Cowles Foundation Paper 987
Reprinted from Colin P. Hargreaves (ed.), Nonstationary Time Series Analysis and Cointegration, 1994

3

Bayes models and forecasts of Australian macroeconomic time series

Peter C.B. Phillips[*]

This paper provides an empirical implementation of some recent work by the author and Werner Ploberger on the development of *Bayes models* for time series. The methods offer a new data-based approach to model selection, to hypothesis testing and to forecast evaluation in the analysis of time series. A particular advantage of the approach is that modelling issues such as lag order, parameter constancy, and the presence of deterministic and stochastic trends all come within the compass of the same statistical methodology, as do the evaluation of forecasts from competing models. The paper shows how to build parsimonious empirical *Bayes models* using the new approach and applies the methodology to some Australian macroeconomic data. *Bayes models* are constructed for thirteen quarterly Australian macroeconomic time series over the period 1959(3)–1987(4). These models are compared with certain fixed format models (like an AR(4) + linear trend) in terms of their forecasting performance over the period 1988(1)–1991(4). The *Bayes models* are found to be superior ~~to~~ in these forecasting exercises ~~for two~~ of the thirteen series, while at the same time being more parsimonious in form. for 10

1. INTRODUCTION

Not all econometric models are designed as instruments for forecasting Nevertheless, the capacity of one model to forecast adequately in comparison with competing models is an important element in the evaluation of its overall

[*] 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 Sam Ouliaris for supplying the data, the NSF for research support under Grant No. SES 9122142, and Glena Ames for her skill and effort in keyboarding the manuscript.

performance. Indeed, many of the procedures that are presently used to appraise a model's performance involve summary statistics that depend in one way or another on the model's within-sample and outside-sample tracking behaviour. Thus, in spite of the multiplicity of objectives in econometric modelling, one common characteristic is the attempt each model makes to explain the data, or certain subsets of the data conditionally on other data. This attempted explanation often leads directly, but sometimes indirectly, to a model's 'probability distribution of the data'. Again, this may be a conditional distribution, and the statistical procedures that are employed may mean that only certain characteristics of the distribution rather than the full distribution are modelled. However, this common element of econometric modelling provides a basis by which different models can be compared. Thus, one model's explanation of the data can be compared with that of another model in terms of their implied 'probability distributions of the data'. In a similar way, one model's predictions can be compared with those of a competing model in terms of the respective 'probability distributions of the prediction errors'.

These ideas underlie some recent work by the author (1992) and by the author and Werner Ploberger (1991, 1992) on the development of *Bayes models* for time series. *Bayes models* are essentially location models conditional on the data that is available to the latest observation. In these models, the location estimate or systematic part of the model is nonlinear and time varying even when the underlying 'true model' is linear in parameters and variables like an autoregression. The location estimate is a predictor given by the current best estimate, using prior information and the available data, of the value of the dependent variable in the next period. The predictor is calculated as the conditional mean of the dependent variable given data to the latest available observation. Here the conditional expectation is taken with respect to the probability measure of the data implied by the given model and the prior distribution of the parameters. We call this measure the *Bayes model* measure. As more data accumulates, this *Bayes model* measure of the data becomes independent of the prior and is therefore 'objective' in the well defined sense that it ultimately depends only on the form of the model and the observed data. Since the Bayes measure is distinct for different models, it may be used as the basis for comparing competing models in terms of their implied 'probability distributions or predictive distributions of the data', as discussed in the last paragraph.

This paper implements the above ideas in an empirical study of Australian macroeconomic data. Our purpose is to seek out the best *Bayes models* in a certain generic class of time series models for each data set and then evaluate the adequacy of these models against certain fixed format competitor models in terms of one-period ahead forecasts. The methodology is based on earlier work in Phillips and Ploberger (1991, 1992) and Phillips (1992), which will be briefly reviewed in section 2 of the paper.

2. MODEL AND FORECAST EVALUATION USING THE PIC CRITERION

Our set up is the single equation stochastic linear regression model

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

where the dependent variable y_t and the error ε_t are real valued stochastic processes on a probability space (Ω, F, P). Accompanying y_t is the filtration $F_t \subset F (t = 0, 1, 2, \ldots)$ to which both y_t and ε_t are adapted. Usually it is convenient to think of F_t as the σ-field generated by $\{\varepsilon_t, \varepsilon_{t-1}, \ldots\}$ and in the cases we consider this will always be appropriate. 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 F_{t-1}-measurable. The errors ε_t satisfy $E(\varepsilon_t | F_{t-1}) = 0$, so that the conditional mean function in (1) is correctly specified under the probability measure P.

An example of (1) that is frequently empirically relevant is the 'ARMA (p,q) + trend(r)' model. This model can be written in difference format as

$$\Delta y_t = h y_{t-1} + \sum_{i=1}^{p-1} \varphi_i \Delta y_{t-i} + \sum_{j=1}^{q} \psi_j \varepsilon_{t-j} + \sum_{k=0}^{r} \delta_k t^k + \varepsilon_t \tag{2}$$

which is especially convenient because it accommodates an autoregressive unit root under the simple restriction $h = 0$. We call the parameter $a = 1+h$ the 'long run autoregressive coefficient' since this parameter is instrumental in determining the shape of the spectrum of y_t at the origin — see Phillips (1991) for elaboration on this point.

In (2) there are $k = p + q + r + 1$ parameters. When $q > 0$, some of the regressors, *viz.* the ε_{t-j}, are not observed. Recursive techniques are then required, either to construct the likelihood as in the use of the Kalman filter, or in repeated linear regressions that involve the construction of estimates of the lagged errors ε_{t-j} as in the Hannan and Rissanen (1982, 1983) recursion. When $q = 0$ in (2) the model is an 'AR(p) + trend (r)'. When $r = -1$ there is no intercept in the model, when $r = 0$ there is a fitted intercept, and when $r = 1$ there is a fitted linear trend. These are the specialisations of (2) that are of primary interest in empirical applications.

The order parameters p, q and r in (2) are not known in practical applications and the model is in any event best regarded as just an approximate generating mechanism. Various methodologies for dealing with this complication are available. Those that concern us here are based on formal statistical order selection methods such as the commonly used criteria AIC and BIC. These criteria and their statistical properties in stationary systems are discussed in detail in the recent book by Hannan and Deistler (1988). When the system is potentially nonstationary as in (2) with $h = 0$ the properties of these criteria are less well understood although they have been studied, notably by Paulsen (1984), Tsay (1984) and Pötscher (1989).

Our approach to the order selection problem is based on the analysis in Phillips and Ploberger (1992) and is closely related to the principle underlying the BIC criterion of Schwarz (1978), viz. to select the model with the highest *a posteriori* probability. This approach has a compelling advantage over AIC and BIC in that it naturally accommodates models of nonstationary time series and has generally superior sampling performance (see Phillips and Ploberger (1992) for simulation evidence on this point). The probability measure used to determine our criterion is the measure associated with the *Bayes model* corresponding to (1). This model is formally derived for the case of Gaussian errors $\varepsilon_t = iid\,\mathrm{N}(0, \sigma^2)$ in (1) and has the form

$$y_t = \hat{\beta}'_{t-1}x_t + v_t, \quad where \quad v_t\big|_{F_{t-1}} \equiv \mathrm{N}(0, f_t) \tag{3}$$

with

$$f_t = \sigma^2\left\{1 + x'_t A_{t-1}^{-1} x_t\right\}, \quad A_t = \sum_1^t x_s x'_s \tag{4}$$

and where $\hat{\beta}_{t-1}$ is the least squares estimate of β based on information in F_{t-1}.

The Phillips and Ploberger analysis shows that, under a uniform prior on β and a Gaussian likelihood the passage via Bayes rule to the posterior density of β implies the replacement of the model (1) by the time varying parameter model (3). We therefore call (3) the *Bayes model* corresponding to (1). Note that the systematic part of (3), $\hat{\beta}'_{t-1}x_t$, is the best estimate or predictor of the location of y_t given information in 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. Thus, the *Bayes model* is identical to the classical model that is actually used to make predictions (in place of (1)). From this perspective there is no difference between the Bayesian and classical approaches. However, we can go further in our approach and find the probability measure associated with the *Bayes model* (3). This is a forward looking measure that can be described by its conditional density given F_{t-1}. This density is given by the Radon Nikodym (RN) derivative of the measure at t (say Q_t) with respect to the measure at $t-1$ (Q_{t-1}), i.e.

$$\begin{aligned} dQ_t/dQ_{t-1} &= pdf_Q(y_t|F_{t-1}) \\ &= (2\pi f_t)^{-1/2}\exp\{-(1/2f_t)v_t^2\} \equiv N(0,f_t), \qquad t=k+1,\ k+2,\ \dots\ . \end{aligned} \tag{5}$$

We use the notation $\mathrm{pdf}_Q(\cdot)$ here to signify that this is the density corresponding to Q-measure. Note that it is defined as soon as there are enough observations in a trajectory to estimate the k-vector β. Thus, (3), (4) and (5) are defined for $t \geq k+1$. The measure Q_t that appears in (5) is called the *Bayes model* measure, i.e. the measure corresponding to the *Bayes model* (3). This measure is σ-finite and, as shown in Phillips and Ploberger (1992), can also be defined in terms of the following RN derivative

$$dQ_t/dP_t = \left|(1/\sigma^2)A_t\right|^{-1/2}\exp\left\{(1/2\sigma^2)\hat{\beta}_t' A_t\hat{\beta}_t\right\}, \tag{6}$$

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).

Associated with every *Bayes model* of the form (3) is a σ-finite measure Q_t. Different models of the same data may be compared in terms of the *Bayes model* measures that are associated with them. The natural mechanism for making such comparisons is the likelihood ratio. Suppose, for example, that we have two models of the form given in (3), one with k parameters and the other with $K \geq k$ parameters. Indexing the variables in (3) by the number of parameters we now have two *Bayes models* of the data: one with k parameters that we write as

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

and the second, more complex model with K parameters

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

The likelihood ratio of the measures associated with $\mathrm{H}(Q_n^k)$ and $\mathrm{H}(Q_n^K)$ is given by the RN derivative dQ_n^k/dQ_n^K. This quantity can be calculated by taking the ratio of the RN derivatives that define Q_n^k and Q_n^K in terms of the reference measure P_n, i.e. the ratio of the corresponding expressions given by (6) for each model. Thus,

$$dQ_n^k/dQ_n^K = \left(dQ_n^k/dP_n\right)/\left(dQ_n^K/dP_n\right)$$
$$= \left|\tfrac{1}{\sigma^2}A_n(k)\right|^{-1/2}\left|\tfrac{1}{\sigma^2}A_n(K)\right|^{-1/2}\exp\left\{\left(\tfrac{1}{2\sigma^2}\right)\left[\hat{\beta}_n(k)'A_n(k)\hat{\beta}_n(k)-\hat{\beta}_n(K)'A_n(K)\hat{\beta}_n(K)\right]\right\}. \quad (7)$$

This likelihood ratio measures the support in the data for the more restrictive model $H(Q_n^k)$ against that of the more complex model $H(Q_n^K)$. When we assign equal prior odds to the two competing models our decision criterion is to accept $H(Q_n^k)$ in favour of $H(Q_n^K)$ when $dQ_n^k/dQ_n^K > 1$.

Since σ^2 in (7) is usually unknown we must supply an estimate of this scale parameter before the criterion can be used in practice. Phillips and Ploberger (1992) suggests the use of $\hat{\sigma}_K^2$, the least squares estimate of σ^2 from the more complex model $H(Q_n^K)$. Our order estimator is then given by

$$\hat{k} = \operatorname{argmin}_k \mathrm{PIC}_k \quad (8)$$

where

$$\mathrm{PIC}_k = \left(dQ_n^K/dQ_n^k\right)\left(\hat{\sigma}_K^2\right). \quad (9)$$

Observe that $\hat{k}$ maximises $1/\mathrm{PIC}_k = dQ_n^k/dQ_n^K(\hat{\sigma}_K^2)$ and thereby selects the model most favoured over $H(Q_n^K)$ according to the density of the data.

An alternative form of the PIC criterion (9) that is given in Phillips and Ploberger (1992) is based on the predictive densities of the competing Bayes models, i.e. $H(Q_n^k)$ and $H(Q_n^K)$. By comparing the densities for these models over the same subsample of data, say $n > K$, we have

$$\mathrm{PICF}_k = \frac{dQ_n^k}{dQ_n^K}\left(\hat{\sigma}_K^2\right)\Big|F_K = \prod_{K+1}^{n}\left(\hat{f}_t^K/\hat{f}_t^k\right)^{1/2}\exp\left\{\sum_{K+1}^{n}\left[v_t(K)^2/2\hat{f}_t^K - v_t(k)^2/2\hat{f}_t^k\right]\right\} \quad (10)$$

where

$$\hat{f}_t^k = \hat{\sigma}_K^2\left(1+x_t(k)'A_{t-1}(k)^{-1}x_t(k)\right),\quad \hat{f}_t^K = \hat{\sigma}_K^2\left(1+x_t(K)'A_{t-1}(K)^{-1}x_t(K)\right);$$
$$v_t(k) = y_t - \hat{\beta}_{t-1}(k)'x_t(k),\qquad v_t(K) = y_t - \hat{\beta}_{t-1}(K)'x_t(K).$$

Note that (10) differs from (9) only to the extent that in (10) the density ratio is conditional on data in F_K. In effect, the initialisation in (9) is at $t = 0$ on the field F_0, whereas in (10) the initialisation is at $t = K$ on the field F_K where there is enough sample data to estimate both models $H(Q_t^k)$ and $H(Q_t^K)$. Both criteria PIC and PICF are scale invariant because of the presence of the error variance estimate $\hat{\sigma}_K^2$ in their definitions. Note, however, that PICF is invariant to linear transformations of the regressors $x_t(k)$ and $x_t(K)$ in the two models $H(Q_t^k)$ and

$H(Q_t^K)$, whereas PIC is not. Use of PIC therefore presumes that there is some natural form of the regressor variables, as there is for example in AR and ARMA models with deterministic time trends. This is the class of model that will be used in the application of our methods that is reported below.

As indicated earlier, our approach is related to the principle underlying the BIC criterion, which leads to the order estimator

$$\operatorname{argmin}_k\left[\mathrm{BIC}_k = \ln\hat{\sigma}_k^2 + k\ln(n)/n\right].$$

When the data are stationary and ergodic it is easily shown that our criterion PIC is asymptotically equivalent to BIC (see Phillips and Ploberger, 1992). However, when the data are nonstationary the criterion PIC imposes a greater penalty than BIC on the presence of additional nonstationary regressors. The simulations in Phillips and Ploberger (1991) show that PIC generally outperforms BIC as an order estimator criterion for both stationary and nonstationary data, at least in Gaussian models.

The PIC and PICF criteria are also related to the MDL and PMDL criteria of Rissanen (1986, 1987*a*, 1987*b*), *viz.*

$$\operatorname{argmin}_k\left[\mathrm{MDL}_k = \ln\hat{\sigma}_k^2 + \ln|A_n(k)|/n\right],\ \operatorname{argmin}_k\left[\mathrm{PMDL}_k = \sum_{t=K+1}^{n}\left\{\ln\hat{\sigma}_{kt} + \hat{e}_{t+1}^2/\hat{\sigma}_{kt}^2\right\}\right],$$

(see e.g. Mills and Prasad, 1992, for these formulae). Clearly, the MDL criterion is closely related to PIC in that the penalty term $\ln|A_n(k)|/n$ involves the data rather than simply a parameter count as in the BIC criterion penalty $k\ \ln(n)/n$. Note, however, that the PIC penalty involves the term $|A_n(k)/\hat{\sigma}_K^2|$ and is therefore scale invariant. In the Rissanen predictive criterion PMDL, $\hat{e}_{t+1}$ and $\hat{\sigma}_{kt}^2$ are defined by

$$\hat{e}_{t+1}^2 = \sum_{j=0}^{t}\left(y_{j+1} - x_{j+1}(k)'\hat{\beta}_t\right)^2,$$

and

$$\hat{\sigma}_{kt}^2 = \sum_{j=1}^{t}\left(y_j - x_j(k)'\hat{\beta}_t\right)^2\Big/t.$$

Writing

$$\exp\{-(½)\mathrm{PMDL}_k\} = \prod_{t=k}^{n-1}(1/\hat{\sigma}_{kt})\exp\{-\hat{e}_{t+1}^2/2\hat{\sigma}_{kt}^2\},$$

we see that PMDL is related in form to our PICF. The criteria differ, however, because (i) PICF employs the recursive one step ahead squared forecast errors $v_t^2(k) = (y_t - x_t(k)'\beta_{t-1})^2$ rather than the sum of squared predictive errors $\hat{e}_{t+1}^2$ as in PMDL; and (ii) PICF employs the forecast error variance $\hat{f}_t^k$ rather than the error variance estimator $\hat{\sigma}_{kt}^2$ as in PMDL.

The PICF criterion (10) has a very interesting interpretation as a form of encompassing test statistic. For, if $dQ_n^k/dQ_n^K(\hat{\sigma}_K^2)|_{F_K} > 1$ 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. Thus, when $dQ_n^k/dQ_n^K(\hat{\sigma}_K^2)|_{F_K} > 1$, the *Bayes model* $\mathrm{H}(Q_n^k)$ encompasses the *Bayes model* $\mathrm{H}(Q_n^K)$ in terms of the probability distribution of the sample data over the period $t = K+1, \ldots, n$. This might be called distributional encompassing for $t \in [K+1, n]$.

Obvious extensions of this principle apply for subperiods of the overall sample. Moreover, the principle can be extended to evaluate the forecasts from competing models. For instance, $\mathrm{H}(Q_n^k)$ and $\mathrm{H}(Q_n^K)$ can be compared in terms of their respective performance in one-period ahead forecasts over the period $t = n+1, \ldots, N$. The *Bayes model* forecast encompassing test statistic for this period is

$$dQ_n^k/dQ_n^K\left(\hat{\sigma}(K)^2\right)\Big|_{F_n} = \prod_{t=n+1}^{N}\left(g_t^K/g_t^k\right)^{1/2}\exp\left\{-\left(1/2\hat{\sigma}_t^2(K)g_t^k\right)v_t(k)^2 + \left(1/2\hat{\sigma}_t^2(K)g_t^K\right)v_t(K)^2\right\}. \tag{11}$$

Note that in this formulation the variance estimate $\hat{\sigma}_t^2(K)$ evolves recursively over the forecast period. Again, $\mathrm{H}(Q_n^k)$ encompasses $\mathrm{H}(Q_n^K)$ in terms of forecast performance over the period $[n+1, N]$ when $dQ_n^k/dQ_n^K(\hat{\sigma}_K^2)|_{F_K} > 1$.

Bayes models like $\mathrm{H}(Q_n^k)$ may be permitted to evolve in a natural way as more observations become available. Thus, period by period we may employ the PIC criterion (8) to select the appropriate value, $\hat{k}_t$, of k for the sample data up to observation $t-1$, prior to making the one-period ahead forecast of the value of y_t. This leads to an evolving sequence of best *Bayes models*

$$\mathbf{H}\left(Q_t^B\right): y_t = \hat{\beta}_{t-1}\left(\hat{k}_{t-1}\right)' x_t\left(\hat{k}_{t-1}\right) + v_t\left(\hat{k}_{t-1}\right)$$

which are determined recursively using the PIC criterion (8) period after period. It is then possible to compare the best *Bayes model* sequence $\mathrm{H}(Q_t^B)$ with a fixed format *Bayes model* sequence $\mathrm{H}(Q_t^F)$ that employs a fixed number of parameters (F). The comparison can be made in terms of their respective predictive densities over a forecast horizon such as $t \in [n+1, N]$. In this case the forecast-encompassing test statistic is

$$dQ_N^B / dQ_N^F \left(\hat{\sigma}^2(\hat{k})\right)\Big|_{F_n}$$
$$= \prod_{t=n+1}^{N} \left(g_t^F / g_t^{\hat{k}_{t-1}}\right)^{1/2} \exp\left\{-\left(1/2\hat{\sigma}_t^2(\hat{k}_{t-1})g_t^{\hat{k}_{t-1}}\right)v_t(\hat{k}_{t-1})^2 + \left(1/2\hat{\sigma}_t^2(\hat{k}_{t-1})g_t^F\right)v_t(F)^2\right\}. \quad (12)$$

We would favour the best *Bayes model* sequence $\{H(Q_t^B)\}_{n+1}^N$ over the sequence of fixed format models $\{H(Q_t^F)\}_{n+1}^N$ if

$$dQ_n^B / dQ_N^F \left(\hat{\sigma}^2(\hat{k})\right)\Big|_{F_n} > 1, \quad (13)$$

that is, if the sequence $H(Q_t^B)$ generates forecasts over $t = n+1, \ldots, N$ that encompass the forecasts of the fixed format sequence of models $H(Q_t^F)$. Note that in (12) and (13) we use recursive estimates of the error variance from the best *Bayes model* sequence, since these are consistent for σ^2 when (1) is the actual generating mechanism.

The most important property of $H(Q_t^B)$ is that this sequence of models adapt to the data. When fewer parameters are needed to model the data the sequence will respond by eliminating unnecessary parameters. When more are needed, the sequence adapts by enlarging the model, either by adding more lags or by adding deterministic trend polynomial regressors, as appropriate. Since the PIC criterion can also be used to test for the presence of a unit autoregressive root the best *Bayes model* sequence $H(Q_t^B)$ can also be designed to include unit roots whenever these are supported by the data.

In general, we may expect $H(Q_t^B)$ to have fewer parameters than $H(Q_t^F)$, especially when F incorporates a linear trend and several lags. Reasonable choices for $H(Q_t^F)$ depend on the time interval of observation. Thus, for annual data a fixed format model of the type 'AR(3) + linear trend' may seem a sensible baseline competitor. For seasonally adjusted quarterly data an 'AR(4) + linear trend' may be reasonable and for monthly data one might choose models with longer lag lengths including an 'AR(12) + linear trend' as a baseline competitor. Some of these alternatives will be used in the empirical work that follows.

3. AUSTRALIAN MACROECONOMIC DATA

The data we use are quarterly and monthly Australian macroeconomic time series. The quarterly series cover the period 1959(3)–1991(4) and the monthly series cover the periods 1959(1)–1991(12) and 1967(7)–1991(12). All variables except interest rates and stock prices are seasonably adjusted. Table 1 gives details of the thirteen series that we use and the variable notation that we employ.

Table 1: Macroeconomic Variable Notation and Description

Variable	Description	Frequency	Sample Period	Forecast Period
C	Aggregate private final consumption expenditure ($m; sa)	Quarterly	1959(3)–1987(4)	1988(1)–1991(4)
RC	Aggregate real private final consumption exp. ($m average 1984/5 prices; sa)	"	"	"
GDP	Gross domestic product ($m; sa)	"	"	"
RGDP	Real gross domestic product ($m average 1984/5 prices; sa)	"	"	"
PGDP	Implicit price deflator for GDP (sa)	"	"	"
CPI	Consumer price index (1981 = 100; sa)	"	"	"
U	Unemployment rate (%; sa)	"	"	"
WR	Wage rate: Average earnings of non-farm wage and salary earners ($/week)	"	"	"
RWR	Real wage rate (= WR/PGDP)	"	"	"
SP500	Australian share price index: all ordinaries (31 December 1979 = 500)	monthly	1959(1)–1987(12)	1988(1)–1991(12)
Int1	Money market 13 week Treasury Notes (% pa, yield)	"	1969(7)–1987(12)	"
Int2	Capital market 2 year Treasury Bonds (% pa, yield)	"	"	"
Int3	Capital market 10 year Treasury Bonds (% pa, yield)	"	"	"

All of the series except interest rates are lagged. Interest rates are taken in levels (% pa) and reciprocals of levels. The latter transformation (i.e. $x \rightarrow 1/x$) is variance stabilising and reduces the volatility in the series that tends to occur at higher interest rate levels. The reciprocal transformation was found to work well for US bond yields in Phillips (1992) and is therefore used again here. All of the series are graphed in Figures 1(a)–13(a). Figure 11'(a), for instance, shows the short term interest rate, Int1, over the period 1969(7)–1991(12). The increased volatility in this series at higher levels of Int1 is apparent. Figure 11(a) graphs the series in reciprocals, i.e. 1/Int1, over the same period. The effects of stabilising the volatility in this case are quite clear from the two figures. This feature of interest rate data is less apparent for the intermediate rate, Int2, and the long term rate, Int3. However, the transformation is used for both these series as well and the graphs are shown in Figures 12(a), 12'(a) and 13(a), 13'(a), respectively.

4. *BAYES MODELS* FOR THE DATA

Using the PIC model selection criterion we set out to find the best *Bayes model* for the time series described in the last section. Two classes of models were considered. The first was the 'ARMA(p,q) + trend(r)' model given in (2) and the second was the simpler 'AR(p) + trend(r)' model.

The algorithm for determining the trend degree and lag orders of the ARMA model is the one given in Phillips and Ploberger (1992). This algorithm involves the following steps:

Step 1. Set maximum orders for the AR, MA and trend components.

Step 2. Run a long autoregression with maximum trend degree and use PIC or BIC to select the AR order ($\hat{p}$).

Step 3. Select the trend degree ($\hat{r}$) in the model chosen in Step 2 using PIC or BIC. Calculate the residuals $\hat{\varepsilon}_t$ from this regression.

Step 4. Run an array of ARMA(p,q) + trend($\hat{r}$) regressions using $\hat{\varepsilon}_{t-j}$ in place of ε_{t-j} for the MA variable. Choose the orders ($\hat{p},\hat{q}$) using either PIC or BIC.

Step 5. If $\hat{p} > 0$, compare the *Bayes model* selected in Steps 1–4, viz. 'ARMA($\hat{p},\hat{q}$) + trend($\hat{r}$)' with a *Bayes model* of the same order having a unit autoregressive root. Choose the restricted 'ARMA($\hat{p},\hat{q}$) with unit root + trend($\hat{r}$)' model if the posterior odds criterion PIC favours this model (i.e. is greater than unity) over the reference model 'ARMA($\hat{p},\hat{q}$) + trend($\hat{r}$)'. If $\hat{p} = 0$, then there is no autoregressive component and hence no autoregressive unit root.

The algorithm for selecting the best *Bayes model* in the 'AR(p) + trend(r)' class is the same as the above, but simply omits the MA component and hence Step 4. One of our interests is to discover whether this simpler class of models is adequate for most economic time series.

These algorithms of model selection were applied to the thirteen Australian macroeconomic time series described earlier. The empirical results are shown in Table 2. All of these series are found to be stochastically nonstationary. Twelve of the series have a unit autoregressive root while one series, GDP, has a mildly explosive long-run autoregressive coefficient of 1.001 that is preferred to a competing model with a unit root.

Only one of the series (the GDP price deflator) is found to have trend degree $r = 1$. Since the best *Bayes model* for this series also has a unit root, the implied

Table 2: Best *Bayes models* for Australian Macro Time Series

	Block A Model class = ARMA(p,q) + trend(r)				Block B Model class = AR(p) + trend(r)			
Series	Model selected Dynamics	r	ρ[a]	Odds[b]	Model selected Dynamics	r	ρ[a]	Odds[b]
C	ARMA(2,1)	–1	1.000	3313.723	AR(3)	–1	1.000	102.212
RC	AR(1)	0	0.994	11.221				
GDP	AR(2)	–1	1.001	0.000				
RGDP	AR(1)	0	0.992	8.936				
PGDP	ARMA(2,1)	1	0.993	84.493	AR(4)	–1	1.001	255.238
CPI	AR(4)	–1	1.001	472.010				
U	AR(4)	–1	1.005	117.851				
WR	AR(3)	–1	1.002	3.998				
RWR	AR(1)	0	0.986	19.489				
SP500	ARMA(1,1)	–1	1.001	174.232	AR(2)	–1	1.001	745.438
Int1	ARMA(1,1)	–1	0.992	78.325	AR(2)	–1	0.999	52.635
Int2	AR(2)	–1	0.997	196.366				
Int3	AR(1)	–1	0.996	92.544				

[a] Long-run autoregressive coefficient; [b] Posterior odds in favour of a unit root
Note: If the model selected in Block A is an AR(p), the Block B result is identical and hence not repeated

model for the series is a stochastic trend around a quadratic. Three of the series (real consumption, real GDP and the real wage rate) are found to have trend degree $r = 0$, leading to a stochastic trend with drift as the best *Bayes model* for these series.

The dynamics are generally well modelled by autoregressions. But for four series (consumption, the GDP deflator, stock prices and the short term interest rate) low order ARMA models are chosen in place of autoregressions. The choice of dynamic model has no effect on the decision in favour of a unit root for these series. Block B of Table 2 shows the model choice outcomes in the 'AR(p) + trend(r)' class and these can be compared with the outcomes selected in the 'ARMA(p,q) + trend(r)' class given in Block A of the table. There is only one important change from restricting the model class to be autoregressive. For the

GDP deflator series an 'AR(4) + trend(−1)' process is selected as distinct from an 'ARMA(2,1) + trend(1)' process when the model class is wider. Note that the long run autoregressive coefficient is larger for the AR(4) model than the ARMA(2,1) (1.001 as distinct from 0.993) but that a unit root *Bayes model* is chosen in each case. Looking at the graph of this series in Figure 6(a), it is apparent that both models can be rationalised in terms of the historical trajectory. The 'AR(4) + trend(−1)' Bayes model with a unit root is, in fact, the more parsimonious of the two (3 parameters as distinct from 4).

5. *BAYES MODEL* FORECAST PERFORMANCE

The four final years of the sample data (1988–1991) were used for an *ex post* forecasting exercise. This involves sixteen observations for the quarterly series and forty-eight observations for the monthly series. The best *Bayes model* sequence $\{\mathrm{H}(Q_t^B)\}_{t=1988(1)}^{1991(4 or 12)}$ was determined recursively using the PIC criterion. For the quarterly data an 'AR(p) + trend(r)' class was used with $p \leq 5$, $r \leq 1$. For the monthly series the parameters were prescribed as $p \leq 12$, $r \leq 1$. The autoregressive model class was chosen in place of the ARMA class because most of the series seemed to be well modelled within this class as discussed in the previous section. The best *Bayes model* sequence was compared with a fixed format *Bayes model* sequence in terms of their respective one-period ahead forecasting capabilities. For the quarterly data an 'AR(4) + linear trend' fixed format model was used. For the monthly data series, we used both 'AR(4) + linear trend' and 'AR(12) + linear trend' fixed format rules.

Figures 1–13 show the one-period ahead forecast performance of these *Bayes model* sequences over the period 1988–1991 inclusive. For each series Figure (a) displays the data and the relevant forecast period, and Figure (b) shows the period by period forecast errors from the two rival models (the solid line is the Bayes model error and the dashed line is the fixed model error while the origin is given by a dotted line). Figure (c) gives details of the evolving form of the best *Bayes model*: the solid line on the graph shows the autoregressive lag order selected (0–6 lags), the dashed line shows the trend degree (−1 = no intercept; 0 = fitted intercept; 1 = fitted linear trend), and the dotted line shows whether or not a unit autoregressive root is selected (−1 = yes, 0 = no). Figure (d) gives a recursive plot of the forecast encompassing test statistic dQ^B / dQ^F over the forecast period. Table 3 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*.

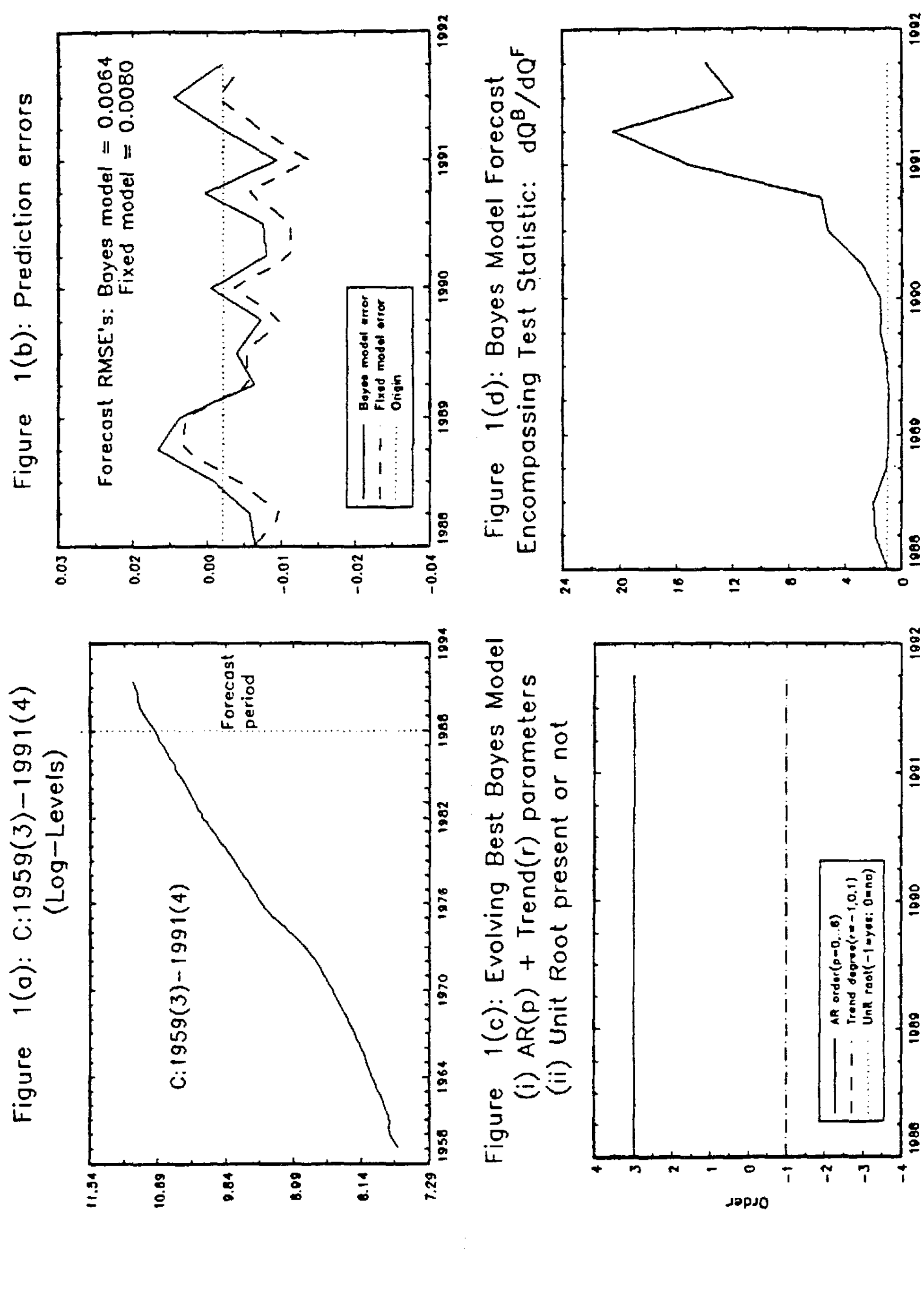


Figure 1(a): C:1959(3)–1991(4)
(Log–Levels)
C:1959(3)–1991(4)
Forecast period
Figure 1(b): Prediction errors
Forecast RMSE's: Bayes model = 0.0064
Fixed model = 0.0080
Bayes model error
Fixed model error
Origin
Figure 1(c): Evolving Best Bayes Model
(i) AR(p) + Trend(r) parameters
(ii) Unit Root present or not
Order
Figure 1(d): Bayes Model Forecast
Encompassing Test Statistic: dQ^B/dQ^F

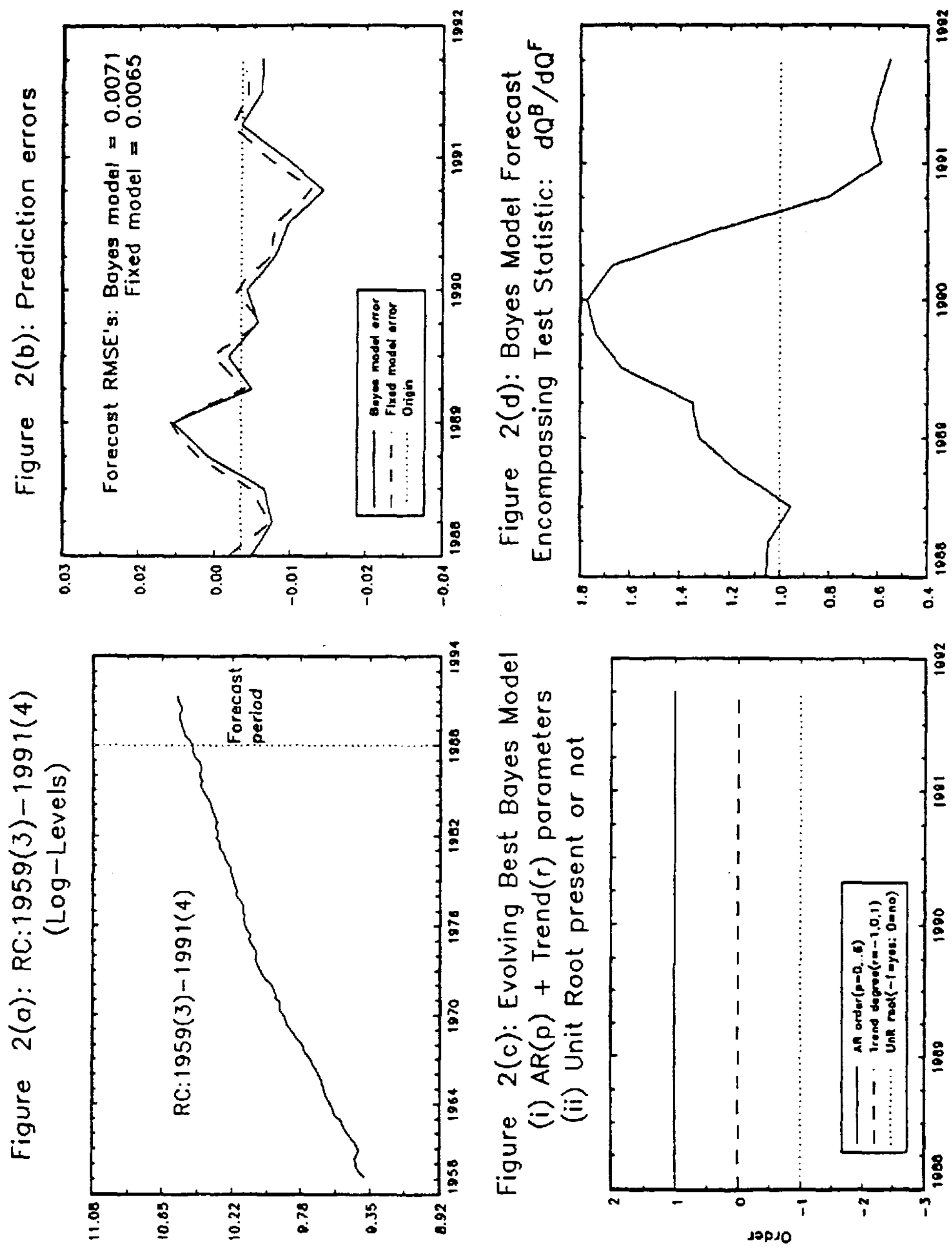


Figure 2(a): RC:1959(3)-1991(4)
(Log-Levels)
RC:1959(3)-1991(4)
Forecast period
Figure 2(b): Prediction errors
Forecast RMSE's: Bayes model = 0.0071
Fixed model = 0.0065
Bayes model error
Fixed model error
Origin
Figure 2(c): Evolving Best Bayes Model
(i) AR(p) + Trend(r) parameters
(ii) Unit Root present or not
Order
Figure 2(d): Bayes Model Forecast
Encompassing Test Statistic: dQ^B/dQ^F

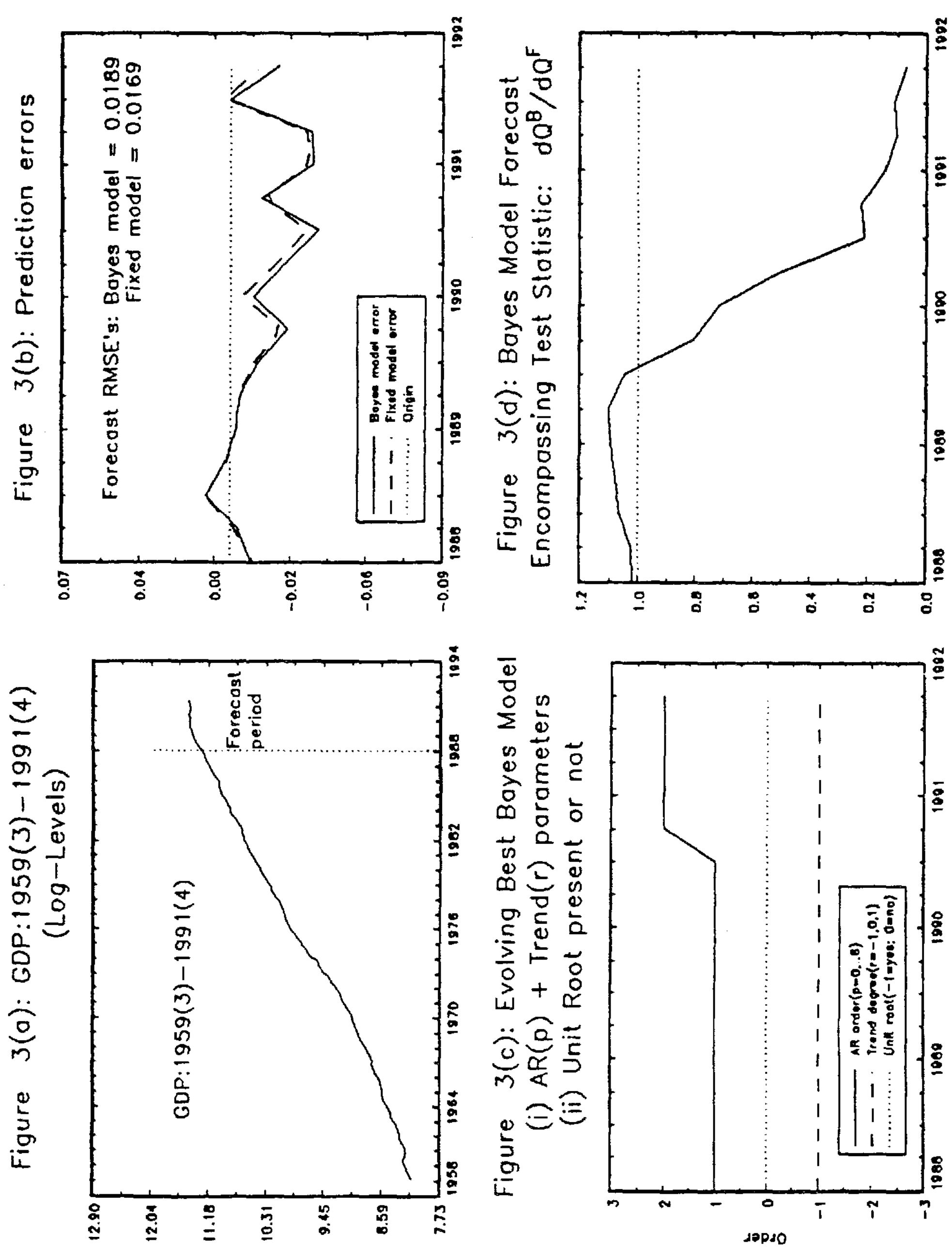


Figure 3(a): GDP:1959(3)-1991(4) (Log-Levels)
GDP:1959(3)-1991(4)
Forecast period
Figure 3(b): Prediction errors
Forecast RMSE's: Bayes model = 0.0189
Fixed model = 0.0169
Bayes model error
Fixed model error
Origin
Figure 3(c): Evolving Best Bayes Model
(i) AR(p) + Trend(r) parameters
(ii) Unit Root present or not
Order
Figure 3(d): Bayes Model Forecast
Encompassing Test Statistic: dQ^B/dQ^F

Figure 4(a): RGDP:1959(3)–1991(4) (Log–Levels)

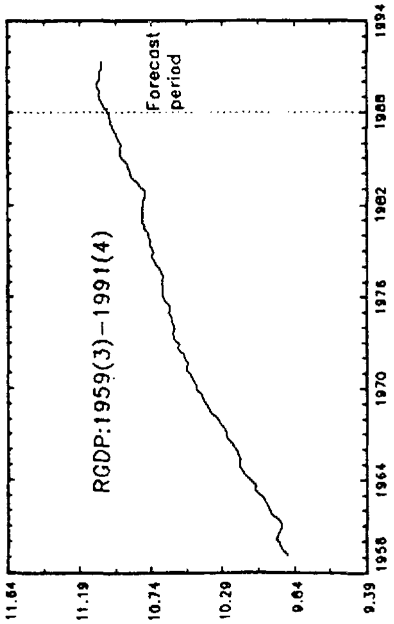


Figure 4(b): Prediction errors

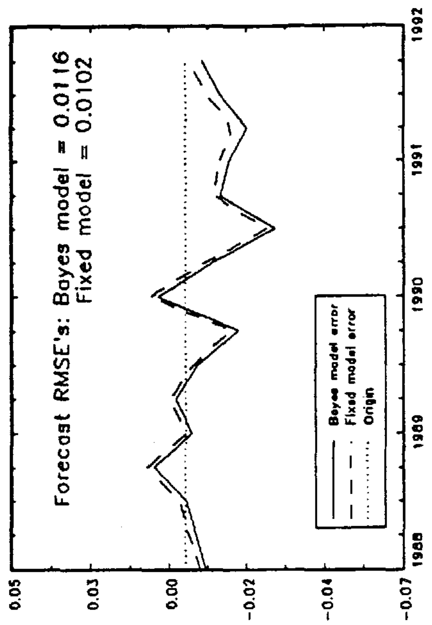


Figure 4(c): Evolving Best Bayes Model (i) AR(p) + Trend(r) parameters (ii) Unit Root present or not

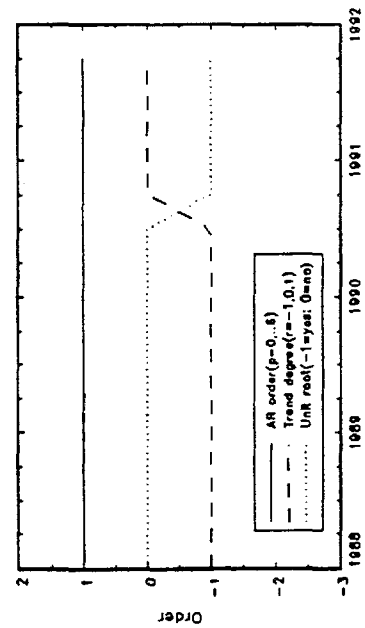


Figure 4(d): Bayes Model Forecast Encompassing Test Statistic: dQ^B/dQ^F

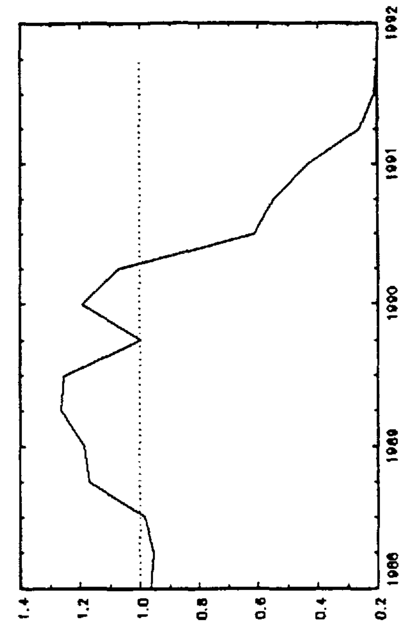


Figure 5(a): PGDP:1959(3)–1991(4) (Log–Levels)

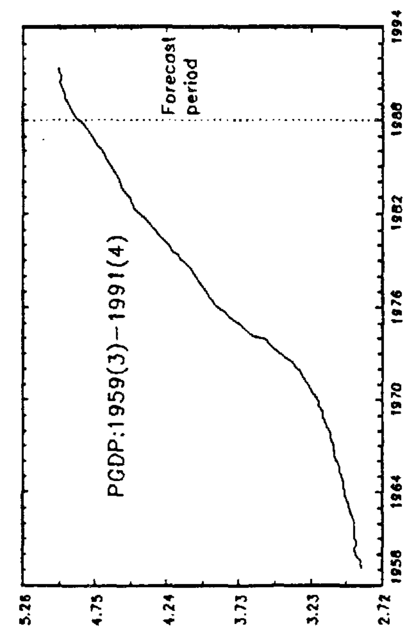


Figure 5(b): Prediction errors

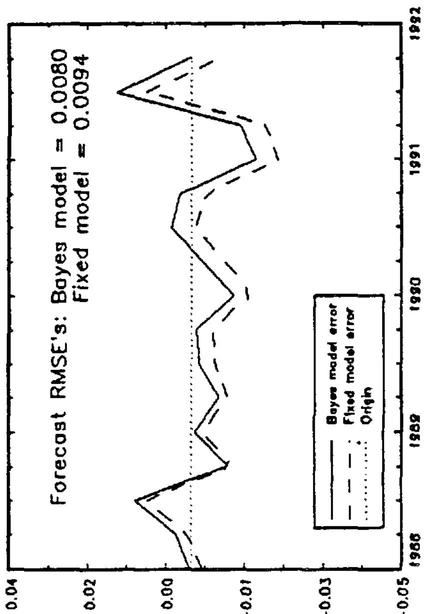


Figure 5(c): Evolving Best Bayes Model (i) AR(p) + Trend(r) parameters (ii) Unit Root present or not

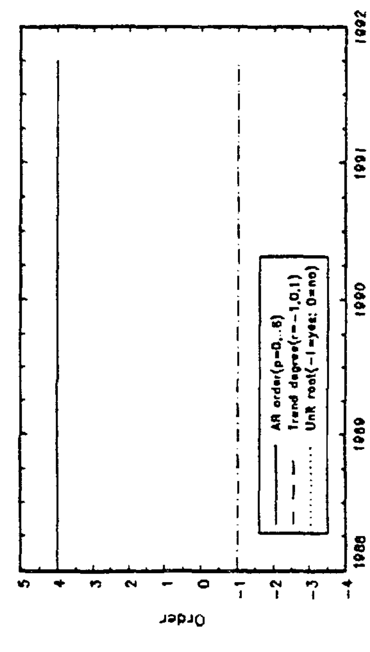
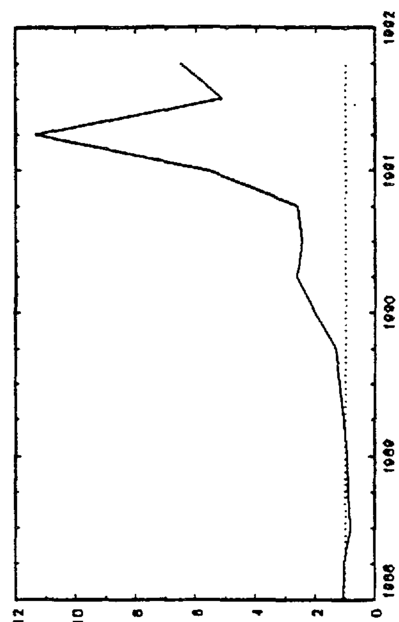


Figure 5(d): Bayes Model Forecast Encompassing Test Statistic: dQ^B/dQ^F



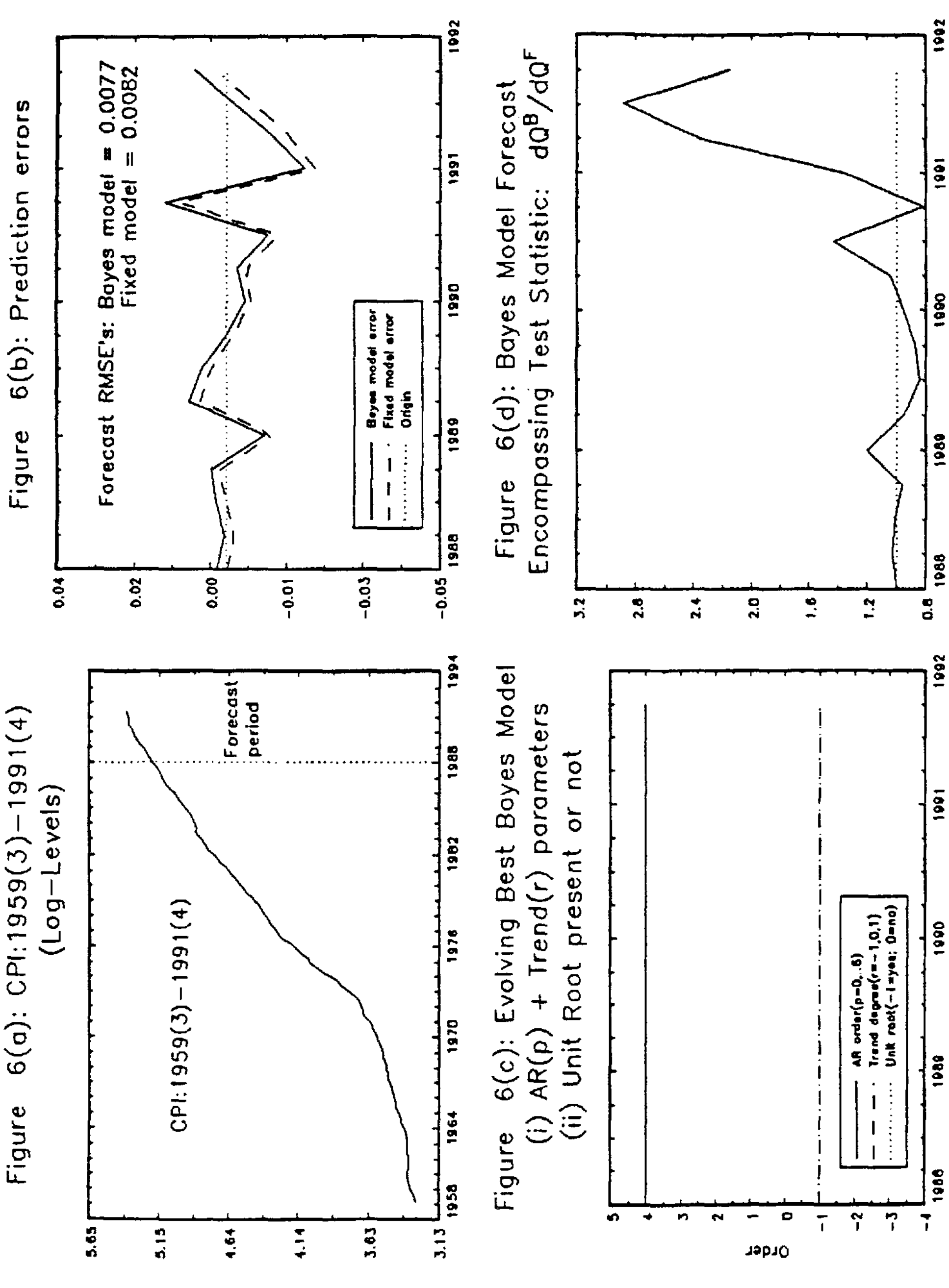


Figure 6(a): CPI:1959(3)-1991(4) (Log-Levels)

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

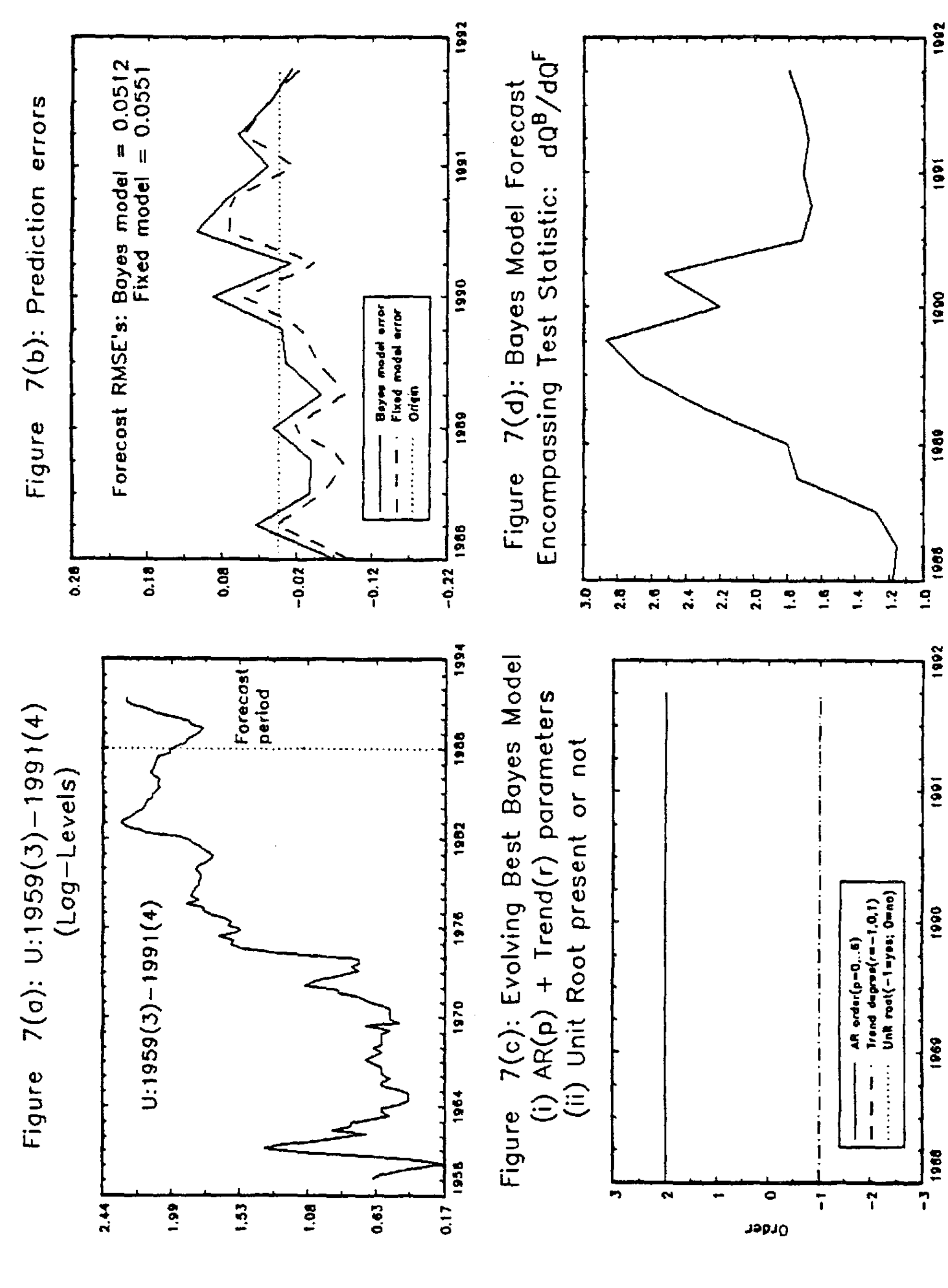


Figure 7(a): U:1959(3)-1991(4) (Log-Levels)

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

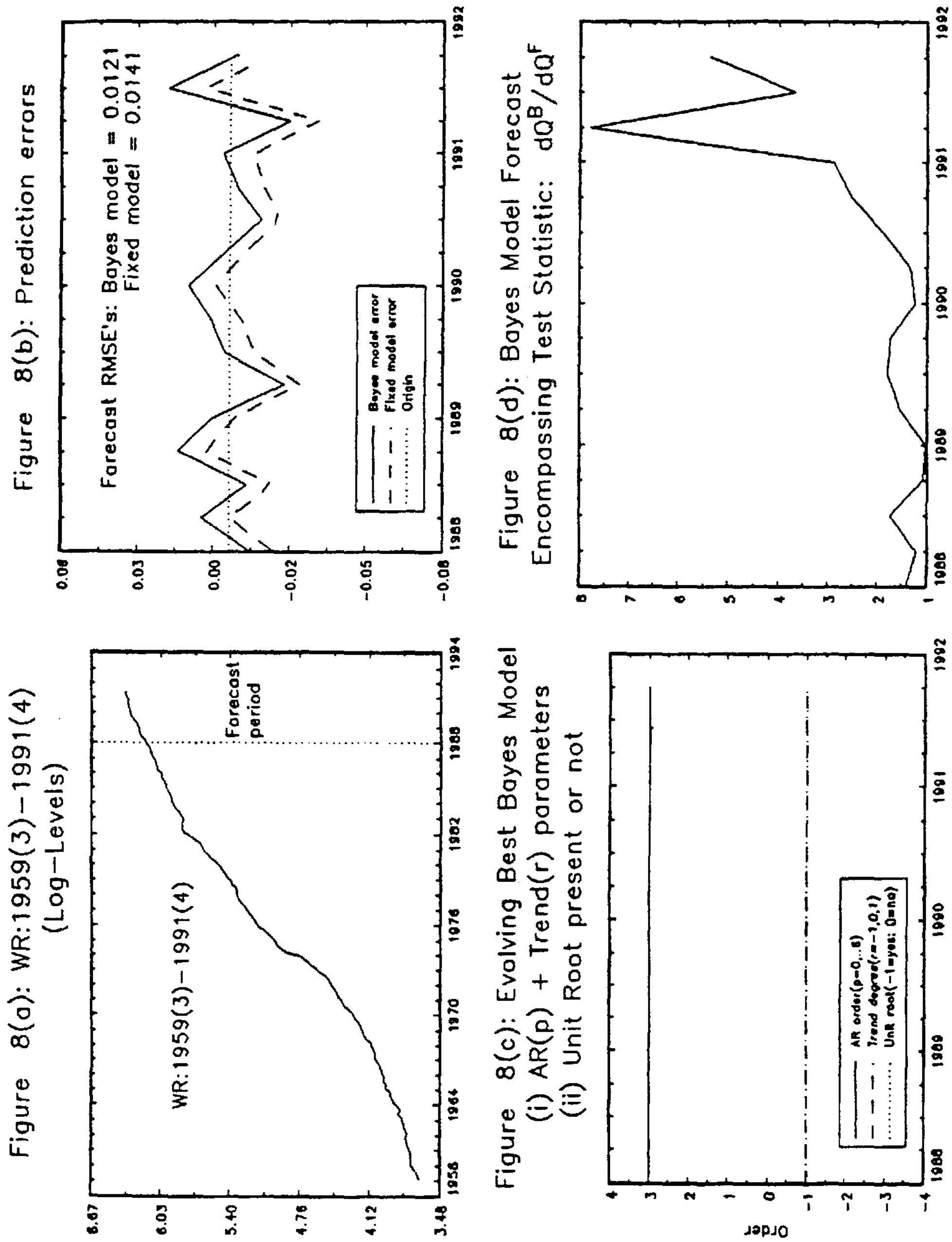


Figure 8(a): WR:1959(3)–1991(4) (Log–Levels)

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

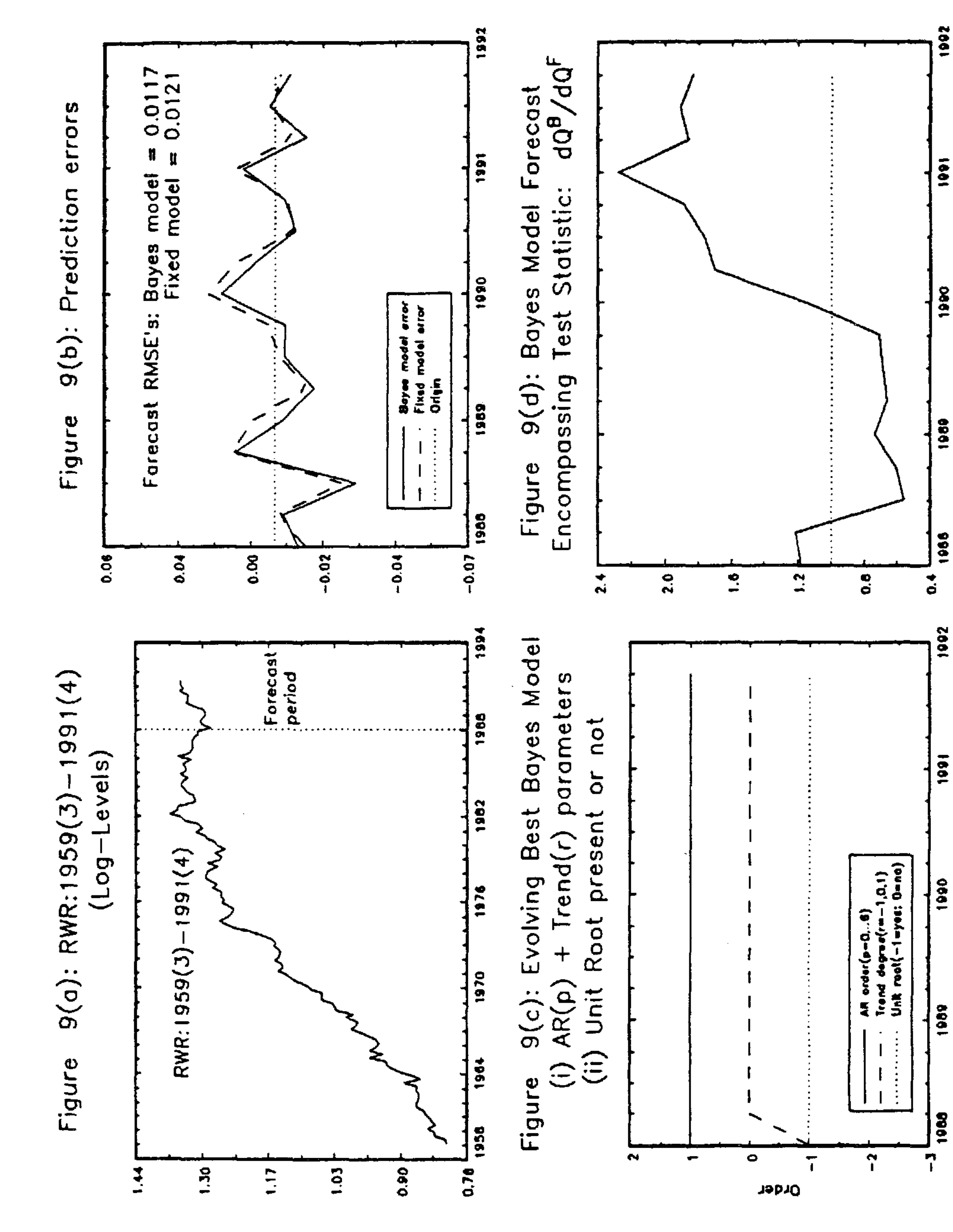


Figure 9(a): RWR:1959(3)–1991(4) (Log–Levels)

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

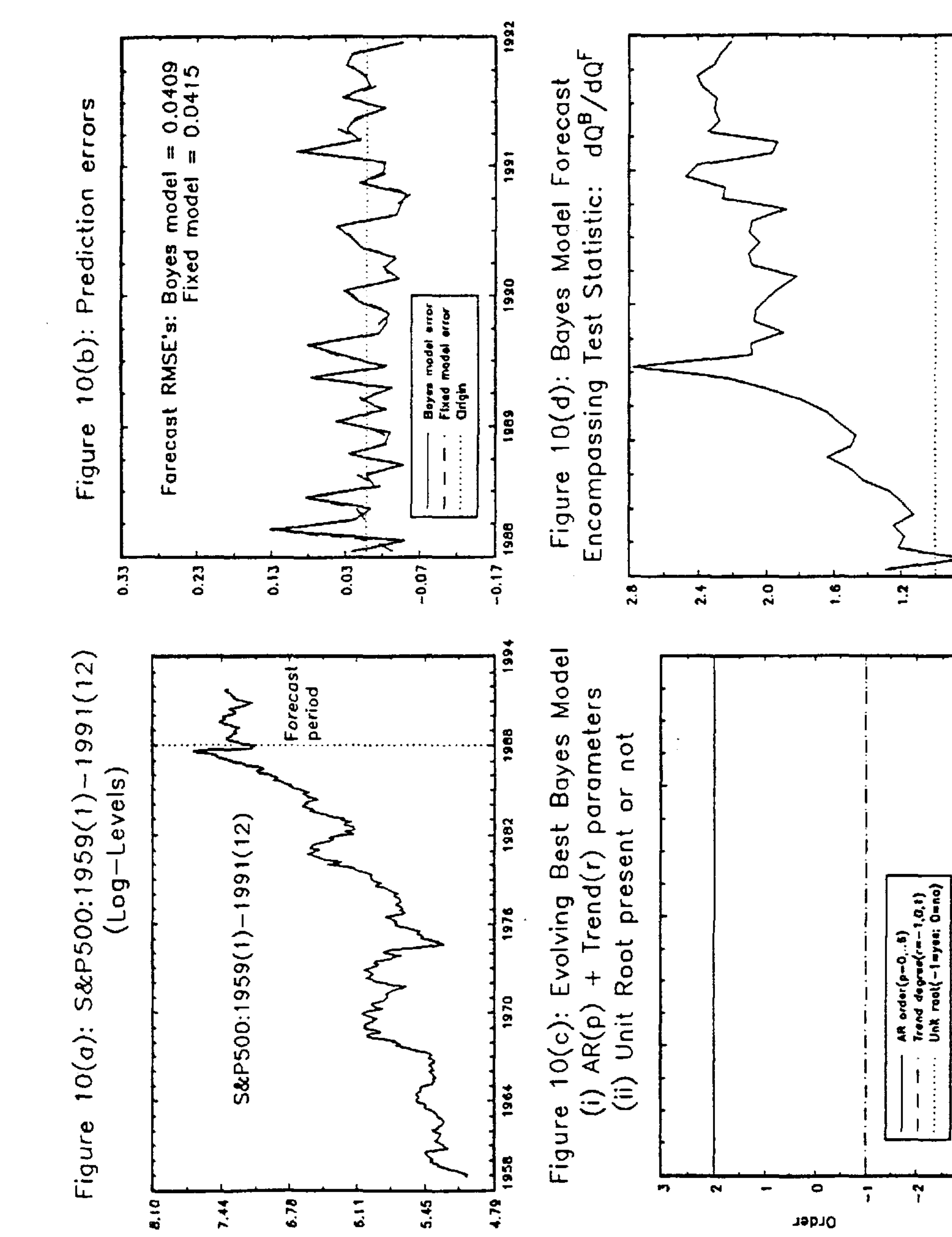


Figure 10(a): S&P500:1959(1)-1991(12) (Log-Levels)

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

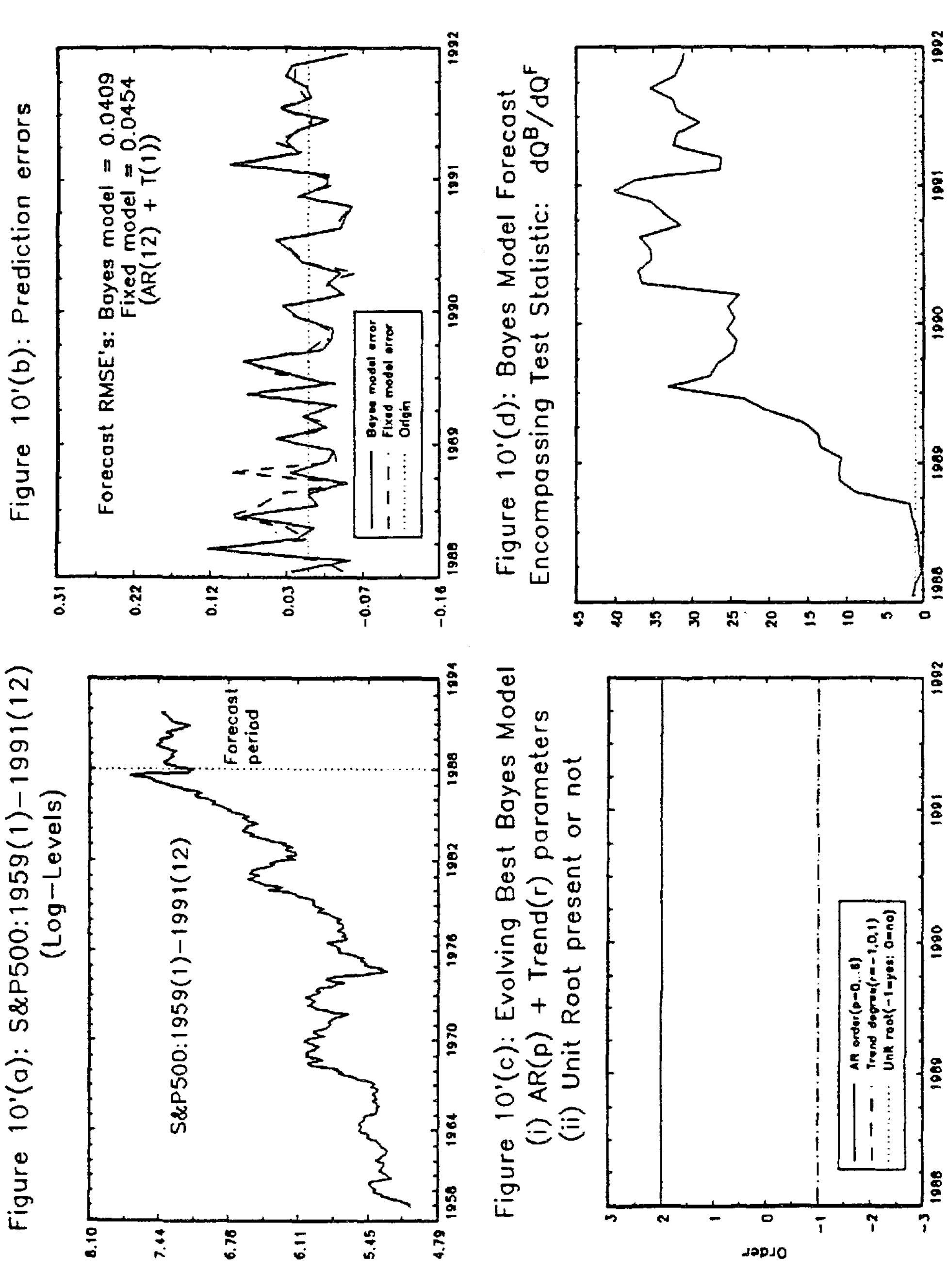


Figure 10'(a): S&P500:1959(1)–1991(12) (Log–Levels)

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

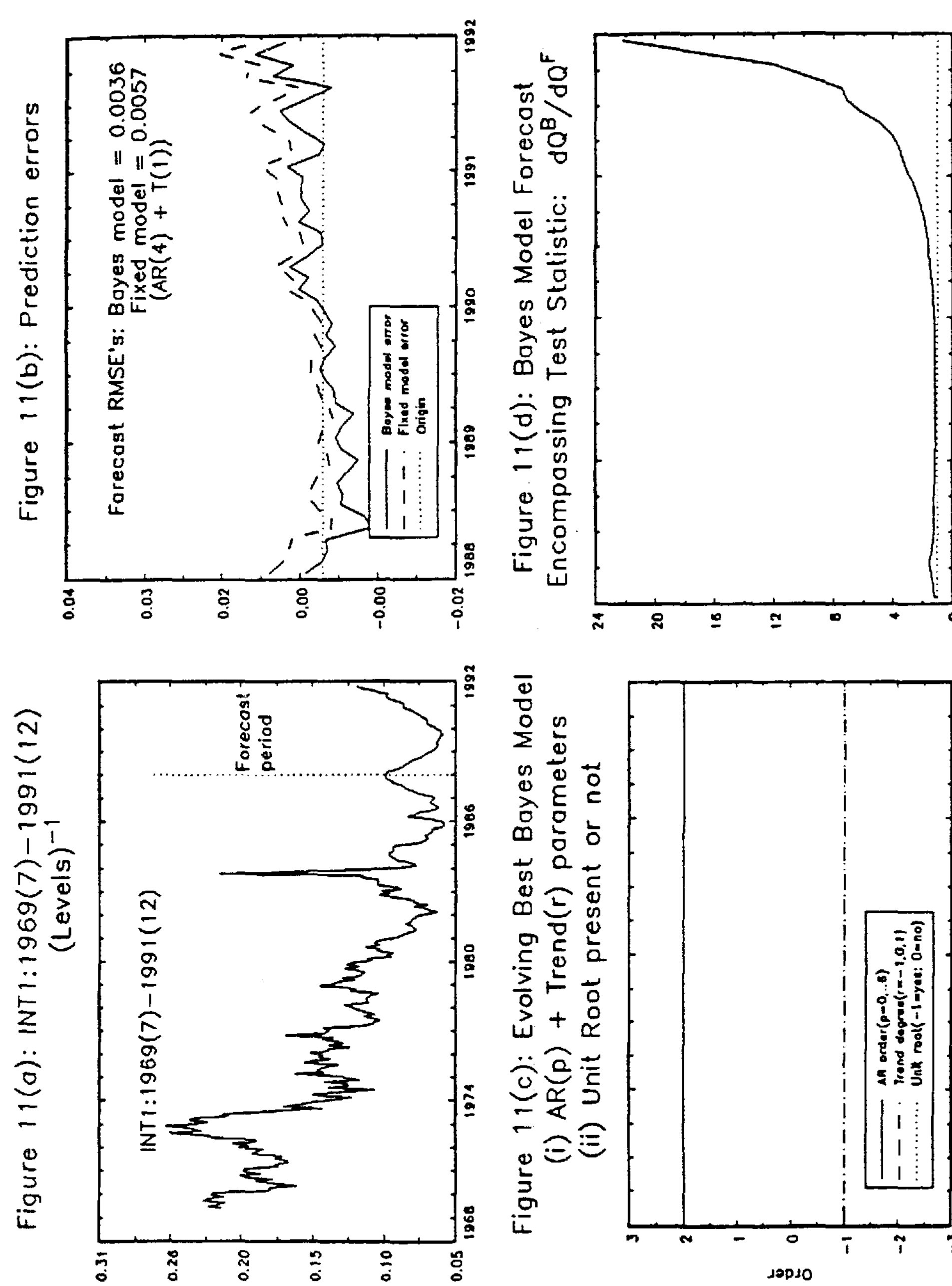


Figure 11(a): INT1:1969(7)–1991(12) (Levels)$^{-1}$

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 11'(a): INT1:1969(7)–1991(12)
Levels

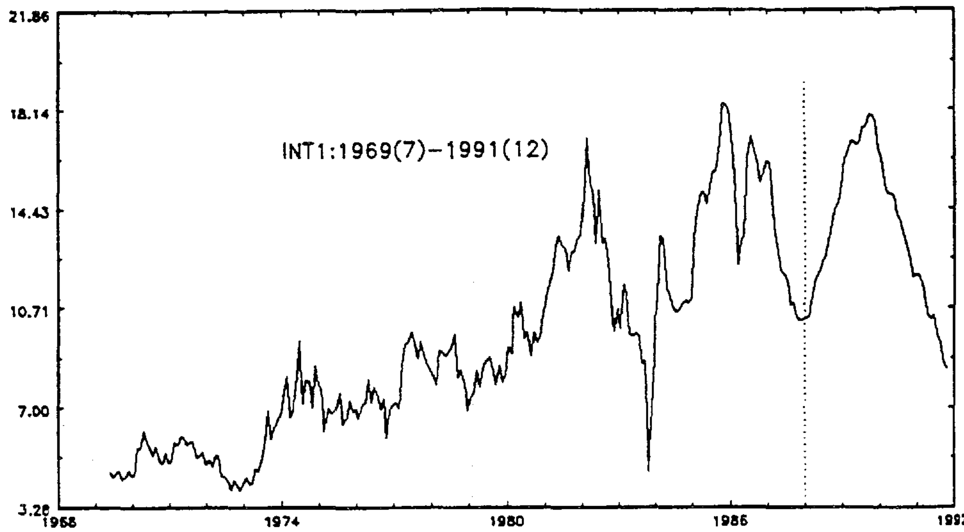
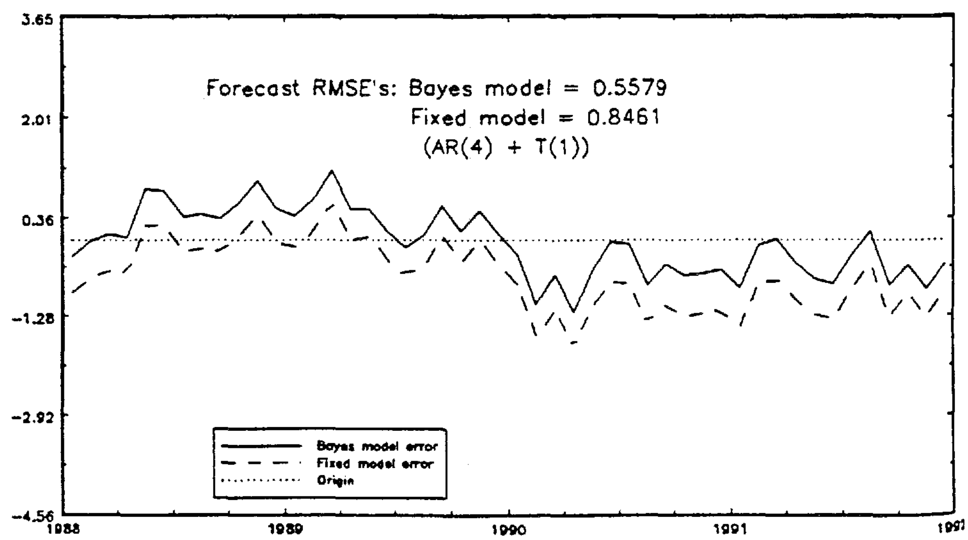


Figure 11'(b): Prediction errors



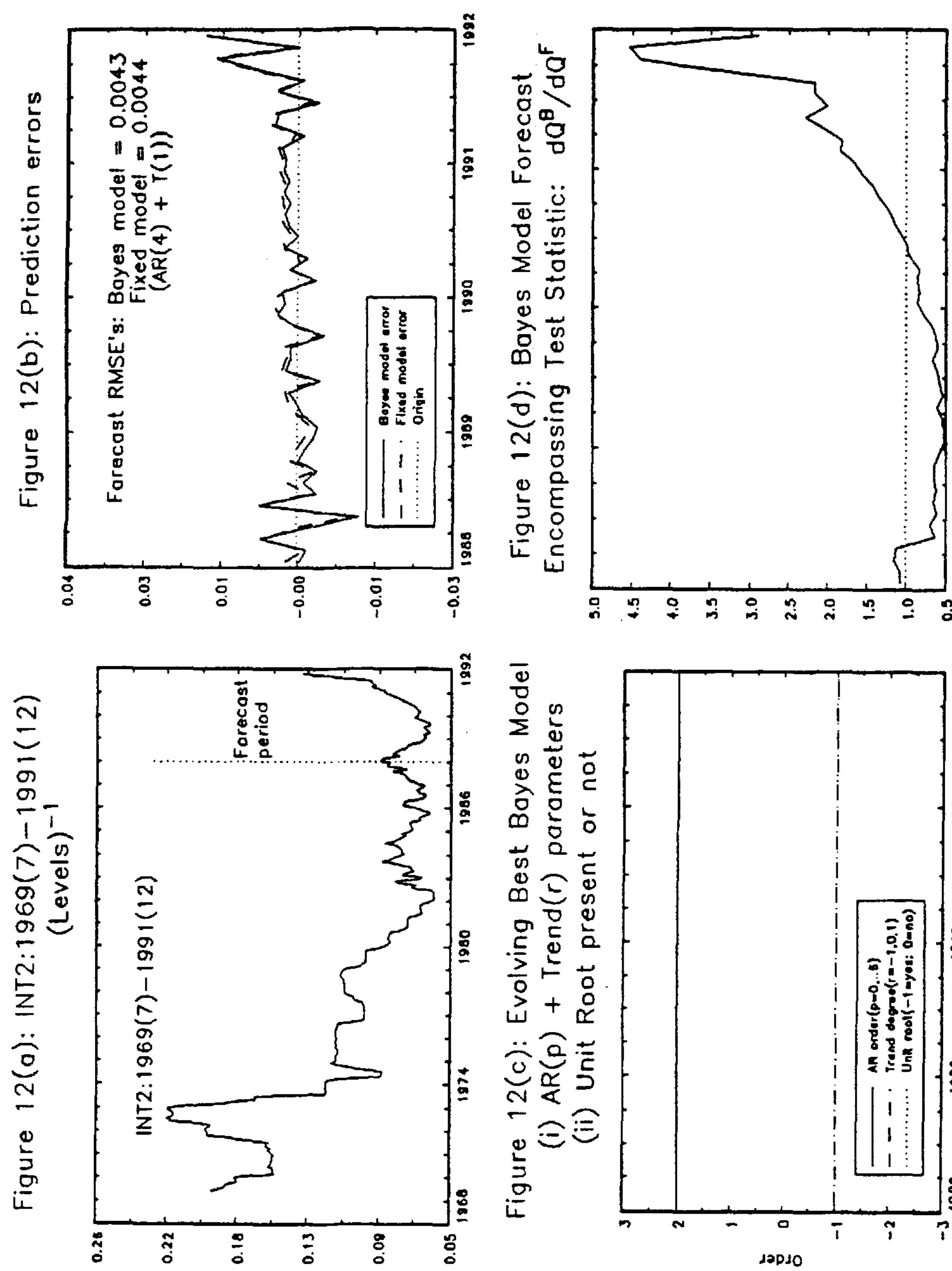


Figure 12(a): INT2:1969(7)−1991(12) $(\text{Levels})^{-1}$

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 12'(a): INT2:1969(7)−1991(12)
Levels

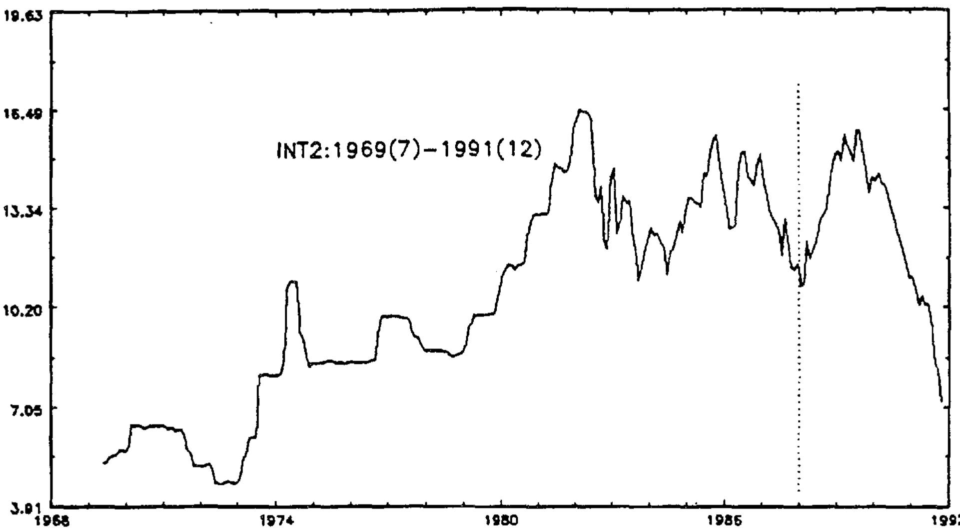
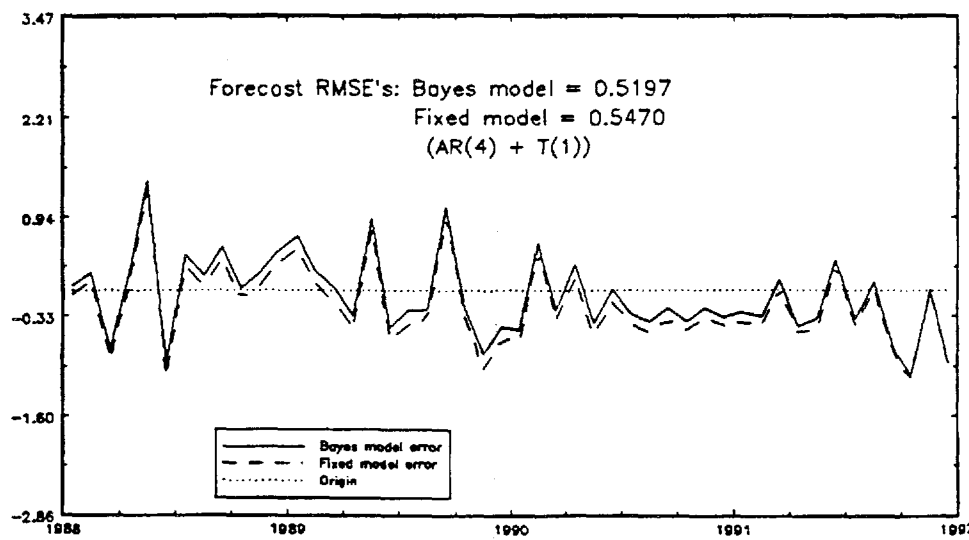


Figure 12'(b): Prediction errors



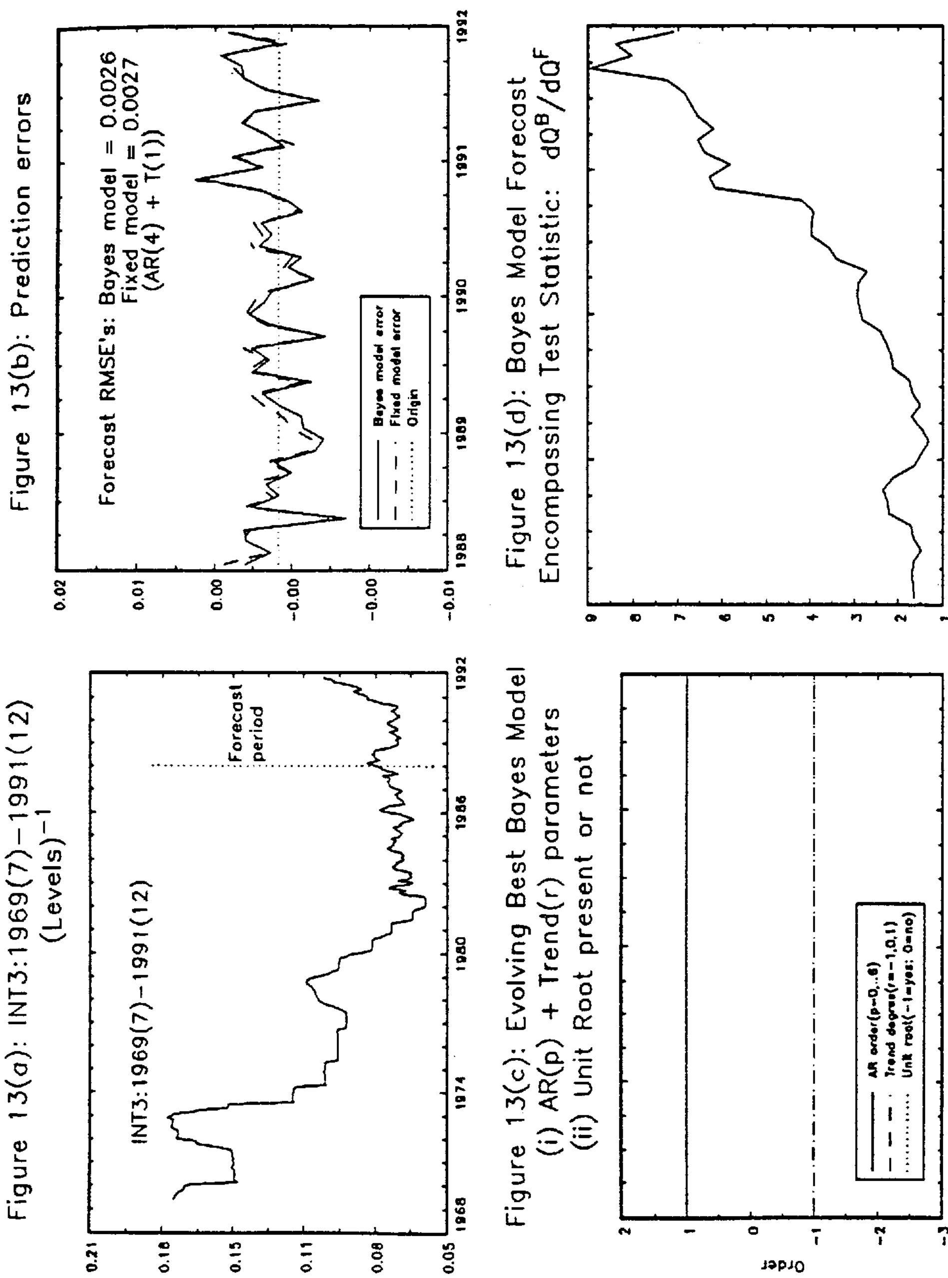


Figure 13(a): INT3:1969(7)−1991(12) (Levels)$^{-1}$

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): INT3:1969(7)–1991(12)
Levels

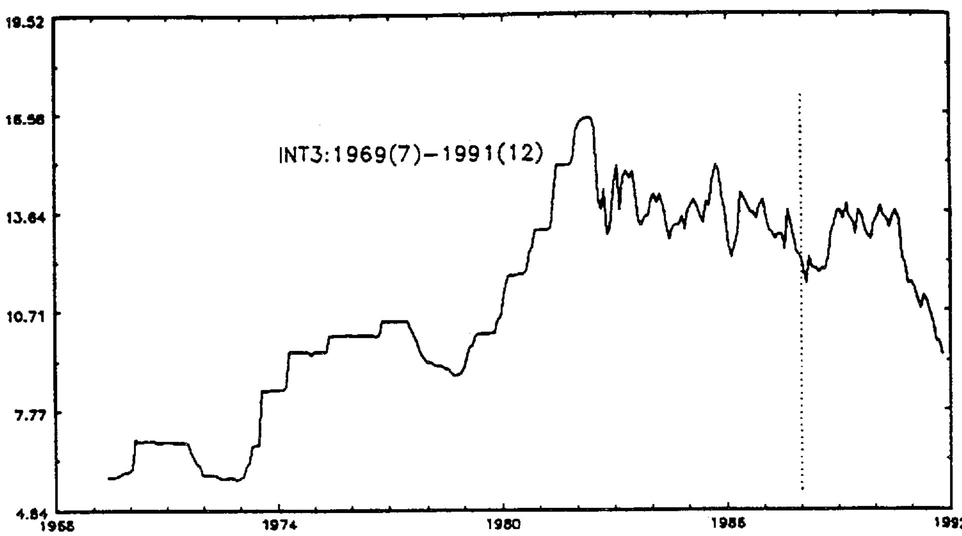


Figure 13'(b): Prediction errors

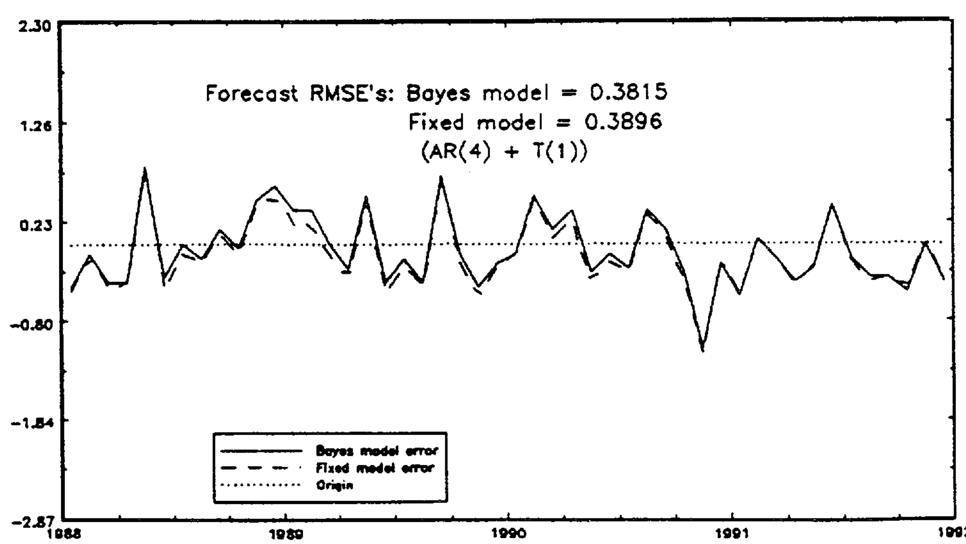


Table 3: Forecasting Exercises for Australian Macroeconomic Time Series, 1988–91

Series	Forecast RMSE Bayes *model*	Fixed model	Number of model changes (date)	Best *Bayes model*	Parameter count ratio *Bayes model* /fixed model	Forecast encompassing test dQ^B/dQ^F in 1991
C	0.0064	0.0080	0	$AR(3)^{-1}$	2/6	13.8880
RC	0.0071	0.0065	0	$AR(1)^{-1}+T(0)$	1/6	0.5491
GDP	0.0189	0.0169	1('90(4))	AR(1); AR(2)	1/6; 2/6	0.0679
RGDP	0.0116	0.0102	1('90(3))	AR(1); $AR(1)^{-1}+T(0)$	1/6; 1/6	0.2006
PGDP	0.0080	0.0094	0	$AR(4)^{-1}$	3/6	6.5094
CPI	0.0077	0.0082	0	$AR(4)^{-1}$	3/6	2.1490
U	0.0512	0.0551	0	$AR(2)^{-1}$	1/6	1.7985
WR	0.0121	0.0141	0	$AR(3)^{-1}$	2/6	5.3556
RWR	0.0117	0.0121	1('88(2))	$AR(1)^{-1}$; $AR(1)^{-1}+T(0)$	0/6; 1/6	1.8320
SP500	0.0409	0.0415	0	$AR(2)^{-1}$	1/6	2.2081
		0.0454[b]			1/14	31.1012
$(Int1)^{-1}$	0.0036	0.0057	0	$AR(2)^{-1}$	1/6	22.1763
Int1[a]	0.5579	0.8461				
$(Int2)^{-1}$	0.0043	0.0044	0	$AR(2)^{-1}$	1/6	2.9060
Int2[a]	0.5197	0.5470				
$(Int3)^{-1}$	0.0026	0.0027	0	$AR(1)^{-1}$	0/6	7.1099
Int3[a]	0.3815	0.3896				

Notes: [a] forecasts for the 'Inti' series were obtained from models for the series in reciprocals, i.e. '$(Inti)^{-1}$'
[b] forecast RMSE for fixed model of form 'AR(12) + T(1)'; $AR(p)^{-1}$ = AR(p) model with a unit root

autoregressive root is selected (–1 = yes, 0 = no). Figure (d) gives a recursive plot of the forecast encompassing test statistic dQ^B/dQ^F over the forecast period. Table 3 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 outcomes from this empirical forecasting exercise are as follows:

(i) For none of the series and for no subperiod is the fixed format 'AR(4) + trend(1)' model a chosen *Bayes model.* Three series (real consumption, real wage rate, real GDP) are chosen to have a unit root with drift. All series except for GDP and real GDP show evidence of a unit root throughout the entire forecasting period. Moreover, the best *Bayes model* for real GDP has a unit root from the 1990(3) quarter and, as noted in the discussion of Table 1, the *Bayes model* for GDP has a mildly explosive long-run autoregressive coefficient. Thus, all series are found to be stochastically nonstationary.

(ii) The best *Bayes model* sequence encompasses the forecasts of the fixed model for all three of the series, these being real consumption, GDP and real GDP. Note from the recursive graphs shown in Figures 2(d), 3(d) and 4(d) that the best *Bayes model* forecasts encompass those of the fixed model for these series also in the first half of the forecast period. For some series the forecast dominance of the best *Bayes model* sequence is substantial and uniform over the forecast period. This is especially notable for consumption, where $dQ^B/dQ^F = 13.888$, the short run interest rate (Int1) where $dQ^B/dQ^F = 22.1763$, and the long run interest rate (Int3) where $dQ^B/dQ^F = 7.1099$.

(iii) From Table 3 it is clear that the best *Bayes models* have a substantial advantage in parsimony over the fixed models. For all series the *Bayes models* have at most 50 percent of the parameters of the fixed model and for ten of the thirteen series the parameter ratio is at most 1/6. Note that the presence of a unit root in the best *Bayes models* for the different series also plays a role in reducing the parameter count. For the long-run interest rate (Int3), the parameter ratio is 0/6 yet the best *Bayes model* – here a martingale – uniformly dominates the fixed model in terms of the forecast encompassing test.

(iv) Root mean squared errors (RMSEs) of forecasts over the period 1988–1991 are given in Table 3. In the graphs, Figure (b) for each series tracks the forecast error generated by each model over the forecast period. By the traditional RMSE criterion the best *Bayes model* is the superior model for ten of the series (consumption, GDP deflator, CPI, unemployment rate, wage rate, real wage rate, stock prices, and the three interest rates). For real consumption, GDP and real GDP the best *Bayes model* has a larger RMSE. For these same series, the *Bayes model* forecasts do not encompass those of the fixed model. So the two criteria reach the same conclusion on which model is superior for each of the thirteen series.

(v) It is worth noting that for some of the series the forecast performance of the best *Bayes model* is quite remarkable given its economical form. Thus, for the consumption series, the *Bayes model* reduces the RMSE of forecast by twenty percent. Looking at Figure 1(b) it is apparent that the *Bayes model* forecasts are substantially and almost uniformly better than those of the fixed model from 1989(3)–1991(4). The *Bayes models* also do very well for the GDP deflator and CPI series. The most dramatic improvement in forecasts comes for the short run interest rate series (Int1). For this series the *Bayes model* (an AR(2) with only one fitted parameter) reduces the RMSE of the fixed model by thirty-six percent from 0.0057 to 0.0036. Figure 11(b) shows that for the subperiod 1990(1)–1991(4) the *Bayes model* is uniformly superior to the fixed model, which consistently underpredicts through this subperiod (leading to a persistently positive forecast error). The reason for this underprediction by the fixed model is clear from the graph of the series in Figure 11(a): a model with a linear trend, like the fixed model, is misspecified. Even though the trend coefficient in the model is revised each period with the latest observation this is not enough to prevent a serious and persistent forecast error. The more parsimonious best *Bayes model* is more flexible, adapts more quickly and convincingly outperforms the fixed model in this case.

(vi) As discussed in section 3, models for the interest rate series are constructed in reciprocals of levels to make the volatility of the series more homogeneous over the sample. Forecasts for both reciprocals of levels and levels are then generated for these series. The results are tabulated in Table 3 and shown in Figure 11', 12' and 13'. In spite of their parsimony, the best *Bayes models* do exceedingly well and dominate the fixed model for all three series both in levels and in reciprocals of levels. In terms of forecasts the odds in favour of the best *Bayes model* are 22.18:1, 2.91:1 and 7.11:1 for Int1, Int2 and Int3 respectively.

(vii) For the monthly series we also considered a fixed model with the format 'AR(12) + trend(1)' to allow for calendar year effects. In each case, this fixed model performed worse than the 'AR(4) + trend(1)' model. Results for this choice of fixed model are shown only for stock prices – see Figures 10'(a)–(d) and Table 3. The best *Bayes model* remains the same in this case and now does even better than before in comparison with the fixed model.

6. CONCLUSION

This paper shows that best *Bayes models* of parsimonious form can be constructed for Australian macroeconomic time series that do very well in competition with fixed format models. For ten out of the thirteen series considered here the *Bayes models* not only improve on the forecasts of more richly parameterised models but also encompass those forecasts. In effect, the predictive distribution of the best *Bayes model* explains the forecasts delivered by the rival model. According to our Bayesian forecast-encompassing statistic and given the actual forecast history, the posterior odds favour the *Bayes models*, sometimes by a factor as high as 30:1, as in the case of the short-term interest rate series.

The models we have considered in this paper are scalar time series models. However, all of the ideas we have employed extend in a natural way to multivariate time series; the statistical theory for this extension will be provided in a subsequent paper. And we hope to conduct some empirical exercises with these multivariate methods on Australian macroeconomic data at a later date.

REFERENCES

HANNAN, E. J. and M. DEISTLER (1988), *The Statistical Theory of Linear Systems*, John Wiley and Sons, New York.

HANNAN, E. J. and J. RISSANEN (1982), 'Recursive Estimation of ARMA Order', *Biometrika*, 69, pp. 273–80 [Corrigenda, *Biometrika*, 1983, p. 70].

MILLS, J. A. and K. PRASAD (1992), 'A Comparison of Model Selection Criteria', *Econometric Reviews*, 11, pp. 201–33.

PAULSEN, J. (1984), 'Order Determination of Multivariate Autoregressive Time Series with Unit Roots', *Journal of Time Series Analysis*, 5, pp. 115–27.

PHILLIPS, P. C. B. (1991), 'Bayesian Routes and Unit Roots: de Rebus Prioribus Semper est Disputandum', *Journal of Applied Econometrics*, 6(4), pp. 435–74.

— (1992), 'Bayesian Model Selection and Prediction with Empirical Applications', *mimeo*, Yale University.

PHILLIPS, P. C. B. and W. PLOBERGER (1991), 'Time Series Modeling with a Bayesian Frame of Reference: I. Concepts and Illustrations', Discussion Paper No. 980, Cowles Foundation.

— (1992), 'Posterior Odds Testing for a Unit Root with Data-based Model Selection', Discussion Paper No. 1017, Cowles Foundation, to appear in *Econometric Theory*, 1994.

PÖTSCHER, B. M. (1989), 'Model Selection Under Nonstationarity: Autoregressive Models and Stochastic Linear Regression Models', *Annals of Statistics*, 17, pp. 1257–74.

RISSANEN, J. (1986), 'Stochastic Complexity and Modeling', *Annals of Statistics*, 14, pp. 1080–100.

— (1987*a*), 'Stochastic Complexity', *Journal of the Royal Statistical Society, B*, pp. 223–39 and 252–65.

— (1987*b*), 'Stochastic Complexity and the MDL Principle', *Econometric Reviews*, pp. 85–102.

SCHWARZ, G. (1978), 'Estimating the Dimension of a Model', *Annals of Statistics*, 6, pp. 461–64.

TSAY, R. S. (1984), 'Order Selection in Nonstationary Autoregressive Models', *Annals of Statistics*, 12, pp. 1425–33.