CAMBRIDGE
UNIVERSITY PRESS

Priors for Macroeconomic Time Series and Their Application

Author(s): John Geweke

Source: *Econometric Theory*, Vol. 10, No. 3/4, Symposium Double Issue: Bayes Methods and Unit Roots (Aug. - Oct., 1994), pp. 609-632

Published by: Cambridge University Press

Stable URL: http://www.jstor.org/stable/3532552

Accessed: 01-05-2017 03:10 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at http://about.jstor.org/terms

*Cambridge University Press* is collaborating with JSTOR to digitize, preserve and extend access to *Econometric Theory*

This content downloaded from 128.83.155.248 on Mon, 01 May 2017 03:10:07 UTC
All use subject to http://about.jstor.org/terms

*Econometric Theory*, **10**, 1994, 609–632. Printed in the United States of America.

# PRIORS FOR MACROECONOMIC TIME SERIES AND THEIR APPLICATION

JOHN GEWEKE
*University of Minnesota*
*and*
*Federal Reserve Bank of Minneapolis*

This paper takes up Bayesian inference in a general trend stationary model for macroeconomic time series with independent Student-*t* disturbances. The model is linear in the data, but nonlinear in parameters. An informative but nonconjugate family of prior distributions for the parameters is introduced, indexed by a single parameter that can be readily elicited. The main technical contribution is the construction of posterior moments, densities, and odd ratios by using a six-step Gibbs sampler. Mappings from the index parameter of the family of prior distribution to posterior moments, densities, and odds ratios are developed for several of the Nelson–Plosser time series. These mappings show that the posterior distribution is not even approximately Gaussian, and they indicate the sensitivity of the posterior odds ratio in favor of difference stationarity to the choice of the prior distribution.

## 1. INTRODUCTION

Beginning with the investigation of Nelson and Plosser [19], the propositions that most macroeconomic aggregates are trend stationary or, alternatively, that they are difference stationary have captured the attention of applied and theoretical econometricians as have few other issues. These ideas have accelerated the development of the sampling theory of estimators in the presence of nonstationarity and near-nonstationarity (Dickey and Fuller [3], Said and Dickey [28], Phillips [23], Sims, Stock, and Watson [34]). More recently, these questions have renewed research in Bayesian inference for time series

Full technical details for this work are provided in Geweke [10], which is available by request from the Federal Reserve Bank of Minneapolis, Research Department, 250 Marquette Avenue, Minneapolis, MN 55480, USA. Software may be requested by electronic mail (geweke@atlas.socsci.umn.edu). This work was supported in part by National Science Foundation grant SES-9210070. Thanks go to Kenneth Wallis and Giovanni Amisano for pointing out an error in an earlier version; to two referees to the editors, whose comments improved the paper; to Zhenyu Wang for providing research assistance; and to Charles Nelson for supplying the data. Any remaining errors are my responsibility. The views expressed in this paper are not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

© 1994 Cambridge University Press 0266-4666/94 $5.00 + .00



This content downloaded from 128.83.155.248 on Mon, 01 May 2017 03:10:07 UTC
All use subject to http://about.jstor.org/terms

(Zellner and Tiao [39], Sims [32], DeJong and Whiteman [2], Phillips [24], Sims and Uhlig [35], Schotman and van Dijk [29–31]). That basic questions about methodology are being taken up in the context of a specific empirical issue testifies to the intellectual health and vigor of econometrics. Contemporaneously with these developments, there have been rapid advances in Bayesian multiple integration, which can enrich time series econometrics. The objective of this paper is to show some ways in which these advances can help address the issues of trend and difference stationarity. In doing so, it builds on a number of recent contributions by Geman and Geman [8], Gelfand and Smith [7], and Geweke [9,11].

This paper breaks new methodological ground in several directions. First, it takes up Bayesian inference in an improved specification of the model by Schotman and van Dijk [29–31], which cannot be attacked by the essentially analytical methods of those papers or Phillips [24]. Second, it employs informative and nonconjugate priors for the parameters of interest. Third, motivated by evidence in Geweke [11], disturbances in the model are leptokurtic. Finally, the paper shows how to construct exact highest posterior density regions for a model that is a nontrivial variant of the standard linear specification.

This work makes two primary substantive contributions. First, it introduces a single-parameter family of informative prior distributions for the autoregressive component of the trend stationary model. The choice of this parameter is implied by the answer to the question, At what time interval is a uniform prior density on the unit interval for the autoregressive component plausible? As this time interval increases, the prior distribution places increasing probability on a near-nonstationary configuration, and, as a corollary, the posterior odds ratio in favor of difference stationarity will approach the prior odds ratio, regardless of the sample. This convergence is illustrated by using the data of Nelson and Plosser [19]. Second, this work presents posterior moments, posterior densities, and highest posterior density regions for these data and priors that indicate near-nonstationarity. Posterior odds ratios in favor of difference stationarity are sensitive to the choice of the parameter for the prior for the autoregressive coefficient but never fall much below the prior odds ratio and often greatly exceed it. Posterior distributions are non-Gaussian.

The paper is organized as follows. Section 2 introduces the model and the likelihood function. A family of informative prior distributions is developed, and the posterior density function is derived. The posterior distribution is interpreted through successive conditioning of each of several subsets of parameters on all the other subsets of parameters. In Section 3 these conditional posterior distributions are used to construct a six-step Gibbs sampler, which generates a Markov process on the parameter space. The limiting and invariant distribution of this Markov process is the joint posterior distribution of the parameters. This section of the paper also discusses the evalua-

This content downloaded from 128.83.155.248 on Mon, 01 May 2017 03:10:07 UTC
All use subject to http://about.jstor.org/terms

tion of the numerical accuracy of Gibbs sampling-based approximations of posterior moments, shows how to construct the posterior odds ratio in favor of difference stationarity as the posterior expectation of a function of interest in the trend stationary model, and develops methods for computation of exact highest posterior density regions. Section 4 reports the empirical findings for the Nelson–Plosser data set. These include an investigation of sensitivity to the specification of the prior distribution, posterior moments and densities for the parameters of interest, and posterior odds ratios in favor of difference stationarity. The last section places the results and methods of this work in the context of a broader research agenda.

## 2. PRIOR AND POSTERIOR DISTRIBUTIONS IN THE TREND STATIONARY MODEL

The trend stationary model used in this research is

$$y_t = \gamma + \delta t + u_t, \tag{1}$$

$$A^*(L)u_t = \epsilon_t; \qquad A^*(L) = (1 - \rho L) + A(L)(1 - L); \qquad 0 \le \rho < 1, \tag{2}$$

$$\{\epsilon_t\} \text{ i.i.d.}, \ \epsilon_t \sim t(0, \sigma^2; \nu). \tag{3}$$

This model is an alternative to the specifications employed by Nelson and Plosser [19], Phillips [24], and others, of the form

$$y_t = \alpha + \beta t + \rho y_{t-1} + \epsilon_t, \tag{4}$$

or elaborations of this form with more lagged values of the dependent variable. An important attraction of (1)–(3) relative to (4) is that $\delta$ is the mean growth rate of $\{y_t\}$ in (1)–(3), whereas $\beta/(1-\rho)$ is the growth rate of $\{y_t\}$ in (4). Priors about mean growth rate and about potential near-nonstationarity are therefore easier to state in the context of (1)–(3) than they are in (4). Throughout we use $A(L) = \sum_{j=1}^{4} a_j L^j$. The truncation of $\{a_j\}$ after $j = 4$ is consistent with the findings of Nelson and Plosser [19], and in conjunction with the prior distribution described in Section 2.1, it loses some of the knife-edge character it might otherwise have. Simple manipulation of (1)–(3) yields

$$y_t = \gamma(1-\rho) + \delta(\rho - \sum_{j=1}^{4} a_j) + \delta(1-\rho)t + \rho y_{t-1} + \sum_{j=1}^{4} a_j(y_{t-j} - y_{t-j-1}) + \epsilon_t, \tag{5}$$

and the likelihood function may be expressed as

$$\sigma^{-T} \prod_{t=1}^{T} \{(1 + \epsilon_t^2)/\nu\sigma^2\}^{-(\nu+1)/2}. \tag{6}$$

This content downloaded from 128.83.155.248 on Mon, 01 May 2017 03:10:07 UTC
All use subject to http://about.jstor.org/terms

The i.i.d. Student-$t$ specification requires the disturbances to be leptokurtic, but for larger values of $\nu$ the distinction is inconsequential. This specification has been lightly used in applications, although it dates back at least to work in astronomy by Jeffreys [15], who used it for mean estimation. It should not be confused with the alternative specification that the entire vector $\epsilon \equiv (\epsilon_1, \ldots, \epsilon_T)'$ has a multivariate-$t$ distribution (taken up by Zellner [38] and Osiewalski and Steel [21], among others) in which the $\epsilon_t$ are not mutually independent. This specification is employed here as an alternative to the normal distribution because posterior odds ratios (reported in detail in Geweke [11]) strongly favor $t(0, \sigma^2; \nu)$ over $N(0, \sigma^2)$ for the data set employed subsequently in this paper.

The work reported here exploits the equivalence between the likelihood function in (6) and the alternative specification

$$\epsilon_t \sim \text{i.d. } N(0, \sigma^2 v_t) \qquad (t = 1, \ldots, T), \tag{7}$$

in conjunction with the independent prior distributions

$$v_t^{-1} \sim \frac{\chi^2(\nu)}{\nu} \qquad (t = 1, \ldots, T),$$

for the fixed, but unknown, relative variance parameters $\{v_t\}$. The prior density kernel for each $w_t = v_t^{-1}$ is $w_t^{(\nu-2)/2} \exp(-\nu w_t/2)$, and the prior density kernel for $v_t$ is

$$v_t^{-(\nu+2)/2} \exp \frac{-\nu}{2v_t}. \tag{8}$$

The posterior density kernel for $\sigma$ and the $v_t$ in the model in (7) with prior densities from (8) is therefore

$$\prod_{t=1}^{T} v_t^{-(\nu+3)/2} \exp\left[ \sum_{t=1}^{T} \frac{\sigma^{-2}\epsilon_t^2 + \nu}{2v_t} \right]. \tag{9}$$

Integrate this expression with respect to $v_1, \ldots, v_T$ to obtain the kernel

$$\prod_{t=1}^{T} (\sigma^{-2}\epsilon_t^2 + \nu)^{-(\nu+1)/2},$$

which is proportional to (6) as a function of the $\epsilon_t$. The specification in (6) is therefore equivalent to (7) and (8), and either specification employed in conjunction with the same prior distributions for the other parameters in the model will produce exactly the same posterior distribution.

This equivalence between the independent Student-$t$ distribution and the normal mixture model with appropriate priors is fully developed in Geweke [11] and is built on related earlier work by De Finetti [1], Fraser [5,6], Harrison and Stevens [13], Maronna [18], Ramsay and Novick [26], West [37],

This content downloaded from 128.83.155.248 on Mon, 01 May 2017 03:10:07 UTC
All use subject to http://about.jstor.org/terms

and Lange, Little, and Taylor [16]. It is a key element of the computational procedure described in Section 3.

### 2.1. Prior Distributions

The parameter of paramount interest, on which recent Bayesian studies of trend and difference stationarity have concentrated, is $\rho$. Here we motivate a simple family of prior distributions, which has also been discussed by Sims [32,33]. Begin with a simplified version of (2),

$$u_t = \rho u_{t-1} + \epsilon_t, \qquad 0 \le \rho < 1, \qquad \epsilon_t \sim \text{i.i.d. } N(0, \sigma^2), \quad \sigma \text{ known.} \tag{10}$$

If $\{u_t\}$ pertains to a point in time (rather than an average over a time interval), then (10) implies that

$$u_t = \alpha u_{t-r} + \epsilon_t^{(r)}, \qquad \epsilon_t^{(r)} \sim N(0, \sigma^2 \frac{1-\rho^{2r}}{1-\rho^2}, \qquad \alpha = \rho^r,$$

$$\operatorname{cov}(\epsilon_t^{(r)}, \epsilon_{t-nr}^{(r)}) = 0 \quad \text{for any non-zero integer } n. \tag{11}$$

A uniform prior distribution on [0, 1) for $\alpha$ in (11) implies a prior distribution for $\rho$ with density $r\rho^{r-1}$ on [0, 1) in (10). Similarly, were the prior distribution on the autoregressive parameter uniform on [0, 1) for a time interval $\tau = n^{-1}$, $n$ integer, the implied prior distribution for $\rho$ in (10) would have density $\tau\rho^{\tau-1}$ on the unit interval. Thus the notion of a "flat prior" for $\rho$ is meaningless without reference to a sampling interval for the time series. These considerations motivate the family of prior densities,

$$\pi_\rho(\rho) = (s+1)\rho^s I_{[0,1)}(\rho); \tag{12}$$

$s = 0$ corresponds to a flat prior on the autoregressive parameter for annual data, $s = 29$ for data recorded every 30 years, $s = -11/12$ for monthly data, etc. The temporal aggregation argument is only motivating: if taken literally in (11) one would have to deal with the presence of the $y_{t-j} - y_{t-j-1}$, the nonnormality of the disturbances, and interaction between prior distributions for the other parameters and $\rho$, any one of which presents technical challenges. The empirical work here is carried through to completion by using several different values of $s$.

If the time interval between measurements is many periods, a uniform distribution for the autoregressive parameter assigns high probability to strong persistence from one period to the next. As $s \to \infty$, the effect of the prior distribution becomes the same as a reformulation of (1)–(3), with $\rho = 1$. As a corollary, the posterior odds ratio in favor of this reformulation must approach the prior odds ratio, as $s \to \infty$; and a posterior odds ratio for (10) with $s = s^*$ in favor of $s = s^* + q$, $s^*$ fixed, must approach the posterior odds ratio in favor of $\rho = 1$, as $q \to \infty$. The operational ramifications of these facts will be seen in Section 4.

This content downloaded from 128.83.155.248 on Mon, 01 May 2017 03:10:07 UTC
All use subject to http://about.jstor.org/terms

The trend coefficient $\delta$ displays no such sensitivity to time aggregation. The prior specification employed here is

$$\delta \sim N(\bar{\delta}, \sigma_\delta^2).$$

The same prior distributions are used for all macroeconomic time series studied. Because the data are in logarithms, $\delta$ indicates mean growth rate. The empirical work is carried out with $\bar{\delta} = 0$ and $\sigma_\delta = .05$. Some checks for sensitivity are reported in Section 4.1.

The prior specification for $a_1, \ldots, a_4$ is

$$a_j \sim N(0, \pi_0 \pi_1^j), \qquad \pi_0 > 0, \qquad 0 < \pi_1 \leq 1.$$

This reflects the belief that these coefficients are not likely to be large in magnitude and that they are smaller, the greater the lag. A similar specification was employed by Doan, Litterman, and Sims [4] for vector autoregressions. In the empirical work, $\pi_0 = .731$ and $\pi_1 = .342$: this implies a standard deviation of .5 for $a_1$ and .1 for $a_4$. Checks for sensitivity are reported in Section 4.1.

For the intercept $\gamma$ of (1), consider prior distributions of the form

$$\gamma | (\rho, a_1, \ldots, a_4, \sigma, y_0) \sim N[m(y_0), \qquad V(\sigma^2, \rho)],$$

where $V(\sigma^2,\rho)$ has the property $\lim_{\rho\to 1}(1-\rho)^2 V(\sigma^2,\rho) = 0$. Conditional on $(\rho, a_1, \ldots, a_4, \sigma, y_0)$, the prior distribution for the intercept in the reduced form (5) is

$$\begin{aligned}
\gamma(1-\rho) + \delta\left(\rho - \sum_{j=1}^{4} a_j\right) \\
\sim N\Bigg[(1-\rho)m(y_0) + \bar{\delta}\left(\rho - \sum_{j=1}^{4} a_j\right), \\
(1-\rho)^2 V(\sigma^2,\rho) + \sigma_\delta^2\left(\rho - \sum_{j=1}^{4} a_j\right)^2\Bigg],
\end{aligned}$$

and for the trend term it is

$$\delta(1-\rho) \sim N[(1-\rho)\bar{\delta},\ (1-\rho)^2\sigma_\delta^2)].$$

The limiting distributions, as $\rho \to 1$, are $N[\bar{\delta}(1 - \sum_{j=1}^4 a_j),\ \sigma_\delta^2(1 - \sum_{j=1}^4 a_j)^2]$ for the intercept and $\delta = 0$ for the trend. This implies the limiting model

$$y_t = \delta(1 - \sum_{j=1}^{4} a_j) + y_{t-1} + \sum_{j=1}^{4} a_j(y_{t-j} - y_{t-j-1}) + \epsilon_t,$$

or, equivalently,

$$y_t = \delta + y_{t-1} + u_t, \tag{13}$$

This content downloaded from 128.83.155.248 on Mon, 01 May 2017 03:10:07 UTC
All use subject to http://about.jstor.org/terms

$$u_t = \sum_{j=1}^{4} a_j u_{t-j} + \epsilon_t, \tag{14}$$

with prior distribution $\delta \sim N(\bar{\delta}, \sigma_\delta^2)$. As $\rho \to 1$, the trend stationary model in (1)–(3) therefore passes smoothly to the difference stationary model in (13)–(14). In the empirical work reported in Section 4, the prior distribution

$$\gamma | (y_0, \rho, \sigma) \sim N[y_0, \sigma_\gamma^2]$$

is employed by using the parameter $\sigma_\gamma = 10$. Because data are in natural logarithms, this prior has little precision. Sensitivity checks reported in Section 4.1 indicate that conclusions about unit roots are essentially invariant with respect to the choice of $\sigma_\gamma$.

With regard to the dispersion of $\epsilon_t$, the reference prior distribution with density

$$\pi_\sigma(\sigma) \propto \sigma^{-1}$$

is assumed for $\sigma$. An exponential prior distribution with density

$$\pi_\nu(\nu) = \omega \exp(-\omega\nu) \tag{15}$$

is taken for $\nu$. In the empirical work, $\omega = .25$, which implies a prior mean of 4 and median of 2.77 for $\omega$. These values are consistent with related findings in Geweke [11], and the exponential form of the prior density allows ample probability for very fat tails in the distribution.

## 2.2. Conditional Posterior Distributions

Consideration of conditional posterior distributions provides both insight into the structure of the posterior distribution and a basis for efficient computation described in Section 3.

*Conditional posterior distribution of* $\gamma$ *and* $\delta$. Write

$$\begin{aligned} w_t &\equiv y_t - \rho y_{t-1} - \sum_{j=1}^{4} a_j (y_{t-j} - y_{t-j-1}) \\ &= \gamma(1-\rho) + \delta\left[\rho - \sum_{j=1}^{4} a_j + (1-\rho)t\right] + \epsilon_t \equiv \gamma z_{1t} + \delta z_{2t} + \epsilon_t, \qquad (16) \\ &\qquad \epsilon_t \sim \text{i.d. } N(0, \sigma^2 v_t) \qquad (t = 1, \ldots, T), \end{aligned}$$

$$w_{T+1} \equiv y_0 = \gamma + \epsilon_{T+1} \equiv \gamma z_{1,T+1} + \epsilon_{T+1}, \qquad \epsilon_{T+1} \sim N(0, \sigma_\gamma^2),$$

$$w_{T+2} \equiv 0 = \delta + \epsilon_{T+2} \equiv \delta z_{2,T+2} + \epsilon_{T+2}, \qquad \epsilon_{T+2} \sim N(0, \sigma_\delta^2).$$

The conditional posterior distribution for $\gamma$ and $\delta$ is therefore bivariate normal, with mean and variance given by the usual generalized least-squares (GLS) expressions.

This content downloaded from 128.83.155.248 on Mon, 01 May 2017 03:10:07 UTC
All use subject to http://about.jstor.org/terms

*Conditional posterior distribution of $a_1, \ldots, a_4$.* Write

$$
\begin{aligned}
w_t &\equiv y_t - \gamma(1-\rho) - \delta\rho - \delta(1-\rho)t - \rho y_{t-1} \\
&= \sum_{j=1}^{4} a_j(y_{t-j} - y_{t-j-1} - \delta) \equiv \sum_{j=1}^{4} a_j z_{jt} + \epsilon_t, \qquad (17)\\
&\qquad \epsilon_t \sim \text{i.d. } N(0, \sigma^2 v_t) \qquad (t = 1, \ldots, T),\\
w_{T+j} &\equiv 0 = \rho_j + \epsilon_{T+j} \equiv \rho_j z_{j,T+j} + \epsilon_{T+j},\\
&\qquad \epsilon_{T+j} \sim \text{i.d. } N(0, \pi_0 \pi_1^{j-1}) \qquad (j = 1, \ldots, 4).
\end{aligned}
$$

Again, the posterior distribution is conditionally multivariate normal, with mean and variance given by the GLS expressions.

*Conditional posterior distribution of $\rho$.* The conditional posterior distribution results from the combination of the simple linear regression model

$$
\begin{aligned}
w_t &\equiv y_t - \gamma + \delta \sum_{j=1}^{4} a_j - \delta t - \sum_{j=1}^{4} a_j(y_{t-j} - y_{t-j-1})\\
&= \rho(y_{t-1} - \gamma + \delta - \delta t) + \epsilon_t \equiv \rho z_t + \epsilon_t,\\
&\qquad \epsilon_t \sim \text{i.d. } N(0, \sigma^2 v_t) \qquad (t = 1, \ldots, T),
\end{aligned}
$$

with the prior density from (12) for $\rho$. The conditional posterior distribution, therefore, has kernel density

$$\rho \exp\left[-\frac{(\rho - \hat{\rho}_1)^2}{2\lambda^2}\right] I_{[0,1)}(\rho), \qquad (18)$$

where $\hat{\rho} = \sum_{t=1}^{T} v_t^{-1} w_t z_t / \sum_{t=1}^{T} v_t^{-1} z_t^2$ and $\lambda^2 = \sigma^2 / \sum_{t=1}^{T} v_t^{-1} z_t^2$.

*Conditional posterior distribution of $\nu$.* From (9) and (15) this distribution has kernel density

$$\left(\frac{\nu}{2}\right)^{T\nu/2} \Gamma\left(\frac{\nu}{2}\right)^{-T} \exp(-\eta\nu), \qquad (19)$$

where $\eta = \frac{1}{2}\sum_{t=1}^{T} [\log(v_t) + v_t^{-1}] + \omega$.

*Conditional posterior distribution of $v_1, \ldots, v_T$.* From (9), the conditional posterior density of $\psi \equiv (\sigma^{-2}\epsilon_t^2 + \nu)/v_t$ is proportional to $\psi^{-(\nu-1)/2} \exp(-\psi/2)$. Hence,

$$\frac{\sigma^{-2}\epsilon_t^2 + \nu}{v_t} \sim \chi^2(\nu + 1). \qquad (20)$$

This result may be obtained heuristically by noting that in the prior distribution $\nu/v_t \sim \chi^2(\nu)$ the likelihood function for (7) $\sigma^{-2}\epsilon_t^2/v_t$ enters in the

This content downloaded from 128.83.155.248 on Mon, 01 May 2017 03:10:07 UTC
All use subject to http://about.jstor.org/terms

form of the kernel density of the $\chi^2(1)$ distribution and appeals to the reproductive property of the chi-square distribution.

*Conditional posterior distribution of $\sigma$.* Given all the other parameters, the posterior density kernel for $\sigma$ is

$$\sigma^{-(T+1)} \exp\left(-\sum_{t=1}^{T} \epsilon_t^2/2\sigma^2 v_t\right).$$

The kernel density of $\phi \equiv \sum_{t=1}^{T} (\epsilon_t^2/v_t)/\sigma^2$ is $\phi^{-(T+1)/2}\exp(-\phi/2)$. Consequently,

$$\sum_{t=1}^{T} (\epsilon_t^2/v_t)/\sigma^2 \sim \chi^2(T), \tag{21}$$

which has an obvious heuristic.

## 3. COMPUTATION OF POSTERIOR MOMENTS AND DENSITIES

In this study, the Gibbs sampler (Gelfand and Smith [7]) is used to produce a sequence of drawings from the parameter space that is neither independent nor identically distributed but converges in distribution to the posterior distribution whose kernel density is in (9). Consistent with the discussion in Section 2.2, adopt the following notation and groupings of parameters:

$$\theta_1' = (\gamma,\delta); \qquad \theta_4 = \nu;$$

$$\theta_2' = (\rho_2,\rho_3,\rho_4,\rho_5); \qquad \theta_5' = (v_1,\ldots,v_T);$$

$$\theta_3 = \rho; \qquad \theta_6 = \sigma;$$

$$\theta' = (\theta_1',\theta_2',\theta_3,\theta_4,\theta_5',\theta_6).$$

The Gibbs sampling algorithm for the posterior distribution is easy to construct. Begin with an arbitrary initial value

$$\begin{aligned}\theta^{(0)\prime} &= (\theta_1^{(0)\prime},\theta_2^{(0)\prime},\theta_3^{(0)},\theta_4^{(0)},\theta_5^{(0)\prime},\theta_6^{(0)})\\ &\in \Theta_1 \times \Theta_2 \times \Theta_3 \times \Theta_4 \times \Theta_5 \times \Theta_6\\ &= R^2 \times R^4 \times \{[0,1]\} \times R^+ \times R^{T+} \times R^+ = \Theta.\end{aligned}$$

A convenient choice is the ordinary least-squares estimate for $\theta_1$, $\theta_2$, and $\theta_3$ (forcing the appropriate constraint on $\rho$ if need be), $\nu = 4$, $\theta_5 = (1,1,\ldots,1)'$, $\theta_6 = s^2$. These initial values were used for all results reported in this paper, but any element of $\Theta$ may be chosen. Given $\theta^{(j)}$,

(i) draw $\theta_1^{(j)} = (\gamma^{(j)},\delta^{(j)})'$ from the bivariate normal distribution for $\gamma$ and $\delta$ indicated by the regression in (16);
(ii) draw $\theta_2^{(j)} = (\rho_2^{(j)},\rho_3^{(j)},\rho_4^{(j)},\rho_5^{(j)})'$ from the multivariate normal distribution for $\rho_2$, $\rho_3$, $\rho_4$, and $\rho_5$ indicated by the regression in (17);
(iii) draw $\rho$ from the distribution whose kernel density is given by (18).

This content downloaded from 128.83.155.248 on Mon, 01 May 2017 03:10:07 UTC
All use subject to http://about.jstor.org/terms

A computationally efficient method is described in [10, Appendix A].

(iv) draw $\nu$ from the distribution whose kernel density is given by (19); see [10, Appendix A].
(v) draw $v_1, \ldots, v_T$ successively and independently according to the conditional posterior distributions in (20).
(vi) draw $\sigma^2$ using (21).

These six steps constitute a single pass of the Gibbs sampler. After each pass, a function of interest $g(\theta^{(j)})$ can be computed, and after $m$ passes, $m^{-1}\sum_{j=1}^{m} g(\theta^{(j)})$ provides a numerical approximation to $E[g(\theta)]$.

This procedure is superficially similar to the EM algorithm, which has been used to maximize the likelihood function in a related, but simpler, situation by Lange et al. [16]. Leonard [17] used a similar approach to find an approximate posterior mode in a related problem. The superficial similarity stems from similar conditioning in each iteration. However, the Gibbs sampler produces the entire posterior distribution by defining a continuous-state Markov chain on $\Theta$. Each conditional density in the chain $p_i(\theta_i | \theta_r, r \neq i)$ is strictly positive for all $\theta_i$ and $\{\theta_r, r \neq i\}$. Theorem 3.8 of Nummelin [20] or Corollary 1 of Tierney [36] therefore implies that $\{\theta^{(j)}\}$ converges in distribution to the posterior distribution. The posterior moments of parameters like $\rho$ and $\delta$, posterior probabilities of intervals, and posterior odds ratios can all be cast in the form of a posterior expectation $E[g(\theta)]$ for an appropriate function of interest $g(\theta)$. So long as $E[|g(\theta)|] < \infty$, $\bar{g}_m \equiv m^{-1}\sum_{j=1}^{m} g(\theta^{(j)}) \rightarrow E[g(\theta)] \equiv \bar{g}$ (Revuz [27, Theorem 4.3.1]; Tierney [36, Theorem 3]).

The numerical accuracy of $\bar{g}_m$ as an approximation of $E[g(\theta)]$ may be assessed by exploiting the sampling theoretic result $m^{1/2}(\bar{g}_m - \bar{g}) \Rightarrow N[0, S(0)]$ (Hannan [12, Theorem 4.11]), where $S(0)$ is the spectral density of $\{g^{(j)}\}$ at frequency zero. Given a consistent estimate $\hat{S}_m(0)$ of $S(0)$, the numerical standard error (NSE) of $\bar{g}_m$, $(m^{-1}\hat{S}_m(0)]^{-1/2}$ may be used to assess the accuracy of $\bar{g}_m$ in the obvious way. A full elaboration is provided in Geweke [9], and technical details for the computations in the work reported here are given in Geweke [10].

### 3.1. Posterior Odds Ratios

Much of the recent empirical literature concerning unit roots has addressed hypotheses about $\rho$ or its conceptual equivalent in other models. The hypotheses studied have been (a) $\rho < 1$; (b) $\rho = 1$; (c) $\rho \geq 1$. For reasons well stated in Schotman and van Dijk [30], (b) is much more compelling than (c). Here we focus on (a) and (b); the prior distributions adapted in Section 2.1 further restrict $\rho \geq 0$ in the case of (a), but trivial variants on these procedures would easily cope with prior distributions for $\rho$ extending to $(-1, 0)$. The discussion in Section 2 considered only hypothesis (a). Here, we construct (b) as a limit of prior distributions under (a).

This content downloaded from 128.83.155.248 on Mon, 01 May 2017 03:10:07 UTC
All use subject to http://about.jstor.org/terms

Begin by considering the general case of alternative hypotheses for the same model with likelihood function $L(\theta)$, $\theta \in \Theta$, that can be described by alternative prior distributions with densities $\pi_A(\theta)$ and $\pi_B(\theta)$. To fix ideas, prior distributions for $\rho$ in the trend stationary model, with the density function from (12) and different values of $s$, are examples. In the general case, the posterior odds ratio in favor of hypothesis (a) is

$$\frac{\int_\Theta L(\theta)\pi_A(\theta)\,d\theta}{\int_\Theta L(\theta)\pi_B(\theta)\,d\theta} = \frac{\int_\Theta \left[\frac{\pi_A(\theta)}{\pi_B(\theta)}\right] L(\theta)\pi_B(\theta)\,d\theta}{\int_\Theta L(\theta)\pi_B(\theta)\,d\theta} = E_B\left[\frac{\pi_A(\theta)}{\pi_B}\right]. \tag{22}$$

Thus, the methods described above may be used to compute (22) by taking $g(\theta) = \pi_A(\theta)/\pi_B(\theta)$ as a function of interest. The quality of the approximation will depend on the behavior of $\pi_A(\theta)/\pi_B(\theta)$ on the part of $\Theta$ where the mass of the posterior density under hypothesis (b) is concentrated and, in general, will be better when the posterior variance of $\pi_A(\theta)/\pi_B(\theta)$, under hypothesis (b), is smaller. (In general, this variance need not even be finite, and this is a consideration in determining the roles of (a) and (b) in the computations.)

In the case of the alternative prior distributions,

$$\pi_A(\rho) = (t+1)\rho^t I_{[0,1)}(\rho), \qquad \pi_B(\rho) = (s+1)\rho^s I_{[0,1)}(\rho), \qquad t > s, \tag{23}$$

the function of interest pertinent to computation of the posterior odds ratio is

$$\left[\frac{t+1}{s+1}\right]\rho^{t-s}. \tag{24}$$

Because this function is bounded on the unit interval, all its posterior moments exist. If the roles of $t$ and $s$ are reversed, these moments, in general, will not exist, and hence the configuration in (23) is maintained in the empirical work.

In the case of the alternative distributions

$$\pi_A(\rho) = \epsilon^{-1} I_{(1-\epsilon,1)}(\rho), \qquad \pi_B(\rho) = (s+1)\rho^s I_{[0,1)}(\rho),$$

the function of interest $(s+1)^{-1}\rho^s\epsilon^{-1}I_{(1-\epsilon,1)}(\rho)$, which for small values of $\epsilon$ is indistinguishable from

$$(s+1)^{-1}\epsilon^{-1}I_{(1-\epsilon,1)}(\rho). \tag{25}$$

The Gibbs sampler may be applied to approximate the posterior expectation of (25), but as $\epsilon$ decreases, this method becomes increasingly inefficient because a very small fraction of the draws of $\rho$ from the distribution whose kernel density is given by (18) will occur in the interval $(1-\epsilon, 1)$. Follow-

This content downloaded from 128.83.155.248 on Mon, 01 May 2017 03:10:07 UTC
All use subject to http://about.jstor.org/terms

ing the discussion of Section 3.1, it is computationally much more efficient to choose as the function of interest the conditional expectation of (25):

$$(s+1)^{-1}\epsilon^{-1}\,\frac{\displaystyle\int_{1-\epsilon}^{1}\exp\left[\frac{-(\rho-\hat{\rho})^2}{2\lambda^2}\right]d\rho}{\displaystyle\int_{0}^{1}\rho^{s}\exp\left[\frac{-(\rho-\hat{\rho})^2}{2\lambda^2}\right]d\rho}. \tag{26}$$

As $\epsilon \to 0$, the posterior expectation of (25) approaches the posterior odds ratio in favor of $\rho = 1$. Taking the same limit in (26), the posterior odds ratio is the posterior expectation of the function of interest

$$(s+1)^{-1}\,\frac{\exp\left[\dfrac{-(1-\hat{\rho})^2}{2\lambda^2}\right]}{\displaystyle\int_{0}^{1}\rho_1^{s}\exp\left[\frac{-(\rho-\hat{\rho})^2}{2\lambda^2}\right]d\rho} \tag{27}$$

in the difference stationary model. This function of interest is computed in each iteration of the Gibbs sampler by using a 21-point Gauss–Kronrod rule (IMSL [14, pp. 569–572]).

### 3.2. Posterior Densities

For public reporting, presentation of posterior densities is often desirable. Because the Gibbs sampler generates a representative sample of points from the posterior distribution, these densities may be approximated by using conventional kernel density methods. But if the function of interest involves only a single $\theta_i$ from the sixfold partition of the parameter space, and the posterior density of the function of interest conditional on the remaining parameters is known analytically, then a much better approximation is possible. These conditions are satisfied here for the parameters $\rho$, $\delta$, and $\nu$, which constitute functions of interest of the parameter subvectors $\theta_3$, $\theta_1$, and $\theta_4$, respectively.

Suppose that $g(\theta)$ is a function of $\theta_i$ alone and that the posterior density function for $g(\theta)$ conditional on $\{\theta_r, r \neq i\}$, $p_g(d|\{\theta_r, r \neq i\})$ is known. The unconditional posterior density function for $g(\theta)$ is $p_g(d) = E[p_g(d|\{\theta_r, r \neq i\}]$. This expectation may be approximated by the Gibbs sampler in the same way as that of any other function of interest. Moreover, in most cases smoothness of $p_g(d)$ is reflected in $p_g(d|\{\theta_r, r \neq i\})$, and the numerical approximation of $p_g(d)$ will therefore be appropriately smooth. Precisely this method is used to provide (very good) numerical approximations to the univariate posterior densities of $\rho$, $\delta$, and $\nu$ in Section 4.3. (In the cases of $\rho$ and $\nu$, numerical one-dimensional integration of the kernel densities from (18) and (19), respectively, is required. That for $\rho$ is available from (26), and

This content downloaded from 128.83.155.248 on Mon, 01 May 2017 03:10:07 UTC
All use subject to http://about.jstor.org/terms

that for $\nu$ is computed by transformation of $(0, \infty)$ into $(0,1)$ followed by evaluation by using a 21-point Gauss–Kronrod rule (IMSL [14, pp. 577–580]).)

If two functions of interest are functions of the same $\theta_i$ alone, and if the conditional bivariate posterior density function for the functions of interest is known analytically, the same method may be applied. Furthermore, the numerical approximations of the densities may be used in conjunction with the Gibbs sample itself to compute highest posterior density regions to arbitrary accuracy. Given the Gibbs sample $\{\theta^{(j)}\}_{j=1}^{m}$, compute the corresponding approximations to the probability densities evaluated at these points: $p^{(q)} \equiv \hat{p}[\theta_i^{(q)}] \equiv m^{-1} \sum_{j=1}^{m} p_g(\theta_i^{(q)} | \{\theta_r^{(j)}, r \neq i\})$. Then sort the $p^{(q)}$ into ascending order, and compute the $\alpha$th quantile $p_\alpha^*$ in the obvious way. An approximate $100(1-\alpha)\%$ highest posterior density region consists of all $d$ for which $m^{-1} \sum_{j=1}^{m} p_g(d | \{\theta_r^{(j)}, r \neq i\}) > p_\alpha^*$. This procedure produces the exact $100(1-\alpha)\%$ highest posterior density region as $m \to \infty$.

The bivariate posterior density of greatest interest here is that of $(\rho, \delta)$. The procedure just described may still be applied, but this is complicated by the fact that the joint conditional distribution of $\rho$ and $\delta$ involves two subvectors of the parameter space, $\theta_1$ and $\theta_3$. The principal idea is to express the conditional density function for $\rho$ and $\delta$ by using a combination of analytical and numerical integration techniques. Sufficient statistics for these conditional distributions are recorded in each pass of the Gibbs sampler, and the bivariate density and highest posterior density regions are then constructed at the end. Technical details are provided in Appendix B of Geweke [10].

## 4. EMPIRICAL RESULTS FOR THE NELSON–PLOSSER DATA SET

These methods were applied to six of the time series studied by Nelson and Plosser [19]: real GNP, nominal GNP, real per capita GNP, unemployment, consumer prices, and velocity. Data were furnished by Charles Nelson, and the least-squares estimates reported in Nelson and Plosser [19, Table 5] were reproduced to all reported places. The sample period for the results here is the same as that used by Nelson and Plosser [19], except that a few early observations could not be used because the model here involves five values of the lagged dependent variable, whereas the number of lags used by Nelson and Plosser [19] varied but did not exceed four for any of these six series.

### 4.1. Sensitivity to the Prior Distribution

Examination of the sensitivity of results to the prior distribution for $\rho$ is a principal objective of this research, which is taken up in Sections 4.2 and 4.3. Before presenting this analysis, we report briefly on sensitivity to other parameters of the prior distribution subsequently fixed (Table 1).

This content downloaded from 128.83.155.248 on Mon, 01 May 2017 03:10:07 UTC
All use subject to http://about.jstor.org/terms

**TABLE 1.** Sensitivity of some posterior moments to some parameters of the prior distribution

| | Posterior odds ratios in favor of | | $\rho$ | | $\delta \times 100$ | | $\nu$ | |
|---|---|---|---|---|---|---|---|---|
| | $\rho = 1$ | Next $\acute{s}$ | Mean | S.D. | Mean | S.D. | Mean | S.D. |
| Real GNP, $s = 0$ | | | | | | | | |
| Base case[a] | 1.6 | 2.95 | .848 | (.076) | 3.067 | (.275) | 5.5 | (3.6) |
| S.D.$(a_1) = 1.0$ | 1.5 | 2.80 | .840 | (.080) | 3.123 | (.279) | 5.0 | (3.5) |
| S.D.$(a_1) = .25$ | 1.5 | 3.07 | .856 | (.070) | 3.076 | (.292) | 5.8 | (3.4) |
| S.D.$(a_1) = .005$ | 2.0 | 3.57 | .874 | (.065) | 3.104 | (.300) | 5.9 | (3.7) |
| S.D.$(a_j) = .5$ | 2.0 | 3.05 | .849 | (.079) | 3.105 | (.276) | 4.7 | (3.4) |
| $\omega = .05$ | 2.1 | 3.14 | .854 | (.077) | 3.090 | (.326) | 16 | (16) |
| $\omega = 1.0$ | .62 | 2.32 | .826 | (.073) | 3.125 | (.227) | 2.75 | (1.12) |
| $\sigma_\delta = 1.0$ | 1.3 | 2.72 | .840 | (.076) | 3.121 | (.271) | 5.3 | (3.6) |
| $\bar{\delta} = .03$, $\sigma_\delta = .02$ | 1.0 | 2.59 | .836 | (.075) | 3.140 | (.267) | 5.3 | (3.5) |
| $\sigma_\gamma = .10$ | 1.6 | 2.77 | .841 | (.076) | 3.102 | (.285) | 5.4 | (3.4) |
| $\sigma_\gamma = 1{,}000$ | 1.4 | 2.85 | .844 | (.076) | 3.092 | (.278) | 5.6 | (3.7) |
| Real GNP, $s = 9$ | | | | | | | | |
| Base case[a] | .53 | .74 | .900 | (.064) | 3.046 | (.377) | 6.0 | (3.6) |
| S.D.$(a_1) = 1.0$ | .64 | .82 | .906 | (.065) | 3.052 | (.378) | 5.9 | (4.0) |
| S.D.$(a_1) = .25$ | .44 | .68 | .898 | (.062) | 3.072 | (.416) | 6.2 | (3.8) |
| S.D.$(a_1) = .005$ | .56 | .84 | .914 | (.056) | 3.103 | (.381) | 6.6 | (3.9) |
| S.D.$(a_j) = .5$ | .65 | .84 | .907 | (.065) | 3.090 | (.374) | 5.1 | (3.6) |
| $\omega = .05$ | .66 | .84 | .908 | (.064) | 3.038 | (.436) | 20 | (19) |
| $\omega = 1.0$ | .35 | .61 | .887 | (.067) | 3.075 | (.307) | 2.9 | (1.1) |
| $\sigma_\delta = 1.0$ | .56 | .78 | .903 | (.064) | 3.057 | (.387) | 6.0 | (3.8) |
| $\bar{\delta} = .03$, $\sigma_\delta = .02$ | .46 | .70 | .899 | (.067) | 3.080 | (.354) | 5.9 | (3.9) |
| $\sigma_\gamma = .10$ | .55 | .73 | .898 | (.067) | 3.068 | (.386) | 6.2 | (4.1) |
| $\sigma_\gamma = 1{,}000$ | .53 | .72 | .898 | (.066) | 3.070 | (.394) | 5.9 | (3.7) |

[a]Configuration of prior parameters for the base case: for $a_j$ ($j = 1,\ldots,4$), priors are independent zero-mean normal, with standard deviations declining geometrically from 0.5 for $a_1$ to 0.1 for $a_4$; for $\gamma$, prior is normal with mean $\bar{\gamma} = y_0$ and standard deviation $\sigma_\gamma = 10.0$; for $\delta$, prior is normal with mean $\bar{\delta} = 0$ and standard deviation $\sigma_\delta = .05$; for $\nu$, prior is exponential with parameter $\omega = .25$ (mean $\omega^{-1} = 4.0$). The prior density for $\rho$ is $(s + 1)\rho^s$ on the unit interval.

For $j = 2,\ldots,5$, the prior distribution of $\rho_j$ is i.d. $N(0, \pi_0\pi_1^{j-1})$. For the empirical work with the six time series, $\pi_0 = .731$ and $\pi_1 = .342$, which implies a standard deviation of .5 for $\rho_2$, .1 for $\rho_5$, and geometrically declining standard deviations in between. This is the "base case" of Table 1. We examine four alternative settings of these parameters while keeping the other parameters of the prior distribution fixed at the values used in the empiri-

This content downloaded from 128.83.155.248 on Mon, 01 May 2017 03:10:07 UTC
All use subject to http://about.jstor.org/terms

cal work. First, $\pi_0$ is increased by a factor of 4, doubling all standard deviations; second, $\pi_0$ is decreased by a factor of 4; third, $\pi_0$ is decreased to $2.5 \times 10^{-5}$ so that the prior standard deviation of $\rho_2$ is .005, thus effectively constraining the coefficients on all $y_{t-j+1} - y_{t-j}$ to be zero; fourth, $\pi_0 = .25$ and $\pi_1 = 1.0$, so that the prior distribution for each of these coefficients is $N(0,.5^2)$. These four settings correspond to the four lines below the "base case" line in Table 1.

The prior distribution for the degrees-of-freedom parameter $\nu$ of the Student-$t$ density of the disturbances is exponential with parameter $\omega$ and therefore has mean $\omega^{-1}$. In the empirical work and "base case," $\omega = .25$. We examine two settings: $\omega = .05$ (mean 20, or "thin tails") and $\omega = 1.0$ (mean 1, or "fat tails"). These two alternative settings correspond to lines five and six below the "base case" in Table 1.

The prior distribution for the trend coefficient $\delta$ is $N(\bar{\delta}, \sigma_\delta^2)$. In the empirical work and "base case," the distribution has mean 0 and standard deviation .05. Because the data are in logarithms, this corresponds to a centered 95% prior confidence interval that extends from a growth rate of −10% to one of +10%. Therefore, this prior distribution is rather diffuse. We examine two alternative prior distributions for $\delta$: $N(0, 1)$, which is even more diffuse, and $N(.03,.02^2)$, which is less diffuse and approximately centered on the mean growth rate of real GNP over the sample period.

The prior distribution for the intercept term is $N(y_0, 10^2)$. Because the data are in logarithms, a centered 95% confidence interval for $\gamma$ has a range equivalent to over 600 years of growth in real GNP at the observed mean growth rate of about 3%. This prior distribution is thus very diffuse. In the last two lines of each panel of Table 1, we examine two alternative prior distributions: $N(y_0, 0.1^2)$, in which the standard deviation amounts to a few years of real GNP growth, and $N(y_0, 1{,}000^2)$, which is flat for all practical purposes.

The effects of these alternative settings on eight posterior moments were examined by employing the Gibbs sampler as described in Section 3. (Some technical details of implementation are provided in Geweke [10, Appendix C].) For the posterior odds ratios and posterior expectations, the number of figures reported is, at most, one more than warranted by the numerical standard error computed as described in Section 3. The posterior odds ratio in favor of $\rho = 1$ is computed as the posterior expectation of the function of interest in (27). The posterior odds ratio in favor of "Next $s$" is computed as the expected value of the function of interest in (24) by using $s = 0$ and $t = 9$ when $s = 0$ (Table 1, top) and $s = 9$ and $t = 29$ when $s = 9$ (Table 1, bottom). Posterior means and standard deviations for the other parameters are computed in a straightforward fashion. Several observations may be made about the results reported in Table 1.

The alternative prior distributions for $\gamma$ (last two lines of each panel in Table 1) have essentially no effect on posterior moments. Differences are of the same order of magnitude as numerical standard errors. Changes in the

This content downloaded from 128.83.155.248 on Mon, 01 May 2017 03:10:07 UTC
All use subject to http://about.jstor.org/terms

prior distributions of the $a_j$ have scarcely larger impacts. Effectively eliminating these parameters by setting their prior standard deviations to .005 or less increases the posterior mean of $\rho$ and the posterior odds ratio in favor of $\rho = 1$ slightly, but beyond this no systematic effects are evident. A more diffuse prior distribution for $\delta$ ($\sigma_\delta = 1.0$) has no discernible effect, but a more informative distribution centered on the actual growth rate ($\bar{\delta} = .03$, $\sigma_\delta = .02$) diminishes the posterior probability of the unit root hypothesis. The latter effect is explained in Section 4.3.

Changes in the prior distribution for $\nu$, the degrees of freedom parameter in the Student-$t$ distribution of the disturbances, have large and systematic effects. When greater prior probability is given to lower degrees of freedom ($\omega = 1.0$), then the posterior odds in favor of $\rho = 1$ drop by more than one-half for $s = 0$ and by one-third for $s = 9$, relative to the base case ($\omega = 4.0$). When greater prior weight is given to higher degrees of freedom ($\omega = .05$), odds ratios in favor of $\rho = 1$ increase by about one-fourth in each panel. If one identifies highly leptokurtic distributions with large outliers, then these results seem consistent with Perron's [22] study, which finds frequentist evidence in favor of unit roots substantially weakened by allowing for structural breaks. Detailed evidence presented in Geweke [11] shows that, given a flat prior for $\nu$, the mode of the posterior density for $\nu$ occurs between 3 and 5 for most of the time series of Nelson and Plosser [19]. Therefore, this work proceeds with a prior for $\nu$ whose mean is $\nu = 4$.

### 4.2. Posterior Odds Ratios and Moments

By using the "base case" priors for all other parameters, posterior odds ratios and moments were computed for the six indicated macroeconomic time series of Nelson and Plosser [19]. Six different prior distributions indexed by $s$ were employed for the autoregressive coefficient $\rho$. As explained in Section 2.1, the choices for $s$ correspond approximately to prior densities for the autoregressive coefficient that are flat on the unit interval for data recorded at various hypothetical intervals: $s = -11/12$, monthly; $-3/4$, quarterly; 0, annually; 9, every decade; 29, every 30 years; and 99, every century. Of course, the actual data used are annual in each case.

Results are reported in Table 2. For the posterior odds ratios, "Next $s$" refers to the value of $s$ in the next row: e.g., in the row labeled $s = -3/4$, the odds ratio is in favor of the prior specification with $s = 0$. Simple arithmetic shows that, except for error due to numerical approximation, the "Next $s$" odds ratio should be the ratio of the "$\rho = 1$" odds ratio for that row to the "$\rho = 1$" odds ratio for the next row, a relationship that is borne out up to the number of places accuracy that numerical standard errors would indicate. These indicators of numerical accuracy are not reported here, but they are used to choose the number of digits reported in Table 2, just as they were in Table 1.

This content downloaded from 128.83.155.248 on Mon, 01 May 2017 03:10:07 UTC
All use subject to http://about.jstor.org/terms

**Table 2.** Posterior odds ratios and moments for six macroeconomic time series[a]

| | Posterior odds ratios in favor of | | $\rho$ | | $\delta \times 100$ | | $\nu$ | |
|---|---|---|---|---|---|---|---|---|
| $s$ | $\rho = 1$ | Next $s$ | Mean | S.D. | Mean | S.D. | Mean | S.D. |
| Real GNP | | | | | | | | |
| −11/12 | 15 | 2.913 | .841 | (.078) | 3.082 | (.264) | 5.4 | (3.9) |
| −3/4 | 5.8 | 3.516 | .841 | (.076) | 3.089 | (.278) | 5.4 | (3.6) |
| 0 | 1.6 | 2.95 | .848 | (.076) | 3.067 | (.275) | 5.5 | (3.6) |
| 9 | .53 | .74 | .900 | (.064) | 3.046 | (.377) | 6.0 | (3.6) |
| 29 | .70 | .75 | .956 | (.039) | 3.070 | (.556) | 6.2 | (3.8) |
| 99 | .931 | | .9894 | (.0107) | 3.098 | (.795) | 7.0 | (4.1) |
| Nominal GNP | | | | | | | | |
| −11/12 | 77.9 | 2.9699 | .9416 | (.0355) | 5.604 | (.612) | 3.20 | (1.75) |
| −3/4 | 25.5 | 3.825 | .9423 | (.0347) | 5.559 | (.638) | 3.11 | (1.58) |
| 0 | 6.83 | 6.17 | .9430 | (.0345) | 5.603 | (.607) | 3.26 | (1.80) |
| 9 | 1.22 | 1.394 | .9539 | (.0317) | 5.65 | (.69) | 3.23 | (1.76) |
| 29 | .842 | .92 | .9695 | (.0239) | 5.76 | (.79) | 3.24 | (1.87) |
| 99 | .891 | | .9891 | (.0103) | 6.00 | (1.03) | 3.4 | (2.2) |
| Real per capita GNP | | | | | | | | |
| −11/12 | 12.3 | 2.906 | .829 | (.079) | 1.774 | (.253) | 5.5 | (1.6) |
| −3/4 | 5.106 | 3.479 | .831 | (.081) | 1.786 | (.268) | 5.5 | (3.3) |
| 0 | 1.032 | 2.42 | .826 | (.079) | 1.849 | (.270) | 6.0 | (4.0) |
| 9 | .526 | .72 | .896 | (.068) | 1.739 | (.378) | 6.1 | (3.8) |
| 29 | .729 | .77 | .9568 | (.0383) | 1.736 | (.572) | 6.9 | (4.2) |
| 99 | .932 | | .9892 | (.0110) | 1.78 | (.81) | 6.9 | (4.3) |
| Unemployment | | | | | | | | |
| −11/12 | 9.77 | 2.8886 | .801 | (.089) | −1.15 | (1.03) | 6.0 | (3.4) |
| −3/4 | 2.51 | 3.370 | .797 | (.090) | −1.02 | (1.02) | 6.6 | (3.8) |
| 0 | .691 | 2.15 | .809 | (.080) | −1.13 | (1.03) | 6.0 | (3.4) |
| 9 | .360 | .603 | .881 | (.074) | −1.33 | (1.43) | 5.9 | (3.4) |
| 29 | .599 | .701 | .9543 | (.0391) | −1.59 | (2.04) | 6.4 | (3.9) |
| 99 | .861 | | .9888 | (.0111) | −1.28 | (2.75) | 6.6 | (4.0) |
| Consumer prices | | | | | | | | |
| −11/12 | $1.58 \times 10^3$ | 2.9976 | .9950 | (.0041) | 1.134 | (.642) | 2.61 | (1.07) |
| −3/4 | $5.1 \times 10^2$ | 3.9845 | .9948 | (.0043) | 1.105 | (.598) | 2.69 | (1.19) |
| 0 | 126 | 9.560 | .9949 | (.0041) | 1.118 | (.644) | 2.67 | (1.10) |
| 9 | 14.1 | 2.736 | .9953 | (.0039) | 1.132 | (.681) | 2.57 | (1.02) |
| 29 | 5.13 | 2.53 | .9957 | (.0036) | 1.138 | (.626) | 2.48 | (0.97) |
| 99 | 1.95 | 1.44 | .9963 | (.0032) | 1.192 | (.681) | 2.54 | (0.98) |

*(continued)*

This content downloaded from 128.83.155.248 on Mon, 01 May 2017 03:10:07 UTC
All use subject to http://about.jstor.org/terms

**TABLE 2.** Continued

| $s$ | Posterior odds ratios in favor of $\rho = 1$ | Next $s$ | $\rho$ Mean | $\rho$ S.D. | $\delta \times 100$ Mean | $\delta \times 100$ S.D. | $\nu$ Mean | $\nu$ S.D. |
|---|---|---|---|---|---|---|---|---|
| Velocity | | | | | | | | |
| $-11/12$ | 158 | 2.9774 | .9560 | (.0321) | $-.891$ | (.371) | 7.2 | (4.1) |
| $-3/4$ | 59 | 3.876 | .959 | (.031) | $-.941$ | (.366) | 7.6 | (4.3) |
| 0 | 14.8 | 7.08 | .959 | (.031) | $-.922$ | (.372) | 7.3 | (4.4) |
| 9 | 2.09 | 1.712 | .9661 | (.0277) | $-.930$ | (.401) | 7.5 | (4.4) |
| 29 | 1.22 | 1.25 | .9780 | (.0198) | $-.911$ | (.435) | 7.7 | (4.3) |
| 99 | .993 | 1.012 | .9906 | (.0090) | $-.890$ | (.522) | 8.3 | (4.7) |

[a]For the posterior odds ratios and posterior expectations, at most the rightmost digit is uncertain because of the inaccuracy of the numerical approximation, as indicated by the numerical standard error, which was computed but is not reported here.

As the prior parameter $s$ increases, the posterior mean of $\rho$ increases, and its posterior standard deviation decreases monotonically (within the limits of numerical accuracy) in every case. The posterior odds ratio in favor of $\rho = 1$ shows a general tendency to move toward 1 as it must in the limit, but the increase is not monotone for the first four series. For these series, the evidence against the unit root hypothesis is strongest when $s = 9$ or $s = 29$.

Comparison of results for different time series in Table 2 is generally consistent with the findings of other investigators for these data: e.g., unemployment and real per capita GNP show less evidence of difference stationarity than do consumer prices and velocity. Specific comparisons provide more insight. The only published work, to my knowledge, that reports posterior probabilities or odds ratios for these data involving the hypothesis $\rho = 1$ is found in Schotman and van Dijk [30] and Phillips [25]. Schotman and van Dijk use essentially the same model as the one here and a prior distribution for $\rho$ that is uniform on [.8, 1.0) under the hypothesis of trend stationarity. They report posterior odds ratios in favor of a unit root as follows: real GNP, .57; nominal GNP, 1.3; real per capita GNP, .53; unemployment, .20; consumer prices, 7.6; velocity, 3.1. Except for unemployment, these figures are very close to the posterior odds ratios corresponding to $s = 9$ reported in Table 2. In view of the placement of the Schotman–van Dijk prior and the prior density $10\rho^9$ in the case $s = 9$, this correspondence is quite reasonable. Any differences between Schotman and van Dijk [31] and Table 2 seem plausibly attributable to the richer model specification and different priors used here. Phillips [25] uses a variant of the model (4) and obtains posterior odds ratios in favor of a unit root: real GNP, .34; nominal GNP, 3.4; real per cap-

This content downloaded from 128.83.155.248 on Mon, 01 May 2017 03:10:07 UTC
All use subject to http://about.jstor.org/terms

ita GNP, .25; unemployment, .015; consumer prices, 7.7; velocity, 2.0. These findings are substantially different from any of those reported in Table 2 and from the results of Schotman and van Dijk, which indicates (as one would expect) the sensitivity of the posterior odds ratio to the specification of the model and prior distributions.

Posterior moments for $\rho$ and $\delta$ are consistent with those reported by Schotman and van Dijk [31] and with parameter estimates for more distantly related models taken up by other investigators. Posterior means and variances for $\nu$ provide new evidence on the dispersion of the disturbances for these time series. For consumer prices, these moments strongly suggest a highly leptokurtic distribution, for which fourth moments do not exist. Nominal GNP disturbances are almost as leptokurtic, whereas for the other series the posterior expectation of $\nu$ exceeds the prior mean of 4.0.

### 4.3. Posterior Densities and Highest Posterior Density Regions

Aspects of posterior densities for real GNP are presented in Figures 1–3. Each figure consists of four panels. The upper-right panels and the two lower panels each present a prior density (thin line) and a marginal posterior density (heavy line) for the parameter indicated on the horizontal axis. All of these densities are proper and normalized, i.e., they integrate to one. The upper-left panels present highest posterior density regions for the joint distribution of the trend coefficient, $\delta$, and the autoregressive coefficient, $\rho$. The interior of the contour line labeled "1" has posterior probability .99, the line labeled "5" has posterior probability .60; other probabilities are listed with the caption for Figure 1. Scales differ from one figure to the next, and axes must be examined when making comparisons, especially for the bivariate marginal densities in the upper-left panels.

None of these densities is even approximately Gaussian. Marginal densities for $\delta$ are nearly symmetric, but clearly leptokurtic. As $\rho \to 1$, $\text{var}(\delta | \rho)$ increases. The source of the increase may be found in the vanishing sample variance of the term $(1 - \rho)t$ of the reduced-form equation (5) as $\rho \to 1$. The effects may be seen in a comparison of the marginal posterior densities for $\delta$ plotted in the upper-right panels: as $s$ increases, so does dispersion in $\delta$ (see also Table 1). The effects are also discernible if the bivariate densities in the upper-left panels are examined closely: as $\rho$ increases, so does the relative dispersion in a horizontal "slice" of the density.

An interesting aspect of the posterior distributions is the asymmetry of the bivariate density with respect to $\rho$. So too is the fact that the marginal posterior density of $\rho$ either has a local minimum near (but not at) $\rho = 1$, or else it increases monotonically despite the existence of only a single interior mode of the bivariate density for $\rho$ and $\delta$. The source of this behavior may be found by considering the bivariate densities. Condition on $\delta$, and let $\hat{\delta}$ denote the posterior mean of $\delta$. As $|\delta - \hat{\delta}|$ increases, the deteriorating "fit" of the trend

This content downloaded from 128.83.155.248 on Mon, 01 May 2017 03:10:07 UTC
All use subject to http://about.jstor.org/terms

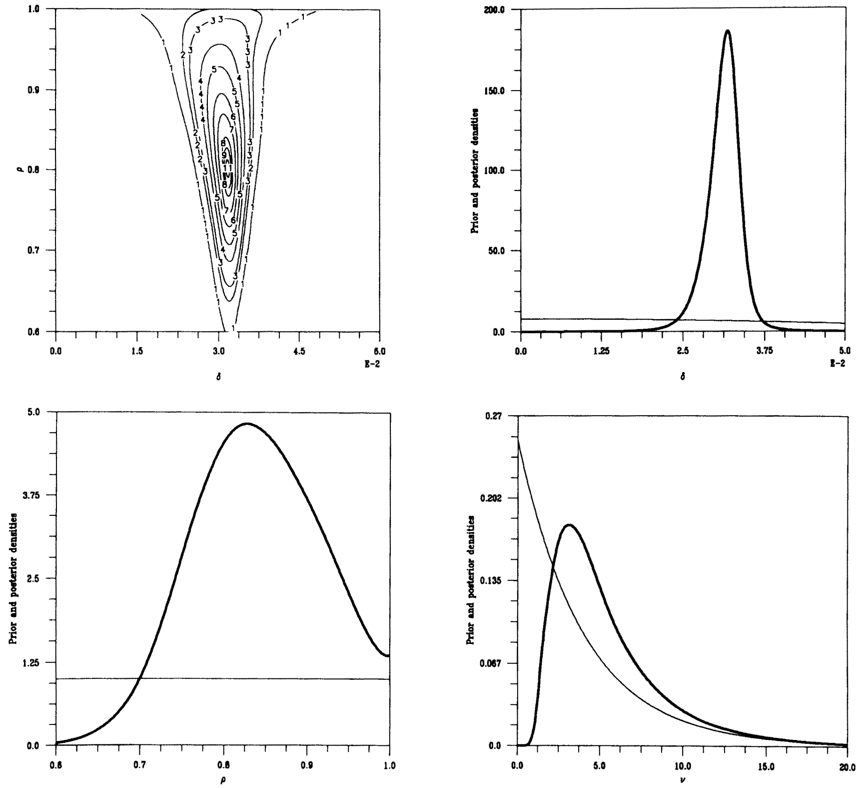


FIGURE 1. Prior and posterior densities for real GNP, $s = 0$.

KEY TO FIGURES 1–3. These figures each have four panels. The upper-right panels and the two lower panels each present a prior density (thin line) and a marginal posterior density (heavy line) for the parameter indicated on the horizontal axis. All of these densities are proper and normalized, i.e., they integrate to one. The upper-left panels present highest posterior density regions for the joint distribution of the trend coefficient, $\delta$, and the autoregressive coefficient, $\rho_1$. The correspondence between the numbered contour lines and the probabilities for this panel is: 1, .99; 2, .95; 3, .90; 4, .75; 5, .60; 6, .40; 7, .25; 8, .10; 9, .05; 10, .01.

line $\gamma + \delta t$ increases the probability of more persistent departures from trend. This is exhibited in the upward shift of mass in the bivariate marginal density along a vertical line, as that line is moved left or right of the center of mass. This effect is also evident in Table 1, where the prior distribution for $\delta$ centered at .03 with a standard deviation of .02 provides a smaller posterior mean for $\rho$ and a lower posterior odds ratio in favor of $\rho = 1$ than in the base case.

This content downloaded from 128.83.155.248 on Mon, 01 May 2017 03:10:07 UTC
All use subject to http://about.jstor.org/terms

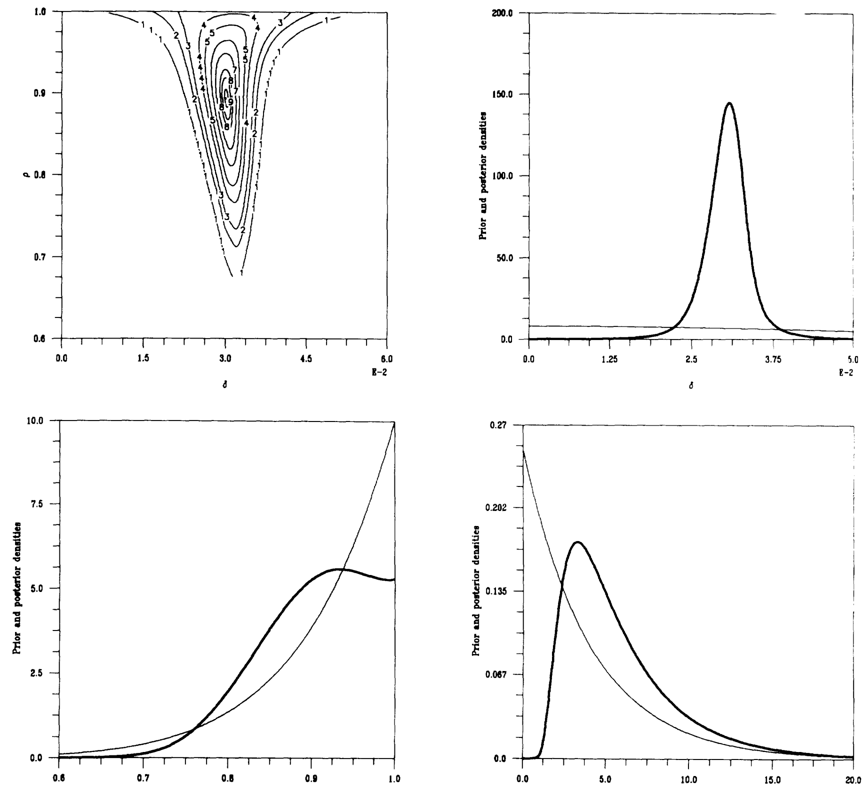


**FIGURE 2.** Prior and posterior densities for real GNP, $s = 9$.

The marginal posterior density for $\nu$ is strongly skewed and little affected by changes in $s$. The posterior density has essentially no mass on $\nu \in (0,1)$, the posterior probability density to the right of $\nu = 3$ exceeds the prior probability density, and the ratio of posterior to prior density attains its maximum ratio of almost 2:1 at around $\nu = 5$. These results are consistent with the strong evidence for leptokurtosis reported in Geweke [11].

Corresponding posterior densities for the other five time series are qualitatively similar, but with changes in location and scale suggested by the posterior means and standard deviations presented in Table 2. These densities are displayed in Geweke [10].

## 5. CONCLUSION

The main technical contribution of this work is to the practical application of Bayesian methods to macroeconomic time series. Beginning with a non-

This content downloaded from 128.83.155.248 on Mon, 01 May 2017 03:10:07 UTC
All use subject to http://about.jstor.org/terms

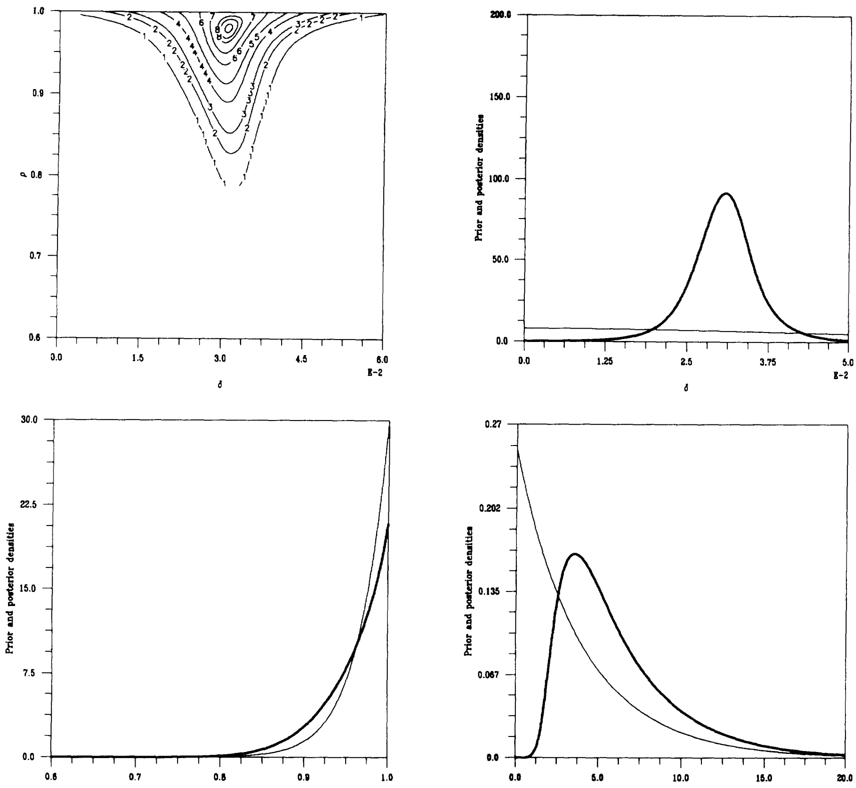


**FIGURE 3.** Prior and posterior densities for real GNP, $s = 29$.

conjugate prior distribution in a nonlinear model with leptokurtic disturbances, posterior moments, odds ratios, and highest posterior densities may be computed methodically. With respect to the problem, a single-parameter prior distribution for the key parameter in the trend stationary model was introduced, with the parameter implied by the answer to the question, At what time interval is a uniform prior density on the unit interval for the autoregressive parameter plausible? For four of the time series examined, the posterior odds ratio in favor of difference stationarity is smallest when this time interval is in the range of 10 to 20 years. The ratios over this range run about 2:1 in favor of trend stationarity for real GNP, real per capita GNP, and unemployment, and barely above 1:1 for nominal GNP. For consumer prices and velocity, the posterior odds heavily favor difference stationarity over this range. As the time interval becomes shorter, the prior distribution makes persistence sufficiently implausible, conditional on trend stationarity, that difference stationarity receives most of the posterior probability. As the time

This content downloaded from 128.83.155.248 on Mon, 01 May 2017 03:10:07 UTC
All use subject to http://about.jstor.org/terms

interval becomes longer, the posterior odds ratio must necessarily converge to 1:1, and that effect is evident for all six time series. This sensitivity reflects the fact that the hypotheses of trend and difference stationarity address the "long run," for which there is never even a single complete observation. The macroeconomic time series record is short enough that different reasonable prior distributions may dominate the data and imply different posterior odds ratios.

These conclusions appear sensitive to only one other aspect of the prior distribution, that pertaining to the degree of leptokurtosis in the disturbance term. Methods introduced in Geweke [11] permitted the specification of independent, identically distributed Student-$t$ disturbances. There is evidence that posterior odds in favor of difference stationarity decline as lower degrees of freedom are given greater prior probability. This sensitivity bears further investigation, perhaps by using a wider array of disturbance distributions than was the case here.

## REFERENCES


1. De Finetti, B. The Bayesian approach to the rejection of outliers. *Proceedings of the Fourth Berkeley Symposium on Mathematical Probability and Statistics, Vol. 1*, pp. 199–210. Berkeley: University of California Press, 1961.
2. DeJong, D.N. & C.H. Whiteman. Reconsidering "Trends and random walks in macroeconomic time series." *Journal of Monetary Economics* 28 (1991): 221–254.
3. Dickey, D.A. & W.A. Fuller. Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica* 49 (1981): 1057–1072.
4. Doan, T., R. Litterman & C.A. Sims. Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews* 3(1) (1984): 1–56.
5. Fraser, D.A.S. Necessary analysis and adaptive inference (with discussion). *Journal of the American Statistical Association* 71 (1976): 99–113.
6. Fraser, D.A.S. *Inference and Linear Models.* New York: McGraw-Hill, 1979.
7. Gelfand, A.E. & A.F.M. Smith. Sampling based approaches to calculating marginal densities (with discussion). *Journal of the American Statistical Association* 85 (1990): 398–409.
8. Geman, S. & D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6 (1984): 721–741.
9. Geweke, J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J.O. Berger, J.M. Bernardo, A.P. Dawid & A.F.M. Smith (eds.), *Bayesian Statistics 4*, pp. 169–194. Oxford: Oxford University Press, 1992.
10. Geweke, J. Priors for Macroeconomic Time Series and Their Application. Minneapolis: Federal Reserve Bank of Minneapolis, Institute for Empirical Macroeconomics discussion paper #64, 1992.
11. Geweke, J. Bayesian treatment of the independent Student-$t$ linear model. *Journal of Applied Econometrics* 8 (1993): 519–540.
12. Hannan, E.J. *Multiple Time Series.* New York: Wiley, 1970.
13. Harrison, P.J. & C.F. Stevens. Bayesian forecasting (with discussion). *Journal of the Royal Statistical Society Series B* 38 (1976): 205–247.
14. IMSL. *IMSL Math/Library User's Manual: FORTRAN Subroutines for Mathematical Applications.* Houston: Author, 1989.
15. Jeffreys, H. *Theory of Probability.* Oxford: Clarendon Press, 1939.


This content downloaded from 128.83.155.248 on Mon, 01 May 2017 03:10:07 UTC
All use subject to http://about.jstor.org/terms

16. Lange, K.L., R.J.A. Little & J.M.G. Taylor. Robust statistical modeling using the *t* distribution. *Journal of the American Statistical Association* 84 (1989): 881-896.
17. Leonard, T. A Bayesian approach to the linear model with unequal variances. *Technometrics* 17 (1975): 95-102.
18. Maronna, R.A. Robust M-estimators of multivariate location and scatter. *Annals of Statistics* 4 (1976): 51-67.
19. Nelson, C.R. & C.I. Plosser. Trends and random walks in macroeconomic time series: Some evidence and implications. *Journal of Monetary Economics* 10 (1982): 139-162.
20. Nummelin, E. *General Irreducible Markov Chains and Non-Negative Operators.* Cambridge: Cambridge University Press, 1984.
21. Osiewalski, J. & M.F.J. Steel. Robust Bayesian inference in elliptical regression models. *Journal of Econometrics* 57 (1993): 345-363.
22. Perron, P. The great crash, the oil price shock, and the unit root hypothesis. *Econometrica* 57 (1989): 1361-1401.
23. Phillips, P.C.B. Time series regressions with a unit root. *Econometrica* 55 (1987): 277-301.
24. Phillips, P.C.B. To criticize the critics: An objective Bayesian analysis of stochastic trends. *Journal of Applied Econometrics* 6 (1991): 333-364.
25. Phillips, P.C.B. Bayesian routes and unit roots: De rebus prioribus semper est disputandum. *Journal of Applied Econometrics* 6 (1991): 435-473.
26. Ramsay, J.O. & M.R. Novick. PLU robust Bayesian decision theory: Point estimation. *Journal of the American Statistical Association* 75 (1980): 901-907.
27. Revuz, D. *Markov Chains.* Amsterdam: North-Holland, 1975.
28. Said, S.E. & D.A. Dickey. Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* 71 (1984): 599-607.
29. Schotman, P.C. & H.K. van Dijk. A Bayesian analysis of the unit root in real exchange rates. *Journal of Econometrics* 49 (1991): 195-238.
30. Schotman, P.C. & H.K. van Dijk. On Bayesian routes to unit roots. *Journal of Applied Econometrics* 6 (1991): 387-401.
31. Schotman, P.C. & H.K. van Dijk. Posterior analysis of possibly integrated time series with an application to real GNP. In D. Brillinger et al. (eds.), *New Directions in Time Series Analysis, Part II*, pp. 341-362. Berlin: Springer-Verlag, 1992.
32. Sims, C.A. Bayesian skepticism on unit root econometrics. *Journal of Economic Dynamics and Control* 12 (1988): 436-474.
33. Sims, C.A. Comment by Christopher A. Sims on "To criticize the critics," by Peter C.B. Phillips. *Journal of Applied Econometrics* 6 (1991): 423-434.
34. Sims, C.A., J.H. Stock & M.W. Watson. Inference in linear time series models with some unit roots. *Econometrica* 58 (1990): 113-144.
35. Sims, C.A. & H. Uhlig. Understanding unit rooters: A helicopter tour. *Econometrica* 59 (1991): 1591-1599.
36. Tierney, L. Markov Chains for Exploring Posterior Distributions. University of Minnesota School of Statistics technical report #560, 1991.
37. West, M. Outlier models and prior distributions in Bayesian linear regression. *Journal of the Royal Statistical Society Series B* 46 (1984): 431-439.
38. Zellner, A. Bayesian and non-Bayesian analysis of the regression model with multivariate Student-*t* error terms. *Journal of the American Statistical Association* 71 (1976): 400-405.
39. Zellner, A. & G.C. Tiao. Bayesian analysis of the regression model with autocorrelated errors. *Journal of the American Statistical Association* 59 (1964): 763-778.

This content downloaded from 128.83.155.248 on Mon, 01 May 2017 03:10:07 UTC
All use subject to http://about.jstor.org/terms