Working Paper 93-13
Statistics and Econometrics Series 11
May 1993

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (341) 624-9849

A DECISION THEORETIC ANALYSIS OF THE UNIT ROOT HYPOTHESIS USING MIXTURES OF ELLIPTICAL MODELS

Gary Koop and Mark F.J. Steel*

Abstract

This paper develops a formal decision theoretic approach to testing for a unit root in economic time series. The approach is empirically implemented by specifying a loss function based on predictive variances; models are chosen so as to minimize expected loss. In addition, the paper broadens the class of likelihood functions traditionally considered in the Bayesian unit root literature by: i) Allowing for departures from normality via the specification of a likelihood based on general elliptical densities; ii) allowing for structural breaks to occur; iii) allowing for moving average errors; and iv) using mixtures of various submodels to create a very flexible overall likelihood. Empirical results indicate that, while the posterior probability of trend-stationarity is quite high for most of the series considered, the unit root model is often selected in the decision theoretic analysis.

Key Words
Bayesian; Monte Carlo Integration; Loss Function; Prediction

*Koop, Department of Economics, University of Toronto; Steel, Department of Statistics and Econometrics, Universidad Carlos III de Madrid. The first author enjoyed the hospitality of CentER, Tilburg University. The second author gratefully acknowledges financial support from the "Cátedra Argentaria" at the Universidad Carlos III de Madrid. We thank Jacek Osiewalski, two referees and an associate editor for many helpful suggestions as well as Herman van Dijk, who kindly provided the data used in this study.

The economic literature devoted to the issue of unit roots in economic time series has grown immensely since the seminal papers of Dickey and Fuller (1979) and Nelson and Plosser (1982). Although the majority of the literature assumes a classical econometric perspective, a growing Bayesian unit root literature has emerged (see DeJong and Whiteman (1991a,b), Phillips (1991), Sims (1988), Koop (1991, 1992), Schotman and van Dijk (1991a,b), Wago and Tsurumi (1990), Zivot and Phillips (1991)). In many cases, Bayesian results differ substantially from their classical counterparts.

This paper makes a contribution to this growing body of Bayesian unit root literature. It considers more general classes of models and methods of drawing inferences than are presently used. The paper uses models that are mixtures over various submodels with general elliptical distributions that differ in both their covariance structure and their treatment of structural breaks. The resulting mixed model is very flexible and encompasses a wide variety of dynamic structures. In addition, the paper uses a formal decision theoretic framework based on predictive means and variances and the conservative notion that it is worse to underestimate than to overestimate predictive variances. This approach accords naturally with a Bayesian paradigm and provides an explicit forum for choosing between stationary, unit root, and explosive models.

Our approach is applied to the fourteen Nelson-Plosser data series (see Data Appendix). Results indicate that, although the probability of trend-stationarity is high for most series, our decision analysis often selects the unit root model.

Section 1 of the paper introduces our hypothesis of interest and the methodology we use to test it. Section 2 discusses the sampling model, Section 3 the prior density, and Section 4 the posterior density. Section 5 treats the decision problem while Section 6 applies the methods to the extended Nelson-Plosser data set. Section 7 concludes.

Section 1: What Are We Testing?

Our aim is to determine whether a unit root is present, ie. to test an exact restriction. One obvious way is to calculate posterior odds comparing the model with a unit root imposed against the unrestricted model. This method requires that an informative (proper) prior be placed over ρ, the coefficient which equals one under a unit root. Koop (1991, 1992) calculates posterior odds using informative natural conjugate priors. Schotman and van Dijk (1991a) use proper priors that require less subjective prior input but at the cost that their priors are data-based.

Rather than test explicitly for a unit root, an alternative methodology (DeJong and Whiteman (1991a,b) and Phillips (1991)) is to calculate the posterior probability that ρ is in some region near 1. This method has the advantage that proper priors are no longer necessary and thus the analysis may be made more "objective". The disadvantage is that the definition of ρ as "close to one" is highly subjective. By way of example, consider calculating the probability that $|\rho| \geq .975$ and $|\rho| \geq 1$. The former is highly subjective whereas the latter is suitable for testing for nonstationarities (ie. unit root or explosive behavior) but not for the presence of a unit root *per se*. In this paper, we use proper priors on ρ which allow us to compute posterior odds for the exact unit root null.

Furthermore we use a decision theoretic framework to carry out the unit root tests. The loss function used in the decision analysis is based on predictive behavior which can differ crucially for stationary (H_1: $|\rho| < 1$), unit root (H_2: $\rho = 1$), and explosive (H_3: $|\rho| > 1$) models. Our decision problem will be set up with respect to these three regions for ρ.

Section 2: The Likelihood Function

Bayesian methods require the specification of a likelihood function. Phillips (1991), for example, bases his likelihood function on the following specification:

$$y_t = \mu + \beta t + \rho y_{t-1} + \sum_{i=1}^{k-1} \phi_i \Delta y_{t-i} + \varepsilon_t \qquad (1)$$

with ε_t i.i.d. $N(0,\tau^{-2})$, while DeJong and Whiteman (1991a) use a different parameterization. Since analytical results cannot always be obtained for their parameterization and extensive Monte Carlo integration is required, Phillips' parameterization is preferred (see Phillips (1991)). In the empirical section we use Phillips' parameterization and set k=3.

We expand the class of considered likelihoods by relaxing i) the normality and ii) the i.i.d. assumption; and iii) by allowing for structural breaks (Perron (1989), Banerjee, Lumsdaine and Stock (1990) and Zivot and Phillips (1990)). We let y (where $y=(y_1,\ldots,y_T)'$) have any density within the class of multivariate elliptical densities, and thereby cover such densities as the multivariate normal, multivariate-t and Pearson type-II. Moreover, we allow the covariance matrix to take the form $\tau^{-2}V(\eta)$, where $V(\eta)$ is any positive definite symmetric matrix parameterized by a finite vector η. Techniques for handling extensions i) and ii) are described in Osiewalski and Steel (1993a,b).

In this paper, no single model need be selected for final analysis. Several different structural breaks and structures for $V(\eta)$ can be chosen, and a supermodel, which is a finite mixture of the various submodels, used. We allow $V(\eta)$ to have various structures in the ARMA class. The motivation behind the inclusion of a moving average component is discussed in Schwert (1987).

Formalizing the ideas described in the preceding paragraphs, we begin with the model with no structural breaks (M_N). We mix over m different correlation structures so that each individual model is labelled M_{Ni} ($i=1,\ldots,m$). For each model M_{Ni} we take:

$$P(y \mid \theta_N, \tau^2, \eta, y_{(0)}, M_{Ni}) = \mid \tau^{-2} V_{Ni}(\eta) \mid^{-1/2} \\ g_{Ni}[(y-h_{Ni}(\theta_N))' \tau^2 V_{Ni}^{-1}(\eta)(y-h_{Ni}(\theta_N))] \tag{2}$$

where $g_{Ni}(.)$ is a nonnegative function which satisfies (for all i and T),

$$\int_0^\infty u^{\frac{T}{2}-1} g_{Ni}(u)\, du = \Gamma(T/2)\, \pi^{-T/2}. \tag{3}$$

In other words, we assume y has a T-variate elliptical density (see, for instance, Dickey and Chen (1985) or Fang, Kotz and Ng (1990)). Note that (3) is a necessary and sufficient condition for (2) to be a proper density, $y_{(0)}$ is the vector of initial observations $(y_{1-k}, \ldots, y_0)'$, and $h_{Ni}(\theta_N)$ is a vector of length T with property:

$$[h_{Ni}(\theta_N)]_t = \mu + \beta t + \rho y_{t-1} + \sum_{j=1}^{k-1} \phi_j \Delta y_{t-j}.$$

Since we assume this function to be identical for all covariance structures, we drop the i subscript and write:

$$h_N(\theta_N) = \rho y_{-1} + X_N \alpha_N \tag{4}$$

where

$$y_{-1} = (y_0, y_1, \ldots, y_{T-1})',$$
$$X_N = (X_1^N, \ldots, X_T^N)',$$
$$X_t^N = (1, t, \Delta y_{t-1}, \ldots, \Delta y_{t-k+1})',$$
$$\alpha_N = (\mu, \beta, \phi_1, \ldots, \phi_{k-1})' = (\mu, \beta, \phi')',$$

and hence $\theta_N = (\rho, \alpha_N')'$. For future reference we define:

$$d_{Ni}(\theta_N,\eta) = (y-h_N(\theta_N))' V_{Ni}^{-1}(\eta) (y-h_N(\theta_N)) . \tag{5}$$

The model without structural breaks (M_N) is then given by the mixture of the probabilities in (2) over the m covariance structures:

$$\begin{aligned} &P(y \mid \theta_N,\tau^2,\eta,\delta,y_{(0)},M_N) \\ &=\sum_{i=1}^{m} \delta_i\, P(y \mid \theta_N,\tau^2,\eta,y_{(0)},M_{Ni}) , \end{aligned} \tag{6}$$

where $\delta = (\delta_1,\ldots,\delta_m)'$ is a vector of mixing parameters with $\delta_i \geq 0$ and $\sum \delta_i = 1$.

We obtain the model with structural breaks (M_S) by mixing over various covariance structures (j=1,...,m) and breakpoints (q=1,...,T-1). Note that we use the same covariance structures as in the previous model. Although not necessary, doing so simplifies the notation such that $V_{Ni}=V_{Sj}=V_j$ for i=j. Moreover, we conceptually allow for the structural breaks to occur at any point in our sample. Two types of structural breaks, level breaks and trend breaks, can occur at any time q=1,...,T-1. Perron (1989) argues for the presence of a level break in 1929 and a trend break in 1973 for most U.S. macroeconomic time series. To reduce the burden of computation only the latter two breaks are included in the empirical analysis although the general notation is retained throughout this section.

Note that we could compute the posterior probability of a model with a particular break at a particular time against any other structural break model (or the model without structural breaks). Thus, we could let the data decide which breakpoints should be favored. However, in practice, we limit ourselves to those breakpoints which we consider reasonable from an economic point of view (thus substantially reducing computational effort).

M_S is a mixture over models with different structural breakpoints and covariance structures (M_{Sjq}). Note that each of these submodels has the likelihood function:

$$\begin{aligned} &P(y \mid \theta_S,\tau^2,\eta,y_{(0)},M_{Sjq}) = \mid \tau^{-2}V_j(\eta) \mid^{-\frac{1}{2}} \\ &g_{Sjq}[(y-h_{Sjq}(\theta_S))'\tau^2 V_j^{-1}(\eta) (y-h_{Sjq}(\theta_S))] \end{aligned} \tag{7}$$

where $g_{Sjq}(.)$ satisfies (3) for all j, q and T, and

$$h_{Sjq}(\theta_S) = h_{Sq}(\theta_S) = \rho y_{-1} + X_N\alpha_N + X_{Dq}\alpha_D = \rho y_{-1} + X_{Sq}\alpha_S .$$

We define $\quad X_{Dq} = (X_1^{Dq},\ldots,X_T^{Dq})'$

and $X_t^{Dq} = (DU(q)_t, DT(q)_t)'$,

where $DU(q)_t = 1$ if $t>q$ and 0 otherwise (level break)

and $DT(q)_t = t-q$ if $t>q$ and 0 otherwise (trend break)

Furthermore $\alpha_D = (d_\mu, d_\beta)'$ and hence

$$X_{Sq} = (X_N \; X_{Dq}) \quad \text{and} \quad \alpha_S = (\alpha_N \; \alpha_D)'.$$

In this setup $\theta_s = (\rho \; \alpha_S')' = (\rho \; \alpha_N' \; \alpha_D')' = (\theta_N' \; \alpha_D')'$ and the structural break models have two parameters more than those lacking structural breaks. In our empirical setup we restrict d_μ to be zero for 1973 and take d_β to be zero for 1929, leaving just one parameter in α_D for each of the structural break models. For future reference we define

$$d_{Sjq}(\theta_S, \eta) = (y - h_{Sq}(\theta_S))' \, V_j^{-1}(\eta) \, (y - h_{Sq}(\theta_S)) \, . \tag{8}$$

The overall model (M_S), mixed over structural breaks and covariance structures, is:

$$\begin{aligned} &P(y \mid \theta_S, \tau^2, \eta, \gamma, \kappa, y_{(0)}, M_S) \\ &= \sum_{j=1}^{m} \gamma_j \sum_{q=1}^{T-1} \kappa_q \; P(y \mid \theta_S, \tau^2, \eta, y_{(0)}, M_{Sjq}) \end{aligned} \tag{9}$$

where $\gamma = (\gamma_1, \ldots, \gamma_m)'$ and $\kappa = (\kappa_1, \ldots, \kappa_{T-1})'$ are mixing parameters with γ_j, $\kappa_q \geq 0 \;\forall$ j, q and $\sum \gamma_j = \sum \kappa_q = 1$.

Finally, we mix over the no-structural-break and structural-break models to obtain the sampling model

$$\begin{aligned} P(y \mid \theta_S, \tau^2, \eta, \lambda, \delta, \gamma, \kappa, y_{(0)}) &= \lambda \sum_{i=1}^{m} \delta_i \; P(y \mid \theta_N, \tau^2, \eta, y_{(0)}, M_{Ni}) \\ &+ (1-\lambda) \sum_{j=1}^{m} \gamma_j \sum_{q=1}^{T-1} \kappa_q \; P(y \mid \theta_S, \tau^2, \eta, y_{(0)}, M_{Sjq}) \end{aligned} \tag{10}$$

with $0 \leq \lambda \leq 1$.

To summarize: (10) is the overall sampling model to be used in this paper. It mixes over two models, one with and one without structural breaks. We weight the model with no structural breaks over covariance structures (see (6)) and the model with structural breaks over covariance structures and structural breakpoints (see (9)). Each of the mT submodels in (10) can have a different type of elliptical density. Not only do our likelihoods allow for normal, Cauchy and Student t densities, but for densities with truncated tails (eg. Pearson type-II densities) as well.

It remains to specify the choices for $V_j(\eta)$. Since most, if not all the residual autocorrelation will be removed by including the lagged Δy_ts in the model, V_j is restricted to two choices (m=2): $V_1 = I_T$ and $V_2 = (1 + \eta^2)I_T - \eta A$, where $\eta \in (-1,1)$ and A is a tridiagonal matrix with 2's on the diagonal and -1's on the off-diagonal. In other words, we allow the errors to be uncorrelated (which, for only the normal distribution, implies independence) under V_1 and to exhibit MA(1) behavior under V_2. Choi (1990) argues that ignoring the MA(1) component of the errors results in a bias in classical estimates of ρ equal to $\eta(1-\rho)/(1+\eta)$ for infinite k which tends to drive results towards the unit root for $\eta > 0$.

Section 3: The Prior Density

A controversy surrounding the use of Bayesian methods is the role of prior information. Many researchers use priors that are noninformative or objective in order to avoid the issue (see DeJong and Whiteman (1991a,b), Koop (1992) and Phillips (1991)). Koop and Steel (1991) discuss the hazards involved in the use of such "objective" priors. Moreover, improper noninformative priors make it impossible to calculate posterior odds required to test for unit roots (see Section 1). For the reasons noted, noninformative priors for ρ are not used in this paper.

An alternative, following Schotman and van Dijk (1991a) and Koop (1991), is to introduce explicit prior information into the analysis. Schotman and van Dijk minimize the amount of subjective prior information by allowing the prior to depend on the data, an approach which violates the likelihood principle and thus is avoided here. Koop (1991) uses natural conjugate priors centered over the unit root restriction and performs a sensitivity analysis with respect to the prior covariance matrix of the regression parameters. In this paper a prior is used which is uniform in $\log(\tau^{-2})$ and in the regression parameters other than ρ. As well as being improper, the prior is noninformative in certain dimensions in that the posterior is proportional to the likelihood function. However, before posterior odds can be calculated the prior must be made proper in the remaining dimensions by bounding it, for example. A sensitivity analysis can easily be performed over the choice of bounding region. These steps are formalized in the remainder of this section.

The prior density for the parameters of the sampling model can be written as:

$$P(\theta_S,\tau^2,\eta,\lambda,\delta,\gamma,\kappa) = P(\theta_S,\tau^2,\eta)P(\lambda)P(\delta)P(\gamma)P(\kappa) \tag{11}$$

That is, we *a priori* assume the mixing parameters to be independent of each other and of the parameters in each submodel. Since the mixing parameters are of no interest to us, we need only specify prior means whose existence is assumed (see Osiewalski and Steel (1993b)). In order to be as noninformative as possible, all models receive equal prior weight. Specifically, we set

$$E(\lambda)=1/3;\ E(\delta_1)=E(\delta_2)=E(\gamma_1)=E(\gamma_2)=1/2 \text{ and } E(\kappa_q)=1/2 \text{ for } q=1,2.$$

Full robustness with respect to the choices for $g_{Sjq}(.)$ and $g_{Ni}(.)$ is achieved by assuming (see Osiewalski and Steel (1993a)):

$$P(\theta_S,\tau^2,\eta) = c_1\tau^{-2}P(\theta_S,\eta).$$

This assumption implies a uniform prior for $\log(\tau^{-2})$. Note that c_1 is a constant which cancels out of the posterior odds ratio and hence is irrelevant for our analysis. All that remains is to specify $P(\theta_S,\eta)$:

$$P(\theta_S,\eta) = P(\theta_N,\alpha_D,\eta) = P(\theta_N \mid \alpha_D,\eta)P(\alpha_D,\eta).$$

Since the parameters α_D and η are not present in all models, we must ensure that $P(\alpha_D,\eta)$ is proper. For the sake of convenience we assume that $P(\alpha_D,\eta)=P(\alpha_D)P(\eta)=P(d_\mu)P(d_\beta)P(\eta)$ and specify:

$$\begin{aligned} P(d_\mu) &= 1/(A_2-A_1) \text{ on } [A_1,A_2] \text{ and } 0 \text{ elsewhere} \\ P(d_\beta) &= 1/(B_2-B_1) \text{ on } [B_1,B_2] \text{ and } 0 \text{ elsewhere} \\ P(\eta) &= 1/2 \text{ on } (-1,1) \text{ and } 0 \text{ elsewhere}. \end{aligned} \tag{12}$$

In practice, $[A_1,A_2]$ and $[B_1,B_2]$ are chosen so as to cover the area where the likelihood function is *a priori* assumed to be appreciable (see Prior Appendix). Finally, it remains to specify

$$P(\theta_N \mid \alpha_D,\eta) = P(\alpha_N \mid \rho,\alpha_D,\eta)P(\rho \mid \alpha_D,\eta).$$

Since the parameters α_N are present in all models we allow them to have an unbounded uniform prior.

We assume that the parameter of interest, ρ, is independent of α_D and η. Under the hypothesis that a unit root is present (H_2) we set $\rho=1$. Under the hypothesis that a unit root is not present we try two priors for ρ. Our first choice is a bounded uniform prior which, for the stationary region (H_1), takes the form:

$$\begin{aligned} P(\rho) &= 2\tfrac{2}{9} \quad \text{if } \rho \in [.55, 1.00) \\ &= 0 \quad \text{otherwise}, \end{aligned}$$

and for the explosive region (H_3):

$$\begin{aligned} P(\rho) &= 10 \quad \text{if } \rho \in (1.00, 1.10] \\ &= 0 \quad \text{otherwise}. \end{aligned}$$

This type of bounded uniform prior leads to a truncated Student t posterior for ρ under V_1 and for $\rho|\eta$ under V_2. Alternatively, an independent Student t prior for ρ with first two moments identical to our uniform can be used to yield a 2-0 poly-t posterior density for ρ (or $\rho|\eta$) (Dreze (1977)). Note that pseudo-random drawings to be used in the Monte Carlo integration can easily be made from all these densities (Richard and Tompa (1980)). However, in our empirical section, we draw values of ρ from an importance function, which can be more efficient due to the truncation. In particular, we use truncated (at $\rho=1$) Student priors for both H_1 and H_3 which are constructed in such a way that their untruncated counterparts mimic the moments of the relevant uniform prior mirrored around $\rho=1$. This yields half-Students with a mode at $\rho=1$. Finally, the degrees of freedom parameter is chosen to be 3 so that this alternative prior has fat tails yet still allows the first two moments to exist.

Although the first two posterior moments of ρ may not be crucially affected by the difference between priors, Koop *et al.* (1992) show that results for n-step ahead prediction can differ dramatically. That is, predictive means and variances will exist for any horizon (n) in the case of a bounded uniform prior; however the Student prior for ρ allows only for finite predictive means (given η) for n up to approximately T, and for finite predictive variances if n is less than approximately T/2. In Section 5 we introduce a loss function based on predictive variances whose behavior is expected to differ across priors.

This concludes our development of a prior for the parameters of our sampling model. It is worth emphasizing that, with four exceptions, ρ, η, d_μ and d_β, the priors for all our parameters are noninformative. We believe that the priors we specify for these exceptions will not be considered unreasonable by other researchers.

Section 4: The Posterior Density

Combining our results from the two previous sections yields our Bayesian model:

$$P(y,\theta_S,\tau^2,\eta,\lambda,\delta,\gamma,\kappa \mid y_{(0)}) = c_1\tau^{-2}P(\rho)P(\alpha_D)P(\eta)P(\lambda)$$
$$P(\delta)P(\gamma)P(\kappa)\{\lambda\sum_{i=1}^{2}\delta_i P(y \mid \theta_N,\tau^2,\eta,y_{(0)},M_{Ni})$$
$$+(1-\lambda)\sum_{j=1}^{2}\gamma_j\sum_{q=1}^{2}\kappa_q P(y \mid \theta_S,\tau^2,\eta,y_{(0)},M_{Sjq})\} \tag{13}$$

Using results from Osiewalski and Steel (1993a), we integrate out τ^2 and the mixing parameters, which yields:

$$P(y,\theta_S,\eta \mid y_{(0)}) = c_2 P(\rho)P(\alpha_D)P(\eta)$$
$$\{\frac{1}{3}|V_i(\eta)|^{-1/2}[d_{Ni}(\theta_N,\eta)]^{-T/2}$$
$$+\frac{2}{3}\sum_{j=1}^{2}\frac{1}{2}\sum_{q=1}^{2}\frac{1}{2}\ |V_j(\eta)|^{-1/2}[d_{Sjq}(\theta_S,\eta)]^{-T/2}\} \tag{14}$$

where $c_2 = c_1\Gamma(T/2)\pi^{-T/2}$ and definitions (5) and (8) are used.

For the individual models we obtain:

$$P(\theta_N,\eta \mid y,y_{(0)},M_{Ni}) =$$
$$C_{Ni}^{-1}\ P(\rho)P(\eta)\,|V_i(\eta)|^{-\frac{1}{2}}[d_{Ni}(\theta_N,\eta)]^{-T/2} \tag{15}$$

and

$$P(\theta_S,\eta \mid y,y_{(0)},M_{Sjq}) =$$
$$C_{Sjq}^{-1}\ P(\rho)P(\eta)\,|V_j(\eta)|^{-\frac{1}{2}}[d_{Sjq}(\theta_S,\eta)]^{-T/2}\ , \tag{16}$$

where C_{Ni} and C_{Sjq} are the integrating constants needed to construct posterior odds (ie. $C_{Ni} = P(y|y_{(0)},M_{Ni})$ and $C_{Sjq} = P(y|y_{(0)},M_{Sjq})$). Although the integrating constants may be calculated directly, it should be noted that α_N may be integrated out of (15) and (16) analytically using the properties of multivariate Student distributions. Once α_N is integrated out, the C_{Ni}'s and C_{Sjq}'s may be calculated using Monte Carlo integration. One-dimensional integration is required for calculation of C_{N1}; two-dimensional integration for C_{N2} and C_{S1q}; and three-dimensional integration for C_{S2q}. Formally, the posterior density for α_D, given ρ and η, is a truncated Student-t over the region given in (12). If this region covers most of the parameter space where the likelihood function is appreciable, the truncation will not matter. In this case we can integrate out the full α_S vector as a joint Student density, leaving only one and two dimensional integrals for C_{S1q} and C_{S2q} which we calculate using Monte Carlo integration. A check on this approximation is to perform the integration with respect to α_D numerically by direct simulation with rejection.

The integrating constant for the sampling model, $C = P(y|y_{(0)})$, is given by:

$$C=\frac{1}{6}\left[\sum_{i=1}^{2} C_{Ni}+\sum_{j=1}^{2}\sum_{q=1}^{2} C_{Sjq}\right]. \tag{17}$$

These integrating constants can be used to calculate the posterior probabilities of the various submodels.

$$P(M_{Ni} \mid y, y_{(0)}) = C_{Ni}/6C$$
$$P(M_{N} \mid y, y_{(0)}) = (C_{N1}+C_{N2})/6C$$
$$P(M_{Sjq} \mid y, y_{(0)}) = C_{Sjq}/6C$$
$$P(M_{Sj} \mid y, y_{(0)}) = (C_{Sj1}+C_{Sj2})/6C$$
$$P(M_{Sq} \mid y, y_{(0)}) = (C_{S1q}+C_{S2q})/6C.$$

The posterior model probabilities may indicate, among other things, whether structural breaks are present or if errors exhibit MA(1) behavior. Although not given here, inference on the parameters could be obtained from weighted averages of (15) and (16), where the weights are the relevant model probabilities.

Under all hypotheses, we use the same general mixture of submodels for the sampling density. Note, however, that in all cases, the relative posterior weights given to the submodels depend on the data.

Section 5: Decision Theory

In the previous sections we have described how the posterior probabilities of various hypotheses can be calculated using Bayesian methods. However, econometricians must frequently make decisions. For instance, in a pre-testing exercise a decision must frequently be made as to whether a unit root is present in a series. If present, the series may have to be differenced in a larger VAR model. The Bayesian paradigm provides a formal framework for making such decisions. To make a decision the researcher specifies a loss function and chooses the action which minimizes expected loss (see Zellner (1971)). By focussing on posterior probabilities, previous Bayesian researchers have implicitly used a very simple loss function where all losses attached to incorrect decisions are equal. (That is, the loss associated with the choice of a unit root when the series is stationary is equal to that associated with the assumption of stationarity when a unit root is present). Classical researchers use a loss function where losses are asymmetric, viz. where the choice of a level of significance implicitly defines the loss function. Lacking a measure over the parameter space, classical researchers are forced to look for, say, dominating strategies (which are rare) or minimax solutions. It is this lack of formal development and justification of a loss function which is, in our opinion, a serious weakness of previous Bayesian and classical unit root studies. This section proposes two loss functions which we use to make decisions on whether to accept or reject the unit root hypothesis.

In practice, the choice of loss function depends on the exact nature of the empirical exercise. For example, if the purpose of the analysis is to pretest for a unit root in each series before beginning a multivariate analysis, then a different loss function might be suitable relative to the case where the unit root hypothesis is of interest in and of itself. Hence, although we believe that the loss functions we propose are very sensible, we acknowledge that other researchers may choose other loss functions. Indeed, we believe this to be an advantage of a decision theoretic approach since researchers are forced to explicitly state and defend the assumptions of their analysis. We cannot overemphasize that a classical analysis has, in the choice of testing procedure and significance level, an implicitly defined loss function. However, because almost every researcher uses the same implicit loss function they are rarely forced to justify it.

Our criterion for the evaluation of losses associated with incorrect decisions is prediction, an important one in that the macroeconomic time series in this study are frequently used for prediction (eg. to forecast from VAR models). We develop two loss functions, the first of which is based on predictive means and variances, and the second of which is a computationally convenient approximation to the first. A special case of our first loss function is simply the normalized difference in mean-squared errors (MSEs) between the chosen model and the "correct" model. However, the cost of assuming stationarity when the series are really nonstationary may be drastically different from the converse. Since differences between nonstationary and stationary models are more pronounced for predictive variances than for predictive means, predictive variances are emphasized here. We allow for asymmetries in our loss function which imply that it is more costly to underestimate predictive variances (and give a false sense of accuracy) than to overestimate predictive variances. Given that the precision of forecasts is often a crucial issue we believe this approach to be a sensible one.

In this paper we use predictive results for the simple AR(1) model with intercept and trend, which amounts to conditioning on ϕ, η and α_D. It would be computationally demanding to integrate out ϕ, η and α_D, and results would be almost identical since our predictive variances are mainly affected by ρ. The predictive variance, conditional on ρ, for forecasting n periods ahead is given in Koop, Osiewalski and Steel (1992) (for T>4):

$$Var(y_{T+n} \mid y, y_{(0)}, \rho) = \frac{SSE_\rho}{T-4}\left[\sum_{i=0}^{n-1}\rho^{2i}+\frac{2}{T(T^2-1)}\sum_{j=1}^{n}\sum_{i=1}^{n} r(i,j)\rho^{2n-i-j}\right] \tag{18}$$

when we integrate out μ, β and τ^2 using the noninformative priors given in Section 3. In (18) we use $r(i,j) \equiv 6ij + 3(T-1)(i+j)+ 2T^2 - 3T+ 1$ and $SSE_\rho = (y-\rho y_{-1})' M(y-\rho y_{-1})$ where $M = I-X(X'X)^{-1}X'X$ is a Tx2 matrix containing observations on the intercept and trend. The corresponding predictive mean conditional on ρ is given by:

$$E(y_{T+n} \mid y, y_{(0)}, \rho) = \rho^n y_T + q_\rho' \beta_\rho, \tag{19}$$

where

$$\beta_\rho = (X'X)^{-1}X'(y-\rho y_{-1}),$$

and

$$q_\rho' = \left[\sum_{i=1}^{n}\rho^{n-i} \quad \sum_{i=1}^{n}(T+i)\rho^{n-i}\right].$$

For notational simplicity we refer to the predictive mean and variance, with ρ marginalized out, at horizon n under H_j (j=1,2,3) as $mean_n^j$ and var_n^j, respectively.

Our first loss function takes the form:

$$l_{d,s}^{n,1}=\max(1, var_n^d/var_n^s)+\delta \max(1, var_n^s/var_n^d)-(1+\delta) + \frac{[mean_n^d-mean_n^s]^2}{var_n^s},$$

where H_d is the hypothesis chosen; H_s is the "correct" hypothesis; and δ, which is greater than or equal to 1, reflects our aversion to underestimating the predictive variance. In order to deal with model uncertainty, we calculate the predictive mean and variance for each of the six submodels under H_1, H_2 and H_3, respectively, and then average across models using the relevant posterior model probabilities. For each decision, d, we compute the expected loss:

$$l_d^{n,1}=\sum_{s=1}^{3} l_{d,s}^{n,1}\, p(H_s \mid y, y_{(0)}),$$

and choose d for which the loss is minimal for a given forecast horizon, n.

Our loss function has some attractive properties. Note first, that if the correct model is chosen (ie. d=s), then the loss is zero. Secondly, the loss increases as predictive bias increases or as the predictive variance of the selected model diverges from the "correct" model. Thirdly, if $var_n^d > var_n^s$ and we overestimate the predictive variance, then

$$l_{d,s}^{n,1}=\frac{MSE_n^d-MSE_n^s}{MSE_n^s}, \qquad (20)$$

where

$$MSE_n^j=var_n^j+(Bias_n^j)^2,$$

and

$$Bias_n^j=mean_n^j-mean_n^s.$$

Note that the bias equals zero if the "correct" model is chosen. It would, of course, be possible to use (20) as our loss function for cases where $var_n^d < var_n^s$ as well. However, we feel that it is important to allow for losses to be asymmetric, which we do by introducing δ.

The parameter δ plays an important role in our loss function. If $\delta=1$, the loss function is symmetric in the sense that underestimating and overestimating the predictive variance are equally costly. For values of δ greater than one underestimating the predictive variance (and giving a researcher excessive confidence in her forecasts) is more costly than overestimating the predictive variance. The loss function is constructed such that losses are: i) equal to the relative overestimation of the predictive variance plus the scaled squared bias if the chosen model has a bigger variance than the "correct" model; and ii) equal to δ times the relative underestimation of the predictive variance plus the scaled squared bias if the chosen model has a smaller variance than the "correct" model.

Koop *et al.* (1992) discusses the properties of predictive means and variances in great detail. For present purposes, it is sufficient merely to note that the predictive means do not differ much across hypotheses but that the predictive variances do. Under H_1 the predictive variance is at least of $O(n^2)$. Note that the trend-stationary model does not lead to bounded predictive variances if parameter uncertainty is taken into account (see Sampson (1991) for a classical analysis). Under H_2 and H_3, the predictive variance grows, respectively, at rate $O(n^4)$ and exponentially.

It is crucial to consider multi-period predictions since they bring out the differences in predictive behavior between stationary, unit root, and explosive models (see Chow (1973) for some specific problems when $n > 1$). At short horizons the losses do not differ much across models (unless δ is very large) and the model is chosen largely on the basis of its posterior probability. At long forecast horizons, the differences in predictive variances between stationary and nonstationary models grow large; and assuming $\delta > 1$, nonstationary models grow concomitantly more attractive. So, if there is any chance that the correct model is nonstationary, our loss function will choose it at some forecast horizon. In other words, the cost of incorrectly choosing the stationary model and seriously underestimating the predictive variance will eventually dominate at some forecast horizon. H_2 will be chosen if n goes to infinity and δ is held constant; and if δ goes to infinity and n is held constant, H_3 will be chosen. Since the decision taken depends crucially on the choice of n and δ, a sensitivity analysis is performed over these two parameters.

This first loss function combines predictive means and predictive variances in a plausible way. Despite these characteristics, the loss function is computationally burdensome because the calculation of var_n^j requires that (18) be evaluated at each draw in our Monte Carlo procedure. For this reason, we introduce a second loss function which ignores the SSE_ρ term which is very similar across models, and the bias term. Furthermore, we replace the powers of ρ in (18) by their expected values; that is, we replace ρ^j with $E(\rho^j)$ calculated using Monte Carlo integration. This strategy amounts to approximating var_n^j by $E(var(y_{T+n} | y, y_{(0)}, \rho))$, where the expectation is taken over ρ. By not fully marginalizing with respect to ρ, we ignore an additional term which would have been added to the predictive variances under H_1 and H_3. For this reason, predictive variances for the trend-stationary and explosive models are slightly underestimated relative to the unit root model, a characteristic which we observe and discuss in our empirical results.

More formally we define:

$$g_n^j(\rho) \equiv \left[\sum_{l=0}^{n-1} \rho^{2l} + \frac{2}{T(T^2-1)} \sum_{l=1}^{n} \sum_{i=1}^{n} r(i,l)\, \rho^{2n-i-l} \right] \mid H_j .$$

For each H_j, we can use the marginal posterior of ρ to calculate the posterior mean:

$$Eg_n^j = E(g_n^j(\rho) \mid H_j, y, y_{(0)}) .$$

Our second loss function can be written as:

$$l_{d,s}^{n,2} = \max(1, Eg_n^d / Eg_n^s) + \delta \max(1, Eg_n^s / Eg_n^d) - (1+\delta) .$$

This loss function is much simpler to compute and has approximately the same properties as the first loss function. Evidence presented in the empirical section indicates that the approximation is a good one.

The decision theoretic approach is based on the assumption that researchers are interested in choosing a particular region for ρ since they may wish, for instance, to difference the data. However, in cases where such a pretest strategy is not required, we suggest basing predictions on a mixture over regions for ρ weighted with the relevant posterior probabilities.

Section 6: Empirical Results

This section presents evidence on the existence of a unit root in the Nelson-Plosser series. The data used are extended to cover the period until 1988 (see Data Appendix). Tables 1 and 2 present posterior means and standard deviations for ρ and η under H_1 and H_3, while Table 3 presents evidence on the presence of structural breaks and moving average errors. Table 4 contains the posterior probabilities of H_1, H_2 and H_3, and Table 5 summarizes the results of the decision analysis. Posterior odds are calculated for testing the various hypotheses with respect to ρ by using the sampling model weighted over all the submodels. Although our primary focus is on the unit root hypothesis, two subsidiary questions are simultaneously addressed: (1) Is there evidence of one or more structural breaks in our economic time series? (2) Is there evidence of MA(1) behavior in the error terms?

Since parameter estimates are only slightly relevant to the issues we address in this paper, we discuss these only briefly. Note first that Tables 1 and 2 support the conclusions of Choi (1990): Omitting the MA(1) component of the error term does indeed tend to drive estimates of ρ towards one in a manner consistent with the asymptotic bias derived by Choi. Table 3 contains the probability that an MA(1) error term is present as well as the AR(3) structure already allowed for in our specification. For many series this probability is very high and for no series is it small enough to be ignored. Thus Choi's results are more than just theoretically interesting. The inclusion of a moving average error term would appear to be an important part of any specification. With respect to structural breaks in Table 3, note that, although our results are consistent with Perron's contention that a level break occurred in 1929 in many macroeconomic time series, we find virtually no evidence for the presence of a trend break in 1973 for any of the series. This latter finding is not inconsistent with Perron's since he only considered the trend-break model when using post-war quarterly data. As Perron (1989) notes, models with structural breaks tend to yield less evidence of a unit root.

We do not discuss Table 4 in detail but we do use the results to calculate the expected losses required for our decision analysis. For our purposes it is sufficient to note that results show that trend-stationarity (H_1) is the most probable hypothesis for most of the series (notable exceptions are the CPI and velocity); however, without a formal loss function it would be rash to rule out the unit root model at this time.

Results from our decision theoretic analysis are presented in Table 5, which contains only those for our computationally attractive second loss function, $l_d^{n,2}$. To judge the difference between the two loss functions, a decision theoretic analysis was carried out using our first loss function, $l_d^{n,1}$, for the uniform prior for two series which exhibit very different behavior: real GNP and the CPI. Using $l_d^{n,1}$ and real GNP, we choose, for $\delta=1,10$ and 100, respectively: i) H_1 for $n<84$, else H_2; ii) H_1 for $n<61$, else H_2; and iii) H_2. A comparison with Table 5 indicates that results using the second loss function are qualitatively the same; that is, a researcher would choose H_1 unless $\delta=100$ or she were interested in very long-run predictions. As expected, however, using the approximation slightly biases results in favor of H_2. For CPI conclusions are exactly the same for our two loss functions. These findings indicate that the second loss function provides a good approximation to the first; consequently, we use only $l_d^{n,2}$ for the other series.

It is worth emphasizing that our loss functions have two key properties. First, as long as δ is greater than 1, it is better to overestimate than to underestimate predictive variances. This property tends the researcher toward favoring H_2 over H_1 and H_3 over H_2 and H_1. Indeed as δ goes to infinity and holding n constant, H_3 will always be chosen. Second, there is a tendency in our loss functions to favor H_2. H_2 lies between H_1 and H_3 such that a researcher will generally never go too far wrong in choosing H_2. (Potential losses would be very large if, say, H_1 were chosen when H_3 were the correct model). In fact, as n goes to infinity and holding δ constant H_2 will always be chosen. These two properties account for most of the findings in Table 5, which presents the model chosen for different values of n and δ. With the exception of the CPI and velocity series and, to a lesser extent, the GNP deflator and real wage series, H_1 is the model chosen (so long as δ or n is not large). However, clear scope exists for choosing nonstationarity if underestimating predictive variances is felt to be a serious problem. If $\delta=100$ a researcher would almost never select the trend-stationary model. Figures 1 and 2 graphically depict the behavior of $l_d^{n,2}$ (which is defined analogously to $l_d^{n,1}$) as n varies for the uniform prior for real GNP per capita and CPI with $\delta=10$.

There appears to be less sensitivity of our loss function with respect to n. If we restrict attention to short- or medium-term forecasts (eg. $n < 10$), only a few cases exist where different values of n yield different conclusions. A typical example is real GNP, where, unless the researcher is interested in forecasting four or more decades into the future, the trend-stationary model is chosen for $\delta=1$ or 10. Only if $\delta=100$ (a strong penalty for underestimating predictive variances) is the unit root model selected.

It is interesting to compare our results to those of other researchers, although it should be stressed from the outset that different authors use data series of different lengths so that some results are not directly comparable. A very nice summary of previous work using the Nelson-Plosser data set is given in Table 1 of Schotman and van Dijk (1991b). Note that: i) Nelson and Plosser (1982)) fail to reject the null hypothesis of a unit root in thirteen of the fourteen series, the exception being the unemployment rate. Other classical papers exhibit similar patterns. ii) Most Bayesian papers find much more evidence for trend-stationarity than do their classical counterparts. Even with a prior heavily weighted towards nonstationarity, Phillips finds that the probability of nonstationarity is greater than a half for only one or two series (see Phillips (1991), Table IV).

Our Table 4 yields similar results to many of the other Bayesian analyses; that is, we find strong evidence of trend-stationarity for some series, but not for all. For around half of the series, evidence is decidedly mixed. This finding is hardly surprising, given the difficulty in distinguishing between unit root models and persistent trend-stationary models in finite samples. In our opinion, previous Bayesian analyses have not gone far enough, and in view of this uncertainty, we believe that researchers must specify a loss function if a single model is to be selected.

Our results in Table 5 are closer to those obtained in classical analyses. For example, for $\delta=10$ and $n>6$ either the unit root or the explosive model would be selected for eight of our fourteen series under the uniform prior. It is also interesting to note that our results for $\delta=10$ with the uniform prior correspond closely to those given in Phillips (1991, reply), who uses the Bayes model likelihood ratio test on the same data. This test can be thought of as an "objective" Bayesian test for a unit root, and is described in more detail in Phillips and Ploberger (1991). Phillips' results differ from ours chiefly in that he finds the nominal wage series to contain a unit root, whereas we only match this finding if n is very large or $\delta=100$. Note, however, that our results are obtained using a formal decision theoretic approach based on a strong aversion to underestimating predictive variances. Researchers who do not wish to include such an aversion in their analysis will tend to choose trend-stationarity more often.

A final issue worth discussing is the sensitivity of our results to various priors. As described in Section 4, we use two different priors for ρ: a half-Student and a bounded uniform prior. The first and second order moments of the half-Student prior are chosen so as to match the uniform prior. The differences between the two priors occur in third and higher moments. Tables 1 and 2 indicate that posterior first and second moments do not differ much across the two priors. The remaining tables, however, indicate somewhat larger differences. This is especially true of Table 5, where in some cases, the two very similar priors yield different conclusions (eg. Nominal GNP and Money Stock for $\delta=10$ or the GNP deflator for $\delta=1$). As described in Section 3, predictive variances exist only for n less than approximately T/2 when a Student t prior under H_1 and H_3 is used. More precisely, for this Student t case, Table 5 reports results only for $n<(T+1)/2$. Our decision analysis depends upon high order moments of ρ and our priors differ in these higher moments. Recall that, while all moments exist for our bounded uniform prior, none beyond 2 exist for our half-Student prior. Although Bayesians who use informative priors typically do not worry about third or higher order prior moments, our analysis suggests that care should be taken in eliciting priors when a decision analysis which involves prediction is carried out. The effect of prior moments on the existence of predictive variances for multi-period forecasting is formally analyzed in Koop *et al.* (1992).

Section 7: Conclusions

The paper develops a formal decision theoretic approach to testing for unit roots which involves the use of a loss function based on predictive moments. It also extends the class of likelihood functions in the Bayesian unit root literature by using a likelihood function which is a mixture over submodels which differ in covariance structure and in the treatment of structural breaks. Each of the individual likelihoods mixed into the overall likelihood function belongs to the class of general elliptical densities.

Our empirical results indicate that a high posterior probability of trend-stationarity exists for most of the economic time series. This finding is consistent with most previous Bayesian analyses. However, if there is a high cost to underestimating predictive variances, our ensuing decision analysis indicates that trend-stationarity is not necessarily the preferred choice. Thus, our loss function can lead to results similar to those of many classical analyses; that is, it can often select the unit root model.

These findings highlight the importance of the decision theoretic part of our, or of any, analysis. In general, the researcher must think clearly about the consequences of selecting a hypothesis in an empirical context. A Bayesian analysis which merely presents posterior model probabilities is incomplete; and a classical analysis which accepts passively the loss structure implicit in the choice of a significance level may be misleading.

Table 1: Posterior Means for ρ and η under H_1 (Standard deviations in parentheses)

			Uniform Prior ρ			Student Prior ρ	
		No MA ρ	MA ρ	MA η	No MA ρ	MA ρ	MA η
Real GNP	nb	0.8134	0.7462	0.4416	0.8291	0.7836	0.3483
		(.0570)	(.0889)	(.3377)	(.0594)	(.0894)	(.3705)
	lb	0.7409	0.6941	0.3815	0.7669	0.7242	0.3484
		(.0681)	(.0829)	(.2880)	(.0689)	(.0999)	(.3127)
	tb	0.8127	0.7338	0.5178	0.8288	0.7732	0.4372
		(.0562)	(.0862)	(.2943)	(.0547)	(.0877)	(.3365)
Nominal GNP	nb	0.9411	0.9031	0.6737	0.9434	0.9025	0.7512
	lb	(.0296)	(.0448)	(.1683)	(.0287)	(.0485)	(.1290)
		0.7777	0.7555	0.3228	0.7991	0.7862	0.3168
	tb	(.0630)	(.0763)	(.2410)	(.0634)	(.0760)	(.2612)
		0.9209	0.8514	0.7762	0.9251	0.8728	0.7744
		(.0371)	(.0659)	(.1206)	(.0355)	(.0625)	(.1182)
Real per cap. GNP	nb	0.8032	0.7363	0.4321	0.8201	0.7782	0.3303
		(.0579)	(.0889)	(.3407)	(.0577)	(.0914)	(.3838)
	lb	0.7564	0.7022	0.4263	0.7813	0.7345	0.3753
		(.0671)	(.0845)	(.2970)	(.0688)	(.0984)	(.3260)
	tb	0.8032	0.7256	0.5152	0.8205	0.7636	0.4365
		(.0583)	(.0866)	(.3004)	(.0579)	(.0918)	(.3383)
Ind. Prod.	nb	0.8256	0.7626	0.3843	0.8392	0.7985	0.3003
		(.0523)	(.0859)	(.3072)	(.0515)	(.0832)	(.3356)
	lb	0.7498	0.6952	0.3530	0.7743	0.7244	0.3181
		(.0678)	(.0811)	(.2401)	(.0666)	(.0984)	(.2620)
	tb	0.8149	0.7386	0.4430	0.8296	0.7731	0.3819
		(.0536)	(.0847)	(.2833)	(.0538)	(.0849)	(.2976)
Employment	nb	0.8637	0.8024	0.4442	0.8734	0.8273	0.4160
		(.0473)	(.0747)	(.2357)	(.0458)	(.0694)	(.2253)
	lb	0.7982	0.7300	0.4209	0.8150	0.7599	0.3953
		(.0563)	(.0767)	(.1916)	(.0555)	(.0773)	(.1954)
	tb	0.8578	0.7866	0.4873	0.8679	0.8148	0.4525
		(.0484)	(.0774)	(.2190)	(.0471)	(.0739)	(.2260)
Unempl. Rate	nb	0.7454	0.6586	0.5935	0.7747	0.6644	0.6001
		(.0736)	(.0748)	(.1303)	(.0750)	(.1117)	(.1242)
	lb	0.7144	0.6523	0.5866	0.7459	0.6412	0.5912
		(.0764)	(.0740)	(.1244)	(.0824)	(.1170)	(.1278)
	tb	0.7378	0.6587	0.5922	0.7682	0.6542	0.6055
		(.0758)	(.0739)	(.1362)	(.0770)	(.1140)	(.1275)
GNP Deflator	nb	0.9634	0.9474	0.4973	0.9640	0.9468	0.5646
		(.0189)	(.0294)	(.3127)	(.0188)	(.0294)	(.2417)
	lb	0.9166	0.8843	0.5462	0.9194	0.8909	0.5313
		(.0289)	(.0423)	(.2314)	(.0285)	(.0396)	(.2400)
	tb	0.9321	0.8942	0.6154	0.9347	0.9095	0.4408
		(.0300)	(.0477)	(.2196)	(.0295)	(.0427)	(.3704)

Table 1 (continued): Posterior Means for ρ and η under H_1 (Standard deviations in parentheses)

			Uniform Prior ρ			Student Prior ρ	
		No MA ρ	MA ρ	MA η	No MA ρ	MA ρ	MA η
CPI	nb	0.9886 (.0077)	0.9804 (.0134)	0.6531 (.1406)	0.9887 (.0078)	0.9808 (.0129)	0.6286 (.1665)
	lb	0.9888 (.0077)	0.9804 (.0134)	0.6539 (.1474)	0.9888 (.0077)	0.9838 (.0099)	0.4881 (.3427)
	tb	0.9820 (.0114)	0.9679 (.0120)	0.6582 (.1456)	0.9820 (.0115)	0.9694 (.0198)	0.6412 (.1718)
Wages	nb	0.9373 (.0279)	0.9053 (.0459)	0.5068 (.3128)	0.9393 (.0273)	0.9032 (.0487)	0.5165 (.3155)
	lb	0.7999 (.0471)	0.7818 (.0593)	0.2137 (.2292)	0.8120 (.0472)	0.7822 (.0596)	0.2225 (.2338)
	tb	0.9212 (.0345)	0.8725 (.0596)	0.6076 (.2479)	0.9247 (.0332)	0.8723 (.0591)	0.5960 (.2659)
Real Wages	nb	0.9280 (.0395)	0.8818 (.0659)	0.6506 (.2152)	0.9322 (.0377)	0.9038 (.0560)	0.5466 (.2737)
	lb	0.9276 (.0397)	0.8751 (.0672)	0.7391 (.2361)	0.9324 (.0377)	0.8867 (.0614)	0.7954 (.1909)
	tb	0.8112 (.0574)	0.7159 (.0807)	0.6047 (.2278)	0.8316 (.0502)	0.7668 (.0886)	0.4624 (.3292)
Money Stock	nb	0.9402 (.0233)	0.9070 (.0380)	0.5721 (.1882)	0.9415 (.0229)	0.9123 (.0357)	0.5534 (.2060)
	lb	0.8807 (.0318)	0.8454 (.0446)	0.4773 (.2041)	0.8848 (.0316)	0.8550 (.0432)	0.4623 (.2158)
	tb	0.9187 (.0270)	0.8726 (.0440)	0.5924 (.1789)	0.9210 (.0269)	0.8811 (.0424)	0.5700 (.2008)
Velo-city	nb	0.9629 (.0212)	0.9395 (.0356)	0.5648 (.3094)	0.9635 (.0207)	0.9437 (.0341)	0.5481 (.3220)
	lb	0.9635 (.0209)	0.9383 (.0360)	0.6035 (.2607)	0.9642 (.0206)	0.9418 (.0342)	0.5976 (.2668)
	tb	0.9580 (.0253)	0.9289 (.0431)	0.6083 (.2823)	0.9594 (.0246)	0.9329 (.0412)	0.6248 (.2509)
Bond Yield	nb	0.9466 (.0299)	0.9195 (.0466)	0.4860 (.1987)	0.9488 (.0289)	0.9277 (.0427)	0.4560 (.2314)
	lb	0.8931 (.0441)	0.8386 (.0674)	0.5516 (.2024)	0.9003 (.0430)	0.8583 (.0638)	0.5355 (.2088)
	tb	0.9449 (.0410)	0.9152 (.0647)	0.4917 (.2198)	0.9501 (.0380)	0.9283 (.0538)	0.4897 (.2005)
Stock Pri-ces	nb	0.9297 (.0333)	0.8991 (.0527)	0.3569 (.3018)	0.9329 (.0320)	0.9080 (.0493)	0.3339 (.3032)
	lb	0.9135 (.0351)	0.8829 (.0512)	0.3322 (.2367)	0.9175 (.0346)	0.8932 (.0480)	0.3152 (.2333)
	tb	0.9069 (.0378)	0.8581 (.0619)	0.4342 (.2663)	0.9120 (.0362)	0.8751 (.0579)	0.3934 (.2811)

* nb = no break, lb = level break, tb = trend break.

Table 2: Posterior Means for ρ and η under H_3 (Standard deviations in parentheses)

			Uniform Prior ρ			Student Prior ρ	
		No MA ρ	MA ρ	MA η	No MA ρ	MA ρ	MA η
Real GNP	nb	1.0167	1.0285	0.4292	1.0134	1.0155	0.3609
		(.0155)	(.0292)	(.4019)	(.0126)	(.0143)	(.4043)
	lb	1.0189	1.0266	0.6654	1.0153	1.0184	0.4962
		(.0175)	(.0227)	(.2331)	(.0144)	(.0168)	(.4123)
	tb	1.0172	1.0219	0.4941	1.0136	1.0172	0.4811
		(.0159)	(.0191)	(.3357)	(.0125)	(.0162)	(.3491)
Nominal GNP	nb	1.0138	1.0196	0.5850	1.0117	1.0155	0.5143
		(.0124)	(.0177)	(.2637)	(.0105)	(.0138)	(.3284)
	lb	1.0186	1.0260	0.6703	1.0144	1.0180	0.5999
		(.0174)	(.0221)	(.2414)	(.0136)	(.0163)	(.3356)
	tb	1.0158	1.0225	0.6064	1.0129	1.0181	0.6181
		(.0142)	(.0200)	(.2720)	(.0177)	(.0153)	(.2546)
Real per cap. GNP	nb	1.0174	1.0230	0.4118	1.0135	1.0159	0.3805
		(.0163)	(.0217)	(.3828)	(.0126)	(.0151)	(.4010)
	lb	1.0193	1.0241	0.5183	1.0152	1.0184	0.5182
		(.0181)	(.0201)	(.4057)	(.0138)	(.0163)	(.3998)
	tb	1.0177	1.0230	0.5088	1.0138	1.0169	0.4824
		(.0166)	(.0202)	(.3344)	(.0128)	(.0154)	(.3488)
Ind. Prod.	nb	1.0150	1.0184	0.2981	1.0125	1.0140	0.2737
		(.0142)	(.0178)	(.3381)	(.0115)	(.0132)	(.3477)
	lb	1.0179	1.0215	0.3533	1.0141	1.0159	0.3517
		(.0169)	(.0195)	(.4593)	(.0131)	(.0139)	(.4285)
	tb	1.0153	1.0193	0.3767	1.0124	1.0143	0.3615
		(.0144)	(.0183)	(.3064)	(.0117)	(.0127)	(.3066)
Employment	nb	1.0150	1.0192	0.3709	1.0128	1.0151	0.3713
		(.0141)	(.0176)	(.2335)	(.0115)	(.0147)	(.2287)
	lb	1.0156	1.0199	0.4118	1.0127	1.0150	0.4007
		(.0152)	(.0187)	(.2173)	(.0117)	(.0135)	(.2260)
	tb	1.0154	1.0193	0.4164	1.0123	1.0153	0.4163
		(.0144)	(.0180)	(.2198)	(.0115)	(.0134)	(.2069)
Unemp. Rate	nb	1.0211	1.0272	0.5908	1.0161	1.0192	0.5823
		(.0189)	(.0228)	(.1231)	(.0149)	(.0172)	(.1394)
	lb	1.0226	1.0287	0.6006	1.0166	1.0203	0.6018
		(.0205)	(.0241)	(.1287)	(.0153)	(.0198)	(.1285)
	tb	1.0218	1.0256	0.5998	1.0165	1.0188	0.5942
		(.0198)	(.0213)	(.1306)	(.0156)	(.0164)	(.1286)
GNP Deflator	nb	1.0091	1.0138	0.5207	1.0083	1.0116	0.5074
		(.0082)	(.0136)	(.3016)	(.0075)	(.0108)	(.3097)
	lb	1.0096	1.0131	0.5662	1.0087	1.0110	0.5730
		(.0089)	(.0125)	(.2805)	(.0079)	(.0103)	(.2737)
	tb	1.0120	1.0175	0.5726	1.0105	1.0137	0.5647
		(.0110)	(.0153)	(.2737)	(.0095)	(.0120)	(.2793)

Table 2 (continued): Posterior Means for ρ and η under H_3

			Uniform Prior ρ			Student Prior ρ	
		No MA ρ	MA ρ	MA η	No MA ρ	MA ρ	MA η
CPI	nb	1.0067 (.0055)	1.0095 (.0083)	0.5818 (.1977)	1.0065 (.0053)	1.0075 (.0071)	0.3119 (.4508)
	lb	1.0069 (.0056)	1.0103 (.0085)	0.6316 (.1626)	1.0065 (.0054)	1.0081 (.0074)	0.3985 (.4384)
	tb	1.0081 (.0069)	1.0119 (.0105)	0.6430 (.1466)	1.0078 (.0065)	1.0097 (.0088)	0.4469 (.3990)
Wages	nb	1.0111 (.0104)	1.0153 (.0144)	0.4781 (.3230)	1.0122 (.0109)	1.0126 (.0118)	0.4522 (.3378)
	lb	1.0125 (.0123)	1.0184 (.0169)	0.5453 (.3167)	1.0112 (.0103)	1.0143 (.0138)	0.5508 (.3138)
	tb	1.0132 (.0124)	1.0189 (.0171)	0.5277 (.3097)	1.0120 (.0111)	1.0147 (.0133)	0.5181 (.3045)
Real Wages	nb	1.0207 (.0181)	1.0258 (.0221)	0.4960 (.3116)	1.0159 (.0139)	1.0188 (.0174)	0.4289 (.3881)
	lb	1.0204 (.0178)	1.0293 (.0229)	0.8209 (.1540)	1.0161 (.0143)	1.0209 (.0188)	0.7014 (.3332)
	tb	1.0171 (.0162)	1.0237 (.0205)	0.5541 (.3000)	1.0137 (.0127)	1.0173 (.0165)	0.5212 (.3473)
Money Stock	nb	1.0082 (.0077)	1.0116 (.0108)	0.5326 (.2143)	1.0078 (.0070)	1.0102 (.0095)	0.5142 (.2366)
	lb	1.0085 (.0082)	1.0123 (.0120)	0.5477 (.2010)	1.0079 (.0074)	1.0107 (.0100)	0.5431 (.2059)
	tb	1.0087 (.0081)	1.0124 (.0120)	0.5525 (.2007)	1.0079 (.0073)	1.0106 (.0098)	0.5504 (.1982)
Velocity	nb	1.0120 (.0104)	1.0170 (.0157)	0.5176 (.3398)	1.0106 (.0091)	1.0137 (.0120)	0.4942 (.3588)
	lb	1.0122 (.0107)	1.0177 (.0160)	0.5781 (.2870)	1.0109 (.0093)	1.0144 (.0129)	0.5535 (.3158)
	tb	1.0156 (.0135)	1.0230 (.0209)	0.5782 (.3056)	1.0132 (.0113)	1.0167 (.0150)	0.5696 (.2953)
Bond Yield	nb	1.0163 (.0143)	1.0224 (.0191)	0.4531 (.2135)	1.0136 (.0120)	1.0222 (.0196)	0.4531 (.2173)
	lb	1.0162 (.0151)	1.0209 (.0187)	0.4942 (.2095)	1.0134 (.0120)	1.0217 (.0196)	0.4966 (.2062)
	tb	1.0401 (.0267)	1.0423 (.0274)	0.4346 (.1976)	1.0255 (.0188)	1.0314 (.0275)	0.4362 (.2063)
Stock Prices	nb	1.0141 (.0130)	1.0164 (.0155)	0.2389 (.3432)	1.0120 (.0108)	1.0130 (.0116)	0.2203 (.3461)
	lb	1.0130 (.0121)	1.0159 (.0150)	0.2940 (.2713)	1.0110 (.0102)	1.0133 (.0133)	0.2846 (.2722)
	tb	1.0137 (.0126)	1.0168 (.0157)	0.3215 (.3071)	1.0115 (.0106)	1.0135 (.0122)	0.3092 (.3099)

* nb = no break, lb = level break, tb = trend break.

Table 3: Posterior Probabilities of Elements in Mixtures

		Uniform Prior for ρ			Student Prior for ρ	
	Level Break	Trend Break	Moving Average	Level Break	Trend Break	Moving Average
Real GNP	0.0614	1.2E-5	0.5856	0.1556	3.6E-5	0.4977
Nominal GNP	0.6639	2.9E-5	0.4740	0.8489	4.5E-5	0.4146
Real per cap. GNP	0.1728	2.3E-5	0.5719	0.1391	2.0E-5	0.4991
Indust. Product.	0.2449	0.0001	0.5211	0.1636	9.6E-5	0.4829
Employment	0.4488	1.0E-5	0.7137	0.3626	1.1E-5	0.6454
Unempl. Rate	0.4447	4.9E-4	0.9930	0.4102	4.8E-4	0.9857
GNP Defla-tor	0.2676	1.4E-4	0.5892	0.3034	1.7E-4	0.5950
CPI	0.0438	5.7E-5	0.4718	0.0461	5.9E-5	0.5148
Wages	0.9453	2.4E-6	0.3240	0.9427	5.3E-6	0.2960
Real Wages	0.1428	0.0033	0.5554	0.1252	0.0085	0.5773
Money Stock	0.5586	1.2E-4	0.7550	0.5194	1.3E-4	0.7399
Velocity	0.0108	0.0002	0.7345	0.0108	0.0002	0.7341
Bond Yield	0.7527	0.0125	0.7763	0.9180	0.0070	0.7661
Stock Prices	0.2838	0.0018	0.4587	0.2935	0.0018	0.4485

Table 4: Posterior Probabilities of Regions for ρ

		Uniform Prior for ρ			Student Prior for ρ	
	H_1: $\rho<1$	H_2: $\rho=1$	H_3: $\rho>1$	H_1: $\rho<1$	H_2: $\rho=1$	H_3: $\rho>1$
Real GNP	0.9824	0.0147	0.0029	0.9816	0.0133	0.0051
Nominal GNP	0.8275	0.1371	0.0354	0.9232	0.0476	0.0292
Real per cap. GNP	0.9883	0.0094	0.0023	0.9852	0.0099	0.0049
Industr.Product.	0.9869	0.0108	0.0023	0.9863	0.0099	0.0039
Employment	0.9678	0.0263	0.0059	0.9657	0.0242	0.0101
Unempl. Rate	0.9968	0.0023	0.0009	0.9905	0.0059	0.0036
GNP Deflator	0.4940	0.4339	0.0722	0.6107	0.2896	0.0997
CPI	0.0789	0.8087	0.1124	0.1376	0.6192	0.2432
Wages	0.9794	0.0176	0.0030	0.9803	0.0162	0.0035
Real Wages	0.4355	0.4198	0.1447	0.5515	0.2720	0.1765
Money Stock	0.9097	0.0789	0.0011	0.9360	0.0495	0.0145
Velocity	0.3198	0.5650	0.1152	0.4389	0.4141	0.1470
Bond Yield	0.7057	0.2315	0.0627	0.8790	0.0746	0.0464
Stock Prices	0.6114	0.3243	0.0643	0.7131	0.2068	0.0801

Table 5: Results of Decision Analysis Using $L_d^{n,2}$ (n=2,..,100)

		Uniform Prior for ρ			Student Prior for ρ	
	$\delta=1$	$\delta=10$	$\delta=100$	$\delta=1$	$\delta=10$	$\delta=100$
Real GNP	$n<67$: H_1 else: H_2	$n<49$: H_1 else: H_2	H_2	H_1	H_1	H_2
Nominal GNP	$n<62$: H_1 else: H_2	H_2	$n<13$: H_3 else: H_2	H_1	$n<25$: H_1 else: H_2	$n<10$: H_3 else: H_2
Real per cap. GNP	$n<81$: H_1 else: H_2	$n<62$: H_1 else: H_2	H_2	H_1	H_1	H_2
Ind. Prod.	$n<82$: H_1 else: H_2	$n<65$: H_1 else: H_2	H_2	H_1	H_1	H_2
Employment	$n<80$: H_1 else: H_2	$n<57$: H_1 else: H_2	H_2	H_1	H_1	$n<7$: H_1 else: H_2
Unempl. Rate	$n<80$: H_1 else: H_2	$n<63$: H_1 else: H_2	$n<41$: H_1 else: H_2	H_1	$n<49$: H_1 else: H_2	H_2
GNP Deflator	H_2	H_2	H_3	H_1	$n<4$: H_3 else: H_2	H_3
CPI	H_2	H_3	H_3	H_2	H_3	H_3
Wages	H_1	$n<78$: H_1 else: H_2	H_2	H_1	H_1	H_2
Real Wages	H_2	$n<13$: H_3 else: H_2	H_3	$n<16$: H_1 else: H_2	$n<20$: H_3 else: H_2	H_3
Money Stock	H_1	$n<6$: H_1 else: H_2	$n<3$: H_3 else: H_2	H_1	H_1	$n<5$: H_3 else: H_2
Velocity	H_2	$n<15$: H_3 else: H_2	H_3	H_2	$n<26$: H_3 else: H_2	H_3
Bond Yield	$n<47$: H_1 else: H_2	H_2	$n<33$: H_3 else: H_2	$n<44$: H_1 else: H_2	H_2	$n<23$: H_3 else: H_2
Stock Prices	$n<50$: H_1 else: H_2	H_2	$n<58$: H_3 else: H_2	$n<56$: H_1 else: H_2	H_2	H_3

Notes to Table 5: All moments of the predictive exist for the uniform prior, but for the Student t prior second order moments of the predictive exist up to n=41,41,41,65,50,50,51,65,45,45,51,61,45,60 for our 14 series, respectively. Hence, our loss comparisons in the last three columns only go up to these horizons. Our loss comparisons for the uniform prior go up to n=100.

FIGURE 1: Simplified Loss Function

GNP per capita, delta=10, uniform prior

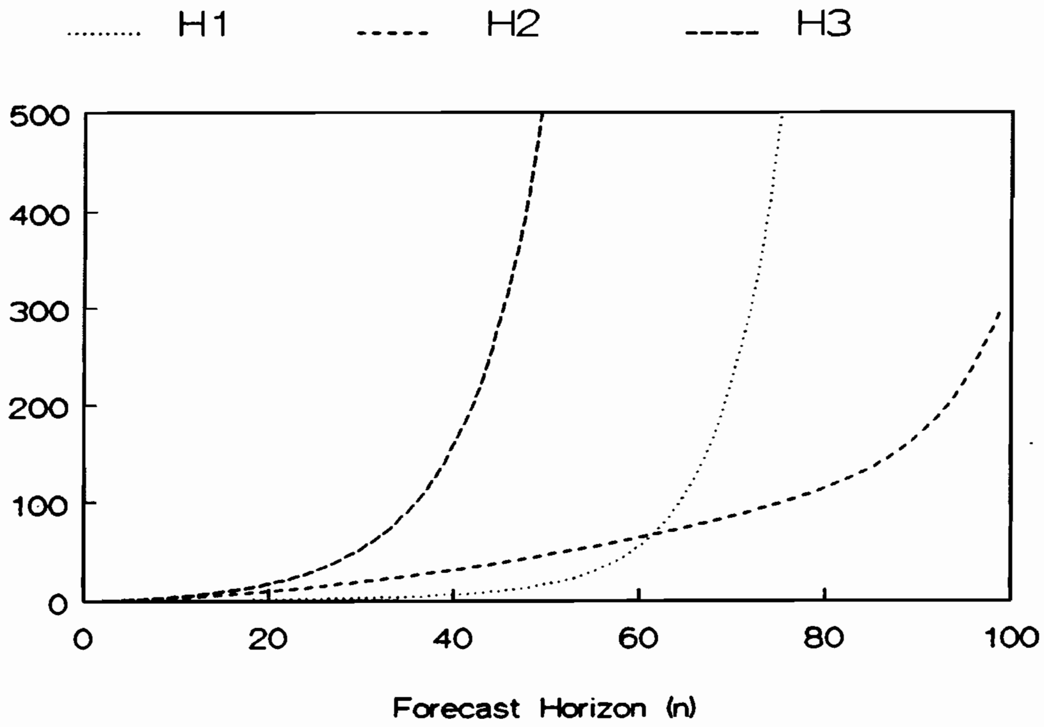
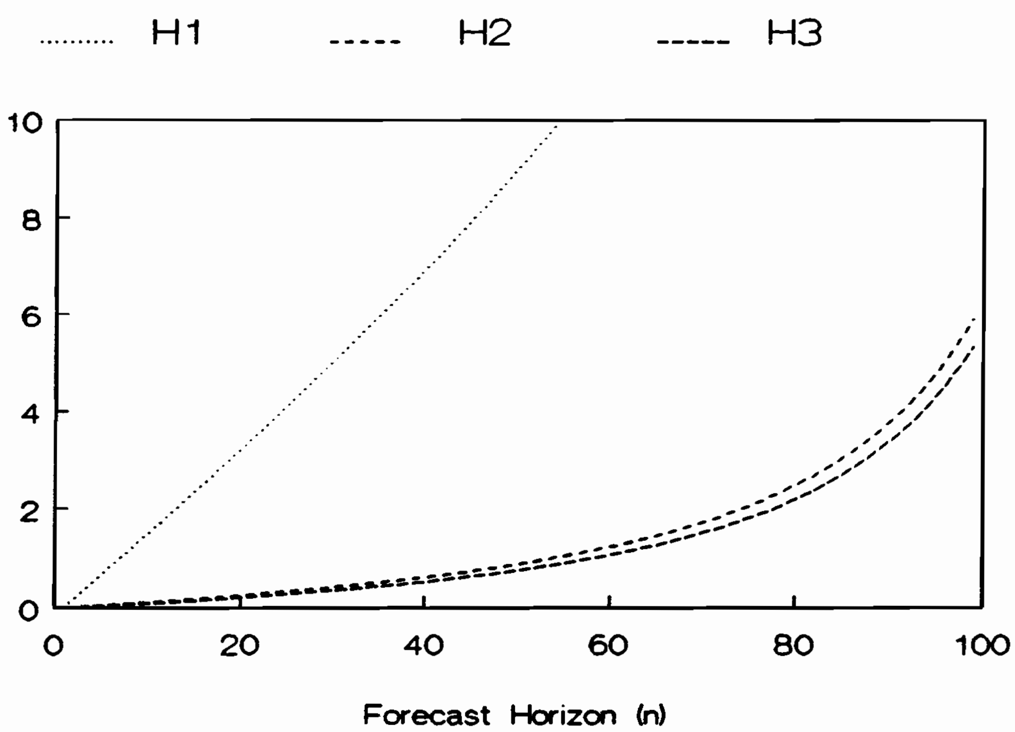


FIGURE 2: Simplified Loss Function

CPI, delta=10, uniform prior



Data Appendix

The data used in this paper are that of Nelson and Plosser (1982) updated to 1988 by Herman van Dijk. Primary data sources are listed in Schotman and van Dijk (1991b). All data are annual U.S. data. We take natural logs of all series except for the bond yield. The fourteen series are:

1) Real GNP (1909-1988).
2) Nominal GNP (1909-1988).
3) Real per capita GNP (1909-1988).
4) Industrial production (1860-1988).
5) Employment (1890-1988).
6) Unemployment rate (1890-1988).
7) GNP deflator (1889-1988).
8) Consumer Price Index (1860-1988).
9) Nominal wages (1900-1988).
10) Real wages (1900-1988).
11) Money stock (1889-1988).
12) Velocity (1869-1988).
13) Bond yield (1900-1988).
14) Common stock prices (1871-1988).

Prior Appendix

The Appendix discusses the selection of the bounded uniform priors for d_μ and d_β in (12). We use symmetric priors for all cases (A_1=-A_2 and B_1=-B_2) and set $A_2=\zeta_1 y_{q-1}$ and $B_2=\zeta_2(y_T-y_0)/T+1$. Since a level break of 10% is deemed to be highly unlikely, we set $\zeta_1=.10$ for all series except the bond yield and unemployment rate (for these series $\zeta_1=.4$). ζ_2 is more difficult to elicit. Looking at $(y_T-y_0)/T+1$, we set $\zeta_2=.1$ for real GNP, wages, employment, industrial production, money stock, and GNP per capita; $\zeta_2=.2$ for nominal GNP; $\zeta_2=.4$ for the Consumer Price Index and the GNP deflator; $\zeta_2=1$ for real wages, velocity, unemployment and common stock prices; and $\zeta_2=4$ for the bond yield. For no series is the posterior mean close to any of these boundaries.

References

Baillie, R. (1979), "The Asymptotic Mean Squared Error of Multistep Prediction from the Regression Model with Autoregressive Errors," *Journal of the American Statistical Association*, 74, 175-184.

Banerjee, A., Lumsdaine, R. and Stock, J. (1990), "Recursive and Sequential Tests of the Unit Root and Trend Break Hypotheses: Theory and International Evidence," NBER Working Paper No. 3510.

Choi, I. (1990), "Most U.S. Economic Time Series Do Not Have Unit Roots: Nelson and Plosser's (1982) Results Reconsidered," manuscript.

Chow, G. (1973), "Multiperiod Predictions from Stochastic Difference Equations by Bayesian Methods," *Econometrica* , 41, 109-118 and 796 (Erratum).

DeJong, D. and Whiteman, C. (1991a), "Reconsidering Trends and Random Walks in Macroeconomic Time Series," *Journal of Monetary Economics*, 28, 221-254.

DeJong, D. and Whiteman, C. (1991b), "The Temporal Stability of Dividends and Stock Prices: Evidence from the Likelihood Function," *American Economic Review, 81*, 600-617.

Dickey, D. and Fuller, F. (1979), "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 74, 427-431.

Dickey, J.M. and Chen, C.H. (1985), "Direct Subjective-Probability Modelling using Ellipsoidal Distributions," in *Bayesian Statistics* 2, ed. J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith, Amsterdam: North Holland.

Dreze. J. (1977), "Bayesian Regression Analysis Using Poly-t Densities," *Journal of Econometrics*, 6, 329-354.

Fang, K.-T., Kotz, S., and Ng, K. W. (1990), *Symmetric Multivariate and Related Distributions*, London: Chapman and Hall.

Koop, G. (1992), "'Objective' Bayesian Unit Root Tests," *Journal of Applied Econometrics*, 7, 65-82.

Koop, G. (1991), "Intertemporal Properties of Real Output: A Bayesian Approach," *Journal of Business and Economic Statistics*, 9, 253-265.

Koop, G., Osiewalski, J. and Steel, M.F.J. (1992), "Bayesian Long-Run Prediction in Time Series Models," Carlos III Working Paper 9210, Carlos III University, Madrid, Spain.

Koop, G. and Steel, M.F.J. (1991), "A Comment on: To Criticize the Critics: An Objective Bayesian Analysis of Stochastic Trends by Peter C. B. Phillips," *Journal of Applied Econometrics*, 6, 365-370.

Nelson, C. and Plosser, C. (1982), "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications," *Journal of Monetary Economics*, 10, 139-162.

Osiewalski, J. and Steel, M.F.J. (1993a), "Robust Bayesian Inference in Elliptical Regression Models," *Journal of Econometrics*, forthcoming.

Osiewalski, J. and Steel, M.F.J. (1993b), "Regression Models Under Competing Covariance Matrices: A Bayesian Perspective," manuscript.

Perron, P. (1989), "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis," *Econometrica*, 57, 1361-1402.

Phillips, P. (1991), "To Criticize the Critics: An Objective Bayesian Analysis of Stochastic Trends," *Journal of Applied Econometrics*, 6, 333-364.

Phillips, P. and Ploberger, W. (1991), "Time Series Modelling with a Bayesian Frame of Reference: I. Concepts and Illustrations," Cowles Foundation Discussion Paper No. 980.

Richard, J. and Tompa, H. (1980), "On the Evaluation of Poly-t Density Functions," *Journal of Econometrics*, 12, 335-351.

Sampson, M. (1991), "The Effect of Parameter Uncertainty on Forecast Variances and Confidence Intervals for Unit Root and Trend-stationary Time Series," *Journal of Applied Econometrics*, 6, 67-76.

Schotman, P. and van Dijk, H. (1991a), "A Bayesian Analysis of the Unit Root in Real Exchange Rates," *Journal of Econometrics*, 49, 195-238.

Schotman, P. and van Dijk, H. (1991b), "On Bayesian Routes to Unit Roots," *Journal of Applied Econometrics*, 6, 387-401.

Schwert, W. (1987), "Effects of Model Specification on Tests for Unit Roots in Macroeconomic Data," *Journal of Monetary Economics*, 20, 75-103.

Sims, C. (1988), "Bayesian Skepticism on Unit Root Econometrics," *Journal of Economic Dynamics and Control*, 12, 463-474.

Wago, H. and Tsurumi, H. (1990), "A Bayesian Analysis of Stationarity and the Unit Root Hypothesis," presented at the 6th World Congress of the Econometric Society.

Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*, New York: John Wiley.

Zivot, E. and Phillips, P. (1990), "A Bayesian Analysis of Trend Determination in Economic Time Series," manuscript.