Testing for Unit Roots in a Bayesian Framework *

Michel LUBRANO[1]
GREQAM, CNRS, Marseille

March 1994

Abstract

Classical tests for unit roots have been criticized for their unusual asymptotic theory leading to disconnected confidence intervals and their lack of power in small samples. Such critiques were initiated by Sims (1988) who promoted a Bayesian approach to the question. Various papers then appeared, which took up this idea. The Bayesian approach is however not exempt from criticism, as shown by Phillips (1991) who emphasizes the role of the prior as a bias toward stationarity. In this paper, I emphasize the role of the model and of the initial conditions. I show that with a structural model and an adequate treatment of the first observation, a Bayesian unit root test based on the computation of $\Pr(\rho \geq 1|y)$ can produce results which are more or less in accordance with classical results. The extended Nelson-Plosser data set serves as an illustration.

key words: Bayesian, unit root tests, prior information, initial conditions.
JEL classification: C11, C12, C22

1 INTRODUCTION

The first interest in unit root testing came from an old debate about business cycle theory: are output fluctuations transitory? The traditional and Keynesian answer is that fluctuations in output are transitory deviations from a deterministic trend. The detrended real output can be modeled as a stationary AR process. As a consequence economic policy has just

*This paper is a very extensive revision of Lubrano (1991). I am grateful to a number of persons for helpful comments and suggestions while writing and correcting this paper: Luc Bauwens, Jean-Pierre Florens, Franz Palm, Peter Schotman, Mark Steel, Herman van Dijk. Thanks are also due to anonymous referees for their constructive suggestions. Remaining errors and shortcomings are solely mine. The initial writing was done while I was visiting the Economic Department of the Catholic University of Louvain. I would like to acknowledge the nice stay I had there. I am grateful to Peter Schotman who provided me with the extended Nelson-Plosser data set.

[1]GREQAM, Centre de la Vieille Charité, rue de la Charité 2, F-13002 Marseille, France. Phone: +33.91.14.07.45. E-mail: lubrano@ehess.cnrs-mrs.fr.

to correct the transitory fluctuations and it will have no lasting effects. Starting with the seminal paper by Nelson and Plosser (1982) many authors challenged this view. If there is a unit root in the GNP, output fluctuations are not transitory and contribute to the stochastic trend contained in the series. Short run corrections of the business cycle become much more dangerous as they may have a permanent effect. The traditional Keynesian economic policy (stop and go) is invalidated. And many classical studies failed to reject the null hypothesis of a unit root in the GNP.
However a number of authors have questioned the power of unit root tests in small samples. Bayesians have been very active in this questioning beginning with Sims (1988). The point is the following. The classical asymptotic theory is complicated, leading to difficulties for inference in the presence of a unit root and to disconnected confidence intervals for testing. In some papers rather extreme opinions were expressed saying that the Bayesian solution was clear and simple when the classical approach was logically unsound. In a series of empirical applications, using a Bayesian approach, DeJong and Whiteman (1989, 1991a, b) obtained results, rejecting the presence of a unit root in various economic series, which contradicted those of Nelson and Plosser.

These results were deeply questioned by Phillips (1991), saying that if the Bayesian tests rejected so often the unit root hypothesis, this was because they used a flat prior on the auto-regressive parameter which very much favored the stationary hypothesis. So their empirical results were declared to be non-robust to the prior specification.

The aim of this paper is to discuss the Bayesian treatment of unit root tests. In a first section, I show that the choice of a model which reflects the properties of the stochastic process is essential. The model I use is similar to that of Schotman and van Dijk (1991a) and is identical to a regression model with auto-correlated errors. In a second section, I discuss the properties of the likelihood function around the unit root, showing that the treatment of the initial condition is essential. The method for testing and the prior density are discussed in the next section. The paper concludes with a comparison of three Bayesian approaches on the extended Nelson-Plosser data set.

2 THE "RIGHT" MODEL FOR TESTING FOR UNIT ROOTS

The literature on time series processes has taught us that a stochastic processes with a unit root displays very peculiar characteristics as underlined in Nelson and Plosser (1982). One key characteristic is that the mean of the process does not exist. Consequently there is no deterministic trend that serves as a long-run anchor around which the process would have a tendency to revert. In this section I shall introduce two different parameterizations of the traditional autoregressive regression model used to test for the presence of a unit root in a series and show that one reproduces the key characteristics of a unit root stochastic process, when the other does not. Following the terminology used in Schotman (1992), one parameterization will be called a structural parameterization, the other a reduced form parameterization.

2.1 Structural model for testing unit roots

A unit root test is useful for discriminating between a difference stationary (DS) and a trend stationary (TS) model. These models can be imbedded into a general model which is very convenient for testing:

$$\begin{cases} y_t & = & \mu + \delta t + u_t \\ A(L) u_t & = & \epsilon_t \end{cases} \tag{1}$$

where $A(L)$ is a lag polynomial of order p. As a simple model is sufficient for illustrating my argument, I shall concentrate on the case $p = 1$ in this paragraph, and discuss the general case in the next paragraph. For $p = 1$, I define $A(L) = 1 - \rho L$ and solve (1) into:

$$y_t = \rho\, y_{t-1} + [(1-\rho)\,\mu + \rho\,\delta] + (1-\rho)\,\delta\, t + \epsilon_t \tag{2}$$

We have the DS model for $\rho = 1$ which will be the null hypothesis. The alternative hypothesis is more complicated as the pure TS model is obtained only for $\rho = 0$. For $\rho < 1$ the model is a stable auto-regressive process with a deterministic trend. For $|\rho| > 1$ we have an explosive process. This parameterization is interesting as it displays one characteristic feature of a processes with a unit root: the absence of a mean. For $\rho = 1$ the intercept μ of the trend disappears, the trend also disappears and only the drift δ is left. This model has been used and discussed by Schotman and van Dijk (1991a, b), Schotman (1992) for unit root testing and by Koop at al (1992b) for prediction.

An alternative parameterization is widely used for testing for unit roots, especially in the classical framework, but also in a Bayesian framework, notably by Phillips (1991). It is the reduced form of (1) with:

$$y_t = \beta_0 + \beta_1 t + \rho y_{t-1} + \epsilon_t \tag{3}$$

where:

$$\beta_0 = (1-\rho)\,\mu + \rho\,\delta \qquad \beta_1 = (1-\rho)\,\delta$$

This parameterization does not imply that β_1 vanishes under the null. So and as underlined in Davidson and MacKinnon (1993, p. 709-710) y_t will not be of the same order of magnitude under the null and under the alternative. With (3), y_t is $O(T)$ under the alternative ($\rho < 1$) and $O(T^2)$ under the null as the presence of a unit root increases the order of y_t. In a classical framework, it is not vital to distinguish between (2) and (3) for practical testing as the estimator for ρ and its standard deviation will be the same in both cases[1] . However this distinction is vital for asymptotic theory. The distributional results of Dickey and Fuller are obtained only in this case [see Davidson and Mackinnon (1993 p.710)]. In a Bayesian framework, the only final tool we have is the complete posterior density. With (3) and a flat prior, we would be back into the standard regression framework with distributional results similar to the classical ones in the stationary case. With (1), we imbed in the joint posterior density one of the essential features of stochastic processes with a unit root and thus have a chance to get inference results which are nearer to the classical ones in this case.

[1] The estimation results for the constant and trend terms are of course different. But if lags of Δy_t are added, their coefficients will be the same in both parameterizations.

2.2 The properties of the general model

The interesting case is of course when p may be greater than one. A larger value of p may cope with remaining auto-correlation and leads to an augmented Dickey-Fuller type of test. If we allow for at most one unit root in the lag polynomial $A(L)$ of order p , $A(L)$ can be factored into:

$$A(L) = (1 - \rho L) + (1 - L)\, A^*(L) \tag{4}$$

where:

$$\rho = 1 - A(1)$$

and $A^*(L)$ is a polynomial of order $p-1$ without a constant term with generic element[2]:

$$a_j^* L^j = (\sum_{i=j+1}^{p} a_i) L^j \qquad j = 1, \ldots, p-1$$

The possible unit root corresponds to $\rho = 1$ (i.e. $A(1) = 0$). With factorization (4), model (1) is written as:

$$y_t = \rho\, y_{t-1} + [(1-\rho)\,\mu + \rho\,\delta] + (1-\rho)\,\delta\, t - A^*(L)(\Delta y_t - \delta) + \epsilon_t \tag{5}$$

The model seems to be the same as before with a non-linearity in ρ. However the extra term $A^*(L)(\Delta y_t - \delta)$ introduces an extra non-linearity in δ. This is the model which is used for instance by Schotman and van Dijk (1991a, b).

In this paper, I shall consider a slight simplification of model (5) in order to get a model with a single non-linearity, whatever the value of p. The idea[3] is to replace $A^*(L)(\Delta y_t - \delta)$ by $A^*(L)(\Delta y_t - \hat{\delta})$ where $\hat{\delta}$ is an estimator of δ which could be for instance the empirical mean of Δy_t. This simplification does not change the spirit of the model[4]. Model (5) is thus simplified into:

$$y_t = \rho\, y_{t-1} + [(1-\rho)\,\mu + \rho\,\delta] + (1-\rho)\,\delta\, t - A^*(L)\Delta y_t^* + \epsilon_t \tag{6}$$

with $\Delta y_t^* = \Delta y_t - \hat{\delta}$.

[2] This result can be proved using an Euclidean division of polynomials. See also Fuller (1976) page 374.

[3] A theoretical justification for this simplification could be an argument of linearization applied to the product $\alpha\,\delta$ around two estimates of α and δ where α stands for $A^*(1)$:

$$\alpha\,\delta \simeq \hat{\alpha}\,\hat{\delta} + (\hat{\alpha} - \alpha)\hat{\delta} + (\hat{\delta} - \delta)\hat{\alpha}$$

For $\hat{\alpha} = 0$, this reduces to: $\alpha\,\delta \simeq \alpha\,\hat{\delta}$. The choice of $\hat{\alpha} = 0$ can be justified as follows. In the time series literature, it is a model choice problem to select the number of lags in an AR model. This is done with an information criterion. If $A(L) = 1 - a_1 L - a_2 L - \ldots - a_p L$, then $A^*(1) = -\sum_{j=2}^{p}(j-1)a_j$. A sufficient, but not necessary condition for $A^*(1) = 0$ is that $a_2 = a_3 = \ldots = a_p = 0$. So linearizing the model around $A^*(1) = 0$ is equivalent to implicitly restricting the model choice problem between an $AR(1)$ and an $AR(p)$. Linearizing arguments can be found elsewhere in the Bayesian literature such as in Zellner (1971 p. 96).

[4] In a classical non-linear regression model, this simplification has only an influence on estimation results for μ.

3 THE IMPORTANCE OF INITIAL CONDITIONS

In order to write the likelihood function, I propose a reparameterization of the model which separates the linear and the non-linear parameters. We have to decide how to deal with the initial conditions: fixed or random. Integrating out the linear parameters (σ^2, μ, δ) under a diffuse prior, we get the marginal likelihood function as a function of ρ alone. I show in this section that the marginal likelihood function in ρ of the model with a constant term is unbounded at $\rho = 1$ when the initial conditions are taken as fixed. It is bounded everywhere when the initial conditions are taken as random and are observed.

3.1 A convenient reparameterization

Let us consider first the simple model (1). It is a regression model with auto-correlated errors where the exogenous variables are $(1, t)$. This model has been analyzed for instance in Zellner and Tiao (1964) and Richard (1977). It is noted for $p = 1$:

$$\begin{cases} y_t = x_t'\beta + u_t \\ (1 - \rho L)\, u_t = \epsilon_t \end{cases} \tag{7}$$

with $x_t' = [1, t]$ and $\beta' = [\mu, \delta]$. The size of the sample is $T + 1$. The error term ϵ_t is distributed as a Normal with zero mean and variance σ^2. In the stationary case the variance of u_t is $\sigma^2/(1 - \rho^2)$. Let us define the reparameterization $y_t(\rho)$ and $x_t(\rho)$ by:

$$\begin{cases} y_t(\rho) &= (1 - \rho L)\, y_t \\ x_t(\rho) &= (1 - \rho L)\, x_t \end{cases} \tag{8}$$

The model is now written as:

$$y_t(\rho) = x_t'(\rho)\beta(\rho) + \epsilon_t \tag{9}$$

When $p \geq 1$ our model is (6) and can be written in the same simple form as (9) provided we extend the definition of $x_t(\rho)$. We have:

$$\begin{cases} y_t = x_t\beta + u_t \\ (1 - \rho L)\, u_t + A^*(L)\, \Delta u_t = \epsilon_t \end{cases} \tag{10}$$

and:

$$\begin{cases} y_t(\rho) &= (1 - \rho L)\, y_t \\ x_t'(\rho) &= [x_t' - \rho x_{t-1}', \Delta y_{t-1}^*, \Delta y_{t-2}^*, \ldots, \Delta y_{t-p+1}^*] \end{cases} \tag{11}$$

Now the parameter β in (9) contains also the $p - 1$ elements of $A^*(L)$ and is consequently of dimension $p + 1$. The sample size becomes $T + p$.

3.2 Initial conditions

Initial conditions may be modeled by considering the process[5] at time $t = 0$:

$$y_0 = \mu + \delta\, 0 + u_0 \tag{12}$$

[5]Other solutions are possible. Zellner (1971) for instance considers that $p(y_0) = f_N(\mu + M, \sigma^2)$ or that y_0 is normal with parameters independent of those of the model.

Many authors, including Phillips (1991) (with model (3)), consider the initial conditions y_0 as fixed and derive the likelihood function of their model conditionally on the first observation, i.e. $L(y|y_0, \rho, \beta, \sigma^2)$.

We can suppose alternatively that the initial conditions are random with:

$$p(y_0|\mu, \sigma^2, \rho) = f_N\left(y_0|\mu, \sigma^2/(1-\rho^2)\right) \tag{13}$$

as $(1-\rho L)u_0 = \epsilon_0$. Calling y the vector of the T observations of y_t, the complete likelihood function is:

$$L(y, y_0|\rho, \beta, \sigma^2) = L(y|y_0, \rho, \beta, \sigma^2).p(y_0|\mu, \sigma^2, \rho) \tag{14}$$

Remark

For $p > 1$ the initial condition (12) should be altered. However, it would become rather complicated to handle in this case and usually a simple model like (12) is kept whatever the value of p. This is what is done for instance in Zivot (1992) or Schotman and van Dijk (1991a, b, c). This amounts to suppose that a part of the initial conditions are kept fixed and the complete likelihood function considered is:

$$L(y|y_{p-1}, y_{p-2}, \ldots, y_0, \rho, \beta, \sigma^2).p(y_0|\mu, \sigma^2, \rho)$$

I shall not detail further in the sequel the distinction between the cases $p = 1$ and $p > 1$ as they induce no difference for the computations.

Zellner (1971 p.88) shows that, in the general stationary case without a constant term, it is of no importance to consider fixed or random initial conditions. What I shall show next, is that this choice is of prime importance in the unit root case when there is a constant term in the model.

3.3 Properties of the likelihood function in the case of fixed initial conditions

For $T + p$ observations, we can define the T vector $y(\rho)$ and the $T \times (p+1)$ matrix $X(\rho)$ in order to write the model as:

$$y(\rho) = X(\rho)\,\beta + \epsilon \tag{15}$$

The conditional moments of y are:

$$\mathrm{E}(y(\rho)|\rho, y_0) = X(\rho)\,\beta \qquad \mathrm{Var}(y(\rho)|\rho, y_0) = \sigma^2\, I_T$$

which lead to the conditional data density:

$$L(y|y_0, \beta, \sigma^2, \rho) \propto \sigma^{-T} \exp -\frac{1}{2\sigma^2}[y(\rho) - X(\rho)\beta]'[y(\rho) - X(\rho)\beta] \tag{16}$$

So conditionally on ρ and y_0, we have the data density of a standard regression model. The analysis can be pursued conditionally on the parameter ρ and within the natural conjugate framework we shall recover all the standard analytical results. This is quite convenient to integrate out β and σ^2 in order to get the marginalized likelihood function of ρ.

The prior density can be decomposed into:

$$D'(\beta,\sigma^2,\rho) = D'(\beta|\sigma^2,\rho).D'(\sigma^2|\rho).D'(\rho)$$

The prior density on β and σ^2 may be an inverted gamma2 Normal density, but I shall use a non-informative prior instead with:

$$D'(\beta,\sigma^2|\rho) \propto 1/\sigma^2 \tag{17}$$

The choice of an analytical form for $D'(\rho)$ is free as anyway we shall have to integrate out numerically this parameter. This prior will be the subject of an extended discussion in the next section.

Let us integrate out σ^2. We get a first expression for the marginalized likelihood function. The posterior density of β is obtained as a by-product. It has the usual Student form:

$$D"(\beta|\rho,y,y_0) = f_t(\beta|\tilde{\beta}_*(\rho),\tilde{M}_*(\rho),\tilde{s}_*(\rho),\tilde{\nu}_*) \tag{18}$$

with:

$$\begin{cases} \tilde{M}_*(\rho) & = & X'(\rho)\,X(\rho) \\ \tilde{\beta}_*(\rho) & = & \tilde{M}_*^{-1}(\rho)\,X(\rho)\,y(\rho) \\ \tilde{s}_*(\rho) & = & y'(\rho)\,y(\rho) - \tilde{\beta}_*'(\rho)\,\tilde{M}_*(\rho)\,\tilde{\beta}_*(\rho) \\ \tilde{\nu}_* & = & T-(p+1) \end{cases} \tag{19}$$

Integrating out the kernel of this posterior density in β , we get a kernel of the marginalized likelihood function as a function of ρ which is:

$$L(\rho;y|y_0) \propto [\tilde{s}_*(\rho)]^{-\tilde{\nu}_*/2}.|\tilde{M}_*(\rho)|^{-1/2} \tag{20}$$

A kernel of the posterior of ρ is obtained by multiplying (20) by $D'(\rho)$. The marginalized likelihood (20) is unbounded for $\rho = 1$ as shown in the next theorem:

Theorem 1 *Let us consider the structural model defined by (9), (10), and (11) with a constant term. The marginal likelihood function of ρ computed with the non-informative prior (17) and the likelihood function (16) is unbounded at $\rho = 1$.*

Proof: the proof is adapted from O'Brien (1970). For $\rho = 1$, $X(\rho)$ has a column of zeros which makes the matrix $X'(\rho)\,X(\rho)$ singular. Let us decompose $X(\rho)$ into:

$$X(\rho) = [(1-\rho),\overline{X}(\rho)] = [1,\overline{X}(\rho)].\begin{bmatrix}(1-\rho) & 0 \\ 0 & I_p\end{bmatrix} = W\,Q$$

Then:

$$\begin{array}{rcl} |\tilde{M}_*(\rho)|^{-1/2} & = & |Q'W'WQ|^{-1/2} = |Q'Q|^{-1/2}.|W'W|^{-1/2} \\ & = & (1-\rho)^{-1}.|W'W|^{-1/2} \end{array}$$

is unbounded for $\rho = 1$. On the contrary the multiplicative term $(1-\rho)$ is not present in $\tilde{s}_*(\rho)$ as:

$$\begin{array}{rcl} \tilde{s}_*(\rho) & = & y'(\rho)(I_T - X(\rho)(X'(\rho)X(\rho))^{-1}X'(\rho))y(\rho) \\ & = & y'(\rho)(I_T - W(W'W)^{-1}W')y(\rho) \end{array}$$

which is finite at $\rho = 1$ as $\tilde{s}_*(1)$ is a residual sum of squares.

□

O'Brien was the first to discuss extensively the model with auto-correlated errors when ρ approaches 1 and there is a constant term. He proposed to use the likelihood function (15) together with a prior for ρ proportional to $(1-\rho)$ as an expression of the posterior density of ρ. This of courses cures the problem, but in a very brutal way as this prior density implies a prior mean for ρ equal to 1/3 and a prior variance equal to 1/18, following the properties of the Beta density[6].

Consequently, if we stick to a non-informative prior for ρ, the marginal likelihood function (20) cannot be used for testing for a unit root[7].

3.4 Properties of the likelihood function in the case of random initial conditions

With (13) we add an observation y_0 to the sample with $\mathrm{E}(y_0) = \mu$ and $\mathrm{Var}(y_0) = \sigma^2/(1-\rho^2)$. This is similar to what is usually done in a classical framework when one computes a GLS estimator for the model with auto-correlated errors at the order one [see for instance Judge et al (1985) p. 285]. So:

$$L(y, y_0|\rho, \beta, \sigma^2) = L(y|y_0, \rho, \beta, \sigma^2).\sqrt{1-\rho^2}\,\sigma^{-1}\ \exp -\frac{1-\rho^2}{2\sigma^2}(y_0-\mu)^2$$

or equivalently:

$$\begin{array}{l} \sqrt{1-\rho^2}\,\sigma^{-(T+1)}. \\ \exp -\frac{1}{2\sigma^2}\left((1-\rho^2)(y_0-\mu)^2 + [y(\rho)-X(\rho)\beta]'[y(\rho)-X(\rho)\beta]\right) \end{array} \tag{21}$$

The marginalized likelihood function of ρ is obtained easily thanks to the properties of the natural conjugate framework. So:

$$L(\rho; y, y_0) \propto [s_*(\rho)]^{-\nu_*/2}.|M_*(\rho)|^{-1/2}.(1-\rho^2)^{1/2} \tag{22}$$

with this time:

$$\left\{ \begin{array}{lcl} M_*(\rho) & = & X'(\rho)X(\rho) + \begin{bmatrix} 1-\rho^2 & 0 \\ 0 & 0 \end{bmatrix} \\ \beta_*(\rho) & = & M_*^{-1}(\rho).\left(X(\rho)\,y(\rho) + \begin{bmatrix} (1-\rho^2)\,y_0 \\ 0 \end{bmatrix}\right) \\ s_*(\rho) & = & y'(\rho)\,y(\rho) + y_0^2(1-\rho^2) - \beta_*'(\rho)\,M_*(\rho)\,\beta_*(\rho) \\ \nu_* & = & T-p \end{array} \right. \tag{23}$$

And now the marginalized likelihood function is well behaved when approaching the unit root as shown by theorem 2:

[6] $D'(\rho) \propto (1-\rho)$ is a special case of $D'(\rho) \propto \rho^{p-1}(1-\rho)^{q-1}$ with $p = 1$ and $q = 2$.

[7] and even for inference on ρ as numerical problems may occur for values of ρ as small as 0.85 as explained in O'Brien (1970).

Theorem 2 *Let us consider the structural model defined by (9), (10), and (11) and the random initial condition (13). The marginal likelihood function of ρ computed with the non-informative prior (17) and the joint likelihood function (21) has a finite value at $\rho = 1$ even when there is a constant term.*

Proof: the proof follows the same lines as before. We have to prove that the factor $|M_*(\rho)|^{-1/2}.(1-\rho^2)^{1/2}$ is now bounded in (22). We have:

$$M_*(\rho) = X'(\rho)\,X(\rho) + \begin{bmatrix} 1-\rho^2 & 0 \\ 0 & 0 \end{bmatrix} = Q'W'WQ + M_0$$

Using the formulae of determinant of partitioned matrices, we can write alternatively the determinant of $M_*(\rho)$ as $|M_*(\rho)| = |W'W|.|A|$ using the partitioned matrix A defined as:

$$A = \begin{bmatrix} M_0 & -Q' \\ Q & (W'W)^{-1} \end{bmatrix}$$

and the formula $|A| = |A_{22}|.|A_{11} - A_{12}A_{22}^{-1}A_{21}|$. However, we cannot use directly the alternative decomposition $|A| = |A_{11}|.|A_{22} - A_{21}A_{11}^{-1}A_{12}|$ as A_{11} (here M_0) is singular. But we can apply it to a new partition of A defined by:

$$A = \begin{bmatrix} 1-\rho^2 & 0_{1p} & \rho-1 & 0_{1p} \\ 0_{p1} & 0_{pp} & 0_{p1} & -I_p \\ 1-\rho & 0_{1p} & (ww)^{11} & (wW)^{12} \\ 0_{p1} & I_p & (W'w)^{21} & (W'W)^{22} \end{bmatrix}$$

and we obtain:

$$\begin{aligned} |A| &= (1-\rho^2).\left| \begin{pmatrix} 0_{pp} & 0_{p1} & -I_p \\ 0_{1p} & (ww)^{11} & (wW)^{12} \\ I_p & (W'w)^{21} & (W'W)^{22} \end{pmatrix} + \tfrac{1-\rho}{1+\rho}. \begin{pmatrix} 0_{pp} & 0_{p1} & 0_{pp} \\ 0_{1p} & 1 & 0_{1p} \\ 0_{pp} & 0_{p1} & 0_{pp} \end{pmatrix} \right| \\ &= (1-\rho^2).|B| \end{aligned}$$

Consequently:

$$|M_*(\rho)|^{-1/2}.(1-\rho^2)^{1/2} = |W'W|^{-1/2}.|B|^{-1/2}.(1-\rho^2)^{-1/2}.(1-\rho^2)^{1/2}$$

and:

$$\lim_{\rho\to 1} |M_*(\rho)|^{-1/2}.(1-\rho^2)^{1/2} = |W'W|^{-1/2}/\sqrt{(ww)^{11}}$$

which is finite. In order to prove that $s_*(\rho)$ has a finite value when $\rho = 1$, it suffices to note that:

$$\lim_{\rho\to 1} s_*(\rho) = \lim_{\rho\to 1} \tilde{s}_*(\rho) = \tilde{s}_*(1)$$

□

Remarks:

- I have implicitly supposed that $\rho \leq 1$ as a value of ρ greater than one implies a negative value for $s_*(\rho)$. But truncating the range of definition of ρ is very inconvenient for testing. I will thus discuss in the next paragraph how it is possible to consider an un-truncated range for ρ.

- As μ and y_0 enter in a symmetric way in the likelihood function, their role can be switched. (13) can then be interpreted as a prior density on μ with mean y_0 and variance $\sigma^2/(1-\rho^2)$ as Schotman and van Dijk (1991a, b) do. The corollary of that interpretation is that the initial conditions have to be fixed. This approach is formally equivalent to mine as I have taken a non-informative prior on μ. But combining an informative prior on μ with (21) would give a different marginal likelihood function in ρ from that used by Schotman and van Dijk.

3.5 Allowing for slightly explosive values of ρ

Zellner (1971) treats the explosive case, but assumes that there is no constant term. In (22) there is a constant term, but the model used for the random initial conditions is such that the multiplicative term $(1-\rho^2)$ puts the likelihood function to zero for $\rho = 1$. For $\rho > 1$, this term is negative when it should continue to be zero in order to truncate the range of integration. This suggests replacing it by $\mathrm{Max}(1-\rho^2, 0)$. But this still excludes inference for explosive values of ρ which are needed for testing in the next section. In order to extend slightly the range of ρ in the direction of explosive values, I replace $\max(1-\rho^2, 0)$ by a smooth transition between the two arguments of the Max operator. Tishler and Zang (1981) proposed in a different context to approximate $\max(z, 0)$ by the smooth transition function $q_v(z)$ defined by:

$$q_v(z) = \begin{cases} 0 & \text{if} \quad z < -v \\ (z+v)^2/4v & \text{if} \quad -v \leq z \leq v \\ z & \text{if} \quad v < z \end{cases}$$

where v is a small positive number. In our case this gives:

$$q_v(1-\rho^2) = \begin{cases} 0 & \text{if} \quad |\rho| > \sqrt{1+v} \\ (1-\rho^2+v)^2/4v & \text{if} \quad \sqrt{1-v} \leq |\rho| \leq \sqrt{1+v} \\ 1-\rho^2 & \text{if} \quad |\rho| < \sqrt{1-v} \end{cases} \tag{24}$$

A useful consequence of (24) is that the domain of definition of ρ can be extended to $[-\sqrt{1+v}, \sqrt{1+v}]$. With this transition function, the model for the initial condition is transformed from (13) to:

$$p(y_0|\mu, \sigma^2, \rho, v) = f_N\left(y_0|\mu, \frac{\sigma^2}{q_v(1-\rho^2)}\right)$$

and the likelihood function (22) becomes:

$$\begin{aligned} &\sqrt{q_v(1-\rho^2)}\,\sigma^{-(T+1)}. \\ &\exp -\tfrac{1}{2\sigma^2}(q_v(1-\rho^2)(y_0-\mu)^2 + [y(\rho) - X(\rho)\beta]'[y(\rho) - X(\rho)\beta]) \end{aligned} \tag{25}$$

The marginal likelihood function of ρ is transformed into:

$$L(\rho; y, y_0) \propto [s_*(\rho)]^{-\nu_*/2}.|M_*(\rho)|^{-1/2}.q_v(1-\rho^2)^{1/2} \tag{26}$$

with:

$$\begin{cases} M_*(\rho) &=& X'(\rho)\,X(\rho) + \begin{bmatrix} q_v(1-\rho^2) & 0 \\ 0 & 0 \end{bmatrix} \\ \beta_*(\rho) &=& M_*^{-1}(\rho)\left(X'(\rho)\,y(\rho) + \begin{bmatrix} q_v(1-\rho^2)y_0 \\ 0 \end{bmatrix}\right) \\ s_*(\rho) &=& y'(\rho)\,y(\rho) + y_0^2 q_v(1-\rho^2) - \beta_*'(\rho)\,M_*(\rho)\,\beta_*(\rho) \end{cases} \tag{27}$$

It is easy to show, using the same argument as the one developed for the proof of theorem 2, that (26) is bounded when $\rho \to 1$. This marginalized likelihood function does not seem very natural at first sight. Say that it is a compromise between the necessities coming from the introduction of explosive values and the need to have a random initial condition. It is conditional on a given value for v. But the choice of a value for v proves not to be crucial. A value of $v = 0.5$ leads to an interval of $[-1.225, 1.225]$ for ρ which is largely sufficient. Other values have been tried in the empirical applications of section 5 without producing very different results. Consequently, v will be fixed to this value in the rest of the paper.

4 TESTING THE POINT NULL HYPOTHESIS OF A UNIT ROOT

4.1 Testing a point hypothesis with posterior odds

A unit root test is based on the marginal likelihood function of ρ and corresponds to the basic hypothesis:

$$H_0 : \rho = 1 \qquad \text{against} \qquad H_1 : \rho < 1 \tag{28}$$

which implicitly assumes that the domain of definition of ρ is truncated at 1. In a Bayesian framework, an obvious way to test a point hypothesis is to compare the likelihood of the model when it incorporates the restriction $\rho = 1$ and the likelihood averaged over the complementary region. This is the posterior odds procedure. One has to give a positive prior weight ξ to the point hypothesis H_0 and share the remaining weight $(1-\xi)$ over the range of ρ covered by H_1 by mean of $(1-\xi)\,D'(\rho)$. With equal prior weights, the posterior odds is given by:

$$K_1 = \frac{L(\rho = 1; y, y_0)}{\int L(\rho; y, y_0)\,D'(\rho)\,d\rho} \tag{29}$$

using (22). The likelihood function corresponding to H_1 is averaged with the prior $D'(\rho)$ as a weighting function. A proper prior density is thus needed and the value of K_1 is dependent on the prior information. Sims (1988), Schotman and van Dijk (1991a, b, c), Koop and Steel (1992a) use posterior odds. The prior density $D'(\rho)$ can be Normal (which is very convenient if the model is parameterized as in (3)) and centered on the value one. Schotman and van Dijk prefer a uniform prior on the interval $[a, 1[$ which leads to:

$$K_1 = \frac{L(\rho = 1; y, y_0)}{\frac{1}{1-a}\int L(\rho; y, y_0)\,d\rho} \tag{30}$$

The unit root model is preferred if $K_1 > 1$ or $\Pr(H_0|y, y_0) \geq 0.50$, treating thus the null and the alternative in a symmetric way. Formula (30) has the advantage of showing that an arbitrary big weight can be given to the alternative, depending on the choice of a. This arbitrariness is related to the Lindley's paradox [Lindley (1957)]. In order to calibrate correctly the prior and also to avoid Lindley's paradox, Schotman and van Dijk choose a such that the prior covers say 95% of the surface contained under the marginal likelihood function $L(\rho; y, y_0)$. The procedure is simple, but it violates the likelihood principle.

4.2 Testing with a diffuse prior

One may be reluctant in using informative priors for testing. This is especially true in the domain of unit roots as classical and Bayesian procedure are reputed to give different answers as underlined in DeJong and Whiteman (1989, 1991a, b). The usual procedure for testing with non-informative prior is to compute a posterior confidence interval based on the posterior density of ρ defined by:

$$D"(\rho|y, y_0) \propto L(\rho; y, y_0)\, D'(\rho) \tag{31}$$

and reject the unit root hypothesis if 1 does not belong to this confidence interval. The two hypotheses are thus:

$$H_0 : \rho = 1 \text{ against } H_1 : \rho \neq 1 \tag{32}$$

which implies that the range of ρ is not truncated at $\rho = 1$ and that in our case the marginal likelihood function to use is (26).

A posterior confidence interval (or credible set) for ρ at the probability level α is a subset C_α of the range of ρ defined by:

$$\Pr(C_\alpha|y, y_0) = \int_{C_\alpha} D"(\rho|y, y_0)\, d\rho \geq 1 - \alpha \tag{33}$$

It is usually desirable to minimize the size of C_α and one way to do this is to choose the HPD (Highest Posterior Density) credible set definition. However it may lead to disconnected intervals if the posterior density is multi-modal. As we are especially interested in what happens near the unit root, one can define C_α such that:

$$\Pr(C_\alpha|y, y_0) = \int_{-\infty}^{\rho_{\text{sup}}} D"(\rho|y, y_0)\, d\rho \geq 1 - \alpha \tag{34}$$

I have chosen $-\infty$, but it could be any lower truncating point of the likelihood function such as $-\sqrt{1+v}$. The test then consists in comparing ρ_{sup} to one. It means that we can alternatively compute:

$$\int_{-\infty}^{1} D"(\rho|y, y_0)\, d\rho \tag{35}$$

and compare this value to $1 - \alpha$. (35) is in fact equivalent to $\Pr(\rho < 1|y, y_0)$ and if $\alpha = 5\%$ in (34), it means that the unit root hypothesis will be rejected if:

$$\Pr(\rho \geq 1|y, y_0) = \int_{1}^{+\infty} D"(\rho|y, y_0)\, d\rho \tag{36}$$

is lower than 5%[8]. Phillips (1991), and implicitly DeJong and Whiteman (1991a, b) also, base their decision on $\Pr(\rho \geq 1|y)$ and I shall proceed in a similar way. In this type of test, the two hypotheses are treated in an asymmetric way. One can say that this is not a pure unit root test as we have now:

$$H_0 : \rho \geq 1 \qquad \text{against} \qquad H_1 : \rho < 1 \tag{37}$$

The null hypothesis is composite as it covers the unit root case (which has a zero probability in this context) and the explosive case. Strictly speaking, this is a test of stationarity against non-stationarity. The only relevant question is to know if empirically this makes a difference with a pure unit root test.

4.3 The debate about the role of the prior on ρ

Either in (29) or in (31) the prior on ρ is present and may influence the result of the test. This is clear for (29), but less evident for (31). The second testing strategy was first proposed by Lindley (1965) who recommends to use it under a non-informative prior density. Most Bayesian tests for unit root use a uniform prior on the parameter ρ. A uniform prior has for long and since the initial paper of Bayes (1764, 1970) been taken as a non-informative prior density for location parameters. This point of view is disputed by Phillips (1991) for time series and unit root testing. Bayesian inference proceeds conditionally on the sample moments. In time series analysis, the behavior of the sample moments may be radically different according to the value of ρ. So a prior which states that all values for ρ (those implying stationarity and those implying non-stationarity and explosiveness) are equally likely is thought to be untenable by Phillips (1991).

The question raised by Phillips is in fact deeper. It concerns the applicability of the likelihood principle in the case of dynamic models. Inference on ρ and testing are done conditionally on the observed sample moments or sufficient statistics. Now we know that they depend on the value taken by ρ and that their behavior differs radically when $|\rho| < 1$ and when $\rho = 1$. There is a contradiction in making inference with a posterior density which is a conditional density, when the condition depends on the parameter. Consequently, the simple regression models (1) or (3) having each the same likelihood function irrespective of the value of ρ should be inadequate for testing unit root, as they should favor very much the stationary hypothesis. To circumvent this difficulty, Phillips uses a prior density that reflects the properties of the sampling process. Ignoring some of the classical sampling properties in the likelihood function, he incorporates them in the prior density of ρ. In the Bayesian literature, there is a class of prior densities which are dependent on the sampling process: this is the Jeffreys' diffuse prior. Phillips derives a testing procedure which uses explicitly the Jeffreys' diffuse prior together with the linear reduced form model (3).

A Jeffreys' prior is derived from the information matrix of the model. If θ represents the parameters of the model and $I(\theta)$ the information matrix, then $D'_J(\theta) \propto |I(\theta)|^{1/2}$. The general principle is appealing, but its application causes certain problems, especially in the

[8] if the likelihood function is (26), the upper bound is truncated to $\sqrt{1+v}$.

multivariate case, as we shall see below. To circumvent the difficulty, Jeffreys recommend in the multivariate case, to apply his principle to each parameter separately[9].

Phillips (1991) computes the Jeffreys prior for model (3) with a constant and a trend. The result is quite complicated. But he finally uses a simplified prior[10], as it is evident from his from his formula (22), which is:

$$D_J'(\rho, \beta_0, \beta_1, \sigma^2) \propto \frac{1}{\sigma^3}\left(\frac{1}{1-\rho^2}\left[T - \frac{1-\rho^{2T}}{1-\rho^2}\right]\right)^{1/2} \tag{38}$$

with $y_0 = 0$. This prior goes to infinity for ρ greater than one and at a rate which depends on the sample size T. For $\rho = 1$ it is proportional to $\sqrt{T(T-1)/2}/\sigma^2$.

If on the contrary we derive the Jeffreys' prior of $(\rho, \mu, \delta, \sigma^2)$ for model (1), we find with Schotman and van Dijk (1991c) a result which is different (assuming $y_0 = \mu$):

$$D_J'(\rho, \mu, \delta, \sigma^2) \propto \frac{1-\rho^2}{\sigma^3}\left(\frac{1}{1-\rho^2}\left[T - \frac{1-\rho^{2T}}{1-\rho^2}\right]\right)^{1/2} \tag{39}$$

so that (39) = $(1-\rho^2)\times$ (38). At $\rho = 1$, this prior is equal to zero and thus puts a zero weight on the unit root hypothesis. Consequently a flat prior for ρ favors much more the unit root hypothesis than the Jeffreys' prior when a structural parameterization like (1) is used.

4.4 A non-informative prior for ρ

If we want to define a diffuse prior for ρ in model (1), we have to rely on a different principle than that of Jeffreys. A prior that goes to infinity beyond the point $\rho = 1$ is quite attractive, but at a speed that is independent of the sample size[11]. With (26) the range of ρ is bounded and ρ belongs to the interval $[-\sqrt{1+\upsilon}, \sqrt{1+\upsilon}]$. A convenient prior for a parameter taking values on a finite interval is the Beta density:

$$D'(\rho|\upsilon) \propto \left(\sqrt{1+\upsilon} + \rho\right)^{p-1} . \left(\sqrt{1+\upsilon} - \rho\right)^{q-1} \tag{40}$$

This family is rather rich, as for instance the uniform density corresponds to $p = q = 1$. But it opens the door to another class of non-informative prior defined by Novick (1969): the non-informative limit of an informative prior density, obtained by letting the parameters of the prior go outside their domain of definition. For instance an inverted Gamma prior density for σ^2 has no converging integral if its number of degrees of freedom is lower than

[9] Let us take the simple case of the Normal process with mean μ and variance σ^2. If σ^2 is known, the Jeffreys' principle yields a prior for μ which is identical to that given by Savage's principle of stable estimation. If σ^2 is unknown, the Jeffreys' prior for μ changes and has no longer this property.

[10] The simplification comes from the fact that he uses a Laplace approximation to compute the posterior density.

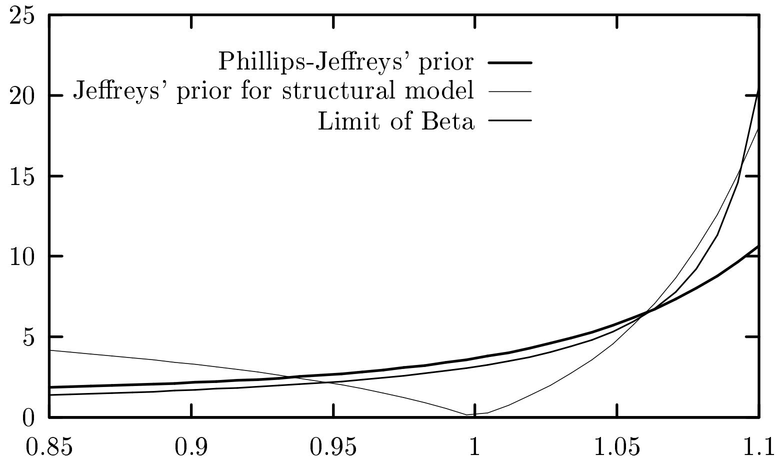
[11] The Jeffreys' prior, with T as an exponent, leads in some cases to inference results which give a totally unreasonable weight to explosive values of ρ , values which are far beyond 1.2.

one. Then two observations suffice to ensure the existence of the posterior density of σ^2 and three the existence of the posterior expectation of σ^2. For the Beta prior (40), the range of the parameters is $p > 0$, $q > 0$. We get a non-informative limit if we let the parameters go outside their range of definition. It suffices to take $p = 1$ and $q = 0$:

$$D'(\rho|v) \propto \left(\sqrt{1+v} - \rho\right)^{-1} \tag{41}$$

This prior has no converging integral or moments. Its shape is similar to that of Jeffreys given in (38). But it is not a function of T and can thus be dominated by the sample[12]. I shall use (41) as a non-informative prior for ρ. Figure 1 compares my prior (41) to the two variants of the Jeffreys' prior (38) and (39). The Jeffreys' priors were computed for $T = 25$ and my prior for $v = 0.25$. For ease of comparison, these priors are normalized to 1 for the range of the graph.

Insert Figure 1 here



We can find in the literature other prior densities for ρ that allow for explosive values and that have not T as an argument. Kadane and Chan (1993) propose a piece-wise informative normal prior with different hyper-parameters below and above $\rho = 1$. Their choice is justified by the fact that the OLS estimator of ρ has very different properties below 1, at 1 and above 1. Berger, as reported in Uhlig (1992), proposes a non-informative prior with a singularity at $\rho = 1$. See the paper of Uhlig for more examples.

5 ANALYSIS OF THE EXTENDED NELSON-PLOSSER DATA SET

The Nelson-Plosser data set has for long been used as a benchmark for various unit root tests [see e.g. Perron (1989), Phillips (1991), ...]. Schotman and van Dijk (1991c) have extended this set of long run annual data[13] up to 1988. The end of the sample proved to be

[12] This prior corresponds to that used in Thornber (1967) for $v = 0$ and $p = q = 0$.
[13] Some series start in 1860!

highly informative in the sense that the conclusion of Nelson and Plosser that all the series of the data set except the unemployment rate were I(1) is now invalidated for three series according to Schotman and van Dijk (1991c). I shall try to illustrate several points with these data. I have reviewed three test procedures:

- the approach proposed in this paper, based on the structural parameterization (1) including a special treatment of the first observation, the degenerated Beta prior (41) on ρ which has a range truncated to 1.22 and a test based on $\Pr(\rho \geq 1|y, y_0)$;
- the approach of Phillips (1991) with the reduced form parameterization (3), the Jeffreys' prior (38) on ρ with no truncation at all and a test based on $\Pr(\rho \geq 1|y)$.
- the approach of Schotman and van Dijk (1991b) with a structural parameterization, fixed initial conditions, but an informative prior on the constant term, a flat prior on ρ which has a range truncated to one and a computation of posterior odds of the hypothesis $\rho = 1$.

To analyze the data set, I selected the number of additional lags with a Schwarz criterion computed with the first model and I added a trend whenever the graph of the data suggested to do so. The results are regrouped into Table 1:

`Insert Table 1 here`

We can note, at a first glance, that the classical and the Bayesian results are in a fair agreement if we adopt a 5% probability level. The classical test rejects the unit root in five cases out of fourteen, my test in six cases, the Phillips' test in seven cases and the Schotman van Dijk's test in eight cases.

`Insert Figure 2 here`

Fig. 2.
Three posterior densities for
the extended Nelson-Plosser data set: series 1 to 4.

`Insert Figure 3 here`

Fig. 3.
Three posterior densities for
the extended Nelson-Plosser data set: series 5 to 8.

`Insert Figure 4 here`

Fig. 4.
Three posterior densities for
the extended Nelson-Plosser data set: series 9 to 12.

`Insert Figure 5 here`

Fig. 5.
Three posterior densities for
the extended Nelson-Plosser data set: series 13 and 14.

A look at the graphs of the posterior densities (figures 2 to 5) is quite instructive. The truncation point for my test is at 1.22 , but is never active. My test and that of Schotman and van Dijk use the same model, but not the same posterior density. The difference comes from the truncation point and the prior for ρ. Truncating and using posterior odds is not anodyne at it leads to a method that rejects the unit root hypothesis more often than the other methods. Note however that a different lower bound for a can change some results as illustrated in table 2:

`Insert Table 2 here`

The first column reproduces the previous results which use a variable a as proposed in Schotman and van Dijk (1992a, b). The second column uses a fixed a equal to 0.80 as in Schotman and van Dijk (1992c). With this new prior, the Nominal GNP becomes integrated of order one, in conformity with the finding of the other tests.

The graphs show that the results of Phillips, obtained with a linear model, rely heavily on the Jeffreys' prior which distorts the posterior density so as to produce a mode very different from the mean every time the other models strongly favor the unit root hypothesis. This is indirectly an illustration of the bias contained in the results of DeJong and Whiteman (1991b).

6 CONCLUSION

Even with a diffuse prior on ρ which is not flat, Bayesian procedures for testing for unit roots may deliver a slightly different message than the classical ones. They may have a slight tendency to give less weight to the unit root hypothesis.

Provided that the model used to describe the data adopts a structural parameterization, the flat prior defamed by Phillips (1991) does not seem to be guilty. Even more, we have shown in Bauwens and Lubrano (1993) that the model used here, coupled with a flat prior, produces a posterior density that is more distant (in the sense of an entropy measure) from the posterior density of DeJong and Whiteman than the posterior of Phillips actually was.

The testing procedure (posterior odds or posterior confidence interval) does not seem also to be so crucial at the light of our empirical results. Out of fourteen series, there are only one or two cases where the results diverge. So empirically, it does not matter to distinguish between testing for a point null hypothesis (unit root) with posterior odds from testing for a one sided hypothesis (non-stationarity).

Finally, we must be conscious that the results obtained may not be very robust if we are interested at qualifying the deep structure[14] of a series. If the data set is stopped in 1970 as the original data set of Nelson and Plosser was, the results are different for a significant number of series and in particular for the much debated real GNP. The test of the unit root hypothesis is also sensitive to the proposed alternative hypothesis. Here under H_1 the series are trend stationary. More refined alternatives can be envisaged as that of broken trends introduced by Perron (1989) and illustrated in a Bayesian framework in Lubrano (1992).

[14] as given for instance by its spectral density.

So what is the use of unit root tests, either classical or Bayesian? Intuitively it is very hazardous to discriminate between two competing economic theories on the basis of a univariate model. However, unit root tests may be helpful when they are used in a more complete modeling strategy as a protection against gross errors, as well as miss- specification tests. On personal grounds, I would of course recommend testing procedures which are Bayesian.

REFERENCES

Bauwens, Luc and Michel Lubrano, 1993, Trends and breaking points of the Bayesian econometric literature, *Document de travail* 93A06 (GREQE, Marseille).

Bayes, Thomas, 1764, 1970, An essay toward solving a problem in the doctrine of chances. Reprinted in E.S. Pearson and M.G.Kendall, eds., *Studies in the History of Statistics and Probability* (Hafner, New York).

Davidson, Russell and James G. MacKinnon, 1993, *Estimation and Inference in econometrics* (Oxford, Oxford University Press).

DeJong, David N. and Charles H. Whiteman, 1989, Trends and cycles as unobserved components in US real GNP: a Bayesian perspective, *Journal of the American Statistical Association Papers and Proceedings*, 63-70.

DeJong, David N. and Charles H. Whiteman, 1991a, The temporal stability of dividends and stock prices: evidence from the likelihood function, *American Economic Review* 81, 600-617.

DeJong, David N. and Charles H. Whiteman, 1991b, Reconsidering "Trends and random walks in macro economic time series", *Journal of Monetary Economics* 28, 221-254.

Fuller, Wayne A., 1976, *Introduction to statistical time series* (Wiley, New-York).

Judge, George G., W.E. Griffiths, R.Carter Hill, Helmut Lütkepohl and Tsoung Chao Lee, 1985, *The theory and practice of econometrics* (Wiley, New-York).

Kadane, Joseph B. and Ngai Hang Chan, 1993, Priors for unit root models, *mimeo*, Carnegie Mellon University, Pittsburgh.

Koop, Gary and Mark Steel, 1992a, A decision theoretic analysis of the unit root hypothesis using mixtures of elliptical models, *mimeo*, (Tilburg University).

Koop, Gary, Jacek Osiewalski and Mark Steel, 1992b, Bayesian long-run prediction in time series models, *Journal of Econometrics, this issue* ??, ???-???.

Lindley, Dennis V., 1957, A statistical paradox, *Biometrika* 44, 187-192.

Lindley, Dennis V., 1965, *Introduction to probability and statistics from a Bayesian viewpoint* (Cambridge University Press, Cambridge).

Lubrano, Michel, 1991, Testing for unit roots and co-integration in a Bayesian framework, DP 9103 (IRES, Louvain-la-Neuve).

Lubrano, Michel, 1992, Bayesian tests for single equation co-integration in the case of structural breaks, DP 92-10 (UCSD, San Diego).

Nelson, Charles R. and Charles I. Plosser, 1982, Trends and random walks in macro-economic time series: some evidence and implications, *Journal of Monetary Economics* 10, 139-162.

Novick, Melvin, 1969, Multi-parameter Bayesian indifference procedures, *Journal of the Royal Statistical Society* B 31, 29-51.

O'Brien, R.J., 1970, Serial correlation in econometric models in: K. Hilton and D.F. Heathfield eds., *The Econometric Study of the United Kingdom* (MacMillan, London) 375-437.

Perron, Pierre, 1989, The great crash, the oil price shock and the unit root hypothesis, *Econometrica* 57, 1361-1401.

Phillips, Peter C.B., 1991, To criticize the critics: an objective Bayesian analysis of stochastic trends, *Journal of Applied Econometrics* 6, 333-364.

Richard, Jean-François, 1977, Bayesian analysis of the regression model when the disturbances are generated by an auto-regressive process, in: A. Aykaç and C. Brumat, eds., *New Developments in the Application of Bayesian Methods* (North Holland, Amsterdam) 185-209.

Schotman, Peter, 1992, Objective information and ignorance: Notes on some priors in the Bayesian analysis of the AR(1) model, *mimeo*, Princeton University.

Schotman, Peter and Herman K. van Dijk, 1991a, A Bayesian analysis of the unit root in real exchange rates, *Journal of Econometrics* 49, 195- 238.

Schotman, Peter and Herman K. van Dijk, 1991b, Posterior analysis of possibly integrated time series with an application to real GNP, *to appear* in: E. Parzen et al. eds, *New Directions in Time Series Analysis*, IMA Volume in Mathematics and its Applications (Springer Verlag, New-York).

Schotman, Peter and Herman K. van Dijk, 1991c, On Bayesian routes to unit roots, *Journal of Applied Econometrics* 6, 387-401.

Sims, Christopher A., 1988, Bayesian skepticism on unit root econometrics, *Journal of Economic Dynamics and Control* 12, 463-474.

Thornber H., 1967, Finite sample Monte Carlo studies: an autoregressive illustration, *Journal of the American Statistical Association*, 801-818.

Tishler, Asher and Israel Zang, 1981, A new maximum likelihood algorithm for piecewise regression, *Journal of the American Statistical Association* 76, 980-987.

Uhlig, Harald, 1992, What macro-economists should know about unit roots as well: the Bayesian perspective, *Research Memorandom No 367*, Princeton University.

Zellner, Arnold, 1971, *An introduction to Bayesian inference in econometrics* (Wiley, New-York).

Zellner, Arnold and George C. Tiao, 1964, Bayesian analysis of the regression model with auto-correlated errors, *Journal of the American Statistical Association* 59, 763-778.

Zivot, Eric, 1992, Bayesian analysis of the unit root hypothesis within an unobserved component model, *mimeograph.*

Table 1

Unit root tests on the extended Nelson-Plosser data set

Series	Start	p	trend	ADF	This paper $\Pr(\rho \geq 1 \mid y)$	Phillips $\Pr(\rho \geq 1 \mid y)$	S-vD $\Pr(\rho = 1 \mid y)$	
Real GNP	1900	2	yes	-3.52*	0.031*	0.002*	0.21*	(.73)
Nominal GNP	1909	2	yes	-2.06	0.149	0.066	0.49*	(.89)
Real GNP per cap.	1909	2	yes	-3.59*	0.029*	0.002*	0.19*	(.72)
Industrial prod.	1860	2	yes	-3.62*	0.015*	0.004*	0.13*	(.71)
Employment	1890	2	yes	-3.47*	0.012*	0.003*	0.11*	(.75)
Unemployment rate	1890	4	no	-4.04*	0.001*	0.001*	0.01*	(.54)
GNP deflator	1889	2	yes	-1.62	0.267	0.133	0.60	(.95)
Consumer prices	1860	4	yes	-1.22	0.231	0.185	0.56	(.98)
Wages	1900	2	yes	-2.40	0.104	0.029*	0.43*	(.89)
Real wages	1900	2	yes	-1.71	0.225	0.292	0.58	(.87)
Money stock	1889	2	yes	-2.91	0.014*	0.007*	0.14*	(.89)
Velocity	1869	2	yes	-1.62	0.486	0.211	0.66	(.94)
Bond yield	1900	4	no	-1.35	0.208	0.240	0.53	(.91)
Stock prices	1871	2	yes	-2.44	0.214	0.063	0.57	(.88)

* indicates when the unit root hypothesis is rejected at the 5% probability level.

The last column under S-vD indicates the lower bound a of the uniform prior on ρ used by Schotman and van Dijk. It was here selected as $a = \mathrm{E}(\rho|y) - 2.75 \times Sd(\rho|y)$.

Table 2

Sensitivity of posterior odds to the prior on ρ

	S-vD adjusted prior		S-vD fixed prior	
Series	$\Pr(\rho = 1 \mid y)$		$\Pr(\rho = 1 \mid y)$	
Real GNP	0.21*	(.73)	0.16*	(.80)
Nominal GNP	0.49*	(.89)	0.63	(.80)
Real GNP per cap.	0.19*	(.72)	0.15*	(.80)
Industrial prod.	0.13*	(.71)	0.10*	(.80)
Employment	0.11*	(.75)	0.09*	(.80)
Unemployment rate	0.01*	(.54)	0.00*	(.80)
GNP deflator	0.60	(.95)	0.85	(.80)
Consumer prices	0.56	(.98)	0.92	(.80)
Wages	0.43*	(.89)	0.59	(.80)
Real wages	0.58	(.87)	0.68	(.80)
Money stock	0.14*	(.89)	0.22*	(.80)
Velocity	0.66	(.94)	0.86	(.80)
Bond yield	0.53	(.91)	0.70	(.80)
Stock prices	0.57	(.88)	0.68	(.80)