MATHEMATICAL ENGINEERING
TECHNICAL REPORTS

A Superharmonic Prior for the Autoregressive Process of the Second Order

Fuyuhiko TANAKA and Fumiyasu KOMAKI

METR 2006–30 May 2006

DEPARTMENT OF MATHEMATICAL INFORMATICS
GRADUATE SCHOOL OF INFORMATION SCIENCE AND TECHNOLOGY
THE UNIVERSITY OF TOKYO
BUNKYO-KU, TOKYO 113-8656, JAPAN

WWW page: http://www.i.u-tokyo.ac.jp/mi/mi-e.htm

A Superharmonic Prior for the Autoregressive Process of the Second Order

Fuyuhiko TANAKA and Fumiyasu KOMAKI

Department of Mathematical Informatics
Graduate School of Information Science and Technology
The University of Tokyo
ftanaka@stat.t.u-tokyo.ac.jp

May 19th, 2006

Abstract

The Bayesian estimation of the spectral density of the AR(2) process is considered. We propose a superharmonic prior on the model as a noninformative prior rather than the Jeffreys prior. Theoretically the Bayesian spectral density estimator based on it dominates asymptotically the one based on the Jeffreys prior under the Kullback-Leibler divergence. In the present paper, an explicit form of a superharmonic prior for the AR(2) process is presented and compared with the Jeffreys prior in computer simulation.

1. INTRODUCTION

The p-th order autoregressive (AR(p)) model is widely-known in the time series analysis. It consists of the data $\{x_t\}$ and satisfying $x_t = -\sum_{i=1}^{p} a_i x_{t-i} + \epsilon_t$, where $\{\epsilon_t\}$ is a Gaussian white noise with mean 0 and variance σ^2. The point estimation of the AR parameters $a_1, \cdots, a_p$ is well understood and has been vigorously investigated for a long time. On the other hand, the AR(p) model is surprizingly challenging to objective Bayesians, even in the comparatively simple case of known σ^2, as is discussed in Phillips (1991). The AR(1) model was investigated in such a Bayesian viewpoint by Berger and Yang (1994). Their approach is based on the reference prior method and the stationarity is not assumed. In contrast, assuming the stationarity, we focus on the Bayesian estimation of the spectral density of the AR(2) model and propose a superharmonic prior as a noninformative prior.

The spectral density estimation in the Bayesian framework and the superharmonic prior approach is briefly reviewed in Section 2.1. Our argument is mainly based on the geometrical properties of the AR(p) model manifolds endowed with the Fisher metric. General definition of the AR(p) model manifolds are given in Section 2.2. We also rewrite the Fisher metric in the simple form and give the Jeffreys prior for the AR(2) process by introducing a new coordinate. A superharmonic prior for the AR(p) process is defined in Section 3.1 and an explicit form is given in the AR(2) process in Section 3.2. Although in an asymptotic sense superharmonic prior distributions are shown to be better than the Jeffreys prior distributions in our setting, we perform the numerical simulation in the AR(2) process and the result ensures this fact in Section 4.

2. THEORETICAL PRELIMINARY

2.1. Our motivation

Let us consider a parametric model of stationary Gaussian process with its mean parameter zero. It is known that a stationary Gaussian process corresponds to its spectral density one-to-one. Thus, we focus on the estimation of the true spectral density $S(\omega|\theta_0)$ in a parametric family of spectral densities

$$\mathcal{M} := \{S(\omega|\theta) : \theta \in \Theta \subseteq \mathbf{R}^k\}.$$

The performance of a spectral density estimator $\hat{S}(\omega)$ is evaluated by the Kullback-Leibler divergence.

$$D(S(\omega|\theta_0)||\hat{S}(\omega)) := \int_{-\pi}^{\pi} \frac{\mathrm{d}\omega}{4\pi} \left\{ \frac{S(\omega|\theta_0)}{\hat{S}(\omega)} - 1 - \log\left(\frac{S(\omega|\theta_0)}{\hat{S}(\omega)}\right) \right\}.$$

Then, let us consider minimizing the average risk. First we assume that a proper prior distribution $\pi(\theta)$ is known in advance. The spectral density estimator minimizing the average risk,

$$\begin{aligned} &E^{\Theta}E^{X}[D(S(\omega|\theta)||\hat{S}(\omega))] \\ &:= \int \mathrm{d}\theta\pi(\theta) \int \mathrm{d}x_1 \ldots \mathrm{d}x_n p_n(x_1, \ldots, x_n|\theta) D(S(\omega|\theta)||\hat{S}(\omega)), \end{aligned}$$

is given by the Bayesian spectral density (with respect to $\pi(\theta)$), which is defined by

$$S_{\pi}(\omega) := \int S(\omega|\theta)\pi(\theta|x)\mathrm{d}\theta. \quad (1)$$

We call the r.h.s. of the above definition (1) a *Bayesian spectral density* even when an improper prior distribution is considered.

If one has no information on the unknown parameter θ, it is natural to adopt a noninformative prior in the Bayesian framework. Although there are arguments on the choice of a noninformative prior, usually the Jeffreys prior seems to be adopted. However, Tanaka and Komaki (2005) show that in the ARMA process there exists another candidate as a noninformative prior if there exists a positive superharmonic function on the ARMA model manifold with the Fisher metric. We call this candidate *a superharmonic prior*, whose definition is given later. It is a non-i.i.d. extension of the work by Komaki (2006). In his paper, Komaki (2006) also pointed out that a sufficient condition for the existence of superharmonic priors is that the model manifold has a non-positive sectional curvature.

Now we focus on the AR process. Tanaka and Komaki (2003) evaluated the sign of the sectional curvature of the AR model manifold with the above motivation. They showed that the sectional curvature on the AR(2) model always has a negative sign, while those on the AR(p) models ($p \geq 3$) could have a positive sign at some points. Thus, it was shown that there exists a superharmonic prior (at least) for the AR(2) process.

But we emphasize that the above result only ensures the existence of a superharmonic prior for the AR(2) process. When estimating the spectral density of the AR process from the observed data, one needs an explicit form of a positive superharmonic function on the model manifold. Thus, the purpose of this paper is to give the explicit form of a positive superharmonic function on the AR(2) model manifold and the superharmonic prior derived from it. Numerical simulation is also performed and its result illustrates the validity of our method.

2.2. Fisher metric on the AR(p) model manifold

Here, we briefly summarize the AR(p) model manifold and some geometrical facts needed in the present paper.

The Riemannian metric of a model specified by a parametric family of spectral densities $\mathcal{M} := \{S(\omega|\theta)|\theta \in \Theta\}$, where θ is a finite-dimensional parameter, is

$$g_{ij} := g\left(\frac{\partial}{\partial\theta^i}, \frac{\partial}{\partial\theta^j}\right) = \frac{1}{4\pi}\int_{-\pi}^{\pi} \mathrm{d}\omega \frac{\partial_i S(\omega|\theta)}{S(\omega|\theta)} \frac{\partial_j S(\omega|\theta)}{S(\omega|\theta)}, \qquad (2)$$

where $\partial_i := \frac{\partial}{\partial\theta^i}$. We call it *the Fisher metric*. In the information geometry, considering θ as a coordinate, we call $\mathcal{M}$ *a model manifold*, or shortly *a model*, see Amari (1987) and Amari and Nagaoka (2000).

The explicit form of the spectral density of the AR model is given by

$$S(\omega, a_1, \cdots, a_p, \sigma^2) = \frac{\sigma^2}{2\pi} \frac{1}{|H_a(z)|^2}, \quad z = e^{i\omega}, \qquad (3)$$

where $H_a(z)$ is the charateristic polynomial, z is the shift operator, i.e., $zx_t = x_{t+1}$, and

$$x_t = H_a(z)^{-1}\epsilon_t, \quad H_a(z) := \sum_{i=0}^{p} a_i z^{-i} \text{ with } a_0 = 1 \ .$$

Here, we use another coordinate system, which brings us a more convenient form to consider it. Equation $z^p H_a(z) = z^p + a_1 z^{p-1} + \cdots + a_{p-1} z + a_p$ is a p-degree polynomial and has p complex roots, $z_1, z_2, \cdots, z_p$ (Note that $|z_i| < 1$ from the stationarity condition). Since $a_1, a_2, \cdots, a_p$ are all real, it consequently has the conjugate roots. Thus, we can put them in the order like, $z_1, \cdots, z_q, z_{q+1}, \cdots, z_{2q} \in \mathbf{C}, z_{2q+1}, \cdots, z_{2q+r} \in \mathbf{R}$ and $z_{q+j} = \bar{z_j} (1 \leq j \leq q)$ (for simplicity, we assume that there are no multiple roots). The roots $z_1, z_2, \cdots, z_p$ correspond to the original parameter $a_1, a_2, \cdots, a_p$ one-to-one. Let us introduce a coordinate system $(\theta^0, \theta^1, \cdots, \theta^p)$ using these roots

$$\theta^0 := \sigma^2, \quad \theta^1 := z_1, \quad \theta^2 := z_2, \ \cdots, \theta^p := z_p.$$

In the remainder of the paper indices $I, J, K, \cdots$ run $0, 1, \cdots, p$ and indices $i, j, k, \cdots$ run $1, 2, \cdots, p$. The formal complex derivatives are defined by

$$\frac{\partial}{\partial z_i} := \frac{1}{2}\left(\frac{\partial}{\partial x_i} + i\frac{\partial}{\partial y_i}\right), \quad \frac{\partial}{\partial \bar{z}_i} := \frac{1}{2}\left(\frac{\partial}{\partial x_i} - i\frac{\partial}{\partial y_i}\right),$$

see, for example, Gunning and Rossi (1965). Since the conjugate complex coordinates z_i and $\bar{z}_i$ correspond to x_i and y_i one-to-one, we are always able to go back to the real coordinates x_i and y_i once we need. In the coordinate system given above, the Fisher metric on the AR(p) model g_{IJ} is given by

$$g_{IJ} = \left(\begin{array}{c|ccc} g_{00} & \cdots & g_{0i} & \cdots \\ \hline \vdots & \cdots & \cdots & \cdots \\ g_{i0} & \vdots & g_{ij} & \vdots \\ \vdots & \vdots & \cdots & \vdots \end{array}\right) \quad \text{and} \quad \left\{\begin{array}{ccc} g_{00} & = & \frac{1}{2(\theta^0)^2} = \frac{1}{2\sigma^4} \\ g_{0i} & = & g_{i0} = 0 \\ g_{ij} & = & \frac{1}{1 - z_i z_j} \end{array}\right., \qquad (4)$$

see Komaki (1999) for more details of information geometric quantities.

Now we can easily obtain the Jeffreys prior $\pi_J(\theta)$ for the AR(2) process using the Fisher metric (4). The Jeffreys prior $\pi_J(\theta)$ is improper prior and given by

$$\begin{aligned} \pi_J(\theta) &\propto \sqrt{|g|} \\ &= \sqrt{\frac{1}{2(\theta^0)^2}} \left| \frac{(\theta^1 - \theta^2)^2}{\{1 - (\theta^1)^2\}\{1 - (\theta^2)^2\}(1 - \theta^1\theta^2)^2} \right|^{\frac{1}{2}}, \end{aligned}$$

where $g := \det(g_{IJ})$. In the next section, we derive a superharmonic prior on the AR(2) model using these forms.

3. SUPERHARMONIC PRIOR FOR THE AR(2) PROCESS

In this section, we describe the definition of a superharmonic prior and give an example of it for the AR(2) process.

3.1. Superharmonic prior

Let $\mathcal{M}$ denote a Riemannian manifold with a coordinate θ. A scalar function $\phi(\theta)$ on $\mathcal{M}$ is called *a superharmonic function* if it satisfies,

$$\Delta\phi(\theta) \leq 0 \quad \forall\theta,$$

where Δ is the Laplace-Beltrami operator. Let g_{IJ} be a Riemannian metric (Fisher metric), g^{IJ}, the inverse of g_{IJ}, and $g := \det(g_{IJ})$. The Laplace-Beltrami operator is defined by

$$\Delta\phi := \frac{1}{\sqrt{g}}\frac{\partial}{\partial\theta^I}\left(\sqrt{g}g^{IJ}\frac{\partial}{\partial\theta^J}\phi\right).$$

Note that we adopt Einstein's summation convention: if an index occurs twice in any one term, once as an upper and once as a lower index, summation over that index is implied.

If a superharmonic function is positive, i.e., $\phi(\theta) > 0$, $\forall\theta$, then it is called *a positive superharmonic function*. When a model manifold with the Fisher metric has a positive superharmonic function $\phi(\theta)$, we call $\pi_H(\theta) := \pi_J(\theta)\phi(\theta)$ *a superharmonic prior*. Note that not all model manifolds with the Fisher metric have a superharmonic prior while all of them have the Jeffreys prior.

In the AR(p) model manifold, Δ can be decomposed into two parts. One part is relevant with $\theta^0 = \sigma^2$ and the others with $\theta^1, \cdots, \theta^p$. We show it in the following for later convenience. First, we divide the metric into two componets, 00 and ij. Let us define $h_{ij} := g_{ij}, \quad h^{ij} := h_{ij}^{-1}, \quad h := \det(h_{ij})$. Then,

$$g = \det(g_{IJ}) = g_{00}h,$$

$$g^{IJ} := (g_{IJ})^{-1} = \begin{pmatrix} g_{00}^{-1} & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & h^{ij} & \\ 0 & & & \end{pmatrix}.$$

We obtain the decomposition of Δ using h_{ij}.

$$\Delta\phi \;=\; \frac{1}{\sqrt{g_{00}}\sqrt{h}}\frac{\partial}{\partial\theta^I}\left(\sqrt{g_{00}}\sqrt{h}\left(g^{I0}\partial_0\phi + g^{Ij}\partial_j\phi\right)\right)$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{g_{00}}\sqrt{h}} \left\{ \frac{\partial}{\partial\theta^0} \left(\sqrt{g_{00}}\sqrt{h} g^{00} \partial_0 \phi \right) + \frac{\partial}{\partial\theta^i} \left(\sqrt{g_{00}}\sqrt{h} h^{ij} \partial_j \phi \right) \right\} \\
&= \frac{1}{\sqrt{g_{00}}} \frac{\partial}{\partial\theta^0} \left(\sqrt{g_{00}} g^{00} \partial_0 \phi \right) + \frac{1}{\sqrt{h}} \frac{\partial}{\partial\theta^i} \left(\sqrt{h} h^{ij} \partial_j \phi \right) \\
&= \Delta_0 \phi + \Delta_h \phi,
\end{aligned}
$$

where we used the identies below.

$$
g^{0j} = 0, \quad \frac{\partial}{\partial\theta^0}\sqrt{h} = 0, \quad \frac{\partial}{\partial\theta^i}\sqrt{g_{00}} = 0.
$$

In the last line, both terms are defined as

$$
\begin{aligned}
\Delta_0 \phi &:= \frac{1}{\sqrt{g_{00}}} \frac{\partial}{\partial\theta^0} \left(\sqrt{g_{00}} g^{00} \partial_0 \phi \right), \\
\Delta_h \phi &:= \frac{1}{\sqrt{h}} \frac{\partial}{\partial\theta^i} \left(\sqrt{h} h^{ij} \partial_j \phi \right).
\end{aligned}
$$

We can, thus, consider each term Δ_0 and Δ_h separately. Observing that

$$
g_{00} = \frac{1}{2(\theta^0)^2}, \quad g^{00} = 2(\theta^0)^2, \quad \theta^0 := \sigma^2 > 0
$$

and Δ_0 is rewritten as

$$
\Delta_0 \phi = 2(\theta^0)^2 \left(\frac{\partial}{\partial\theta^0} \right)^2 \phi + 2\theta^0 \frac{\partial\phi}{\partial\theta^0}.
$$

Therefore, in order to find out a positive superharmonic function, it is enough to find $\phi_h(\theta^1, \cdots, \theta^p)$ out satisfying

$$
\begin{cases} \phi_h(\theta^1, \cdots, \theta^p) > 0 \\ \Delta_h \phi_h(\theta^1, \cdots, \theta^p) \le 0 \end{cases} \quad \text{for } |\theta^i| < 1, \quad i = 1, \cdots, p \,. \tag{5}
$$

It is easily seen that ϕ_h satisfies the original condition $\Delta\phi_h \le 0$ since $\Delta_0\phi_h = 0$.

3.2. An example of superharmonic priors π_H for the AR(2) process

From here on, we consider the AR(2) model only and fix $\sigma^2 = 1$. Since we assume the stationarity condition, the parameter region on the AR(2) model is given by

$$
\Omega := \{\theta = (\theta^1, \theta^2) = (z_1, z_2) : \ |z_1| < 1, |z_2| < 1\}
$$

Fortunately, we find out one positive superharmonic function on this model.

Proposition 3.1. *A positive superharmonic function on the AR(2) model is given by*

$$\phi(z_1, z_2) := 1 - z_1 z_2.$$

Thus, a superharmonic prior for the AR(2) process is given by

$$\pi_H(z_1, z_2) \propto \left| \frac{(z_1 - z_2)^2}{(1 - z_1^2)(1 - z_2^2)} \right|^{\frac{1}{2}}.$$

Proof.
From the stationarity condition, it is easily seen that $\phi(z_1, z_2) > 0$. Next, we will show that

$$\Delta_h \phi(z_1, z_2) = -\phi(z_1, z_2) < 0. \tag{6}$$

Both (6) and $\Delta_0 \phi = 0$ indicate that $\phi(z_1, z_2)$ is a superharmonic function. In the AR(2) model, $\Delta_h \phi$ is written in the following way.

$$\begin{aligned}
\Delta_h \phi &= \frac{1}{\sqrt{h}} \frac{\partial}{\partial z_i} \left(\sqrt{h} h^{ij} \frac{\partial}{\partial z_j} \phi \right) \\
&= \frac{\partial}{\partial z_1} \left(h^{11} \frac{\partial \phi}{\partial z_1} + h^{12} \frac{\partial \phi}{\partial z_2} \right) + \frac{\partial}{\partial z_2} \left(h^{21} \frac{\partial \phi}{\partial z_1} + h^{22} \frac{\partial \phi}{\partial z_2} \right) \\
&\quad + \frac{1}{2h} \left(\frac{\partial h}{\partial z_1} \right) \left(h^{11} \frac{\partial \phi}{\partial z_1} + h^{12} \frac{\partial \phi}{\partial z_2} \right) + \frac{1}{2h} \left(\frac{\partial h}{\partial z_2} \right) \left(h^{21} \frac{\partial \phi}{\partial z_1} + h^{22} \frac{\partial \phi}{\partial z_2} \right),
\end{aligned}$$

where h_{ij} with $p = 2$ is represented by

$$h_{ij} = g_{ij} = \begin{pmatrix} \frac{1}{1-z_1^2} & \frac{1}{1-z_1 z_2} \\ \frac{1}{1-z_1 z_2} & \frac{1}{1-z_2^2} \end{pmatrix}.$$

Determinant h and the inverse matrix h^{ij} are given below.

$$h = h_{11} h_{22} - h_{12} h_{21} = \frac{(z_1 - z_2)^2}{(1 - z_1^2)(1 - z_2^2)(1 - z_1 z_2)^2},$$

$$h^{ij} = \frac{1}{h} \begin{pmatrix} \frac{1}{1-z_2^2} & -\frac{1}{1-z_1 z_2} \\ -\frac{1}{1-z_1 z_2} & \frac{1}{1-z_1^2} \end{pmatrix}.$$

Using the above, it is straightforward to obtain (6) *Q.E.D.*

4. SIMULATION

In the last section, we obtained a superharmonic prior $\pi_H(\theta)$ for the AR(2) process. It is shown to be better than the Jeffreys prior $\pi_J(\theta)$ in an asymptotic sense in Tanaka and Komaki (2005). From numerical simulation we

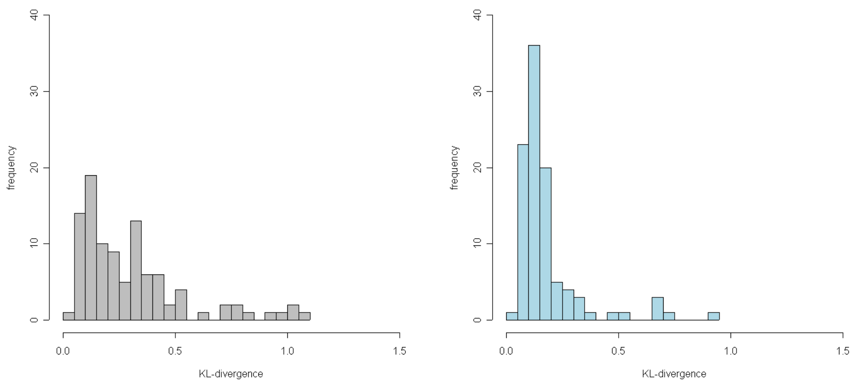

Figure 1: Histgram of the Kullback-Leibler divergence (Upper:$D(S_T, \hat{S}_J)$, Lower:$D(S_T, \hat{S}_H)$)

can see directly that the Bayesian estimator of the spectral density based on our superharmonic prior gives a better estimating method than that based on the Jeffreys prior.

Our simulation goes as follows: First we generate 100 groups of the observed data $\mathbf{x} = (x_1, x_2, \cdots, x_N)$ under the AR(2) process with a_1, a_2 fixed. For simplicity, we fix $\sigma^2 = 1$, and the length of the time series is set $N = 10$. Then, we construct a posterior distribution $\pi_J(\theta|\mathbf{x})$ using $\pi_J(\theta)$ and another one $\pi_H(\theta|\mathbf{x})$ using $\pi_H(\theta)$. Two kinds of the Bayesian estimator of the spectral density, $\hat{S}_J(\omega)$, $\hat{S}_H(\omega)$ are given by

$$\hat{S}_{J(H)}(\omega) := \int\int S(\omega|a_1, a_2)\pi_{J(H)}(a_1, a_2|\mathbf{x})\mathrm{d}a_1\mathrm{d}a_2.$$

Finally we calculate the Kullback-Leibler divergence from the true spectral density $S_T(\omega|a_1, a_2)$ to $\hat{S}_J(\omega), \hat{S}_H(\omega)$ for each observation. For example, Figure 1 shows the histgram of the Kullback-Leibler divergence with $a_1 = 0.5, a_2 = -0.3$. Obviously $\hat{S}_H(\omega)$ performs better than $\hat{S}_J(\omega)$ in the sense that the former estimator of the spectral density is more concenterated on the true density $S_T(\omega)$. Similar results hold with other sets of the AR parameters. This result illustrates the validity of our method in a practical application .

5. CONCLUDING REMARKS

We indicated how to estimate the spectral density on the AR(2) process in the Bayesian framework by introducing a superharmonic prior.

Our attention was paid only to the AR(2) model. We briefly describe the MA models. Generally, the MA models are completely different from the AR models as a stochastic process and in the information geometrical viewpoint they are known to have the different structures. These models, however, have the same structure in the Riemaniann geomerical viewpoint, i.e., the Fisher metric in the same order coincides under the appropriate coordinate transformation. In particular, the same result as the AR(2) model holds in the MA(2) model.

We also note that a superharmonic prior could exist for the higher order (i.e. $p \geq 3$) AR processes and other stochastic processes such as ARMA processes. Thus, we will try to enlarge our method to them. It is also important in practice to give a construction algorithm to obtain an explicit form of a positive superharmonic function once it is shown to exist on a statistical model manifold.

ACKNOWLEDGEMENT

One of the authors, F. Tanaka, was supported by the JSPS Research Fellowships for Young Scientists.

REFERENCES

AMARI, S. (1987) Differential geometry of a parametric family of invertible linear systems - Riemannian metric, dual affine connections, and divergence. *Mathematical System Theory* 20, 53–82.

AMARI, S. and NAGAOKA, H. (2000) *Methods of Information Geometry.* Oxford: American Mathematical Society.

BERGER, J. YANG, R.Y. (1994) Noninformative priors and Bayesian testing for the AR(1) model. *Econotmeric Theory* 10, 461–482.

GUNNING, C. and ROSSI, H. (1965) *Analytic Functions of Several Complex Variables.* Englewood Cliffs, New Jersey: Prentice-Hall.

KOMAKI, F. (1999) Estimating method for parametric spectral densities. *Journal of Time Series Analysis* 20, 31–50.

KOMAKI, F. (2006) Shrinkage priors for Bayesian prediction. to appear in *Annals of Statistics.*

TANAKA, F. and KOMAKI, F. (2003) The sectional curvature of AR model manifolds. *Tensor* 64, 131–143.

TANAKA, F. and KOMAKI, F. (2005) Asymptotic expansion of the risk difference of the Bayesian spectral density in the ARMA model. *Mathematical Engineering Technical Reports* 2005–31.

PHILLIPS, P. C. B. (1991) To criticize the critics: An objective Bayesian analysis of stochastic trends. *Journal of Applied Econometrics* 6, 333–364.