Unit Roots: Bayesian Significance Test

M. Diniz [a] , C. A. B. Pereira [b] & J. M. Stern [c]

[a] Department of Statistics, Universidade Federal de Sao Carlos, Sao Carlos, Brazil

[b] Department of Statistics, Universidade de Sao Paulo, Sao Paulo, Brazil

[c] Department of Applied Mathematics, Universidade de Sao Paulo, Sao Paulo, Brazil

Available online: 03 Nov 2011

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis
Taylor & Francis Group

Unit Roots: Bayesian Significance Test

M. DINIZ[1], C. A. B. PEREIRA[2], AND J. M. STERN[3]

[1]Department of Statistics, Universidade Federal de Sao Carlos,
Sao Carlos, Brazil
[2]Department of Statistics, Universidade de Sao Paulo, Sao Paulo, Brazil
[3]Department of Applied Mathematics, Universidade de Sao Paulo,
Sao Paulo, Brazil

The unit root problem plays a central role in empirical applications in the time series econometric literature. However, significance tests developed under the frequentist tradition present various conceptual problems that jeopardize the power of these tests, especially for small samples. Bayesian alternatives, although having interesting interpretations and being precisely defined, experience problems due to the fact that that the hypothesis of interest in this case is sharp or precise. The Bayesian significance test used in this article, for the unit root hypothesis, is based solely on the posterior density function, without the need of imposing positive probabilities to sets of zero Lebesgue measure. Furthermore, it is conducted under strict observance of the likelihood principle. It was designed mainly for testing sharp null hypotheses and it is called FBST for Full Bayesian Significance Test.

1. Introduction

Testing precise or sharp hypotheses, particularly in nonenumerable parameter spaces, has been one of the major difficulties in the field of statistics in either the frequentist or Bayesian paradigm. We refer to the following as important situations: (i) the need for nuisance parameter elimination, (Basu, 1977); and (ii) Lindley paradox, (Lindley, 1957). A hypothesis is called sharp if its dimension is smaller than the dimension of the whole parameter space.

The Full Bayesian Significance Test, FBST, developed by Pereira and Stern (1999) was designed mainly to deal properly with sharp hypotheses. For a recent review, see Pereira et al. (2008). For the absolutely continuous case, the FBST is solely based on posterior densities without the need of one of the following

practices: elimination of nuisance parameters or adoption of positive probabilities in zero Lebesgue measure sets.

Challenged by econometrists, the authors applied the FBST to an important Time Series Econometric problem: inference on unit roots. The frequentist alternative tests may have questionable interpretations and other known problems[1]. The results obtained by this application of the FBST seem to be quite suitable and convincing. The FBST is compared with both Bayesian and frequentist alternative tests.

Section 2 describes briefly the FBST and discusses how to implement it computationally. Section 3 discusses the Bayesian alternatives already known in the literature. Section 4 presents numerical examples of the literature and compares the FBST with its alternatives. Section 5 is a discussion on some deeper issues and fine details.

2. FBST

The FBST was introduced by Pereira and Stern (1999). It was created mainly to test sharp hypotheses which is a matter of discussion and controversies. This article assumes that one accept and is interested in testing sharp hypotheses.

Let us now consider general statistical spaces, where the parameter space is $\Theta \subset \mathfrak{R}^m$ and the sample space $X^k \subset \mathfrak{R}$. A sharp hypothesis H states that θ belongs to a sub-manifold Θ_H of smaller dimension than Θ. The subset Θ_H has null Lebesgue measure whenever H is sharp.

In the FBST construction the posterior probability density on the parameter space is used as an ordering system and all sets of the same nature are dealt with accordingly in the same way. As a consequence, the sets that define sharp hypotheses keep having null probabilities. Instead of changing the nature of H by assigning positive probability to it, we consider the tangential set T of points having posterior density values higher than any θ in Θ_H. We then reject H if the posterior probability of T is large. We will formalize these ideas in the sequel.

Let us consider a standard parametric statistical model: $\theta \in \Theta \subset \mathfrak{R}^m$ is the parameter, $g(\cdot)$ a probability prior density over Θ, x is the observation (a scalar or a vector), and $L_x(\cdot)$ is the likelihood generated by data x. Posterior to the observation of x, the sole relevant entity for the evaluation of the Bayesian evidence, ev, is the posterior probability density for θ given x, denoted by

$$g_x(\theta) = g(\theta \,|\, x) \propto g(\theta) L_x(\theta).$$

We are of course restricted to the case where the posterior probability distribution over Θ is absolutely continuous, that is, $g_x(\theta)$ is a density over Θ. For simplicity, we may use H for Θ_H in the sequel. Now, let $r(\theta)$ be a reference density on Θ such that the function $s(\theta) = g_x(\theta)/r(\theta)$ is called the "relative surprise"[2].

Definition 2.1 (Evidence). Consider a sharp hypothesis $H : \theta \in \Theta_H$ and let

$$s^* = \sup_{\theta \in H} s(\theta) \tag{1}$$

---

[1]See Phillips and Xiao (1998).
[2]See Good (1983).

and

$$T = \{\theta \in \Theta : s(\theta) > s^*\}. \tag{2}$$

The Bayesian evidence value against H is defined as the posterior probability of the tangential set, i.e., $\overline{ev} = Pr\{\theta \in T \mid x\} = \int_T g_x(\theta)d\theta$.

Notice that the tangential set T is the highest relative surprise set. It is the set of points $\theta \in \Theta$ with higher relative surprise $s(\theta)$ than any point in H. Therefore, the set is "tangential" to H. This approach does not exclude or avoid the model considered in the hypothesis being tested but just uses the concept of "tangent" to define an evidential measure favoring the hypothesis.

One must also note that the evidence value supporting H, $ev = 1 - \overline{ev}$, is not evidence against A, the alternative hypothesis (which is not sharp anyway). Equivalently, $\overline{ev}$ is not evidence in favor of A, although it is against H.

Definition 2.2 (TEST). The FBST (Full Bayesian Significance Test) is the procedure that rejects H whenever $ev = 1 - \overline{ev}$ is smaller than a critical level, ev_c.

Being a statistic, ev has a sampling distribution. For well-behaved likelihood and posterior densities[3], Pereira et al. (2008) showed that, asymptotically, the evidence follows a χ^2 distribution with degrees of freedom given by the dimension of the parameter space. This fact gives a way to define, at least asymptotically, a critical level to reject the hypothesis being tested.

A major practical issue for the use of the FBST is the determination of the critical level. Ev being a statistic defined on a zero to one scale does not ease the matter (the same occurs with p-values). The formal identification of the FBST as a Bayes test of hypothesis yields critical values derived from loss functions allowing this identification. In fact, Madruga et al. (2001) showed that there are loss functions the minimization of which makes ev a Bayes estimator of $\phi = I(\theta \in H)$. Hence, the FBST is in fact a Bayes procedure in the formal sense of Wald (1950).

By using a reference density in the definition of the tangential set T, the FBST formulation above presented is explicit invariant under general coordinate transformations of the parameter space[4]. For the FBST application on unit root tests discussed in the sequel, we will use the (improper) uniform density as reference density on Θ. Madruga et al. (2003) remarked that it is possible to generalize the procedure using other reference densities such as neutral or reference priors if one is available.

2.1. *Numerical Calculus*

The evidence calculus supporting H defined in the last section is performed numerically in two steps. The first one involves the optimization of $s(\theta)$ under H and, the second one, the integration of the posterior, $g_x(\theta)$, over T.

The optimization step consists of finding the parameter space point in H that maximizes $s(\theta)$. It is, therefore, a maximization under constraint problem:

$$\theta^* = \arg\max_{\theta \in \Theta_H} s(\theta), \quad s^* = s(\theta^*)$$

---

[3]See (Schervish, 1995, p. 436).
[4]See the Appendix.

To solve this problem, we use a numerical optimizer. To calculate the integral, it is possible to use various numerical techniques. We introduce a method based on Laplace approximation and Monte Carlo techniques that were able to deal with the majority of the problems under discussion.

Let θ be the parameter vector and x the observations vector as above. The posterior distribution is given by:

$$g(\theta \mid x) = \frac{g(\theta)L_x(\theta)}{\int_\Theta g(\theta)L_x(\theta)d\theta}.$$

To calculate the *e*-value we need to integrate the posterior over the tangential set, i.e., $T = \{\theta \in \Theta : s(\theta) \geq s^*\}$:

$$\int_{\{\theta \in \Theta \mid s(\theta) \geq s^*\}} g_x(\theta)d\theta = \frac{\int_T g(\theta)L_x(\theta)d\theta}{\int_\Theta g(\theta)L_x(\theta)d\theta}. \tag{3}$$

One way to approximate integrals like the denominator above is to use the Laplace approximation. Consider the integral

$$I = \int_\Theta b(\theta)\exp[-Nh(\theta)]d\theta$$

in which N is the sample size, θ is $(k \times 1)$, $\Theta = \mathcal{R}^k$, and $-h(\cdot)$ is a twice differentiable function with only one maximum in $\hat{\theta}$, $\partial h(\theta)/\partial\theta|_{\theta=\hat{\theta}} = 0$, and $H(\theta) = \frac{\partial^2 h(\theta)}{\partial\theta\partial\theta'}$ is positive definite. Furthermore, $b(\cdot)$ is continuous on the neighborhoods of $\hat{\theta}$ with $b(\hat{\theta}) \neq 0$. Expanding $h(\theta)$ as a second-order Taylor series for $\hat{\theta}$ we have an approximation of $\exp[-Nh(\theta)]$ proportional to a normal density. By doing the same with $b(\theta)$ we arrive at the following approximation for the above integral:

$$\widehat{I} = (2\pi)^{k/2}b(\hat{\theta})|NH(\hat{\theta})|^{-1/2}\exp[-Nh(\hat{\theta})]$$

since the $O(N^{-1/2})$ terms from the expansions of $b(\theta)$ and $h(\theta)$ disappear when we integrate.

Now we use the Tierney and Kadane (1986) method to calculate (3). Let us consider $b(\theta) = 1$ and the restriction $\exp[-Nh(\theta)] = g(\theta)L_x(\theta)$. If $h(\theta)$ satisfy the conditons given above, we have that the value of (3) is approximated by:

$$= \frac{\exp[-Nh(\hat{\theta})]\int_{T(\theta)}\exp\left[-\frac{N}{2}(\theta-\hat{\theta})'H(\hat{\theta})(\theta-\hat{\theta})\right]d\theta}{\exp[-Nh(\hat{\theta})](2\pi)^{k/2}|NH(\hat{\theta})|^{-1/2}}$$

$$= (2\pi)^{-k/2}|NH(\hat{\theta})|^{1/2}\int_{T(\theta)}\exp\left[-\frac{N}{2}(\theta-\hat{\theta})'H(\hat{\theta})(\theta-\hat{\theta})\right]d\theta.$$

The last expression is the integral over the tangential set of the θ multivariate normal density with mean $\hat{\theta}$ and variance $(NH(\hat{\theta}))^{-1}$. Therefore, to evaluate the integral we can generate a large number of vectors with this distribution and evaluate the posterior with these vectors. The proportion of them that belongs to T is the approximate value for (3).

3. Bayesian Unit Root Tests

Bayesians seem to have started the investigation of unit root problems in the late 1980s. Sims (1988) and Sims and Uhlig (1991), as far as we know, published the first Bayesian articles to carefully deal with the unit root problem of hypothesis testing. The frequentist critics of these works received a proper answer in Phillips (1991a,b). This was the starting point of a fruitful debate that generates a long list of articles in the literature of Bayesian Time Series.

The present section introduces the main Bayesian procedures that have been used to test unit roots. The notation is of Bauwens et al. (1999), who also presented a great summary of Bayesian articles on the subject.

Consider the model:

$$y_t = \mu + \delta t + (1 - \rho)y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + \varepsilon_t \tag{4}$$

with ε_t i.i.d. $N(0, \sigma^2)$ for $t = 1, \ldots, T$. This is a complete model tested by the ADF. Being $\theta = (\rho, a^*)$ the parameters vector, in which $\rho = \sum_{i=1}^{p} \rho_i$ and $a^* = (\mu, \delta, \Gamma_1, \ldots, \Gamma_{p-1})$, and assuming σ^2 fixed, the prior density of θ can be factorized as

$$p(\theta) = p(\rho)p(a^* \mid \rho).$$

The marginal likelihood for ρ is:

$$l(\rho \mid D) \propto \int l(\theta|D)p(a^* \mid \rho)da^*,$$

where D is the observations vector. This function, associated with a prior for ρ is the main ingredient used by standard Bayesian procedures to test the existence of unit roots. Even though the procedure varies among authors according to some specfic aspects, mentioned below, basically all of them use Bayes factors and posterior probabilities.

One issue is about the specification of the null hypothesis: some authors, starting from Schotman and van Dijk (1991), consider $H_0 : \rho = 1$ against $H_1 : \rho < 1$. This is the way the frequentist school addresses the problem, as do Dickey and Fuller (1979). The main problem is that no explosive value for ρ is considered. The problem is solved by the standard Bayesian approach using the posterior probabilities ratio:

$$B_{01} = \frac{l(\rho = 1 \mid D)}{\int_0^1 l(\rho \mid D)p(\rho)d\rho}.$$

It is said that one of the advantages of this approach is that the null and the alternative hypotheses are treated equally weighted: with equal probabilities. However, the expression above is not defined if $p(\rho)$ is not a proper density since the Bayes factor denominator is equal to the predictive density, defined just if $p(\rho)$ is a proper density. There are also problems if $l(\rho = 1 \mid D)$ is zero or infinite.

Other authors like Phillips (1991a) and Lubrano (1995) considered the problem as being the test of $H_0 : \rho \geq 1$ against $H_1 : \rho < 1$, considering explicitly explosive

values for ρ. The main advantage of this approach is the possibility to calculate posterior probabilities like

$$\mathrm{P}(\rho > 1 \,|\, D) = \int_1^\infty p(\rho \,|\, D) d\rho$$

defined to whatever prior over ρ. Some authors, like DeJong and Whiteman (1991), did not choose ρ as the parameter of interest. They examine the largest value in module of the roots of the characteristic equation concerning the autoregressive part of the model, i.e.,

$$1 - \sum_{i=1}^p \rho_i L^i = 0,$$

and verify if it is smaller or bigger than one. Usually, this value is slightly smaller than ρ. The authors argue that this small difference can be important. When this approach is used, unit roots are found less frequently. For an AR(3) model with a constant and trend, DeJong and Whiteman (1991) derived the posterior density for the dominant root for the 14 series used by Nelson and Plosser (1982) and concluded the following: for 11 of them, the dominant root was smaller than one, that is to say, the series were trend-stationary. These results were based on a flat prior for the autoregressive parameters and the deterministic trend coefficient.

Another controversy is about the prior over ρ. Phillips (1991a) argued that the difference between the results given by the frequentist and Bayesian inferences is due to the fact that the flat prior proposed by Sims (1988) overweights the stationary region. Hence, he derived a Jeffreys prior for the AR(1) model: this prior quickly tends to infinite as ρ raises and becomes bigger than one. The obtained posterior produces the same results as Nelson and Plosser (1982). The next section discusses these results in detail. The critics of Phillips approach[5] judged the Jeffreys prior as unrealistic, from a subjective point of view.

A final controversial point concerns the modeling of the initial observations. If they are directly included in the likelihood, implicitly, the process is considered stationary. In fact, when one is sure about it and it is believed that the data generating process is working for a long period, it is reasonable to assume that the dynamic model parameters determine the marginal distribution of the initial observations. In the simplest AR(1) model, this would imply that $y_1 \sim N(0, \sigma^2/(1 - \rho^2))$. In this case, to perform the inference conditional on the first observation would discard relevant information. On the other hand, in a non stationary model there is no marginal distribution defined for y_1. Then, it is valid to conditionate on initial oservations. For the models presented here, we always work with the conditional likelihood – conditional on the initial observations. As argued by Sims (1988), inferences for stationary models are little affected by considering these conditional likekelihoods, especially for large samples. He compared these inferences with the ones based on the exact likelihoods under explicit modeling for intial observations.

---

[5]See Bauwens et al. (1992, p. 162) and comments about Phillips (1991a).

4. Applications

Now we show how to implement the FBST for unit roots. First, we describe the way the procedure is implemented to test unit roots in a real data set and then perform simulations to compare the power of the two tests: FBST and ADF.

Consider a time series generated by an AR(p) process with constant and deterministic trend:

$$y_t = \mu + \delta t + \rho_1 y_{t-1} + \cdots + \rho_p y_{t-p} + \varepsilon_t, \tag{5}$$

with and $\varepsilon_t \sim N(0, \sigma^2)$ $\forall t = 1, \ldots, T$. Alternatively, the process can be written as:

$$y_t = \mu + \delta t + \Gamma_0 y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + \varepsilon_t, \tag{6}$$

where $\Delta y_t = y_t - y_{t-1}$, $\Gamma_0 = \rho_1 + \cdots + \rho_p - 1$, and $\Gamma_i = -\sum_{j=i+1}^{p} \rho_j$, for $i = 1, \ldots, p - 1$. So, the series has a unit root if $\Gamma_0 = 0$.

Considering $\theta = (\mu, \delta, \Gamma_0, \ldots, \Gamma_{p-1})$ and $Y_p = (y_1, \ldots, y_p)$ the first p observations, we have that the likelihood function is

$$f(\varepsilon_p, \ldots, \varepsilon_T \mid \theta, \sigma, Y_p) = (2\pi)^{-T/2} \sigma^{-T} \exp\left\{-(1/2\sigma^2) \sum_{t=p+1}^{T} \varepsilon_t^2\right\}, \tag{7}$$

in which $\varepsilon_t = \Delta y_t - \mu - \delta t - \Gamma_0 y_{t-1} - \Gamma_1 \Delta y_{t-1} - \cdots - \Gamma_{p-1} \Delta y_{t-p+1}$.

Assuming a flat prior for $(\theta, \log\sigma)$ we obtain the following non informative prior for (θ, σ):

$$\pi(\theta, \sigma) \propto 1/\sigma. \tag{8}$$

We are aware of the problems surrounding this kind of prior applied to this problem, as mentioned by Bauwens et al. (1999). However, one of our goals here is to show how the FBST can be implemented even for an improper and controversial prior like this one. To write the joint posterior we use the following notation:

$$Y = \begin{pmatrix} \Delta y_{p+1} \\ \Delta y_{p+2} \\ \vdots \\ \Delta y_T \end{pmatrix}$$

$$X = \begin{bmatrix} 1 & 1 & y_p & \Delta y_p & \ldots & \Delta y_2 \\ 1 & 2 & y_{p+1} & \Delta y_{p+1} & \ldots & \Delta y_3 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & T & y_{T-1} & \Delta y_{T-1} & \ldots & \Delta y_{T-p+1} \end{bmatrix}$$

$$\theta = \begin{pmatrix} \mu \\ \delta \\ \Gamma_0 \\ \vdots \\ \Gamma_{p-1} \end{pmatrix}$$

being Y of dimension $(T - p \times 1)$, X, $(T - p \times p + 2)$, and θ, $(p + 2 \times 1)$.

Using this matrix notation, the sum of squared errors is given by $(Y - X\theta)'(Y - X\theta)$. By the fact that the least squares estimator of θ is given by $\hat{\theta} = (X'X)^{-1}X'Y$ and the predicted values by $\widehat{Y} = X\hat{\theta}$, it is possible to observe that:

$$(Y - X\theta)'(Y - X\theta) = (Y - \widehat{Y})'(Y - \widehat{Y}) + (\theta - \hat{\theta})'X'X(\theta - \hat{\theta}). \quad (9)$$

Equation (9) is useful to write the joint posterior:

$$f(\theta, \sigma \mid Y, Y_p) \propto \sigma^{-(T+1)}\exp\left\{-\frac{1}{2\sigma^2}[(Y - \widehat{Y})'(Y - \widehat{Y}) + (\theta - \hat{\theta})'X'X(\theta - \hat{\theta})]\right\}. \quad (10)$$

This is the Normal-Gamma density that simplifies the use of the Gibbs sampler.

Using the above posterior and the uniform density as reference density on the parameter space[6], we test for unit roots 14 U.S. macroeconomic time series first mentioned in Nelson and Plosser (1982). Here, we use the extended series of Schotman and van Dijk (1991).

The following table lists the FBST e-values and the ADF *p*-values for the aforementioned time series. We have used the computer procedure described in MacKinnon (1994) to find the ADF *p*-values. In order to obtain comparable results, we follow the specification given by Bauwens et al. (1999) for all models.

As can be seen from the posterior expression, the conditional posteriors are $\pi(\theta|\sigma, Y, Y_p) \propto N(\hat{\theta}, \sigma^2 V)$ and $\pi(1/\sigma^2|\theta, Y, Y_p) \propto \Gamma\left(\frac{T-p+3}{2}, B\right)$, where $B = 0.5(Y - \widehat{Y})'(Y - \widehat{Y}) + (\theta - \hat{\theta})'X'X(\theta - \hat{\theta})$ and $V = (X'X)^{-1}$. For the FBST computations, various solvers can be used in the optimization step, as those developed by Birgin et al. (2004), Corana et al. (1987), or Goffe et al. (1994). For the integration step we used standard Monte Carlo sampling, see Lauretto et al. (2003)[7].

Table 1 shows that the nonstationary posterior probabilities are quite distant from the ADF *p*-values. These results were highlighted by Sims (1988) and Sims and Uhlig (1991). Considering the simplest AR(1) model, they argued that, once classical inference is based on the distribution of $\hat{\rho} \mid \rho = 1$, it reaches counterintuitive conclusions because the referred distribution is skewed. Their argument is as follows: Bayesian inference uses the distribution which is not skewed.

As said before, Phillips (1991a) claimed that the difference in results between classical and Bayesian approaches is due to the flat prior that puts much weight on the stationary region. He proposed the use of Jeffreys priors, which restored the conclusions drawn by the classical test. Phillips argued that the flat prior was, actually, informative when used in time series models like those for unit root tests. Using simulations he shows that:

" [the use of a] flat prior has a tendency to bias the posterior towards stationarity. ... even when [the estimator] is close to unity, there may still be a non negligible downward bias in the [flat] posterior probabilities".

Tables 2 and 3 display some maximum likelihood estimators and the respective standard errors assuming unit roots. Tables 4 and 5 show the maximum likelihood estimators for the same series for the unrestricted model. Tables 6 and 7 give the number of series which rejected the unit root hypothesis in 100 generated samples

---

[6]Our reference density is, therefore, the improper density, $r(\theta, \sigma) \propto 1$.

[7]In the simulations described below, we combined Monte Carlo sampling with the Laplace approximation techniques described in Sec. 2 to perform the integration step.

Table 1
Unit root tests for Nelson and Plosser data

Series	Start	p	Trend	ADF	p-value	$P(\Gamma_0 \geq 0 \mid Y)$	e-value
Real GNP	1909	2	yes	-3.52	0.044	0.0005	0.040
Nominal GNP	1909	2	yes	-2.06	0.559	0.0238	0.523
Real GNP per capita	1909	2	yes	-3.59	0.037	0.0004	0.034
Industrial prod.	1860	2	yes	-3.62	0.032	0.0003	0.028
Employment	1890	2	yes	-3.47	0.048	0.0004	0.043
Unemployment rate	1890	4	no	-4.04	0.019	0.0001	0.020
GNP deflator	1889	2	yes	-1.62	0.778	0.0584	0.762
Consumer prices	1860	4	yes	-1.22	0.902	0.1154	0.983
Nominal wages	1900	2	yes	-2.40	0.377	0.0106	0.341
Real wages	1900	2	yes	-1.71	0.739	0.0475	0.715
Money stock	1889	2	yes	-2.91	0.164	0.0029	0.147
Velocity	1869	2	yes	-1.62	0.779	0.0620	0.777
Bond yield	1900	4	no	-1.35	0.602	0.0962	0.936
Stock prices	1871	2	yes	-2.44	0.357	0.0103	0.349

Table 2
MLE under $H_0 : \Gamma_0 = 0$

Parameters	Real GNP	Ind. Prod.	GNP def.	Wage
μ	0.01543	0.049427	0.00187	0.01494
δ	0.00011	-0.00014	0.00027	0.00020
Γ_1	0.33146	0.03636	0.44992	0.46687
σ	0.05558	0.09682	0.04364	0.05545

assuming that there was (Table 6) or not (Table 7) a unit root. We used three criteria to reject the hypothesis: the ADF asymptotic p-value for 5% significance, the exact ADF p-value for 5% significance and the 5% asymptotical, as described in Sec. 2.

It is important to remember that finite sample critical values for unit root tests depend on the assumption that the error terms are $N(0, \sigma^2 I)$. Recall that these simulated results were generated using this assumption. The asymptotic critical values are valid much more generally, since they do not require normality or homoskedasticity. Therefore, for small samples, it is safer to rely on asymptotic critical values.

Table 3
Standard error of MLE under $H_0 : \Gamma_0 = 0$

Parameters	Real GNP	Ind. Prod.	GNP def.	Wage
μ	0.01320	0.01806	0.00902	0.01247
δ	0.00028	0.00024	0.00016	0.00024
Γ_1	0.10895	0.08966	0.09163	0.09661

Table 4
MLE – unrestricted model

Parameters	Real GNP	Ind. Prod.	GNP def.	Wage
μ	0.81849	0.05221	0.09086	0.39792
δ	0.00567	0.00718	0.00112	0.00309
Γ_0	-0.17631	−0.17658	-0.03164	-0.06494
Γ_1	0.41106	0.12432	0.46979	0.50130
σ	0.05193	0.09252	0.04329	0.05392

Table 5
Standard error of MLE – unrestricted model

Parameters	Real GNP	Ind. Prod.	GNP def.	Wage
μ	0.23279	0.01727	0.05667	0.16301
δ	0.00163	0.00206	0.00056	0.00125
Γ_0	0.05104	0.04941	0.01990	0.02756
Γ_1	0.10436	0.08915	0.09175	0.09522

Table 6
Simulated series rejecting H_0 in 100 generated assuming H_0

Series	$<ADF_{5\%}(\infty)$	$<ADF_{5\%}(ex.)$	$ev < 0.091$
Real GNP	4	4	7
Ind. Prod.	7	5	10
GNP def.	6	6	8
Wage	2	2	5

Table 6 shows that the FBST, even using the flat prior, has a power similar to the ADF test. Hence, the argument used by Phillips to criticize conclusions based on posterior probabilities, under flat priors, does not hold for the FBST.

We perform more numerical simulations to compare the ADF and the FBST powers. The exercise was the following. After simulating 1,000 series with the data

Table 7
Simulated series rejecting H_0 in 100 generated assuming the unrestricted model

Series	$<ADF_{5\%}(\infty)$	$<ADF_{5\%}(ex.)$	$ev < 0.091$
Real GNP	84	83	94
Ind. Prod.	86	81	95
GNP def.	19	19	25
Wage	22	21	38

Table 8

Number of times in which $H_0 : \rho = 1$ was rejected for the
1,000 series generated by (11)

Parameters	ADF	$Ev < 0.144$	$Ev < 0.145$
$\rho = 1$	42	50	50
$\rho = 0.99$	65	63	63
$\rho = 0.975$	83	84	84
$\rho = 0.95$	139	132	132

Table 9

Number of times in which $H_0 : \rho = 1$ was rejected for the
1,000 series generated by (12) $- \mu = 0.5$

Parameters	ADF	$Ev < 0.15$	$Ev < 0.112$
$\rho = 1$	10	50	30
$\rho = 0.99$	33	101	73
$\rho = 0.975$	65	188	145
$\rho = 0.95$	110	300	229

generating processes:

$$y_t = \rho y_{t-1} + \varepsilon_t \tag{11}$$

$$y_t = \mu + \rho y_{t-1} + \varepsilon_t \tag{12}$$

$$y_t = \mu + \delta t + \rho y_{t-1} + \varepsilon_t, \tag{13}$$

and 50 observations each, we calculate the ADF p-values and the FBST e-values. To decide in favor or against the hypothesis, we used the ADF 5% significance level for samples of 50 observations. For the FBST we used one level defined empirically, i.e., the e-value for the fifth percentile when the hypothesis is true (second column) and the asymptotical 5% level described in Sec. 2 (third column). Tables 8–10 summarize the results.

The ADF and FBST have a similar power for the model without deterministic terms. For either models, with only constant or with both constant and deterministic

Table 10

Number of times in which $H_0 : \rho = 1$ was rejected for the
1,000 series generated by (13) $- \mu = 0.5$, $\delta = 0.02$

Parameters	ADF	$Ev < 0.137$	$Ev < 0.099$
$\rho = 1$	22	50	31
$\rho = 0.99$	46	221	165
$\rho = 0.975$	55	270	196
$\rho = 0.95$	55	293	232

trend, the FBST has a better performance even if we consider the statistic asymptotic levels.

5. Concluding Remarks

In the past few decades, the econometrics literature presented tests to identify unit roots. The greatest advances have been made by the frequentist tradition and just after the 1990's, the Bayesian approach on this topic advanced and presented interesting alternatives.

However, Bayes factor tests for unit roots have many difficulties to deal with time series. To overcome these dificulties, various alternative Bayes factor tests have been proposed but their performances are still in question[8]. There have also been specially designed priors that show better performance. However, the use of these priors departs from some basic paradigms of Bayesian statistics, such as the likelihood principle. Moreover, these techniques have to be fine tuned to each particular problem type or application.

Frequentist tests depend on hypotheses about the models used on the tests and on the distribution of the error terms to derive the asymptotic results. The present work shows how to use a genuine Bayesian procedure[9], FBST, to test sharp hypotheses in time series analysis: the unit root sharp hypothesis. The FBST has showed its versatility: (a) the *e*-value derivation and implementation are straightforward from its general definition; (b) it uses absolutely no artificial restrictions, like a special prior, or a probability measure on the hypothesis set, induced by some specific parameterization; (c) it is in strict compliance with the likelihood principle; (d) it can conduct the test with any prior distribution; (e) it does not need closed conjectures concerning error distributions, even for small samples; (f) it is an exact procedure, since it does not use asymptotic restrictrions; and (g) it is invariant with respect to the null hypothesis parameterization and with respect to the parameter space parameterization.[10]

Although the authors are aware of the problems involving the prior chosen for this work, our goal was to show the possibility to implement the test even with this prior, because its use causes problems to other Bayesian procedures like the Bayes factor tests. To proceed with this research agenda, it would be interesting to do more studies with the FBST applied to unit root testing with semi-parametric distributions. Another suggestion is to broaden the group of models used for unit root testing. For instance, we could include moving average terms and work with ARMA models. For the FBST, this extension is not as difficult as it is for the frequentists. The frequentist tradition should derive another statistic with its respective distribution and critical values.

Appendix

The definition of the evidence against some sharp hypothesis H given in Sec. 2 is invariant with respect to a proper reparameterization. For instance, let $\omega = \phi(\theta)$ where $\phi(\cdot)$ is a measurable and integrable function. For purpose of

---

[8]See Bauwens et al. (1999).
[9]See Pereira et al. (2008).
[10]See the Appendix.

illustration, assume that it is bijective and continuously differentiable. Under the reparameterization, the Jacobian, surprise, posterior, and e-value (against the hypothesis) are, respectively, $J(\omega)$, $\tilde{r}(\omega)$, $\tilde{g}_x(\omega)$, and $\widetilde{ev}(H)$, given by:

$$J(\omega) = \left[\frac{\partial \theta}{\partial \omega}\right] = \left[\frac{\partial \phi^{-1}(\omega)}{\partial \omega}\right] = \begin{bmatrix} \frac{\partial \theta_1}{\partial \omega_1} & \cdots & \frac{\partial \theta_1}{\partial \omega_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial \theta_n}{\partial \omega_1} & \cdots & \frac{\partial \theta_n}{\partial \omega_n} \end{bmatrix}$$

$$\tilde{s}(\omega) = \frac{\tilde{g}_x(\omega)}{\tilde{r}(\omega)} = \frac{g_x(\phi^{-1}(\omega))|J(\omega)|}{r(\phi^{-1}(\omega))|J(\omega)|}.$$

Let $\Omega_H = \phi(\Theta_H)$. It follows that

$$\tilde{s}^* = \sup_{\omega \in \Omega_H} \tilde{s}(\omega) = \sup_{\theta \in \Theta_H} s(\theta) = s^*$$

hence, the tangential set under the reparameterization is, $T \mapsto \phi(T) = \tilde{T}$, and

$$\widetilde{ev}(H) = \int_{\tilde{T}} \tilde{g}_x(\omega)d\omega = \int_T g_x(\theta)d\theta = \overline{ev}(H).$$

We remark that the FBST is also invariant with respect to the null hypothesis parameterization. This is not a trivial issue because some statistical procedures do not satisfy this property. The reader interested in a broader discussion of the FBST properties can see Madruga et al. (2003) and Pereira et al. (2008).

References

Basu, D. (1977). On the elimination of nuisance parameters. *J. Amer. Statist. Assoc.* 72:355–366.

Bauwens, L., Lubrano, M. (1996). *Advances in Econometrics*. Vol. 11, part B. London: JAI Press, p. 328.

Bauwens, L., Lubrano, M., Richard, J.-F. (1999). *Bayesian Inference in Dynamic Econometric Models*. Oxford: Oxford University Press.

Birgin, E. G., Castillo, R., Martinez, J. M. (2004). Numerical comparison of augmented lagrangian algorithms for nonconvex problems. *Computat. Optimiz. Applic.* 31:1–56. Software and documentation available at: www.ime.usp.br/~egbirgin/tango/.

Corana, A., Marchesi, M., Martini, C., Ridella, S. (1987). Minimizing multimodal functions of continuous variables with the simulated annealing algorithm. *ACM Trans. Mathemat. Software* 13:262–280.

DeJong, D., Whiteman, C. H. (1991). Reconsidering trends and random walks in macroeconomic time series. *J. Monetary Econ.* 28:221–254.

Dickey, D. A., Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *J. Amer. Statist. Assoc.* 74:427–431.

Goffe, W. L., Ferrier, G. D., Rogers, J. (1994). Global optimization of statistical functions with simulated annealing. *J. Econometrics* 60:65–99.

Good, I. J. (1983). *Good Thinking: the Foundations of Probability and its Applications*. Minneapolis: University of Minnesota Press.

Lauretto, M., Pereira, C. A. B., Stern, J. M., Zacks, S. (2003). Full Bayesian significance test applied to multivariate normal structure models. Comparing parameters of two bivariate normal distributions using the invariant full bayesian significance test. *Braz. J. Probabi. Statist.* 17:147–168.

Lindley, D. V. (1957). A Statistical Paradox. *Biometrika* 44:187–192.

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

MacKinnon, J. G. (1994). Approximate asymptotic distribution functions for unit-root and cointegration tests. *J. Bus. Econ. Statist.* 12:167–176.

Madruga, M. R., Esteves, L. G., Wechsler, S. (2001). On the Bayesianity of Pereira-Stern tests. *Test* 10:291–299.

Madruga, M. R., Pereira, C. A. B., Stern, J. M. (2003). Bayesian evidence test for precise hypotheses. *J. Statistical Plan. Infer.* 117:185–198.

Nelson, C., Plosser, C. (1982). Trends and random walks in macroeconomic time series: Some evidence and implications. *J. Monetary Econ.* 10:139–162.

Pereira, C. A. B., Stern, J. M. (1999). Evidence and credibility: full Bayesian significance test for precise hypotheses. *Entropy J.* 1:69–80.

Pereira, C. A. B., Stern, J. M., Wechsler, S. Can a significance test be genuinely bayesian. *Bayesian Anal.* 1:79–100.

Phillips, P. C. (1991a). To criticize the critics: an objective Bayesian analysis of stochastic trends. *J. Appl. Econometrics* 6:333–364.

Phillips, P. C. (1991b). Bayesian routes and unit roots: de rebus prioribus semper est disputandum. *Journal of Applied Econometrics* 6:435–474.

Phillips, P. C., Xiao, Z. (1998). A primer on unit root testing. *J. Econ. Surv.* 12:423–469.

Schervish, M. (1995). *Theory of Statistics*. New York: Springer.

Schotman, P. C., van Dijk, H. K. (1991). On Bayesian routes to unit roots. *J. Appl. Econometrics* 49:387–401.

Sims, C. A. (1988). Bayesian skepticism on unit root econometrics. *J. Econ. Dynam. Contr.* 12:463–474.

Sims, C. A., Uhlig, H. (1991). Understanding unit rooters: a helicopter tour. *Econometrica* 59:1591–1600.

Tierney, L., Kadane, J. B. (1986). Accurate approximation for posterior moments and marginal densities. *J. Amer. Statist. Assoc.* 81:82–86.

Wald, A. (1950). *Statistical Decision Functions*. New York: John Wiley and Sons.