AN ALTERNATIVE BAYES FACTOR FOR TESTING FOR UNIT AUTOREGRESSIVE ROOTS

Caterina Conigliani and Fulvio Spezzaferri
Working Paper n° 31 2002

Comitato Scientifico

Marisa Cenci
Andrea Gheno
Valentina Guizzi
Julia Mortera
Carlo D. Mottura
Luciano Pieraccini
Paola Vicard
- I “Working Papers” del Dipartimento di Economia svolgono la funzione di divulgare tempestivamente, in forma definitiva o provvisoria, i risultati di ricerche scientifiche originali. La loro pubblicazione è soggetta all’approvazione del Comitato Scientifico.


- Copie della presente pubblicazione possono essere richieste alla Redazione.

REDAZIONE:
Dipartimento di Economia
Università degli Studi di Roma Tre
Via Ostiense, 139 - 00154 Roma
Tel. 0039-6-57374003 fax 0039-6-57374093
E-mail: dip_eco@uniroma3.it
To Marco
AN ALTERNATIVE BAYES FACTOR FOR TESTING FOR UNIT AUTOREGRESSIVE ROOTS*

Caterina Conigliani** and Fulvio Spezzaferri***

* Research partly supported by MURST, Italy.
** Università Roma Tre, Italy.
*** Università di Roma “La Sapienza”, Italy.
<table>
<thead>
<tr>
<th>CONTENTS</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>2. Models and notation</td>
<td>2</td>
</tr>
<tr>
<td>3. Bayes factors and alternative Bayes factors</td>
<td>4</td>
</tr>
<tr>
<td>4. The choice of the prior distributions</td>
<td>8</td>
</tr>
<tr>
<td>5. The geometric partial Bayes factor for testing</td>
<td>15</td>
</tr>
<tr>
<td>For unit autoregressive roots factors</td>
<td></td>
</tr>
<tr>
<td>5.1 Computing the marginal distributions corresponding to the full likelihood</td>
<td>15</td>
</tr>
<tr>
<td>5.2 Computing the marginal distributions corresponding to the geometric training likelihoods</td>
<td>18</td>
</tr>
<tr>
<td>6. Testing for a unit root in a first order autoregressive process</td>
<td>21</td>
</tr>
<tr>
<td>6.1 Computing the marginal distributions under the ARMA(1,1) model</td>
<td>22</td>
</tr>
<tr>
<td>6.2 A simulation experiment</td>
<td>24</td>
</tr>
<tr>
<td>References</td>
<td>28</td>
</tr>
</tbody>
</table>
Abstract. 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.

Key words: Autoregressive model, Bayes factors, model selection, noninformative prior distributions, time series, unit root.

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 (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. In particular, great attention in the literature has been devoted to the nature of suitable noninformative priors for the autoregressive coefficients. Just to mention a few examples: Sims (1988) and Sims and Uhlig (1991) advocate the use of flat priors; Phillips (1991) finds that flat priors bias the inference towards stationary models, and suggests instead the use of Jeffreys priors; for the $AR(1)$ model, Berger and Yang (1994) consider a reference prior approach; Marinucci and Petrella (1999), again only for the $AR(1)$ model, derive a noninformative prior by considering the functional relationship between the autoregressive parameter and the variance of the process. It is particularly interesting to notice 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 for instance 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 non-stationary.

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 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. The models and notation are introduced in Section 2, while 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 informations and to alternative Bayes factors. Section 4 deals with the choice of the prior distributions in this setting. The maginal distributions for computing Bayes factors and posterior odds are derived in Section 5 for the general case, and in Section 6 for the simple problem of testing for a unit root in a first order autoregressive process.

2 Models and notation

Let \( \{Y_t; t \in T\} \) 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

\[
\phi_p(B)(Y_t - \mu) = \eta_t
\]

where \( \phi_p(B) = 1 - \phi_1 B - \ldots - \phi_p B^p \), \( B \) is the backshift operator such that \( B^k y_t = y_{t-k} \), and \( \{\eta_t\} \) is a Gaussian white noise process with variance \( \tau^2 \). The stationarity condition constrains the parameter vector \( \phi^{(p)} = (\phi_1, \phi_2, \ldots, \phi_p) \) to lie in the stationarity region \( \{\phi^{(p)} : \phi_p(z) = 0 \text{ implies } |z| > 1\} \).
Suppose we observe the time series $y = (y_1, \ldots, y_n)$, a finite realization of an autoregressive process of order $p$, with $p$ unknown; the goal is the identification of $p$, and the contemporary 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, for a given order $p$, we can write

$$\varphi_p(B) = \phi_{p-1}(B)(1 - B),$$

where $\phi_{p-1}(B)$ is a stationary autoregressive polynomial of order $p - 1$ with parameter vector $\phi^{(p-1)} = (\phi_1, \phi_2, \ldots, \phi_{p-1})$. Then letting $Z_t = Y_{t+1} - Y_t$, from (1) we can write

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

so that $\{Z_t\}$ is a stationary $ARMA(p - 1, 0)$ process.

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 write also 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 $ARMA(p, 1)$, with mean $E(Z_t) = 0$, autoregressive polynomial $\varphi_p(B)$, and moving average coefficient equal to 1:

$$\varphi_p(B)Z_t = (1 - B)\eta_t.$$  \hspace{1cm} (3)

Now for a given order $p$, the test for a unit root can be performed by comparing the $ARMA(p, 1)$ model (3) and the $ARMA(p - 1, 0)$ 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 $y = (y_1, \ldots, y_n)$ and testing for a unit root can be seen as the one of choosing a model for $z = (z_1, \ldots, z_{n-1})$ in the set

$$M_z = \{ARMA(p, 1), ARMA(p - 1, 0) \mid p \in \mathcal{P}\},$$  \hspace{1cm} (4)

where $\mathcal{P}$ is a set of possible values for $p$. 

3
3 Bayes factors and alternative Bayes factors

The choice of the model that better fit 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 $M_j$, for a set of observations $y = (y_1, ..., y_n)$, and let $f_r(y \mid \theta_r)$ and $\pi_r(\theta_r)$ be respectively the distribution of the data and the prior distribution of the parameters $\theta_r$ under model $M_r$ ($r = i, j$). The Bayes factor for $M_i$ against $M_j$ is defined as

$$B_{ij}(y) = \frac{\int f_i(y \mid \theta_i) \pi_i(\theta_i) \, d\theta_i}{\int f_j(y \mid \theta_j) \pi_j(\theta_j) \, d\theta_j},$$

and represents the weight of evidence in the data in favour of $M_i$ against $M_j$. If a priori we assume that $P(M_i) = P(M_j) = 1/2$, the posterior probabilities of $M_i$ and $M_j$ are then

$$P(M_i \mid y) = \frac{B_{ij}(y)}{1 + B_{ij}(y)},$$

and

$$P(M_j \mid y) = \frac{1}{1 + B_{ij}(y)},$$

respectively.

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 penalised 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, $y = (y(l), y(n-l))$, of size $l$ and $n-l$ respectively, with $0 < l < n$. First, subsample $y(l)$ is used to obtain the posterior distributions $\pi_r(\theta_r \mid y(l))$; in the second step, taking
these as prior distributions, the remaining data $y(n-l)$ are used to compute a Bayes factor:

$$B_{ij}(y(n-l) \mid y(l)) = \frac{\int f_i(y(n-l) \mid \theta_i) \pi_i(\theta_i \mid y(l)) d\theta_i}{\int f_j(y(n-l) \mid \theta_j) \pi_j(\theta_j \mid y(l)) d\theta_j} = \frac{\int f_i(y \mid \theta_i) \pi_i(\theta_i) d\theta_i}{\int f_i(y(l) \mid \theta_i) \pi_i(\theta_i) d\theta_i} \left/ \frac{\int f_j(y \mid \theta_j) \pi_j(\theta_j) d\theta_j}{\int f_j(y(l) \mid \theta_j) \pi_j(\theta_j) d\theta_j} \right..$$

(5)

As widely 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 $y(l)$ is such that the updated prior distributions are proper, they 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 $B_{ij}(y)$ 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 $y(l)$ from the data.

To avoid the arbitrariness of choosing a particular training sample, O'Hagan (1991, 1995) suggested instead the use of a proportion $b$ of the data for training. Formally, suppose $y$ is a sequence of i.i.d. data, and let $b = l/n$. If both $l$ and $n$ are large, the likelihood $f_r(y(l) \mid \theta_r)$ based only on the training sample $y(l)$ will approximate to the full likelihood $f_r(y \mid \theta_r)$ raised to the power of $b$:

$$f_r(y(l) \mid \theta_r) \approx f_r(y \mid \theta_r)^b,$$

(6) 

($r = i, j$). The above approximation lead to the definition of the fractional Bayes factor.

A different proposal to remove the dependence of the partial Bayes factor on a specific training sample was suggested by De Santis and Spezzaferri (1997, 2001). They noted that one could replace the likelihood $f_r(y(l) \mid \theta_r)$ in (5) with the geometric training likelihood $f_r^l(y \mid \theta_r)$, defined as the geometric mean of the $f_r(y(l) \mid \theta_r)$ as $y(l)$ varies over the set of all possible proper training samples of a given size $l$ ($r = i, j$). Clearly, for i.i.d.
data the geometric training likelihood over the set of training samples of size \( l \) is equal to the fractional likelihood \( f_r \left( y | \theta_r \right)^b \) with \( b = l/n \), so that the proposal by De Santis and Spezzaferri leads to O’Hagan’s fractional Bayes factor. But in more general situations, the geometric training likelihood can be used to introduce a new tool for Bayesian model choice with weak prior information, that we can refer to as the \textit{geometric partial Bayes factor}.

Now consider the problem of the identification of an autoregressive model for an observed time series and the detection of a unit root in its characteristic polynomial; as we discussed earlier, this problem can be seen as the one of choosing a model in the set (4) for the differenced time series \( z = (z_1, \ldots, z_{n-1}) \).

In order to simplify the notations, let \( M_{p,q} \) denote the stationary \( ARMA \left( p, q \right) \) model for the first differences. Moreover, let \( f_{p,q} \left( z \mid \cdot \right) \) be the distribution of the data, \( \theta_{p,q} \) be the vector of parameters, and \( \pi_{p,q} (\theta_{p,q}) \) be the the prior distribution of the parameters under model \( M_{p,q} \left( p \in \mathcal{P}; q = 0, 1 \right) \); clearly, from Section 2, is

\[
\theta_{p,1} = \begin{cases} (\tau^2, \varphi(p)) & p > 0 \\ \tau^2 & p = 0 \end{cases}
\]

and

\[
\theta_{p,0} = \begin{cases} (\tau^2, \phi(p)) & p > 0 \\ \tau^2 & p = 0, \end{cases}
\]

and both under \( M_{p,1} \) and \( M_{p,0} \), we assume the autoregressive coefficients and the variance \( \tau^2 \) of the white noise process \( \{ \eta_t \} \) to be \textit{a priori} independent.

It is important to note that, because of the stationarity constraints, the choice of the prior distribution for the autoregressive parameters both under \( M_{p,1} \) and \( M_{p,0} \) does not represent a problem for the evaluation of Bayes factors and of posterior odds in this setting. 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 \( \tau^2 \) is improper, the Bayes factor for comparing any two models in (4) is undefined. Now, it may be argued that since \( \tau^2 \) appears in \( M_{p,1} \) and \( M_{p,0} \) for any \( p \in \mathcal{P} \), the (unknown) multiplicative constants in the Bayes factor would cancel; however, as pointed out in Berger and Pericchi (1996), this is a dangerous practice since 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 $\tau^2$ is improper, it was shown by Marriott and Newbold (1998) that the standard Bayesian analysis is extremely non robust with respect to the choice of the prior distribution for the autoregressive coefficients.

Thus, for the problem of choosing a model in (4), we will adopt a robust approach and focus our attention on alternative Bayes factors. In particular, since it was shown in Barbieri and Conigliani (2000) that in the case of dependent data (6) does not hold, so that the use of fractional Bayes factors is not justified in this setting, we will employ the geometric partial Bayes factor.

Formally, let $f_{p,q}^l (z | \theta_{p,q})$ be the the geometric training likelihood under $M_{p,q}$ corresponding to a specific size $l$ of the training samples. According to (5), the geometric partial Bayes factor for comparing any two models in (4), say $M_{i,\delta}$ and $M_{j,\gamma}$, is

$$B_{(i, \delta)(j, \gamma)}^l (z) = \frac{m_{i,\delta}^l (z)}{m_{i,\delta}^l (z)} \frac{m_{j,\gamma}^l (z)}{m_{j,\gamma}^l (z)}$$

(i, j \in \mathcal{P}; \delta, \gamma = 0, 1), where

$$m_{p,q}^l (z) = \int f_{p,q}^l (z | \theta_{p,q}) \pi_{p,q} (\theta_{p,q}) d\theta_{p,q}$$

and

$$m_{p,q} (z) = \int f_{p,q} (z | \theta_{p,q}) \pi_{p,q} (\theta_{p,q}) d\theta_{p,q}$$

are the marginal distributions under $M_{p,q}$ corresponding to the full likelihood and to the geometric training likelihood respectively ($p \in \mathcal{P}; q = 0, 1$).

One last issue is worth a few considerations. The choice of the training sample size, $l$ in our notation, has been widely discussed in the literature; see for instance O'Hagan (1995, 1997). One simple and obvious guidance, that has proved to be reliable in a range of problems involving improper priors, is to set $l$ equal to the size of the minimal training sample. On the other hand, it is interesting to notice 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 $l$. Here, if we
consider an improper prior for $\tau^2$, a minimal training sample is any subset of $z$ made of a single observation $z_t$ ($t = 1, 2, ..., n - 1$). Being a *proper* training sample, each $z_t$ carries enough informations to update the prior distribution for $\tau^2$ 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, in order 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 The choice of the prior distributions

The evaluation of the marginal densities (8) and (9) clearly requires the full specification of the various models in (4), and in particular the specification of the prior distributions for the parameters under the different models. However, in general the integrals with respect to the autoregressive coefficients, assuming either informative or noninformative priors for these parameters, cannot be solved analytically and numerical procedures need to be employed. Thus, for the moment we do not specify a particular form for $\pi_{p,1} (\varphi^{(p)})$ and $\pi_{p,0} (\phi^{(p)})$, while for $\tau^2$ under $M_{p,q}$ we consider the standard noninformative prior:

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

($p \in P; q = 0, 1$). However, there is one issue that is worth underlying, since it turns out to be crucial when actually choosing the prior distributions for the autoregressive parameters in this problem, namely the fact that all the models in (4) are nested models.

In order to see this, for a given order $p$ consider the *ARMA* ($p, 1$) process in (3), depending on $\varphi^{(p)} = (\varphi_1, \varphi_2, ..., \varphi_p)$ through the autoregressive polynomial $\varphi_p (B)$; moreover, consider the following reparametrization in terms of the vector $\phi^{(p)}_0 = (\phi_0, \phi_1, ..., \phi_{p-1})$:
\[
\begin{align*}
\varphi_1 &= \varphi_0 & p &= 1 \\
\varphi_i &= \varphi_0 + \varphi_1 & p &> 1 \\
\varphi_p &= -\varphi_{p-1} & p &> 1 \\
\varphi_j &= \varphi_j - \varphi_{j-1} & j &= 2, \ldots, p - 1, \quad p > 2
\end{align*}
\]

that is
\[
\begin{align*}
\phi_0 &= \varphi_1 & p &= 1 \\
\phi_0 &= \sum_{i=1}^{p} \varphi_i & p &> 1 \\
\phi_j &= -\sum_{h=j+1}^{p} \varphi_h & j &= 1, \ldots, p - 1.
\end{align*}
\]

Then \(M_{p,1}\) can be rewritten as
\[
\phi_p^0(B)Z_t = (1 - B)\eta_t,
\]

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

and it is straightforward to see that all models \(M_{p-h,q}\), with \(p > 1\), \(q = 0, 1\) and \(h = 1, 2, \ldots, p - 1\), are nested in \(M_{p,1}\). In fact, as it is summarised in Figure 1, for any \(1 \leq h \leq p - 1\), \(M_{p-h,1}\) can be obtained from \(M_{p,1}\) simply by letting \(\phi_{p-1} = \ldots = \phi_{p-h+1} = \phi_{p-h} = 0\); moreover, \(M_{p-1,0}\) can be obtained from \(M_{p,1}\) by letting \(\phi_0 = 1\), while for \(1 < h \leq p - 1\), \(M_{p-h,0}\) can be obtained from \(M_{p,1}\) by letting \(\phi_{p-1} = \ldots = \phi_{p-h+1} = 0\) and \(\phi_0 = 1\). It follows that once \(\pi_{p,1}(\phi_0, \phi_1, \ldots, \phi_{p-1})\) is introduced under \(M_{p,1}\) (\(p > 1\)), extending \(\pi_{p,1}\) by continuity on the set \(\{\phi_0 = 1\}\), the priors to be considered under the remaining models are the conditional distributions:
\[
\pi_{p-h,1}(\phi_0, \ldots, \phi_{p-h-1}) = \pi_{p,1}(\phi_0, \ldots, \phi_{p-h-1}|\phi_{p-1} = \ldots = \phi_{p-h} = 0)
\]
\((1 \leq h \leq p - 1),\)
\[
\pi_{p-1,0}(\phi_1, \ldots, \phi_{p-1}) = \pi_{p,1}(\phi_1, \ldots, \phi_{p-1}|\phi_0 = 1)
\]

and
\[
\pi_{p-h,0}(\phi_1, \ldots, \phi_{p-h}) = \pi_{p,1}(\phi_1, \ldots, \phi_{p-h}|\phi_{p-1} = \ldots = \phi_{p-h+1} = 0, \phi_0 = 1)
\]
\((1 < h \leq p - 1).\)
Moreover, if $\pi_{p,1}(\phi_0, \phi_1, \ldots, \phi_{p-1})$ is defined on the stationarity region of $M_{p,1}$, then the above conditional distributions are defined on the stationarity regions of the corresponding models. In order to see this, 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

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

where

$$
\begin{align*}
\phi_i^{p-1} &= \frac{\phi_i^p + \phi_p^p \phi_{p-i}^p}{1 - (\phi_p^p)^2} & i = 1, \ldots, p - 1 \\
\phi_i^p &= \phi_i & i = 1, \ldots, p.
\end{align*}
$$
Now, a similar result can be obtained also for $C_p^0$. Recall that the one-to-one transformation which reparametrizes $\varphi^{(p)}$ in terms of the partial autocorrelations $r_1, ..., r_p$ of the autoregressive process of order $p$ is:

\[
\begin{align*}
\varphi_k & = r_k, & k = 1, ..., p; & p \geq 1 \\
\varphi_i & = \varphi_i^{k-1} - r_k \varphi_{k-i}^{k-1}, & 1 \leq i < k \leq p; & p \geq 2
\end{align*}
\]

where

\[
\varphi_i = \varphi_i^p, \quad i = 1, ..., p; \quad p \geq 1.
\]

It follows, from (11), that also the coefficients $\phi_0, \phi_1, ..., \phi_{p-1}$ of $\phi_p^0(B)$ can be written in terms of partial autocorrelations:

\[
\begin{align*}
\phi_0^0 & = r_1 \\
\phi_k^0 & = -r_{k+1} \\
\phi_i^k & = \phi_i^{k-1} + r_{k+1} (\phi_0^{k-1} + \phi_{k-i}^{k-1} - 1) & 1 \leq i < k \leq p - 1; & p > 1 \\
\phi_0^k & = r_{k+1} + \phi_0^{k-1} (1 - r_{k+1}) & 2 \leq k \leq p - 1; & p > 2 \\
\phi_i^k & = r_2 - \phi_0^0 (1 - r_2) \\
\phi_i & = \phi_i^{p-1}
\end{align*}
\]

and by inverting (12) we obtain the analogous of Monahan’s result.

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

\[
|\phi_{p-1}^0| < 1 \quad \text{and} \quad \phi_0^{p-2}, \phi_1^{p-2}, ..., \phi_{p-2}^{p-2} \in C_{p-1}^0,
\]

where

\[
\begin{align*}
i) \quad \phi_i^{p-2} & = \frac{\phi_i^{p-1} + \phi_{p-1}^{p-1} (\phi_0^{p-1} + \phi_{p-1}^{p-1} - 1)}{1 - (\phi_{p-1}^{p-1})^2} & i = 1, ..., p - 2; & p > 2 \\
ii) \quad \phi_0^{p-2} & = \frac{\phi_0^{p-1} + \phi_{p-1}^{p-1}}{1 + \phi_{p-1}^{p-1}} & p > 2 \\
iii) \quad \phi_0^0 & = \frac{\phi_0^1 + \phi_1^1}{1 + \phi_1^1} & p = 2 \\
iv) \quad \phi_i^{p-1} & = \phi_i & i = 0, ..., p - 1.
\end{align*}
\]
Example 1. Let $p = 2$; we apply Theorem 2 in order to determine $C^0_2$. First note that from (12) we can write $\phi_0^{p-1} = 1 - \prod_{j=1}^p (1 - r_j)$, so that

$$1 - 2^p < \phi_0^{p-1} < 1.$$

(13)

It follows that here is $\phi_0 = \phi_1^0 \in (-3, 1)$. Moreover, $|\phi_0^0| < 1$ and from iii) is

$$\phi_0 = -\frac{\phi_0 + \phi_1}{1 + \phi_1},$$

so that

$$\phi_1 > -\frac{1 + \phi_0}{2}.$$

Being $|\phi_1| = |\phi_1^0| < 1$, the latter inequality implies that $C^0_2$ is identified by

$$\left\{ \begin{array}{l}
-3 < \phi_0 < 1 \\
-\frac{1 + \phi_0}{2} < \phi_1 < 1.
\end{array} \right.$$  

(14)

Note that (14) can equivalently be expressed by

$$\left\{ \begin{array}{l}
-2\phi_1 - 1 < \phi_0 < 1 \\
-1 < \phi_1 < 1
\end{array} \right.$$  

(15)

so that, by applying (10) to (15), we obtain

$$\left\{ \begin{array}{l}
\varphi_2 - 1 < \varphi_1 < 1 - \varphi_2 \\
-1 < \varphi_2 < 1
\end{array} \right.$$  

which is the usual stationarity region for an autoregressive process of order $p = 2$.

Example 2. Let $p = 3$; we apply Theorem 2 in order to determine $C^0_3$. First note that, from (13), is

$$\phi_0 = \phi_2^2 \in (-7, 1).$$

(16)

Moreover, from ii) is

$$\phi_0^3 = \frac{\phi_0 + \phi_2}{1 + \phi_2},$$

and because of (12) is $\phi_0^3 \in (-3, 1)$, so that

$$\phi_2 > -\frac{3 + \phi_0}{4}.$$
Being $|\phi_2| = |\bar{\phi}_2^2| < 1$, the latter inequality implies that

$$-\frac{3 + \phi_0}{4} < \phi_2 < 1. \quad (17)$$

Finally, from i) is

$$\phi_1 = \phi_1 + \phi_2 (\phi_0 + \phi_1 - 1) \frac{1 - (\phi_2)^2}{1 - (\phi_2)^2},$$

and because of iii) is $\phi_1^2 \in \left(-\frac{\phi_0^2 + 1}{2}, 1\right)$, so that

$$\phi_2 - \frac{1 + \phi_0}{2} < \phi_1 < 1 - \phi_2 + \frac{\phi_2 (1 - \phi_0)}{1 + \phi_2}. \quad (18)$$

The stationarity region of the ARMA $(3,1)$ process is identified by (16), (17), and (18).

We can now prove the following two results concerning the behaviour of the conditional distributions $\pi_{p-h,q} (q = 0, 1; h = 1, 2, ..., p - 1)$ in terms of stationarity conditions.

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

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

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

\begin{align*}
a) & \quad \phi_i^{p-2} = \frac{\phi_i^{p-1} + \phi_{p-1}^{p-1}\phi_{p-i-1}^{p-1}}{1 - (\phi_{p-1}^{p-1})^2}, \\ b) & \quad \phi_0^{p-2} = 1 \\ c) & \quad (1, \phi_1^{p-2}, ..., \phi_{p-2}^{p-2}) \in C_{p-1}^0 \cup \{\phi_0^{p-2} = 1\}.
\end{align*}

It follows from c) that $(\phi_1^{p-2}, ..., \phi_{p-2}^{p-2}) \in C_{p-2}$, so that, from Monahan (1984), is $(\phi_1, ..., \phi_{p-1}) \in C_{p-1}$. \(\square\)
Theorem 4. Let $\pi_{p,1}(\phi_0, ..., \phi_{p-1})$ be defined on the stationarity region of the autoregressive component of $M_{p,1}$ ($p > 1$). Then $\pi_{p-1,1}(\phi_0, ..., \phi_{p-2} | \phi_{p-1} = 0)$ is defined on the stationarity region of the autoregressive component of $M_{p-1,1}$:

$$\{(\phi_0, ..., \phi_{p-2}) : (\phi_0, ..., \phi_{p-2}, 0) \in C^0_p \} \equiv C^0_{p-1}$$

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

$$\phi_i^{p-2} = \phi_i^{p-1} = \phi_i, \quad i = 0, ..., p - 2,$$

hence the thesis. $\square$

Example 1 (continued). Let $\pi_{2,1}(\phi_0, \phi_1)$ be a prior distribution on $C^0_2$; recall that $(\phi_0, \phi_1) \in C^0_2$ implies (14) or equivalently (15).

First, we want to verify that the conditional distribution $\pi_{1,0}(\phi_1 | \phi_0 = 1)$ is defined on $C_1$. In fact, from (14) when $\phi_0 = 1$ is $-1 < \phi_1 < 1$, which is exactly the stationary region of the ARMA $(1, 0)$ process $(1 - \phi_1 B) Z_t = \eta_t$.

Now consider the conditional distribution $\pi_{1,1}(\phi_0 | \phi_1 = 0)$; we want to verify that it is defined on $C^0_1$. In fact, from (15) when $\phi_1 = 0$ is $-1 < \phi_0 < 1$, which is exactly the stationary region of the ARMA $(1, 1)$ process $(1 - \phi_0 B) Z_t = (1 - B) \eta_t$.

Example 2 (continued). Let $\pi_{3,1}(\phi_0, \phi_1, \phi_2)$ be a prior distribution on $C^0_3$; recall that $(\phi_0, \phi_1, \phi_2) \in C^0_3$ implies (16), (17) and (18).

First, we want to verify that the conditional distribution $\pi_{2,0}(\phi_1, \phi_2 | \phi_0 = 1)$ is defined on $C_2$. In fact when $\phi_0 = 1$ from (17) is $-1 < \phi_2 < 1$, and from (18) is $\phi_1 < 1$ and it is well known that these two inequalities identify the stationary region of the ARMA $(2, 0)$ process $(1 - \phi_1 B - \phi_2 B^2) Z_t = \eta_t$.

Now consider the conditional distribution $\pi_{2,1}(\phi_0, \phi_1 | \phi_2 = 0)$; we want to verify that it is defined on $C^0_2$. In fact, (16) and (17) can be equivalently written as

$$\begin{cases}
-1 < \phi_2 < 1 \\
-3 - 4\phi_2 < \phi_0 < 1.
\end{cases} \quad (19)$$

Then when $\phi_2 = 0$ from (19) is $-3 < \phi_0 < 1$, and from (18) is $-(1 + \phi_0)/2 < \phi_1 < 1$ and we have seen in Example 1 that these two inequalities identify the
stationarity region of the ARMA(2,1) process \((1 - (\phi_0 + \phi_1) B + \phi_1 B^2) Z_t = (1 - B) \eta_t\).

The above results show that in this setting, especially as far as the choice of the prior distribution is concerned, it is particularly convenient to refer to the parametrization of all the models in (4) in terms of \(\phi_0^{(p)}\). In particular, if \(p_{\text{max}}\) is the maximum value of the autoregressive order \(p\) in (4), we can reparametrize \(M_{p_{\text{max}},1}\) in terms of \(\phi_0^{(p_{\text{max}})}\), and introduce the prior distribution \(\pi_{p_{\text{max}},1}(\phi_0, \phi_1, \ldots, \phi_{p_{\text{max}} - 1})\); then all the other models in (4) are nested in \(M_{p_{\text{max}},1}\), and all the corresponding prior distributions follow from \(\pi_{p_{\text{max}},1}\). Moreover, the use of this reparametrization allows to specify \(\pi_{p_{\text{max}},1}(\phi_0, \phi_1, \ldots, \phi_{p_{\text{max}} - 1})\) by introducing a prior distribution for \(\phi_0\) and a prior distribution for the remaining parameters conditionally on \(\phi_0\); this has the advantage of allowing the prior belief on the presence of a unit root to be expressed explicitly.

5 The geometric partial Bayes factor for testing for unit autoregressive roots

We now determine the various marginal distributions required for comparing any two models in (4) by means of the geometric partial Bayes factor (7). In order to do so, for each model in (4) we first need to write explicitly the likelihood function and the geometric training likelihood corresponding to \(l = 1\) and to \(l > 1\); note that because of the stationarity constraints, both under \(M_{p,1}\) and \(M_{p,0}\) we can define the exact likelihood function (Box, Jenkins and Reinsel, 1994).

5.1 Computing the marginal distributions corresponding to the full likelihood

Consider first \(M_{p,0}\), 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) is:

\[
v_{ii} = 1 + \sum_{h=1}^{i-1} \phi_h^2 - \sum_{h=p+1-i}^{p} \phi_h^2, \quad 1 \leq i \leq p
\]

and

\[
v_{ij} = -\phi_{j-i} + \sum_{h=1}^{i-1} \phi_h \phi_{h+j-i} - \sum_{h=p+1-j}^{p+i-j} \phi_h \phi_{h+j-i}, \quad 1 \leq i < j \leq p,
\]

while the other elements of the matrix may be determined by noting that \( V_p \) is symmetric with respect to both its principal diagonals. Then, if we let

\[
a_t = z_t - \sum_{i=1}^{p} \phi_i z_{t-i},
\]

the exact likelihood function under model (2) can be written as

\[
f_{p,0} (z | \tau^2, \phi^{(p)}) = (2\pi \tau^2)^{-(n-1)/2} |V_p|^{1/2} \exp \left\{ -\frac{1}{2\tau^2} \left[ \sum_{t=1}^{p} \sum_{s=1}^{p} v_{ts} z_t z_s + \sum_{t=p+1}^{n-1} a_t^2 \right] \right\}, \quad (20)
\]

and the corresponding marginal distribution is

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

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

\[
f_{0,0} (z | \tau^2) = (2\pi \tau^2)^{-(n-1)/2} \exp \left\{ -\frac{1}{2\tau^2} \sum_{t=1}^{n-1} z_t^2 \right\}, \quad (21)
\]

and

\[
m_{0,0} (z) = \Gamma \left( \frac{n - 1}{2} \right) \left( \pi \sum_{t=1}^{n-1} z_t^2 \right)^{-(n-1)/2}, \quad (22)
\]
Consider now model $M_{p,1}$ and suppose that $p > 0$. Let $\tau^2 W^{-1}_p$ be the variance and covariance matrix of $p$ consecutive variables of the process (3), $b = (0, \ldots, 0, 1)$ 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

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

Then the exact likelihood function under $M_{p,1}$ is

$$f_{p,1}(\mathbf{z}|\phi_0^{(p)}, \tau^2) = (2\pi \tau^2)^{-(n-1)/2} |D|^{1/2} |D + H'H|^{-1/2} \times \exp\left\{-\frac{1}{2\tau^2} \sum_{t=1}^{n-1} \sum_{s=1}^{n-1} r_{ts} z_t z_s \right\},$$

(23)

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

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

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

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

(where we use the convention that for all $j = 1, \ldots, p$, $h_{ij} = 0$ if $i < 1$). It follows that the corresponding marginal distribution is

$$m_{p,1}(\mathbf{z}) = \pi^{-(n-1)/2} \Gamma\left(\frac{n-1}{2}\right) \times \int \left(\frac{|D|}{|D + H'H|}\right)^{1/2} \left[\sum_{t=1}^{n-1} \sum_{s=1}^{n-1} r_{ts} z_t z_s \right]^{-(n-1)/2} \pi_{p,1}\left(\phi_0^{(p)}\right) d\phi_0^{(p)}.$$

17
Instead, if \( p = 0 \), \( z = (z_1, \ldots, z_{n-1}) \) is a realization of a moving average process of order 1 with mean equal to 0 and moving average parameter equal to 1; following Galbraith and Galbraith (1974) the exact likelihood is then

\[
f_{0,1} (z | \tau^2) = (2\pi \tau^2)^{-(n-1)/2} n^{-1/2} \exp \left\{ -\frac{1}{2\tau^2} \sum_{t=1}^{n-1} \sum_{s=1}^{n-1} u_{ts} z_t z_s \right\}, \tag{24}
\]

where

\[
u_{ts} = \begin{cases} \frac{t (n - s)}{n} & t \leq s \\ u_{st} & t > s, \end{cases}
\]

so that

\[
m_{0,1} (z) = \Gamma \left( \frac{n - 1}{2} \right) n^{-1/2} \left[ \prod_{t=1}^{n-1} \prod_{s=1}^{n-1} \left( \frac{n s_{ts} z_t z_s}{2} \right)^{(n-1)/2} \right].
\]

5.2 Computing the marginal distributions corresponding to the geometric training likelihoods

First, for each model in (4), we write the geometric training likelihood corresponding to \( l = 1 \). Recall that under \( M_{p,q} \), \( z_i \) has a normal distribution with zero mean and variance \( \tau^2 \gamma_{pq} \) that can be determined with standard methods from \( \phi^{(p)} \) and \( \phi_0^{(p)} \) when \( q = 0 \) and \( q = 1 \) respectively. Then the geometric training likelihood is

\[
f_{1,1} (z | \tau^2, \phi^{(p)}) = (2\pi \tau^2 \gamma_{pq})^{-1/2} \exp \left\{ -\frac{1}{2 \tau^2 \gamma_{pq} \sum_{t=1}^{n} z_t^2} \right\}
\]

and the corresponding marginal distribution is

\[
m_{1,1} (z) = \left[ \frac{1}{(n-1)} \sum_{t=1}^{n-1} z_t^2 \right]^{-1/2}. \tag{25}
\]

It is interesting to note that \( m_{1,1} (z) \) turns out to be independent of \( p \) and \( q \), as well as of the prior distribution for the autoregressive parameters \( \phi^{(p)} \) and \( \phi_0^{(p)} \), confirming that minimal training samples do not carry enough informations about the autoregressive parameters. Moreover, it implies that
when \( l = 1 \), the geometric partial Bayes factor for comparing any two models in (4) equals the corresponding standard Bayes factor.

Now suppose that in order to achieve robustness with respect to the choice of the prior distributions for the autoregressive coefficients, we let \( l > 1 \). Note, following Varshavsky (1996) and Barbieri and Conigliani (1998, 2000), that in the time series set up it seems natural to preserve the time structure for the training samples; for this reason, when computing a geometric training likelihood, instead of considering all possible subsets of the data of a given size, we will consider only subsamples made of successive observations. In particular, let \( \mathbf{z}_k(l) = (z_{k+1}, \ldots, z_{k+l}) \) be any subsample of \( \mathbf{z} \) of size \( l \) made of successive observations; clearly \( k = 0, 1, \ldots, n - l - 1 \), so that there are \( n - l \) training samples of this kind. In what follows, we will assume that \( l > p_{\text{max}} \), so that we can write the exact likelihood function for \( \mathbf{z}_k(l) \) under each model in (4), and the geometric training likelihood accordingly.

Consider first model \( M_{p,0} \); when \( p > 0 \), from (20) is

\[
f_{p,0}^{l}(\mathbf{z} | \tau^2, \phi^{(p)}) = \left(2\pi\tau^2\right)^{-l/2} |V_p|^{1/2} \times \exp\left\{-\frac{1}{2(n-l)\tau^2}\sum_{k=0}^{n-l-1} \left[ \sum_{t=1}^{p} \sum_{s=1}^{p} v_{ts} z_{k+t} z_{k+s} + \sum_{t=p+1}^{l} a_{k+t}^2 \right]\right\},
\]

so that

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

Instead, if \( p = 0 \), according to (21) is

\[
f_{0,0}^{l}(\mathbf{z} | \tau^2) = \left(2\pi\tau^2\right)^{-l/2} \exp\left\{-\frac{1}{2(n-l)\tau^2}\sum_{k=0}^{n-l-1} \sum_{t=1}^{l} z_{k+t}^2 \right\},
\]

and

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

(26)
Consider now model $M_{p,1}$, with $p > 0$; from (23) we have

$$
\begin{align*}
& f_{p,1}^{l} \left( z \mid \phi_0^{(p)}, \tau^2 \right) = \left( 2\pi \tau^2 \right)^{-l/2} \left| D \right|^{1/2} \left| D + \tilde{H}' \tilde{H} \right|^{-1/2} \\
& \quad \times \exp \left\{ -\frac{1}{2 (n - l) \tau^2} \sum_{k=0}^{n-l-1} \sum_{t=1}^{l} \sum_{s=1}^{l} \tilde{r}_{ts} z_{k+t} z_{k+s} \right\},
\end{align*}
$$

where $\tilde{r}_{ts}$ is the $(t, s)$ element of the $l \times l$ matrix $\tilde{R}$ defined by

$$
\tilde{R} = \tilde{A}' \tilde{A} - \left( \tilde{H}' \tilde{A} \right)' \left( D + \tilde{H}' \tilde{H} \right)^{-1} \left( \tilde{H}' \tilde{A} \right).
$$

$\tilde{A}$ 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 $H$. The corresponding marginal distribution is then

$$
\begin{align*}
& m_{p,1}^{l} (z) = \left[ \frac{\pi}{(n - l)} \right]^{-l/2} \Gamma \left( \frac{l}{2} \right) \\
& \quad \times \int \left| D \right|^{1/2} \left| D + \tilde{H}' \tilde{H} \right|^{-1/2} \left[ \sum_{k=0}^{n-l-1} \sum_{t=1}^{l} \sum_{s=1}^{l} \tilde{r}_{ts} z_{k+t} z_{k+s} \right]^{-l/2} \\
& \quad \times \pi_{p,1} \left( \phi_0^{(p)} \right) d\phi_0^{(p)}.
\end{align*}
\quad (27)
$$

Instead, if $p = 0$, according to (24) is

$$
\begin{align*}
& f_{0,1}^{l} \left( z \mid \tau^2 \right) = \left( 2\pi \tau^2 \right)^{-l/2} \left( l + 1 \right)^{-1/2} \\
& \quad \times \exp \left\{ -\frac{1}{2 (n - l) \tau^2} \sum_{k=0}^{n-l-1} \sum_{t=1}^{l} \sum_{s=1}^{l} \tilde{u}_{ts} z_{k+t} z_{k+s} \right\},
\end{align*}
$$

where

$$
\tilde{u}_{ts} = \begin{cases} 
  t (l + 1 - s) & t \leq s \\
  l + 1 & t > s,
\end{cases}
$$

so that

$$
\begin{align*}
& m_{0,1}^{l} (z) = \Gamma \left( \frac{l}{2} \right) \left( l + 1 \right)^{-1/2} \left[ \frac{\pi}{(n - l)} \sum_{k=0}^{n-l-1} \sum_{t=1}^{l} \sum_{s=1}^{l} \tilde{u}_{ts} z_{k+t} z_{k+s} \right]^{-l/2}.
\end{align*}
$$
6 Testing for a unit root in a first order autoregressive process

Let \( y = (y_1, \ldots, y_n) \) 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 \( z = (z_1, \ldots, z_{n-1}) \), 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 details 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. Here we consider the same model selection problem adopting a robust approach, and solve it by means of the geometric partial Bayes factor \( B^l_{(1,1)(0,0)} (z) \).

Note that the marginal distributions under the integrated model corresponding to the full likelihood and to the geometric training likelihoods with \( l = 1 \) and \( l > 1 \) are given by (22), (25) and (26) respectively. Moreover, the marginal distribution under the stationary model corresponding to the geometric training likelihood with \( l = 1 \) is given by (25); recall that (25) holds regardless of the prior distribution for the autoregressive parameter. In order to compute the remaining marginal densities under the stationary model, we first need to write explicitly the exact likelihood function and the geometric training likelihood corresponding to \( l > 1 \); again, as the integrals with respect to the autoregressive coefficient in general cannot be solved analytically, for the moment we do not specify a prior distribution for \( \phi_0 \).
6.1 Computing the marginal distributions under the ARMA(1,1) model

Consider the ARMA(1,1) model, and recall that in this case is

\[ \mathbb{E}(Z_t^2) = \frac{2\tau^2}{(1 + \phi_0)}. \]

It follows, using the notation of Section 5, that:

\[ D = \frac{1 + \phi_0}{1 - \phi_0} \begin{bmatrix} 1 & -1 \\ -1 & 2(1 + \phi_0)^{-1} \end{bmatrix} \]

and

\[ |D| = \frac{1 + \phi_0}{1 - \phi_0}; \]

moreover, \( A \) is a lower triangular \((n - 1) \times (n - 1)\) matrix with \((r, s)\) element given by

\[ a_{rs} = \begin{cases} \begin{array}{cl} 0 & r - s < 0 \\ 1 & r - s = 0 \\ 1 - \phi_0 & r - s > 0 \end{array} \end{cases} \]

and \( H \) is the \((n - 1) \times 2\) matrix given by

\[ H' = \begin{bmatrix} -\phi_0 & -\phi_0 & \ldots & -\phi_0 \\ 1 & 1 & \ldots & 1 \end{bmatrix}', \]

so that

\[ |D + H'H| = \frac{1 + \phi_0}{1 - \phi_0} + n - 1. \]

It follows from (23), that the full likelihood under the ARMA(1,1) model can be written as

\[
    f_{1,1}(z \mid \phi_0, \tau^2) = (2\pi \tau^2)^{-(n-1)/2} \left[ 1 + \frac{(n - 1)(1 - \phi_0)}{1 + \phi_0} \right]^{-1/2} 
    \times \exp \left\{ -\frac{1}{2\tau^2} \sum_{t=1}^{n-1} \sum_{s=1}^{n-1} r_{ts} z_t z_s \right\},
\]

where for \( t < s \) is

\[ r_{ts} = \frac{(1 - \phi_0) [\phi_0 + t (1 - \phi_0)] [1 + (n - 1 - s) (1 - \phi_0)]}{2\phi_0 + n (1 - \phi_0)} ,\]
for \( t > s \) is \( r_{ts} = r_{st} \), while for \( t = s \) is
\[
r_{ss} = \frac{(1 - \phi_0)^3 (s - 1) (n - 1 - s) + 2\phi_0 + (n - 1)(1 - \phi_0)}{2\phi_0 + n(1 - \phi_0)}
\]
\((t, s = 1, \ldots, n - 1)\). The corresponding marginal distribution is then
\[
m_{1,1}(z) = \pi^{-(n-1)/2} \Gamma \left( \frac{n - 1}{2} \right)
\times \int_{-1}^{1} \left[ 1 + \frac{(n - 1)(1 - \phi_0)}{1 + \phi_0} \right]^{-1/2} \left( \sum_{t=1}^{n-1} \sum_{s=1}^{n-1} r_{ts} z_t z_s \right)^{-(n-1)/2} \pi_{1,1}(\phi_0) d\phi_0.
\]

Similarly, from (27), under \( M_{1,1} \) the geometric training likelihood corresponding to \( l > 1 \) is given by
\[
f_{1,1}(z | \phi_0, \tau^2) = (2\pi \tau^2)^{-l/2} \left[ 1 + \frac{l(1 - \phi_0)}{1 + \phi_0} \right]^{-1/2}
\times \exp \left\{ -\frac{1}{2(n - l) \tau^2} \sum_{k=0}^{n-l-1} \sum_{t=1}^{l} \sum_{s=1}^{l} \tilde{r}_{ts} z_{k+t} z_{k+s} \right\},
\]
where for \( t < s \) is
\[
\tilde{r}_{ts} = \frac{(1 - \phi_0) [\phi_0 + t(1 - \phi_0)][1 + (l - s)(1 - \phi_0)]}{2\phi_0 + (l + 1)(1 - \phi_0)},
\]
for \( t > s \) is \( \tilde{r}_{ts} = \tilde{r}_{st} \), while for \( t = s \) is
\[
\tilde{r}_{ss} = \frac{(1 - \phi_0)^3 (s - 1)(l - s) + 2\phi_0 + l(1 - \phi_0)}{2\phi_0 + (l + 1)(1 - \phi_0)}
\]
\((t, s = 1, \ldots, n - 1)\). The corresponding marginal distribution is then
\[
m_{1,1}(z) = \left[ \frac{\pi}{(n - l)} \right]^{-l/2} \Gamma \left( \frac{l}{2} \right)
\times \int_{-1}^{1} \left[ 1 + \frac{l(1 - \phi_0)}{1 + \phi_0} \right]^{-l/2} \left( \sum_{k=0}^{n-l-1} \sum_{t=1}^{l} \sum_{s=1}^{l} \tilde{r}_{ts} z_{k+t} z_{k+s} \right)^{-l/2} \pi_{1,1}(\phi_0) d\phi_0.
\]

23
6.2 A simulation experiment

As we pointed out in the Introduction, Marriott and Newbold (1998) 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 non-zero probability mass at $\phi_0 = 1$. Instead, they advocate the use of sharp, informative prior distributions, with densities giving more prior weight to values of $\phi_0$ 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 behaviour 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.

Thus, they generated 1000 samples of 100 observations from first order autoregressive processes with different values of $\phi_0$, and for each sample they computed the posterior probability of the random walk model $M_{0,0}$ using the standard Bayes factor and equal prior model’s probabilities, assuming either the uniform prior, the beta prior or the sharp beta prior for $\phi_0$. Then they computed the proportion of the samples for which $P(M_{0,0} | z)$ was greater than 0.5. The results are presented in Table 1, and led the authors to conclude that in this setting the choice of the prior distribution for $\phi_0$ is particularly relevant, so that special care should be devoted to elicit such prior and “authoromatic” choices should be avoided. In fact, while the sharp beta prior performed extremely well when the “true” process was stationary autoregressive, the opposite was the case for a random walk generating process.

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

<table>
<thead>
<tr>
<th>$\phi_0$</th>
<th>Prior for $\phi_0$</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>uniform</td>
<td>beta</td>
<td>beta sharp</td>
</tr>
<tr>
<td>1</td>
<td>0.968</td>
<td>0.875</td>
<td>0.677</td>
</tr>
<tr>
<td>0.95</td>
<td>0.827</td>
<td>0.554</td>
<td>0.244</td>
</tr>
<tr>
<td>0.85</td>
<td>0.212</td>
<td>0.018</td>
<td>0.000</td>
</tr>
</tbody>
</table>
For purposes of comparison, we consider here the same model selection problem and solve it by means of the geometric partial Bayes factor $B_{(1,1)(0,0)}^t(z)$. We also explore the behaviour of a triangular prior of the form

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

which is somehow intermediate between the uniform and the beta. The results, corresponding to different sizes $l$ of the training samples, are presented in Table 2.

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

<table>
<thead>
<tr>
<th>$\phi_0$</th>
<th>uniform</th>
<th>triangular</th>
<th>beta</th>
<th>beta sharp</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.984</td>
<td>0.924</td>
<td>0.871</td>
<td>0.814</td>
</tr>
<tr>
<td>0.95</td>
<td>0.916</td>
<td>0.703</td>
<td>0.520</td>
<td>0.333</td>
</tr>
<tr>
<td>0.85</td>
<td>0.392</td>
<td>0.083</td>
<td>0.045</td>
<td>0.030</td>
</tr>
</tbody>
</table>

Two issues raise clearly from the table above. First, notice that as $l$ increases, the differences in the behaviour of the various prior distributions are less relevant. In fact, while 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.

However, this gain in robustness is not obtained when the beta sharp prior is used. Indeed, the sharper the prior, the less the updated prior distribution is sensitive to the training likelihood. Therefore the differences in the performance of the various priors as $l$ increases are consequences of the fact that these distributions have mass concentrated on intervals of very different width. On the other hand the priors we used represent very different prior belief on the possible alternatives to the value $\phi_0 = 1$. In fact, if the aim of the test is to choose between, say, $\phi_0 = 0.95$ and $\phi_0 = 1$, then a distribution extremely concentrated around 1 is certainly a reasonable prior. On the other hand, if
the uncertainty on $\phi_0$ concerns the interval $(-1, 1)$, a distribution with mass on the whole interval will generally perform better. In order to see this, 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 $\phi_0$, 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.

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

<table>
<thead>
<tr>
<th>$\phi_0$</th>
<th>uniform</th>
<th>triangular</th>
<th>beta</th>
<th>beta sharp</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td>0.291</td>
<td>0.170</td>
<td>0.045</td>
<td>0.015</td>
</tr>
<tr>
<td>0</td>
<td>0.090</td>
<td>0.044</td>
<td>0.010</td>
<td>0.004</td>
</tr>
<tr>
<td>-0.3</td>
<td>0.022</td>
<td>0.010</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

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 2 that under either one of the beta priors the empirical cdfs for $P(M_{0,0} \mid \mathbf{z})$ exhibit a steep growth around the median, so that only very rarely $P(M_{0,0} \mid \mathbf{z})$ will 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 a strong evidence in favour of the stationary model as the uniform prior does. The behaviour of the four cdfs in Figure 2 can also be summarized by averages and standard deviations of the simulated values of $P(M_{0,0} \mid \mathbf{z})$. 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.
Figure 2. Empirical cdf for $P(M_0,0|\mathbf{z})$ for series of 10 observations
$(l = 1, \phi_0 = 0)$.

Finally, note that although our approach is fully Bayesian, its performance can be interpreted also 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 et al. (1994) concerning the usual Dickey-Fuller (D-F) test and a test based on maximum likelihood estimation (MLE).

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

<table>
<thead>
<tr>
<th>$\phi_0$</th>
<th>Prior for $\phi_0$</th>
<th>D-F test</th>
<th>MLE test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>uniform</td>
<td>beta sharp</td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>$l=1$ $l=10$ $l=20$</td>
<td>$l=1$ $l=10$ $l=20$</td>
<td></td>
</tr>
<tr>
<td>0.016</td>
<td>0.076 0.129</td>
<td>0.421 0.426 0.435</td>
<td>0.050</td>
</tr>
<tr>
<td>0.95</td>
<td>0.082 0.297 0.480</td>
<td>0.875 0.875 0.875</td>
<td>0.117</td>
</tr>
<tr>
<td>0.85</td>
<td>0.608 0.917 0.955</td>
<td>0.999 0.999 0.998</td>
<td>0.621</td>
</tr>
</tbody>
</table>
References


<table>
<thead>
<tr>
<th>Year</th>
<th>Author(s)</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - 1997</td>
<td>Mariano D’Antonio e Margherita Scarlato</td>
<td>Struttura economica e commercio estero: un’analisi per le province italiane</td>
</tr>
<tr>
<td>2 - 1997</td>
<td>Pierangelo Garegnani e Antonella Palumbo</td>
<td>Accumulation of capital</td>
</tr>
<tr>
<td>3 - 1997</td>
<td>Elio Cerrito</td>
<td>Crisi di cambio e problemi di politica monetaria nell’Italia di fine Ottocento. Appunti su alcune evidenze empiriche</td>
</tr>
<tr>
<td>4 - 1998</td>
<td>Francesco Manni</td>
<td>Struttura delle fonti di finanziamento: Un’indagine sulle principali società italiane produttrici di bevande</td>
</tr>
<tr>
<td>5 - 1998</td>
<td>J. O. Berger e J. Morlita</td>
<td>Default Bayes factors for one-sided hypothesis testing</td>
</tr>
<tr>
<td>6 - 1998</td>
<td>Attilio Trezzini</td>
<td>Capacity utilisation in the long run: a reply to Serrano</td>
</tr>
<tr>
<td>7 - 1998</td>
<td>A. M. Ferragina</td>
<td>Quality product differentiation in CEE-EU Intra-Industry trade</td>
</tr>
<tr>
<td>8 - 1998</td>
<td>M. F. Renzi, L. Cappelli, G. Salerno</td>
<td>Outsourcing: Opportunità e limiti per le aziende che operano con sistemi di qualità conformi alle norme ISO 9000</td>
</tr>
<tr>
<td>9 - 1998</td>
<td>Margherita Scarlato</td>
<td>The impact of international trade on employment and wage differentials: some evidence from the Italian macro-regions</td>
</tr>
<tr>
<td>10 - 1998</td>
<td>Attilio Trezzini</td>
<td>Some notes on long-run capacity utilisation, steady state and induced investment</td>
</tr>
<tr>
<td>11 - 1999</td>
<td>Salvatore Monni</td>
<td>A Convergence analysis of human development</td>
</tr>
<tr>
<td>12 - 1999</td>
<td>Guido M. Rey</td>
<td>Informazione e politiche pubbliche: non è mai troppo tardi</td>
</tr>
<tr>
<td>13 - 2000</td>
<td>Maria Maddalena Barbieri e Caterina Conigliani</td>
<td>Fractional bayes factors for the analysis of autoregressive models with possible unit roots</td>
</tr>
<tr>
<td>14 - 2000</td>
<td>Margherita Scarlato</td>
<td>Capitale Sociale e Sviluppo Economico</td>
</tr>
<tr>
<td>15 - 2000</td>
<td>Anna M. Ferragina</td>
<td>Price versus quality competition in Italy’s trade with Central and Eastern Europe over the Transition</td>
</tr>
<tr>
<td>16 - 2000</td>
<td>Mariano D’Antonio e Margherita Scarlato</td>
<td>Capitale umano e sviluppo economico Un modello di equilibrio economico generale per il Centro-Nord e il Mezzogiorno d’Italia</td>
</tr>
<tr>
<td>17 - 2000</td>
<td>Marisa Cenci e Luana Foffa Ciucci</td>
<td>Gli effetti della regolamentazione sull’attività di Insider Trading</td>
</tr>
<tr>
<td>18 - 2000</td>
<td>Andrea Gheno</td>
<td>Alberi binomiali e struttura della volatilità</td>
</tr>
<tr>
<td>19 - 2000</td>
<td>Marisa Cenci</td>
<td>Modelli per la struttura a termine con volatilità stocastica (Una rassegna critica)</td>
</tr>
<tr>
<td>20 - 2000</td>
<td>Marisa Cenci e Andrea Gheno</td>
<td>Metodologie per la valutazione delle obbligazioni convertibili in ipotesi di evoluzione stocastica della struttura per scadenza</td>
</tr>
<tr>
<td>21 - 2001</td>
<td>Mariano D’Antonio</td>
<td>L’interdipendenza tra economia e ambiente: un’analisi applicata all’Italia</td>
</tr>
<tr>
<td>22 - 2001</td>
<td>Alessia Naccarato</td>
<td>Uno studio sulla diffusione di fenomeni metereologici mediante particolari modelli spazio-temporali</td>
</tr>
</tbody>
</table>
23 - 2001  Marisa Cenci e Margherita Scarlato
Settore sommerso e politiche di emersione: un approccio stocastico

24 - 2001  Sebastiano Fadda  - Per un nuovo sostegno allo sviluppo locale: patti territoriali e servizi al territorio nel nuovo scenario economico

25 - 2002  Fabrizio De Filippis e Daniela Storti - Le politiche di sviluppo rurale nell’Unione europea: un “secondo pilastro” tutto da inventare

26 - 2002  Fabrizio De Filippis, Luca Salvatici - The eastward Enlargement of the European Union and the Common Agricultural Policy

27 - 2002  Simona Fiumara ed Elena Bellisario - La rilevanza della certificazione di qualità nella responsabilità del produttore di giocattoli

28 - 2002  Sebastiano Fadda - Does the change of Economic institutions require a change in values?

29 - 2002  Andrea Corese - La mediateca: approcci gestionali

30 - 2002  Saverio M. Fratini - Il teorema dell’indice e l’unicità dell’equilibrio con produzione: una discussione dei limiti del procedimento.

31 - 2002  Caterina Conigliani and Fulvio Spezzaferri - An alternative Bayes factor for testing for unit autoregressive roots