COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale University
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 1023

BAYESIAN MODEL SELECTION AND PREDICTION WITH EMPIRICAL APPLICATIONS

by

Peter C.B. Phillips

July 1992

BAYESIAN MODEL SELECTION AND PREDICTION WITH EMPIRICAL APPLICATIONS

by

Peter C. B. Phillips*
Cowles Foundation for Research in Economics, Yale University
and
University of Auckland

O. ABSTRACT

This paper builds on some recent work by the author and Werner Ploberger (1991, 1992) on the development of "Bayes models" for time series and on the authors' new model selection criterion "PIC". The PIC criterion is used in this paper to determine both the lag order and the trend degree in an autoregression with deterministic trend. A new forecast encompassing test for "Bayes models" is developed which allows one Bayes model to be compared with another on the basis of their respective forecasting performance. The paper reports an extended empirical application of the new methodology to the Nelson-Plosser (1982)/Schotman-Van Dijk (1991) data. It is shown that simple, parsimonious "Bayes models" forecast-encompass fixed "Bayes models" of the "AR(3) + linear trend" variety for most of these series. In some cases, the forecast performance of the parsimonious "Bayes models" is substantially superior. The results cast doubt on the value of working with fixed format time series models in empirical research and demonstrate the practical advantages of evolving format "Bayes models." The paper makes a new suggestion for modelling interest rates in terms of reciprocals of levels rather than levels (which display more volatility) and shows that the best data-determined model for this transformed series is a martingale.

JEL Classification No. 211
Keywords: Bayes model; Bayes measure; BIC; Forecast; Forecast-encompass; Model selection; PIC; Unit root

First Draft: May, 1992

This Version: July, 1992

*All of the computations reported in this paper were performed by the author on a 486-33 PC using programs written in GAUSS-386i (Version 2.2). The author thanks the NSF for research support under Grant No. SES 9122142 and thanks Glena Ames for her skill and effort in keyboarding the manuscript.

1. INTRODUCTION

A feature of the Bayesian approach to inference that is especially important in time series applications is that the analysis is conducted conditional on the realized history of the series. In two earlier papers (1991, 1992), the author and Werner Ploberger have studied the implications of this data conditioning and have shown that its mathematical effect is to translate the underlying reference model to a "Bayes model" whose parameters are time varying and data dependent and whose error process is conditionally heterogeneous. This "Bayes model" is in fact just a location model, where the systematic part of the model is time varying and gives the current best estimate (using prior information and the available data) of the *location* of the dependent variable in the next period. Associated with this "Bayes model" is a σ-finite measure. The Phillips-Ploberger (1992) paper shows how to use this measure to construct posterior odds for evaluating one "Bayes model" against another. The resulting statistic is a new model selection criterion "PIC" (a posterior odds information criterion) which can be used to compare models and to test hypotheses (like the presence of a unit root).

The purpose of the present paper is to show that the "PIC" criterion can also be used to compare models on the basis of their respective forecasting performance. We give a version of the PIC criterion that is a forecast encompassing test of one Bayes model against another. This test determines whether one model encompasses another on the basis of its forecast performance. In this sense, it is a Bayesian version of classical forecast encompassing tests -- see Chong-Hendry (1986) in particular.

Our methodology, however, is rather different from the usual Bayesian and classical approaches. We use our PIC criterion to select what the data supports as the best "Bayes model" period by period. As new data arrives, we allow our model itself to evolve. Not only are the parameters updated as the new data arrives but also the form of the "Bayes model" itself may change with the learning process. For instance, whereas a model with a unit root and no deterministic trend may be selected in one period, in future periods we may find a

"Bayes model" with a linear trend or a mildly explosive autoregressive coefficient favored over the unit root specification. In our methodology, therefore, the "Bayes model" is part of the learning process and our forecast encompassing test allows us to compare such a sequence of Bayes models against corresponding "Bayes models" of fixed format in which the parameters are updated but the model format is not. The test enables us to determine empirically whether parsimonious models of possibly evolving format can outperform fixed format models. The advantage of our criterion is that in making such model comparisons there is a built-in penalty for employing more parameters in making a forecast.

These ideas and our model selection methodology are implemented empirically with the historical time series for the USA used by Nelson-Plosser (1982) and Schotman-Van Dijk (1991). Bayes models are constructed and estimated with the data up to 1969. These models are then allowed to evolve as data over the period 1970-1988 accumulates. The form of the evolving model is monitored and its forecasting performance is tracked against that of a "Bayes model" of fixed format. Encompassing tests are then constructed to determine whether the evolving model outperforms the fixed model in terms of its forecasting performance. The empirical results are striking. For all but four of the series (industrial production, employment, consumer prices and stock prices) it is possible to encompass the forecasts of a fixed format "AR(3) + linear trend" model using a parsimonious, evolving "Bayes model" that often has only one fitted parameter.

2. MODEL SELECTION BY "PIC"

The model framework of this paper is the same as that in Phillips-Ploberger (1992). The set up is the linear regression

$$y_t = \beta' x_t + \varepsilon_t\,, \quad (t = 1, 2, \dots) \tag{1}$$

whose dependent variable y_t and error ε_t are real valued stochastic processes on a probability space $(\Omega, \mathcal{F}, P)$. Accompanying y_t is a filtration $\mathcal{F}_t \subset \mathcal{F}$ $(t = 0, 1, 2, \dots)$ to which both y_t and ε_t are adapted. The regressors x_t $(k \times 1)$ in (1) are defined on the same space and are assumed to have the property that x_t is $\mathcal{F}_{t-1}$-measurable. The errors ε_t satisfy $E(\varepsilon_t|\mathcal{F}_{t-1}) = 0$, so that the conditional mean function in (1) is correctly specified.

A general example of (1) is the "ARMA(p, q) + trend (r)" model, which is convenient to write in difference format as

(2) $$\Delta y_t = hy_{t-1} + \Sigma_{i=1}^{p-1}\varphi_i \Delta y_{t-i} + \Sigma_{j=1}^{q}\psi_j \varepsilon_{t-j} + \Sigma_{k=0}^{r}\delta_t t^k + \varepsilon_t .$$

Here, there are $k = p + q + r + 1$ parameters. In this case some of the regressors, viz. the ε_{t-j}, are not directly observable. The difficulty can be accommodated by means of recursive techniques such as the Hannan-Rissanen (1982) algorithm which permit the use of estimates of the ε_{t-j} from a preliminary long autoregression. This method was implemented in the present context in the Phillips-Ploberger (1992) paper, to which the reader is referred for a full description of the algorithm.

When $q = 0$ in (2) the model is an AR(p) + trend (r). When $h = 0$ the model has one autoregressive unit root. When $r = -1$ there is no intercept in the model, when $r = 0$ there is an intercept, and when $r = 1$ there is a linear trend. These are the main specializations of (2) that are of interest in empirical applications.

The theory in Phillips-Ploberger (1992) is developed for the model (1) with Gaussian errors $\varepsilon_t \equiv$ iid $N(0, \sigma^2)$. The treatment of the nuisance parameter σ^2 is classical. In particular, our model selection criterion "PIC" is developed conditional on σ^2, and then for practical implementation σ^2 is replaced by its least squares (or maximum likelihood) estimate from the most complex model in the class under consideration.

Phillips-Ploberger show that there is a "Bayes model" corresponding to the "classical" parametric model (1). (Here we use the term "classical" to signify that in (1) there is a "true value" of the parameter β under which ε_t has the properties ascribed to it.) We use the regression notation $Y_n' = [y_1, ..., y_n]$, $X_n' = [x_1, ..., x_n]$ and set $A_n = X_n' X_n$. Then, the "Bayes model" corresponding to (1) has the form

(4) $$y_t = \hat{\beta}_{t-1}' x_t + v_t , \text{ where } v_t|\mathcal{F}_{t-1} \equiv N(0, f_t)$$

with

(5) $$f_t = \sigma^2\{1 + x_t' A_{t-1}^{-1} x_t\}$$

and where $\hat{\beta}_{t-1} = (X_{t-1}'X_{t-1})^{-1}X_{t-1}'Y_{t-1}$ is the least squares estimate of β based on information in $\mathcal{F}_{t-1}$.

The Phillips-Ploberger analysis shows that, under a uniform prior on β and a Gaussian likelihood for Y_n, the passage via Bayes rule to the posterior density of β implies the replacement of the model (1) by the time varying parameter model (4). That is, the appropriate reference frame for a Bayesian analysis is the model (4), thereby justifying the terminology "Bayes model." In the earlier paper we interpreted (4) as a location model where $y_{t|t-1} = E(y_t|\mathcal{F}_{t-1}) = \hat{\beta}_{t-1}' x_t$ is the best estimate of the location of y_t given information in $\mathcal{F}_{t-1}$. This location estimate is identical to the maximum likelihood estimate of the best predictor of the next period observation, i.e. it is precisely the predictor we would use in classical inference. From this perspective and as far as the model that is actually used to make predictions is concerned, there is no difference between the Bayesian and classical approaches.

The probability measure associated with the "Bayes model" (4) is a forward looking measure than can be described by the conditional density of y_t given $\mathcal{F}_{t-1}$. This density is given by

(6) $$dQ_t/dQ_{t-1} = \mathrm{pdf}(y_t|\mathcal{F}_{t-1}) = (2\pi f_t)^{-1/2}\exp\{-(1/2f_t)v_t^2\} \equiv N(0, f_t)\ ,\ t = k+1, k+2, \ldots$$

and it is defined as soon as there are enough observations in a trajectory to estimate the k-vector β. Thus, (4) and (6) are defined for $t \geq k+1$. The measure Q_t that appears in (6) is the "Bayes model" measure, i.e. the measure corresponding to the "Bayes model" (4). This measure is σ-finite and, as shown in Phillips-Ploberger (1992), can also be defined in terms of the following Radon Nikodym (RN) derivative

(7) $$\frac{dQ_t}{dP_t} = |(1/\sigma^2)A_t|^{-1/2} \exp\left\{(1/2\sigma^2)\hat{\beta}_t' A_t \hat{\beta}_t\right\},$$

which is taken with respect to the reference measure P_t for the model (1) in which $\beta = 0$ (i.e. the probability measure of the $N(0, \sigma^2 I_t)$ distribution).

Once the measure Q_t is defined, either directly as in (7) or recursively as in (6), the measure can be used to compare models and test hypotheses. The mechanism is simply the likelihood ratio of the respective measures of the two competing models. Thus, if Q_n^k is the measure of a Bayes model such as (4) with k parameters and n observations and Q_n^K is the corresponding measure of a model in the same class but with K parameters then we compare the models using the RN derivative

$$
\begin{aligned}
(8) \qquad dQ_n^k/dQ_n^K &= (dQ_n^k/dP_n)/(dQ_n^K/dP_n) \\
&= |(1/\sigma^2)A_n(k)|^{-1/2}|(1/\sigma^2)A_n(K)|^{-1/2}\exp\left\{(1/2\sigma^2)[\hat{\beta}_n(k)A_n(k)\hat{\beta}(k) - \hat{\beta}_n(K)'A_n(K)\beta_n(K)]\right\} .
\end{aligned}
$$

This likelihood ratio measures the support in the data for the more restrictive model (with k parameters)

$$H(Q_n^k) : y_{n+1} = \hat{\beta}_n(k)'x_{n+1}(k) + v_{n+1}(k) ,$$

against the more complex model (with K parameters)

$$H(Q_n^K) : y_{n+1} = \hat{\beta}_n(K)'x_{n+1}(K) + v_{n+1}(K) .$$

When we assign equal prior odds to the two competing models our decision criterion is to accept $H(Q_n^k)$ in favor of $H(Q_n^K)$ when $dQ_n^k/dQ_n^K > 1$. This posterior odds criterion is, in fact, a method of order selection that compares the posterior predictive densities of the two competing models.

The model selection criterion suggested in Phillips-Ploberger (1992) is based on (8) and uses the more complex model with K regressors to estimate σ^2. Let $\hat{\sigma}_K^2$ be the maximum likelihood estimate of σ^2 from this model. Then the order estimator satisfies

$$(9) \qquad \hat{k} = \text{argmin}_k \, \text{PIC}_k$$

where

$$(10) \qquad \text{PIC}_k = (dQ_n^K/dQ_n^k)(\hat{\sigma}_K^2) .$$

Observe that $\hat{k}$ maximizes $1/\text{PIC}_k = dQ_n^k/dQ_n^K(\hat{\sigma}_K^2)$ and thereby selects the model most favored over $H(Q_n^K)$ according to the predictive density.

An alternative way of writing the PIC criterion (10) is to use the predictive densities implied directly by the competing Bayes models, i.e. $H(Q_n^k)$ and $H(Q_n^K)$. If we compare the densities for these models over the same subsample of data, say $n > K$, we have

$$PIC_k' = dQ_n^k/dQ_n^K(\hat{\sigma}_K^2)|_{\mathcal{F}_K} = \Pi_{K+1}^n\left(\hat{f}_t^K/\hat{f}_t^k\right)^{1/2}\exp\left\{\Sigma_{K+1}^n[v_t(K)^2/2\hat{f}_t^K - v_t(k)^2/2\hat{f}_t^k]\right\} \tag{11}$$

where

$$\hat{f}_t^k = \hat{\sigma}_K^2(1 + x_t(k)'A_{t-1}(k)^{-1}x_t(k)) \ , \ \hat{f}_t^K = \hat{\sigma}_K^2(1 + x_t(K)'A_{t-1}(K)^{-1}x_t(K)) \ ;$$

$$v_t(k) = y_t - \hat{\beta}_{t-1}(k)'x_t(k) \ , \ v_t(K) = y_t - \hat{\beta}_{t-1}(K)'x_t(K)$$

and $\hat{\sigma}_K^2$ is the least squares estimate of the error variance in the more complex model $H(Q_n^K)$. This predictive form of the PIC criterion will be especially useful in the development of the forecast encompassing test in the next section.

As it stands, (10) may already be interpreted as an encompassing test statistic. For, if $dQ_n^k/dQ_n^K(\hat{\sigma}_K^2) > 1$ the evidence in the sample suggests that the density for the model with k parameters exceeds the density of the model with K parameters when both are evaluated at the sample data. This is equivalent to saying that the model with k parameters encompasses the model with K parameters in terms of their respective probability densities.

An alternative Bayesian approach to the encompassing principle has recently been developed in a series of works by Florens (1990), Florens and Mouchart (1989), Florens, Mouchart and Rolin (1990) and Florens, Mouchart and Larribeau-Nori (1992). In this work the distance between the respective posterior distributions of a parameter of interest in two competing models is measured by the Kullback-Liebler divergence of the two densities. Critical values of the divergence statistic are computed by simulation and parameter encompassing is supported by the data when the observed divergence is smaller than the critical value. This procedure may be regarded as a parameter encompassing test, whereby the posterior distribution of a parameter under one model is "explained" or "encompassed" by that of another model if the Kullback-Liebler divergence is small. (The procedure is complicated in practice by the fact that the parameter may not occur naturally in one of the models and must then be replaced by what the authors call a Bayesian pseudo true value.) This alternative Bayesian approach to the concept of encompassing is obviously of interest. Like our approach it keeps sample space considerations alive beyond the computation of the likelihood. Beyond this, our approaches are very different, not only in terms of the measures employed (viz., the Kullback-Liebler divergence as compared with our RN derivative of the respective Bayes measures) but also the objects of comparison (viz., the posterior densities as compared with our predictive densities).

3. A "BAYES MODEL" FORECAST ENCOMPASSING TEST

An important element in evaluating any econometric model's performance is its forecasting capability. Many procedures are now available and the literature on the subject is diverse. An approach to this subject that is closest conceptually, at least, to our own is due to Chong and Hendry (1986). These authors critiqued some of the more traditional methods of evaluation such as those based on a model's dynamic simulation tracking performance and its historical record of forecast accuracy. In place of these measures, Chong-Hendry suggested the use of a simple t-test of forecast encompassing to determine whether one model's forecasts can encompass those of a rival model. The test is mounted as a regression t-test on the coefficient in the regression of the forecast errors from a base model on the forecasts of the rival model. When the test is insignificant the base model's forecasts are said to encompass those of the rival model. The test is justified asymptotically and requires only the forecasts from the competing models together with the ex post sample data.

Our own approach is to assess the forecasting capability of rival models in terms of our model selection criterion "PIC". In such an exercise the predictive form of our criterion "PIC" given in (11) is ideally suited. Let us suppose that we wish to compare the "Bayes models" $H(Q_n^k)$ and $H(Q_n^K)$ in terms of their respective performance in one-period ahead forecasts over the period $t = n+1, ..., N$. Our "Bayes model" forecast encompassing test statistic would be

$$(12) \qquad dQ_N^k/dQ_N^K(\hat{\sigma}(K)^2) = \prod_{t=n+1}^{N} \left(g_t^K/g_t^k\right)^{1/2} \exp\left\{-(1/2\hat{\sigma}_t^2(K)g_t^k)v_t(k)^2 + (1/2\hat{\sigma}_t^2(K)g_t^K)v_t(K)^2\right\}$$

where the notation follows that employed in (11) except that $g_t^k = 1 + x_t(k)'A_{t-1}(k)^{-1}x_t(k)$ and $\sigma_t^2(K) = (Y_t - X_t(K)\hat{\beta}_{t-1}(K))'(Y_t - X_t(K)\hat{\beta}_{t-1}(K))/(t-K)$ is the least squares estimate of the error variance σ^2 in the "Bayes model" $H(Q_t^K)$.

More generally, we want the "Bayes model" to evolve as we accumulate more observations. Thus, over the forecast horizon $t = n+1, ..., N$ we want to allow the new data to assist in selecting the most appropriate model. This can be achieved quite simply by using the PIC criterion to select the best "Bayes model" period by period. Let

$$PIC_k(t) = dQ_t^K/dQ_t^k(\sigma_t^2(K))$$

and for each period $t = n, ..., N-1$ choose the model according to:

$$\hat{k}_t = \operatorname{argmin} PIC_k(t) \ . \tag{13}$$

The model $H(Q_t^{\hat{k}_{t-1}})$ is then used to generate the forecast for the next time period.

Suppose we wish to compare the "Bayes model" sequence $H(Q_t^{\hat{k}_{t-1}})$ with a sequence of "Bayes models" $H(Q_t^F)$ with a fixed number of parameters (F). We can make the comparison as the basis of the respective one-period ahead forecasting performance of the two models over the period $t = n+1, ..., N$. The test statistic is

$$\begin{aligned} &dQ_N^B/dQ_N^F(\hat{\sigma}^2(\hat{k})) \\ &= \prod_{t=n+1}^{N} \left(g_t^F/g_t^{\hat{k}_{t-1}}\right)^{1/2} \exp\left\{-\left(1/2\hat{\sigma}_t^2(\hat{k}_{t-1})g_t^{\hat{k}_{t-1}}\right)v_t(\hat{k}_{t-1})^2 + (1/2\hat{\sigma}_t^2(\hat{k}_{t-1})g_t^F)v_t(F)^2\right\} \end{aligned} \tag{14}$$

where we use Q_n^B in place of $Q_n^{\hat{k}}$ for ease of notation. On the basis of their one-period ahead forecasting performance over $t = n+1, ..., N$ we would favor the sequence of "Bayes models" $\{H(Q_t^{\hat{k}_{t-1}})\}_{n+1}^{N}$ over the sequence of fixed format "Bayes models $\{H(Q_t^F)\}_{n+1}^{N}$ if

$$dQ_n^B/dQ_N^F(\hat{\sigma}^2(\hat{k}))) > 1 \ . \tag{15}$$

This test is a "Bayes model" forecast encompassing test. If (15) holds we conclude that the sequence $H(Q_t^{\hat{k}_{t-1}})$ generates forecasts over $t = n+1, ..., N$ that encompass the forecasts of the fixed format sequence of models $H(Q_t^F)$. Note that in (14) and (15) we use $\hat{\sigma}_t^2(\hat{k}_{t-1})$ to estimate the error variance. This is because $\hat{k}_{t-1}$ is consistent for k and hence $\hat{\sigma}_t^2(\hat{k}_{t-1})$ is consistent for σ^2 in both models when (1) is the actual generating mechanism. This choice of variance estimate allows for the possibility that the number of lags or the trend degree in the fixed model may be too small, whereas this will not be the case in the "Bayes model," at least when the sample size is large enough, because $\hat{k}$ is consistent.

The properties of the forecast encompassing test (15) follow from Theorem 3.3 of Phillips-Ploberger (1992). If the fixed format model sequence $H(Q_t^F)$ is overparameterized in the sense that $F > \hat{k}_{t-1}$ for infinitely many t as $N \to \infty$ then $dQ_n^{\hat{k}}/dQ_n^F(\hat{\sigma}^2(F))$ diverges to ∞. Thus, we will always choose $H(Q_t^{\hat{k}_{t-1}})$ over $H(Q_t^F)$ as $N \to \infty$ if there is a true model of the data with fewer parameters than F.

However, the most important advantage of $H(Q_t^{\hat{k}_{t-1}})$ is that the sequence of models adapt to the data. If for some periods a model with fewer parameters than F is supported by the data then criterion (13) will choose that model. (Equally, if in other periods a more complex model is required then the criterion will choose a model with more parameters than F.) The forecast encompassing test (15) then determines whether we pay a price in forecasting performance for choosing the more parsimonious model. This will be so if in some periods the model reverts to a model with more parameters. Note, however, that the price paid for parsimony is generally small even in this case. For if the generating mechanism does revert to a model with more parameters the learning mechanism in the period by period choice of model using (13) is rapid, so that if the change is an important one the "Bayes model" sequence $H(Q_t^{\hat{k}_{t-1}})$ should quickly accommodate it.

4. EMPIRICAL APPLICATION

The methods of the last two sections were applied to the fourteen historical time series of the USA economy studied originally by Nelson-Plosser (1982) and extended recently by Schotman-Van Dijk (1991). We took advantage of the 18 years' extension of these series to examine "Bayes model" one-period ahead forecasts over this period and to implement our "Bayes model" forecast encompassing test.

In performing this forecasting exercise we evaluate our best "Bayes model" sequence $\{H(Q_t^{k_{t-1}})\}_{t=n+1}^{N}$ against a fixed format "Bayes model" sequence. The best "Bayes model" is chosen from the "AR(p) + trend (r)" class using the PIC procedure described in Section 2. The parameterization chosen for the fixed format model is the "AR(3) + linear trend" model that has been a common choice in recent empirical work with traditional Bayesian methods (e.g. DeJong-Whiteman, 1991). This model is also updated period by period in the sense that the latest data is used to revise parameter estimates as we move through the forecast period. Hence, the difference between the fixed format "Bayes model" and our best "Bayes model" sequence is that in the latter the model orders of both the deterministic trend and the lag order are chosen (by PIC) period by period and in each period the best "Bayes model" incorporates the outcome of a unit root test (again by PIC). Thus, $H(Q_t^{k_{t-1}})$ is an evolving sequence of best "Bayes models" whose form is entirely data-based, being determined by our model selection criterion PIC.

Figures 1-14 show the one period ahead forecasting performance of these two model sequences over the period 1970-1988 inclusive. In each case Figure (a) displays the data and the relevant forecast period, and Figure (b) shows the period by period forecast errors from the two models. Figure (c) gives details of the evolving form of the best "Bayes model": the lines on the graph show the autoregressive lag order selected (0-6 lags), the trend degree (-1 = no intercept; 0 = fitted intercept; 1 = fitted linear trend), and whether or not a unit autoregressive root is selected (-1 = yes, 0 = no). Figure (d) recursively plots the encompassing test statistic dQ^B/dQ^F over the forecast period. Table 1 tabulates these details, gives the root mean squared error (RMSE) of forecasts for the two models over the forecast period, and records the evolving format of the best "Bayes model."

The main items of interest to emerge from this empirical forecasting exercise are as follows:

(i) For only one series (industrial production) is an "AR(p) + T(1)" model (i.e. an autoregression with a linear trend) accepted as the best "Bayes model" over the whole forecasting

TABLE 1: EMPIRICAL RESULTS FOR HISTORICAL U.S. TIME SERIES IN FORECASTING EXERCISES OVER 1970-1988

Series	Forecast RMSE's		Number of changes in "Bayes model" (date of change)	Best "Bayes model"	Parameter count ratio: "Bayes model"/ fixed model	Forecast encompassing test dQ^b/dQ^f in 1988
	Bayes model	Fixed model				
Real GNP	0.0264	0.0239	1 (1979)	$AR(1)^{-1}$; $AR(2)+T(1)$	1/5; 4/5	1.0037
Nominal GNP	0.0444	0.0438	0	$AR(2)^{-1}$	1/5	1.6838
Real p.c. GNP	0.0249	0.0237	1 (1978)	$AR(2)^{-1}$; $AR(2)+T(1)$	1/5; 4/5	1.1123
Industrial production	0.0531	0.0508	0	AR(1)+T(1)	3/5	0.8289
Employment	0.0191	0.0166	1 (1974)	$AR(2)^{-1}$; $AR(4)+T(1)$	1/5; 6/5	0.5143
Unemployment	0.1563	0.1746	0	$AR(4)+T(0)$	5/5	1.8587
GNP deflator	0.0297	0.0323	0	$AR(2)^{-1}$	1/5	3.2322
Consumer prices	0.0322	0.0277	0	$AR(4)^{-1}$	4/5	0.3727
Nominal wages	0.0292	0.0297	0	$AR(2)^{-1}$	1/5	1.8417
Real wages	0.0267	0.0358	0	$AR(2)^{-1}$	1/5	53.7792
Money stock	0.0320	0.0331	1 (1984)	AR(2)+T(1); $AR(2)^{-1}$	4/5; 1/5	1.5866
Velocity	0.0218	0.0260	0	$AR(1)^{-1}$	0/5	2.8919
$(\text{Bond yield})^{-1}$	0.0146	0.0151	0	$AR(1)^{-1}$	0/5	2.2084
Bond yield*	1.1866	1.3077				
Stock price	0.1515	0.1421	3 (1972,1973,1984)	$AR(2)^{-1}$; $AR(1)^{-1}$; $AR(2)^{-1}$; $AR(1)^{-1}$	1/5; 0/5; 1/5; 0/5	0.4232

Legend: * forecasts for "Bond yield" series obtained from models for (Bond yield^{-1})
$AR(p)^{-1}$ = $AR(p)$ model with a unit autoregressive root

period. For four series (real GNP, real p.c. GNP, employment and money stock) an "AR(p) + $T(1)$" model is chosen for certain subperiods of the sample as the best model.

(ii) Bayes models with a unit root are selected for 12 of the series, four of these in subperiods (real GNP, real p.c. GNP, employment and money stock)

(iii) Only two series are chosen to be stationary about a level or linear trend (unemployment and industrial production).

(iv) The best "Bayes model" encompasses the forecasts of the fixed model for ten of the series over the full period, and in some of these cases by a very wide margin (e.g. real wages, GNP deflator and velocity). The case of the real wage series is especially interesting. Here it is apparent from Figure 10(b) that the fixed model produces systematically biased forecasts of real wages. Clearly, there is a substantial cost in terms of forecast capability to including a linear trend in a model for this series. Thus, although the best "Bayes model," which is an AR(2) with a unit root (i.e. a one parameter model), is nested within the fixed model (the five parameter AR(3) + $T(1)$" model) the more parsimonious model has substantially superior forecasting performance. This is explained by the fact that the larger model adapts slowly to the effects of new observations whereas the smaller model is more flexible. The data graph in Figure 10(a) shows clearly that a linear trend is less realistic over the full data set 1900-1988 than it is over the sample period 1900-1968. Thus, even though the trend coefficient in the fixed model is revised period by period, the presence of the trend in this model is a form of misspecification and is thereby responsible for the systematic bias in the model's forecasts. Figure 10(c) shows that our model selection criterion eliminates the trend in the best "Bayes model." In effect, our criterion detects the fact that the penalty from including the trend is too great. The outcome is a parsimonious and flexible Bayes model (with only one fitted parameter) whose forecasting performance almost uniformly dominates that of the fixed model. The forecast encompassing statistic for this series is $dQ_n/dQ_n = 53.7792$. Thus, on the basis of their respective forecast performance the odds in favor of the one parameter "Bayes model" over the "AR(3) + $T(1)$" model are more than 53:1.

(v) Forecast accuracy is measured directly by the root mean squared error (RMSE) of forecasts over the period 1970-1988. In terms of this measure, the best "Bayes models" are superior for seven of the series (unemployment, GNP deflator, real wages, nominal wages, money stock, velocity and bond yields). For some series the reduction in the RMSE of forecast is substantial, as in the case of real wages where the forecast accuracy improves by 25. The improvement in forecast performance is even more dramatic when the parameter ratio (1:5) of the two models is taken into account. Another series where a parsimonious, one parameter "Bayes model" does especially well is the GNP deflator series where the value of the forecast encompassing test statistic dQ_n/dQ_n is 3.2322. For two of the series, velocity and bond yields, the best "Bayes model" is a random walk. This model, with no fitted parameter, outperforms the fitted model for both series in actual forecast performance (i.e., they have smaller forecast RMSE's) and the odds in favor of the random walk model over the fixed model are 2.89:1 and 2.2:1, respectively. From the recursive plots of the dQ_n/dQ_n statistic shown in Figure 12(d) it is apparent that the simple random walk model uniformly dominates the fixed model over the forecast period for the velocity series. The dominance is close to uniform for the bond yield series.

(vi) The bond yield series deserves extra attention. The graph of this data series for the full period 1900-1988 is shown in Figure 13′(a). Clearly this series shows much more volatility over the latter part of the sample. We therefore employed the variance stabilizing transformation $x \to 1/x$. The resulting series in $(\text{levels})^{-1}$ is shown in Figure 13(a), which displays more homogeneous variance over the full sample. Interestingly, the best "Bayes model" for this series is a martingale, whether the series is taken in levels or in reciprocals of levels, i.e. $(\text{levels})^{-1}$. Since our test criterion dQ_n/dQ_n is based on a Gaussian model with homogeneous variance we used the two models for this series taken in $(\text{levels})^{-1}$ form and computed the recursive values of dQ_n/dQ_n shown in Figure 13(d) from these models' forecasts. The models for the series in $(\text{levels})^{-1}$ form were then used to compute forecasts of the series in levels and the resulting forecast performance of the two models is shown in Figure 13′(b). In both cases (i.e. in both levels and $(\text{levels})^{-1}$ form), the best "Bayes model" (which is here a martingale in $(\text{levels})^{-1}$) clearly outperforms the fixed model.

5. CONCLUSION

This paper utilizes a new Bayesian encompassing test to evaluate models on the basis of their one-period ahead forecasting performance. The models compared are "Bayes models" whose estimated coefficients are updated period by period as new data becomes available. One of the models has a fixed parametric form, which for the empirical exercises conducted here is the "AR(3) + linear trend" model that has frequently been used in empirical work with macroeconomic time series. The other model is the best "Bayes model" whose parametric form is determined period by period using the model selection criterion PIC. This model is "best" in the sense that, on the basis of the sample period data, the model chosen has the highest posterior odds in relation to a general model in the "AR(p) + trend (r)" class. The best "Bayes model" is allowed to evolve in form (and, hence, also in terms of its number of fitted parameters) period upon period during the forecast interval as the new data accumulates.

These new modelling methods are applied to the Nelson-Plosser/Schotman-van Dijk historical macroeconomic time series for the USA economy. The best "Bayes models" are found to be parsimonious (often with as few as one or no parameters) and to do very well in actual forecasts over the 1970-1988 period. For ten of the fourteen series, the best "Bayes models" encompass the fitted format models on the basis of their respective forecasting capability over 1970-1988. For some series (like the real wage, GNP deflator and bond yields) the improvement in actual forecasting performance is substantial, especially when the parsimony of these models (which have only one fitted parameter in these cases against the five fitted parameters of the fixed model) is taken into account. The bond yield series is particularly interesting. In contrast to previous empirical investigations which work with levels of this series, we find that reciprocals of levels rather than levels is the form more suited to empirical implementation. The series is, in fact, well modelled by a martingale in both forms but, in levels, has a much more volatile conditional error variance. Forecasts from the best "Bayes model" (an AR(1) with a unit root) outperform those of the fixed model in both cases.

Overall, these results seem very promising for the use of Bayesian model selection principles and data-based evolving "Bayes models" in empirical applications. Further applications of these methods and extensions of the methodology to a wider class of base models are now under way. These results and those of the present paper are sufficiently encouraging for us to put forward a suggestion for empirical econometric modelling. Formally stated, the principle that we suggest cautions against the use of fixed format models that are not data-determined. We state it as follows:

THE EXXON VALDEZ PRINCIPLE. *Fixed format time series models do not adapt fast enough to new data, just as big tankers cannot stop or turn corners in a hurry.* □

In contrast, parsimonious "Bayes models" of the type employed in this paper adapt to new data by evolving in form as well as by updating parameters. When changes occur these models adapt more rapidly than fixed format time series models. As a tool of modelling data they are more flexible and as a tool of prediction they are much less prone to serious error. In the latter connection we observe that evolving "Bayes models" are very reluctant (according to our data-based choice criterion PIC) to include a deterministic trend in an empirical model. Thus, for the Nelson-Plosser data a trend is included in the evolving model only for the industrial production series over the entire period 1970-1988. Deterministic trends are regressors with more leverage than conventional stationary regressors. As a result, they have the potential for being very powerful predictors. On the other hand, when a trend is inappropriate the potential for seriously biased forecasts is substantial (as evidenced by the case of the real wage series). Our results therefore indicate that there are some serious costs to the mechanical inclusion of deterministic trends in time series models. For most (specifically, 13 out of 14) of the Nelson-Plosser/Schotman-van Dijk series the data does not support the inclusion of deterministic trends. When they are included, the result is generally inferior forecasting capability.

6. REFERENCES

Chong, Y. Y. and D. F. Hendry (1986). "Econometric evaluation of linear macro-economic models," *Review of Economic Studies*, 53, 661-690.

DeJong, D. N. and C. H. Whiteman (1991). "Trends and random walks in macroeconomic time series: A reconsideration based on the likelihood principle," *Journal of Monetary Economics*.

Florens, J.-P. (1990). "Parameter sufficiency and encompassing," pp. 115-136 in *Essays in Honour of Edmond Malinvaud*. Cambridge: MIT Press.

Florens, J.-P. and M. Mouchart (1989). "Model selection: Some remarks from a Bayesian viewpoint," pp. 27-44 in J.-P. Florens, M. Mouchart, J.-P. Raoult and L. Simar (eds.), *Model Choice*. Bruxelles: Publications des Facultes Universitaires, Saint Louis.

Florens, J.-P., M. Mouchart and J.-M. Rolin (1990). *Elements of Bayesian Statistics*. New York: Marcel Dekker.

Florens, J.-P., M. Mouchart and Sophie Larribeau-Nori (1992). "Bayesian encompassing tests of a unit root hypothesis," paper presented to Yale-NSF Symposium on "Bayes Methods and Unit Roots," April 1992.

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

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.

Phillips, P. C. B. and W. Ploberger (1991). "Time series modeling with a Bayesian frame of reference: I. Concepts and illustrations," Cowles Foundation Discussion Paper No. 980.

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.

Schotman, P. and H. K. van Dijk (1991). "On Bayesian routes to unit roots," *Journal of Applied Econometrics*, 6, 387-402.

Figure 1(a): RGNP:1909−1988 Log−Levels

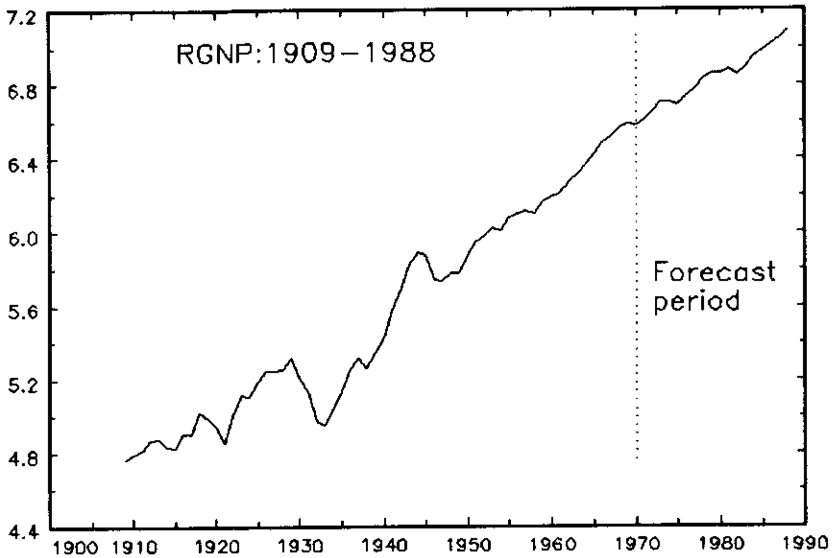

Figure 1(b): Prediction errors

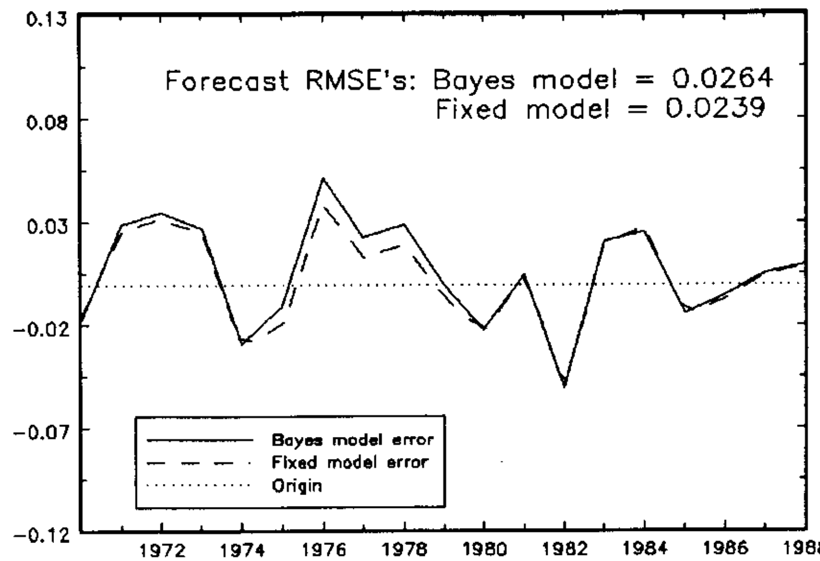

Figure 1(c): Evolving Best Bayes Model
(i) AR(p) + Trend(r) parameters
(ii) Unit Root present or not

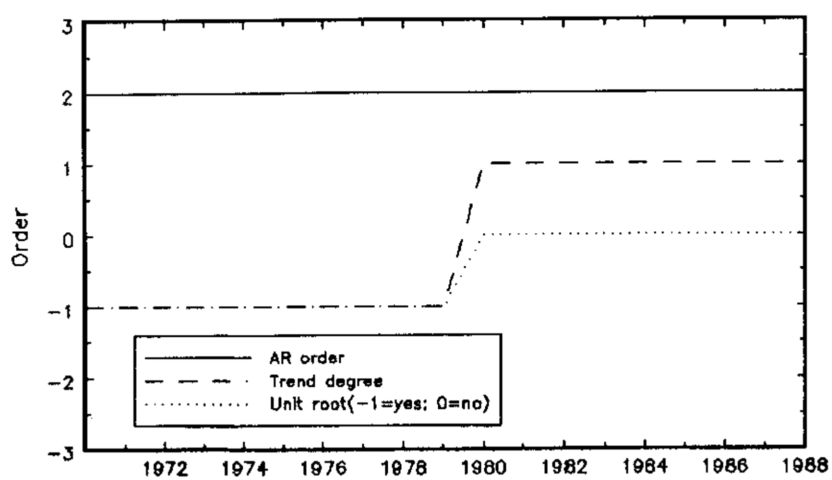

Figure 1(d): Bayes Model Forecast
Encompassing Test Statistic: dQ^B/dQ^F

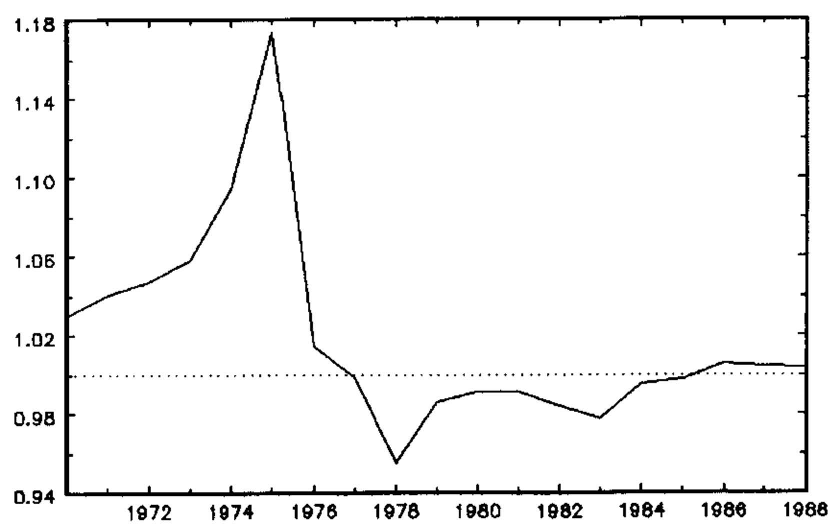

Figure 2(a): NGNP:1909–1988 Log–Levels

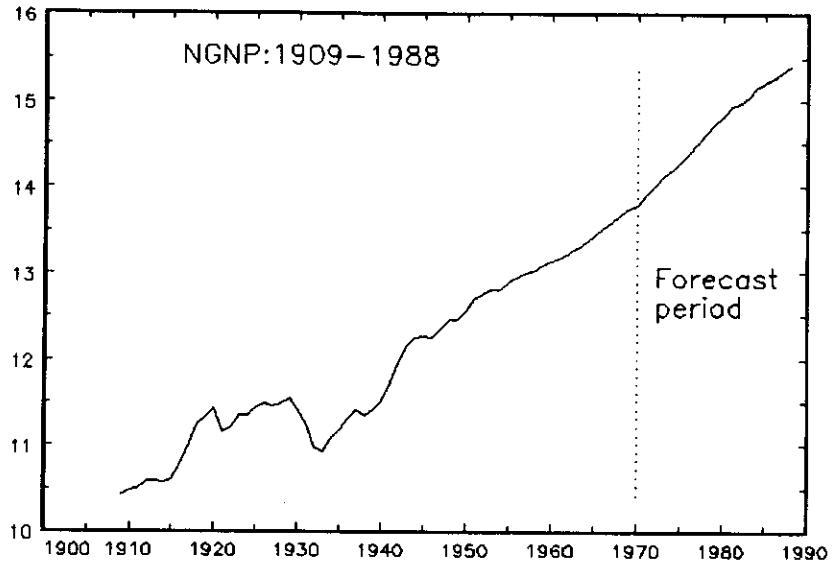

Figure 2(b): Prediction errors

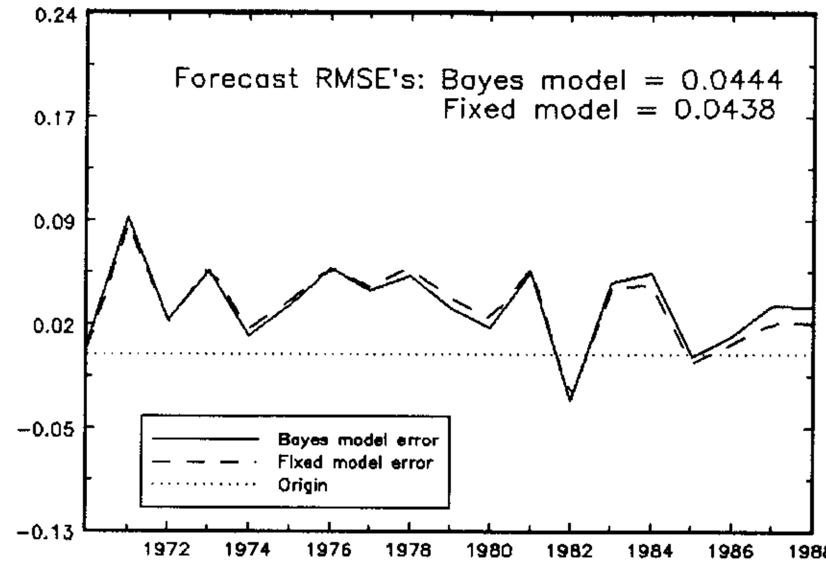

Figure 2(c): Evolving Best Bayes Model
(i) AR(p) + Trend(r) parameters
(ii) Unit Root present or not

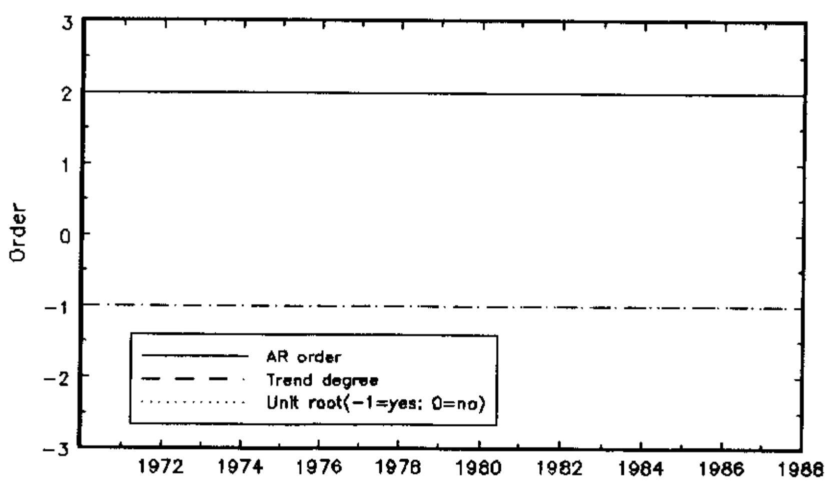

Figure 2(d): Bayes Model Forecast
Encompassing Test Statistic: dQ^B/dQ^F

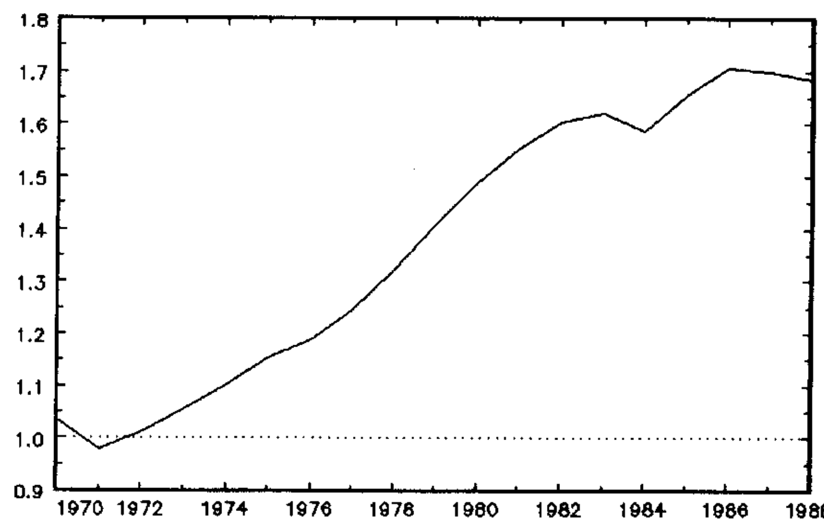

Figure 3(a): RPCGNP:1909–1988 Log–Levels

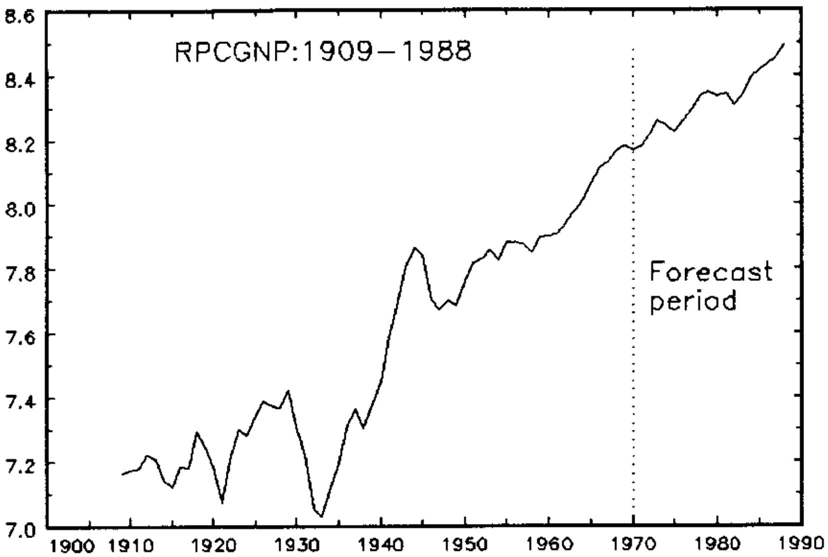

Figure 3(b): Prediction errors

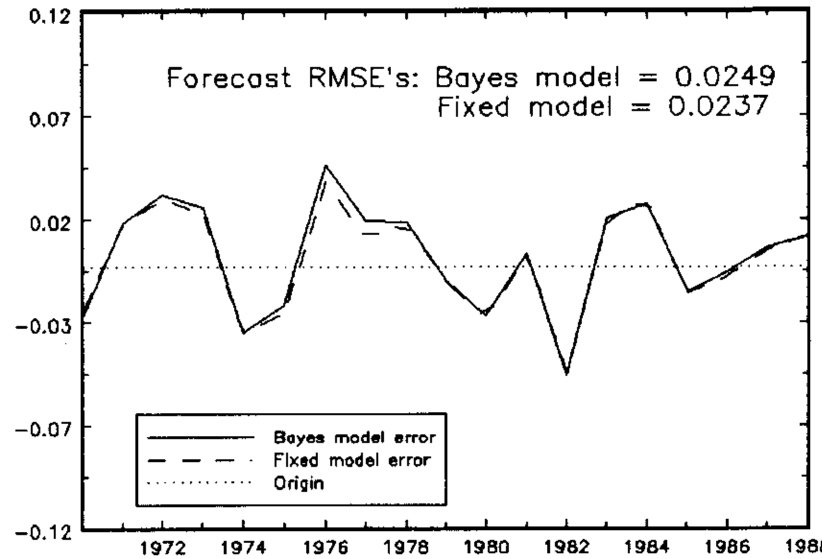

Figure 3(c): Evolving Best Bayes Model
(i) AR(p) + Trend(r) parameters
(ii) Unit Root present or not

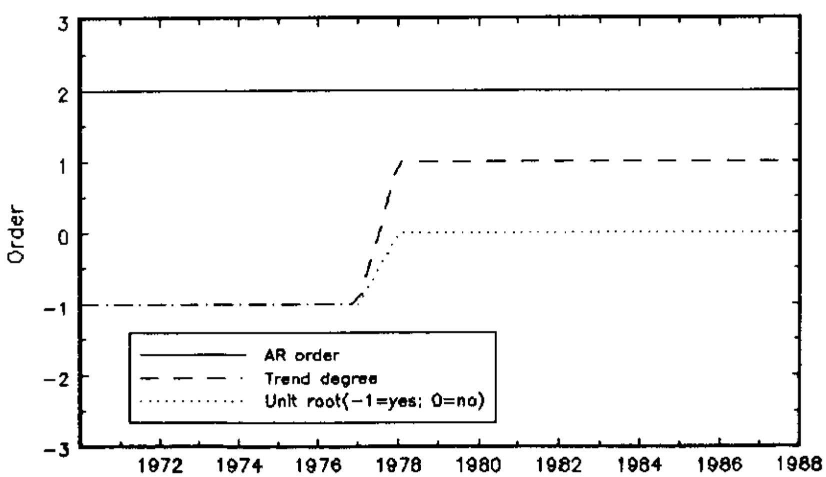

Figure 3(d): Bayes Model Forecast
Encompassing Test Statistic: dQ^B/dQ^F

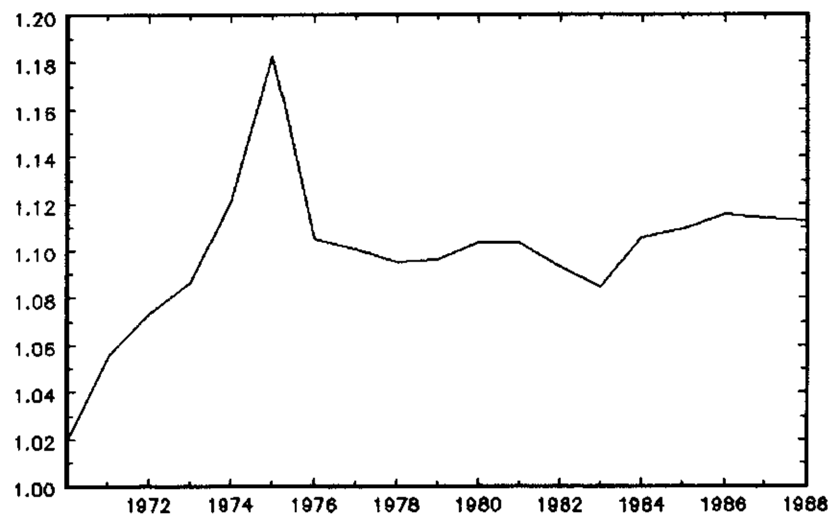

Figure 4(a): IP:1860–1988 Log–Levels

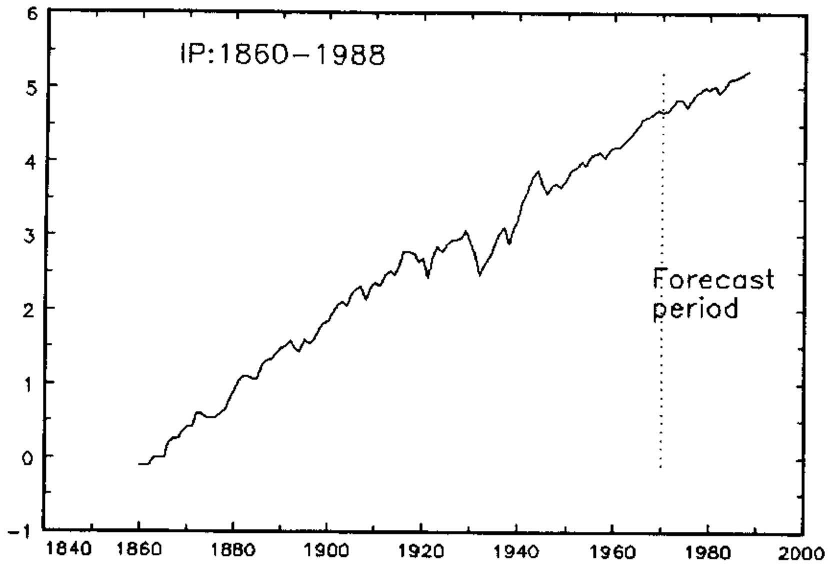

Figure 4(b): Prediction errors

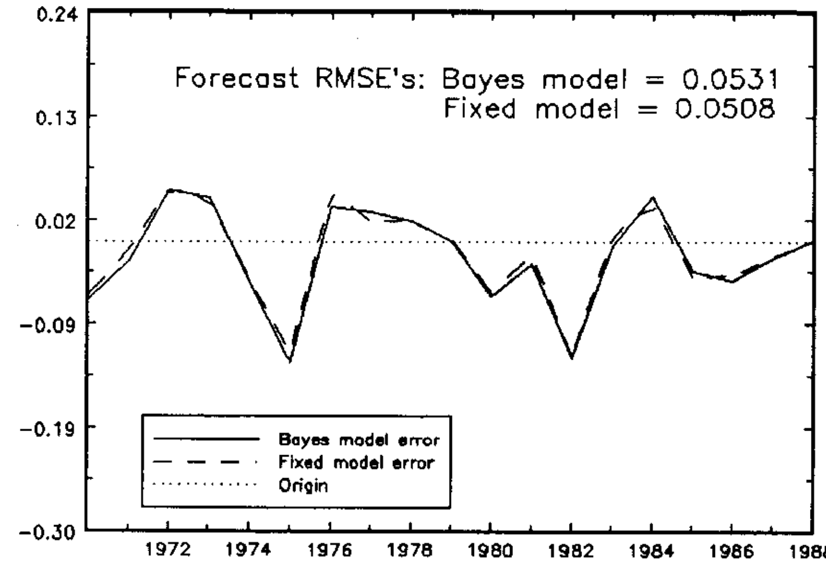

Figure 4(c): Evolving Best Bayes Model
(i) AR(p) + Trend(r) parameters
(ii) Unit Root present or not

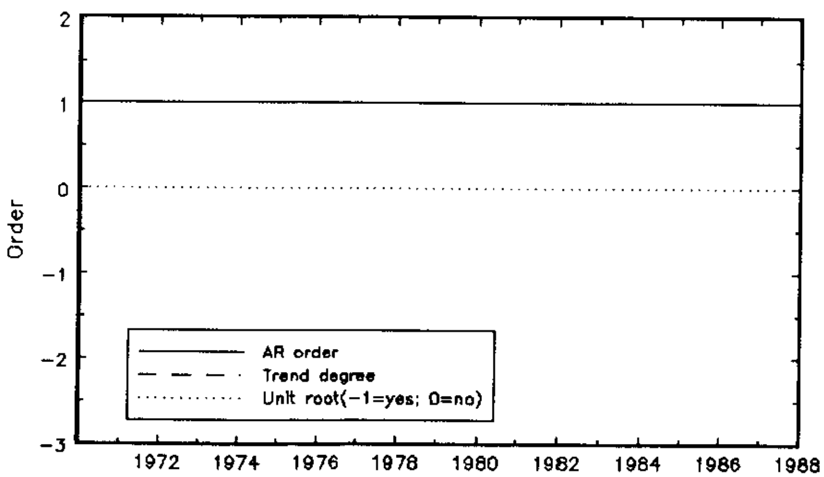

Figure 4(d): Bayes Model Forecast Encompassing Test Statistic: dQ^B/dQ^F

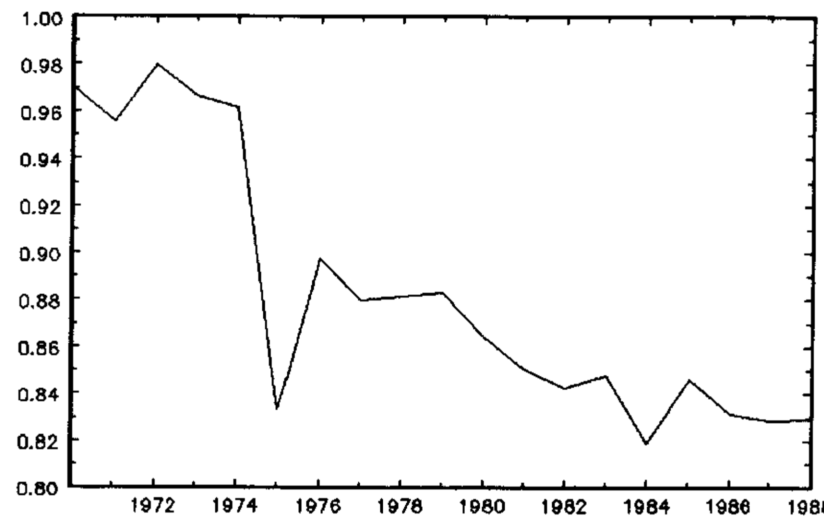

Figure 5(a): E:1890−1988 Log−Levels

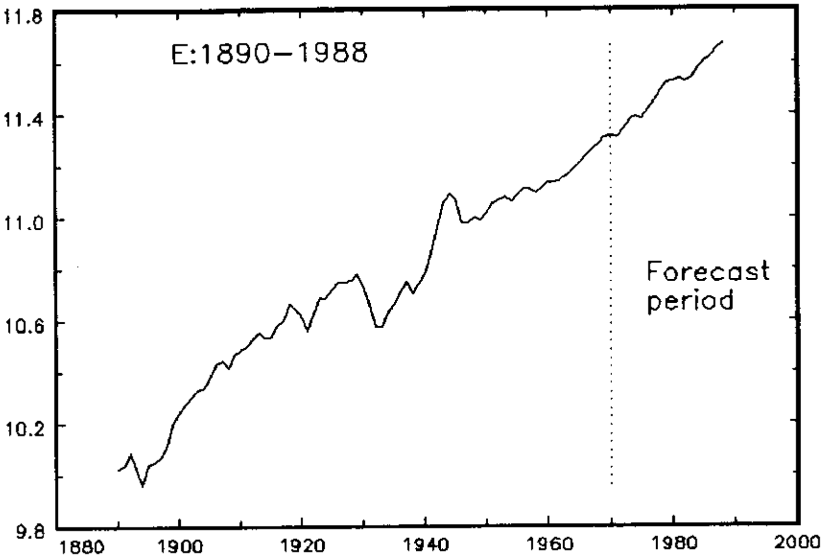

Figure 5(b): Prediction errors

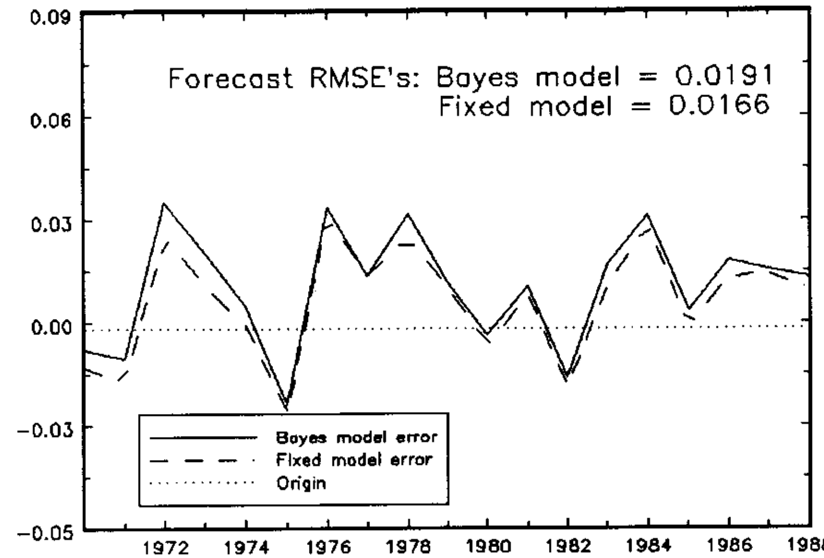

Figure 5(c): Evolving Best Bayes Model
(i) AR(p) + Trend(r) parameters
(ii) Unit Root present or not

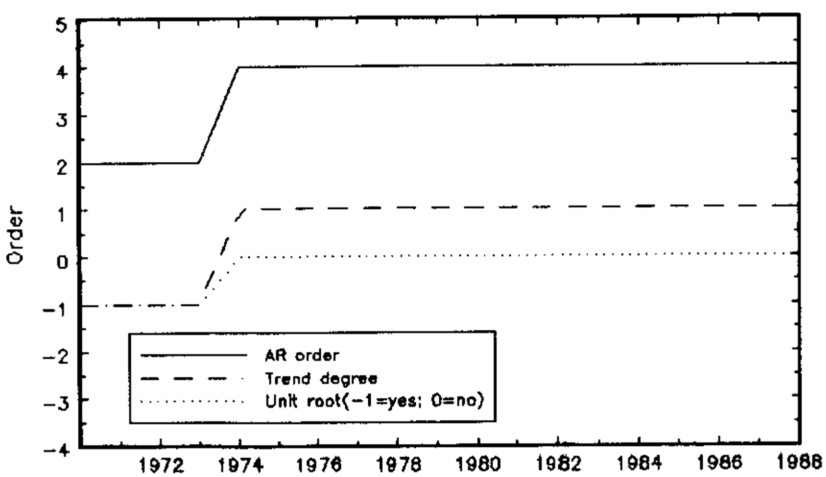

Figure 5(d): Bayes Model Forecast
Encompassing Test Statistic: dQ^B/dQ^F

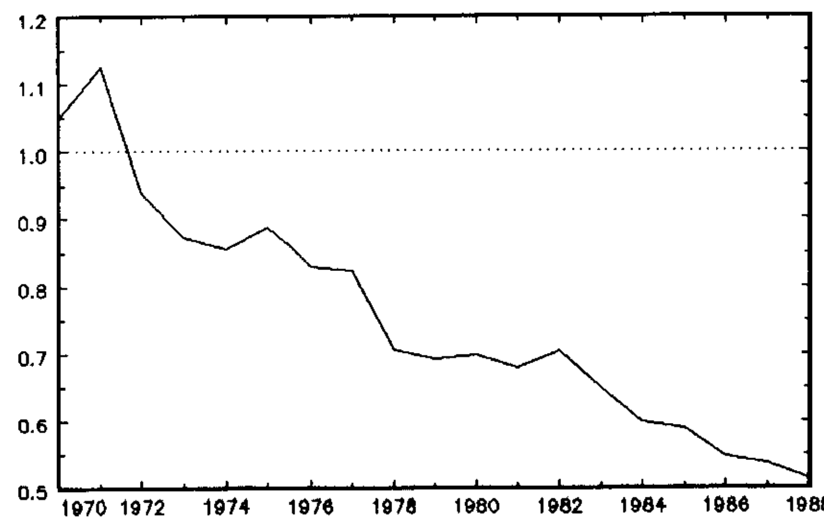

Figure 6(a): UN:1890–1988 Log–Levels

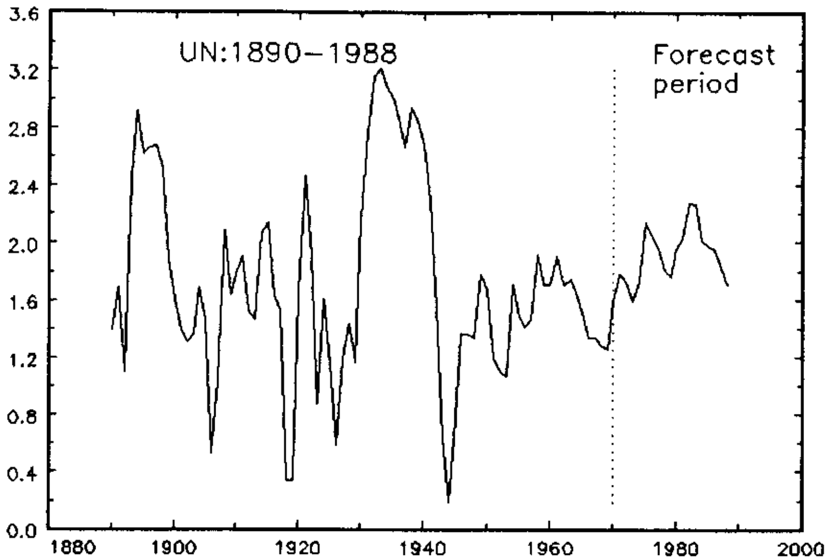

Figure 6(b): Prediction errors

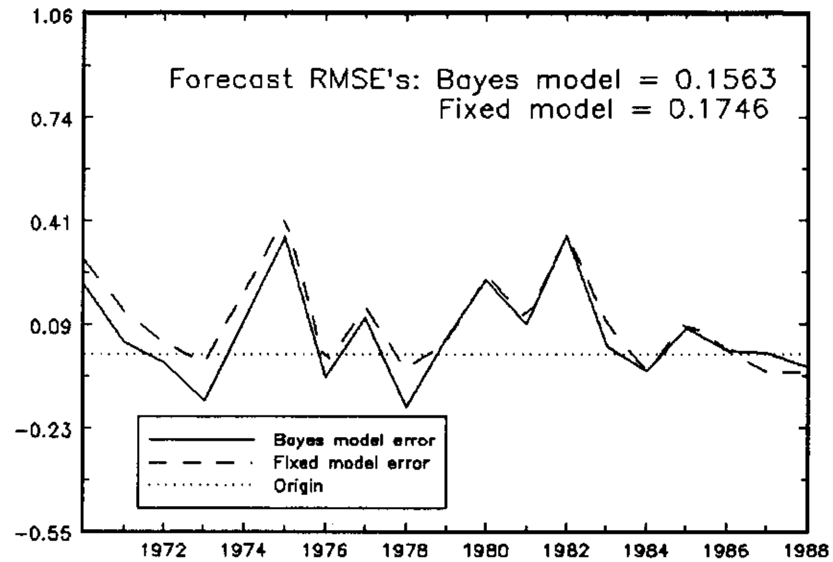

Figure 6(c): Evolving Best Bayes Model
(i) AR(p) + Trend(r) parameters
(ii) Unit Root present or not

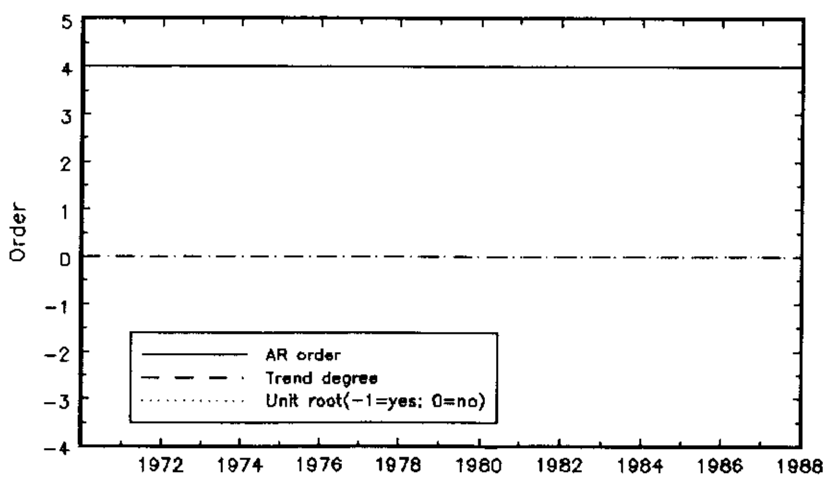

Figure 6(d): Bayes Model Forecast
Encompassing Test Statistic: dQ^B/dQ^F

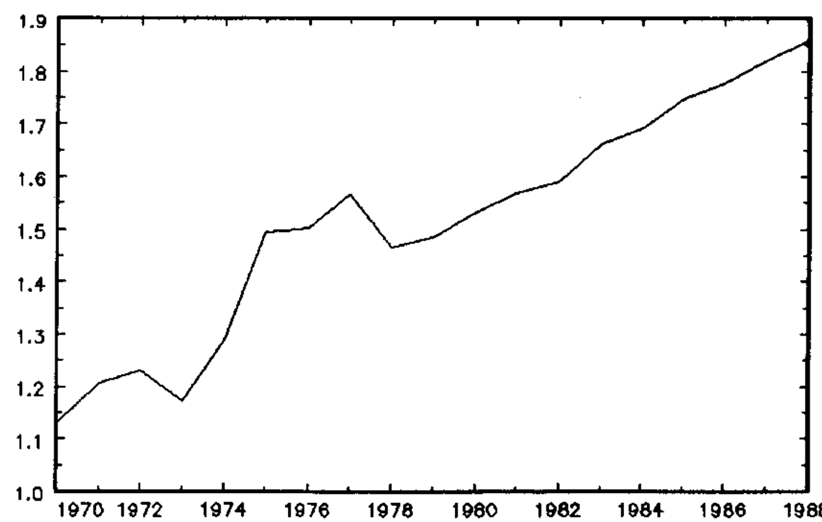

Figure 7(a): PRGNP:1889–1988 Log–Levels

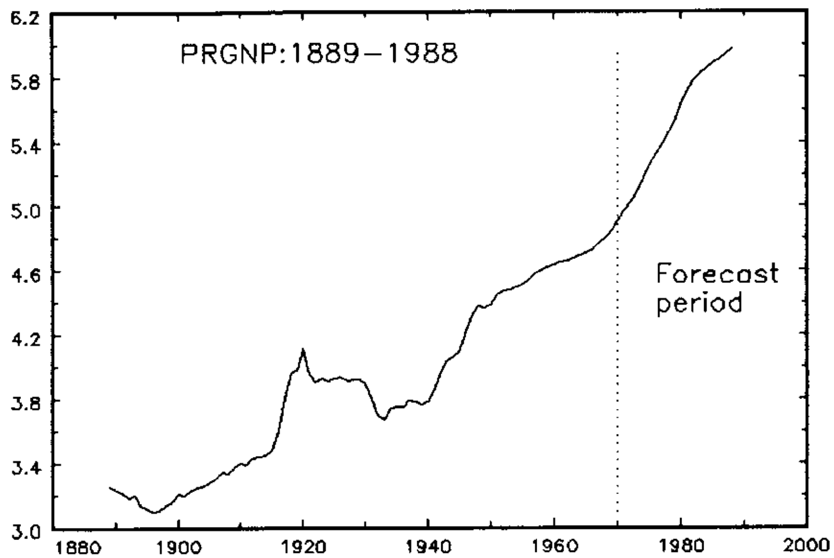

Figure 7(b): Prediction errors

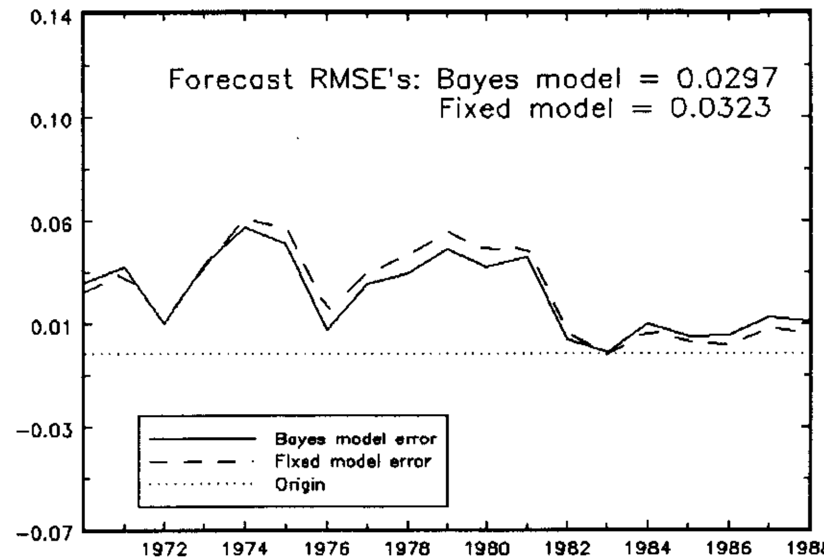

Figure 7(c): Evolving Best Bayes Model
(i) AR(p) + Trend(r) parameters
(ii) Unit Root present or not

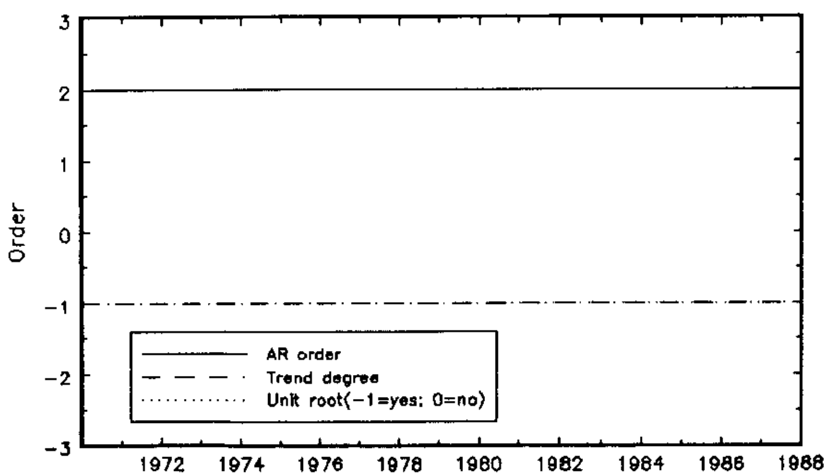

Figure 7(d): Bayes Model Forecast
Encompassing Test Statistic: dQ^B/dQ^F

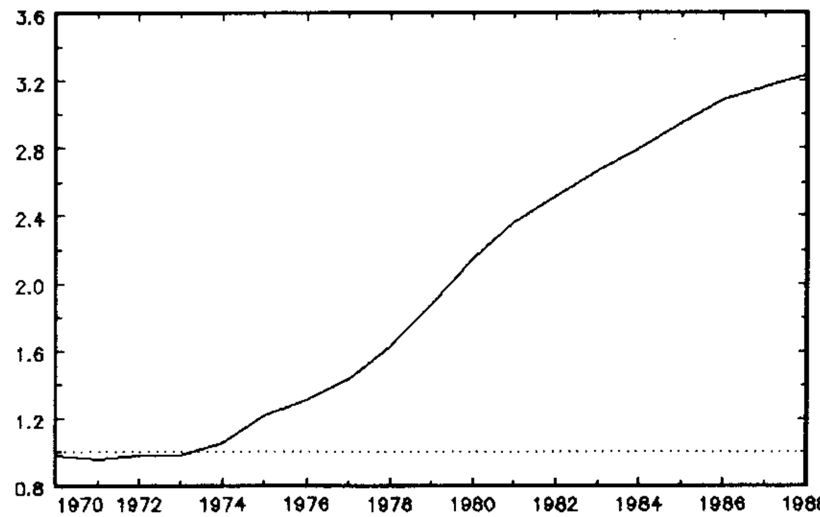

Figure 8(a): CPI:1860–1988 Log–Levels

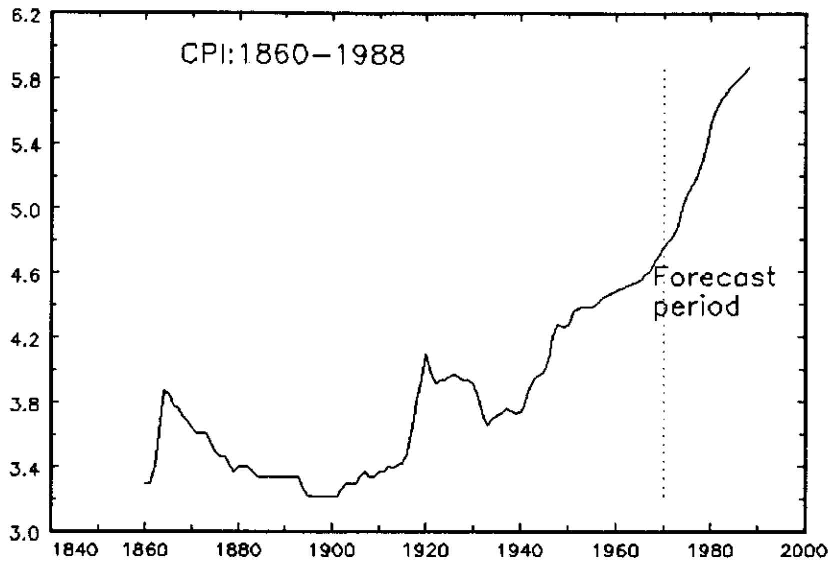

Figure 8(b): Prediction errors

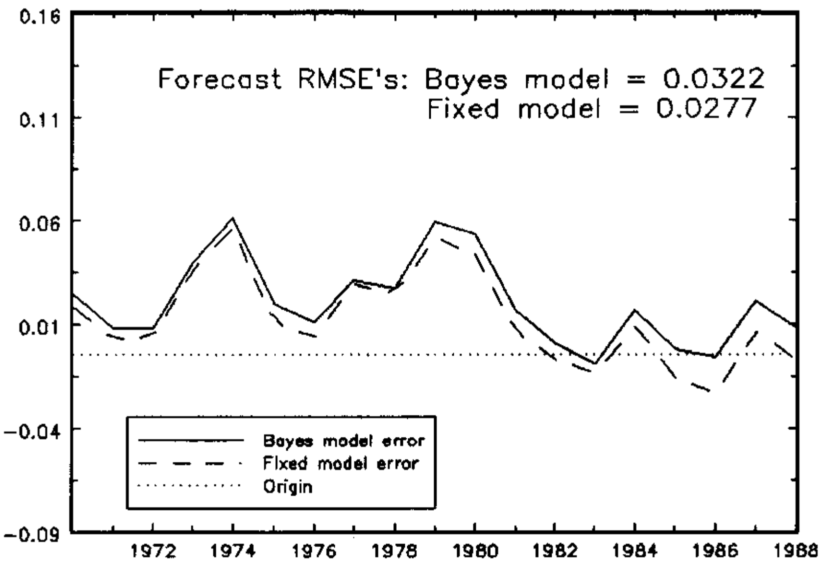

Figure 8(c): Evolving Best Bayes Model
(i) AR(p) + Trend(r) parameters
(ii) Unit Root present or not

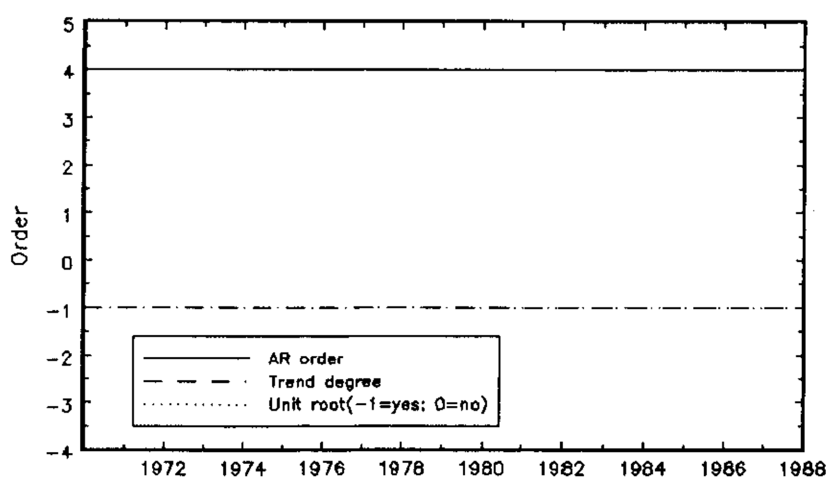

Figure 8(d): Bayes Model Forecast
Encompassing Test Statistic: dQ^B/dQ^F

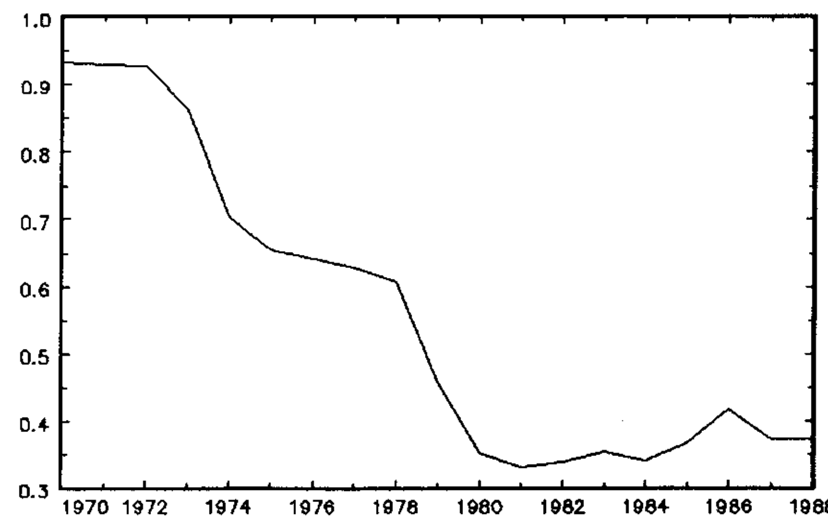

Figure 9(a): NW:1900–1988 Log–Levels

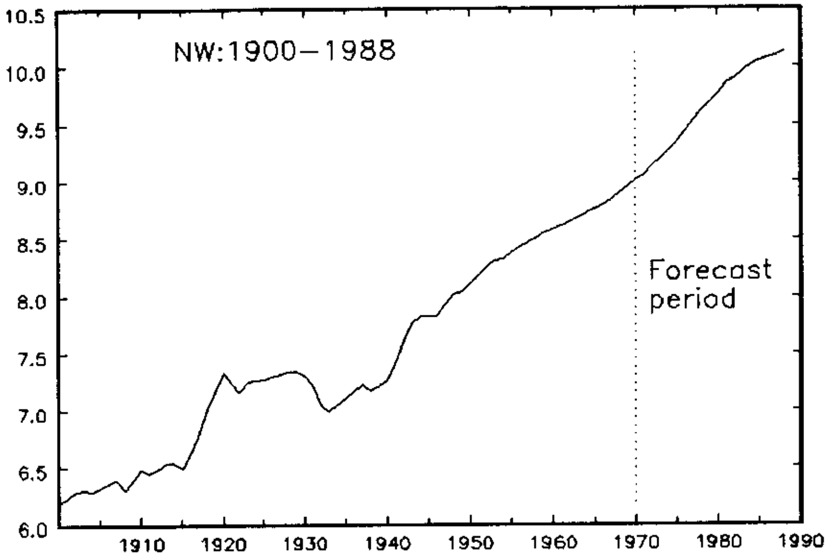

Figure 9(b): Prediction errors

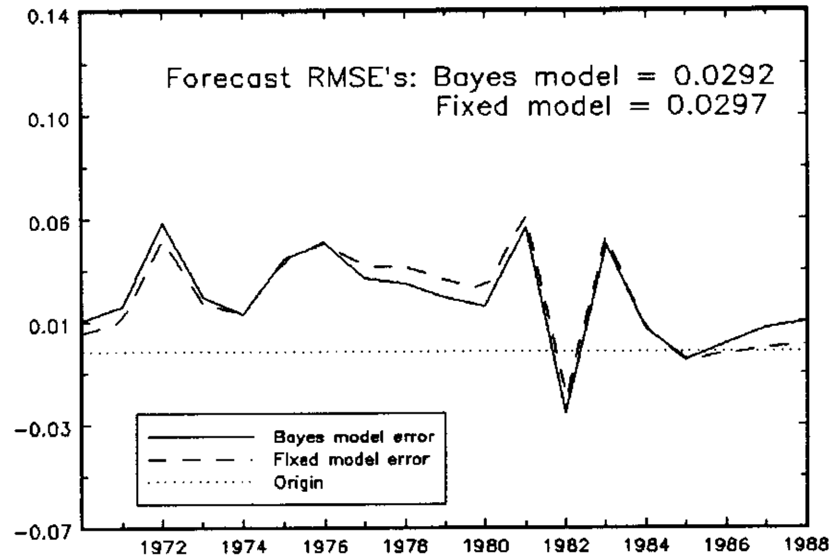

Figure 9(c): Evolving Best Bayes Model
(i) AR(p) + Trend(r) parameters
(ii) Unit Root present or not

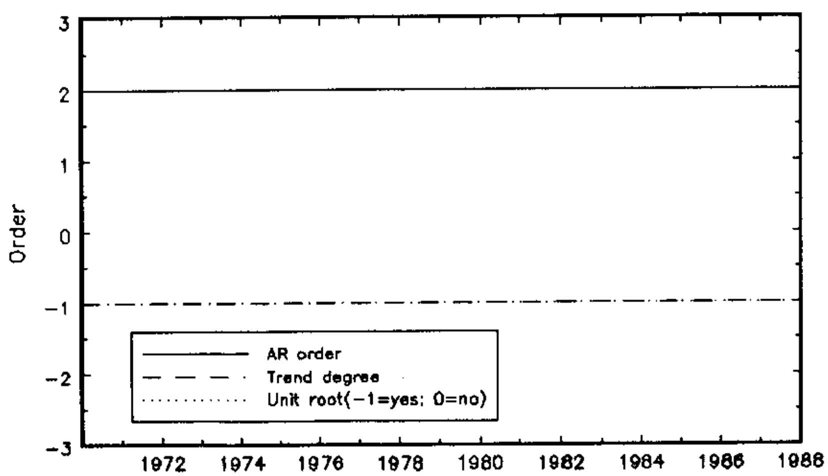

Figure 9(d): Bayes Model Forecast
Encompassing Test Statistic: dQ^B/dQ^F

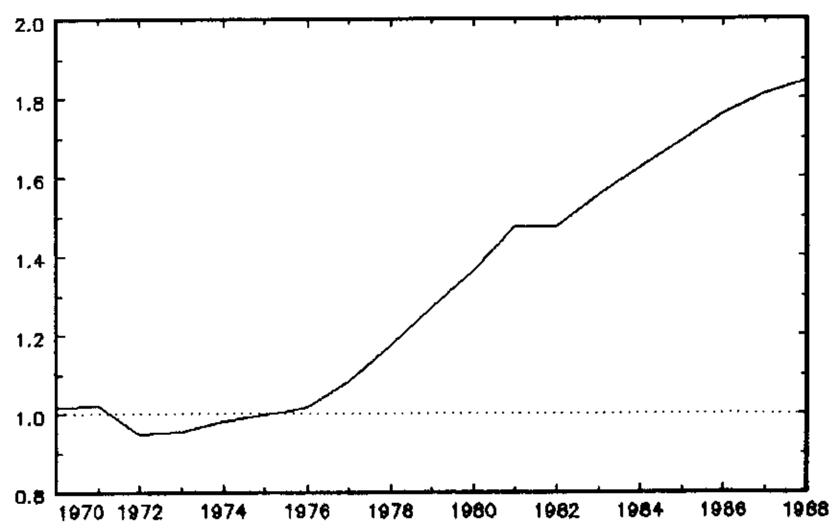

Figure 10(a): RW:1900–1988 Log–Levels

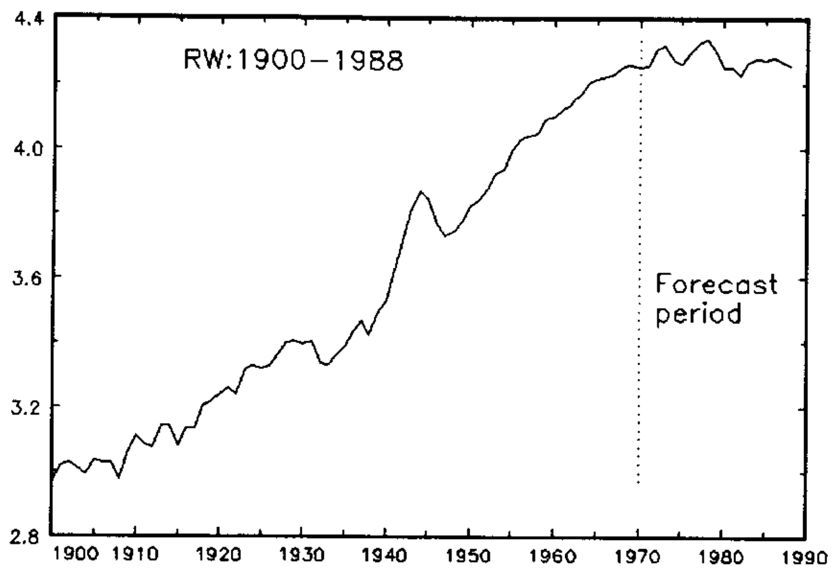

Figure 10(b): Prediction errors

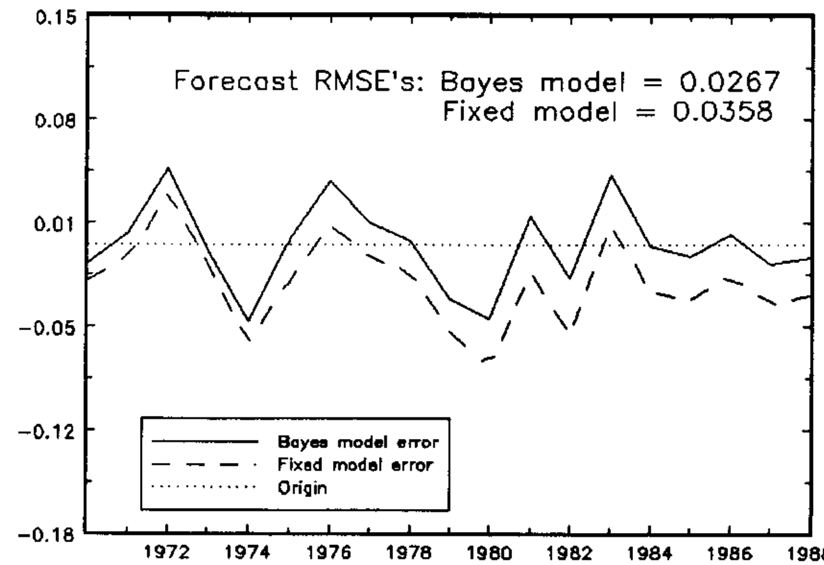

Figure 10(c): Evolving Best Bayes Model
(i) AR(p) + Trend(r) parameters
(ii) Unit Root present or not

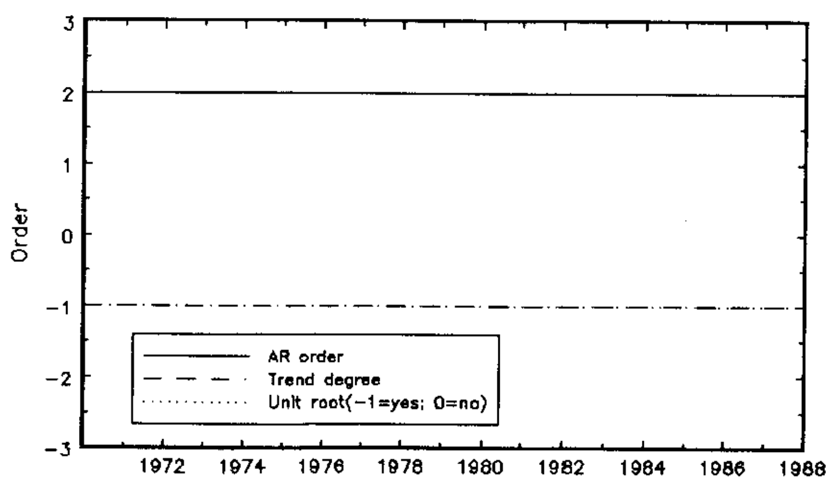

Figure 10(d): Bayes Model Forecast
Encompassing Test Statistic: dQ^B/dQ^F

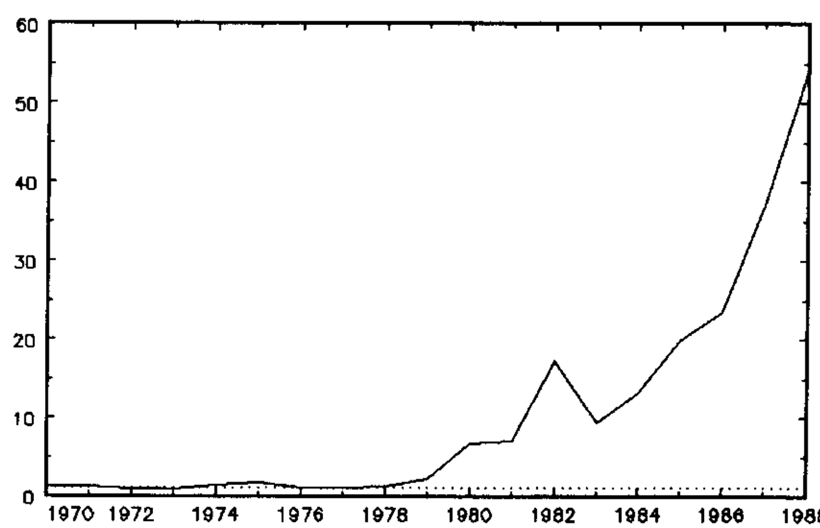

Figure 11(a): MS:1869–1988 Log–Levels

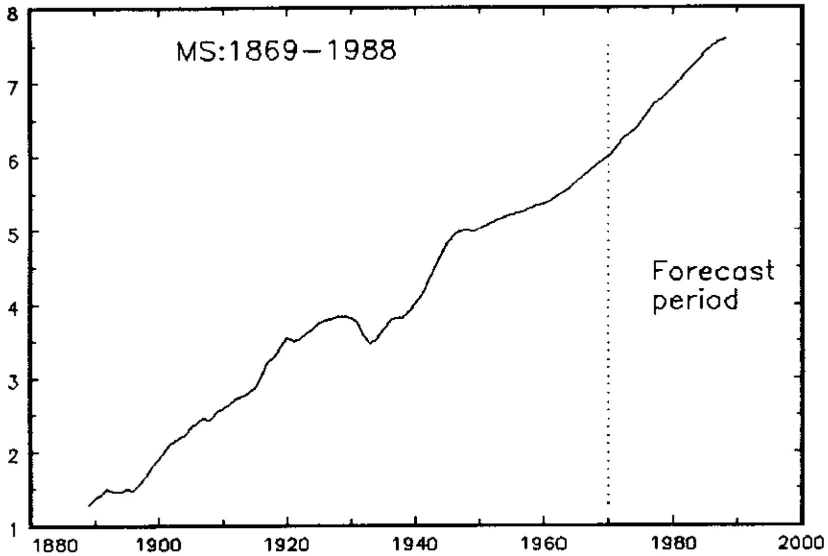

Figure 11(b): Prediction errors

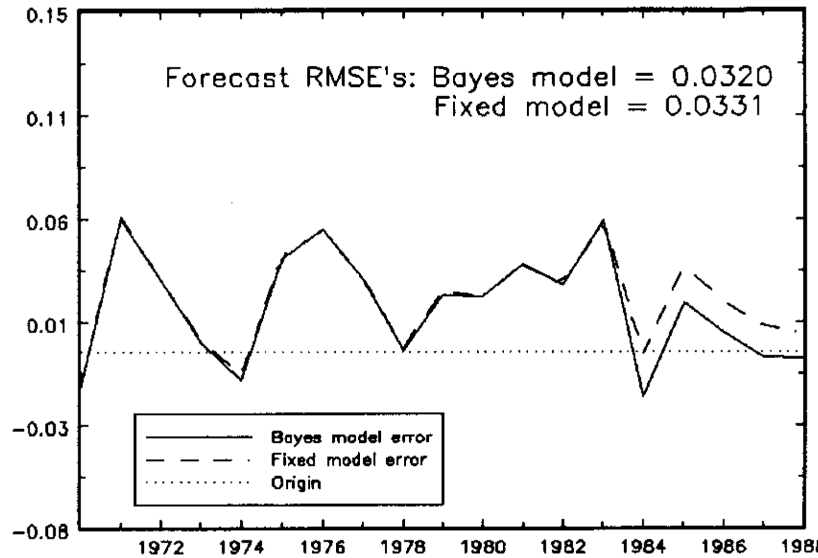

Figure 11(c): Evolving Best Bayes Model
(i) AR(p) + Trend(r) parameters
(ii) Unit Root present or not

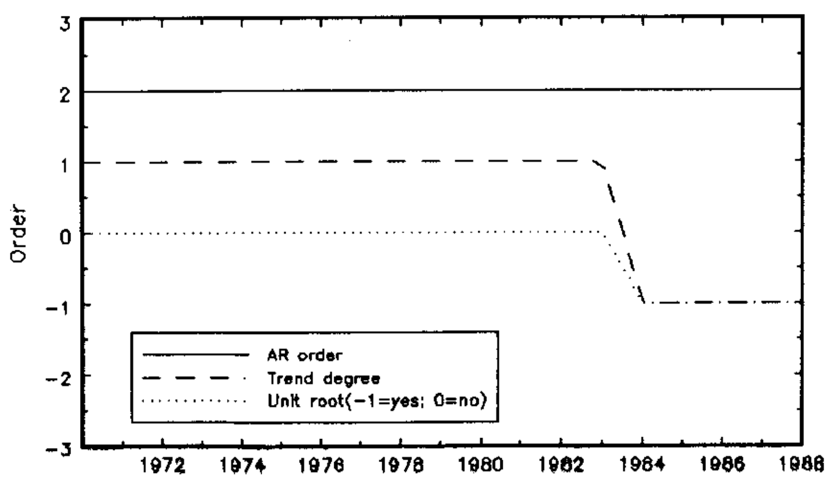

Figure 11(d): Bayes Model Forecast
Encompassing Test Statistic: dQ^B/dQ^F

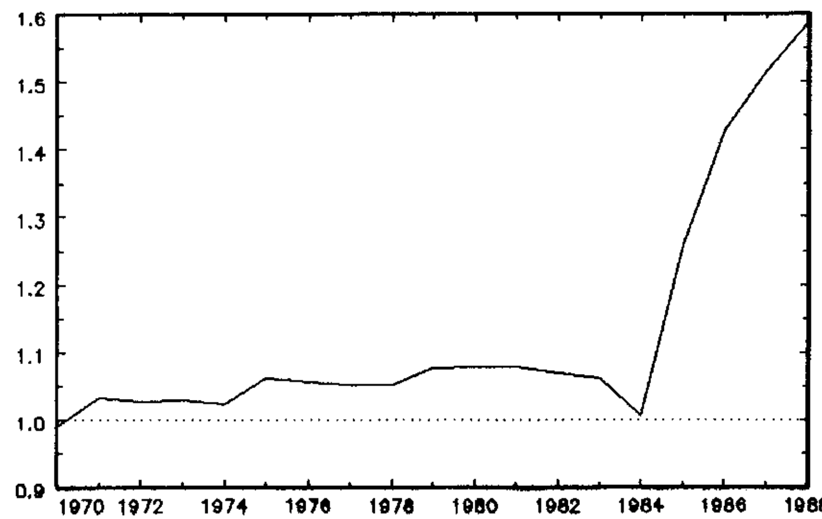

Figure 12(a): V:1869−1988 Log−Levels

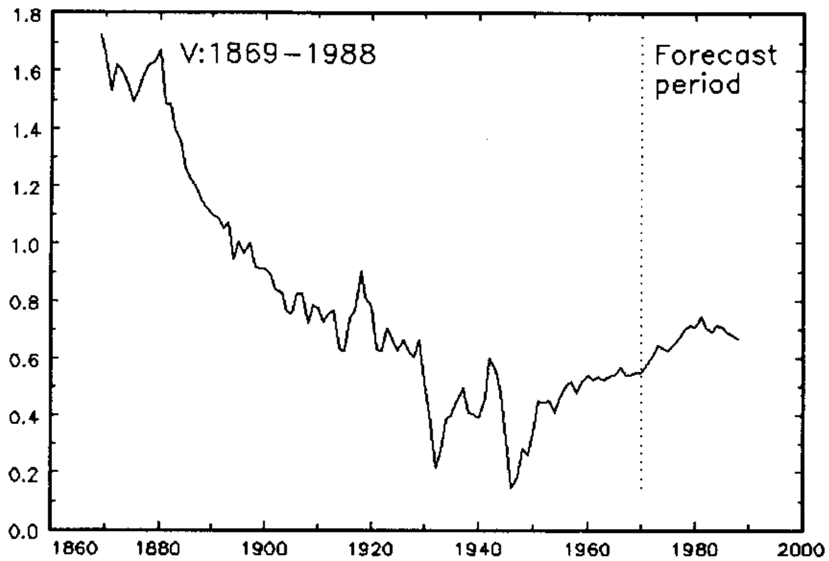

Figure 12(b): Prediction errors

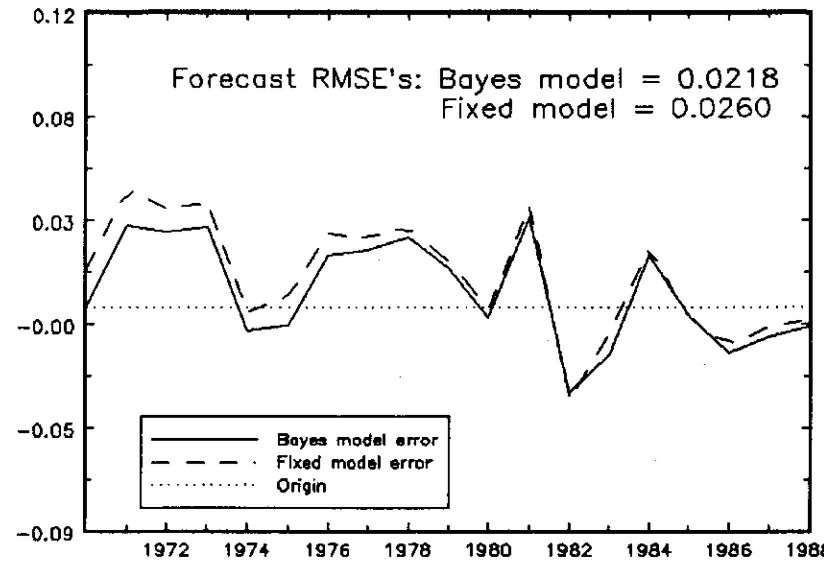

Figure 12(c): Evolving Best Bayes Model
(i) AR(p) + Trend(r) parameters
(ii) Unit Root present or not

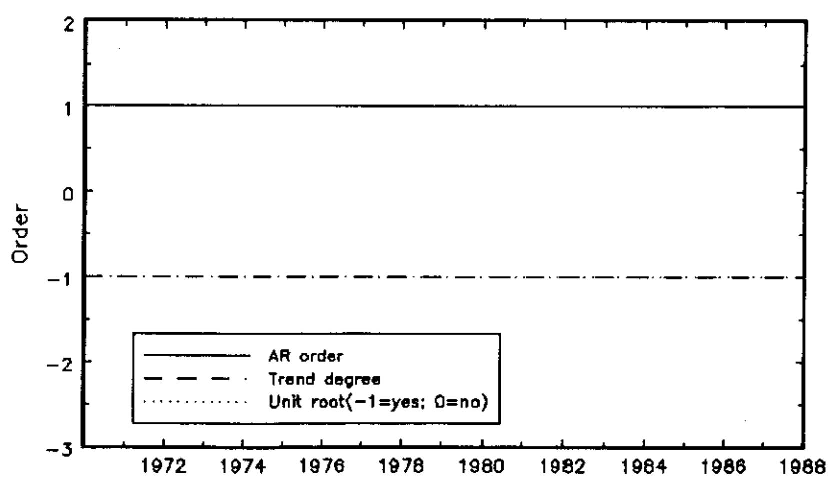

Figure 12(d): Bayes Model Forecast
Encompassing Test Statistic: dQ^B/dQ^F

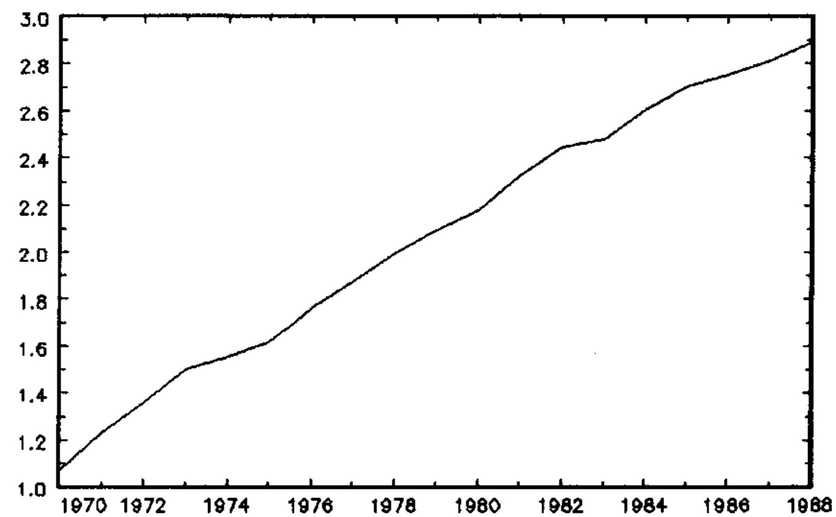

Figure 13(a): BY:1900–1988 $(\text{Levels})^{-1}$

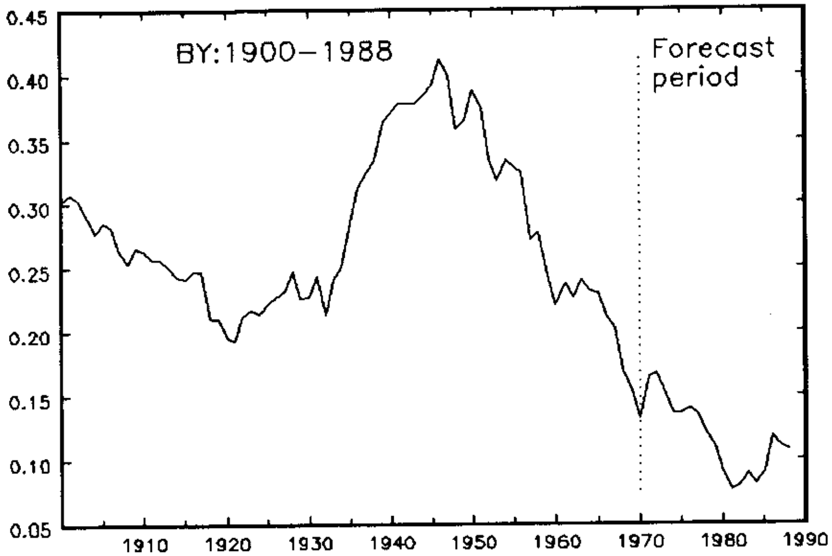

Figure 13(b): Prediction errors

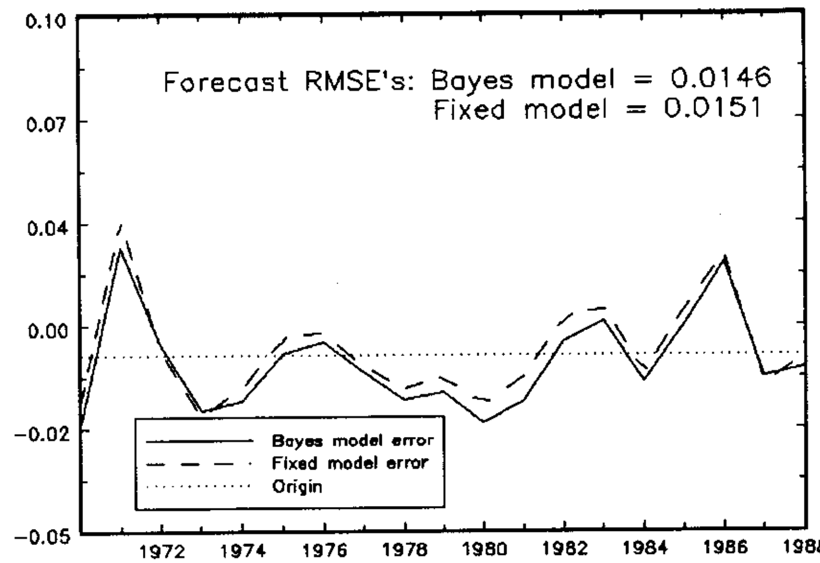

Figure 13(c): Evolving Best Bayes Model
(i) AR(p) + Trend(r) parameters
(ii) Unit Root present or not

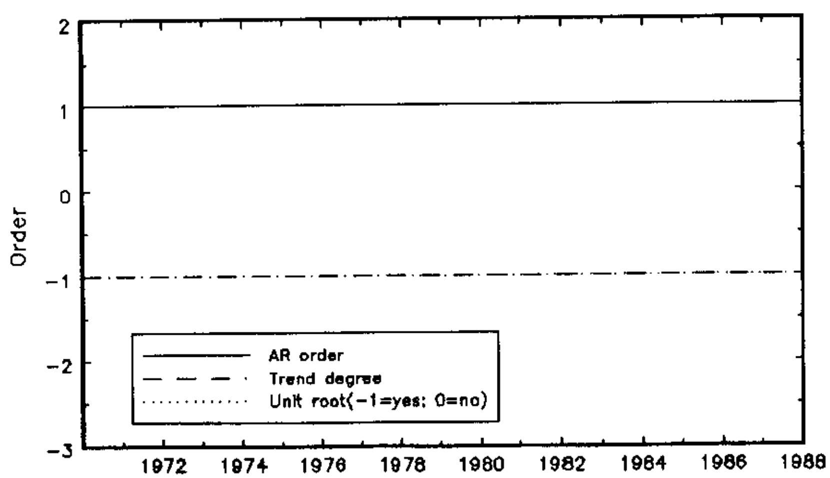

Figure 13(d): Bayes Model Forecast
Encompassing Test Statistic: dQ^B/dQ^F

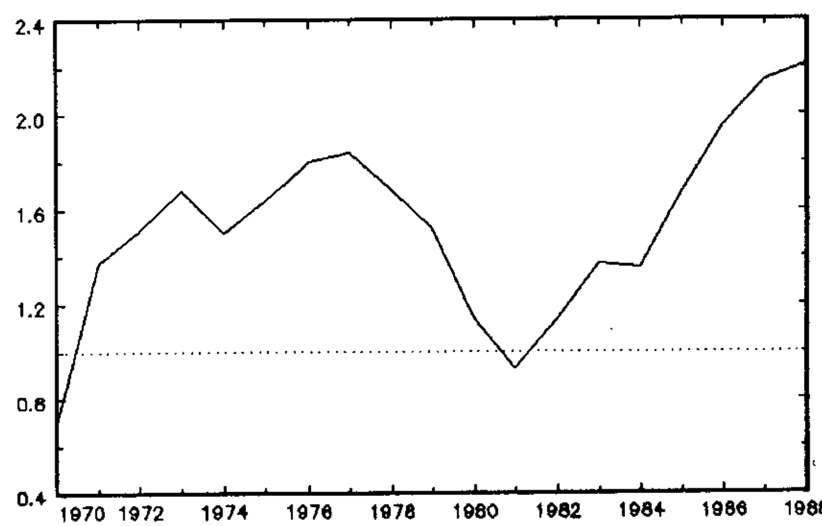

Figure 13'(a): BY:1900–1988 Levels

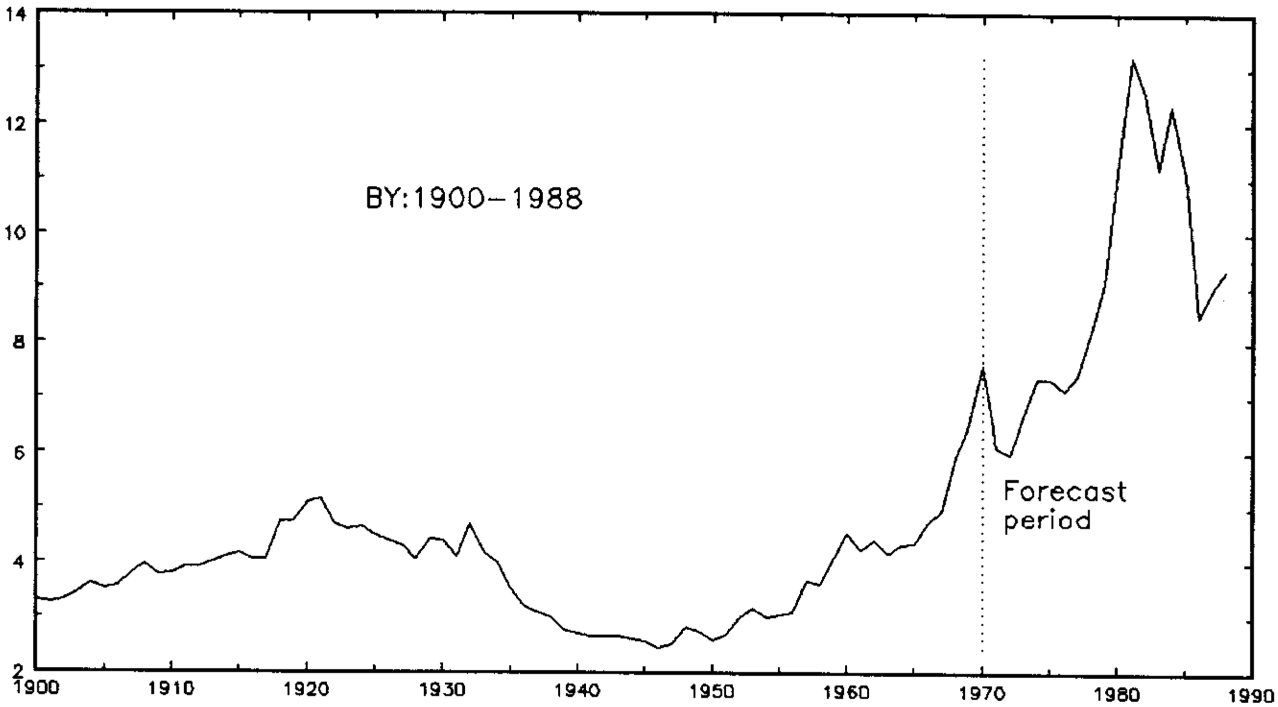

Figure 13'(b): Prediction errors

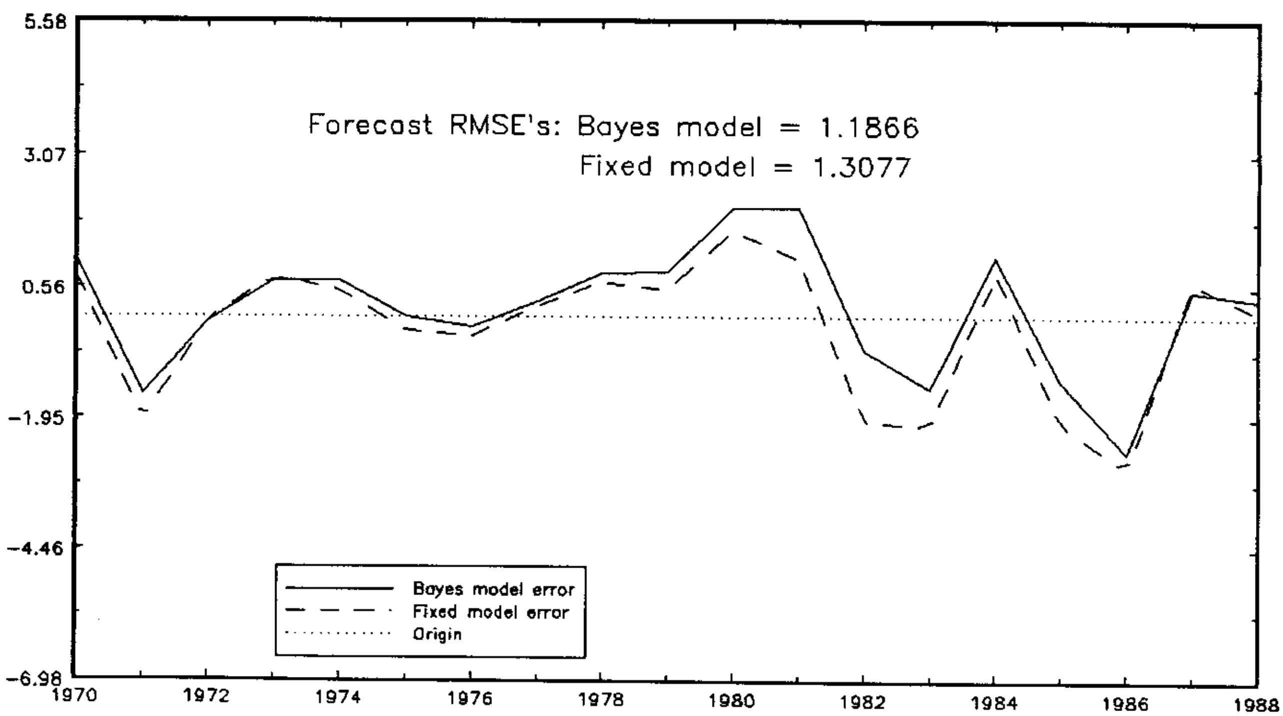

Figure 14(a): SP500:1871–1988 Log–Levels

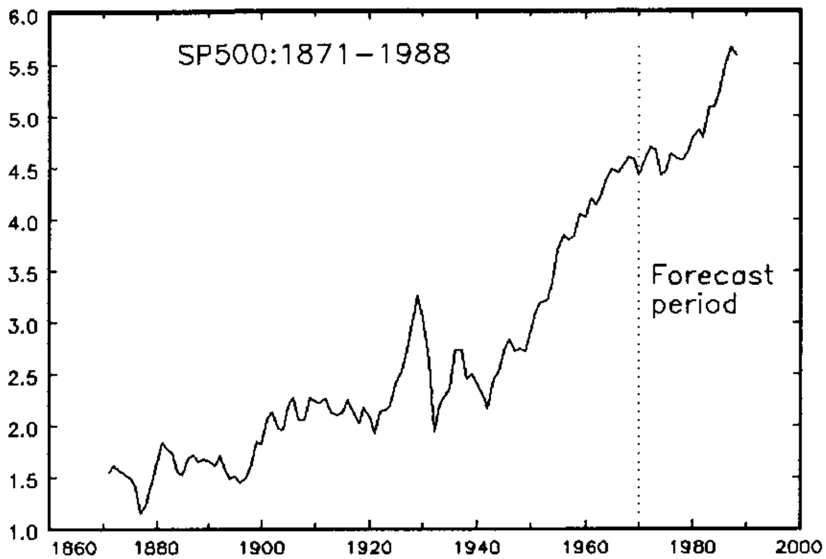

Figure 14(b): Prediction errors

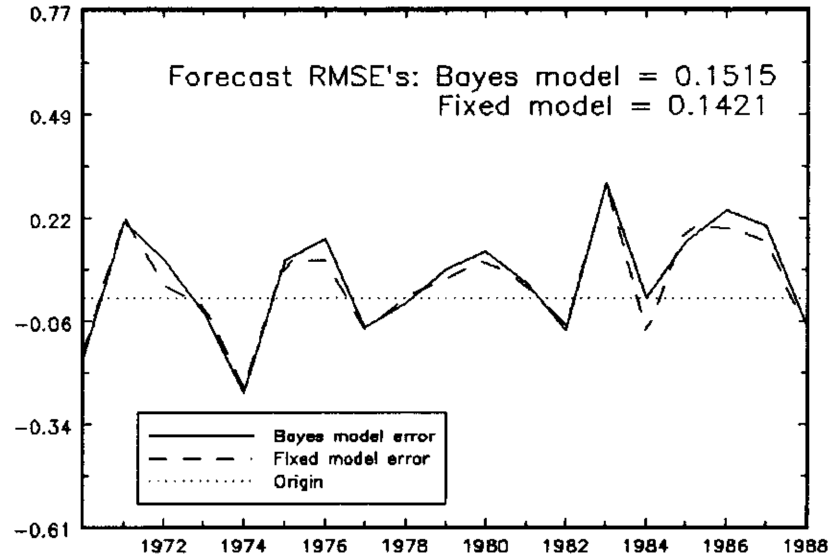

Figure 14(c): Evolving Best Bayes Model
(i) AR(p) + Trend(r) parameters
(ii) Unit Root present or not

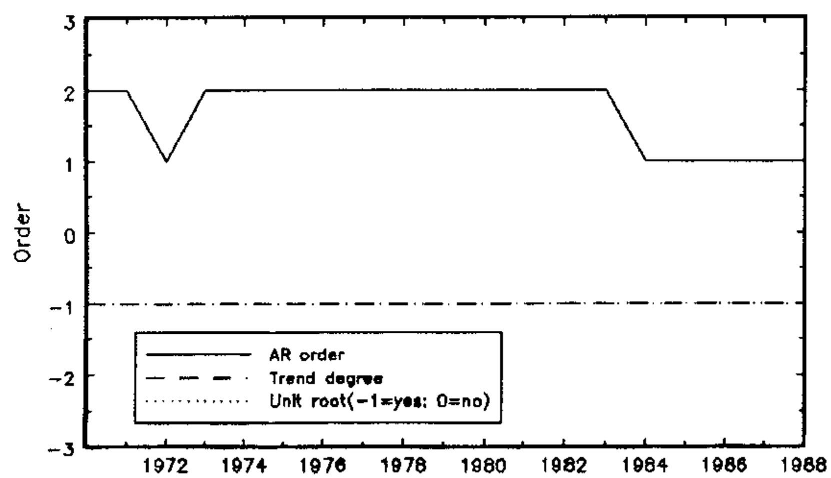

Figure 14(d): Bayes Model Forecast
Encompassing Test Statistic: dQ^B/dQ^F