

Cowles Foundation Reprint 891

ECONOMETRIC REVIEWS, 13(3), 291-336 (1994)

**A BAYESIAN ANALYSIS OF TREND DETERMINATION
IN ECONOMIC TIME SERIES**

Eric Zivot

and

Peter C.B. Phillips

Department of Economics
Wellesley College, Wellesley, MA 02181

Cowles Foundation for Research in Economics
Yale University, New Haven, CT 06520

Key words and Phrases: Bayesian analysis, flat prior, fragile inference, hypergeometric function, ignorance prior, Laplace approximation, structural change, unit root.

ABSTRACT

In this paper we provide a comprehensive Bayesian posterior analysis of trend determination in general autoregressive models. Multiple lag autoregressive models with fitted drifts and time trends as well as models that allow for certain types of structural change in the deterministic components are considered. We utilize a modified information matrix-based prior that accommodates stochastic nonstationarity, takes into account the interactions between long-run and short-run dynamics and controls the degree of stochastic nonstationarity permitted. We derive analytic posterior densities for all of the trend determining parameters via the Laplace approximation to multivariate integrals. We also address the sampling properties of our posteriors under alternative data generating processes by simulation methods. We apply our Bayesian techniques to the Nelson-Plosser macroeconomic data and various stock price and dividend data. Contrary to DeJong and Whiteman (1989a,b,c), we do not find that the data overwhelmingly favor the existence of deterministic trends over stochastic trends. In addition, we find evidence supporting Perron's (1989) view that some of the Nelson and Plosser data are best construed as trend stationary with a change in the trend function occurring at 1929.

1. INTRODUCTION

Most macroeconomic and financial time series appear to exhibit some form of trend behavior and cannot be regarded as stationary in any conventional sense. The correct statistical interpretation and treatment of this important characteristic of economic time series, however, is not obvious and has been the focus of much research in the past few years. Moreover, assumed trend behavior can have profound implications for economic theories purporting to explain observed economic events, for econometric modeling strategies and for forecasting accuracy.

Traditionally, empirical researchers have treated observed trends as deterministic functions of time, which allowed standard statistical techniques to be used for econometric analysis. Further, in statistical theory nonstationary components have most frequently been assumed to be harmonizable (i.e. subject to frequency decomposition, at least asymptotically, rather like a stationary time series) upon suitable normalization, as in the use of Grenander-type conditions (e.g. Hannan (1970) p. 215). Such approaches to the phenomenon of nonstationarity are called trend-stationary and under these views, current stochastic shocks have only a temporary effect on the long-run movement of a series. Consequently, long-run forecasts from such a model may be expected to be fairly precise as long as the trend is consistently estimated. Led in part by the popularity and success of Box and Jenkins' ARIMA modeling methodology, many researchers have challenged this traditional view of trend behavior and argued instead that observed trends are better modeled as stochastic functions of time. According to the simplest version of this theory, stochastic shocks accumulate over time and observed series require differencing in order to ensure stationarity. In this case, current shocks have an enduring effect on the evolution of the series; hence, long-run forecasts are expected to be quite poor. This latter interpretation of trend behavior has set off an explosion of research into the econometric analysis of models with stochastic trends.

Many of the developments in this area have arisen in response to the need for reliable statistical techniques for discriminating between deterministic and stochastic trends in observed economic time series. Many classical statistical procedures have been developed explicitly for this purpose. Most notable are the unit root tests of Dickey and Fuller (1979, 1981), Sargan and Bhargava (1983), Said and Dickey (1984), Phillips (1987) and Phillips and Perron (1988). These tests have been applied to a wide variety of economic time series as in the empirical studies of Nelson and Plosser (1982), Schwert (1987) and Perron (1988), and often the null hypothesis of a unit root cannot be rejected at conventional levels of significance. These empirical results have led many to accept the notion that a wide variety of economic time series contain unit roots and, therefore, stochastic trends.

While there has been a great deal of research on the development of classical methods for determining trend behavior, much less work has been done on the use of Bayesian methods. Recently, however, Sims (1988) and Sims and Uhlig (1989) have touted the superiority of Bayesian flat prior methods over classical methods for the purpose of determining trend behavior

in observed economic series. In particular, Sims (1988) states that flat prior Bayesian procedures are simpler, more reasonable, and provide a logically sounder starting place for inference than classical hypothesis testing procedures. Sims uses Bayesian arguments to critique classical unit root testing methodology in the abstract. By contrast, DeJong and Whiteman (1989a,b,c) conduct empirical research with flat prior Bayesian techniques and directly challenge classical unit root findings in a wide array of cases, including the Nelson and Plosser macroeconomic series, various stock price and dividend data, and postwar quarterly real GNP for the U.S.A. Their main conclusions are that the classical unit root inferences for most of these series are the result of assigning zero prior probability to the alternative hypothesis that the series are trend-stationary and when this "excessively sharp" prior is relaxed, the data tend to support the trend-stationary alternative. In related work, Schotman and van Dijk (1990) use a flat prior Bayesian analysis to investigate the unit root hypothesis for real exchange rate data. They construct a Bayesian unit root test based on the posterior odds ratio of the unit root model without drift against a constant mean stationary model. Using this test, they find more evidence in favor of the stationary model than is suggested by the outcomes of classical unit root tests.

Most recently, Phillips (1991a) has offered an alternative Bayesian approach, confronting the skepticism embodied in the abstract critique by Sims on classical methods and challenging at least some of the empirical findings of DeJong and Whiteman. Phillips shows that the mechanical use of flat priors in autoregressive models that allow for unit and explosive roots ignores important generic information embodied in time series models, such as the way autoregressive coefficients affect the shape of the autocorrelogram and influence the amount of information that is carried in the data and its sample moments. In this context, flat priors are shown to downweight non-stationary models in favor of stationary ones and they do not represent ignorance (or lack of information) in any meaningful sense in time series models, in contrast to the linear regression model with fixed regressors. As an alternative to flat priors, Phillips lays the groundwork for the development of a class of information matrix-based priors that accommodate stochastic nonstationarity and utilize the prior information that is available in autoregressive models about the way the coefficients influence the sample moments and the information they carry about the parameters. Using these model-based priors, Phillips shows that Bayesian inferences have many of the same characteristics as classical inferences in these types of models, particularly that Bayesian analysis often manifests as much uncertainty about the data generating mechanism and the presence or absence of stochastic trends as classical significance testing. Phillips applies this Bayesian methodology to the Nelson and Plosser data and finds, contrary to DeJong and Whiteman (1989a), that the data do not overwhelmingly favor trend-stationary models over difference stationary models. For some of the Nelson-Plosser series, like stock prices, the posterior distributions are very different from those of DeJong and Whiteman.

Phillips' framework for analyzing trend behavior in multiple lag autoregressions with fitted time trends is not complete. The approximate information matrix priors that he employs in his empirical work ignore important interactions between long-run and short-run dynamics that

may be expected to have a substantial influence on posterior inferences. In addition, Phillips only derives posterior densities for the long-run autoregressive component. For analyzing trend behavior, the posterior densities for the coefficients on the deterministic components should also be a focus of interest, as indeed they were in the aforementioned DeJong and Whiteman studies. Moreover, Phillips restricts his attention to autoregressive models with fitted drifts and time trends. Given the recent analysis by Perron (1989), models that allow for broken trends are also clearly of interest. Lastly, Phillips only applies his methodology to the Nelson and Plosser data. A broader range of economic time series needs to be analyzed to determine the robustness of the methodology for empirical purposes.

This paper will complete this research program on the objective Bayesian analysis of trends in economic time series. We provide an exhaustive Bayesian posterior analysis of trend determination in general autoregressive models. We utilize an information matrix-based prior that explicitly takes into account the interaction between long-run and short-run dynamics. In doing so, we uncover an important relationship between the number of transient dynamics terms and the behavior of the implied prior for the long-run autoregressive parameter. Using the Laplace approximation to multivariate integrals and properties of the confluent hypergeometric function of the second kind with multiple arguments, we derive analytic posterior densities for all of the trend determining parameters. The resulting posteriors are shown to be improper due to the dominating behavior of the prior for large values of the long-run autoregressive parameter when there are many lags in the autoregression. To achieve integrable posteriors we modify the information matrix prior by attaching an exponential factor that attenuates extreme values in the unstable region of the long-run autoregressive parameter. We construct the exponential factor so that the degree of nonstationarity permitted is parameterized by a user-specified scalar quantity.

We also expand the class of models considered to include models that allow for certain kinds of structural change in the deterministic components where the point of structural change may or may not be known. Using a uniform prior for the break point to express ignorance about the point of structural change, we derive unconditional posterior densities for all of the trend determining parameters. These are shown to be mixtures of densities of the type derived for the no structural change model where the mixture variate is the marginal posterior mass function of the break point. This extension provides a Bayesian alternative to the classical methodologies used by Christiano (1988), Perron (1989, 1990), Banerjee et al. (1990) and Zivot and Andrews (1990) to test for a unit root in the presence of possible structural change at a known or unknown point in time.

Our Bayesian techniques are illustrated and evaluated through simulation experiments and applications to several sets of widely analyzed economic data series including Nelson and Plosser's (1982) macroeconomic data and various stock price and dividend data. Contrary to DeJong and Whiteman (1989a,b,c), we do not find that the data overwhelmingly favor the existence of deterministic trends over stochastic trends. More importantly, our analysis demonstrates that unit root inference are not, as DeJong and Whiteman assert, contingent on the use of zero prior probability

on trend stationary alternatives. They are so only within a flat prior framework, which we argue is inappropriate for investigating stochastic trend behavior. We also find evidence supporting Perron's view that some of the Nelson and Plosser data are best construed as trend stationary with a change in the trend function occurring at 1929.

The outline of the paper is as follows. Section 2 develops the theory for models that do not allow structural change in the deterministic components. Section 2.1 presents the models to be considered and sections 2.2-2.3 detail the derivations of the priors and posteriors. In section 2.4, we examine the behavior of our posteriors with simulated data and compare our posteriors with the posteriors derived in Phillips (1991a) and those derived from a flat prior. Posterior analysis of trend determination in certain types of structural change models is taken up in section 3. Empirical applications of the no-structural change model to the Nelson-Plosser data and the stock price and dividend data are given in sections 4.1 and 4.2. In section 4.3, we re-examine Perron's (1989) unit root results that are conditional on structural change occurring at 1929 for some of the Nelson-Plosser data from a Bayesian perspective. Section 5 contains some concluding remarks.

2. TREND DETERMINATION IN TIME SERIES MODELS

2.1. Models with fitted drifts and transient dynamics

We consider the following dynamic model for an observed time series $\{y_t\}_1^T$:

$$y_t = \mu + \beta t + \psi(L)y_t + u_t \quad (1)$$

where $\psi(L) = \sum_1^k \psi_i L^i$, $Ly_t = y_{t-1}$ and $u_t = \text{iid } N(0, \sigma^2)$. It will be convenient to employ the following reparameterization of (1) to focus attention on the dominant autoregressive component of $\psi(L)$:

$$y_t = \mu + \beta t + \rho y_{t-1} + \sum_1^{k-1} \varphi_i \Delta y_{t-i} + u_t \quad (1')$$

where $\rho = \sum_1^k \varphi_i$, $\varphi_i = -\sum_1^{k+1} \psi_j$ and $\Delta y_{t-i} = y_{t-i} - y_{t-i-1}$. We may facilitate the transformation from (1) to (1') via the nonsingular matrix H . To see this, let $\psi' = (\psi_1, \dots, \psi_k)$, $y'_{t-1} = (y_{t-1}, \dots, y_{t-k})$, $\varphi' = (\varphi_1, \dots, \varphi_{k-1})$, $\gamma' = (\rho, \varphi')$ and $x'_{t-1} = (y_{t-1}, \Delta y_{t-1}, \dots, \Delta y_{t-k+1})$. Then

$$y_t = \mu + \beta t + (H'\psi)'H^{-1}y_{t-1} + u_t = \mu + \beta t + \gamma'x_{t-1} + u_t$$

so that $\gamma = H'\psi$, $x_{t-1} = H^{-1}y_{t-1}$ and

$$H = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & -1 & 0 & \cdots & 0 \\ 1 & -1 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & -1 & -1 & \cdots & -1 \end{bmatrix}, \quad H^{-1} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -1 \end{bmatrix}$$

We are interested in two types of trend behavior that result from placing certain restrictions on the parameters of (1) (hence, (1')). If the equation $1 - \psi(L) = 0$ has all of its roots

outside the unit circle and $\beta \neq 0$ then the series $\{y_t\}$ exhibits stationary fluctuations about a deterministic time trend and, accordingly, we call the series $\{y_t\}$ trend stationary (TS). In this case, $\rho < 1$ in (1'). If $1 - \psi(L) = 0$ has a single root of unity and $\beta = 0$ then the series of differenced observations $\{\Delta y_t\}$ exhibits stationary fluctuations and we call $\{y_t\}$ difference stationary (DS). In this case, $\rho = 1$ and the φ 's are the parameters of the autoregressive representation for $\{\Delta y_t\}$.

The parameterization (1) was used by DeJong and Whiteman (1989a,b) in their Bayesian analysis of trend determination. Their attention focused on the time trend parameter β and the modulus of the largest root of the equation $1 - \psi(L) = 0$. This setup is awkward for analysis because the roots of $1 - \psi(L) = 0$ are nonlinear functions of the ψ 's, and, as argued in Phillips (1991b), inappropriate because no single root like Λ determines long-run behavior by itself in general. The alternative parameterization (1') was used by Phillips (1991a) in his Bayesian analysis and also corresponds to the regression formulation used in augmented Dickey-Fuller (ADF) unit root tests. The ADF setup is more convenient to use and it isolates the critical parameter ρ that determines long-run behavior of y_t . Of course, several other parameterizations are possible and these may have certain other advantages (see Schotman and van Dijk (1990) and Zivot (1992)).

Let $\theta' = (\mu, \beta, \rho, \varphi', \sigma) \in \mathbb{R}^{k+2} \times \mathbb{R}_+$, $y' = (y_1, \dots, y_T)$ denote the $T \times 1$ vector of sample observations, $y_0' = (y_0, \dots, y_{-k+1})$ the $k \times 1$ vector of fixed initial values, and $f(y|\theta, y_0)$ the joint probability density function (pdf) of the sample given the parameter vector θ and the initial values y_0 . Treated as a function of θ , f is the likelihood function and is denoted by $L(\theta|y, y_0)$. Prior views concerning θ are represented by the pdf $\pi(\theta)$ (which may be proper or improper). Given the sample observations y and the initial values y_0 , these prior views about θ are updated via Bayes' rule to give the posterior pdf $p(\theta|y, y_0)$, where $p(\theta|y, y_0) \propto \pi(\theta)L(\theta|y, y_0)$.

The focus of our analysis is ρ and β since their values determine the trend behavior of $\{y_t\}$. To derive marginal posterior pdf's for ρ and β , we must extract the following integrals:

$$\begin{aligned} p(\rho|y, y_0) &\propto \int \dots \int p(\theta|y, y_0) d\mu d\beta d\varphi d\sigma = \int \dots \int L(\theta|y, y_0) \pi(\theta) d\mu d\beta d\varphi d\sigma \\ p(\beta|y, y_0) &\propto \int \dots \int p(\theta|y, y_0) d\mu d\rho d\varphi d\sigma = \int \dots \int L(\theta|y, y_0) \pi(\theta) d\mu d\rho d\varphi d\sigma \end{aligned}$$

Some of the priors that we employ do not have the convenience of flat or conjugate priors and, as a result, the above integrals cannot be evaluated exactly for these priors. However, very good analytic approximations may be obtained using Laplace's method for approximating multivariate integrals. The method is applicable to integrals of the form

$$I(\lambda) = \int_D \exp\{\lambda\phi(\theta)\} g(\theta) d\theta, \quad \theta' = (\theta_1, \dots, \theta_p)$$

where D is a simply connected domain in p -dimensional space. If $\phi(\theta)$ and $g(\theta)$ are continuously differentiable to second order, if ϕ obtains an interior absolute maximum at θ^* in D , and if the Hessian $\partial^2\phi(\theta)/\partial\theta\partial\theta'$ is negative definite, then it can be shown that

$$I(\lambda) \sim (2\pi/\lambda)^{p/2} \exp\{\lambda\phi(\theta^*)\} g(\theta^*) |\partial^2\phi(\theta^*)/\partial\theta\partial\theta'|^{-1/2}$$

in the sense that $A \sim B$ if $A/B \rightarrow 1$ as $\lambda \rightarrow \infty$. The error on the approximation is of $O[\exp\{\lambda\phi(\theta^*)\} \lambda^{-(p+2)/2}]$ as $\lambda \rightarrow \infty$.

For many problems, the Laplace approximation is a convenient and efficient alternative to simulation-based numerical methods, such as Monte Carlo integration (see Kloek and van Dijk (1978) and Geweke (1989)), for evaluating multivariate integrals. It is particularly useful for models with Gaussian likelihoods. The technique is quite general because a numerical optimization procedure may be used to compute the maximizing value of ϕ and then this value may be utilized in the approximation formula. Phillips (1983) originally suggested this technique as a means of obtaining marginal posterior densities in models with many parameters. Tierney and Kadane (1986) employ it for various Bayesian marginalization problems. In the present context, the analytic approximations to the marginal posteriors provide considerable information concerning the nature of the posteriors that would be difficult to ascertain if simulation-based numerical methods were used to evaluate the integrals. In addition, the marginal posteriors are very cheap to compute and this in turn facilitates extensive posterior analysis under an entire class of priors thereby encouraging an investigator to examine whether inferences are robust or fragile within that class. As argued by Leamer (1978, 1984), testing the sensitivity of posterior inferences to variations in the prior is an essential part of careful Bayesian analysis.

2.2 Prior and posterior analysis

Bayesian analysis begins with the specification of prior beliefs concerning the parameters of interest. In many situations, researchers do not have strong beliefs or much prior knowledge about the parameters of the given model to be analyzed and, consequently, wish to use a prior that reflects this lack of knowledge or ignorance. There is no universally accepted way of expressing ignorance about the parameters of a given model and a number of suggestions for generating priors to represent "knowing little" or ignorance have been put forth. Excellent summaries of the various methods for determining noninformative priors can be found in Zellner (1971) and Berger (1985). Two widely used methods are particularly relevant to our situation.

The first method to express ignorance about the parameters of interest is to use a uniform or flat prior for the parameters or certain transformations of the parameters. In model (1) (or (1')), if we use flat priors for the regression coefficients and a flat prior for the natural logarithm of the scale parameter then $\pi(\theta) \propto \sigma^{-1}$. The use of flat priors is very convenient because they often lead to posterior inferences that agree with inferences drawn from classical procedures. In econometric analysis, Zellner (1971) and Judge et al. (1985) advocate using flat priors for the analysis of the linear regression model with fixed regressors and its various extensions, and for the analysis of certain types of time series models. Flat and truncated flat prior have been utilized in unit root models by Sims (1988), DeJong and Whiteman (1989a,b,c), Schotman and van Dijk (1990) and Sims and Uhlig (1989).

Although giving each point in the parameter space equal treatment is an intuitively appealing way of expressing ignorance, this view of ignorance has been harshly criticized because it lacks certain invariance properties (see Berger (1985) ch. 3 and the references therein). For example, suppose we express our ignorance about a parameter $\theta \in (-\infty, \infty)$ by adopting the

uniform prior $\pi(\theta) \propto c$, where c is some arbitrary constant. Now instead of considering θ , suppose the problem is re-parameterized in terms of $\eta = \exp(\theta)$. This is a 1-1 transformation and ideally should not affect inference. By the change of variables formula, the corresponding prior for η is given by $\pi(\eta) = \eta^{-1} \pi(\ln(\eta))$. So, if the noninformative prior for θ is chosen to be constant, we should choose the noninformative prior for η to be proportional to η^{-1} to maintain consistency in inference. Thus we cannot maintain consistency and choose both the noninformative prior for θ and that for η to be constant.

In addition, Phillips (1991a) argues that flat priors in autoregressive models cannot represent ignorance in an objective sense because they ignore generically available information such as the way the autoregressive coefficients affect the correlation structure of the data and the expected amount of information carried by the data. To illustrate this point, consider the following example taken from Phillips (1991a). Let

$$y_t = \rho y_{t-1} + u_t \quad (t = 1, \dots, T)$$

with $u_t \sim \text{iid } N(0, \sigma^2)$ and consider two intervals for ρ : $[0.5, 0.6]$ and $[0.95, 1.05]$. A flat prior over ρ indicates that ρ values in the two intervals are equally likely. The flat prior, however, ignores information from the AR(1) model that we always have about the way ρ values in these two intervals affect sample behavior. Sample behavior is expected to be very different for ρ values in these two intervals and this represents prior knowledge based on the postulated AR(1) model. Phillips maintains that an objective ignorance prior for ρ in a model that allows for stochastic nonstationarity must incorporate such model-based information.

These criticisms of the flat prior have led to the search for noninformative priors which are appropriately invariant under transformations and which incorporate model-based information. The most widely used method for generating such priors is Jeffreys' rule

$$\pi(\theta) \propto |I_{\theta\theta}|^{1/2} = J(\theta) \quad (2)$$

where $I_{\theta\theta} = -E[\partial^2 \ell n f(y|\theta, \iota_0)/\partial\theta\partial\theta']$ denotes the Fisher information matrix. In addition to flat priors, such information matrix-based priors are an essential feature of our analysis.

We proceed to derive the Jeffreys prior for models (1) and (1'). The log-likelihood function of θ based on the sample y with fixed initial values ι_0 is

$$l(\theta|y, \iota_0) \propto -T \ln(\sigma) - \frac{1}{2} \sigma^{-2} \Sigma_1^T (y_t - \mu - \beta t - \gamma' x_{t-1})^2 \quad (3)$$

and the information matrix is given by

$$I_{\theta\theta} = \begin{bmatrix} \sigma^{-2}T & \sigma^{-2}T(T+1)/2 & \sigma^{-2}\sum_1^T E[x_{t-1}] & 0 \\ \sigma^{-2}T(T+1)/2 & \sigma^{-2}T(T+1)(2T+1)/6 & \sigma^{-2}\sum_1^T tE[x_{t-1}] & 0 \\ \sigma^{-2}\sum_1^T E[x_{t-1}] & \sigma^{-2}\sum_1^T tE[x_{t-1}] & \sigma^{-2}\sum_1^T E[x_{t-1}x_{t-1}'] & 0 \\ 0 & 0 & 0 & \sigma^{-2}T \end{bmatrix} \quad (4)$$

Phillips uses an approximation to the determinant of (4) that ignores the time series effects of the parameters φ and is given by

$$\pi(\theta) \propto \sigma^{-k-2} \{ \alpha_0(\rho) + \alpha_1(\rho, \mu, \beta) / \sigma^2 \}^{1/2}, \quad (5)$$

where

$$\alpha_0(\rho) = T(1-\rho^2) - (1-\rho^2)^{-2}(1-\rho^{2T}), \quad (6)$$

$$\alpha_1(\rho, \mu, \beta) = \Sigma_0^{-1} [\mu(1-\rho)^{-1}(1-\rho^T) + \beta \{ (1-\rho)^{-1}i - \rho(1-\rho)^{-2}(1-\rho^T)i \}]^2. \quad (7)$$

The prior (5) may be interpreted as an approximation to the square root of the product of the diagonal elements of (4) when $\varphi = 0$. It may therefore be regarded as an approximate Jeffreys prior for a general nonstationary model. However, when $\varphi \neq 0$ the simple approximation (5) is expected to "bias" inferences since it is based on a model in which $\varphi = 0$. Clearly, a model-based reference prior that allows $\varphi \neq 0$ is desirable since most empirical models can be expected to have non-trivial transient dynamics with $\varphi \neq 0$. Adding this extra level of generality is important because as more lags are added to the fitted model they explain a larger portion of the variability in the data and the role of the long-run autoregressive parameter is diminished. Various methods of explicitly accounting for the effects of the transient dynamics on the prior are possible. One simple method is as follows.

The dominating term of (4) is $\Sigma_1^T E[x_{t-1} x'_{t-1}]$, and if we ignore the off block diagonal elements of (4), the Jeffreys prior would be

$$\pi(\theta) \propto \sigma^{-3} |\sigma^{-2} \Sigma_1^T E[x_{t-1} x'_{t-1}]|^{1/2}. \quad (8)$$

An explicit representation for (8) can be found by setting up (1) in companion form as

$$Y_t = RY_{t-1} + \mu + \beta t + U_t \quad (9)$$

with

$$Y_t = \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-k+1} \end{bmatrix}, \quad R = \begin{bmatrix} \psi_1 & \psi_2 & \cdots & \psi_k \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad U_t = \begin{bmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \quad (10)$$

We use (9) to evaluate the required information matrix as follows. Back substitution leads to

$$Y_t = \Sigma_0^{-1} R^s [U_{t-s} + \mu + \beta(t-s)] + R^s Y_0$$

and then

$$E[Y_t Y'_t] = \Sigma_0^{-1} R^s \Sigma R^{s'} + \{ \Sigma_0^{-1} R^s [\mu + \beta(t-s)] + R^s Y_0 \} \{ \Sigma_0^{-1} R^s [\mu + \beta(t-s)] + R^s Y_0 \}'$$

where $\Sigma = E[U_t U'_t] = \sigma^2 E_{11}$, where E_{11} has unity in the (1,1) position and zeros elsewhere. We may express the above matrix representation of $E[Y_t Y'_t]$ as

$$\sigma^2 A_{\alpha}(\psi) + A_{1t}(\mu, \beta, \psi) \quad (11)$$

where

$$A_{\alpha}(\psi) = \Sigma_0^{-1} R^s E_{11} R^{s'},$$

$$A_{1t}(\mu, \beta, \psi) = \{ \Sigma_0^{-1} R^s [\mu + \beta(t-s)] + R^s Y_0 \} \{ \Sigma_0^{-1} R^s [\mu + \beta(t-s)] + R^s Y_0 \}'.$$

Since $\psi = H'^{-1}\gamma$ we may recast (11) using the alternative parameterization, viz.

$$\begin{aligned} E[Y_t Y_t'] &= \sigma^2 A_{0t}(H'^{-1}\gamma) + A_{1t}(\mu, \beta, H'^{-1}\gamma) \\ &= \sigma^2 A_{0t}^*(\rho, \varphi) + A_{1t}^*(\mu, \beta, \rho, \varphi). \end{aligned} \quad (12)$$

We now deduce the required "information matrix" component for model (1) as

$$\sigma^{-2} \Sigma_1^T E[Y_{t-1} Y_{t-1}'] = \sigma^{-2} \Sigma_0^T E[Y_t Y_t'] = \Sigma_0^T [A_{0t}(\psi) + \sigma^{-2} A_{1t}(\mu, \beta, \psi)]$$

and the corresponding matrix for model (1') is

$$\begin{aligned} \sigma^{-2} \Sigma_1^T E[x_{t-1} x_{t-1}'] &= \sigma^{-2} \Sigma_0^T H^{-1} E[Y_t Y_t'] H'^{-1} \\ &= H^{-1} \Sigma_0^T [A_{0t}^*(\rho, \varphi) + \sigma^{-2} A_{1t}^*(\mu, \beta, \rho, \varphi)] H'^{-1} \\ &= A_0^*(\rho, \varphi) + \sigma^{-2} A_1^*(\mu, \beta, \rho, \varphi) \end{aligned} \quad (13)$$

where

$$\begin{aligned} A_0^*(\rho, \varphi) &= H^{-1} \Sigma_0^T A_{0t}^*(\rho, \varphi) H'^{-1}, \\ A_1^*(\mu, \beta, \rho, \varphi) &= H^{-1} \Sigma_0^T A_{1t}^*(\mu, \beta, \rho, \varphi) H'^{-1}. \end{aligned}$$

Using (13) we have an explicit representation for the approximate Jeffreys prior given by (8):

$$\begin{aligned} \pi(\theta) &\propto \sigma^{-3} |A_0^*(\rho, \varphi) + \sigma^{-2} A_1^*(\mu, \beta, \rho, \varphi)|^{1/2} \\ &= \sigma^{-3} |A_0^*(\rho, \varphi)|^{1/2} \prod_{j=1}^k (1 + \sigma^{-2} \lambda_j^*(\mu, \beta, \rho, \varphi))^{1/2}, \end{aligned} \quad (14)$$

where $\lambda_j^*(\mu, \beta, \rho, \varphi)$, $j = 1, \dots, k$ are the eigenvalues of $A_0^*(\rho, \varphi)^{-1} A_1^*(\mu, \beta, \rho, \varphi)$.

The general behavior of the prior (14) as a function of ρ is very similar to the simple approximation (5) used earlier by Phillips (see Zivot and Phillips (1991) for a full graphical examination of the prior (14)). Unlike (5), however, (14) is affected by the nature and number of the transient dynamics terms, φ . For example, when $k = 2$ large positive values of φ result in proportionally more weight being given to stationary values of ρ and large negative values of φ result in proportionally more weight being given to nonstationary values of ρ . In this case, $\rho = \psi_1 + \psi_2$ and $\varphi = -\psi_2$ so, given ψ_1 , large positive values of φ reduce ρ and large negative values of φ increase ρ . In addition, as the lag length, k , increases the prior places proportionally more weight on nonstationary value of ρ . In this respect, the information matrix prior works to counteract the increasing downward OLS bias in $\hat{\rho}$ as k increases that occurs when the true value of ρ is unity.

Two additional comments on the prior (14) are in order:

(i) A computational problem occurs in the evaluation of $|A_0^*|$ for values of $\rho > 1$. In particular, numerical estimates of $|A_0^*|$ for ρ values just above unity are unstable even in the $k = 2$ or $k = 3$ case for $T \geq 50$. Thus, in order to use the more general Jeffreys prior in practical applications, a stable numerical approximation to $|A_0^*|$ is required. One such possibility that works well is given by the product of the diagonal entries of A_0^* .

(ii) Since the approximate Jeffreys priors (5) and (14) are defined for all $\rho \in \mathbb{R}$ and place rapidly increasing weight on values of ρ greater than unity, they may not serve as appropriate reference priors for all researchers. In practical applications researchers are often only interested in a restricted range for ρ and with most economic time series researchers would give

very little, if any, subjective prior weight to values of ρ much greater than unity. In particular, if interest focusses on evaluating the unit root hypothesis given the data, many researchers only want to consider values of ρ in the range $(\rho_0, 1+\varepsilon)$ for some $0 < \rho_0 < 1$ and ε small. Such researchers may be skeptical of results that are generated from priors that place heavy weight on a priori extreme and unlikely values. To accommodate such subjective prior views on ρ we can easily modify the Jeffreys prior by limiting the effective range of ρ and, hence, the degree of non-stationarity allowed a priori.

These issues will be addressed in detail below. But first it is informative to derive the marginal posterior densities for ρ and β as well as the joint posterior density for $\alpha' = (\rho, \beta)$. let $\delta' = (\mu, \beta, \varphi')$, $\pi' = (\mu, \rho, \varphi')$, $\eta' = (\mu, \varphi')$ and use y_{-1} , τ , U , V , Z and W to represent the observation matrices of (y_{t-1}) , (t) , (y_{t-1}, t) , $(1, t, \Delta y_{t-1}, \dots, \Delta y_{t-k+1})$, $(y_{t-1}, 1, \Delta y_{t-1}, \dots, \Delta y_{t-k+1})$ and $(1, \Delta y_{t-1}, \dots, \Delta y_{t-k+1})$, respectively. In addition, for any matrix R of full rank and conformable matrix y define $Q_R = I - R(R'R)^{-1}R'$ and $m_R(y) = y'Q_R y$. To facilitate the determination of the posterior densities for ρ , β and α via the Laplace approximation, we employ the following decompositions of the sum of squares $\Sigma_1^T (y_t - \mu - \beta t - \gamma' x_{t-1})^2$:

$$m(\hat{0}) + (\rho - \hat{\rho})^2 m_V(y_{-1}) + (\delta - \hat{\delta}(\rho))' V' V (\delta - \hat{\delta}(\rho)), \quad (15)$$

$$m(\hat{0}) + (\beta - \hat{\beta})^2 m_Z(\tau) + (\pi - \hat{\pi}(\beta))' Z' Z (\pi - \hat{\pi}(\beta)), \quad (16)$$

$$m(\hat{0}) + (\alpha - \hat{\alpha})' m_W(U)(\alpha - \hat{\alpha}) + (\eta - \hat{\eta}(\alpha))' W' W (\eta - \hat{\eta}(\alpha)), \quad (17)$$

where $m(\hat{0}) = \Sigma_1^T \hat{u}_t^2$, $\hat{u}_t = y_t - \hat{\mu} - \hat{\beta}t - \hat{\gamma}' x_{t-1}$ are the OLS residuals and

$$\hat{\delta}(\rho) = \hat{\delta} + (V'V)^{-1}V'y_{-1}(\hat{\rho} - \rho),$$

$$\hat{\pi}(\beta) = \hat{\pi} + (Z'Z)^{-1}Z'\tau(\hat{\beta} - \beta),$$

$$\hat{\eta}(\alpha) = \hat{\eta} + (W'W)^{-1}W'U(\hat{\alpha} - \alpha).$$

The following proposition gives the approximate posterior densities for ρ , β and α under the prior (14).

PROPOSITION 1

- (a) $p(\rho | y, u_0) \propto |A_0^*(\rho, \hat{\varphi}(\rho))|^{1/2} [m(\hat{0}) + (\rho - \hat{\rho})^2 m_V(y_{-1})]^{-(T-k+1)/2}$
- (b) $p(\beta | y, u_0) \propto |A_0^*(\hat{\rho}(\beta), \hat{\varphi}(\beta))|^{1/2} [m(\hat{0}) + (\beta - \hat{\beta})^2 m_Z(\tau)]^{-(T-k+1)/2}$
- (c) $p(\alpha | y, u_0) \propto |A_0^*(\hat{\rho}(\alpha), \hat{\varphi}(\alpha))|^{1/2} [m(\hat{0}) + (\alpha - \hat{\alpha})' m_W(U)(\alpha - \hat{\alpha})]^{-(T-k+2)/2}$

Remarks (i) The derivation of the approximate posterior densities for ρ , β and α follows the same lines as the derivation of the posterior density for ρ outlined in Phillips (1991a). Note that this derivation, which is given in the appendix, makes use of two approximations. First, the reduction of the multivariate integrals to a single integral in σ is performed by means of the Laplace approximation. Next, the integral in σ involves a confluent hypergeometric function of the second kind with multiple arguments (see Phillips (1988) for details about this function). The second level of approximation applies when this function is itself approximated (for large arguments) to facilitate computation. The mathematical details of these derivations are presented in

the Appendix. For our purposes here it is most important to note that the integration over σ eliminates the effect of the second matrix component $A_1^*(\mu, \beta, \rho, \varphi)$ in (14) on the posteriors, as given in (a), (b) and (c) above. The matrix component from the prior that is retained in the posteriors, $A_0^*(\rho, \varphi)$, is a function of ρ and φ only. The prior effects of the deterministic components (μ, β) on the behavior of the posteriors are asymptotically dominated by the prior effects of (ρ, φ) .

(ii) As a result of the Laplace approximation, the posterior densities for ρ , β and α are of the same general form and, therefore, we expect these densities to have similar shape characteristics. In the density formula for ρ , the matrix component \tilde{A}_0^* is a function of ρ since the Laplace approximation produces the term $\tilde{\varphi}(\rho)$. Similarly, in the posterior for β , \tilde{A}_0^* is a function of β and in the density for α it is a function of α .

(iii) The posteriors for ρ , β and α given in Proposition 1 are improper densities over \mathbb{R} ; i.e. they are not integrable over the whole real line. To see this, note that the diagonal elements of \tilde{A}_0^* are of the order $\text{diag}(\rho^{2T-4}, \rho^{2T-6}, \dots, \rho^{2T-2k-2})$ as $|\rho| \rightarrow \infty$ and so

$$|\tilde{A}_0^*|^{1/2} \sim (\rho^{2T-4} \dots \rho^{2T-2k-2})^{1/2} = \rho^{Tk - (k+1)(k+2)/2 + 1}, \quad (18)$$

which clearly dominates the tail behavior of the posteriors. Thus, the determinantal prior leads to improper posteriors for ρ , β and α . There is, in effect, too much information in the Jeffreys prior (14) relative to the actual information contained in the data.

(iv) Phillips (1991a) derived the approximate marginal posterior for ρ based on (5) and it is given by

$$p(\rho|y, y_0) \propto \alpha_0(\rho)^{1/2} [m(0) + (\rho - \hat{\rho})^2 m_V(y_{-1})]^{-T/2}. \quad (19)$$

It is a straightforward exercise to derive the corresponding posteriors for β and α . These are

$$p(\beta|y, y_0) \propto \alpha_0(\beta)^{1/2} [m(0) + (\beta - \hat{\beta})^2 m_Z(\tau)]^{-T/2}, \quad (20)$$

$$p(\alpha|y, y_0) \propto \alpha_0(\rho)^{1/2} [m(0) + (\alpha - \hat{\alpha})' m_W(U)(\alpha - \hat{\alpha})]^{-(T+1)/2}, \quad (21)$$

which are of a similar form to the posteriors given in the proposition. However, since $\alpha_0(\rho)^{1/2}$ is of the order $O(\rho^{T-2})$ for $\rho > 1$ the marginal posteriors have Pareto tails of order $O(|\rho|^{-2})$ as $\rho \rightarrow \infty$ and thus are proper densities upon standardization. The main difference between these posteriors and those based on (14) is that the posteriors based on (14) are influenced by the prior effects of φ and the lag length of the autoregression whereas the posteriors based on (5) are not.

(v) It is interesting to note that the posteriors derived from the prior (14) are almost identical to the posteriors which are derived directly from the prior

$$\pi(\theta) = \pi(\mu, \beta, \sigma)\pi(\rho, \varphi) \propto \sigma^{-1} |A_0^*(\rho, \varphi)|^{1/2}. \quad (22)$$

The only differences between these posteriors and the posteriors given in the proposition are the degrees of freedom in the exponents of the Student-t kernels, the effect of which is small due to the dominating behavior of $|\tilde{A}_0^*|^{1/2}$. The prior (22) results from assuming that μ , β and $\ell n(\sigma)$ are a priori independent of ρ and φ and are independently uniformly distributed over the real line and that the prior for (ρ, φ) is determined by Jeffreys rule applied to the model

$$y_t = \psi(L)y_t + u_t \quad (t = 1, \dots, T)$$

with $u_t \equiv \text{iid } N(0, \sigma^2)$. Given that the posterior densities for ρ and β using the priors (14) and (22) are nearly the same, it is much easier both conceptually and computationally to use the approximate invariant prior (22) than the approximation (14). The joint prior (22) is built up from individual priors in a straightforward way and the prior relationships between the parameters are clearly spelled out. In addition, the independence assumption between (μ, β, σ) and (ρ, φ) means that we get a quick route to the posteriors via the Laplace approximation.

(vi) A similar result holds for the posteriors based on (5). The approximate densities given in (19)–(21) also result from the prior

$$\pi(\theta) \propto \sigma^{-k-2} \alpha_0(\rho)^{1/2} \quad (23)$$

which employs, conditional on σ , flat priors for μ, β and φ and a Jeffreys prior for ρ based on the model

$$y_t = \rho y_{t-1} + u_t \quad (t = 1, \dots, T)$$

with $u_t \equiv \text{iid } N(0, \sigma^2)$. The prior (23) is a more convenient form than the prior (5).

The fact that the approximate posteriors for ρ and β are not integrable and the numerical evaluations of $|A_0^*|$ are unstable greatly hinders the applicability of the information matrix-based prior (22) for general autoregressive models. In the next section, therefore, we shall introduce some modifications to the prior (22) that effectively restricts the relevant range of ρ (hence, the degree of nonstationarity) and that result in numerically stable, fully integrable posteriors.

2.3. Modified priors and posterior analysis

To avoid the numerical instability encountered in evaluating $|A_0^*|$ we may approximate this determinant by taking the product of the diagonal elements of A_0^* . Using this approximation gives the prior

$$\pi(\rho, \varphi) \propto \sigma^{-1} (\prod_{i=1}^k a_{0ii}^*)^{1/2}, \quad (24)$$

where a_{0ii}^* denotes the i 'th diagonal element of A_0^* . This product approximation of the determinant of the information matrix was successfully used in Phillips (1991a) for the case where the transient dynamics were ignored. The product approximation (24) as a function of ρ is graphed in Figure 2(a) for the case $t_0 = 0, \sigma = 1, T = 100, k = 2$ for various values of φ . Comparing this figure with Figure 1(a) shows that (24) behaves much like (22) and suggests that little information will be lost if the product approximation is used. Note that the diagonal elements a_{0ii}^* that enter (24) are still, of course, functionally dependent on the transient dynamics parameters φ .

We now consider the objection to the Jeffreys prior that it places increasing weight on extreme and unlikely values of ρ as far as economic time series are concerned. By dealing with this objection we also produce fully integrable posteriors. The obvious way to restrict the relevant range of ρ and at the same time make the posteriors integrable is to eliminate values of ρ in the extreme unstable regions by truncation of the prior. When we implement truncation methods, however, we still find that as k increases the posterior probability of a unit root or greater generally increases rapidly because the form of (18) still causes the prior to dominate for values

of ρ in the region just beyond unity. Simulation experiments show that the posterior density for ρ is bimodal and that the second mode (for $\rho > 1$) is the dominant one in typical cases such as that in which the true generating process is a unit root with drift or an autoregression with a large stable root. This outcome also occurs empirically with many of the series in the Nelson and Plosser data set. Thus, arbitrary truncation does not appear to be a sensible and practical solution to the integrability problem.

An alternative and rather appealing way of effectively restricting the range of ρ and dealing with the improper posteriors is to introduce an exponential factor into the prior that attenuates values of the prior for $\rho > 1$. The particular modification we have in mind also allows a researcher to control the degree of nonstationarity desired in the prior. We shall consider product priors for (ρ, φ) that are based on (24) but are modified with exponential factors. They take the general form

$$\pi(\rho, \varphi) \propto (\Pi_{j=1}^k a_{0j})^{1/2} \exp\{-\rho^{2c_k}\}, \quad (25)$$

where c_k is some function (to be specified) of the number of lags in the autoregression and the sample size. The selection of c_k determines the shape of the prior in the unstable region $\rho > 1$ and hence affects the degree of nonstationarity allowed. It can have a large influence on posterior inferences. The choice of c_k could be left entirely arbitrary with its value depending on the application and on the prior views of the researcher. However, we propose a method for determining c_k that generates a family of priors parameterized by a scalar quantity (ε). This leads to an ε -family of posteriors for any given application. Thus empirical researchers have the option of reporting a range of posterior inferences so that issues of fragility can be more easily addressed in practice.

Our suggestion for determining c_k is as follows. Recall, from (18) that

$$(\Pi_{j=1}^k a_{0j})^{1/2} \sim \rho^{Tk - (k+1)(k+2)/2 + 1} = \rho^{Tk - d_k},$$

where $d_k = (k+1)(k+2)/2 + 1$. Hence the prior (25) behaves like

$$\rho^{Tk - d_k} \exp\{-\rho^{2c_k}\} \quad (26)$$

for $|\rho| > 1$. We may induce a convenient family of priors by specifying a modal value for (26). This will occur for a value of ρ , say

$$\rho = 1 + \varepsilon$$

for which (26) attains its maximum and beyond which the prior falls away rapidly. Thus when we set $\rho = 1 + \varepsilon$ as the modal value we anticipate low prior probability from (26) for values of ρ much greater than $1 + \varepsilon$. This gives an investigator a convenient means of attaching low prior weight to extreme unstable values of ρ .

To determine the maximizing value of c_k when $\rho = 1 + \varepsilon$ is the modal value, we optimize (26) with respect to ρ , set $\rho = 1 + \varepsilon$ in the first order conditions, and then solve for c_k . Straight-forward calculations reveal that an approximate solution to the optimization problem is

$$c_k = -(4\varepsilon)^{-1} + (4\varepsilon)^{-1}(1 + 4\varepsilon(Tk - d_k))^{1/2} = c_k(\varepsilon). \quad (27)$$

The modified Jeffreys prior for (ρ, φ) is then

$$\pi(\rho, \varphi | \varepsilon) \propto (\Pi_{10}^k \mathbf{a}_{00}^*)^{1/2} \exp\{-\rho^{2c_k(\varepsilon)}\}. \quad (28)$$

Notice that this prior is conditioned on the value of ε . We now have a family of modified ignorance priors for θ , indexed by ε , which are built up from individual priors for (μ, β, σ) and (ρ, φ) :

$$\pi(\theta) = \pi(\mu, \beta, \sigma)\pi(\rho, \varphi | \varepsilon) \propto \sigma^{-1}(\Pi_{10}^k \mathbf{a}_{00}^*)^{1/2} \exp\{-\rho^{2c_k(\varepsilon)}\}. \quad (29)$$

Conditional on ε , the above prior results from assuming independent flat priors for μ, β and $\ln(\sigma)$ and the modified Jeffreys prior for (ρ, φ) . A family of priors, for example, on a grid such as $\{\varepsilon = 0.001, 0.0025, 0.075, 0.01\}$ are easily generated giving a maximum for the priors in the range (1.001, 1.01). The modified priors inherit the general characteristics of the previously discussed priors for ρ values less than $1 + \varepsilon$. Beyond $\rho = 1 + \varepsilon$, however, the exponential factor draws the density rapidly towards zero. Illustrations of these ε -priors are given in Zivot and Phillips (1991) and Phillips (199b).

The following theorem gives the approximate posterior densities of ρ, β and α under the modified Jeffreys prior (29). The proof follows from an application of Laplace's method for approximating multivariate integrals along the line we have earlier discussed and is therefore omitted.

THEOREM 1

- (a) $p(\rho | \varepsilon, y, y_0) \propto (\Pi_{10}^k \mathbf{a}_{00}^*(\rho, \tilde{\varphi}(\rho)))^{1/2} \exp\{-\rho^{2c_k(\varepsilon)}\} [m(\hat{u}) + (\rho - \hat{\rho})^2 m_V(y_{-1})]^{-(T-k-1)/2}$
- (b) $p(\beta | \varepsilon, y, y_0) \propto (\Pi_{10}^k \mathbf{a}_{00}^*(\tilde{\rho}(\beta), \tilde{\varphi}(\beta)))^{1/2} \exp\{-\tilde{\rho}(\beta)^{2c_k(\varepsilon)}\} [m(\hat{u}) + (\beta - \hat{\beta})^2 m_Z(\tau)]^{-(T-k-1)/2}$
- (c) $p(\alpha | \varepsilon, y, y_0) \propto (\Pi_{10}^k \mathbf{a}_{00}^*(\tilde{\rho}(\alpha), \tilde{\varphi}(\alpha)))^{1/2} \exp\{-\tilde{\rho}(\alpha)^{2c_k(\varepsilon)}\} \\ \times [m(\hat{u}) + (\alpha - \hat{\alpha})' m_W(U)(\alpha - \hat{\alpha})]^{-(T-k-2)/2}$

Remarks (i) The above posteriors are fully integrable due to the addition of the exponential weighting factors. They are conditional on the value of ε , which must be specified by the user. The analytic nature of the posteriors, however, makes it easy to generate a family of posteriors for a given range of ε -values. In this manner, the sensitivity of posterior inferences to the amount of nonstationarity allowed a priori can easily be addressed. It is important to note that other types of damping factors could be used to restrict the range of ρ . The flexibility of the Laplace approximation permits easy determination of the posteriors.

(ii) The densities of ρ, β and α are asymmetric. For a given ε value, the modes of the ρ -posteriors are generally skewed towards unity and the modes of the β -posteriors are skewed towards the origin. This asymmetry works to counteract the downward OLS bias in $\hat{\rho}$ and upward bias in $\hat{\beta}$ that occurs when the true values of ρ and β are one and zero, respectively. As the lag length, k , of the fitted model increases the prior has a larger influence on the shapes of the posteriors due to the dominating behavior of (18) for ρ values less than $1 + \varepsilon$. In addition, for large ε the posteriors for ρ and β are often bimodal, like the densities derived from (5), and

display considerable uncertainty about the true values of ρ and β . For more information on the typical shapes of these densities, the reader is referred to Zivot and Phillips (1991).

(iii) Phillips (1991a) investigated the sampling behavior of the ρ -posteriors based on a flat prior and the approximate Jeffreys prior (5) when the true data generating process (DGP) contains a unit root with drift. He showed that, on average, posterior inferences drawn from a flat prior are severely biased away from the unit root model and that inferences from (5) are much more consonant with the true DGP. We have performed similar sampling experiments to illustrate the expected behavior of the posterior densities of ρ computed from a flat prior (F-posterior), Phillips' approximate Jeffreys prior (J-posterior) and the modified Jeffreys prior (29) (MJ-posterior). We briefly summarize here the main results from these experiments. A full account is given in Zivot and Phillips (1991).

First we investigate the sampling behavior of the posteriors based on the simple random walk with drift model with iid $N(0, \sigma^2)$ errors. For a fixed value of k , the expected posterior probabilities of the event $\{\rho \geq 1\}$ based on the MJ-posteriors varies with the value of ε . They are similar to those computed under a flat prior for $\varepsilon = .001$ and but become larger as ε is increased. The magnitude of these expected probabilities for values of ε near .001 are somewhat misleading by themselves because the upper range of ρ is restricted to a small neighborhood beyond unity and they cannot illustrate the rightward skewness of the MJ-posteriors. As k increases in the fitted model two things tend to happen. First, the extra regressors tend to soak up variation in the data and this increases the downward bias in $\hat{\rho}$. In addition, the introduction of extra regressors tends to add more noise to the fitted model and this increases the estimated variance of $\hat{\rho}$. The simulation results indicate that this latter effect dominates the behavior of the posteriors. For all posteriors, the expected probabilities of $\{\rho \geq 1\}$ increase monotonically in k . The MJ-posteriors produce the largest increases, which is what we expect since the prior gives proportionally more weight to nonstationary models as k increases.

Next we consider the random walk with drift model with moving average errors. Phillips (1991a) found that the J-posterior was quite sensitive to the value of the moving average parameter, θ . In particular, for values of θ near -1 he found that $E[P_J(\rho \geq 1)] \approx 1$. This large expected probability is misleading since as θ approaches -1 the true DGP approaches a deterministically trending process with iid errors. As pointed out in Phillips, this outcome is the result of the bias that stems from the fact that (5) is not an ignorance prior when $\varphi \neq 0$. The MJ-posteriors explicitly account for the effects of the transient dynamics terms on ρ and thus we expect posterior inferences to be more objective than inferences based on the J-posterior when the transient dynamics terms play a large role in explaining the data. The results of the simulations support this assertion. For large negative values of θ , the MJ-expected posterior probabilities are small, although they increase as ε or k is increased. These results indicate that one would need a fitted model with a long lag length in order to find evidence for the unit root model. Interestingly, the MJ-expected probabilities increase as θ nears unity. It appears that the MJ-posteriors pick up the added persistence in the data whereas the F and J-posteriors do not.

3. TREND DETERMINATION IN STRUCTURAL CHANGE MODELS

3.1 Introduction

Recently, the appropriateness of the unit root model for a number of macroeconomic and financial time series has been questioned by Perron (1988, 1989, 1990). For example, using the Nelson and Plosser data and a U.S. postwar quarterly real GNP series, Perron (1989) argues that if the observations corresponding to the years of the Great Depression and the 1973 oil price shock are treated as exogenous events (points of structural change) then a flexible trend stationary representation is favored over a flexible difference stationary representation. Contrary to the unit root hypothesis, these results imply that the only observations that have had a permanent effect on the long-run level of most macroeconomic aggregates are those associated with the Great Depression and the first oil price crisis.

Perron's unit root testing methodology, which is conditional on structural change at a known point in time, has been criticized by Christiano (1988), Banerjee et al. (1990) and Zivot and Andrews (1990). These authors argue that Perron's unit root tests are biased against the unit root null because his choices of break points are correlated with the data and, hence, problems associated with pre-testing are applicable to his methodology. They suggest ways for correcting his tests for this bias based on procedures to estimate the structural change points. After these corrections are made, the overall evidence against the unit root hypothesis for the aforementioned series is not nearly as strong as Perron suggests.

The suggestions for correcting Perron's unit root tests made by the above authors, however, are problematic in two respects and only asymptotic results are available. The first problem arises from the fact that when the point of structural change is unknown some procedure, usually ad hoc, must be adopted for estimating it. This break point estimation procedure produces a "pre-testing" bias in any subsequent inference and the bias depends explicitly on the procedure employed. The second problem involves the nonstandard asymptotic distribution theory that results from models with integrated regressors and with estimated structural breaks. As shown in Banerjee et al. (1990), Hansen (1990) and Zivot and Andrews (1990), the resulting limit theory is quite complicated in these types of models. In addition, Zivot and Andrews (1990) and Perron (1990) demonstrate that the finite sample distributions of the unit root test statistics can be very different from their asymptotic counterparts.

An attractive alternative to the classical approach to remedy the "pre-testing" bias caused by the unknown location of the possible break point is to adopt a strictly Bayesian methodology. The Bayesian approach for handling structural change at an unknown point in time is straightforward and the distribution theory is valid in finite samples. Conditional inferences concerning the parameters of the model are avoided by placing a prior distribution over the location of the structural change point and the mechanics of Bayes' rule then produces posterior distributions from which unconditional inferences can be made.

Structural change in the linear model with fixed regressors from a Bayesian point of view has been investigated by several authors including Ferreira (1977), Chin Choy and Broemeling (1980), Smith (1980), Booth and Smith (1982), Holbert (1982) and Broemeling and Tsurumi (1987). Time series models have been addressed but the analysis has been limited to deterministically trending models with stable autoregressive errors. Stochastically trending processes have not, so far, been considered. Also, the primary focus of much of the previous work has been on the detection of structural change. Here, we are more interested in making inferences concerning the trend determining parameters of the model when we allow for the possibility of structural change.

Essentially, there are two problems associated with comparing the trend stationary structural change model under unknown change point with the unit root model. The first problem involves testing the two hypotheses. If it is determined that the data favor the unit root model then we are done. If, on the other hand, the data favor the trend stationary structural change model, then we encounter the second problem of estimating the change point as well as the other parameters of the model. Throughout this discussion we focus on the first of these two problems.

3.2. Prior and posterior analysis

We now expand the model considered in the previous sections to allow for certain forms of structural change. Specifically, we use as our underlying model

$$y_t = \mu + \beta t + d_\mu DU(r)_t + d_\beta DT(r)_t + \psi(L)y_t + u_t, \quad r = 2, \dots, T-2 \quad (30)$$

where $d_\mu = \mu_2 - \mu$ and $DU(r)_t = 1$ if $t > r$, 0 otherwise; $d_\beta = \beta_2 - \beta$ and $DT(r)_t = t - r$ if $t > r$, 0 otherwise. The inclusion of the dummy variables $DU(r)_t$ and $DT(r)_t$ allow for models with broken trends. The dummy variable $DU(r)_t$ allows for a one time change in the level of the series occurring at time r , while $DT(r)_t$ permits a one time change in the slope. If $2 \leq r \leq T-2$ and either $d_\mu \neq 0$ or $d_\beta \neq 0$ then there is exactly one change in the deterministic trend occurring at some unknown point r and we call (30) a structural change (SC) model. If $r = T$ then no change has occurred and (30) collapses to the no structural change (NSC) model (1) considered earlier. As with the NSC model, we employ the following reparameterization of (30) to focus attention on the dominant autoregressive component of $\psi(L)$:

$$y_t = \mu + \beta t + d_\mu DU(r)_t + d_\beta DT(r)_t + \rho y_{t-1} + \Sigma_1^{k-1} \varphi_i \Delta y_{t-i} + u_t, \quad r = 2, \dots, T-2. \quad (30')$$

For the SC model the parameter vector of interest is $\theta' = (\mu, \beta, d_\mu, d_\beta, \rho, \varphi', \sigma, r) \in \mathbb{R}^{k+4} \times \mathbb{R}_+ \times N_{[2, T-2]}$, where $N_{[2, T-2]}$ denotes the set of integers $\{2, 3, \dots, T-2\}$. As before, prior views about θ are given by $\pi(\theta)$. Since r is restricted to integer values, this prior will be a mixture of continuous and discrete parts. Concerning the point of structural change, we make the assumption that we have no information where it occurs. In addition, we assume that the break point, r , is independent of the other parameters in the model. These assumptions result in a uniform prior over r :

$$\pi(r) = (T-3)^{-1}, \quad r = 2, \dots, T-2. \quad (31)$$

As an ignorance prior for the break point r , the uniform prior is intuitively appealing and has been used by Ferreira (1977), Chin Choy and Broemeling (1980), Holbert (1982), Broemeling (1985) and Broemeling and Tsurumi (1987). For the remaining parameters, we use the three sets of priors discussed earlier.

With the addition of the dummy variables, the focus of our analysis expands to include parameters d_μ , d_β , the possible structural change date r as well as ρ and β . We obtain the unconditional marginal posterior pdf's (pmf for r) for these parameters under the SC model in three steps. First we determine the marginal posterior mass function (pmf) for the break point r . This is obtained by integrating the joint posterior with respect to μ , β , d_μ , d_β , ρ , φ and σ for $r = 2, \dots, T-2$ (we exclude the endpoints $r = 1$ and $r = T-1$ to avoid singularities in the data matrix). Next, we derive the marginal posterior pdf's for β , ρ , d_μ and d_β conditional on the break date r being equal to some value. Lastly, we obtain the unconditional marginal densities simply by averaging the conditional marginal pdf's with respect to the marginal pmf of r . For example, the unconditional marginal posterior pdf for ρ is given by

$$p(\rho | y, \iota_0) \propto \Sigma_2^{T-2} p(r | y, \iota_0) p(\rho | r, y, \iota_0) \quad (32)$$

The unconditional marginal posterior densities are mixtures of univariate densities where the mixing variate is the posterior mass function of r , $p(r | y, \iota_0)$. If the change point r is known, then the conditional densities, $p(\rho | r, y, \iota_0)$, \dots , $p(d_\beta | r, y, \iota_0)$, are the appropriate ones to use for making inferences. When the break point is unknown, the Bayesian analysis averages the $T-3$ conditional densities with respect to the marginal posterior pmf of r .

Let $\delta' = (\mu, \beta, d_\mu, d_\beta, \varphi')$, $\pi' = (\rho, \mu, d_\mu, d_\beta, \varphi')$, $\nu' = (\rho, \mu, \beta, d_\beta, \varphi')$, $\eta' = (\rho, \mu, \beta, d_\mu, \varphi')$, $\xi' = (\rho, \delta')$ and use $du(r)$, $d\tau(r)$, $V(r)$, $Z(r)$, $U(r)$, $W(r)$ and $X(r)$ to represent the observation matrices of $(DU(r)_t)$, $(DT(r)_t)$, $(1, t, DU(r)_t, DT(r)_t, \Delta y_{t-1}, \dots, \Delta y_{t-k+1})$, $(y_{t-1}, 1, DU(r)_t, DT(r)_t, \Delta y_{t-1}, \dots, \Delta y_{t-k+1})$, $(y_{t-1}, 1, t, DT(r)_t, \Delta y_{t-1}, \dots, \Delta y_{t-k+1})$, $(y_{t-1}, 1, t, DU(r)_t, \Delta y_{t-1}, \dots, y_{t-k+1})$ and $(y_{t-1}, 1, t, DU(r)_t, DT(r)_t, \Delta y_{t-1}, \dots, \Delta y_{t-k+1})$, respectively. In addition, we have the following decompositions of the sum of squares $\Sigma_1^T (y_t - \xi'x_t)^2$:

$$m(\hat{u}(r)) + (\rho - \hat{\rho}(r))^2 m_{V(r)}(y_{-1}) + (\delta - \hat{\delta}(r))' V(r)' V(r) (\delta - \hat{\delta}(r)), \quad (33)$$

$$m(\hat{u}(r)) + (\beta - \hat{\beta}(r))^2 m_{Z(r)}(r) + (\pi - \hat{\pi}(r))' Z(r)' Z(r) (\pi - \hat{\pi}(r)), \quad (34)$$

$$m(\hat{u}(r)) + (d_\mu - \hat{d}_\mu(r))^2 m_{U(r)}(d_\mu(r)) + (\nu - \hat{\nu}(r))' U(r)' U(r) (\nu - \hat{\nu}(r)), \quad (35)$$

$$m(\hat{u}(r)) + (d_\beta - \hat{d}_\beta(r))^2 m_{W(r)}(d\tau(r)) + (\eta - \hat{\eta}(r))' W(r)' W(r) (\eta - \hat{\eta}(r)), \quad (36)$$

$$m(\hat{u}(r)) + (\xi - \hat{\xi}(r))' X(r)' X(r) (\xi - \hat{\xi}(r)), \quad (37)$$

where $m(\hat{u}(r)) = \Sigma_1^T \hat{u}(r)_t^2$, $\hat{u}(r)_t$ are the OLS residuals for a fixed r and

$$\hat{u}(r) = y - X(r)\hat{\xi}(r),$$

$$\hat{\delta}(r) = \hat{\delta}(r, \rho) = \hat{\delta}(r) + (V(r)' V(r))^{-1} V(r)' y_{-1} (\hat{\rho}(r) - \rho),$$

$$\hat{\pi}(r) = \hat{\pi}(r, \beta) = \hat{\pi}(r) + (Z(r)' Z(r))^{-1} Z(r)' \pi (\hat{\beta}(r) - \beta),$$

$$\hat{\nu}(r) = \hat{\nu}(r, d_\mu) = \hat{\nu}(r) + (U(r)' U(r))^{-1} U(r)' du(r) (\hat{d}_\mu(r) - d_\mu),$$

$$\hat{\eta}(r) = \hat{\eta}(r, d_\beta) = \hat{\eta}(r) + (W(r)' W(r))^{-1} W(r)' d\tau(r) (\hat{d}_\beta(r) - d_\beta).$$

The following theorem gives, for the SC model, the marginal posterior pmf of r and the unconditional marginal posterior pdf's of the trend determining parameters ρ , β , d_μ and d_β based on the priors (29) and (31).

THEOREM 2

- (a) $p(r|\varepsilon, y, \iota_0) \propto \exp\{-\hat{\rho}(r)^{-2c(\varepsilon)}\}(\prod_1^k \alpha_{0ii}(\hat{\rho}(r), \hat{\varphi}(r)))^{1/2} |X(r)'X(r)|^{-1/2} m(\hat{\iota}(r))^{-(T-k-4)/2}$
- (b) $p(\rho|\varepsilon, y, \iota_0) \propto \Sigma_2^{T-2} p(r|y, \iota_0) \exp\{-\rho^{-2c(\varepsilon)}\}(\prod_1^k \alpha_{0ii}(\rho, \hat{\varphi}(r, \rho)))^{1/2} \times [m(\hat{\iota}(r)) + (\rho - \hat{\rho}(r))^2 m_{V(r)}(y_{-1})]^{-(T-k-3)/2}$
- (c) $p(\beta|\varepsilon, y, \iota_0) \propto \Sigma_2^{T-2} p(r|y, \iota_0) \exp\{-\beta^{-2c(\varepsilon)}\}(\prod_1^k \alpha_{0ii}(\beta, \hat{\varphi}(r, \beta)))^{1/2} \times [m(\hat{\iota}(r)) + (\beta - \hat{\beta}(r))^2 m_{Z(r)}(\tau)]^{-(T-k-3)/2}$
- (d) $p(d_\mu|\varepsilon, y, \iota_0) \propto \Sigma_2^{T-2} p(r|y, \iota_0) \exp\{-\hat{\rho}(r, d_\mu)^{-2c(\varepsilon)}\}(\prod_1^k \alpha_{0ii}(\hat{\rho}(r, d_\mu), \hat{\varphi}(r, d_\mu)))^{1/2} \times [m(\hat{\iota}(r)) + (d_\mu - \hat{d}_\mu(r))^2 m_{U(r)}(du(r))]^{-(T-k-3)/2}$
- (e) $p(d_\beta|\varepsilon, y, \iota_0) \propto \Sigma_2^{T-2} p(r|y, \iota_0) \exp\{-\hat{\rho}(r, d_\beta)^{-2c(\varepsilon)}\}(\prod_1^k \alpha_{0ii}(\hat{\rho}(r, d_\beta), \hat{\varphi}(r, d_\beta)))^{1/2} \times [m(\hat{\iota}(r)) + (d_\beta - \hat{d}_\beta(r))^2 m_{W(r)}(d\tau(r))]^{-(T-k-3)/2}$

Remarks (i) Under a flat prior, the posterior pmf of r is

$$p(r|y, \iota_0) \propto |X(r)'X(r)|^{-1/2} m(\hat{\iota}(r))^{-(T-k-4)/2},$$

and the unconditional posterior pdf's of the trend parameters are mixtures of conditional Student-t densities. These results are analogous to those obtained by Ferreira (1977) and Chin Choy and Broemeling (1980) for the switching regression model with fixed regressors (see Holbert (1982), Broemeling (1985) and Broemeling and Tsurumi (1987) for a summary and some extensions). Under the approximate Jeffreys prior (23), the posterior pmf of r is

$$p(r|y, \iota_0) \propto \alpha_0(\hat{\rho}(r))^{1/2} |X(r)'X(r)|^{-1/2} m(\hat{\iota}(r))^{-(T-k-4)/2},$$

and the posteriors of the trend parameters are mixtures of conditional densities of the form derived earlier.

(ii) If we define the Bayes estimator of r to be the break point with the largest posterior mass, then we see that the Bayes estimator is very close to the classical estimator of the break point which minimizes the sum of squared residuals.

To evaluate the evidence for the unit root model given by the data when we allow for the possibility of structural change, we need to combine the SC and NSC posteriors of the trend determining parameters. To do this, let $q \in [0, 1]$ denote the prior probability of no structural change. Then the posterior densities of the trend parameters become a weighted average of the NSC and SC posteriors.

4. EMPIRICAL APPLICATIONS

4.1. The Nelson-Plosser data set

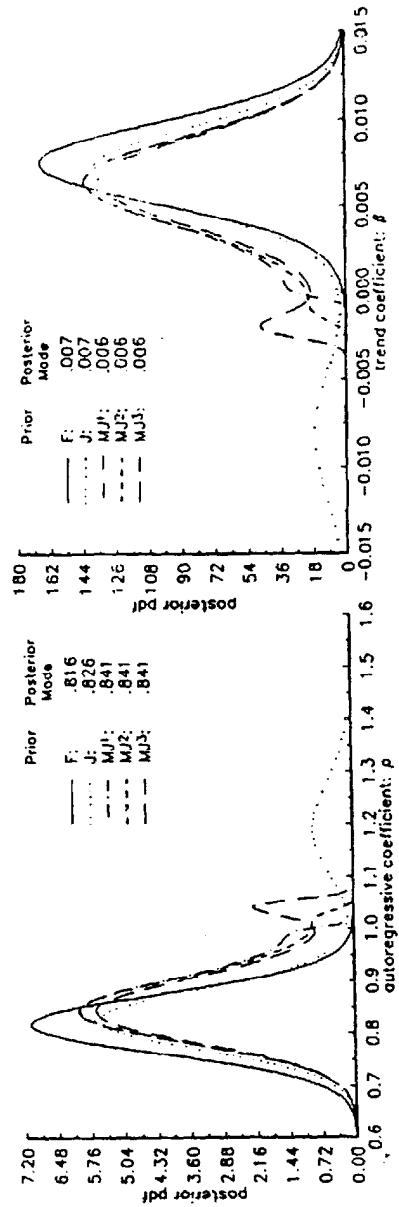
The first data set we analyze is that used by Nelson and Plosser (1982). The data set includes the following fourteen annual macroeconomic series: real GNP, nominal GNP, real per

capita GNP, industrial production, employment, unemployment rate, GNP deflator, consumer prices, nominal wages, real wages, money stock, velocity, common stock prices and bond yields. The start dates for the series vary from 1860 for industrial production and consumer prices to 1909 for the GNP series. All series terminate in 1970. We analyze the natural logarithms of all of the series except bond yields which we analyze in levels form.

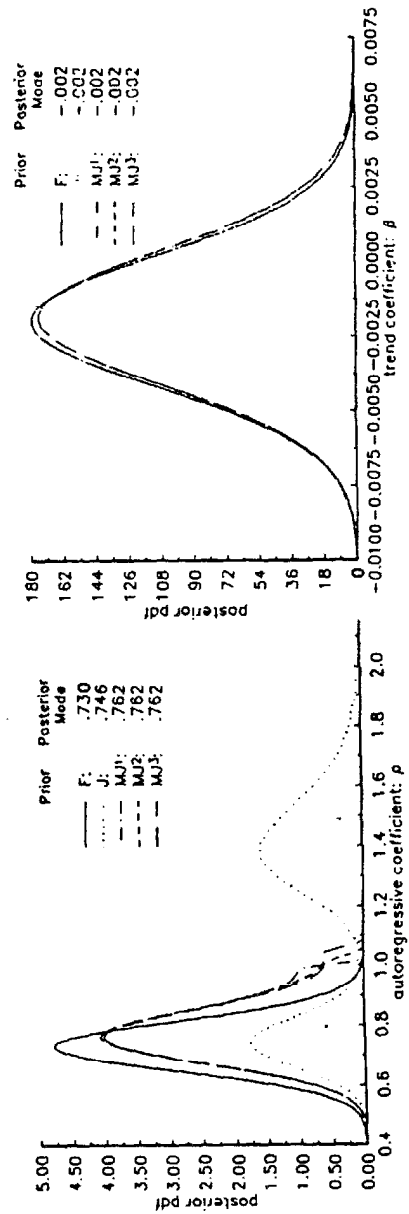
Nelson and Plosser used the model formulation (1'), with k determined from the data, to conduct ADF unit root tests on the series. They could not reject the unit root hypothesis at the 5% level of significance for all of the series except the unemployment rate. Perron (1988) arrived at similar conclusions using the Phillips-Perron unit root tests. In contrast, DeJong and Whiteman (1989a) employed a flat prior Bayesian analysis based on the model (1) with $k = 3$ and found that the data favored a TS representation over a DS representation for all of the series except consumer prices, velocity and bond yields. In the spirit of Geweke (1988), their analysis focused on the modulus of the dominant root, Λ , of the equation $1 - \psi(L) = 0$ and the time trend parameter β and they derived posterior densities for these parameters based on flat priors for the coefficients of (1) by Monte Carlo integration methods (see Phillips (1991b) for more discussion on the implications of this parameterization). Their inferences concerning trend behavior are based on the posterior probabilities of the events $\{\Lambda \geq .975 | A\}$ and $\{\beta \leq .001 | A\}$, where A is the event $\{0 \leq \beta < .016, .55 \leq \Lambda < 1.055\}$. They found appreciable probabilities for these events for a majority of the series only when they restricted β to be equal to zero. Phillips (1991a) also conducted a Bayesian analysis of the Nelson-Plosser data using a flat prior and his approximate Jeffreys prior (5) for the model (1') with $k = 1$ and $k = 3$. Using a flat prior, Phillips, in agreement with DeJong and Whiteman, found insufficient evidence in favor of the unit root model for most of the series. Using the Jeffreys prior, however, he found substantially more support for the unit root model for some series (notably, stock prices, industrial production and nominal GNP).

We expand on the analysis given in Phillips (1991a) by considering different model specifications and by including posterior inferences for the deterministic trend parameter. We utilize model (1') and compute posterior densities for the long-run autoregressive parameter ρ and the time trend parameter β based on the modified Jeffreys prior (29) (MJ-posteriors), Phillips' approximate Jeffreys prior (5) (J-posteriors) and the flat prior $\pi(\theta) \propto \sigma^{-1}$ (F-posteriors). For each series, we compute the posteriors for the case where k is specified as in Nelson and Plosser (1982) and for the case where $k = 3$ (as in DeJong and Whiteman (1989a)) to assess the sensitivity of the results to the specification of the model. In order to use the modified prior (29), we must specify the mode of the implied prior for ρ by selecting a value for ε . This choice determines the degree of nonstationarity allowed a priori and may have a large impact on posterior inferences. Instead of confining ourselves to a single value of ε for each series, we exploit the convenience of the analytic representation of our posteriors and report inferences based on a broad range of ε values. In this manner, we can address the robustness or fragility of our inferences to the specification of the prior.

Industrial Production



Unemployment Rate



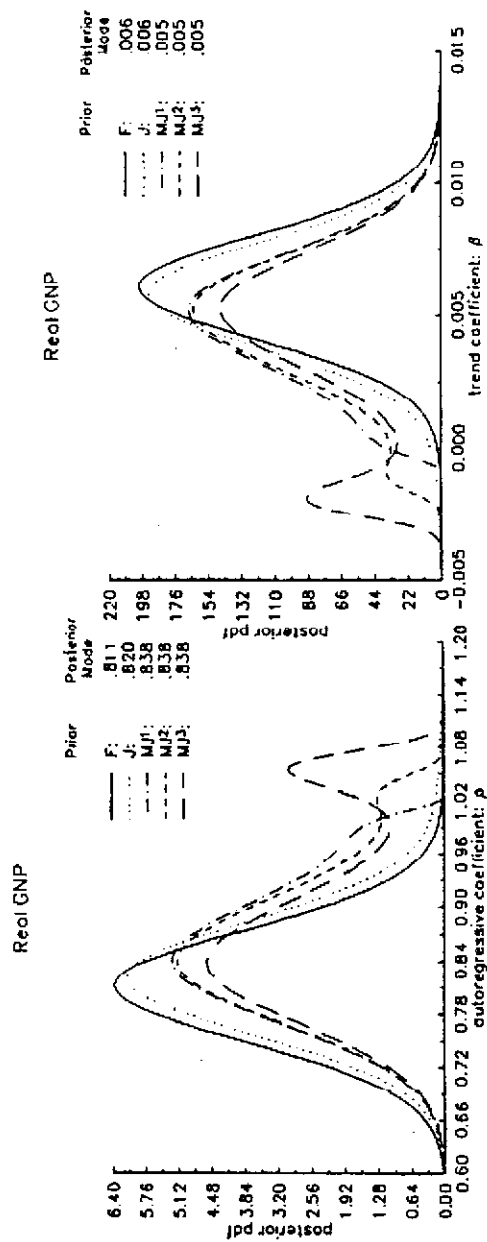
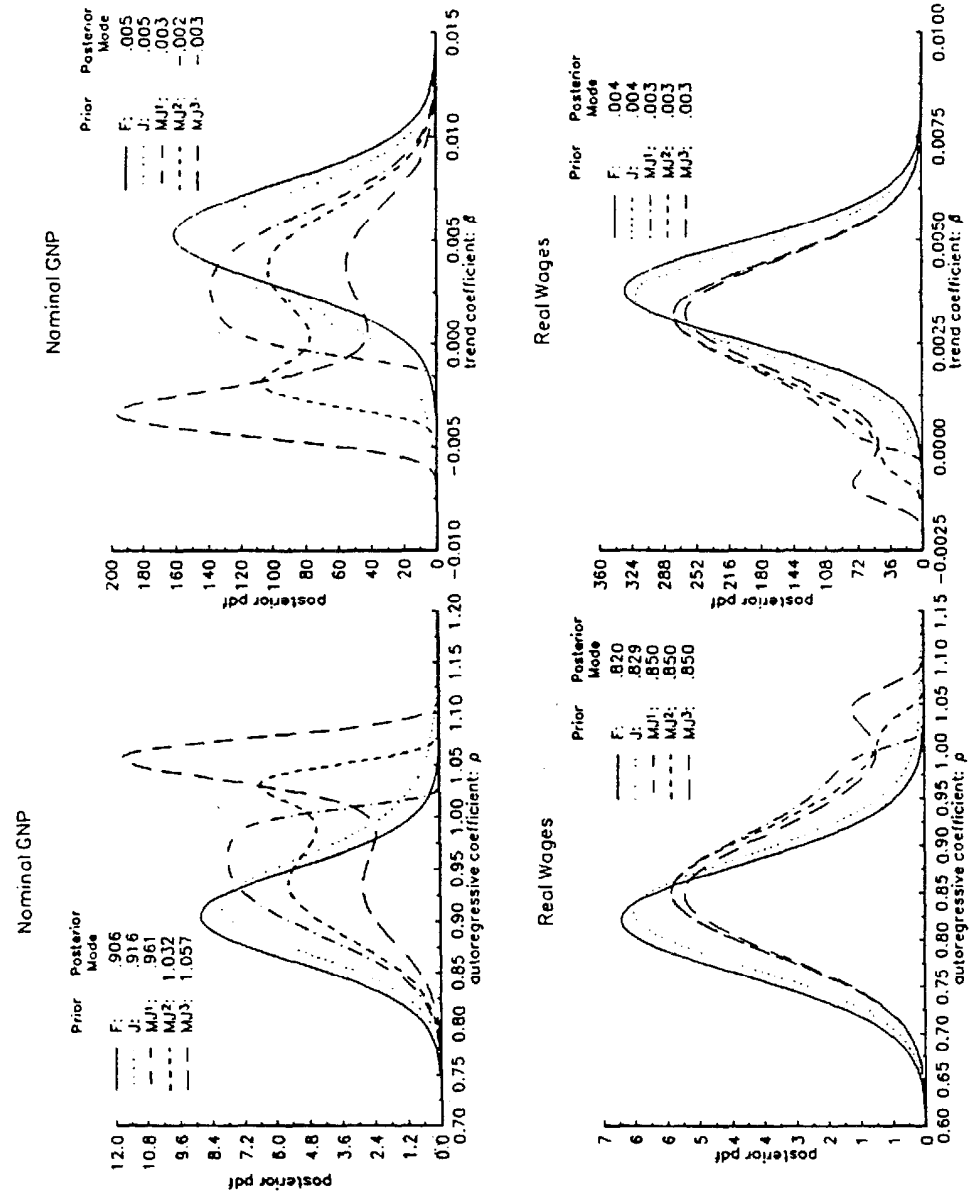


Figure 1



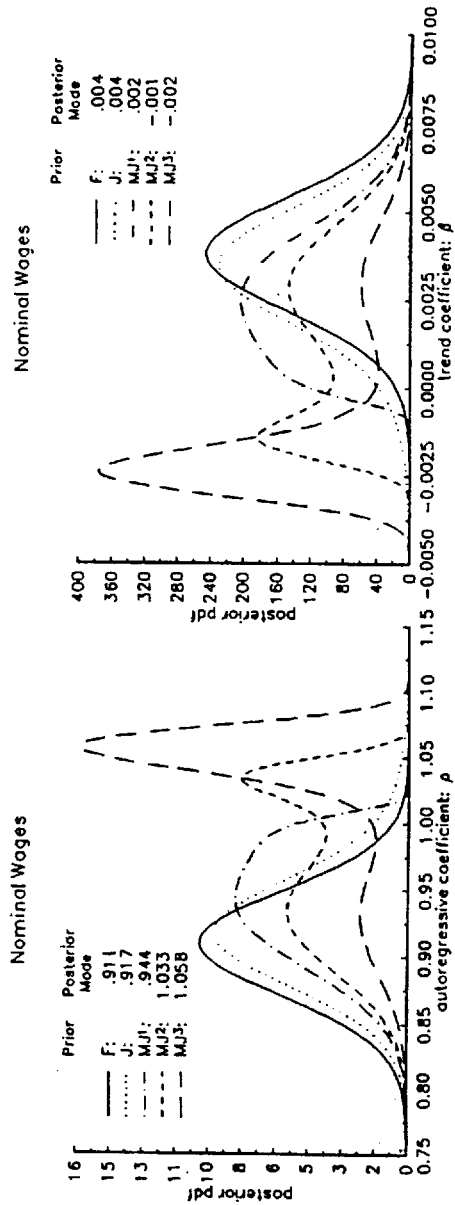


Figure 2

Figures 1 and 2 give the marginal posterior densities of ρ and β based on the three sets of priors for some of the Nelson-Plosser series computed using $k = 3$ (see Zivot and Phillips (1991) for a complete set of results). Each panel displays the MJ-posteriors computed for ε equal to .001, .025 and .050. These values of ε were chosen to illustrate the sensitivity or lack of sensitivity of the posteriors to the *a priori* degree of nonstationarity allowed. To retain comparability with DeJong and Whiteman, we restrict ε to be less than .055. Table 1 reports the posterior probabilities of the events $\{\rho \geq 1\}$ and $\{\beta \leq 0\}$ computed from the three sets of posteriors for each series for Nelson and Plosser's choice of k and $k = 3$. We infer that the unit root hypothesis is not implausible if $P(\rho \geq 1) \geq .05$.

Visual inspection of the posteriors reveals that there can be considerable differences in the shapes for the three sets of priors for some of the series. The F-posteriors are centered about the OLS estimates and, except for consumer prices, velocity and bond yields, they give virtually no evidence for the unit root model. For most of the series, the J-posteriors are quite similar in shape to the F-posteriors, although the modes of the J-posteriors for ρ are always to the right of the corresponding F-posterior modes and the J-posterior modes for β are always to the left of the respective F-posterior modes. From Figure 1, we see that the ρ -posteriors for industrial production and the unemployment rate are bimodal about unity and the bimodality is such that the regions of highest posterior density (HPD) are disjoint, indicating considerable uncertainty about the true value of ρ . The β -posterior for industrial production is bimodal and disjoint about zero. Interestingly, even though the ρ -posterior for the unemployment rate is bimodal and disjoint, the β -posterior is nearly symmetric and indicates that a deterministic trend is most likely not present. The MJ-posteriors for ρ and β vary considerably depending on the series and on the value of ε . In general, the real series appear to be trend stationary whereas the nominal series appear to have a unit root. For all values of ε considered, the posteriors for real GNP, per capita real GNP, industrial production, unemployment, and real wages show little evidence of stochastic trends. In contrast, the posteriors for consumer prices, velocity, bond yields and stock prices indicate that stochastic trends are quite possible. However, unlike the F-posteriors or the J-posteriors, the MJ-posteriors for nominal GNP, employment, GNP deflator, nominal wages and the money stock give some evidence for the unit root model. Moreover, this evidence is sensitive to the value of ε . When $\varepsilon = .001$, the ρ -posteriors are heavily skewed toward unity and the β -posteriors are skewed toward the origin. The posterior modes for ρ (β), for the above series, are .961 (.003), .933 (.002), .892 (.002), .944 (.002) and .933 (.004), respectively. For larger values of ε , the posteriors become bimodal and give substantial evidence of stochastic nonstationarity.

The posterior probabilities presented in Table 1 allow us to compare trend behavior across model specifications and priors. Consider first the model specifications chosen by Nelson and Plosser (1982). Under the flat prior we have $P(\rho \geq 1) > .05$ and $P(\beta \leq 0) > .02$ only for the bond yields series. For Phillips' approximate prior, we include industrial production, the unemployment rate, nominal wages and velocity. For the modified Jeffreys prior with $\varepsilon = .001$ we get the same results as with the Phillips prior except we include consumer prices and exclude the

Table 1
Posterior Probabilities of Stochastic and Deterministic Nonstationary Nelson-Plosser Data

Event		$P\{\rho \geq 1\}$					$P\{\beta \leq 1\}$				
Series	k	F	J	MJ ¹	MJ ²	MJ ³	F	J	MJ ¹	MJ ²	MJ ³
Nominal GNP	2	.011	.042	.031		.245	.01	.035	.031	.104	.206
	3	.020	.081	.069		.647	.01	.068	.074	.310	.625
Real GNP	2	.002	.012	.006		.048	.00	.009	.006	.019	.041
	3	.002	.015	.014		.048	.00	.011	.013	.060	.174
Per Capita Real GNP	2	.002	.010	.006		.042	.00	.007	.008	.014	.026
	3	.002	.013	.012	.055	.156	.00	.008	.020	.047	.120
Industrial Production	6	.005	.964	.067	.287	.518	.00	.955	.182	.353	.556
	3	.001	.190	.006		.099	.00	.155	.012	.033	.094
Employment	3	.004	.044	.024		.418	.00	.039	.044	.131	.360
Unemployment Rate	4	.001	.092	.007	.032	.099	.59	.587	.556	.556	.556
	3	.001	.625	.005	.018	.038	.81	.810	.793	.793	.793
Consumer Prices	4	.024	.063	.173		.996	.00	.007	.030	.662	.954
	3	.073	.181	.221		.992	.02	.050	.082	.654	.894
GNP Deflator	2	.006	.027	.023		.305	.00	.018	.017	.096	.250
	3	.005	.024	.044		.875	.00	.016	.038	.410	.876
Wages	3	.013	.053	.059		.747	.01	.042	.049	.333	.730
Real Wages	2	.001	.012	.004		.024	.00	.009	.003	.010	.020
	3	.002	.017	.011		.101	.00	.013	.007	.037	.097
Velocity	2	.040	.313	.051	.197	.382	.81	.658	.687	.662	.604
	3	.070	.533	.095		.736	.76	.511	.507	.420	.282
Money Stock	2	.001	.005	.012		.381	.00	.005	.012	.114	.328
	3	.003	.010	.045		.982	.00	.011	.048	.720	.978
Bond Yields	3	.758	.996	.506		.999	.03	.020	.685	.032	.018
Stock Prices	2	.004	.043	.012		.147	.00	.015	.011	.028	.063
	3	.017	.220	.046		.548	.00	.077	.024	.161	.428

Notes: "F", "J", "MJ¹", "MJ²" and "MJ³" denote "Flat," "Jeffreys," "Modified Jeffreys: $\epsilon = .001$," "Modified Jeffreys: $\epsilon = .025$ " and "Modified Jeffreys: $\epsilon = .050$," respectively. The first value of k is the value specified in Table 5 of Nelson and Plosser (1982). For some series, this value is equal to 3.

unemployment rate. When $\epsilon = .050$, the probability inequalities hold for all series except real GNP, per capita real GNP and real wages. The large posterior probabilities of stochastic nonstationarity for industrial production are due mostly to the dominating behavior of the prior, which occurs since the lag length is large ($k = 6$). Notice that the MJ-posteriors for unemployment do not indicate that a unit root is present as opposed to the J-posteriors.

For the model with k set equal to three, the posterior probabilities computed under the flat prior satisfy $P(\rho \geq 1) \geq .05$ and $P(\beta \leq 0)$ for consumer prices, velocity and bond yields. Using the J-posteriors we add nominal GNP, industrial production, unemployment, nominal wages and stock prices. The MJ-posteriors for $\epsilon = .001$ give the same results as the J-posteriors except for industrial production, unemployment and stock prices. When $\epsilon = .005$, the unit root model appears plausible for all of the series except real GNP and per capita GNP. Even though the posterior probabilities of stochastic nonstationarity for the GNP deflator, money stock and stock price series are not very large, their ρ -posteriors are heavily skewed toward unity and their β -

posteriors are skewed toward the origin. The posteriors are also very sensitive to the a priori amount of stochastic nonstationarity permitted. Thus, the unit root model is not necessarily implausible for these series as well.

Our empirical results are in general agreement with the results obtained by Phillips (1991a). The only series for which our interpretation of stochastic structure differs substantially from that of Phillips are industrial production and the unemployment rate. For $k = 3$, Phillips' unmodified prior produces posteriors for ρ that have significant second modes beyond unity. Our modified prior that attenuates extreme unstable values of ρ does not produce posteriors that reflect this much uncertainty.

4.2. Stock price and dividend data

In this section we analyze the annual stock price and dividend data examined by DeJong et al. (1988) and DeJong and Whiteman (1989b). These include the Dow Jones Industrial Averages (1928-1978) used by Shiller (1978), the value-weighted New York Stock Exchange Index data (1926-1981) used by Marsh and Merton (1987) and the Standard and Poor's 500 series (1871-1985) used by DeJong et al. (1988).

DeJong et al. (1988) conducted various classical unit root tests on the above data and they could not reject the unit root hypothesis at the 5% level for all of the series except the Standard and Poor's dividend series. Using a trend stationarity test, however, they could not reject the null hypothesis at the 5% level that $\beta \neq 0$ and $\rho = .85$ either. Faced with these conflicting results, DeJong and Whiteman (1989b) resort to a flat prior Bayesian analysis of the type used in their earlier paper to try to settle the issue. Based on their Bayesian analysis, they find that trend-stationarity is strongly supported by the data and only when the trend coefficient is restricted to zero *a priori* do the data admit unit roots.

Our Bayesian analysis of these series is summarized in Table 2, where we compute posterior probabilities of stochastic and deterministic nonstationarity for models with $k = 3$ (as in DeJong and Whiteman (1989b)) and $k = 4$ (as in DeJong et al. (1988)). Figures 3 and 4 display the posterior pdf's of ρ and β computed under the three sets of priors described earlier for a model with $k = 3$. As in the previous section, the MJ-posteriors are computed for three values of ϵ to show the sensitivity of the posteriors to the specification of the prior.

The posteriors computed under the three sets of priors vary considerably. As with the Nelson-Plosser data, the F-posteriors for the stock price and dividend data give very little, if any, evidence in support of the unit root model. For all of the series, the posterior probability that ρ exceeds unity is less than .036 and the probability that β exceeds zero is greater than .938.

The J-posteriors, on the other hand, give considerable support for the unit root model for all of the series with the possible exception of the S&P 500 dividend series. The ρ -posteriors are bimodal about unity, with the stock price series exhibiting more substantial second modes (for $\rho > 1$) than the dividend series. The β -posteriors are not bimodal like the ρ -posteriors but they often have significant density to the left of the origin. When $k = 3$, all of the stock price series

Table 2
Posterior Probabilities of Stochastic and
Deterministic Nonstationary Stock Price and Dividend Data

Event		$P(\rho \geq 1)$					$P(\beta \leq 1)$				
Series	k	F	J	MJ ¹	MJ ²	MJ ³	F	J	MJ ¹	MJ ²	MJ ³
Dow Jones Dividends	3	.003	.368	.009	.025	.042	.008	.027	.054	.048	.052
	4	.001	.445	.010	.029	.050	.003	.016	.048	.048	.051
Dow Jones Stock Prices	3	.015	.563	.026	.070	.114	.051	.091	.186	.175	.175
	4	.022	.899	.048	.147	.244	.068	.154	.353	.351	.367
NYSE Dividends	3	.009	.227	.025	.078	.147	.011	.116	.051	.088	.140
	4	.005	.099	.024	.073	.126	.004	.053	.022	.061	.104
NYSE Stock Prices	3	.035	.530	.045	.131	.218	.054	.265	.165	.197	.243
	4	.025	.389	.047	.144	.232	.029	.181	.142	.186	.238
S&P 500 Dividends	3	.000	.140	.001	.013	.065	.001	.003	.009	.009	.017
	4	.000	.394	.007	.050	.261	.000	.012	.021	.028	.083
S&P 500 Prices	3	.007	.411	.025	.123	.328	.037	.069	.156	.149	.164
	4	.002	.132	.034	.257	.726	.017	.030	.131	.140	.209

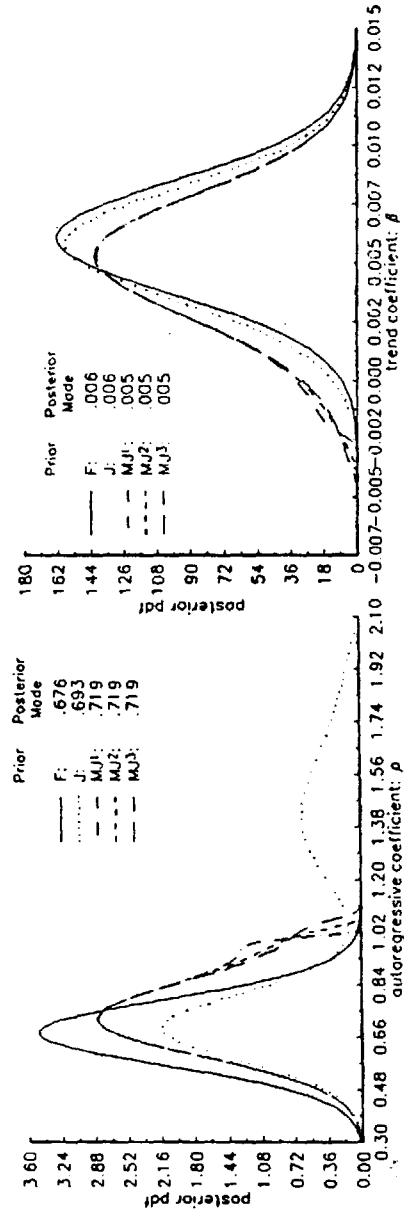
Notes: See the notes for Table 1. The data were generously provided by Charles Whiteman.

have $P(\rho \geq 1) > .41$ and $P(\beta \leq 0) > .69$. The Dow Jones and NYSE dividend series satisfy $P(\rho \geq 1) > .22$ and $P(\beta \leq 0) > .02$. Although the S&P 500 dividend series has $P(\rho \geq 1) = .14$, it has $P(\beta \leq 0) = .003$ which is quite small. The evidence for the unit root model for this series is tenuous.

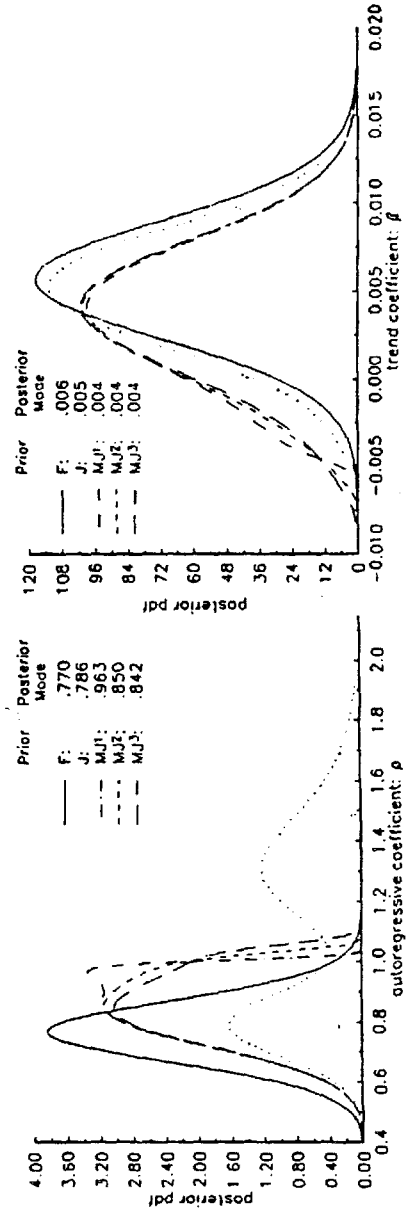
The MJ-posteriors for the stock price series show some support for the unit root model whereas the posteriors for the dividend series, with the possible exception of the NYSE series, give relatively little support. For the stock price series with $k = 3$ and $\epsilon = .001$, the ρ -posteriors are heavily skewed toward unity. The posterior modes are .963 (Dow Jones), .975 (NYSE) and .915 (S&P 500). The β -posteriors are more symmetric than the ρ -posteriors but they have substantial density to the left of the origin: the β -posteriors for the three stock price series satisfy $P(\beta \leq 0) > .15$. In addition, for all of the stock price series except the S&P 500 series, as ϵ is increased the posteriors for ρ and β do not change very much. For the dividend series with $k = 3$ and $\epsilon = .001$, the ρ posterior modes are .719, .855 and .798 and all series have $P(\rho \geq 1) < .025$ and $P(\beta \leq 0) < .051$. The posteriors for the NYSE dividend series are sensitive to the value of ϵ , indicating that inferences concerning the unit root hypothesis are fragile.

As illustrated above, posterior inferences concerning the stochastic structure of these stock price and dividend series are quite sensitive to the specification of the priors. They are not, however, particularly sensitive to the model specification. If one adopts a flat prior, like DeJong and Whiteman (1989b), then one is led to believe that the data are trend stationary. If, on the other hand, one uses a Jeffreys-type prior then one is led to believe that only some of the dividend series are trend stationary. One cannot easily dismiss the unit root model for the stock price series and the NYSE dividend series.

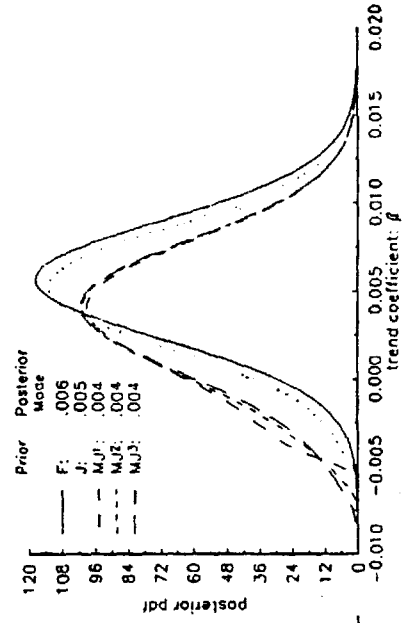
Dow Jones Dividends



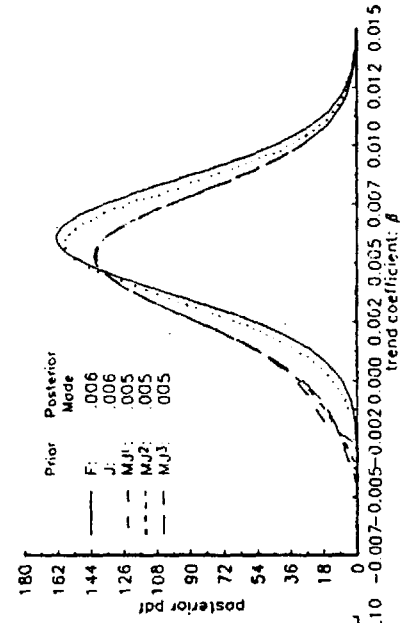
Dow Jones Stock Prices



Dow Jones Stock Prices



Dow Jones Dividends



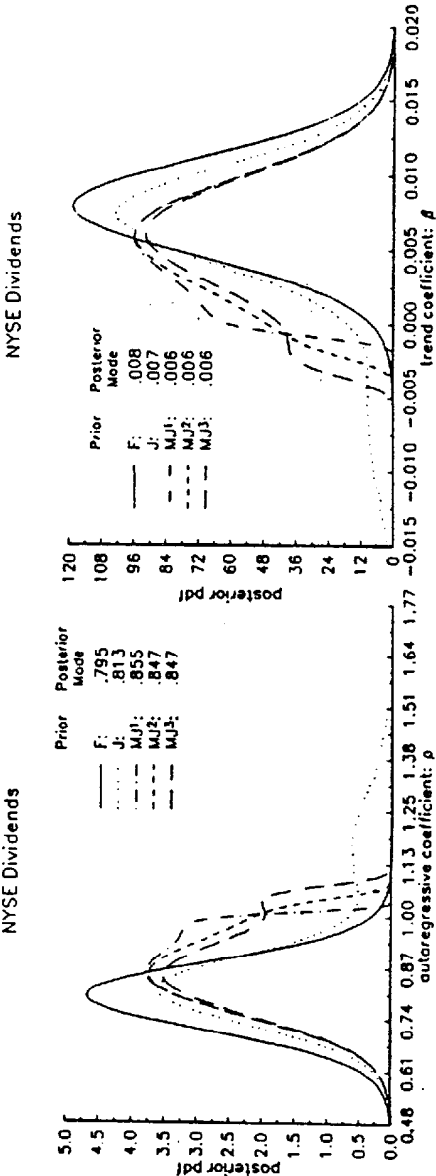
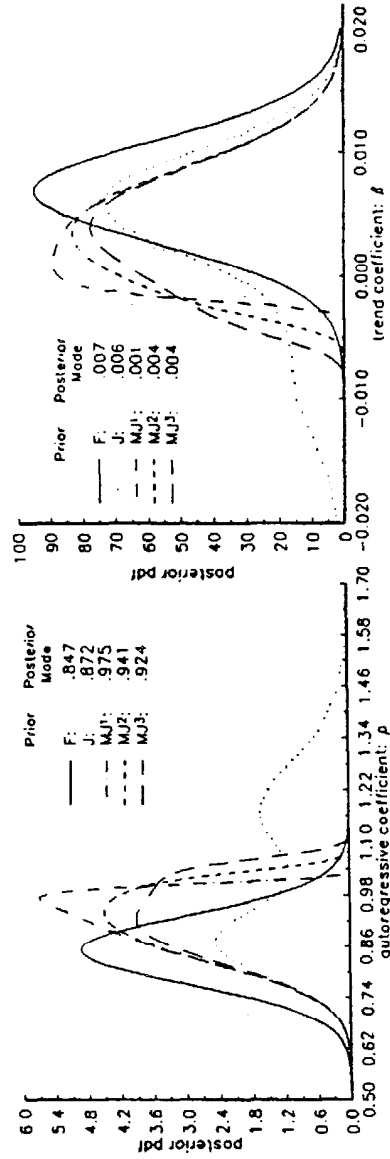
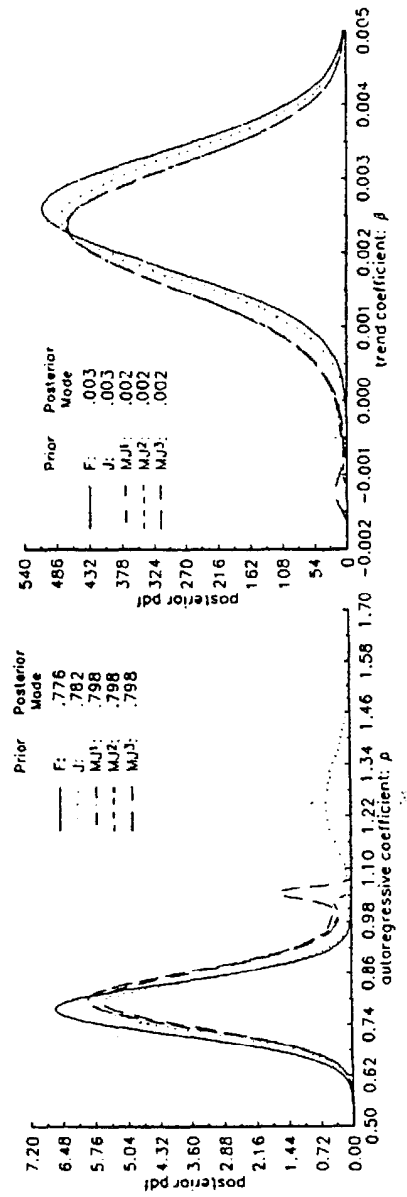


Figure 3

NYSE Stock Prices



S & P 500 Dividends



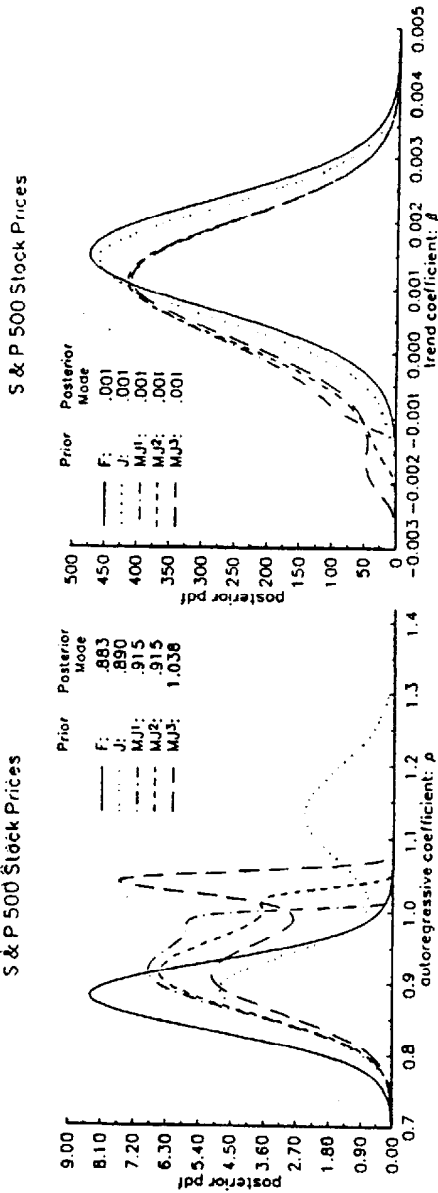


Figure 4

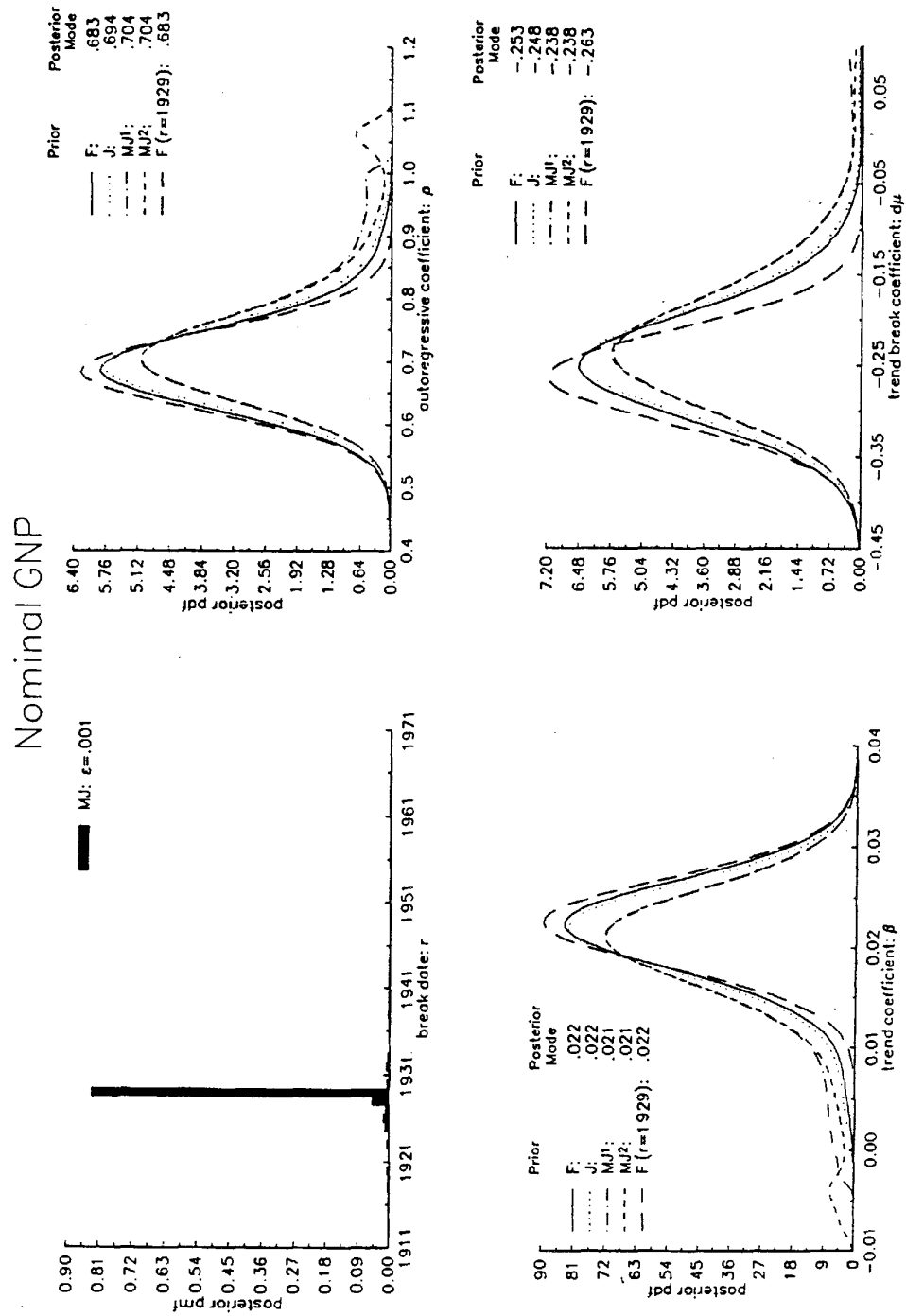


Figure 5

Nominal Wages

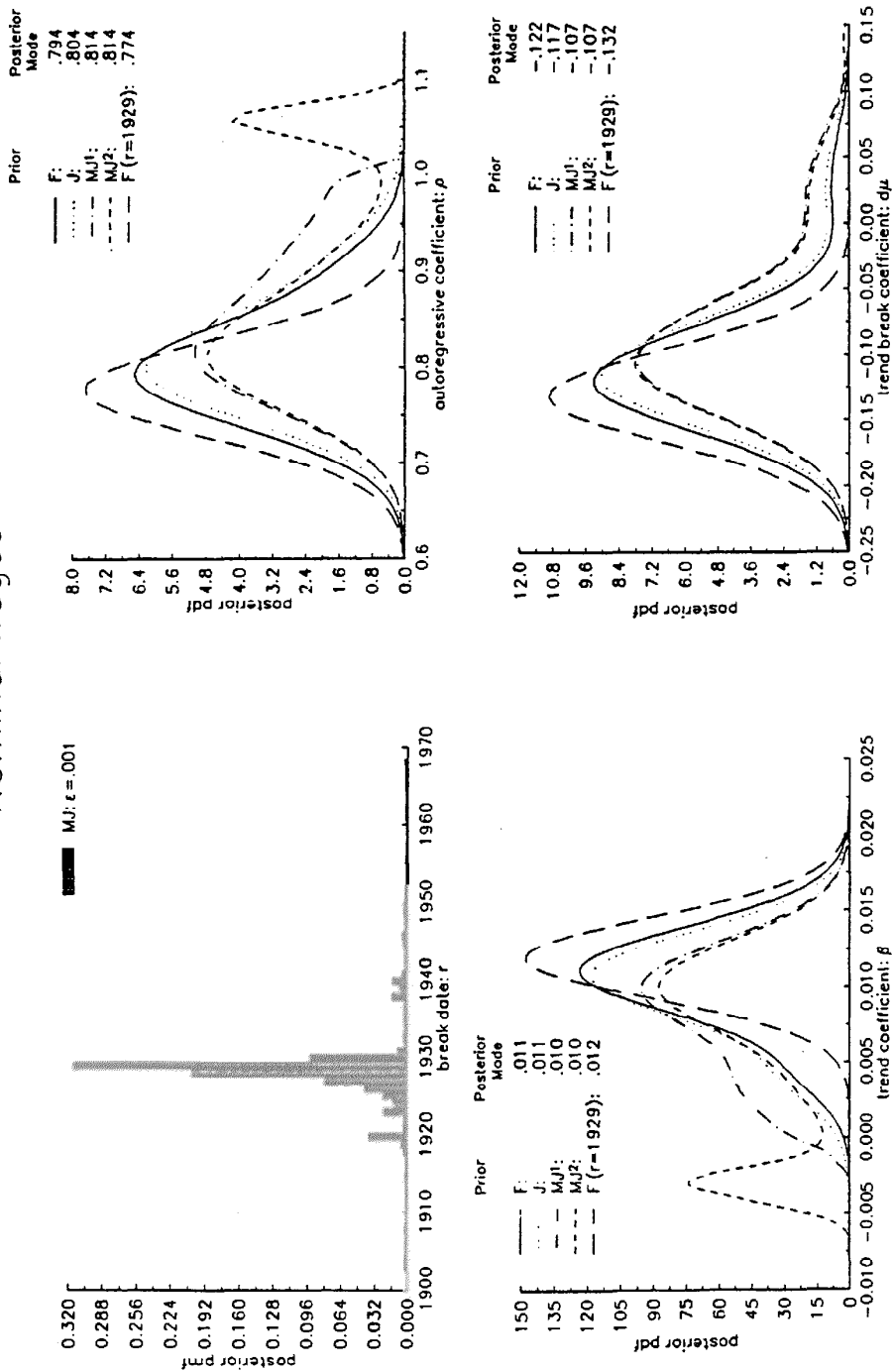


Figure 6

Employment

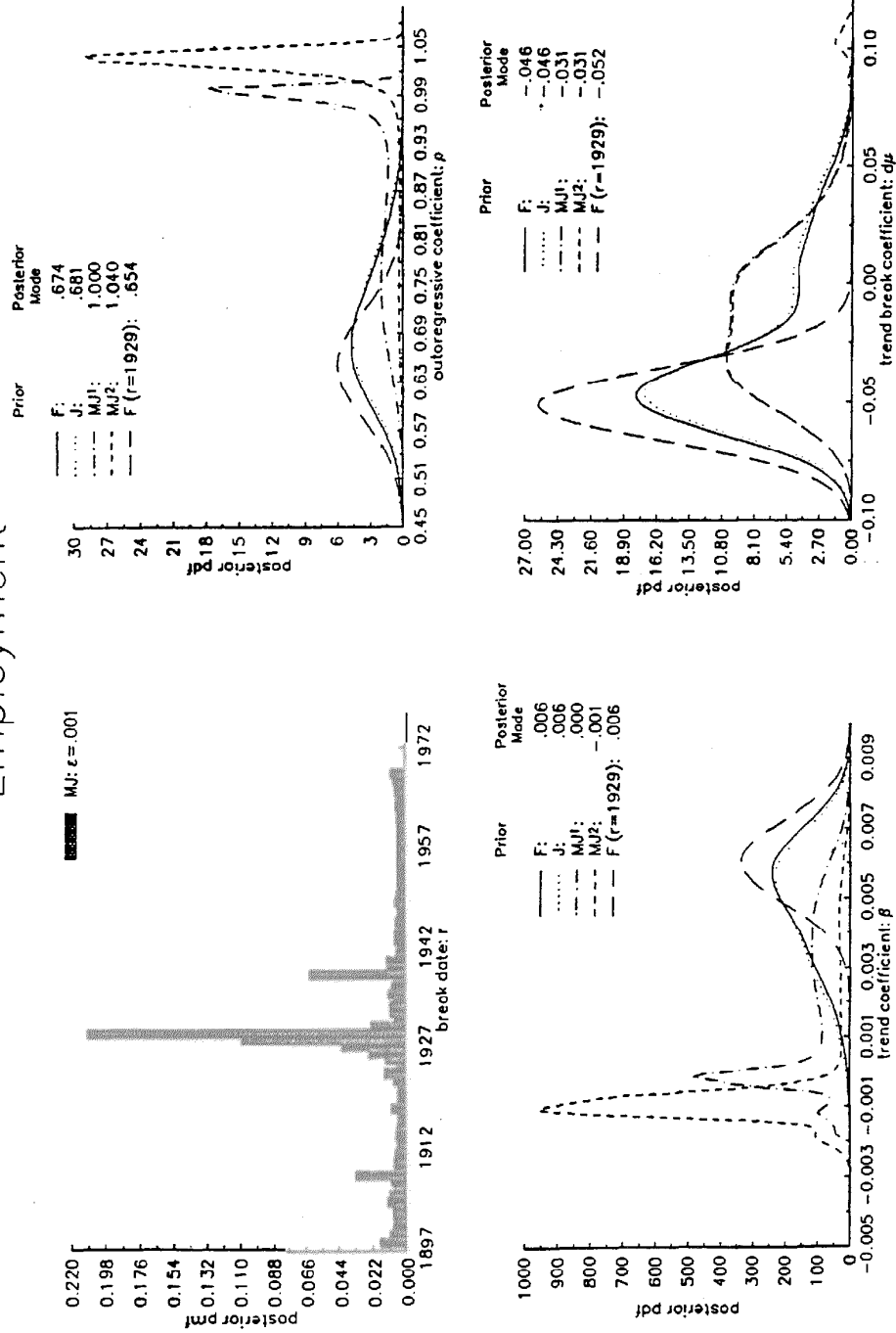


Figure 7

4.3. The Nelson-Plosser data set revisited

In this section we reexamine some of the Nelson-Plosser data by considering models that allow for the possibility of structural change in the deterministic components. Specifically, we utilize two versions of (30'). The first model we analyze, which we call Model (A), has the restriction that $d_\beta = 0$ a priori. This model allows for a one time change in the level of the series occurring at time r . Perron (1989) called this the "crash" model. Our second model, Model (B), allows $d_\mu \neq 0$ and $d_\beta \neq 0$. Perron used a variant of these two models for some of the Nelson-Plosser data to test the null hypothesis of a unit root with structural change against the alternative of trend stationarity with a broken trend where the point of structural change was set a priori at 1929. By allowing for a flexible trend under the alternative, Perron could reject the unit root hypothesis at the 5% level of significance for all of the series except consumer prices, velocity and bond yields.

We now examine the sensitivity of Perron's results to his exogeneity assumption concerning the break date by assuming ignorance about the location of the change point and computing the marginal posterior pmf of r and the unconditional posterior pdf's for ρ , β and d_μ (and d_β for stock prices). We focus our analysis on the series for which we earlier found evidence of stochastic trends. These series include nominal GNP, GNP deflator, consumer prices, employment, nominal wages, money stock, velocity, stock prices and bond yields. Following Perron, we use Model (A) for all of the above series except stock prices, for which we use Model (B). We compute the posteriors with values of k used in section 4.1 and with values of k used by Perron (1989).

Figures 5-7 display the posterior pmf of the break date r , the unconditional marginal pdf's of ρ , β and d_μ (and d_β for stock prices) as well as the conditional densities of these parameters for $r = 1929$ for the nominal GNP, nominal wage and employment series (see Zivot and Phillips (1991) for a full set of results). As in the previous sections, these posteriors are computed under three sets of priors. The mass functions and densities are plotted for q , the prior probability of no structural change, equal to zero to illustrate the most extreme case. Table 3 summarizes the posterior pmf's of r , Table 4 summarizes the posterior pdf's of the trend determining parameters for the Model (A) series and Table 5 gives the results for the Model (B) series (stock prices).

Consider first the posterior pmf's of r . Because the posterior pmf's of r give essentially the same inferences for most of the series, we only discuss the values of the mass functions computed from the MJ-posterior with $\epsilon = .001$. Notice that the pmf's computed under the three sets of priors for a given lag length are very similar and, in most cases, give the highest posterior mass to the same value of r . In addition, the F-posteriors are generally larger than the J or MJ-posteriors. Second, the break date with highest posterior mass varies with the lag specification for some of the series. If we define the Bayes estimator of the break date to be the date with highest posterior mass, then $r = 1929$ is chosen as the Bayes estimate for nominal GNP and nominal wages irrespective of the value of k and it is chosen for employment only when $k = 8$. When $k = 3$, the most likely change date for employment is 1894. With q , the prior probability of no structural change, equal to zero the MJ-posterior ($\epsilon = .001$) masses associated with $r = 1929$ for

Table 3
Break Dates with Highest Posterior Mass: $q = 0$

Rank		1				2				3			
Series	k	F	J	MJ ¹	MJ ³	F	J	MJ ¹	MJ ³	F	J	MJ ¹	MJ ³
Nominal GNP	2	1929	1929	1929	1929	1928	1928	1928	1928	1926	1926	1926	1926
		.884	.886	.845	.849	.055	.058	.060	.060	.015	.017	.019	.018
	3	1929	1929	1929	1929	1928	1928	1928	1928	1926	1926	1926	1926
Employment		.901	.882	.830	.836	.044	.047	.051	.051	.012	.015	.018	.018
	9	1929	1929	1929	1932	1930	1930	1933	1929	1928	1928	1938	1933
		.993	.992	.603	.999	.004	.004	.262	.000	.000	.002	.110	.000
Consumer Prices	3	1894	1894	1894	1894	1893	1893	1893	1893	1897	1897	1897	1897
		.681	.675	.647	.648	.183	.195	.235	.235	.031	.028	.022	.022
	8	1929	1929	1929	1929	1928	1928	1928	1928	1927	1938	1938	1938
GNP Deflator		.449	.422	.208	.208	.167	.161	.108	.108	.04	.057	.063	.063
	4	1864	1864	1864	1864	1920	1920	1920	1920	1873	1867	1919	1919
		.999	.999	.997	.997	.000	.000	.001	.001	.000	.000	.000	.000
Nominal Wages	3	1864	1864	1864	1864	1863	1863	1863	1863	1867	1867	1867	1867
		.999	.999	.999	.999	.000	.000	.000	.000	.000	.000	.000	.000
	2	1920	1920	1920	1920	1929	1929	1928	1928	1928	1928	1929	1929
Velocity		.539	.578	.580	.567	.073	.061	.054	.058	.072	.061	.054	.058
	3	1920	1920	1920	1920	1929	1929	1929	1929	1928	1928	1928	1928
		.419	.472	.496	.486	.133	.111	.085	.089	.103	.088	.069	.071
Money Stock	6	1920	1920	1920	1920	1929	1929	1929	1929	1928	1928	1918	1918
		.488	.549	.626	.625	.179	.143	.045	.045	.082	.069	.036	.036
	3	1929	1929	1929	1929	1928	1928	1928	1928	1930	1930	1930	1930
Bond Yields		.405	.377	.314	.319	.234	.228	.203	.206	.114	.106	.092	.094
	8	1929	1929	1929	1929	1930	1930	1920	1920	1928	1928	1930	1930
		.626	.623	.243	.243	.281	.272	.124	.124	.044	.048	.098	.089
Stock prices	2	1946	1946	1946	1946	1949	1949	1949	1949	1947	1947	1880	1880
		.157	.153	.147	.150	.152	.136	.119	.123	.130	.124	.114	.116
	3	1880	1880	1880	1880	1946	1946	1946	1946	1949	1949	1947	1947
Money Stock		.237	.237	.215	.218	.133	.127	.117	.120	.119	.105	.083	.084
	3	1928	1928	1928	1928	1929	1929	1929	1929	1930	1930	1927	1930
		.140	.127	.095	.098	.103	.092	.068	.071	.062	.055	.043	.044
Bond Yields	7	1928	1928	1896	1896	1929	1929	1928	1928	1930	1930	1929	1929
		.154	.146	.102	.102	.141	.130	.054	.054	.135	.122	.044	.044
	3	1967	1967	1967	1918	1965	1920	1965	1935	1966	1921	1966	1920
Stock prices		.681	.262	.686	.163	.113	.178	.125	.089	.078	.174	.093	.084
	2	1953	1953	1953	1953	1954	1954	1954	1954	1952	1952	1952	1952
		.547	.528	.514	.515	.143	.139	.134	.135	.098	.099	.099	.099
Stock prices	3	1936	1936	1936	1936	1937	1937	1937	1937	1939	1953	1953	1953
		.102	.100	.091	.102	.101	.097	.091	.091	.090	.086	.088	.088

Notes: q denotes the prior probability of no structural change. Model (A) is used for all of the series except stock prices, for which Model (B) is used.

these series are .830, .208 and .314, respectively. The Bayes estimate of r for the money stock is 1928 (1929 had the second highest posterior mass) with posterior mass equal to .095. In addition, from the plots of the pmf's we see that most of the posterior mass is concentrated around $r = 1929$. These results indicate that a structural break most likely occurred near 1929. The results for the other series are mixed. There is no evidence of structural change occurring at 1929 for the consumer prices or bond yields series since the most likely break dates for these series occur at the sample endpoints. For the GNP deflator, structural change is most likely to have occurred in 1920 ($P(r=1920) = .496$) and $r = 1929$ has the second highest posterior mass ($P(r=1929) = .054$). There is no clearly dominant change date for the stock price series. The

Table 4

Posterior Probabilities of Stochastic and Deterministic Nonstationarity: $q = 0$
 Model (A): $y_t = \mu + \beta t + d_\mu DU(r)_t + \theta y_{t-1} + \Sigma_1^{k-1} \varphi_i \Delta y_{t-i} + \varepsilon_t$

Event		$P\{\rho \geq 1\}$				$P\{\beta \leq 0\}$				$P\{d_\mu \geq 0\}$			
Series	k	F	J	MJ ¹	MJ ³	F	J	MJ ¹	MJ ³	F	J	MJ ¹	MJ ³
Nominal GNP	3	.001	.004	.007	.043	.001	.004	.016	.041	.009	.014	.040	.036
		.000	.000	.000	.003	.000	.000	.000	.000	.000	.000	.000	.000
Employment	8	.000	.008	.097	.846	.005	.012	.305	.807	.191	.216	.272	.281
		.000	.000	.025	.763	.000	.000	.059	.663	.001	.001	.009	.009
Consumer Prices	3	.004	.012	.024	.114	.000	.000	.000	.004	.000	.000	.000	.000
		.101	.245	.259	.996	.036	.048	.097	.111	.364	.330	.284	.299
GNP Deflator	3	.002	.010	.020	.339	.001	.002	.010	.295	.050	.054	.081	.081
		.000	.000	.002	.217	.000	.000	.002	.253	.001	.001	.004	.004
Nominal Wages	3	.001	.009	.014	.203	.003	.010	.024	.224	.049	.066	.119	.116
		.000	.000	.001	.042	.000	.000	.001	.001	.000	.000	.002	.002
Velocity	3	.007	.130	.029	.277	.936	.878	.790	.750	.537	.517	.483	.485
		.078	.578	.101	.759	.625	.478	.362	.294	.356	.341	.325	.334
Money Stock	3	.003	.002	.038	.960	.009	.014	.051	.956	.343	.364	.417	.418
Bond Yields	3	.239	.871	.171	.977	.360	.302	.969	.518	.767	.509	.980	.248
		.377	.946	.314	.987	.005	.184	.163	.964	.024	.097	.000	.801

Notes: See the Notes for Table 1. q denotes the prior probability of no structural change. The second line of probabilities are conditional on $r = 1929$.

Table 5

Prior Probabilities of Stochastic and Deterministic Nonstationarity
 for Stock Prices: $k = 3, q = 0$

Model (B): $y_t = \mu + \beta t + d_\mu DU(r)_t + d_\beta DT(r)_t + \rho y_{t-1} + \Sigma_1^{k-1} \varphi_i \Delta y_{t-i} + \varepsilon_t$

Event	F	J	MJ ¹	MJ ³
$P\{\rho \geq 1\}$.000	.123	.001	.016
	.000	.000	.000	.000
$P\{\beta \leq 0\}$.002	.004	.005	.010
	.000	.000	.000	.000
$P\{d_\mu \geq 0\}$.404	.402	.412	.412
	.000	.000	.000	.000
$P\{d_\beta \leq 0\}$.018	.022	.033	.033
	.000	.000	.000	.000

Notes: See the Notes for Table 1. q denotes the prior probability of no structural change. The second line of probabilities are conditional on $r = 1929$.

Bayes estimate of r is 1936 with a posterior mass of .091 and $P(r=1929) = .07$. The posterior pmf of r for the velocity series is bimodal with one mode at 1880 and another at 1946 and there is very little posterior mass at 1929. These results suggest that Perron's conditional unit root testing results are most likely applicable to the nominal GNP, employment, nominal wage and money stock series. His results appear tenuous for the GNP deflator and stock price series.

Now consider the posterior pdf's of the trend determining parameters. For nominal GNP, since the posterior mass at $r = 1929$ is near .8 for all of the priors, the unconditional posterior pdf's (with $q = 0$) are almost identical to the pdf's conditional on $r = 1929$. In addition, the posteriors computed from the three sets of priors are very similar. The ρ -posteriors and β -posteriors give virtually no indication of stochastic nonstationarity and the d_μ -posteriors clearly indicate a drop in the level of the series at 1929. The trend stationarity structural change model appears to be quite plausible for this series unless one has a large prior probability of no structural change. Similar inferences can be made for the employment and nominal wage series based on the F and J posteriors. Since $k = 8$ for employment, the MJ-prior has a large influence on the shape of the posteriors, as is clearly indicated from the plots of the posteriors, and, consequently, the unit root model is given considerable posterior probability. The MJ-posteriors for the nominal wage series indicate more uncertainty about the true values of the trend parameters than the F or J posteriors but not enough to sway inferences toward the unit root model. Inferences concerning the stochastic structure of the money stock series, however, are not as clear cut. The unconditional F and J posteriors for ρ and β are close to the conditional distributions and they give negligible support for the unit root model. The unconditional MJ-posteriors for these parameters display much more uncertainty about the true values of ρ and β . Further, the posteriors are very sensitive to the value of ε . When $\varepsilon = .001$, we have $P(\rho \geq 1) = .038$ and $P(\beta \leq 0) = .051$ and when $\varepsilon = .05$ these probabilities become .960 and .956, respectively. More importantly, however, the unconditional posterior pdf's of d_μ for the three priors are much different from the conditional pdf's. The unconditional pdf's are centered near zero and indicate considerable uncertainty about the presence of structural change. For the three sets of posteriors we have $P(d_\mu > 0) > .34$. For the GNP deflator, the unconditional pdf's give substantial evidence for the TS SC model with a change occurring at 1920 instead of 1929. For $\varepsilon = .05$, however, the MJ-posteriors indicate that the unit root model is still plausible. Last, the posterior pdf's of the trend parameters for the stock price series indicate that the SC model with $r = 1929$ is not very plausible. The unconditional posterior pdf's of d_μ and d_β are quite different from the pdf's conditional on $r = 1929$.

In sum, our Bayesian analysis indicates that the trend stationary structural change model with $r = 1929$ is quite likely for nominal GNP and nominal wages if the prior probability of no structural change is not too large. Further, these inferences are robust across different priors and lag specifications. The TS SC inference for employment is fragile since it depends on a particular lag specification.

5. CONCLUSIONS

This paper expands on the methodology introduced in Phillips (1991a) and provides a comprehensive Bayesian posterior analysis of trend behavior in general autoregressive time series models based on a modified information matrix prior that incorporates the interactions between

short-run and long-run dynamics and which permits the researcher to limit the amount of stochastic nonstationarity allowed. Analytic posterior densities for all of the trend determining parameters are derived using Laplace approximation techniques. Simulations show that Bayesian methods based on our modified information matrix priors work well when there are transient system dynamics. Bayesian methods of trend determination in models that permit structural change or trend breaks are also presented.

These Bayesian techniques are applied to the Nelson-Plosser historical US macroeconomic data and to various stock price and dividend series. Our empirical results for the Nelson-Plosser series are generally in accord with those of Phillips (1991a) concerning the presence of stochastic trends. In addition, we find much more evidence of stochastic trends in the stock price and dividend series than is suggested by the results of DeJong and Whiteman (1989b). Our Bayesian analysis also shows evidence of trend breaks in some of the macroeconomic series with the breaks occurring around 1929, thereby providing some partial support to the conclusion reached by Perron (1989).

APPENDIX

PROOF OF PROPOSITION 1: The joint posterior is given by

$$\begin{aligned} p(\rho, \sigma, \delta | y, \iota_0) &\propto \sigma^{-3} |A_0^*(\rho, \varphi)|^{1/2} \Pi_1^k (1 + \sigma^{-2} \lambda_j^*(\rho, \delta))^{1/2} \sigma^{-T} \\ &\quad \times \exp\{-(1/2\sigma^2) \Sigma_1^T (y_t - \mu - \beta t - \gamma' x_{t-1})^2\} \\ &= \sigma^{-(T+3)} |A_0^*(\rho, \varphi)|^{1/2} \Pi_1^k (1 + \sigma^{-2} \lambda_j^*(\rho, \delta))^{1/2} \exp\{-(1/2\sigma^2)[m(\hat{u}) + (\rho - \hat{\rho})^2 m_V(y_{-1})]\} \\ &\quad \times \exp\{-(1/2\sigma^2)(\delta - \bar{\delta}(\rho))' V' V (\delta - \bar{\delta}(\rho))\}, \end{aligned} \quad (A1)$$

where the last line follows from the decomposition (15).

To determine the marginal posterior density of ρ we must integrate the joint posterior with respect to δ and σ :

$$p(\rho | y, \iota_0) \propto \iint p(\rho, \sigma, \delta | y, \iota_0) d\delta d\sigma. \quad (A2)$$

We first integrate (A1) with respect to δ to give

$$\begin{aligned} p(\rho, \sigma | y, \iota_0) &\propto \sigma^{-(T+3)} \exp\{-(1/2\sigma^2)[m(\hat{u}) + (\rho - \hat{\rho})^2 m_V(y_{-1})]\} \int |A_0^*(\rho, \varphi)|^{1/2} \\ &\quad + \Pi_1^k (1 + \sigma^{-2} \lambda_j^*(\rho, \delta))^{1/2} \exp\{-(1/2\sigma^2)(\delta - \bar{\delta}(\rho))' V' V (\delta - \bar{\delta}(\rho))\} d\delta. \end{aligned} \quad (A3)$$

Recognizing that the major contribution to the integral occurs about the vector of points $\delta' = \bar{\delta}(\rho)' = (\bar{\mu}(\rho), \bar{\beta}(\rho), \bar{\varphi}(\rho)')$, the Laplace approximation reduces the integral to

$$|V'V|^{-1/2} \sigma^{k+1} |A_0^*(\rho, \bar{\varphi}(\rho))|^{1/2} \Pi_1^k (1 + \sigma^{-2} \lambda_j^*(\rho, \bar{\delta}(\rho)))^{1/2}. \quad (A4)$$

Since the elements of $V'V$ are at least $O(T)$, this approximation has a relative error of $O(T^{-1})$.

Combining (A3) with (A4) gives

$$\begin{aligned} p(\rho, \sigma | y, \iota_0) &\propto |A_0^*(\rho, \bar{\varphi}(\rho))|^{1/2} \Pi_1^k (1 + \sigma^{-2} \lambda_j^*(\rho, \bar{\delta}(\rho)))^{1/2} \\ &\quad \times \sigma^{-(T+2-k)} \exp\{-(1/2\sigma^2)[m(\hat{u}) + (\rho - \hat{\rho})^2 m_V(y_{-1})]\}. \end{aligned} \quad (A5)$$

The marginal density of ρ is then given by

$$p(\rho | y, y_0) \propto |A_0^*(\rho, \tilde{\varphi}(\rho))|^{1/2} \int \Pi_1^k(1 + \sigma^{-2} \lambda_j^*(\rho, \tilde{\delta}(\rho)))^{1/2} \sigma^{-(T+2-k)} \\ \times \exp\{-(1/2\sigma^2)[m(0) + (\rho - \hat{\rho})^2 m_V(y_{-1})]\} d\sigma. \quad (A6)$$

Making the change of variables $z = \sigma^{-2}$ in the above integral gives

$$p(\rho | y, y_0) \propto |A_0^*(\rho, \tilde{\varphi}(\rho))|^{1/2} \int \Pi_1^k(1 + z \lambda_j^*(\rho, \tilde{\delta}(\rho)))^{1/2} z^{g-(3/2)} \\ \times \exp\{-(z/2)[m(0) + (\rho - \hat{\rho})^2 m_V(y_{-1})]\} dz, \quad (A7)$$

where $g = (T+2-k)/2$. The integral in (A7) may be expressed as a confluent hypergeometric function of the second kind with multiple arguments (see Phillips (1988)). If $[m(0) + (\rho - \hat{\rho})^2 m_V(y_{-1})]$ is large, then most of the value of the integral comes from a neighborhood of z near the origin. In this case we may approximate $\Pi_1^k(1 + z \lambda_j^*(\rho, \tilde{\delta}(\rho)))^{1/2}$ by 1 and (A7) becomes

$$p(\rho | y, y_0) \propto |A_0^*(\rho, \tilde{\varphi}(\rho))|^{1/2} \int z^{g-(3/2)} \exp\{-(z/2)[m(0) + (\rho - \hat{\rho})^2 m_V(y_{-1})]\} dz. \quad (A8)$$

If we let $w = z[m(0) + (\rho - \hat{\rho})^2 m_V(y_{-1})]$, then using properties of the gamma function and ignoring terms that do not depend on ρ leaves us with

$$p(\rho | y, y_0) \propto |A_0^*(\rho, \tilde{\varphi}(\rho))|^{1/2} [m(0) + (\rho - \hat{\rho})^2 m_V(y_{-1})]^{-(T-k+1)/2}. \quad (A9)$$

This completes the proof for part (a).

The proofs for parts (b) and (c) follow in a similar fashion using the sum of squares decompositions (16) and (17).

PROOF OF THEOREM 2: Part (a). Using the sum of squares decomposition (37), the joint posterior is given by

$$p(\xi, r | \varepsilon, y, y_0) \propto \sigma^{-(T+1)} (\Pi_{1a}^k a_{0ii}^*(\rho, \varphi))^{1/2} \exp\{-\rho^{2c_k(\varepsilon)}\} \\ \times \exp\{-(1/2\sigma^{-2})[m(0(r)) + (\xi - \hat{\xi}(r))' X(r)' X(r) (\xi - \hat{\xi}(r))]\}. \quad (A10)$$

To determine the marginal posterior mass functions of r , $r = 2, \dots, T-2$, we must integrate the joint posteriors with respect to ξ and σ :

$$p(r | \varepsilon, y, y_0) \propto \int \int p(\xi, \sigma, r | y, y_0) d\xi d\sigma, \quad r = 2, \dots, T-2. \quad (A11)$$

We first integrate (A10) with respect to ξ to give

$$p(\sigma, r | \varepsilon, y, y_0) \propto \sigma^{-(T+1)} \exp\{-m(0(r))/2\sigma^{-2}\} \int (\Pi_{1a}^k a_{0ii}^*(\rho, \varphi))^{1/2} \exp\{-\rho^{2c_k(\varepsilon)}\} \\ \times \exp\{-(1/2\sigma^{-2})(\xi - \hat{\xi}(r))' X(r)' X(r) (\xi - \hat{\xi}(r))\} d\xi.$$

Recognizing that the major contribution to the integral occurs about $\xi' = \hat{\xi}(r)' = (\hat{\rho}(r), \hat{\mu}(r), \hat{\beta}(r), \hat{\varphi}(r)')$, the Laplace approximation yields

$$p(\sigma, r | \varepsilon, y, y_0) \propto |X(r)' X(r)|^{-1/2} \sigma^{-(T-k-4)} (\Pi_{1a}^k a_{0ii}^*(\hat{\rho}(r), \hat{\varphi}(r)))^{1/2} \\ \times \exp\{-\hat{\rho}(r)^{2c_k(\varepsilon)}\} \exp\{-m(0(r))/2\sigma^{-2}\},$$

which has a relative error of $O(T^{-1})$ since the elements of $X(r)' X(r)$ are at least $O(T)$. Next we integrate with respect to σ , and find

$$p(r | \varepsilon, y, y_0) \propto |X(r)' X(r)|^{-1/2} (\Pi_{1a}^k a_{0ii}^*(\hat{\rho}(r), \hat{\varphi}(r)))^{1/2} \exp\{-\hat{\rho}(r)^{2c_k(\varepsilon)}\} \\ \times \int \sigma^{-(T-k-4)} \exp\{-m(0(r))/2\sigma^{-2}\} d\sigma. \quad (A12)$$

Making the change of variables $z = m(u(r))/2\sigma^{-2}$ and using the Gamma integral, we have

$$p(r|\varepsilon, y, u_0) \propto |X(r)'X(r)|^{-1/2} (\Pi_{1a}^k \circ_{ii}(\hat{\rho}(r), \hat{\varphi}(r)))^{1/2} \exp\{-\hat{\rho}(r)^{2c_k(\varepsilon)}\} m(u(r))^{-(T-k-4)/2}, \quad (A13)$$

$r = 2, \dots, T-2$. This proves part (a).

Consider now the determination of the posterior density of ρ . Using the sum of squares decomposition (23), we may rewrite the joint posterior as

$$p(\rho, \sigma, \delta, r|\varepsilon, y, u_0) \propto \sigma^{-(T+1)} (\Pi_{1a}^k \circ_{ii}(\rho, \varphi))^{1/2} \exp\{-\rho^{2c_k(\varepsilon)}\} \\ \times \exp\{-(1/2\sigma^{-2})[m(u(r)) + (\rho - \hat{\rho}(r))^2 m_{V(r)}(y_{-1}) + (\delta - \hat{\delta}(r))' V(r)' V(r) (\delta - \hat{\delta}(r))]\}. \quad (A14)$$

We proceed as before and start by integrating (A14) with respect to δ to give

$$p(\rho, \sigma, r|\varepsilon, y, u_0) \propto \sigma^{-(T+1)} \exp\{-\rho^{2c_k(\varepsilon)}\} \exp\{-(1/2\sigma^{-2})[m(u(r)) + (\rho - \hat{\rho}(r))^2 m_{V(r)}(y_{-1})]\} \\ \times \int (\Pi_{1a}^k \circ_{ii}(\rho, \varphi))^{1/2} \exp\{-(1/2\sigma^{-2})(\delta - \hat{\delta}(r))' V(r)' V(r) (\delta - \hat{\delta}(r))\}. \quad (A15)$$

Using the Laplace approximation, and ignoring terms that do not depend on ρ , σ or r , the integral that appears in (A15), may be reduced to

$$|V(r)'V(r)|^{-1/2} \sigma^{k+4} (\Pi_{1a}^k \circ_{ii}(\rho, \hat{\varphi}(r)))^{1/2}, \quad (A16)$$

which has a relative error of $O(T^{-1})$. Combining (A15) with (A16) gives

$$p(\rho, \sigma, r|\varepsilon, y, u_0) \propto |V(r)'V(r)|^{-1/2} \sigma^{-(T-k-3)} (\Pi_{1a}^k \circ_{ii}(\rho, \hat{\varphi}(r)))^{1/2} \exp\{-\rho^{2c_k(\varepsilon)}\} \\ \times \exp\{-(1/2\sigma^{-2})[m(u(r)) + (\rho - \hat{\rho}(r))^2 m_{V(r)}(y_{-1})]\}. \quad (A17)$$

Integration with respect to σ , ignoring terms that do not depend on ρ , then yields

$$p(\rho|\varepsilon, r, y, u_0) \propto (\Pi_{1a}^k \circ_{ii}(\rho, \hat{\varphi}(r)))^{1/2} \exp\{-\rho^{2c_k(\varepsilon)}\} \\ \times [m(u(r)) + (\rho - \hat{\rho}(r))^2 m_{V(r)}(y_{-1})]^{-(T-k-3)/2}. \quad (A18)$$

Since $p(\rho, r|\varepsilon, y, u_0) = p(\rho|\varepsilon, r, y, u_0)p(r|\varepsilon, y, u_0)$, averaging (A18) with respect to (A13) thus gives the marginal posterior density function of ρ . This proves part (b).

The proofs for parts (c)-(e) are straightforward extensions of the above reasoning based on the decompositions (34)-(36). They are therefore omitted.

ACKNOWLEDGMENTS

The first author would like to thank Paula Ann Shamoian for her help with the graphs.

REFERENCES

- Banerjee, A., R.L. Lumsdaine and J.H. Stock (1990): "Recursive and sequential tests for a unit root: Theory and international evidence," unpublished manuscript, Kennedy School of Government, Harvard University, Cambridge, MA.
- Berger, J.O. (1985): *Statistical decision theory and Bayesian analysis*, Second Edition, Springer-Verlag, New York.
- Booth, N.B. and A.F.M. Smith (1982): "A Bayesian approach to retrospective identification of change points," *Journal of Econometrics* 19, 7-22.
- Broemeling, L.D. (1985): *Bayesian analysis of linear models*, Marcel Dekker, New York.

- Broemeling, L.D. and H. Tsurumi (1987): *Econometrics and structural change*, Marcel Dekker, New York.
- Chin Choy, H.H. and L.D. Broemeling (1980): "Some Bayesian inferences for a changing linear model," *Technometrics* 22, 71-78.
- Christiano, L.J. (1988): "Searching for a break in GNP," Working Paper No. 2695, National Bureau of Economic Research, Cambridge, MA.
- DeJong, D.N., J.C. Nankervis, N.E. Savin and C.H. Whiteman (1988): "Integration versus trend-stationarity in macroeconomic time series, Working Paper Series No. 88-27a, Department of Economics, University of Iowa, Iowa City, IA.
- DeJong, D.N. and C.H. Whiteman (1989a): "Trends and random walks in macroeconomic time series: A reconsideration based on the likelihood principle," forthcoming in the *Journal of Monetary Economics*.
- _____ (1989b): "The temporal stability of dividends and stock prices: Evidence from the likelihood function," forthcoming in the *American Economic Review*.
- _____ (1989c): "Trends and cycles as unobserved components in real GNP: A Bayesian perspective," forthcoming in the *Proceedings of the American Statistical Association*.
- Dickey, D.A. and W.A. Fuller (1979): "Distribution of the estimators for autoregressive time series with a unit root," *Journal of the American Statistical Association*, 74, 427-431.
- _____ (1981): "Likelihood ratio tests for autoregressive time series with a unit root," *Econometrica*, 49, 1057-1072.
- Ferreira, P.E. (1977): "A Bayesian analysis of a switching regression model: Known number of regimes," *Journal of the American Statistical Association* 70, 370-374.
- Geweke, J. (1988): "The secular and cyclical behavior of real GDP in nineteen OECD countries, 1957-1983," *Journal of Business and Economic Statistics*, 6, 479-486.
- _____ (1989): "Bayesian inference in econometric models using Monte Carlo integration," *Econometrica*, 57, 1317-39.
- Hannan, E.J. (1970): *Multiple time series*, John Wiley, New York.
- Hansen, B.E. (1990): Testing for structural change of unknown form in models with non-stationary regressors," unpublished manuscript, Department of Economics, University of Rochester, Rochester, NY.
- Holbert, D. (1982): "A Bayesian analysis of a switching linear model," *Journal of Econometrics* 19, 77-87.
- Judge, G.G., W.E. Griffiths, R.C. Hill, H. Lütkepohl and T. Lee (1985): *The theory and practice of econometrics*, John Wiley, New York.
- Kloek, T. and H. van Dijk (1978): "Bayesian estimates of equation system parameters: An application of integration by Monte Carlo," *Econometrica*, 50, 725-736.
- Leamer, E.E. (1978): *Specification searches: Ad hoc inferences with nonexperimental data*, John Wiley & Sons, New York.
- _____ (1984): "Let's take the con out of econometrics," *American Economic Review*, 73, 31-43.

- Marsh, T.A. and R.C. Merton (1986): "Dividend variability and variance bounds tests for the rationality of stock market prices," *American Economic Review*, 76, 483-494.
- Nelson, C.R. and C.I. Plosser (1982): "Trends versus random walks in macroeconomic time series: Some evidence and implications," *Journal of Monetary Economics*, 10, 139-162.
- Perron, P. (1988): "Trends and random walks in macroeconomic time series: Further evidence from a new approach," *Journal of Economic Dynamics and Control*, 12, 297-332.
- (1989): "The great crash, the oil price shock and the unit root hypothesis," *Econometrica* 57, 1361-1401.
- (1990): "Testing for a unit root in a time series with a changing mean," *Journal of Business and Economic Statistics*, Vol. 8, No. 2, 153-162.
- Phillips, P.C.B. (1983): "Time series regression with a unit root," *Econometrica*, 55, 277-301.
- (1988): "The characteristic function of the Dirichlet and multivariate F distributions," Cowles Foundation Discussion Paper No. 865, Yale University, New Haven, CT.
- (1991a): "To criticize the critics: An objective Bayesian analysis of stochastic trends," forthcoming in *Journal of Applied Econometrics*, 6, 333-364.
- (1991b): "Bayesian routes and unit roots: de rebus prioribus semper est disputandum," forthcoming in *Journal of Applied Econometrics*, 6, 435-474.
- Phillips, P.C.B. and P. Perron (1988): "Testing for a unit root in time series regression," *Biometrika*, 75, 335-346.
- Said, S.E. and D.A. Dickey (1984): "Testing for unit roots in ARMA models of unknown order," *Biometrika*, 71, 599-607.
- Schotman, P. and H.K. van Dijk (1990): "A Bayesian analysis of the unit root in real exchange rates," forthcoming in the *Journal of Econometrics*.
- Schwert, G.W. (1987): "Effects of model misspecification on tests for unit roots in macroeconomic data," *Journal of Monetary Economics*, 20, 73-103.
- Shiller, R.J. (1981): "Do stock prices move too much to be justified by subsequent changes in dividends?," *American Economic Review*, 71, 421-436.
- Sims, C.A. (1988): "Bayesian skepticism on unit root econometrics," *Journal of Economic Dynamics and Control* 12, 463-474.
- Sims, C.A. and H. Uhlig (1989): "Understanding unit rooters: A helicopter tour," forthcoming in *Econometrica*.
- Smith, A.F.M. (1980): "Change point problems: Approaches and applications," in J.M. Bernardo et al., eds., *Bayesian Statistics I*, Valencia, University Press, Valencia, 83-98.
- Tierney, L. and J.B. Kadane (1986): "Accurate approximations for posterior moments and marginal densities," *Journal of the American Statistical Association*, 81, 82-86.
- Zellner, A. (1971): *An introduction to Bayesian inference in econometrics*, John Wiley & Sons, New York.
- Zivot, E. and D.W.K. Andrews (1990): "Further evidence on the Great Crash, the oil price shock and the unit root hypothesis," forthcoming in the *Journal of Business and Economic Statistics*.

- Zivot, E. and P.C.B. Phillips (1991): "A Bayesian analysis of trend determination in economic time series," Cowles Foundation Discussion Paper No. 1002, Yale University, New Haven, CT.
- Zivot, E. (1992): "A Bayesian analysis of the unit root hypothesis within an unobserved components model," unpublished manuscript, Department of Economics, Wellesley College, Wellesley, MA.