiao, C. (1979). "Autoregressive modeling of Canadian money and income data," *Journal of the American Statistical Association*, 74, 553-560.

Jeffreys, H. (1961). *Theory of Probability*, 3rd Edition. London: Oxford University Press.

Johansen, S. (1988). "Statistical analysis of cointegration vectors," *Journal of Economic Dynamics and Control*, 12, 231-254.

Johansen, S. (1991). "Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models," *Econometrica*, 59, 1551-1580.

Lindley, D. V. (1990). "The present position in Bayesian statistics," *Statistical Science*, 5, 44-89.

Litterman, R. B. (1986). "Forecasting with Bayesian vector autoregressions: Five years of experience," *Journal of Business and Economic Statistics*, 4, 25-38.

Nelson, C. R. and C. Plosser (1982). "Trends and random walks in macroeconomic time series: Some evidence and implications," *Journal of Monetary Economics*, 10, 139-162.

Osborn, D. R. (1984). "Causality testing and its implications for dynamic econometric models," *Economic Journal*, 94, 82-95.

Phillips, P. C. B. (1991a). "To criticize the critics: An objective Bayesian analysis of stochastic trends," *Journal of Applied Econometrics*, 6(4), 333-364.

Phillips, P. C. B. (1991b). "Bayesian routes and unit roots: de rebus prioribus semper est disputandum," *Journal of Applied Econometrics*, 6(4), 435-474.

Phillips, P. C. B. (1992a). "Bayesian model selection and prediction with empirical applications," Yale University, mimeographed, May 1992.

Phillips, P. C. B. (1992b). "Bayes models and forecasts of Australian macroeconomic time series," Yale University, mimeographed, June 1992.

Phillips, P. C. B. and W. Ploberger (1992). "Posterior odds testing for a unit root with data-based model selection," Cowles Foundation Discussion Paper No. 1017.

Poirier, D. J., ed. (1991a). "Bayesian empirical studies in economics and finance," *Journal of Econometrics*, 49, 1-304.

Poirier, D. J. (1991b). "A Bayesian view of nominal money and real output through a new classical macroeconomic window," *Journal of Business and Economic Statistics*, 9, 125-148.

Rissanen, J. (1978). "Modeling by shortest data description," *Automatica*, 14, 465-471.

Schwarz, G. (1978). "Estimating the dimension of a model," *Annals of Statistics*, 6, 461-464.

Sims, C. A. (1972). "Money, income and causality," *American Economic Review*, 62, 540-552.

Sims, C. A. (1980a). "Macroeconomics and reality," *Econometrica*, 48, 1-48.

Sims, C. A. (1980b). "Comparison of interwar and postwar business cycles: Monetarism reconsidered," *American Economic Review*. 70, 250-257.

Sims, C. A. (1988). "Bayesian skepticism on unit root econometrics," *Journal of Economic Dynamics and Control*, 12, 436-474.

Sims, C. A., J. H. Stock and M. W. Watson (1990). "Inference in linear time series models with some unit roots," *Econometrica*, 58, 113-144.

Sims, C. A. and H. Uhlig (1991). "Understanding unit rooters: A helicopter tour," *Econometrica*, 59, 1591-1600.

Toda, H. and P. C. B. Phillips (1991). "Vector autoregressions and causality," Cowles Foundation Discussion Paper No. 977.

Todd, R. M. (1984). "Improving economic forecasting with Bayesian vector autoregression," *Federal Reserve Bank of Minneapolis Quarterly Review*, 4, 18-29.

Todd, R. M. (1990). "Vector autoregression evidence on monetarism: Another look at the robustness debate," *Federal Research Bank of Minneapolis Quarterly Review*, 19-37.

Wei, C. Z. (1992). "On predictive least squares principles," *Annals of Statistics*, 20, 1-42.

West, M. and P. J. Harrison (1989). *Bayesian Forecasting and Dynamic Models*. New York: Springer-Verlag.

Zellner, A. (1985). "Bayesian econometrics," *Econometrica*, 53, 253-269.

Zellner, A. and C-K. Min (1992). "Bayesian analysis, model selection and prediction," University of Chicago, mimeographed.

Zellner, A. and C. Hong (1989). "Forecasting international growth rates using Bayesian shrinkage and other procedures," *Journal of Econometrics*, 40, 183-202.

Figure 0: US Quarterly Data 1959:1 – 1992:1

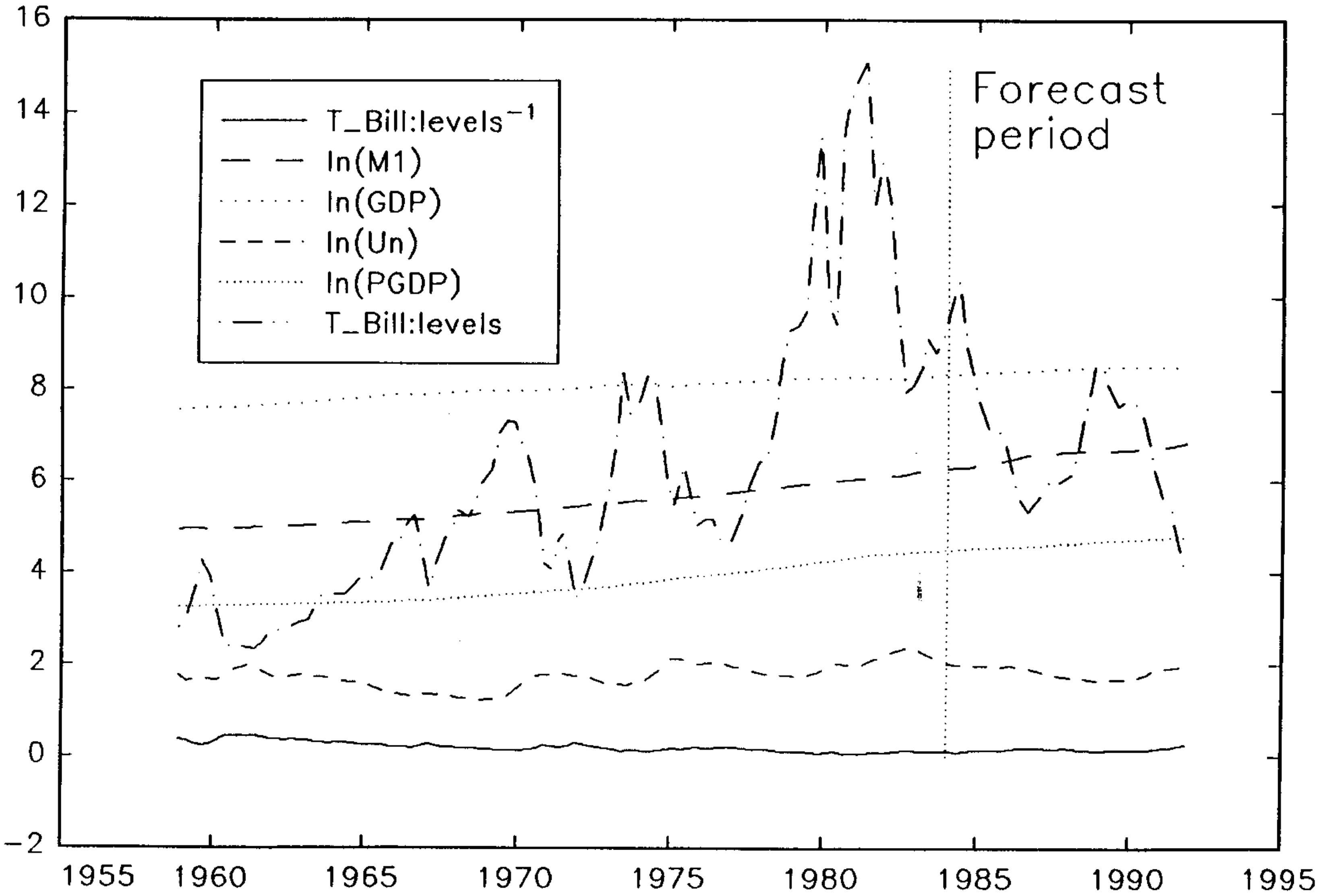


Figure 1(a): T_Bill: 1959:1–1992:1 (Levels)$^{-1}$

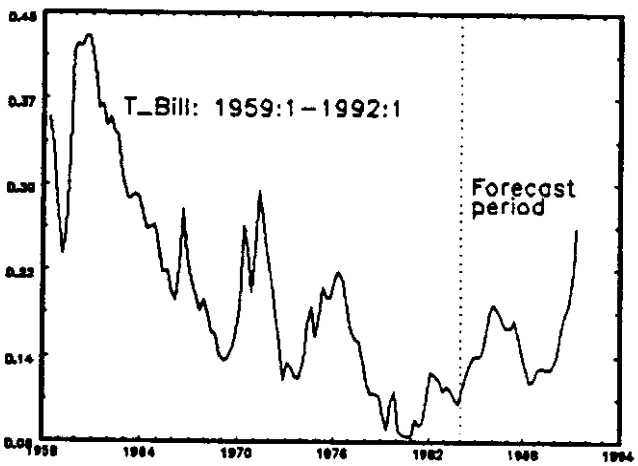


Figure 1(b): Prediction errors

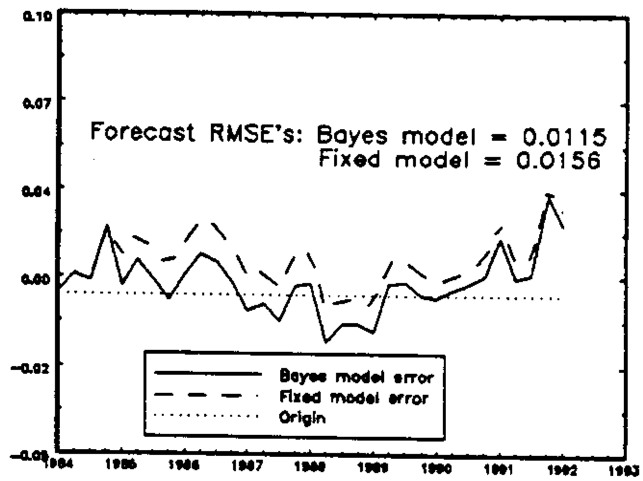


Figure 1(c): Evolving Bayes Model
(i) AR(p) + Trend(r) parameters
(ii) Unit Root present or not

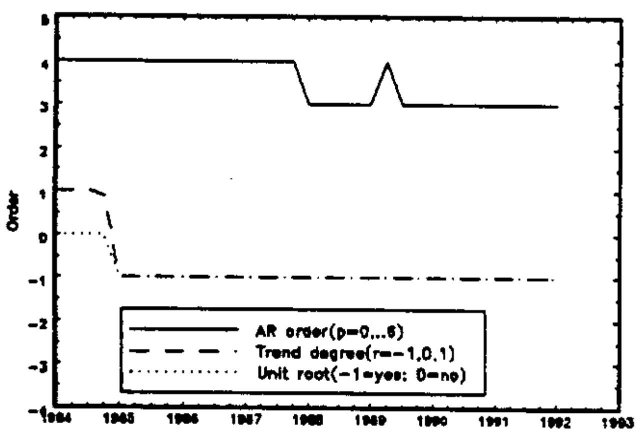


Figure 1(d): Recursive plot of Bayes Model Forecast Encompassing Test

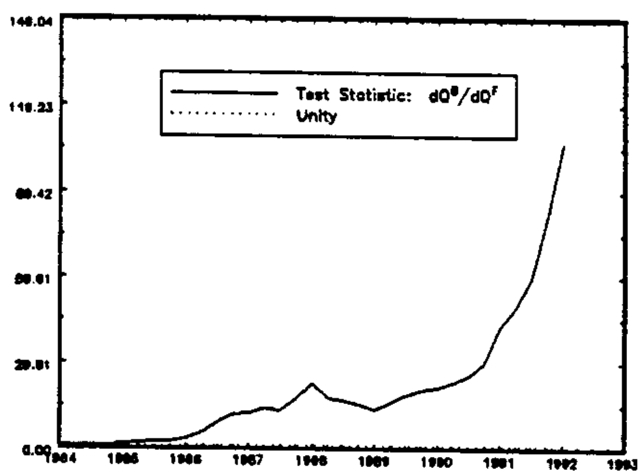


Figure 1(e): Long run AR coefficient

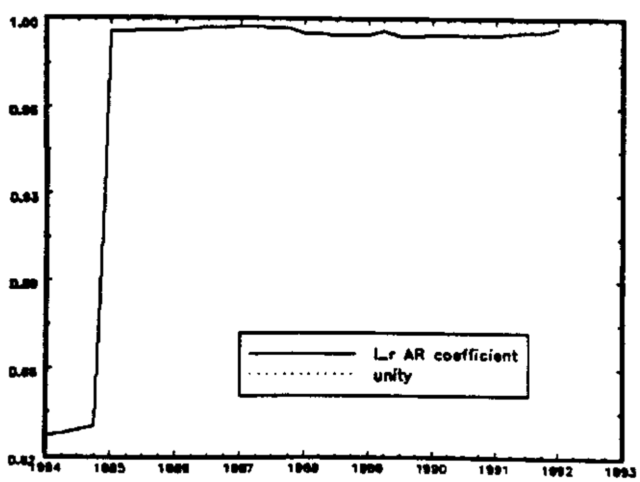


Figure 1(f): Bayes Model Forecasts and 95% confidence bands

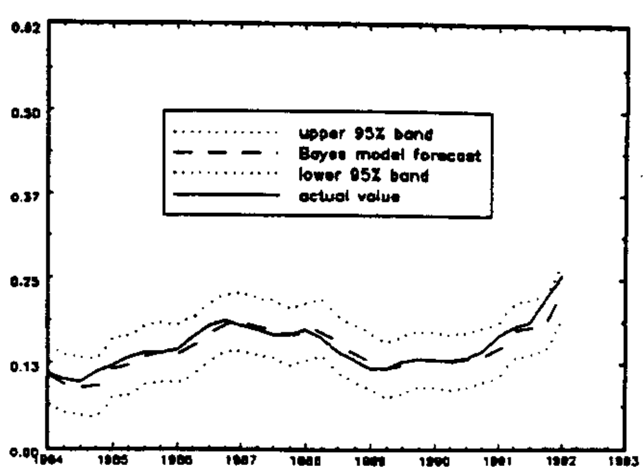


Figure 1'(a): T_Bill: 1959:1-1992:1 Levels

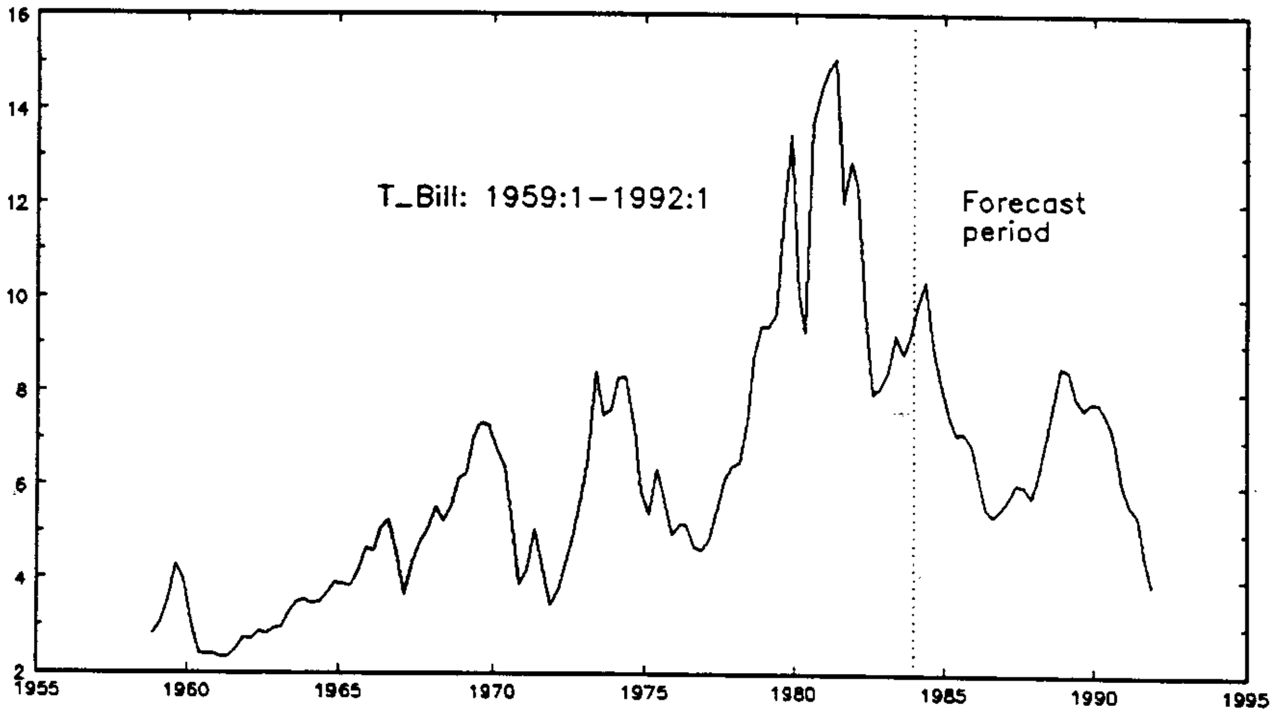


Figure 1'(b): Prediction errors

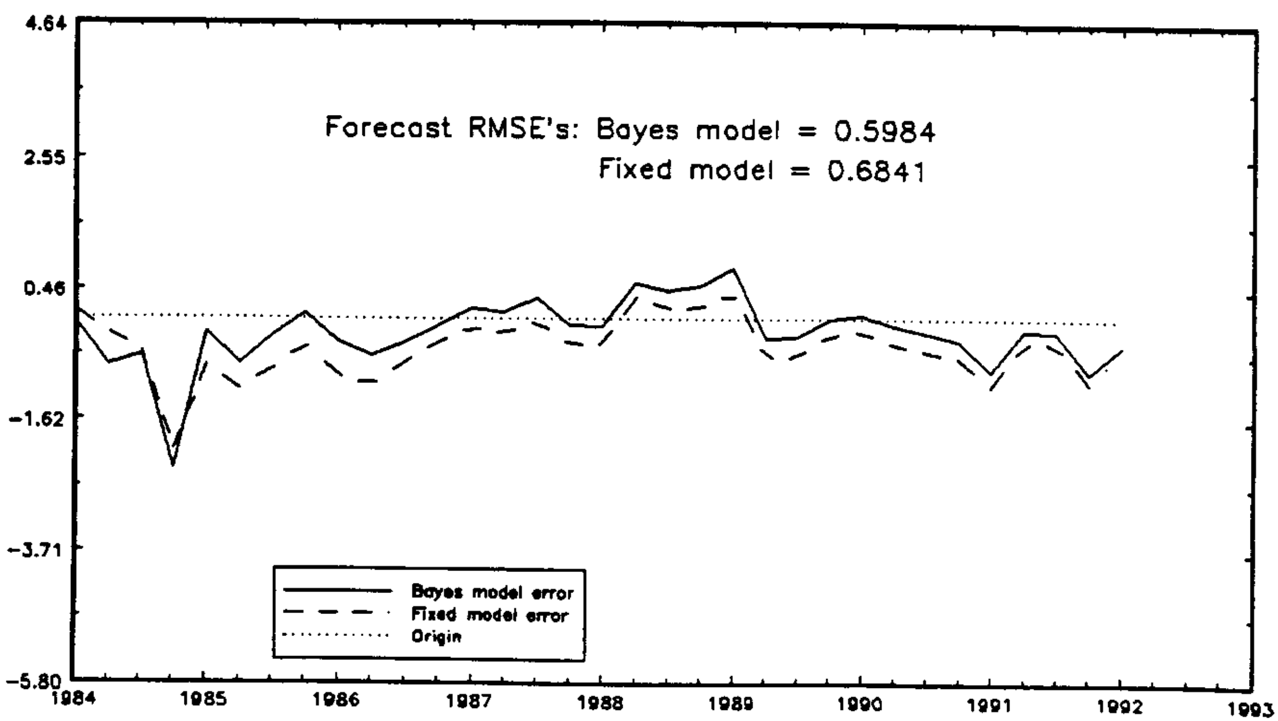


Figure 1''(a): T_Bill: 1959:1−1992:1 (Levels)

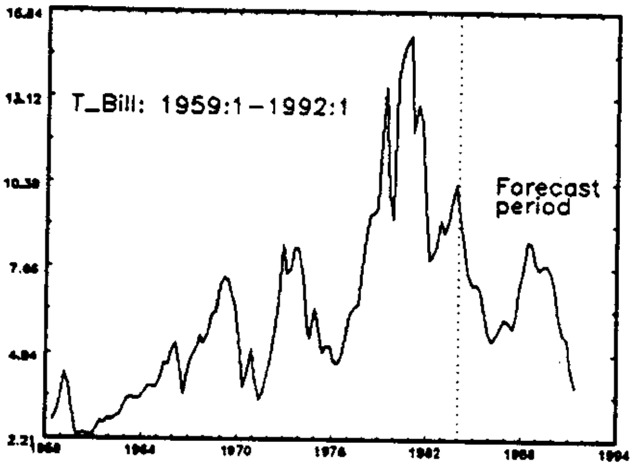


Figure 1''(b): Prediction errors

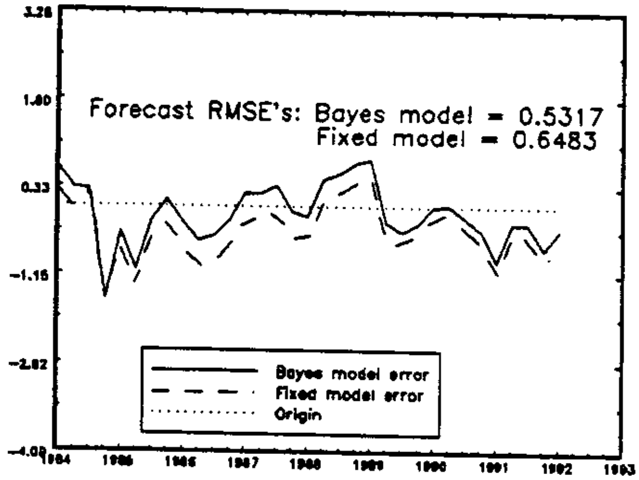


Figure 1''(c): Evolving Bayes Model
(i) AR(p) + Trend(r) parameters
(ii) Unit Root present or not

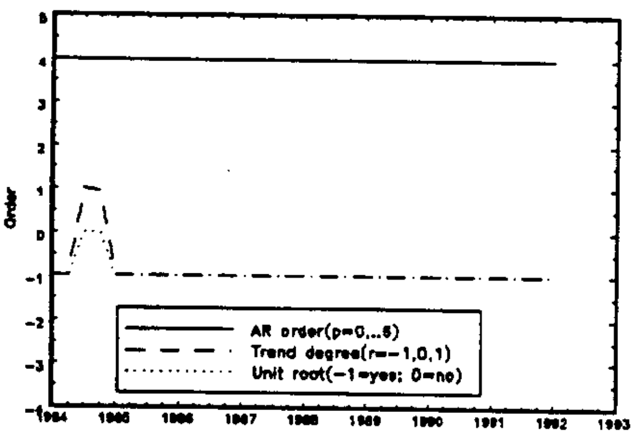


Figure 1''(d): Recursive plot of Bayes Model Forecast Encompassing Test

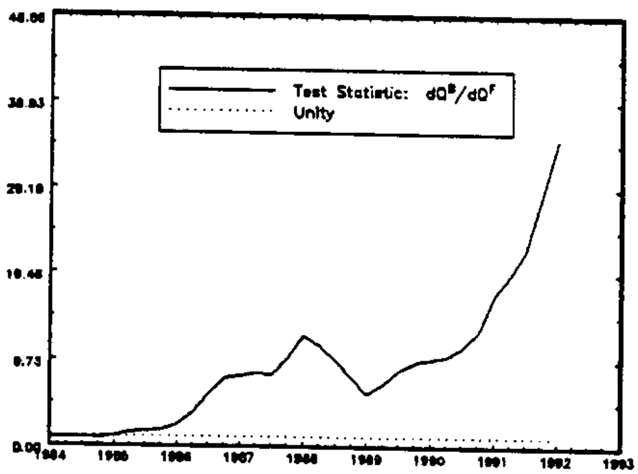


Figure 1''(e): Long run AR coefficient

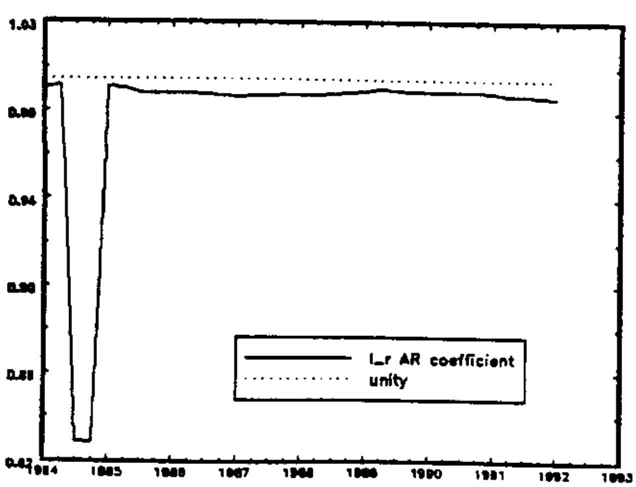


Figure 1''(f): Bayes Model Forecasts and 95% confidence bands

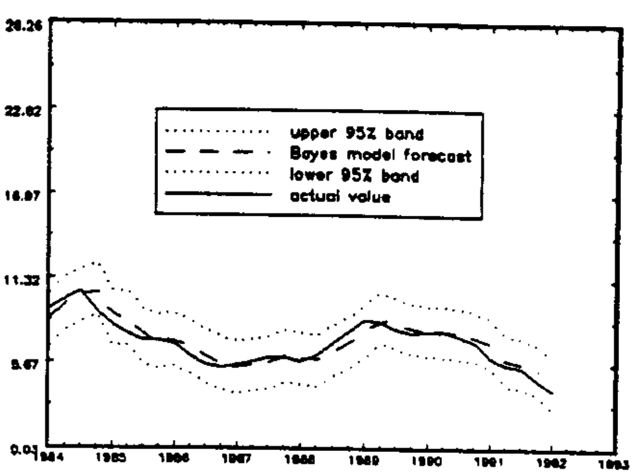


Figure 2(a): M1: 1959:1–1992:1 (Log–Levels)

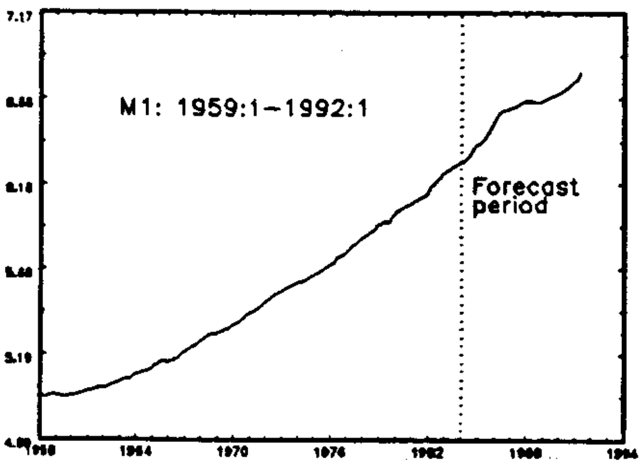


Figure 2(b): Prediction errors

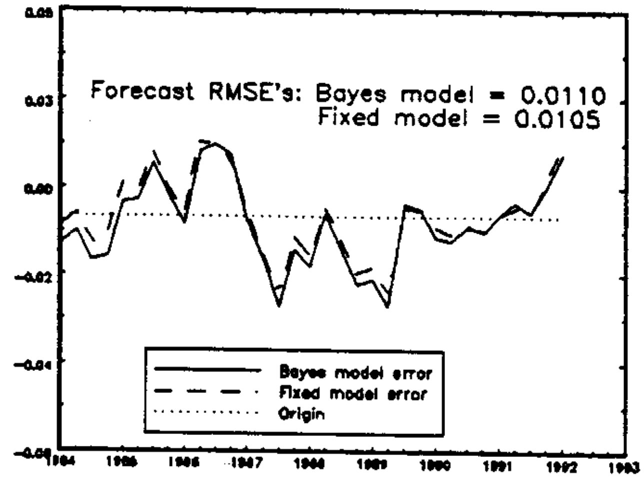


Figure 2(c): Evolving Bayes Model
(i) AR(p) + Trend(r) parameters
(ii) Unit Root present or not

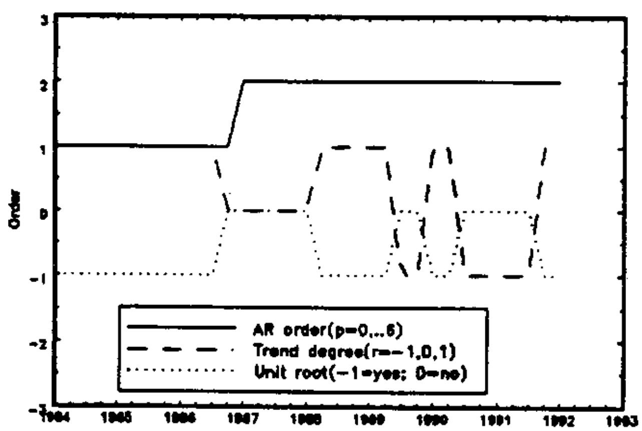


Figure 2(d): Recursive plot of Bayes Model Forecast Encompassing Test

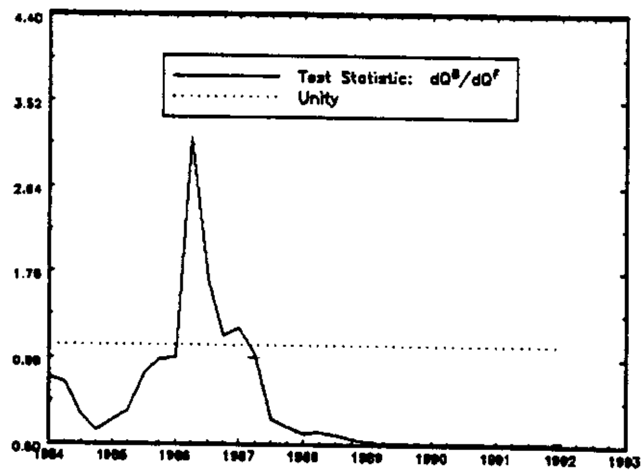


Figure 2(e): Long run AR coefficient

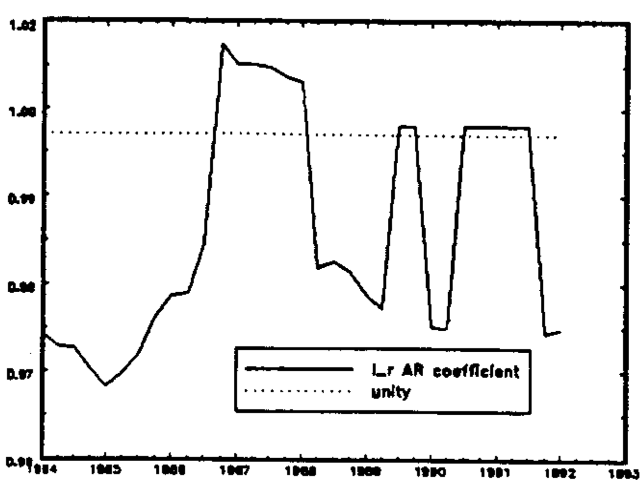


Figure 2(f): Bayes Model Forecasts and 95% confidence bands

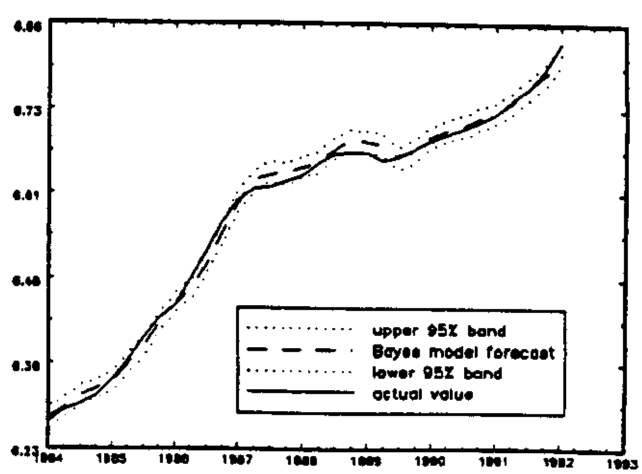


Figure 3(a): GDP: 1959:1–1992:1 (Log–Levels)

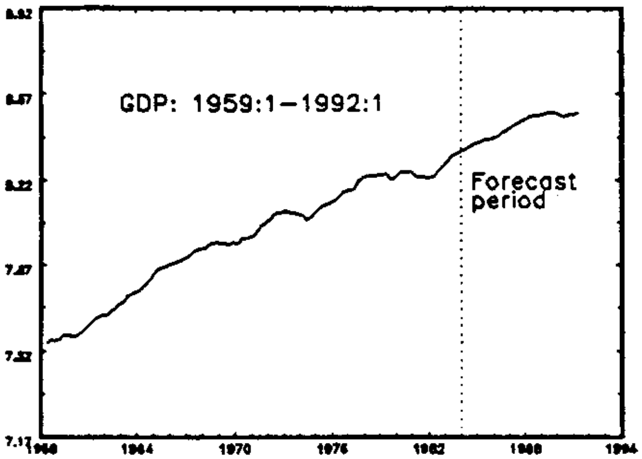


Figure 3(b): Prediction errors

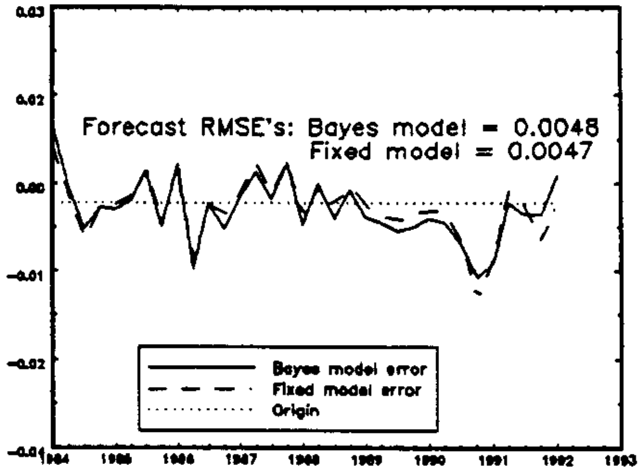


Figure 3(c): Evolving Bayes Model
(i) AR(p) + Trend(r) parameters
(ii) Unit Root present or not

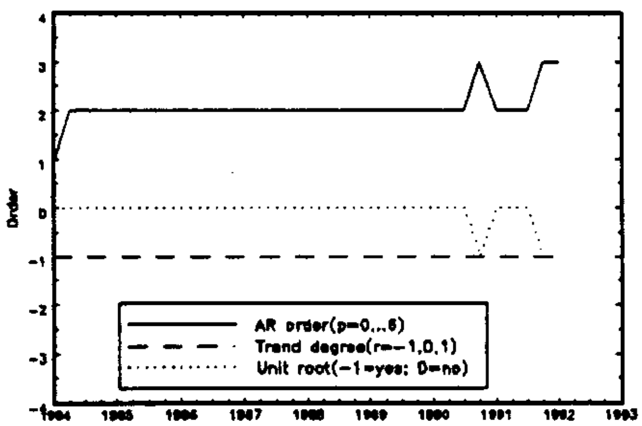


Figure 3(d): Recursive plot of Bayes Model Forecast Encompassing Test

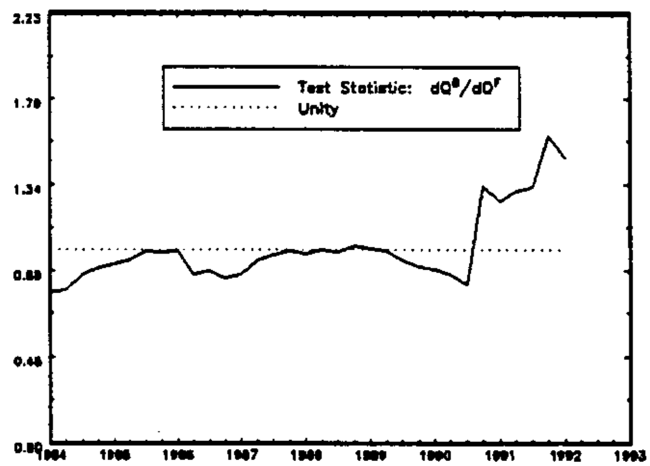


Figure 3(e): Long run AR coefficient

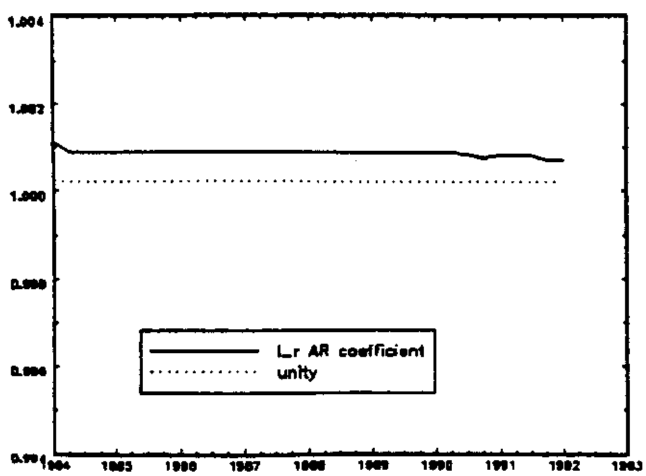


Figure 3(f): Bayes Model Forecasts and 95% confidence bands

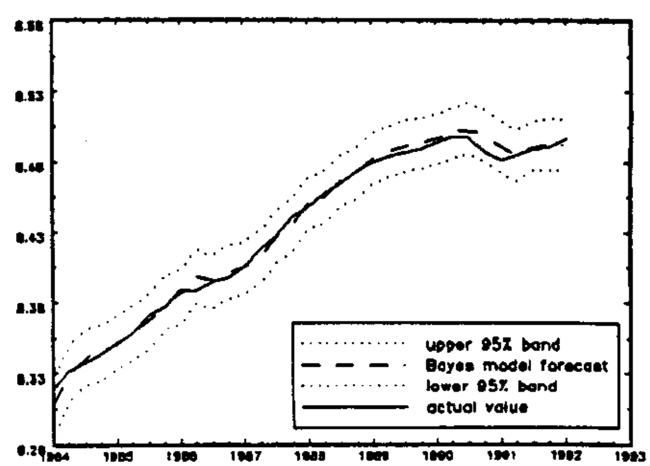


Figure 4(a): Un: 1959:1-1992:1
(Log-Levels)

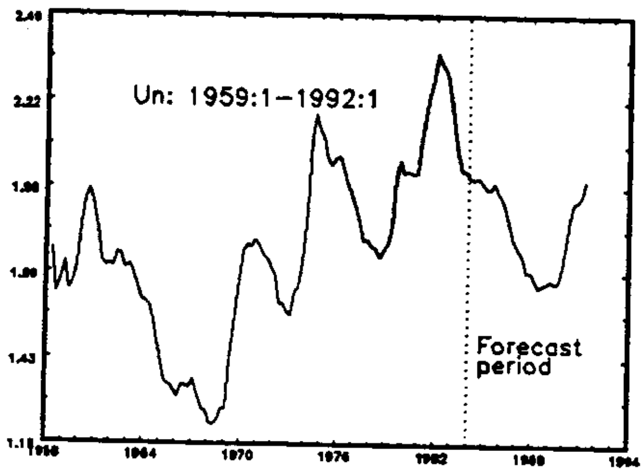


Figure 4(b): Prediction errors

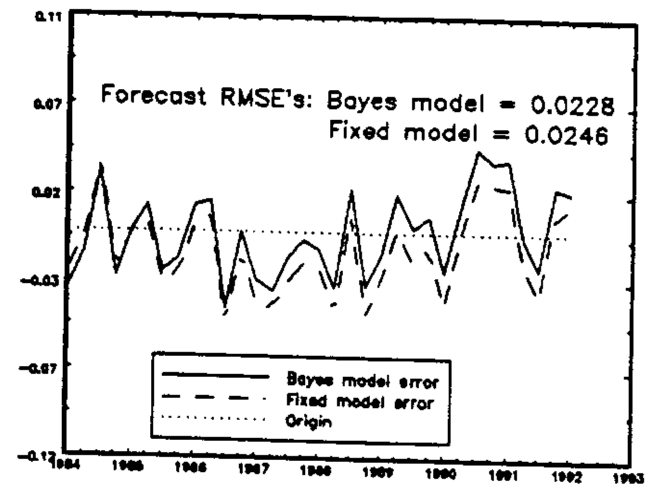


Figure 4(c): Evolving Bayes Model
(i) AR(p) + Trend(r) parameters
(ii) Unit Root present or not

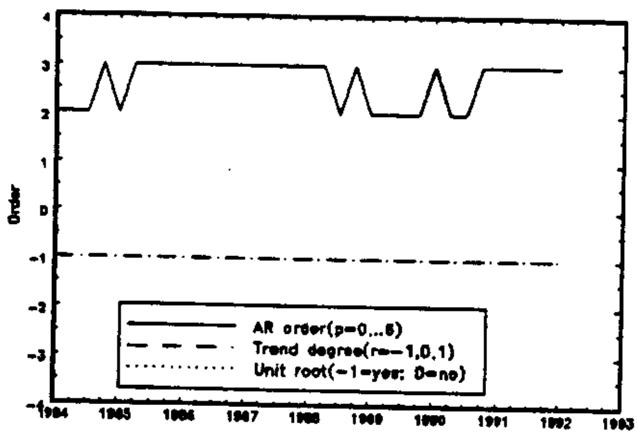


Figure 4(d): Recursive plot of Bayes
Model Forecast Encompassing Test

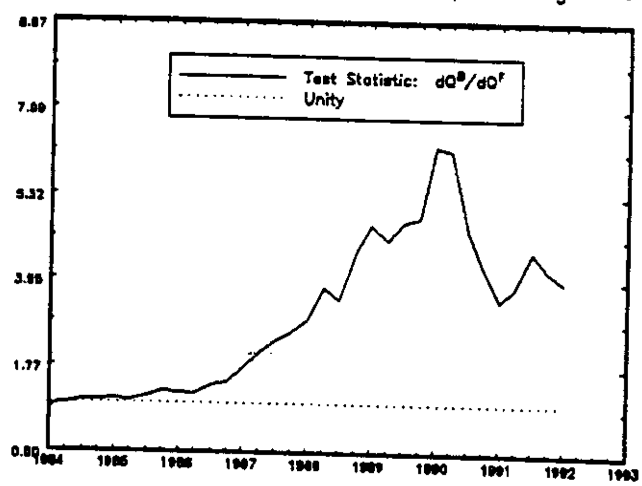


Figure 4(e): Long run AR coefficient

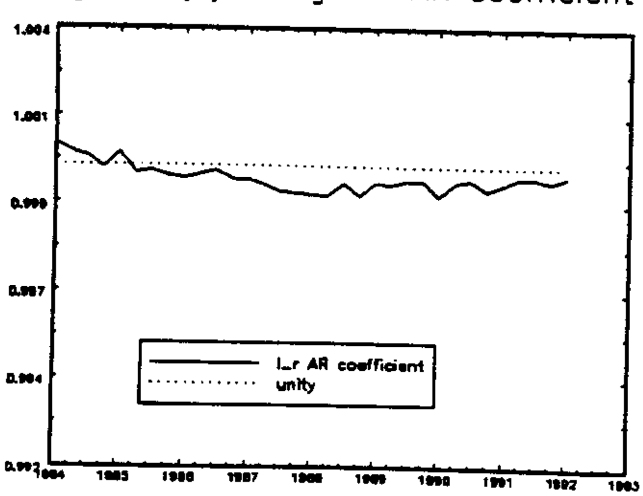


Figure 4(f): Bayes Model Forecasts
and 95% confidence bands

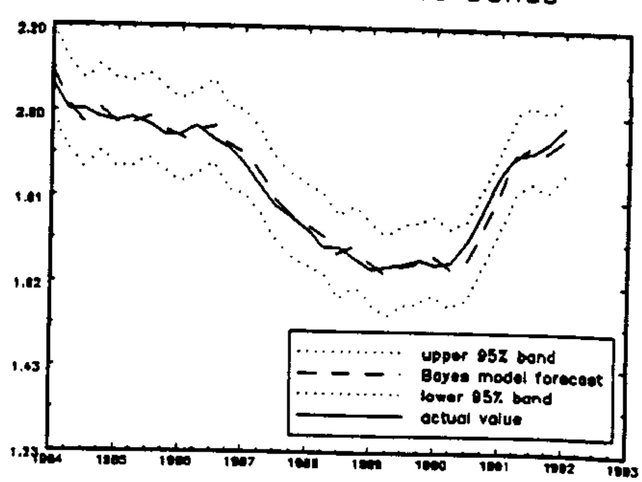


Figure 5(a): PGDP: 1959:1−1992:1 (Log−Levels)

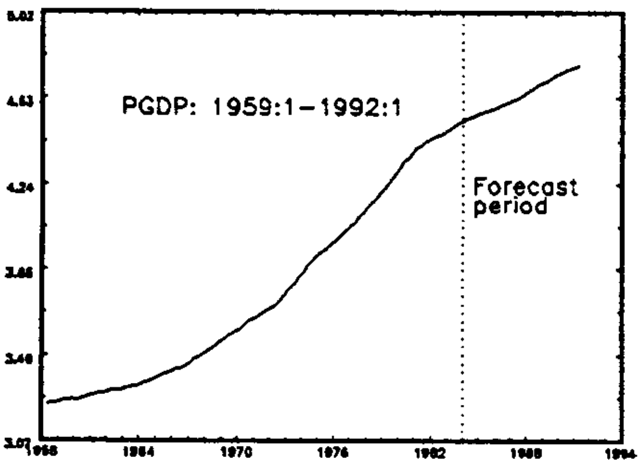


Figure 5(b): Prediction errors

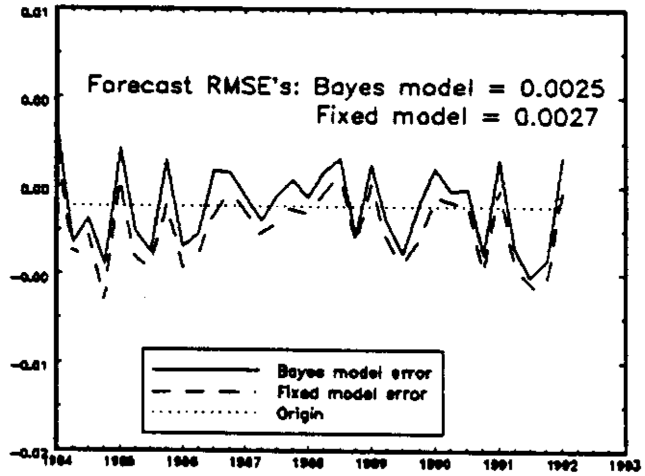


Figure 5(c): Evolving Bayes Model
(i) AR(p) + Trend(r) parameters
(ii) Unit Root present or not

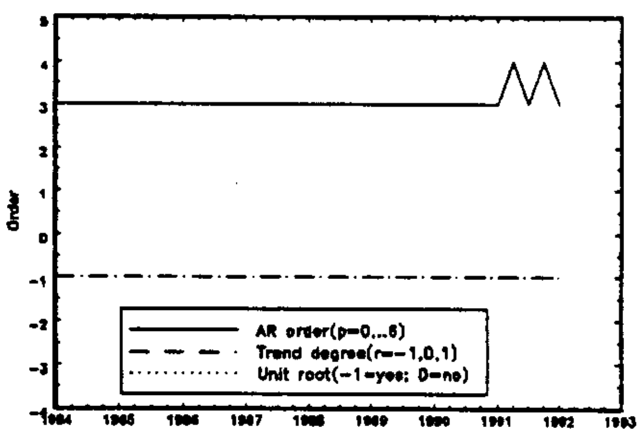


Figure 5(d): Recursive plot of Bayes Model Forecast Encompassing Test

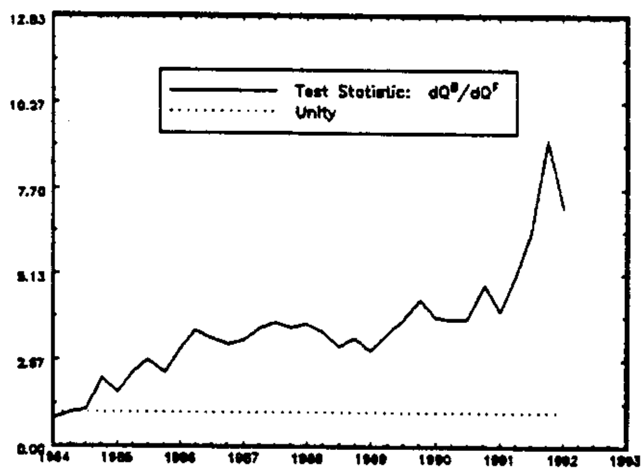


Figure 5(e): Long run AR coefficient

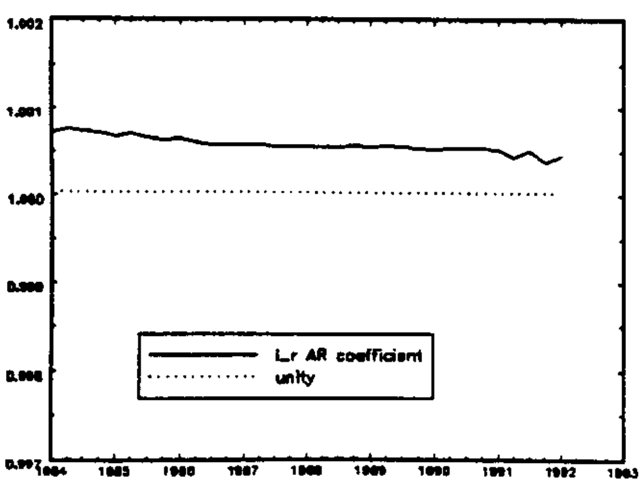


Figure 5(f): Bayes Model Forecasts and 95% confidence bands

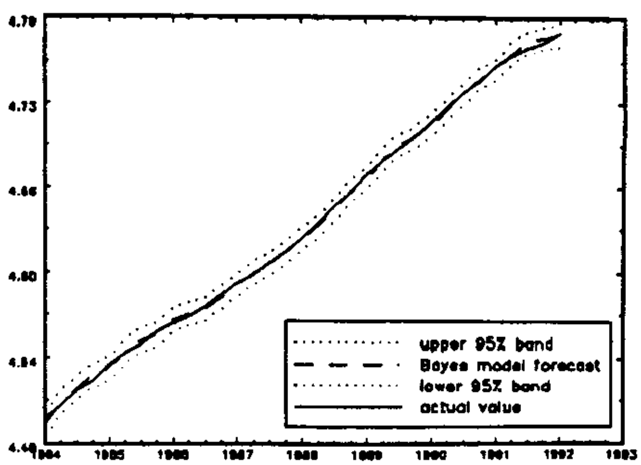


Figure 6: Forecasts from a 'Bayes Model' VAR for T_Bill^{-1}: 1959:1 – 1992:1

Figure 6(a): Prediction errors

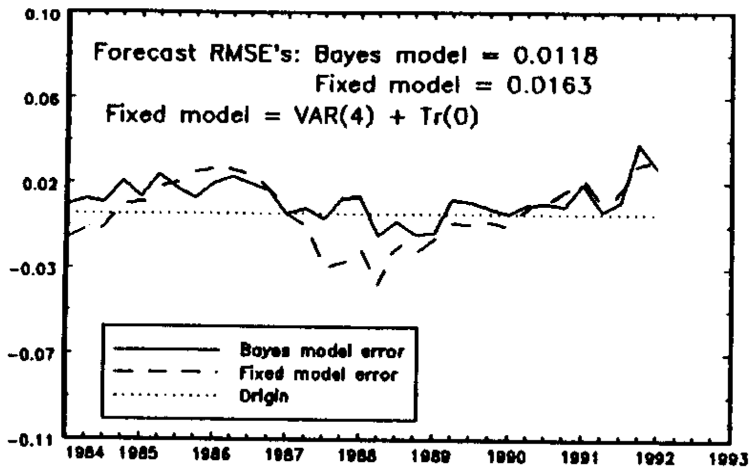


Figure 6(b): Bayes Model Forecasts and 95% confidence bands

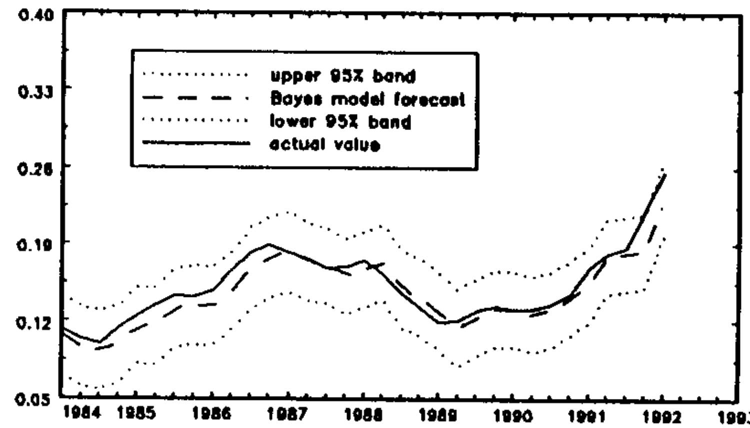


Figure 6(c): Recursive plot of Bayes Model Unit Root Test

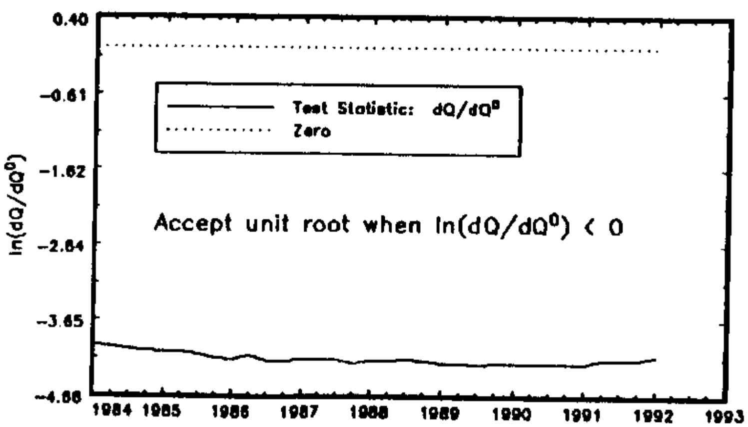


Figure 6(d): Recursive plot of Long run AR coefficient

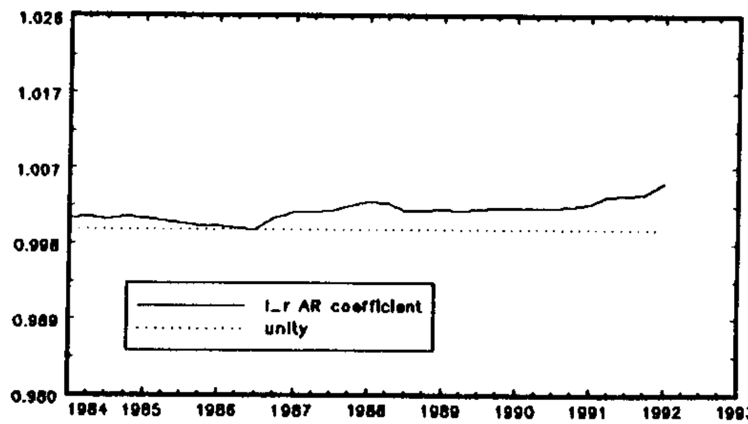


Figure 7: Forecasts from a 'Bayes Model' VAR for M1: 1959:1–1992:1

Figure 7(a): Prediction errors

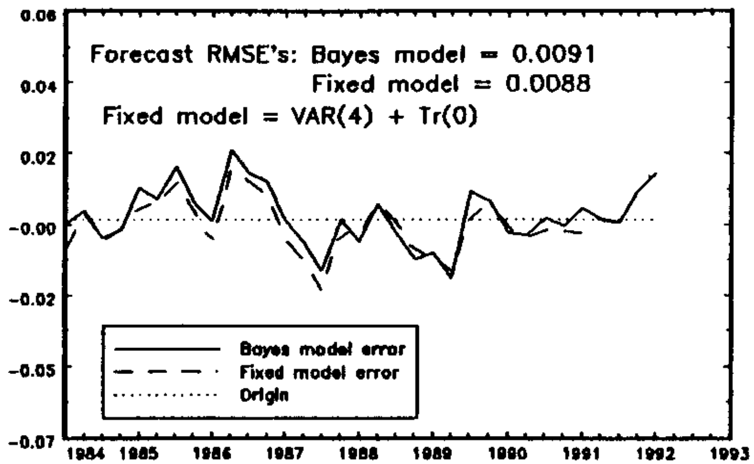


Figure 7(b): Bayes Model Forecasts and 95% confidence bands

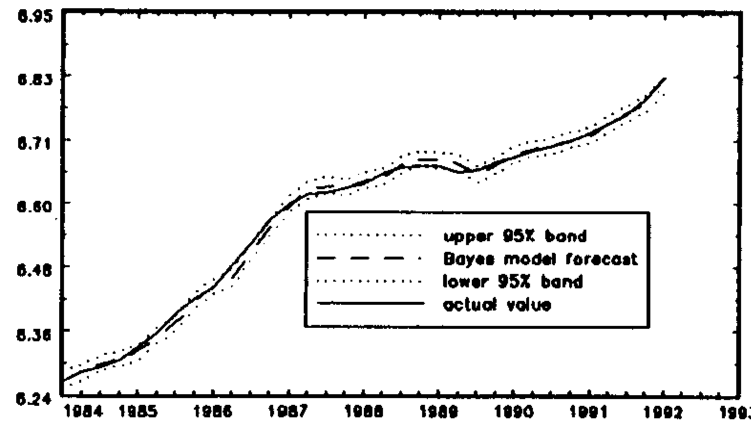


Figure 7(c): Recursive plot of Bayes Model Unit Root Test

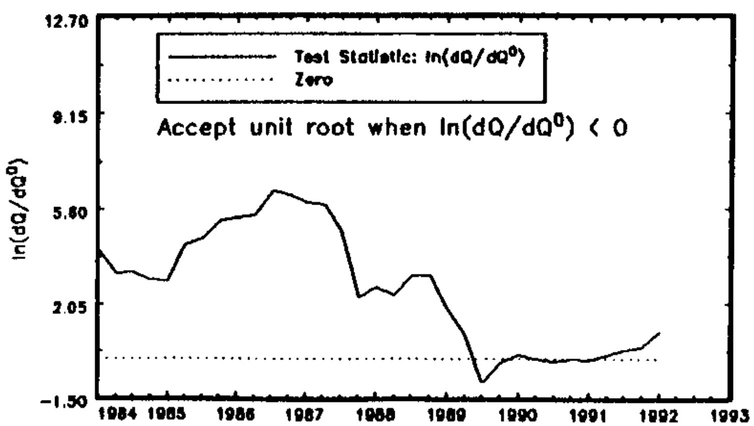


Figure 7(d): Recursive plot of Long run AR coefficient

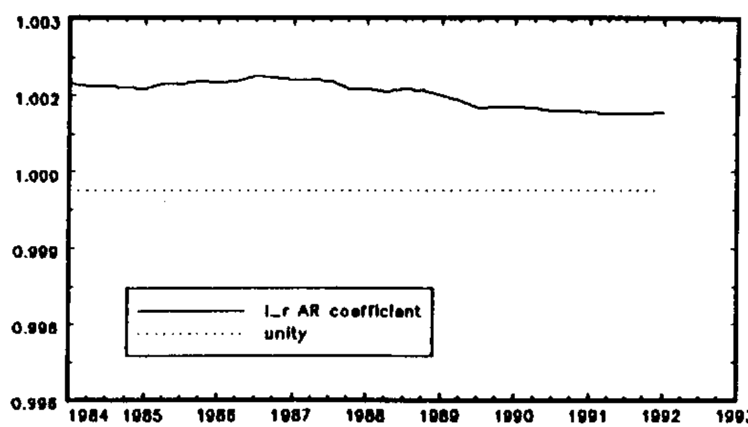


Figure 8: Forecasts from a 'Bayes Model' VAR for GDP: 1959:1–1992:1

Figure 8(a): Prediction errors

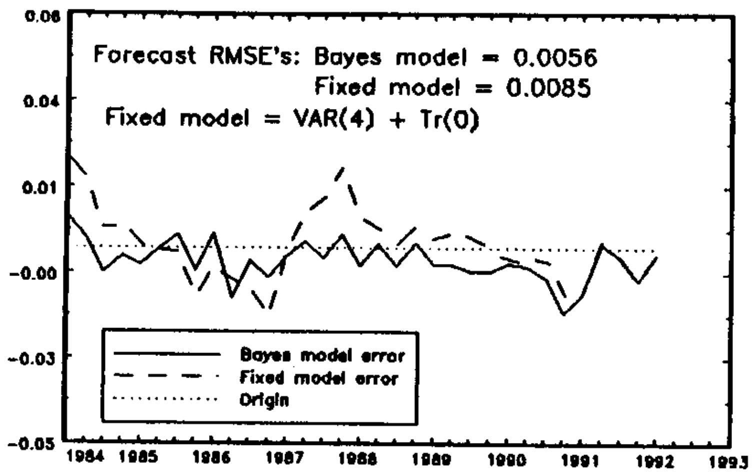


Figure 8(b): Bayes Model Forecasts and 95% confidence bands

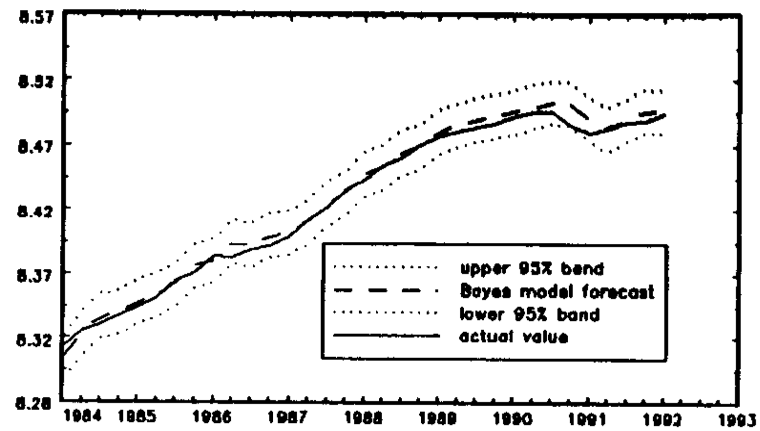


Figure 8(c): Recursive plot of Bayes Model Unit Root Test

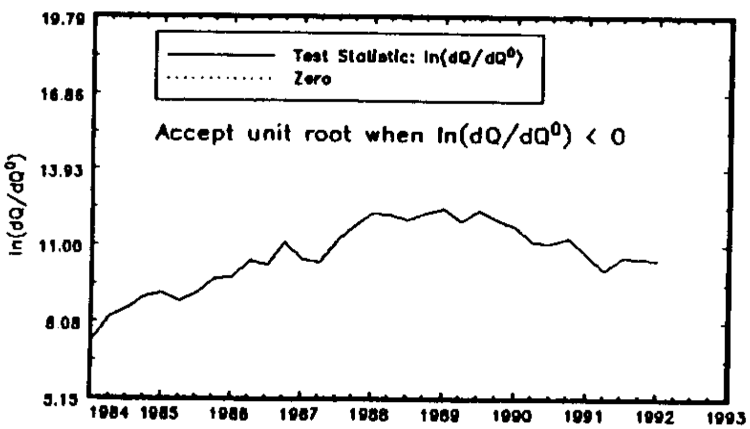


Figure 8(d): Recursive plot of Long run AR coefficient

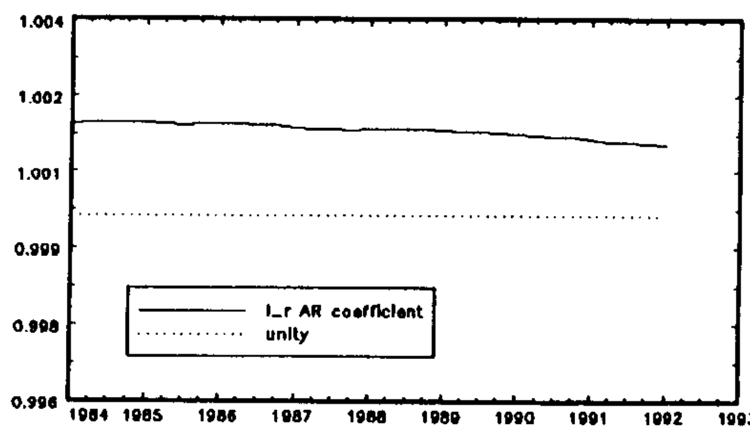


Figure 9: Forecasts from a 'Bayes Model' VAR for Un: 1959:1–1992:1

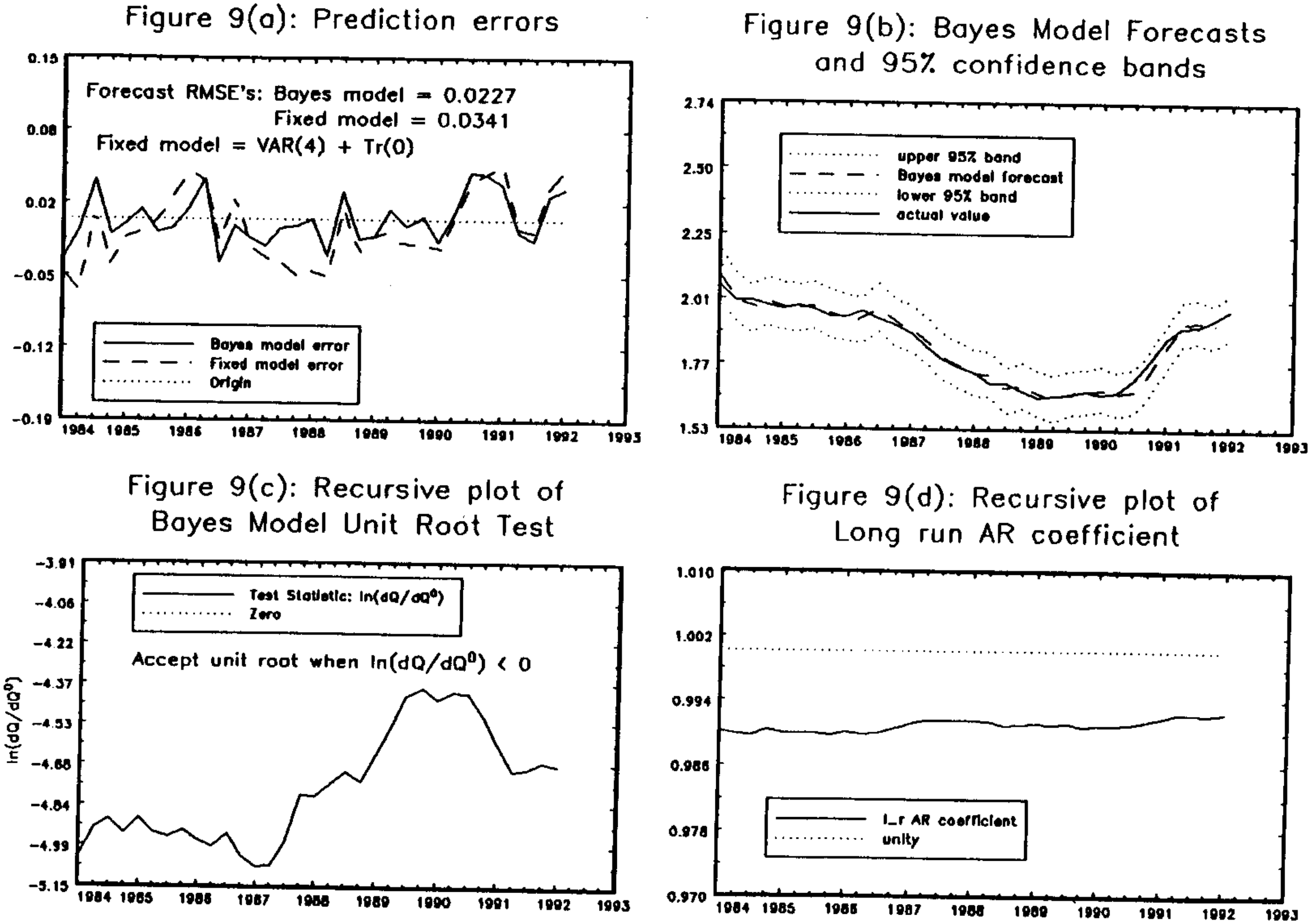


Figure 10: Forecasts from a 'Bayes Model' VAR for PGDP: 1959:1–1992:1

Figure 10(a): Prediction errors

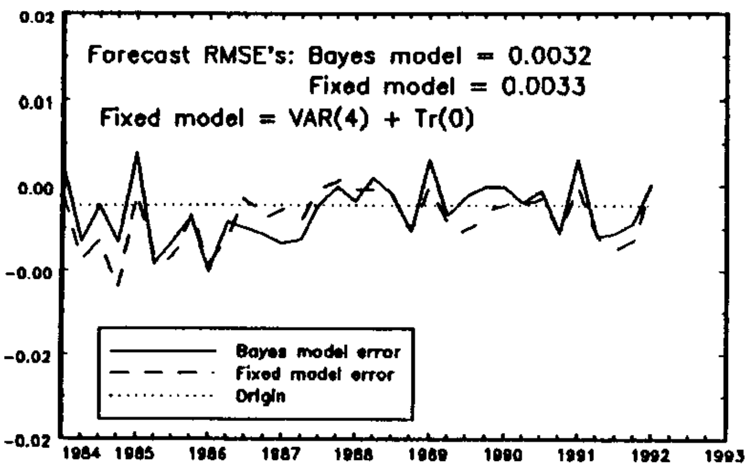


Figure 10(b): Bayes Model Forecasts and 95% confidence bands

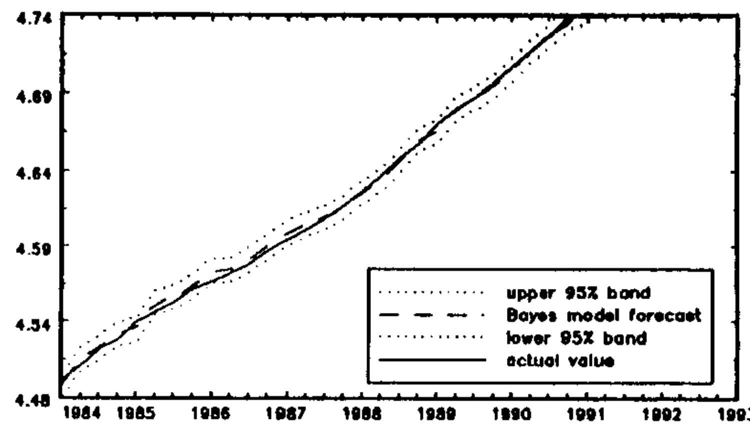


Figure 10(c): Recursive plot of Bayes Model Unit Root Test

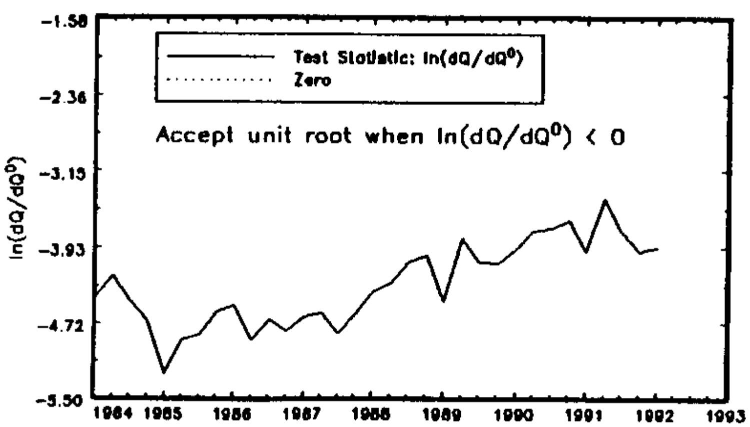


Figure 10(d): Recursive plot of Long run AR coefficient

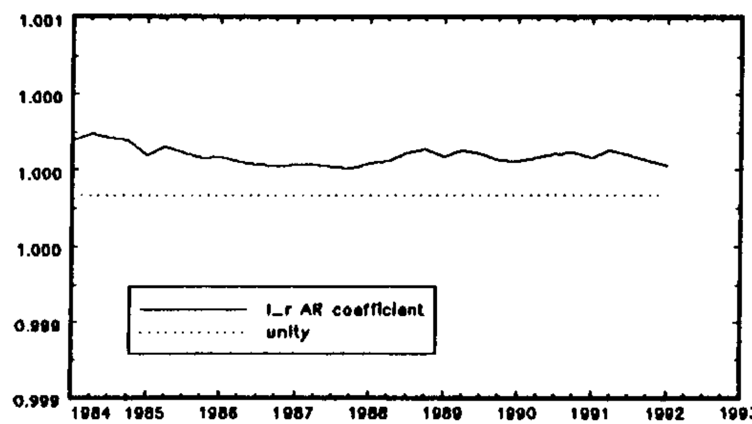


Figure 11: 'Bayes Model' VAR Tests of Causality and Cointegration

Figure 11(a): Recursive plot of Bayes Model Causality Test: Un on M1

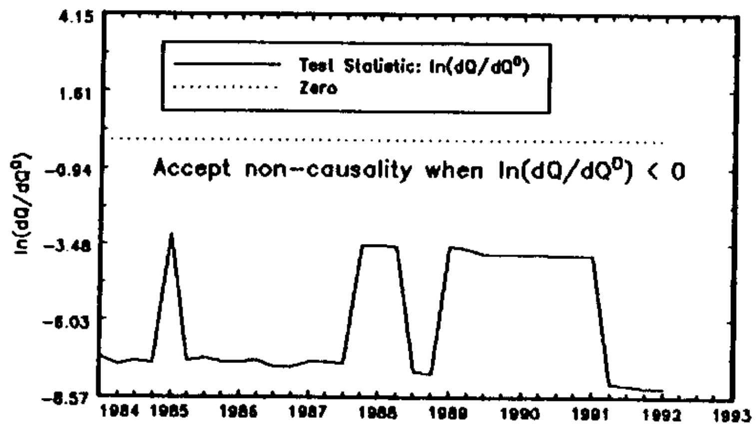


Figure 11(b): Recursive plot of Bayes Model Causality Test: M1 on GDP

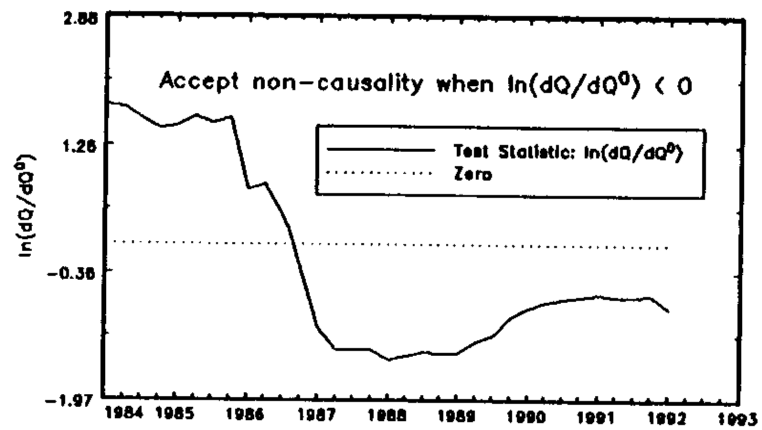


Figure 11(c): Recursive plot of Bayes Model Cointegration Test

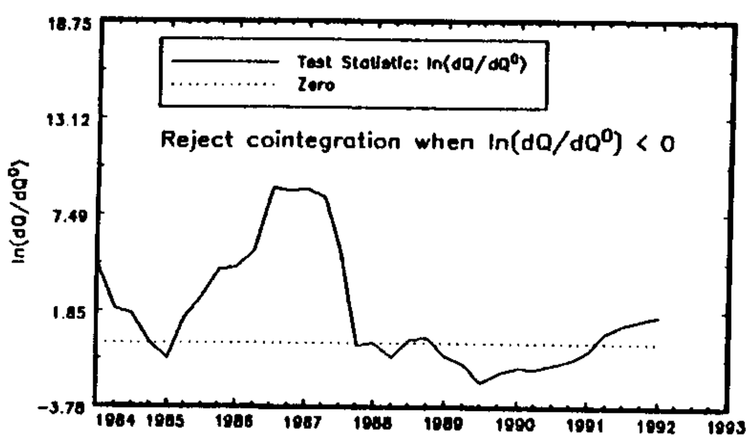


Figure 11(d):Recursive plot of cointegrating coefficients

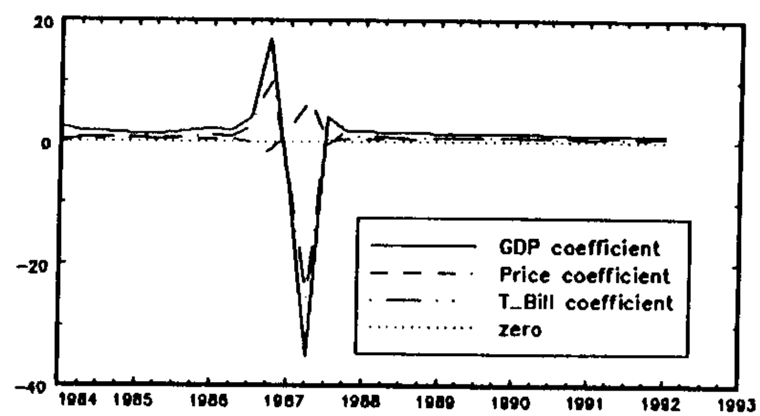


Figure 12: 'Bayes Model' VAR Tests of Forecast Encompassing

Figure 12(a): Recursive plot of Test: 'Bayes model' vs. VAR(4) + Tr(1)

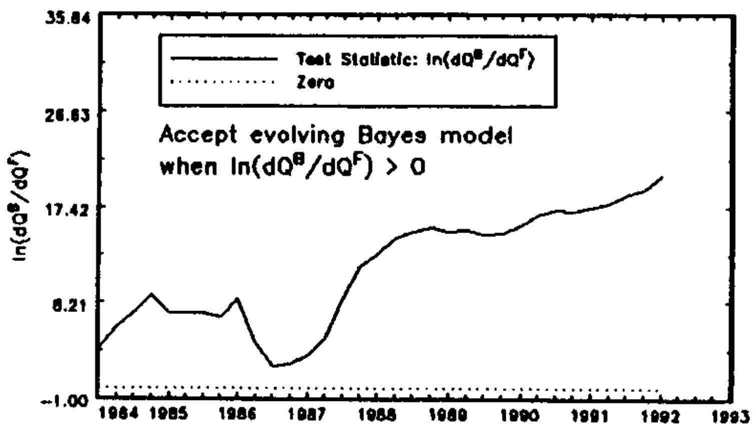


Figure 12(b): Recursive plot of Test: 'Bayes model' vs. VAR(2) + Tr(-1)

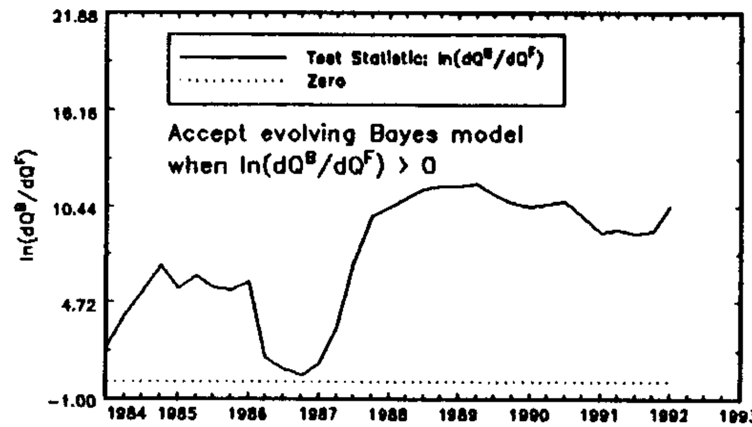


Figure 12(c): Recursive plot of Test: 'Bayes model^{nc} vs. 'Bayes model^{c}'

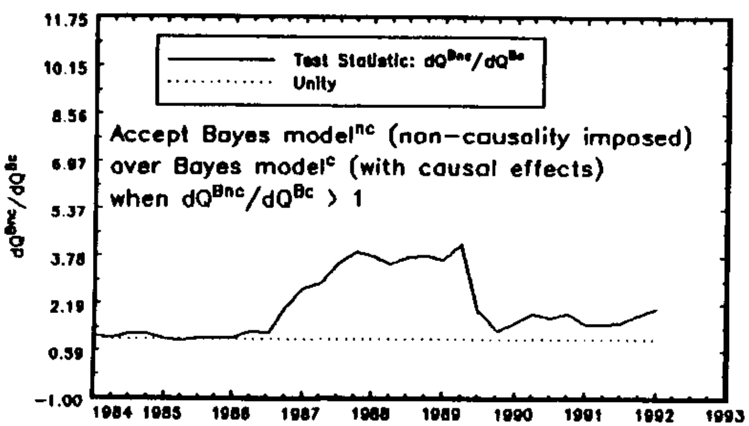
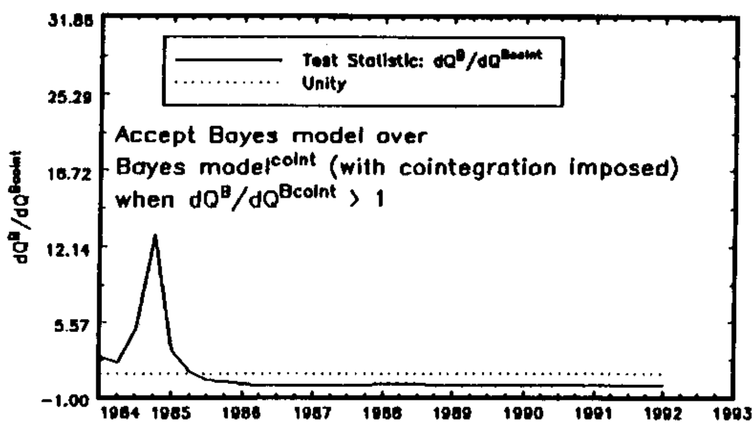


Figure 12(c): Recursive plot of Test: 'Bayes model vs. 'Bayes model^{coint}'



Figures 13–17: Prediction Errors for a Bayes Model^{coint} VAR

Figure 13:T_Bill^{-1}: 1959:1 – 1992:1

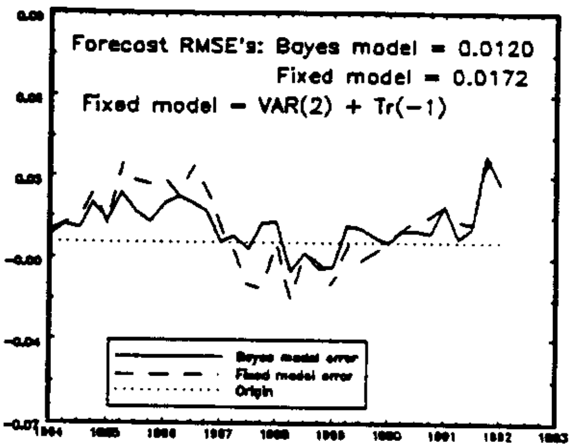


Figure 14: M1: 1959:1–1992:1

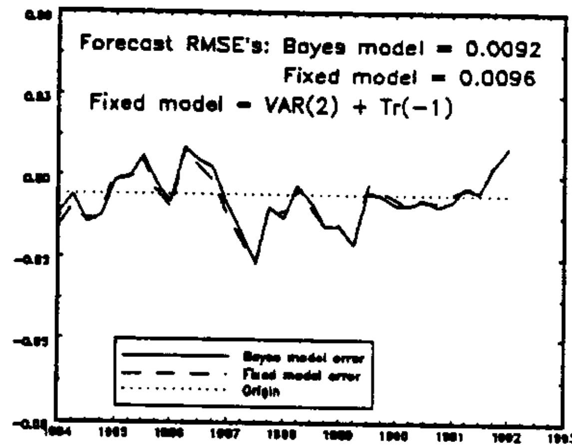


Figure 15: GDP: 1959:1–1992:1

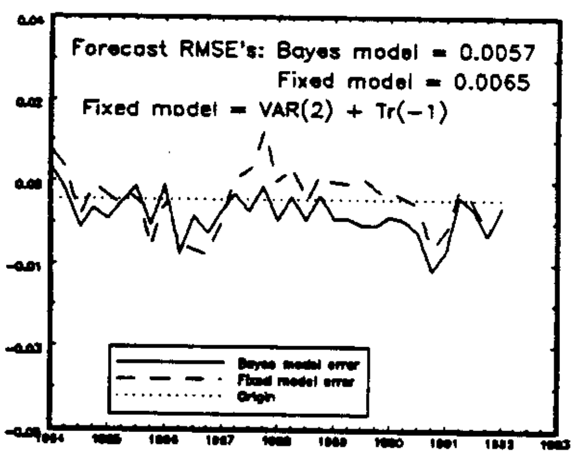


Figure 16: Un: 1959:1–1992:1

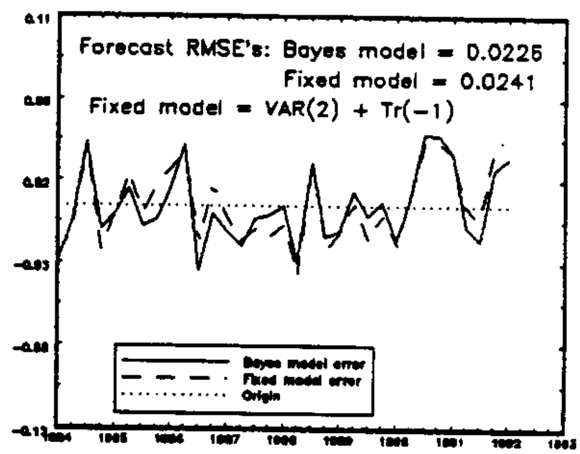
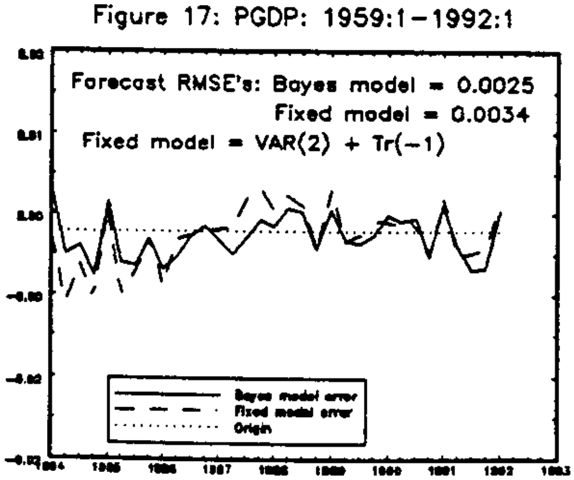


Figure 17: PGDP: 1959:1–1992:1



Figures 18–22: Prediction Errors for a Bayes Model$^{\text{coint. + non-coint.}}$ VAR

Figure 18: T_Bill^{-1}: 1959:1 – 1992:1

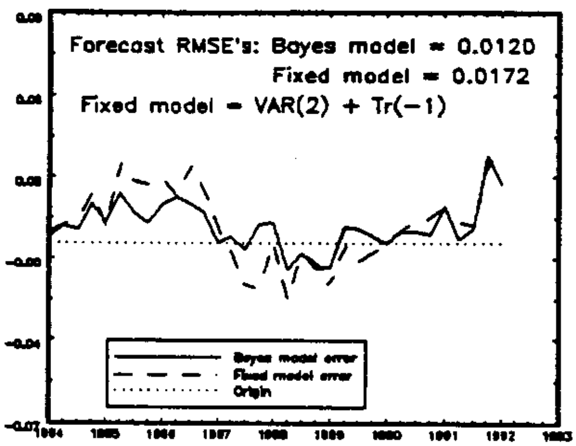


Figure 19: M1: 1959:1–1992:1

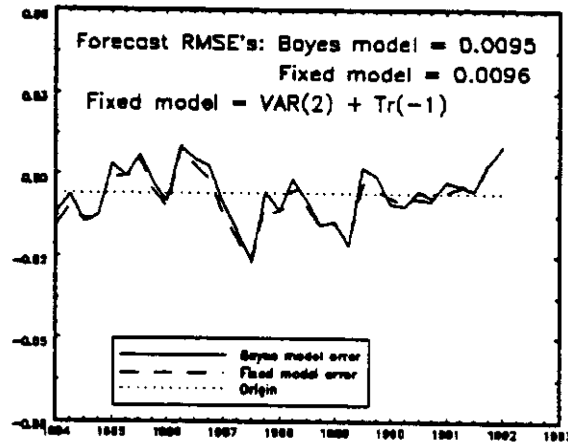


Figure 20: GDP: 1959:1–1992:1

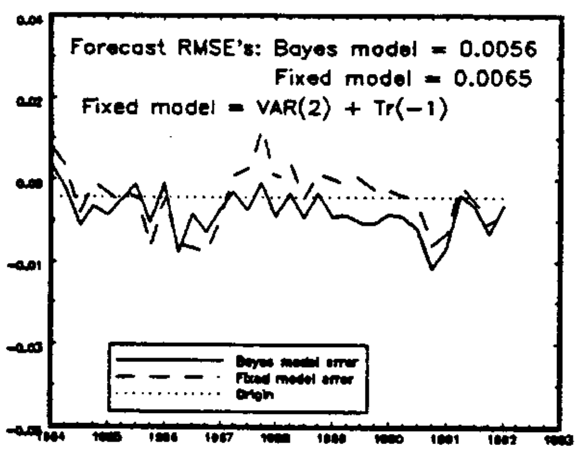


Figure 21: Un: 1959:1–1992:1

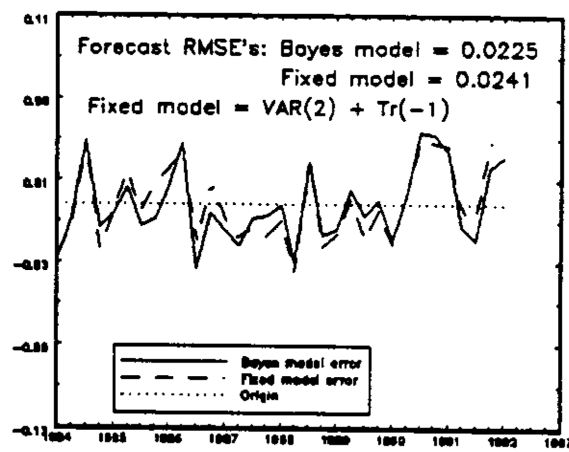
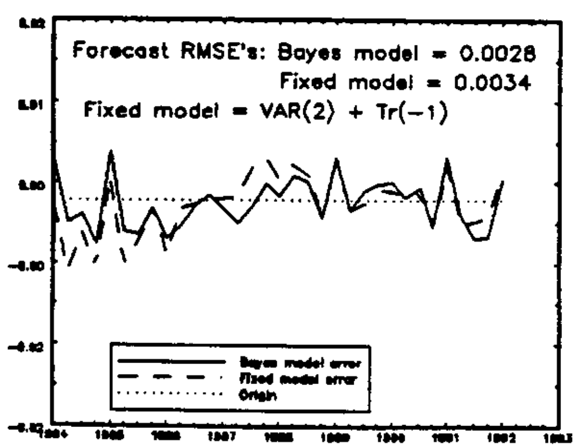


Figure 22: PGDP: 1959:1–1992:1



Figures 23–27: Prediction Errors for a 'Bayes Model^{diag}' VAR & BVARM model

Figure 23: T_Bill^{-1}: 1959:1 – 1992:1

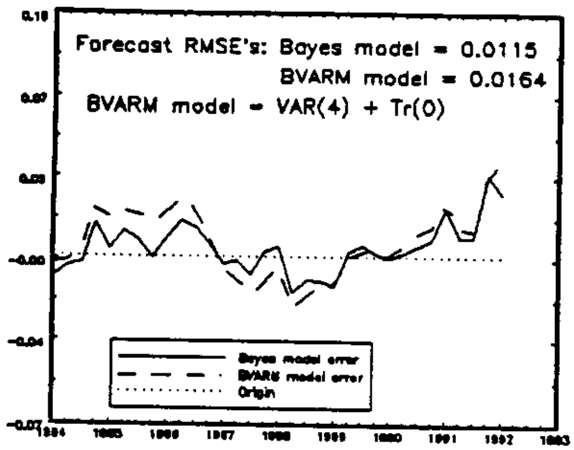


Figure 24: M1: 1959:1–1992:1

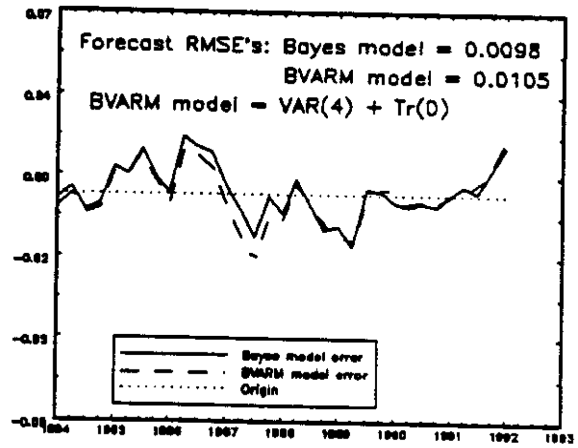


Figure 25: GDP: 1959:1–1992:1

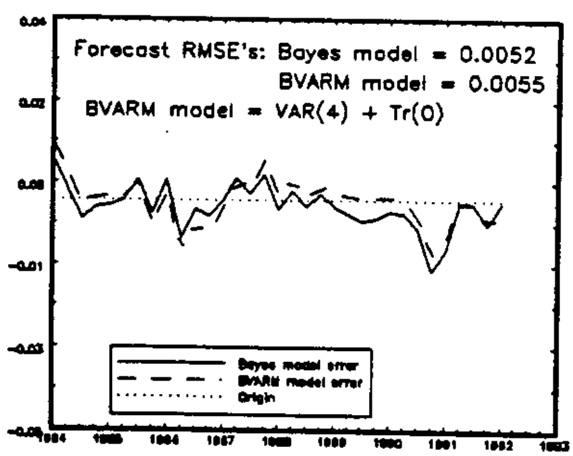


Figure 26: Un: 1959:1–1992:1

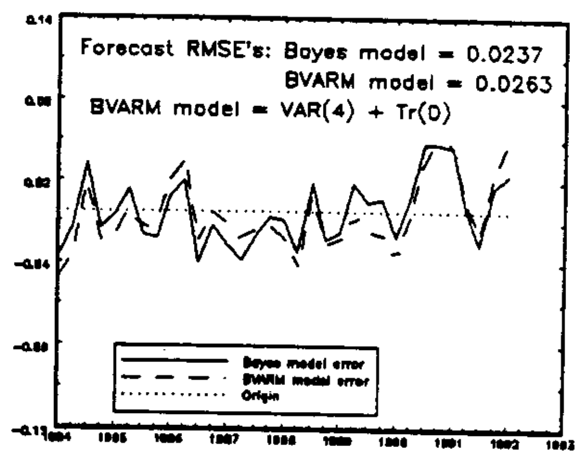


Figure 27: PGDP: 1959:1–1992:1