Econometric Theory, **23**, 2007, 440–463. Printed in the United States of America.
DOI: 10.1017/S0266466607070193

A ROBUST BAYESIAN APPROACH FOR UNIT ROOT TESTING

CATERINA CONIGLIANI
Università Roma Tre

FULVIO SPEZZAFERRI
Università di Roma "La Sapienza"

In this paper we deal with the identification of an autoregressive model for an observed time series and the detection of a unit root in its characteristic polynomial. This is a big issue concerned with distinguishing stationary time series from time series for which differencing is required to induce stationarity. We consider a Bayesian approach, and particular attention is devoted to the problem of the sensitivity of the standard Bayesian analysis with respect to the choice of the prior distribution for the autoregressive coefficients.

1. INTRODUCTION

The detection of a unit root in the characteristic polynomial of an autoregressive model is still one of the most debated issues in time series analysis (see, e.g., Phillips and Xiao, 1998). This is a big issue concerned with distinguishing stationary time series from time series for which differencing is required to induce stationarity, and in practice it is usually addressed through formal hypothesis testing.

In recent years, however, there has been more and more interest in the Bayesian approach to this problem, leading to the comparison of stationary models with those involving a unit autoregressive root via the calculation of posterior odds and of Bayes factors (for a different model selection criterion based on predictive densities see Phillips and Ploberger, 1994).

In particular, the literature has devoted great attention to the nature of suitable noninformative priors for the autoregressive coefficients. As a few examples consider the work of Sims (1988) and Sims and Uhlig (1991), who advocate the use of flat priors. Phillips (1991a) finds that flat priors bias the inference toward stationary models and suggests instead using Jeffreys priors derived from conditional likelihood functions. Also Uhlig (1994a) determines the Jeffreys priors for an AR(1) process from the exact likelihoods and justifies the use of flat priors only in some specific cases. Uhlig (1994b) summarizes most of the

We thank three anonymous referees for their useful comments, which have improved the final version of the paper. Address correspondence to Caterina Conigliani, Facoltà di Economia, Università Roma Tre, via Silvio D'Amico 111, 00145 Roma, Italy; e-mail: caterina.conigliani@eco.uniroma3.it.

 © 2007 Cambridge University Press 0266-4666/07 $15.00

https://doi.org/10.1017/S0266466607070193 Published online by Cambridge University Press

Bayesian contribution to the unit root problem and discusses the sensitivity of the tails of the predictive densities on the prior treatment of explosive roots. Schotman and Van Dijk (1991a) stress the sensitivity of the posterior odds to the size of the stationary region and suggest restricting this size. Many authors (e.g., Schotman and Van Dijk, 1991b; Phillips, 1991b; Zivot, 1994; Marriott and Newbold, 1998) note that, in the unit root case, the lack of identifiability of the process level parameter implies that any improper prior on this parameter generates a posterior probability for the unit root hypothesis that goes to one. Berger and Yang (1994) consider a reference prior approach for the AR(1) model. Marinucci and Petrella (1999) also derive a noninformative prior for the AR(1) model by considering the functional relationship between the autoregressive parameter and the variance of the process. It is particularly interesting to note the paper by Marriott and Newbold (1998), who criticize the use of priors such as the uniform or the Jeffreys prior for the autoregressive coefficients in this setting and advocate the use of sharp, informative prior distributions. However, for the simple problem of testing for a unit root in a first-order autoregressive process, they find that the prior distribution for the autoregressive coefficient has a substantial impact on the posterior odds, so that, e.g., a very sharp beta prior performs extremely well when the generating process is stationary autoregressive, but the uniform prior is preferable when the "true" model is nonstationary. To better clarify this point, consider the two models $Y_t = Y_{t-1} + \eta_t$ versus $Y_t = \varphi_1 Y_{t-1} + \eta_t$ where $\{\eta_t\}$ is a Gaussian white noise process and assume a flat prior for φ_1. If the data are nearly a random walk then the standard procedure is heavily tilted toward finding a unit root, because a flat prior over φ_1 assigns a lot of a priori probability to parameters that we know from the start will have little chance at explaining the data.

The purpose of the present paper is to further investigate the choice of the prior distribution for the autoregressive coefficients in this setting and to derive a *procedure* for the identification of the order of an autoregressive model for an observed time series and the detection of a unit root in its characteristic polynomial, which is *robust* with respect to such prior. To make the model comparison robust to the choice of priors for the autoregressive coefficients, one could use the first l observations for training and let the rest of the data choose a model. In the preceding example, e.g., the updated prior for φ_1 already will be concentrated on the values that are more plausible for explaining the data. Here, to avoid the choice of a specific training sample, we suggest using updated versions of the priors obtained from suitable training likelihoods based on all observations. As in Marriott and Newbold (1998) we remove the problem of the lack of identifiability of the level parameter by considering the first differences of the original series. Furthermore, as the alternative hypothesis to the stationary case, we consider the unit root model, assigning null prior probability to explosive roots (for a discussion on the unit root as hypothesis of interest, see Schotman and Van Dijk, 1991b). Also, we introduce a new parametrization to allow expressing the prior distributions on the parameters of the

https://doi.org/10.1017/S0266466607070193 Published online by Cambridge University Press

models under comparison through conditional distributions obtained from the prior assigned to the largest model.

The models and notation are introduced in Section 2, and the general ideas and key tools for comparing different models in the Bayesian setting are introduced in Section 3; particular attention is devoted to model selection problems for dependent data with weak prior information and to alternative Bayes factors. The marginal distributions for computing the Bayes factors and posterior odds are derived in Section 4. Section 5 considers in detail the simple problem of testing for a unit root in a first-order autoregressive process, and Section 6 deals with the more general case in which the order of the process is also unknown. A few concluding remarks are presented in Section 7.

2. MODELS AND NOTATION

Let $\{Y_t; t \in \mathcal{T}\}$ be a Gaussian stationary stochastic process with mean $\mathbb{E}(Y_t) = \mu$. The process $\{Y_t\}$ is said to be an autoregressive process of order p (ARMA$(p,0)$) if it may be represented as

$$\varphi_p(B)(Y_t - \mu) = \eta_t, \tag{1}$$

where $\varphi_p(B) = 1 - \varphi_1 B - \cdots - \varphi_p B^p$ is the autoregressive polynomial, B is the backshift operator such that $B^k Y_t = Y_{t-k}$, and $\{\eta_t\}$ is a Gaussian white noise process with variance τ^2; the stationarity condition constrains the parameter vector $\boldsymbol{\varphi}^{(p)} = (\varphi_1, \varphi_2, \ldots, \varphi_p)$ to lie in the stationarity region $\{\boldsymbol{\varphi}^{(p)} : \varphi_p(z) = 0 \textit{ implies } |z| > 1\}$.

Suppose we observe the time series $\mathbf{y} = (y_1, \ldots, y_n)$, a finite realization of an autoregressive process of order p, with p unknown; the goal is the identification of p and the simultaneous test for a unit root in the autoregressive polynomial.

Recall that if in the generating process there is a unit autoregressive root, stationarity can be achieved by considering the first differences of the original series. In fact in this case, letting $Z_t = Y_{t+1} - Y_t$, the process $\{Z_t\}$ is stationary ARMA$(p-1,0)$ and can be represented as

$$\phi_{p-1}(B) Z_t = \eta_t, \tag{2}$$

where $\phi_{p-1}(B) = 1 - \phi_1 B - \cdots - \phi_{p-1} B^{p-1}$ and $\phi_j = -\sum_{h=j+1}^{p} \varphi_h$ $(j = 1, \ldots, p-1)$.

It follows that for fixed p, the test for a unit root can be performed by choosing between (1) and (2). However, the use of Bayes factors requires that the comparison between models is done by considering the same observations; thus, following Marriott and Newbold (1998), we also write the stationary process (1) in terms of the first differences. In particular, it is straightforward to verify that if $\{Y_t\}$ is stationary ARMA$(p,0)$, the process $\{Z_t\}$ is stationary

https://doi.org/10.1017/S0266466607070193 Published online by Cambridge University Press

ARMA$(p,1)$ with moving average coefficient equal to 1 and can be represented in terms of the parameter vector $\boldsymbol{\varphi}^{(p)}$ as

$$\varphi_p(B)Z_t = (1-B)\eta_t \tag{3}$$

or equivalently in terms of $\boldsymbol{\phi}^{(p-1)} = (\phi_1, \phi_2, \ldots, \phi_{p-1})$ and ϕ_0 as

$$\phi_p^0(B)Z_t = (1-B)\eta_t, \tag{4}$$

where $\phi_0 = \varphi_1$ if $p = 1$ and $\phi_0 = \sum_{i=1}^{p} \varphi_i$ otherwise, and

$$\phi_p^0(B) = \begin{cases} 1 - \phi_0 B & p = 1 \\ 1 - (\phi_0 + \phi_1)B + \phi_1 B^2 & p = 2 \\ 1 - (\phi_0 + \phi_1)B - \sum_{j=2}^{p-1} (\phi_j - \phi_{j-1})B^j + \phi_{p-1} B^p & p > 2. \end{cases}$$

This latter representation has the advantage of emphasizing that the ARMA$(p-1,0)$ model (2) is nested in the ARMA$(p,1)$ model (4) and can be obtained from (4) by letting $\phi_0 = 1$. More generally, the smaller models ARMA$(p-h,0), h = 2, \ldots, p$, and ARMA$(p-h,1), h = 1, \ldots, p-1$, can be obtained from (4) simply by letting $\phi_{p-1} = \cdots = \phi_{p-h+1} = 0$, $\phi_0 = 1$ and $\phi_{p-1} = \cdots = \phi_{p-h} = 0$, respectively.

Now for a given order p, the test for a unit root can be performed by comparing model (4) and model (2) for the first differences $z_t = y_{t+1} - y_t$ $(t = 1, \ldots, n-1)$; if p is also unknown, the problem of identifying the model for an observed time series and testing for a unit root can be seen as the one of choosing a model for the first differences in the set

$$\mathcal{M}_{\mathbf{z}} = \{M_{p,1}, M_{p-1,0}; p \in \mathcal{P}\}, \tag{5}$$

where $\mathcal{P}$ is a set of possible values for p and $M_{p,q}$ denote the stationary ARMA(p,q) model for $\mathbf{z} = (z_1, \ldots, z_{n-1})$ $(p \in \mathcal{P}; q = 0,1)$.

It should be noted that the primary motivation in Marriott and Newbold (1998) for conducting the analysis in first differences is that this allows them to circumvent having to specify a prior distribution for the mean μ. In fact, Marriott and Newbold (1998) argued that under the stationary model a reasonable prior for μ is the (improper) uniform prior. However, because μ is unidentified under the unit root model, specification of an improper uniform prior will result in an indeterminacy in the Bayes factor. Hence, doing the analysis in the first differences allows them to get around this problem. In the following section we discuss more generally the issue of the sensitivity of the Bayes factor to prior assumptions, and we review some methods recently proposed in the literature for dealing with this problem.

https://doi.org/10.1017/S0266466607070193 Published online by Cambridge University Press

3. BAYES FACTORS AND ALTERNATIVE BAYES FACTORS

The choice of the model that better fits the data is a crucial problem in the various approaches to statistical inference; in the Bayesian setting the key tool for model comparison is the Bayes factor.

Suppose we are comparing two models, M_i and M_j, and let $f_r(\mathbf{z}|\boldsymbol{\theta}_r)$ and $\pi_r(\boldsymbol{\theta}_r)$ be, respectively, the distribution of the data and the prior distribution of the parameters $\boldsymbol{\theta}_r$ under model M_r $(r = i, j)$. The Bayes factor for M_i against M_j is defined as

$$B_{ij}(\mathbf{z}) = \frac{\int f_i(\mathbf{z}|\boldsymbol{\theta}_i)\pi_i(\boldsymbol{\theta}_i)\, d\boldsymbol{\theta}_i}{\int f_j(\mathbf{z}|\boldsymbol{\theta}_j)\pi_j(\boldsymbol{\theta}_j)\, d\boldsymbol{\theta}_j},$$

and it represents the weight of evidence in the data in favor of M_i against M_j. Combining this information with the prior odds we then obtain the posterior probabilities of the models under comparison. Then, e.g., we can predict future observations by taking a weighted average of the optimal predictors for each model, with weights given by these posterior probabilities.

Many problems arise, however, when using the Bayes factor if the prior information is weak, mainly as a consequence of its sensitivity to prior assumptions. In fact, when the prior distributions are proper but diffuse, the more flat the prior, the more penalized the corresponding model; moreover, when they are improper, i.e., defined only up to arbitrary constants, then the Bayes factor is itself a multiple of these constants.

Thus, when the prior information is weak, various authors suggest the use of partial Bayes factors: the idea is to use part of the data as a training sample to update the prior distributions and the remainder of the data to compare the models. Formally, divide the data into two parts. First, subsample $\mathbf{z}(l)$ of size l is used to obtain the posterior distributions $\pi_r(\boldsymbol{\theta}_r|\mathbf{z}(l))$; in the second step, taking these as prior distributions, the remaining data $\mathbf{z}(-l)$ are used to compute a Bayes factor:

$$\begin{aligned} B_{ij}(\mathbf{z}(-l)|\mathbf{z}(l)) &= \frac{\int f_i(\mathbf{z}(-l)|\boldsymbol{\theta}_i)\pi_i(\boldsymbol{\theta}_i|\mathbf{z}(l))\, d\boldsymbol{\theta}_i}{\int f_j(\mathbf{z}(-l)|\boldsymbol{\theta}_j)\pi_j(\boldsymbol{\theta}_j|\mathbf{z}(l))\, d\boldsymbol{\theta}_j} \\ &= \frac{\int f_i(\mathbf{z}|\boldsymbol{\theta}_i)\pi_i(\boldsymbol{\theta}_i)\, d\boldsymbol{\theta}_i}{\int f_i(\mathbf{z}(l)|\boldsymbol{\theta}_i)\pi_i(\boldsymbol{\theta}_i)\, d\boldsymbol{\theta}_i} \Bigg/ \frac{\int f_j(\mathbf{z}|\boldsymbol{\theta}_j)\pi_j(\boldsymbol{\theta}_j)\, d\boldsymbol{\theta}_j}{\int f_j(\mathbf{z}(l)|\boldsymbol{\theta}_j)\pi_j(\boldsymbol{\theta}_j)\, d\boldsymbol{\theta}_j}. \end{aligned} \tag{6}$$

https://doi.org/10.1017/S0266466607070193 Published online by Cambridge University Press

As discussed in O'Hagan (1995, 1997) and in De Santis and Spezzaferri (1997, 1999), partial Bayes factors are less sensitive to prior distributions than Bayes factors. In fact, if $\mathbf{z}(l)$ is a *proper* training sample, i.e., is such that the updated prior distributions $\pi_r(\boldsymbol{\theta}_r|\mathbf{z}(l))$ are proper for $r = i, j$, partial Bayes factors do not depend on arbitrary constants when improper priors are used. Furthermore, they are well defined also with proper priors, and they have the appealing property of balancing the bias introduced to Bayes factors by flat (proper) priors; it follows that they can be successfully used not only as automatic methods when the prior distributions are improper but also in the robust approach to model selection.

There is, however, a difficulty with the use of partial Bayes factors, namely, the selection of the training sample from the data. Among the different proposals to remove the dependence of the partial Bayes factor on a specific training sample, it is particularly interesting to look at that of De Santis and Spezzaferri (1997, 2001), who suggested replacing the likelihood $f_r(\mathbf{z}(l)|\boldsymbol{\theta}_r)$ in (6) with the *geometric training likelihood* $f_r^l(\mathbf{z}|\boldsymbol{\theta}_r)$, defined as the geometric mean of the $f_r(\mathbf{z}(l)|\boldsymbol{\theta}_r)$ as $\mathbf{z}(l)$ varies over the set of all possible proper training samples of a given size l $(r = i, j)$. For independent and identically distributed (i.i.d.) data this proposal leads to the *fractional Bayes factor* (O'Hagan, 1991, 1995), but in more general situations, where the use of fractional Bayes factor is not justified as it is in the case of dependent data, the geometric training likelihood can be used to introduce a new tool for Bayesian model choice with weak prior information, which we can refer to as the *generalized fractional Bayes factor*.

Now, when choosing a model in (5), because of the stationarity constraints, the choice of the prior distribution for the autoregressive parameters under both the stationary model and the integrated one does not represent a problem for the evaluation of Bayes factors and of posterior odds; in fact, even if we assume noninformative priors for these parameters, on the stationarity regions these turn out to be proper. On the other hand, if the prior distribution for τ^2 is improper, the Bayes factor for comparing any two models in (5) is undefined. Now, it may be argued that because τ^2 appears in all the models in (5), the (unknown) multiplicative constants in the Bayes factor would cancel; however, as pointed out in Berger and Pericchi (1996), this is a dangerous practice because what appears to be the same parameter can have very different interpretations, and therefore different prior distributions, under different models. Moreover, whether or not the prior for τ^2 is improper, it was shown by Marriott and Newbold (1998) that the standard Bayesian analysis is extremely nonrobust with respect to the choice of the prior distribution for the autoregressive coefficients. Thus, we will adopt a robust approach and in particular focus our attention on the generalized fractional Bayes factor.

One last issue is worth some consideration. The choice of the size of the training sample has been widely discussed in the literature; see, e.g., O'Hagan (1995, 1997). One simple and obvious choice that has proved to be reliable in a range of problems involving improper priors is to consider the minimal train-

https://doi.org/10.1017/S0266466607070193 Published online by Cambridge University Press

ing sample, i.e., the smallest sample size needed to update an improper prior so as to obtain a proper prior distribution. On the other hand, it is interesting to note the paper by Conigliani and O'Hagan (2000); through the analysis of a suitable sensitivity measure the authors prove that, within the fractional Bayes factor approach, greater robustness with respect to misspecification of the priors can be achieved by increasing the size of the training sample. Here, if we consider an improper prior for τ^2, a minimal training sample is any subset of $\mathbf{z}$ made of a single observation z_t ($t = 1, 2, \ldots, n-1$). Being a *proper* training sample, each z_t carries enough information to update the prior distribution for τ^2 and turn it into a proper distribution. On the other hand, being *minimal*, it does not carry any other information, and in particular it does not carry any information about the autoregressive coefficients; in this sense, to learn also about the autoregressive coefficients and to reduce the impact of their prior distribution on posterior odds and Bayes factors, we expect larger training samples to be more appropriate.

4. COMPUTING THE GENERALIZED FRACTIONAL BAYES FACTOR

Suppose we intend to select a model in (5) for a set of observations $\mathbf{z} = (z_1, \ldots, z_{n-1})$. Let $f_{p,q}(\mathbf{z}|\cdot)$ be the distribution of the data, $\boldsymbol{\theta}_{p,q}$ be the vector of parameters, and $\pi_{p,q}(\boldsymbol{\theta}_{p,q})$ be the prior distribution of the parameters under each model $M_{p,q}$ in (5) ($p \in \mathcal{P}; q = 0, 1$); clearly, from Section 2, and letting $\boldsymbol{\phi}_0^{(p)} = (\phi_0, \phi_1, \ldots, \phi_{p-1})$,

$$\boldsymbol{\theta}_{p,1} = \begin{cases} \tau^2 & p = 0 \\ (\boldsymbol{\phi}_0^{(p)}, \tau^2) & p > 0 \end{cases}$$

and

$$\boldsymbol{\theta}_{p,0} = \begin{cases} \tau^2 & p = 0 \\ (\boldsymbol{\phi}^{(p)}, \tau^2) & p > 0. \end{cases}$$

Moreover, let $\mathbf{z}(l)$ be a training sample of size l and let $f^l_{p,q}(\mathbf{z}|\boldsymbol{\theta}_{p,q})$ be the geometric training likelihood of size l under $M_{p,q}$. The generalized fractional Bayes factor for comparing any two models in (5), say, $M_{i,\delta}$ and $M_{j,\gamma}$, is defined as

$$B^l_{(i,\delta)(j,\gamma)}(\mathbf{z}) = \frac{m_{i,\delta}(\mathbf{z})}{m^l_{i,\delta}(\mathbf{z})} \Big/ \frac{m_{j,\gamma}(\mathbf{z})}{m^l_{j,\gamma}(\mathbf{z})}, \tag{7}$$

where

$$m_{p,q}(\mathbf{z}) = \int f_{p,q}(\mathbf{z}|\boldsymbol{\theta}_{p,q}) \pi_{p,q}(\boldsymbol{\theta}_{p,q})\, d\boldsymbol{\theta}_{p,q}$$

https://doi.org/10.1017/S0266466607070193 Published online by Cambridge University Press

and

$$m^{l}_{p,q}(\mathbf{z}) = \int f^{l}_{p,q}(\mathbf{z}|\boldsymbol{\theta}_{p,q})\pi_{p,q}(\boldsymbol{\theta}_{p,q})\, d\boldsymbol{\theta}_{p,q}$$

are the marginal distributions under $M_{p,q}$ corresponding to the full likelihood, and to the geometric training likelihood, respectively $(p = i, j; q = \delta, \gamma)$.

We now determine the various marginal distributions needed for comparing all models in (5). First, note that because of the stationarity constraints, under both $M_{p,1}$ and $M_{p,0}$ we can define the exact likelihood function (Box, Jenkins, and Reinsel, 1994). Moreover, note that as the integrals with respect to the autoregressive coefficients, assuming either informative or noninformative priors for these parameters, cannot be solved analytically, here we do not specify a particular form for $\pi_{p,1}(\boldsymbol{\phi}_0^{(p)})$ and $\pi_{p,0}(\boldsymbol{\phi}^{(p)})$, whereas for τ^2 we assumed the standard noninformative prior

$$\pi_{p,q}(\tau^2) \propto \frac{1}{\tau^2}$$

$(p \in \mathcal{P}; q = 0,1)$. However, there is one issue concerning the priors $\pi_{p,1}$ and $\pi_{p,0}$ that is worth emphasizing. Let $\bar{p}$ be the maximum value of p in the set $\mathcal{P}$ defined in (5) and consider a prior $\pi_{\bar{p},1}(\boldsymbol{\phi}_0^{(\bar{p})})$ assigned on the stationary region of the largest model $M_{\bar{p},1}$ in (5). As previously noted all the other models $M_{\bar{p}-h,1}$ and $M_{\bar{p}-h,0}(h = 1, \ldots, \bar{p} - 1)$ in (5) can be obtained from $M_{\bar{p},1}$. It follows that extending $\pi_{\bar{p},1}$ by continuity on the set $\phi_0 = 1$, as priors for the parameters $\boldsymbol{\phi}_0^{(\bar{p}-h)}$ and $\boldsymbol{\phi}^{(\bar{p}-h)}$ under models $M_{\bar{p}-h,1}$ and $M_{\bar{p}-h,0}$ we can consider the conditional distributions

$$\pi_{\bar{p},1}(\phi_0, \phi_1, \ldots, \phi_{\bar{p}-h+1}|\phi_{\bar{p}-1} = \cdots = \phi_{\bar{p}-h} = 0)$$

and

$$\pi_{\bar{p},1}(\phi_1, \ldots, \phi_{\bar{p}-h}|\phi_{\bar{p}-1} = \cdots = \phi_{\bar{p}-h+1} = 0, \phi_0 = 1),$$

respectively. It is shown in the Appendix that the preceding distributions are defined on the stationary regions of the corresponding models.

Finally note that when we let $l = 1$, i.e., the size of the minimal training sample, for all p and for all q the marginal distribution corresponding to the geometric training likelihood under $M_{p,q}$ is

$$m^{1}_{p,q}(\mathbf{z}) = \left[\frac{1}{(n-1)}\sum_{t=1}^{n-1} z_t^2\right]^{-1/2},$$

i.e., is independent of p and q, and also of the prior distribution for the autoregressive parameters $\boldsymbol{\phi}^{(p)}$ and $\boldsymbol{\phi}_0^{(p)}$. This is an important point: it confirms

https://doi.org/10.1017/S0266466607070193 Published online by Cambridge University Press

that minimal training samples do not carry enough information about the autoregressive parameters; moreover, it implies that when $l = 1$, the generalized fractional Bayes factor for comparing any two models in (5) equals the corresponding standard Bayes factor. For this reason, to achieve robustness with respect to the choice of the prior distributions for the autoregressive coefficients, we consider the case $l > 1$. In particular, as it was noted in Varshavsky (1996) and Barbieri and Conigliani (1998, 2000) that in the time series setup it seems natural to preserve the time structure for the training samples, instead of considering all possible subsets of the data of a given size, when computing the geometric training likelihood we consider only subsamples $\mathbf{z}_k(l) = (z_{k+1}, \ldots, z_{k+l})$ made of successive observations.

Consider $M_{p,0}$ first and suppose that $p > 0$. Let $\tau^2 V_p^{-1}$ be the variance and covariance matrix of p consecutive variables of the stationary process (2) and let v_{ts} be the (t, s) element of V_p; following Galbraith and Galbraith (1974)

$$v_{ii} = 1 + \sum_{h=1}^{i-1} \phi_h^2 - \sum_{h=p+1-i}^{p} \phi_h^2, \qquad 1 \le i \le p$$

and

$$v_{ij} = -\phi_{j-i} + \sum_{h=1}^{i-1} \phi_h \phi_{h+j-i} - \sum_{h=p+1-j}^{p+i-j} \phi_h \phi_{h+j-i}, \qquad 1 \le i < j \le p,$$

and the other elements of the matrix may be determined by noting that V_p is symmetric with respect to both its principal diagonals. Then the marginal distributions corresponding to the full likelihood and to the geometric training likelihood can be written as

$$m_{p,0}(\mathbf{z}) = \pi^{-(n-1)/2} \Gamma\left(\frac{n-1}{2}\right) \times \int |V_p|^{1/2} \left[\sum_{t=1}^{p}\sum_{s=1}^{p} v_{ts} z_t z_s + \sum_{t=p+1}^{n-1} a_t^2\right]^{-(n-1)/2} \pi_{p,0}(\boldsymbol{\phi}^{(p)})\, d\boldsymbol{\phi}^{(p)}$$

and, for $l > p$,

$$m_{p,0}^{l}(\mathbf{z}) = \left[\frac{\pi}{(n-l)}\right]^{-l/2} \Gamma\left(\frac{l}{2}\right) \times \int |V_p|^{1/2} \left[\sum_{k=0}^{n-l-1} \left(\sum_{t=1}^{p}\sum_{s=1}^{p} v_{ts} z_{k+t} z_{k+s} + \sum_{t=p+1}^{l} a_{k+t}^2\right)\right]^{-l/2} \times \pi_{p,0}(\boldsymbol{\phi}^{(p)})\, d\boldsymbol{\phi}^{(p)},$$

https://doi.org/10.1017/S0266466607070193 Published online by Cambridge University Press

respectively, where

$$a_t = z_t - \sum_{i=1}^{p} \phi_i z_{t-i}.$$

Instead, if $p = 0$, $\mathbf{z} = (z_1, \ldots, z_{n-1})$ is a realization of a white noise process, so that

$$m_{0,0}(\mathbf{z}) = \Gamma\left(\frac{n-1}{2}\right)\left(\pi \sum_{t=1}^{n-1} z_t^2\right)^{-(n-1)/2}$$

and

$$m_{0,0}^{l}(\mathbf{z}) = \Gamma\left(\frac{l}{2}\right)\left(\frac{\pi}{(n-l)} \sum_{k=0}^{n-l-1} \sum_{t=1}^{l} z_{k+t}^2\right)^{-l/2}.$$

Now consider model $M_{p,1}$ and suppose that $p > 0$. Let $\tau^2 W_p^{-1}$ be the variance and covariance matrix of p consecutive variables of the process (4), $b = (0, \ldots, 0, 1)$ be a vector with $p - 1$ elements equal to 0 and the last element equal to 1, and D be the block matrix defined by

$$D = \begin{bmatrix} W_p^{-1} & b' \\ b & 1 \end{bmatrix}^{-1}.$$

Then, following Galbraith and Galbraith (1974), the marginal distributions corresponding to the full likelihood and to the geometric training likelihood can be written as

$$m_{p,1}(\mathbf{z}) = \pi^{-(n-1)/2}\Gamma\left(\frac{n-1}{2}\right) \times \int \left(\frac{|D|}{|D + H'H|}\right)^{1/2} \left[\sum_{t=1}^{n-1} \sum_{s=1}^{n-1} r_{ts} z_t z_s\right]^{-(n-1)/2} \pi_{p,1}(\boldsymbol{\phi}_0^{(p)})\, d\boldsymbol{\phi}_0^{(p)}$$

and, for $l > p$,

$$m_{p,1}^{l}(\mathbf{z}) = \left[\frac{\pi}{(n-l)}\right]^{-l/2}\Gamma\left(\frac{l}{2}\right) \times \int |D|^{1/2} |D + \tilde{H}'\tilde{H}|^{-1/2} \left[\sum_{k=0}^{n-l-1} \sum_{t=1}^{l} \sum_{s=1}^{l} \tilde{r}_{ts} z_{k+t} z_{k+s}\right]^{-l/2} \times \pi_{p,1}(\boldsymbol{\phi}_0^{(p)})\, d\boldsymbol{\phi}_0^{(p)},$$

respectively, where r_{ts} is the (t, s) element of the $(n-1) \times (n-1)$ matrix R defined by $R = A'A - (H'A)'(D + H'H)^{-1}(H'A)$, A is a lower triangular $(n-1) \times (n-1)$ matrix with (r, s) element a_{rs} given by

https://doi.org/10.1017/S0266466607070193 Published online by Cambridge University Press

$$a_{rs} = \begin{cases} 0 & r - s < 0 \\ 1 & r - s = 0 \\ 1 - (\phi_0 + \phi_{r-s}) & 1 \le r - s < p \\ 1 - \phi_0 & r - s \ge p, \end{cases}$$

H is the $(n-1) \times (p+1)$ matrix with (i,j) element h_{ij} given by

$$h_{ij} = \begin{cases} h_{i-1,j} & j - i < 0 \\ h_{i-1,j} - \phi_0 & j - i = 0; \quad p = 1 \\ h_{i-1,j} + \phi_{p-1} & j - i = 0; \quad p > 1 \\ h_{i-1,j} - (\phi_{p+i-j} - \phi_{p+i-j-1}) & 1 \le j - i \le p - 2 \\ h_{i-1,j} - (\phi_0 + \phi_1) & j - i = p - 1; \quad p > 1 \\ 1 & j = p + 1 \end{cases}$$

(where we use the convention that for all $j = 1, \ldots, p$, $h_{ij} = 0$ if $i < 1$), $\tilde{r}_{ts}$ is the (t,s) element of the $l \times l$ matrix $\tilde{R}$ defined by $\tilde{R} = \tilde{A}'\tilde{A} - (\tilde{H}'\tilde{A})'(D + \tilde{H}'\tilde{H})^{-1}(\tilde{H}'\tilde{A})$, $\tilde{A}$ is the lower triangular $l \times l$ matrix obtained from A by considering the first l elements of the first l rows of A, and $\tilde{H}$ is the $l \times (p+1)$ matrix with rows given by the first l rows of H.

5. TESTING FOR A UNIT ROOT IN A FIRST-ORDER AUTOREGRESSIVE PROCESS: A SIMULATION EXPERIMENT

Let $\mathbf{y} = (y_1, \ldots, y_n)$ be a finite realization of a first-order autoregressive process. As discussed in Section 2, the problem of testing for a unit root in its characteristic polynomial can be solved with reference to the differenced time series $\mathbf{z} = (z_1, \ldots, z_{n-1})$ by comparing the ARMA(1,1) model $M_{1,1}$

$$(1 - \phi_0 B) z_t = (1 - B)\eta_t$$

and the ARMA(0,0) model $M_{0,0}$

$$z_t = \eta_t.$$

As pointed out in the Introduction, this problem was considered in detail in Marriott and Newbold (1998), with particular attention devoted to the problem of the sensitivity of the standard Bayesian analysis with respect to the choice of the prior distribution for the autoregressive coefficient. In fact, they criticize the use of priors such as the uniform or the Jeffreys prior for the autoregressive coefficient in this setting, the argument being that it is unreasonable to hold such vague prior belief while at the same time attaching nonzero probability mass at $\phi_0 = 1$. Instead, they advocate the use of sharp, informative prior dis-

https://doi.org/10.1017/S0266466607070193 Published online by Cambridge University Press

tributions, with densities giving more prior weight to values of ϕ_0 close to one; in particular, they find that a realistic representation of the beliefs of an investigator considering the random walk model as a possible explanation for the behavior of a series is a beta distribution with parameters $(\alpha = 5, \beta = 0.5)$ or $(\alpha = 50, \beta = 0.5)$, the latter being a sharper distribution with mean and variance approximately equal to 0.98 and 0.0008.

Thus, they generated 1,000 samples of 100 observations from first-order autoregressive processes with different values of ϕ_0, and for each sample they computed the Bayes factor $B_{(1,1)(0,0)}(\mathbf{z})$ considering either the uniform prior, the beta prior, or the sharp beta prior for ϕ_0. Assuming equal prior probabilities for the two models they computed the proportion of the samples for which the posterior probability of $M_{0,0}$, i.e.,

$$P(M_{0,0}|\mathbf{z}) = \frac{1}{1 + B_{(1,1)(0,0)}(\mathbf{z})},$$

is greater than $\frac{1}{2}$. The results are presented in Table 1 and led the authors to conclude that the sharp beta prior performed extremely well when the "true" process was stationary autoregressive, whereas the opposite was the case for a random walk generating process, in which case the uniform prior should be preferred. The authors conclude that the choice of the prior distribution for ϕ_0 is particularly relevant, so that special care should be devoted to elicit such a prior and "automatic" choices of noninformative priors such as the uniform should be avoided.

For purposes of comparison we consider here the same model selection problem and solve it by means of the generalized fractional Bayes factor $B^l_{(1,1)(0,0)}(\mathbf{z})$. We also explore the behavior of a triangular prior of the form

$$\pi_{1,1}(\phi_0) = \frac{1 + \phi_0}{2}, \qquad |\phi_0| < 1,$$

Table 1. Marriott and Newbold (1998): Proportion of the times that the random walk model would have been chosen for series of 100 observations

	Prior for ϕ_0		
ϕ_0	Uniform	Beta	Beta sharp
1	0.968	0.875	0.677
0.95	0.827	0.554	0.244
0.85	0.212	0.018	0.000

https://doi.org/10.1017/S0266466607070193 Published online by Cambridge University Press

which is somehow intermediate between the uniform and the beta. The results are presented in Table 2, and we will see that they lead to very different conclusions with respect to those in Marriott and Newbold (1998), in favor of the use of noninformative priors.

Two issues clearly arise from Table 2, as a direct consequence of the use of the updated priors in the generalized fractional Bayes factor (7). First, notice that as l increases, the differences in the behavior of the various prior distributions are less relevant. In fact, whereas for $l = 1$ we obtain results in accordance with those in Marriott and Newbold (1998), when $l = 20$ the uniform prior, the triangular prior, and the beta prior give similar results both when the "true" process is stationary autoregressive and when it is a random walk. Second, as l increases the uniform prior performs better than when $l = 1$ in the stationary case and only slightly worse in the random walk case. Similar behavior is shared also by the triangular and beta prior, whereas no improvements are obtained by increasing l when using the beta sharp.

Indeed, the sharper the prior, the less the updated prior distribution is sensitive to the training likelihood, so that we did not expect the behavior of the beta sharp to change significantly by increasing l. Moreover, it is not surprising that if in the stationary case we assume a prior distribution that gives a lot of prior weight in the neighborhood of the true ϕ_0, this will always perform better than a flat prior, no matter what the value of l. And in fact for $l = 20$ the uniform prior still (wrongly) favors the unit root model more often than the beta sharp for both $\phi_0 = 0.95$ and $\phi_0 = 0.85$.

To further investigate this issue, we compare the performance of the different priors simulating series of 10 observations from first-order autoregressive processes with $\phi_0 = 0.3, 0, -0.3$; note that with samples of 100 observations, whatever prior we assume for ϕ_0, the random walk model would never be chosen in these cases. The results are presented in Table 3 for $l = 1$ and again show that the two beta priors perform slightly better than the uniform and the triangular prior when the true process is stationary. However, it is interesting to look also at the shapes of the empirical cumulative distribution function for the posterior probability of $M_{0,0}$. For instance if we consider the case $\phi_0 = 0$, we can see from Figure 1 that under either one of the beta priors the empirical c.d.f.s for $P(M_{0,0}|\mathbf{z})$ exhibit a steep growth around the median, so that only very rarely will $P(M_{0,0}|\mathbf{z})$ be either very large or very small. In other words, although the posterior probabilities corresponding to the beta priors almost always choose the right model, they hardly ever provide strong evidence in favor of the stationary model as the uniform prior does. The behavior of the four c.d.f.s in Figure 1 can also be synthesized by averages and standard deviations of the simulated values of $P(M_{0,0}|\mathbf{z})$. They are, respectively, 0.21858 and 0.17975 under the uniform prior, 0.19879 and 0.14547 under the triangular prior, 0.32313 and 0.07920 under the beta prior, and 0.48758 and 0.00251 under the beta sharp.

These results are particularly interesting and suggest that even in the stationary case very sharp prior distributions perform rather poorly. In fact, for almost

https://doi.org/10.1017/S0266466607070193 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466607070193 Published online by Cambridge University Press

TABLE 2. Proportion of the times that the random walk model would have been chosen for series of 100 observations ($l = 1$, 10, 20)

	Prior for ϕ_0											
	Uniform			Triangular			Beta			Beta sharp		
ϕ_0	$l = 1$	$l = 10$	$l = 20$	$l = 1$	$l = 10$	$l = 20$	$l = 1$	$l = 10$	$l = 20$	$l = 1$	$l = 10$	$l = 20$
1	0.984	0.924	0.871	0.963	0.917	0.868	0.870	0.814	0.759	0.579	0.574	0.565
0.95	0.916	0.703	0.520	0.859	0.673	0.506	0.521	0.431	0.333	0.125	0.125	0.125
0.85	0.392	0.083	0.045	0.240	0.063	0.039	0.030	0.021	0.023	0.001	0.001	0.002

TABLE 3. Proportion of the times that the random walk model would have been chosen for series of 10 observations ($l = 1$)

	Prior for ϕ_0			
ϕ_0	Uniform	Triangular	Beta	Beta sharp
0.3	0.291	0.170	0.045	0.015
0	0.090	0.044	0.010	0.004
−0.3	0.022	0.010	0.000	0.000

all samples generated from the stationary model, under the beta sharp the posterior odds are practically equal to the prior odds, and the information on the true model contained in the data is ignored. The same problem in terms of posterior uncertainty in the model choice can be observed also in the simulation experiment summarized in Table 2. Really, when we use the beta sharp we effectively compare models that are very close to each other and practically indistinguishable. For this reason we argue that the criterion based on the proportion of the samples for which the posterior probability of one model is greater than $\frac{1}{2}$ can be misleading when comparing the different priors. A better understanding of the implications on the posterior odds of the prior specification can be achieved by looking at the shape of the empirical cumulative distribution functions. Moreover, we argue that a more disperse distribution such as the uniform, together with an appropriate size l of the training samples, should be preferred.

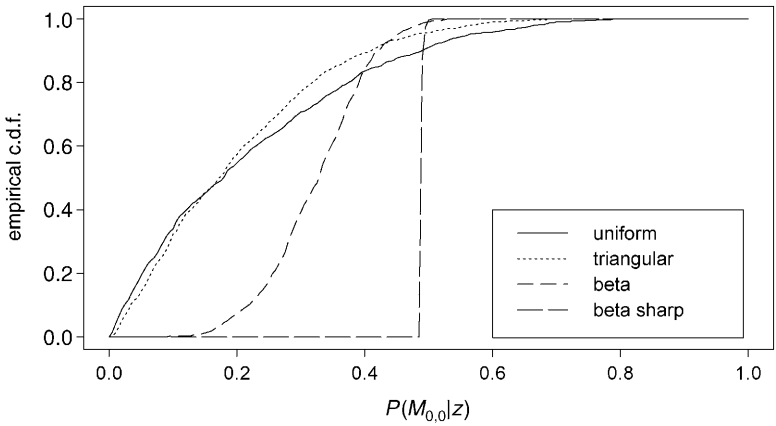


FIGURE 1. Empirical c.d.f. for $P(M_{0,0}|\mathbf{z})$ for series of 10 observations ($l = 1, \phi_0 = 0$).

https://doi.org/10.1017/S0266466607070193 Published online by Cambridge University Press

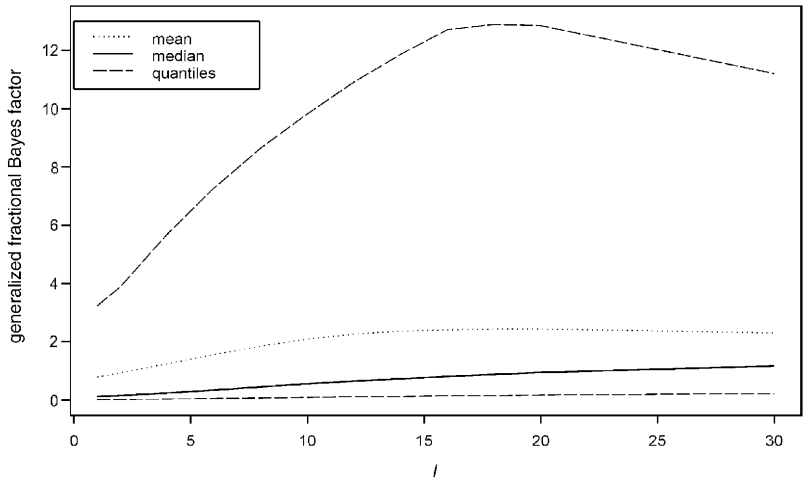


FIGURE 2. Summaries of the distribution of the generalized fractional Bayes factor for different values of l ($\phi_0 = 0.95$).

As for the problem of what an appropriate size of the training samples could be, some useful indications can be obtained from Figures 2 and 3, which show how the distribution of the generalized fractional Bayes factor varies with l when we assume a uniform prior for ϕ_0. In particular, for the simulation experiment of Table 2, the two figures show the mean, the median, and the 2.5% and

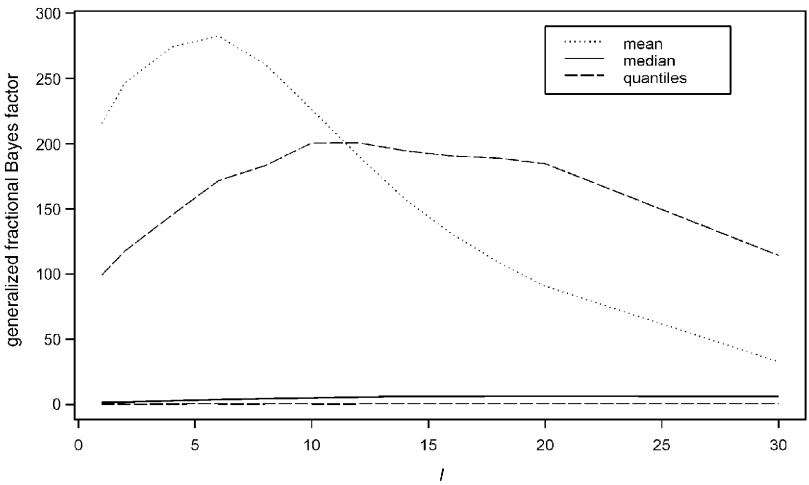


FIGURE 3. Summaries of the distribution of the generalized fractional Bayes factor for different values of l ($\phi_0 = 0.85$).

https://doi.org/10.1017/S0266466607070193 Published online by Cambridge University Press

Table 4. Comparison of unit root test criteria in terms of size and power for series of 100 observations

		Prior for ϕ_0							
		Uniform			Beta sharp			D-F	MLE
	ϕ_0	$l = 1$	$l = 10$	$l = 20$	$l = 1$	$l = 10$	$l = 20$	test	test
α		0.016	0.076	0.129	0.421	0.426	0.435	0.050	0.050
$1 - \beta$	0.95	0.084	0.297	0.480	0.875	0.875	0.875	0.117	0.276
$1 - \beta$	0.85	0.608	0.917	0.955	0.999	0.999	0.998	0.621	0.905

the 97.5% quantiles of $B^l_{(1,1)(0,0)}(\mathbf{z})$ for $\phi_0 = 0.95$ and $\phi_0 = 0.85$, respectively. In both cases we see that at $l = 1$ the distribution of the generalized fractional Bayes factor is significantly concentrated on the interval (0,1). Then, as l starts to increase, it gives more weight to values bigger than 1, so that less often $B^l_{(1,1)(0,0)}(\mathbf{z})$ wrongly chooses the random walk model. Finally, for large values of l, the mass of the distribution tends to be concentrated on 1, because the more observations are used for training, the fewer observations are left to choose between the models; in the limiting case, as l approaches $n = 100$, no model choice can be done.

Finally, note that although our approach is fully Bayesian, its performance also can be interpreted within the usual hypothesis testing framework. Table 4 shows some of the results of Table 2 in terms of size and power of the test, together with those derived in Pantula, Gonzales-Farias, and Fuller (1994) concerning the usual Dickey–Fuller (D-F) test and a test based on maximum likelihood estimation (MLE). Note that in the uniform case considerably more power can be achieved by increasing l, whereas the size of the test increases only slightly.

6. TESTING FOR A UNIT ROOT AND ESTIMATING THE ORDER OF AN AUTOREGRESSIVE PROCESS: A SIMULATION EXPERIMENT

Let $\mathbf{y} = (y_1, \ldots, y_n)$ be a finite realization of an autoregressive process of order p unknown and suppose that p varies in the set $\mathcal{P} = \{1,2\}$. As discussed in Section 2, the problem of identifying the model for this series can be solved with reference to the differenced time series $\mathbf{z} = (z_1, \ldots, z_{n-1})$ by comparing the models summarized in Table 5.

For the specification of the prior distributions under these models, two points should be taken into account. First, note that all models are nested in $M_{2,1}$; in particular, $M_{1,0}$, $M_{1,1}$, and $M_{0,0}$ can be obtained from $M_{2,1}$ by letting $\phi_0 = 1$, $\phi_1 = 0$, and $(\phi_0, \phi_1) = (1,0)$, respectively. It follows that once $\pi_{2,1}(\phi_0, \phi_1)$ is introduced under $M_{2,1}$, the priors to be considered under $M_{1,0}$ and $M_{1,1}$ are the conditional distributions:

https://doi.org/10.1017/S0266466607070193 Published online by Cambridge University Press

TABLE 5. Models under comparison when $p \in \{1,2\}$

Label	Model	Stationarity region
$M_{2,1}$	$(1-(\phi_0+\phi_1)B+\phi_1B^2)z_t=(1-B)\eta_t$	$-3<\phi_0<1;\quad -[(1+\phi_0)/2]<\phi_1<1$
$M_{1,0}$	$(1-\phi_1B)z_t=\eta_t$	$-1<\phi_1<1$
$M_{1,1}$	$(1-\phi_0B)z_t=(1-B)\eta_t$	$-1<\phi_0<1$
$M_{0,0}$	$z_t=\eta_t$	

$$\pi_{1,0}(\phi_1)=\pi_{2,1}(\phi_1|\phi_0=1)$$

and

$$\pi_{1,1}(\phi_0)=\pi_{2,1}(\phi_0|\phi_1=0),$$

respectively (extending $\pi_{2,1}$ by continuity on the set $\{\phi_0=1\}$). Moreover, if $\pi_{2,1}(\phi_0,\phi_1)$ is defined on the stationarity region of $M_{2,1}$, then the preceding conditional distributions are defined on the stationarity regions of the corresponding models (see the Appendix).

Second, to express explicitly the prior belief on the presence of a unit root, under $M_{2,1}$ we specify $\pi_{2,1}(\phi_0,\phi_1)$ by introducing a prior distribution for ϕ_0, $\pi_{2,1}(\phi_0)$, and a prior distribution for the remaining parameter conditionally on ϕ_0, $\pi_{2,1}(\phi_1|\phi_0)$; by doing so, we can compare the performance of the different prior distributions for ϕ_0 considered in the previous section. In particular, we assume $\pi_{2,1}(\phi_1|\phi_0)$ to be uniform on the interval $(-[(1+\phi_0)/2];1)$ and $\pi_{2,1}(\phi_0)$ to be either uniform, beta, or beta sharp on $(-3;1)$; the corresponding prior distributions under the remaining models are a uniform distribution on $(-1;1)$ for ϕ_1 under $M_{1,0}$, and either a uniform, a beta, or a beta sharp on $(-1;1)$ for ϕ_0 under $M_{1,1}$.

The results, corresponding to different sizes l of the training samples, are presented in Tables 6–8. In particular, assuming a uniform prior for ϕ_0, Table 6 illustrates the performance of our procedure with emphasis on the identification of the order of the autoregressive process; as expected, for all p the uncertainty between the stationary model that generated the data and the corresponding integrated one increases as ϕ_0 approaches 1. The results in Tables 7 and 8 emphasize the effects of the different prior distributions for ϕ_0 on the performance of the proposed method in detecting unit root models. These results are in accordance with those observed in Section 5.

7. DISCUSSION

In this paper we derived a procedure for the identification of an autoregressive model for an observed time series and the detection of a unit root in its characteristic polynomial, which is *robust* with respect to the choice of the prior distribution for the autoregressive coefficients. The procedure is based

https://doi.org/10.1017/S0266466607070193 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466607070193 Published online by Cambridge University Press

Table 6. Proportion of the times that the different models would have been chosen for series of 100 observations ($l = 1$, 10, 20)

		$M_{0,0}$			$M_{1,1}$			$M_{1,0}$			$M_{2,1}$		
ϕ_1	ϕ_0	$l = 1$	$l = 10$	$l = 20$	$l = 1$	$l = 10$	$l = 20$	$l = 1$	$l = 10$	$l = 20$	$l = 1$	$l = 10$	$l = 20$
0.5	0.5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	1.000	1.000
0.5	0.85	0.005	0.000	0.000	0.000	0.001	0.001	0.394	0.109	0.096	0.601	0.890	0.903
0.5	0.95	0.003	0.001	0.000	0.000	0.000	0.000	0.972	0.809	0.732	0.025	0.190	0.268
0.5	1	0.002	0.001	0.000	0.000	0.000	0.000	0.990	0.922	0.879	0.008	0.077	0.121
0	0.5	0.000	0.000	0.000	0.976	0.902	0.862	0.001	0.000	0.001	0.023	0.098	0.137
0	0.85	0.686	0.447	0.380	0.257	0.442	0.460	0.050	0.070	0.090	0.007	0.041	0.070
0	0.95	0.837	0.709	0.605	0.021	0.061	0.100	0.142	0.224	0.280	0.000	0.006	0.015
0	1	0.939	0.826	0.713	0.018	0.077	0.125	0.042	0.091	0.141	0.001	0.006	0.021

TABLE 7. Proportion of the times that the ARMA(1,0) model would have been chosen for series of 100 observations ($l = 1, 10, 20$)

		Prior for ϕ_0								
		Uniform			Beta			Beta sharp		
ϕ_1	ϕ_0	$l = 1$	$l = 10$	$l = 20$	$l = 1$	$l = 10$	$l = 20$	$l = 1$	$l = 10$	$l = 20$
0.5	1	0.990	0.922	0.879	0.878	0.797	0.709	0.737	0.695	0.669
0.5	0.95	0.972	0.809	0.732	0.794	0.669	0.619	0.554	0.531	0.528
0.5	0.85	0.394	0.109	0.096	0.085	0.063	0.063	0.007	0.008	0.008

on the generalized fractional Bayes factor, which has been shown to be less sensitive to prior distributions than the standard Bayes factor. Moreover, we pointed out that it is possible to assign coherently the prior distributions on the parameters of the various models under comparison by introducing a suitable parametrization.

Numerical simulations show that as the size of the training sample increases the differences in the behavior of the various prior distributions are less relevant. Moreover, we have shown that by increasing the size of the training sample the uniform prior performs well also when the true process is stationary, whereas no improvements were observed in the behavior of the sharp priors when the data are generated from a nonstationary process. And even when the sharp priors choose the right model, as in the stationary case, they hardly provide strong evidence as the uniform does, so that our analysis clearly favors flat priors with respect to sharp priors.

Finally note that a possible generalization of the models given in Section 2 can be obtained including a deterministic time trend in (1):

$$\varphi_p(B)(Y_t - \mu - \delta t) = \eta_t.$$

TABLE 8. Proportion of the times that the ARMA(0,0) model would have been chosen for series of 100 observations ($l = 1, 10, 20$)

		Prior for ϕ_0								
		Uniform			Beta			Beta sharp		
ϕ_1	ϕ_0	$l = 1$	$l = 10$	$l = 20$	$l = 1$	$l = 10$	$l = 20$	$l = 1$	$l = 10$	$l = 20$
0	1	0.939	0.826	0.713	0.802	0.727	0.632	0.549	0.506	0.461
0	0.95	0.837	0.709	0.605	0.769	0.649	0.559	0.590	0.529	0.465
0	0.85	0.686	0.447	0.380	0.372	0.319	0.296	0.215	0.201	0.205

https://doi.org/10.1017/S0266466607070193 Published online by Cambridge University Press

If we let $y_t^* = y_t - \delta t$ and $z_t^* = y_{t+1}^* - y_t^*$, the series $\{y_t^*\}$ follows an autoregressive process of order p. Then if the generating process is stationary $\{z_t^*\}$ is a stationary ARMA$(p,1)$ process, whereas in the unit root case $\{z_t^*\}$ is a stationary ARMA$(p-1,0)$ process. For example, if $p = 1$, because $z_t^* = z_t - \delta$, we compare the ARMA(1,1) model $z_t - \delta = \phi_0(z_{t-1} - \delta) + \eta_t - \eta_{t-1}$ with the ARMA(0,0) model $z_t = \delta + \eta_t$, where the mean δ is a common nuisance parameter. Using a flat prior for δ (Schotman and Van Dijk, 1991b; Lubrano, 1995) the marginal distributions corresponding to the full likelihoods and to the geometric training likelihoods under both models can be easily obtained substituting z_t with $z_t - \delta$ in the expressions of $m_{p,1}(\mathbf{z})$, $m_{p,1}^l(\mathbf{z})$, $m_{p-1,0}(\mathbf{z})$, and $m_{p-1,0}^l(\mathbf{z})$ given in Section 4 and integrating out δ.

REFERENCES

Barbieri, M. & C. Conigliani (1998) Bayesian analysis of autoregressive time series with change points. *Journal of the Italian Statistical Society* 7, 243–255.

Barbieri, M. & C. Conigliani (2000) Fractional Bayes factors for the analysis of autoregressive models with possible unit roots. In *Proceedings of the XL Riunione Scientifica della Società Italiana di Statistica*, pp. 137–140. ISTAT Centro Stampa Roma.

Berger, J.O. & L.R. Pericchi (1996) The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association* 91, 109–122.

Berger, J.O. & R.-Y. Yang (1994) Noninformative priors and Bayesian testing for the AR(1) model. *Econometric Theory* 10, 461–482.

Box, G.E.P., G.M. Jenkins, & G. Reinsel (1994) *Time Series Analysis, Forecasting and Control*, 3rd ed. Prentice-Hall.

Conigliani, C. & A. O'Hagan (2000) Sensitivity of the fractional Bayes factor to prior distributions. *Canadian Journal of Statistics* 28, 343–352.

De Santis, F. & F. Spezzaferri (1997) Alternative Bayes factors for model selection. *Canadian Journal of Statistics* 25, 503–515.

De Santis, F. & F. Spezzaferri (1999) Methods for default and robust Bayesian model comparison: The fractional Bayes factor approach. *International Statistical Review* 67, 267–286.

De Santis, F. & F. Spezzaferri (2001) Consistent fractional Bayes factor for nested normal linear models. *Journal of Statistical Planning and Inference* 97, 305–321.

Galbraith, R.F. & J.I. Galbraith (1974) On the inverses of some patterned matrices arising in the theory of stationary time series. *Journal of Applied Probability* 11, 63–71.

Lubrano, M. (1995) Testing for unit roots in a Bayesian framework. *Journal of Econometrics* 69, 81–109.

Marinucci, D. & L. Petrella (1999) Noninformative Bayesian analysis for stationary and nonstationary AR(1) time series. In J.M. Bernardo, J.O. Berger, A.P. Dawid, & A.F.M. Smith (eds.), *Bayesian Statistics 6*, pp. 821–828. Oxford University Press.

Marriott, J. & P. Newbold (1998) Bayesian comparison of ARIMA and stationary ARMA models. *International Statistical Review* 66, 323–336.

Monahan, J.F. (1984) A note on enforcing stationarity in autoregressive-moving average models. *Biometrika* 71, 403–404.

O'Hagan, A. (1991) Discussion of *Posterior Bayes Factors* by M. Aitkin. *Journal of the Royal Statistical Society, Series B* 53, 136.

O'Hagan, A. (1995) Fractional Bayes factor for model comparison (with discussion). *Journal of the Royal Statistical Society, Series B* 57, 99–138.

O'Hagan, A. (1997) Properties of intrinsic and fractional Bayes factors. *Test* 6, 101–118.

Pantula, S.G., G. Gonzales-Farias, & W.A. Fuller (1994) A comparison of unit-root test criteria. *Journal of Business & Economic Statistics* 12, 449–459.

https://doi.org/10.1017/S0266466607070193 Published online by Cambridge University Press

Phillips, P.C.B. (1991a) To criticize the critics: An objective Bayesian analysis of stochastic trends. *Journal of Applied Econometrics* 6, 333–364.
Phillips, P.C.B. (1991b) Bayesian routes and unit roots: De rebus prioribus semper est disputandum. *Journal of Applied Econometrics* 6, 435–474.
Phillips, P.C.B. & W. Ploberger (1994) Posterior odds testing for a unit root with data-based model selection. *Econometric Theory* 10, 774–808.
Phillips, P.C.B. & Z. Xiao (1998) A primer on unit root testing. *Journal of Economic Surveys* 12, 423–469.
Schotman, P.C. & H.K. Van Dijk (1991a) A Bayesian analysis of the unit root hypothesis in real exchange rate. *Journal of Econometrics* 49, 195–238.
Schotman, P.C. & H.K. Van Dijk (1991b) On Bayesian routes to unit roots. *Journal of Applied Econometrics* 6, 387–401.
Sims, C. (1988) Bayesian skepticism on unit root econometrics. *Journal of Economic Dynamics and Control* 12, 463–474.
Sims, C. & H. Uhlig (1991) Understanding unit rooters: A helicopter tour. *Econometrica* 59, 1591–1599.
Uhlig, H. (1994a) On Jeffreys prior when using the exact likelihood function. *Econometric Theory* 10, 633–644.
Uhlig, H. (1994b) What macroeconomists should know about unit roots. *Econometric Theory* 10, 645–671.
Varshavsky, J.A. (1996) Intrinsic Bayes factors for model selection with autoregressive data. In J.M. Bernardo, J.O. Berger, A.P. Dawid, & A.F.M. Smith (eds.), *Bayesian Statistics 5*, pp. 757–763. Oxford University Press.
Zivot, E. (1994) A Bayesian analysis of the unit root hypothesis within an unobserved components model. *Econometric Theory* 10, 552–578.

APPENDIX A: Stationarity Regions

In this Appendix, we characterize the stationarity regions of the autoregressive polynomials $\phi_p^0(B)$ defined in (4), and we show that the conditional distributions introduced in Section 4 as priors for the parameters of the models $M_{\bar{p}-h,1}$ and $M_{\bar{p}-h,0}$, $h = 1, \ldots, \bar{p} - 1$, are defined on the appropriate stationarity regions. Let C_p and C_p^0 be the stationarity regions of the autoregressive polynomials $\phi_p(B)$ and $\phi_p^0(B)$ under $M_{p,0}$ and $M_{p,1}$, respectively, and recall the following result by Monahan (1984).

THEOREM 1 (Monahan, 1984). $(\phi_1, \ldots, \phi_p) \in C_p$ $(p > 1)$ *if and only if*

$$|\phi_p^p| < 1 \quad \textit{and} \quad (\phi_1^{p-1}, \ldots, \phi_{p-1}^{p-1}) \in C_{p-1},$$

where

$$\begin{cases} \phi_i^{p-1} = \dfrac{\phi_i^p + \phi_p^p \phi_{p-i}^p}{1 - (\phi_p^p)^2} & i = 1, \ldots, p-1 \\ \phi_i^p = \phi_i & i = 1, \ldots, p. \end{cases}$$

Now, a similar result also can be obtained for C_p^0. Recall that the one-to-one transformation that reparametrizes $\boldsymbol{\varphi}^{(p)}$ in terms of the partial autocorrelations $r_1, \ldots, r_p$ of the autoregressive process of order p is

https://doi.org/10.1017/S0266466607070193 Published online by Cambridge University Press

$$\begin{cases} \varphi_k^k = r_k & k = 1, \ldots, p; \quad p \geq 1 \\ \varphi_i^k = \varphi_i^{k-1} - r_k \varphi_{k-i}^{k-1} & 1 \leq i < k \leq p; \quad p \geq 2 \end{cases}$$

where

$$\varphi_i = \varphi_i^p, \qquad i = 1, \ldots, p; \qquad p \geq 1.$$

It follows that also the coefficients $\phi_0, \phi_1, \ldots, \phi_{p-1}$ of $\phi_p^0(B)$ can be written in terms of partial autocorrelations:

$$\begin{cases} \phi_0^0 = r_1 & p = 1 \\ \phi_k^k = -r_{k+1} & k = 0, \ldots, p-1; \quad p > 1 \\ \phi_i^k = \phi_i^{k-1} + r_{k+1}(\phi_0^{k-1} + \phi_{k-i}^{k-1} - 1) & 1 \leq i < k \leq p-1; \quad p > 2 \\ \phi_0^k = r_{k+1} + \phi_0^{k-1}(1 - r_{k+1}) & 2 \leq k \leq p-1; \quad p > 2 \\ \phi_0^1 = r_2 - \phi_0^0(1 - r_2) & \\ \phi_i = \phi_i^{p-1} & i = 0, \ldots, p-1; \quad p \geq 1 \end{cases} \tag{A.1}$$

and by inverting (A.1) we obtain the analogue of Monahan's result.

THEOREM 2. $(\phi_0, \phi_1, \ldots, \phi_{p-1}) \in C_p^0$ $(p > 1)$ *if and only if*

$$|\phi_{p-1}^{p-1}| < 1 \quad \textit{and} \quad (\phi_0^{p-2}, \phi_1^{p-2}, \ldots, \phi_{p-2}^{p-2}) \in C_{p-1}^0,$$

where

$$\begin{cases} (i) \quad \phi_i^{p-2} = \dfrac{\phi_i^{p-1} + \phi_{p-1}^{p-1}(\phi_0^{p-1} + \phi_{p-i-1}^{p-1} - 1)}{1 - (\phi_{p-1}^{p-1})^2} & i = 1, \ldots, p-2; \quad p > 2 \\ (ii) \quad \phi_0^{p-2} = \dfrac{\phi_0^{p-1} + \phi_{p-1}^{p-1}}{1 + \phi_{p-1}^{p-1}} & p > 2 \\ (iii) \quad \phi_0^0 = -\dfrac{\phi_0^1 + \phi_1^1}{1 + \phi_1^1} & p = 2 \\ (iv) \quad \phi_i^{p-1} = \phi_i & i = 0, \ldots, p-1. \end{cases}$$

We can now prove the following two results concerning the behavior of the conditional distributions $\pi_{p,1}(\phi_1, \ldots, \phi_{p-1} | \phi_0 = 1)$ and $\pi_{p,1}(\phi_0, \phi_1 \ldots, \phi_{p-2} | \phi_{p-1} = 0)$ in terms of stationarity conditions.

THEOREM 3. *Let* $\pi_{p,1}(\phi_0, \ldots, \phi_{p-1})$ *be defined on the stationarity region of the autoregressive component of* $M_{p,1}$ $(p > 1)$. *Then* $\pi_{p,1}(\phi_1, \ldots, \phi_{p-1} | \phi_0 = 1)$ *is defined on the stationarity region of the autoregressive component of* $M_{p-1,0}$*:*

$$\{(\phi_1, \ldots, \phi_{p-1}) : (1, \phi_1, \ldots, \phi_{p-1}) \in C_p^0 \cup \{\phi_0 = 1\}\} \equiv C_{p-1}.$$

Proof. The result is clearly verified for $p = 2$. We now prove it for $p > 2$, assuming that it is true for $p - 1$. With $\phi_0 = \phi_0^{p-1} = 1$, from Theorem 2 we have

https://doi.org/10.1017/S0266466607070193 Published online by Cambridge University Press

(a) $\phi_i^{p-2} = \dfrac{\phi_i^{p-1} + \phi_{p-1}^{p-1}\phi_{p-i-1}^{p-1}}{1 - (\phi_{p-1}^{p-1})^2}$ $i = 1, \ldots, p-2$
(b) $\phi_0^{p-2} = 1$
(c) $(1, \phi_1^{p-2}, \ldots, \phi_{p-2}^{p-2}) \in C_{p-1}^0 \cup \{\phi_0^{p-2} = 1\}$.

It follows from (c) that $(\phi_1^{p-2}, \ldots, \phi_{p-2}^{p-2}) \in C_{p-2}$, so that, from Monahan (1984), $(\phi_1, \ldots, \phi_{p-1}) \in C_{p-1}$. ■

THEOREM 4. *Let $\pi_{p,1}(\phi_0, \ldots, \phi_{p-1})$ be defined on the stationarity region of the autoregressive component of $M_{p,1}$ ($p > 1$). Then $\pi_{p,1}(\phi_0, \ldots, \phi_{p-2} | \phi_{p-1} = 0)$ is defined on the stationarity region of the autoregressive component of $M_{p-1,1}$:*

$$\{(\phi_0, \ldots, \phi_{p-2}) : (\phi_0, \ldots, \phi_{p-2}, 0) \in C_p^0\} \equiv C_{p-1}^0.$$

Proof. With $\phi_{p-1} = \phi_{p-1}^{p-1} = 0$, from Theorem 2 it follows that

$$\phi_i^{p-2} = \phi_i^{p-1} = \phi_i, \qquad i = 0, \ldots, p-2,$$

hence the thesis. ■

By applying Theorems 3 and 4 recursively, the preceding results also hold for the conditional distributions that define the priors of the parameters of each model in (5).

https://doi.org/10.1017/S0266466607070193 Published online by Cambridge University Press