THE UNIVERSITY OF MELBOURNE

Melbourne Institute Working Paper Series

Working Paper No. 20/06

Bayesian Inference for a Threshold Autoregression with a Unit Root

Penelope Smith

MELBOURNE INSTITUTE
of Applied Economic and Social Research

Bayesian Inference for a Threshold Autoregression with a Unit Root*

Penelope Smith
Melbourne Institute of Applied Economic and Social Research
The University of Melbourne

Melbourne Institute Working Paper No. 20/06

ISSN 1328-4991 (Print)

ISSN 1447-5863 (Online)

ISBN 0 7340 3221 8

October 2006

*I am grateful to Peter Summers for many helpful discussions and for comments from James Hamilton and Gael Martin and participants at the "Unit Root and Cointegration Testing" conference held at the University of the Algarve, Portugal September 29-October 1 2005.

Melbourne Institute of Applied Economic and Social Research
The University of Melbourne
Victoria 3010 Australia
***Telephone* (03) 8344 2100**
***Fax* (03) 8344 2111**
***Email* melb-inst@unimelb.edu.au**
***WWW Address* http://www.melbourneinstitute.com**

Abstract

A Bayesian approach to distinguishing between nonlinear and unit root behavior offers several practical advantages over equivalent frequentist procedures. Foremost among these advantages is the simplicity of the test. This paper compares the small sample power and size properties of a joint Bayesian test for a unit root and a threshold effect with Caner and Hansen's (2001) frequentist strategy. The results from Monte Carlo experiments indicate that the simpler Bayesian test performs at least as well as Caner and Hansen's procedure.

1 Introduction

This paper presents a simple Bayesian approach to jointly testing for nonlinear and nonstationary behavior in time series data. It is motivated by a growing body of evidence to suggest that many important macroeconomic variables which are commonly found to be nonstationary, such as real GDP growth and the unemployment rate, may also exhibit nonlinear behavior.[1] The Bayesian test is a simpler alternative to a similar test developed by Caner and Hansen (2001).

It is by now well known that parameter instability can matter a great deal when testing for unit roots. Perron (1989), for example, has noted that standard tests—such as the augmented Dickey-Fuller (ADF) and Phillips-Perron tests—have reduced power against stationary alternatives when there are breaks or switches in either the growth rate or the level of the data generating process (DGP).[2] However, unit root tests are usually conducted under the assumption that the DGP has a constant dynamic structure.

The irregular and complex nature of the testing problem possibly explains why practitioners rarely jointly test for nonlinear and non-stationary behavior in economic time series. As noted by Caner and Hansen (2001) standard Wald tests for nonlinearity have nonstandard asymptotic null distributions when unit roots are present in the DGP. This arises both from the presence of a nuisance parameter which is not identified under the null hypothesis—sometimes known as Davies's (1977) problem—and from discontinuities in the likelihood function associated with unit root processes.

This paper focuses on threshold nonlinearity, although the results can easily be extended to other nonlinear structures such as Markov Switching. Threshold models are, in their simplest form, piece-wise regression models with parameters that switch according to the value of a threshold variable. The threshold variable may be exogenous or some function of the dependent variable, as is the case for Tong's (1978) the self-exciting threshold autoregression.

The most commonly applied approach to jointly testing for unit roots and threshold

[1] See Koop and Potter (1999*b*) for a review of nonlinearity in macreconomic time series.

[2] See Dickey and Fuller (1981) and Phillips and Perron (1988).

effects was developed Caner and Hansen (2001). To overcome the theoretical difficulties associated with the testing problem, Caner and Hansen proposed a sequential testing strategy. This strategy involved first testing for a threshold effect when a unit root may be present and then testing for a unit root against the stationary, unit root and partial unit root alternatives. Critical values of their test statistics were obtained via a bootstrap procedure based on an asymptotic theory which they developed. This testing strategy has been adopted in applied work by Aretis, Cipollini and Fattouh (2002), Bec, Salem and MacDonald (2002), Henry and Shields (2004) and Basci and Caner (2005).

An equivalent test is relatively simple within the Bayesian framework. As described by Koop and Potter (1999*a*) the Bayesian approach is able to circumvent Davies' problem by adopting an informative prior for the threshold parameter. This makes it possible to average over all observable values of the threshold to find and compare the marginal likelihoods associated with the linear and threshold specifications. Furthermore, pre-testing biases are avoided as the Bayesian approach allows the practitioner to jointly test for unit roots and threshold effects.

The benefits of taking the Bayesian approach to inference are the motivation for the development of a simple and efficient Bayesian procedure to jointly test for a threshold effect and a unit root in a univariate time series. This test is based upon Bayesian tests for non-linearity outlined by Koop and Potter (1999*a*) and is closely related to Bayesian tests for unit roots in the presence of structural change developed by Phillips and Zivot (1994) and Marriot and Newbold (2000). Analytical results are available due to the use of a natural conjugate prior. This provides large gains to computational efficiency over Caner and Hansen's (2001) procedure. The performance of the test is benchmarked against the Caner and Hansen's testing strategy via a set of Monte Carlo experiments and by application to the male unemployment rate in the United States.

2 The testing problem

Caner and Hansen (2001) consider the following self-exciting threshold autoregression of order k ($SETAR(k)$):

$$\Delta y_t = \begin{cases} \mu_1 + \beta_1 t + \rho_1 y_{t-1} + \sum\limits_{j=1}^{k} \alpha_{1j} \Delta y_{t-j} + e_t \text{ if } Z_{t-1} < \lambda \\ \mu_2 + \beta_2 t + \rho_2 y_{t-1} + \sum\limits_{j=1}^{k} \alpha_{2j} \Delta y_{t-j} + e_t \text{ if } Z_{t-1} \geqq \lambda. \end{cases} \tag{1}$$

Here Z_{t-1} is the threshold variable constructed from lagged values of the data $\widetilde{Y}_T = \{y_t\}_{t=-k-1}^{T}$ and the threshold λ is unknown. Caner and Hansen specify the threshold variable to be $Z_t = y_t - y_{t-m}$ for some $m \geqq 1$. However, they note that the definition of Z_t is not central to the analysis and that their results extend to any strictly stationary and ergodic Z_t which is pre-determined and has a continuous distribution.

Define $\theta_i = [\mu_i, \rho_i, \beta_i, \alpha_i]$ to be the $1 \times n$ vector of the parameters for alternative threshold values $i \in \{1, 2\}$ and $x_t = [1, t, y_{t-1}, \Delta y_{t-1}, \ldots, \Delta y_{t-k}]'$. Also define $\Psi = [\theta_1, \theta_2, \sigma]$. Assuming independently, identically distributed Gaussian disturbances $e_t \sim i.i.d.N(0, \sigma^2)$ the likelihood function associated with (1) for a given observation t is

$$f_t\left(y_t | y_{t-1}^k; \Psi, \lambda\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{\left(\Delta y_t - \theta_1 x_t I_{\{Z_{t-1}<\lambda\}} - \theta_2 x_t I_{\{Z_{t-1}\geqq\lambda\}}\right)^2}{2\sigma^2}\right).$$

Here $y_{t-1}^k = y_{t-1}, \ldots, y_{t-k-1}$ and $I_{\{\cdot\}}$ is a binary indicator function taking on the value 1 if the identity $\{\cdot\}$ is true and 0 otherwise. For data series $\widetilde{Y}_T$ the likelihood function associated with (1) is

$$f_T(\widetilde{Y}_T; \Psi, \lambda) = \prod_{t=1}^{T} f_t\left(y_t | y_{t-1}^k; \Psi\right).$$

The series $\widetilde{Y}_T$ follows a threshold process if $\theta_1 \neq \theta_2$ and may have a unit root in either regime if $\rho_1 = 0$ or $\rho_2 = 0$. When testing for a threshold effect the relevant null hypothesis is that of a linear autoregression

$$H_0 : \theta_1 = \theta_2 \tag{2}$$

with the alternative

$$H_1 : \theta_1 \neq \theta_2. \tag{3}$$

Let Ψ_0 denote the parameter vector Ψ under the null hypothesis. It is easy to see that the likelihood function under the null hypothesis is independent of the threshold parameter λ so that $f_T(\widetilde{Y}_T; \Psi_0, \lambda) = f_T(\widetilde{Y}_T; \Psi_0)$. This is known as "Davies's (1977) problem". Davies noted that when a nuisance parameter is unidentified under the null hypotheses a consistent estimator of the nuisance parameter will not exist and the distributions of the usual test statistics—such as the Wald (W_T) Lagrange multiplier (LM_T) and likelihood ratio (LR_T) test statistics—will be non-standard. Furthermore, as described by Andrews and Ploberger (1994), Davies' problem places these test statistics outside the domain of standard asymptotic optimality results.

Andrews and Ploberger (1994) consider the problem of optimal testing when a nuisance parameter is unidentified under the null hypothesis against local alternatives of the form $f_T\left(\Psi_0 + B_T^{-1}h\right)$ for some $h \in \mathbb{R}$ and some nonrandom diagonal matrix B. They show that for particular integrable weighting functions $Q_\lambda(h)$ on values of h and a chosen weight function $J(\lambda)$ on values of a nuisance parameter λ, optimal tests of will be of the form

$$Exp\text{-}M_T = (1+c)^{-n/2} \int \exp\left(\frac{1}{2}\frac{c}{1+c}M(\lambda)\right) dJ(\lambda). \tag{4}$$

Here $M(\lambda)$ is one of the standard W_T, LM_T or LR_T test statistics given λ, and $c > 0$ is a scalar constant depending on the weight functions $Q_\lambda(\cdot)$. Andrews and Ploberger show that tests of this form will be optimal in the sense of having the greatest average weighted power over all tests φ_T of asymptotic level α. That is, they are the class of tests that maximize

$$\overline{\lim_{T\to\infty}} \int \Pr\left(\varphi_T \text{ rejects } H_0 \left| \Psi_0 + B_T^{-1}h\right., \lambda\right) dQ_\lambda(h) dJ(\lambda),$$

where $\overline{\lim_{T\to\infty}}$ is the average $\lim_{T\to\infty}$.

Despite Andrews and Ploberger's (1994) optimality result, Hansen's (1996) Sup-Wald test is the dominant approach to testing for non-linearity.[3] This is possibly because Andrews and Ploberger do not discuss how critical values for the $Exp\text{-}M_T$ test statistic should be obtained.

Hansen (1996) developed a solution to Davies' problem which approximated asymptotically correct critical values for the null hypothesis by bootstrapping the data under the assumption that (1) is a stationary process. For the testing problem described by (2) and (3) Hansen's Sup-Wald statistic is

$$\sup_{\lambda \in \Lambda} W_T(\lambda) = T\left(\frac{\widehat{\sigma}_0^2}{\widehat{\sigma}^2\left(\widehat{\lambda}\right)} - 1\right). \tag{5}$$

Here $\widehat{\sigma}^2(\widehat{\lambda})$ is an unrestricted estimate of the residual variance from (1) conditional upon the estimated value of the threshold $\widehat{\lambda}$ and $\widehat{\sigma}_0^2$ is an estimate of the residual variance under the null hypothesis. The least squares estimate of the threshold variable $\widehat{\lambda}$ is obtained as the value of λ that minimizes $\sigma^2(\lambda)$. In turn $\sigma^2(\lambda)$ is a vector of estimates of σ^2 for alternative vales of the threshold variable, obtained by sequential least squares over observed values of the threshold variable.

If a unit root is present in (1) then the asymptotic distribution of (5) will be discontinuous in the parameters at the boundary $\rho = 0$ and the bootstrap distribution of (5) will no longer be consistent.[4] To address this problem Caner and Hansen (2001) augmented Hansen's (1996) procedure to allow for a unit root in (1) by imposing a unit root on the bootstrap in order to locate the correct asymptotic distribution for this case. They show that the constrained bootstrap will be first order correct under the null hypothesis of a linear model (2) if the DGP has a unit root, but incorrect otherwise. As the order of integration of the DGP is usually unknown, Caner and Hansen advise that prudence be applied in the implementation of their test. They suggest that the p-values be bootstrapped with and without a unit root, with inference based on the larger, more conservative p-value.

[3] Hansen (1996) demonstrates that the Sup-Wald test has near optimal power against distant alternatives.

[4] Caner and Hansen (2001), p.1564.

Caner and Hansen (2001) also to developed a theory of inference for the complementary case of testing for a unit root when a threshold effect may be present. The relevant null hypothesis in this case is

$$H_0 : \rho_1 = \rho_2 = 0. \tag{6}$$

How the alternative hypothesis of a stationary process should be formulated is less obvious. This is because the region of the parameter space for which $SETAR$ models will be stationary and ergodic is not well understood. Chan and Tong (1985) show that when $k = 1$, the stationarity of the process depends upon the ρ_i parameters. In particular they demonstrate that the necessary and sufficient conditions for the stationarity of (1) are

$$\rho_1 < 0, \rho_2 < 0 \text{ and } (1 + \rho_1)(1 + \rho_2) < 1.$$

However, Chan and Tong note that there are no general conditions for the geometrical ergodicity of higher-order models.[5] Based on this analysis the first alternative hypothesis proposed by Caner and Hansen (2001) is that $\widetilde{Y}_T$ is stationary and ergodic:

$$H_1 : \rho_1 < 0 \text{ and } \rho_2 < 0. \tag{7}$$

and the second alternative is the partial unit root hypothesis:

$$H_2 : \begin{cases} \rho_1 < 0 \text{ and } \rho_2 = 0. \\ \qquad \text{or} \\ \rho_1 = 0 \text{ and } \rho_2 < 0. \end{cases} \tag{8}$$

In order to distinguish between hypotheses (7) and (8), Caner and Hansen (2001) propose two test statistics. The first is the two-sided Wald statistic from the least squares

[5]Chen and Tsay (1991) show that the dynamic properties of (1) are also dependent upon m. In addition to the conditions of Chan and Tong (1985) (1) must also satisfy

$$(1+\rho_1)^{s(m)}(1+\rho_2)^{t(m)} < 1 \text{ and } (1+\rho_1)^{t(m)}(1+\rho_2)^{s(m)} < 1$$

where $t(m)$ and $s(m)$ are nonnegative integers depending on m and odd and even numbers.

estimates of (1). The second is the sum of the standard test of the null hypothesis of a stationary process in each regime against the unrestricted alternative $\rho_1 = 0$ or $\rho_2 = 0$:

$$R_{2T} = t_1^2 + t_2^2.$$

Here t_1 and t_2 are the t ratios of the OLS estimates of ρ_1 and ρ_2. Caner and Hansen note that as the two alternative hypotheses are one-sided, this two-sided version of the test may have less power than the one sided test. Given that it is unclear how to form an optimal one-sided test when unit roots may be present,[6] Caner and Hansen suggest that a one-sided Wald test focusing on negative values of the OLS estimates of ρ_1 and ρ_2 should also be considered:

$$R_{1T} = t_1^2 1_{\{\widehat{\rho}_1 < 0\}} + t_2^2 1_{\{\widehat{\rho}_2 < 0\}}.$$

Rejection of (6) gives no indication of whether H_1 or H_2 is more likely. To overcome this difficultly Caner and Hansen (2001) suggest that $t_1$and t_2 be examined individually as tests of the stationary alternative H_1. The significance of only one of the t statistics would be consistent with the partial unit root hypothesis H_2.

In order to conduct these tests, sampling distributions of the tests under the null hypothesis H_0 are required. Caner and Hansen (2001) show that the asymptotic distributions of the test statistics for H_0 are different depending on whether or not there is a threshold effect in the DGP. This is because the null hypothesis of a unit root is consistent with both $\theta_1 = \theta_2$ and $\theta_1 \neq \theta_2$. Separate bootstrap procedures are therefore required for each case. Caner and Hansen find the unidentified threshold model to be slightly preferred in terms of size. Both the one sided and two sided tests are found to have good power when compared with the standard ADF test. The one sided test R_{1T} performs slightly better than the two sided test R_{2T}.

Caner and Hansen's (2001) approach provides some valuable advances in asymptotic theory and has been extended to the analysis of threshold cointegration by Gouveia and

[6]See Elliot, Rothenberg and Stock (1996)

Rodrigues (2004). However there are several disadvantages associated with this approach. As already noted, the test for a threshold effect is only nearly optimal, even when the DGP is stationary. Furthermore, the procedure is analytically complex and computationally expensive, while the sequential structure of the approach subjects it to a pretesting bias. Finally, Caner and Hansen's testing procedure is not easily generalizable to alternative nonlinear specifications, such as Markov switching processes.

3 Bayesian analysis

There is a clear correspondence between the Bayesian approach to inference when a nuisance parameter is unidentified under the null hypothesis and the approach of Andrews and Ploberger (1994). In fact Andrews and Ploberger affirm that their $Exp\text{-}M_T$ test statistic (4) may be given a Bayesian interpretation if the weight functions $J(\lambda)$ and Q_λ are interpreted as priors. They argue that their approach is preferred, however, as it avoids the necessity of placing a prior on the nuisance parameter under the null hypothesis. They also argue that their approach is computationally more efficient.

Bayesian methods can be computationally complex when a closed-form solution to the posterior distributions of the model parameters does not exist or when the posterior distributions are not known. This is usually the case for more complex nonlinear specifications, such as Markov switching models, which must be estimated using numerical methods of integration. However, these methods are arguably no more complex than the bootstrapping techniques proposed by Hansen (1996) and Caner and Hansen (2001). Furthermore, as will be outlined in the following section of this paper, analytical solutions to the posterior distribution can be found for model (1) when normal, natural conjugate priors are adopted. Under these priors the Bayesian approach to testing for non-linearity will be asymptotically equivalent to Andrews and Ploberger's optimal $Exp\text{-}M_T$ test statistic.

Andrews and Ploberger (1994) also note that Bayesian posterior odds ratios are asymptotically equivalent to the $Exp\text{-}LM_T$ test, but do not describe why, or elucidate the conditions under which this will be true. This result follows from Andrews (1994). Adopting Andrew's notation, let $\pi \in (0,1)$ denote the prior probability assigned to the null hypothesis, and let Q_μ denote the prior distribution function for the data. Define $\mu(\cdot)$ to be a probability distribution on $\mathbb{R}^+ = \{r \in \mathbb{R} : r \geq 0\}$ that depends on the prior Q_μ and let $g(\cdot)$ be a function for $\kappa \in \mathbb{R}$. The following strictly increasing function of $M = W_T, LM_T$ or LR_T can be defined:

$$PO(M,\mu) = \frac{1-\pi}{\pi} \int \exp\left(\frac{-r^2}{2}\right) g\left(Mr^2\right) d\mu\left(r^2\right).$$

The posterior odds statistic associated with the null and alternative hypotheses will be

$$PO_T(Q_\mu) = \frac{1-\pi}{\pi} \int f_T\left(\boldsymbol{\Psi_0} + T^{-1/2}h\right) d\frac{Q_\mu(h)}{f_T(\boldsymbol{\Psi_0})}.$$

Andrews (1994) shows that if the priors and the data satisfy certain regularity conditions, then under both the null (2) and the alternative (3)

$$PO_T(Q_\mu) - PO(M_T,\mu) \xrightarrow{p} 0. \tag{9}$$

Andrews notes that one prior which satisfies these conditions is the multivariate normal distribution with a variance proportional to a scalar, c: $Q_\mu \sim N(0, c\Sigma)$. Andrews shows that if this is the case then $\mu_c = \sqrt{c\chi_n^2}$ and

$$PO(M,\mu) = \frac{1-\pi}{\pi}(1+c)^{-n/2} \exp\left[\frac{1}{2}\frac{c}{(1+c)}M\right]$$

for the two sided testing case where $n \geq 1$. This is Andrews and Ploberger's Exp-M_T statistic (4) for $J(\lambda)$ equal to a point mass at λ_0 and $\pi = \frac{1}{2}$.

More generally, if we define

$$PO_T(Q_\mu, \lambda) = \frac{1-\pi}{\pi} \int\int f_T\left(\boldsymbol{\Psi_0} + T^{-1/2}h, \lambda\right) \frac{dQ_\mu(h)\, dJ(\lambda)}{f_T(\boldsymbol{\Psi_0})}$$

and

$$PO(M,\mu,\lambda) = \frac{1-\pi}{\pi}(1+c)^{-n/2}\int \exp\left[\frac{1}{2}\frac{c}{(1+c)}M(\lambda)\right]dJ(\lambda)$$

then because (9) will hold for all points $\lambda \in \Lambda$, it follows that under the regularity conditions outlined Andrews (1994) and for $\pi = \frac{1}{2}$

$$PO_T(Q_\mu,\lambda) \xrightarrow{p} PO(M,\mu,\lambda)$$

or

$$PO_T \xrightarrow{p} Exp\text{-}LM_T. \tag{10}$$

Recall that Andrews and Ploberger (1994) have shown the $Exp\text{-}LM_T$ to be asymptotically optimal when a nuisance parameter is unidentified under the null. The significance of this result is that under certain regularity conditions and for particular class of priors, the Bayesian PO_T test will be asymptotically equivalent to the optimal $Exp\text{-}LM_T$ test. This suggests that the Bayesian approach to testing for threshold effects will be superior to Hansen's (1996) and Caner and Hansen's (2001) approaches.

The Bayesian approach to testing for a unit root in the presence of non-linearity has received relatively little attention. However, as discussed by Phillips and Zivot (1994) for the case of unknown break points, Bayesian methods avoid a pretesting bias which arises from the fact that the delay lag of the threshold parameter m is unknown and must be estimated before tests can be conducted. Phillips and Zivot also emphasize that the asymptotic distribution theory associated with these test statistics is complex and often at odds with the corresponding finite sample distributions. Furthermore, Bayesian techniques allow joint testing for non-linearity and unit roots through the comparison of posterior model probabilities.

The correspondence between Bayesian and frequentist procedures for inference breaks down when unit roots are present in the DGP. This was initially noted by Sims (1988), who argued that a flat prior Bayesian approach to inference was simpler than the frequentist approach, avoiding complications such as disjoint confidence regions or uncertainty regarding the specification of the trend.[7]

Sims's (1988) argument was largely based on the assertion that inference based on the likelihood principle will be unchanged, even if the DGP is a random walk. The likelihood principle, expressed simply, is that only observed sample outcomes are relevant to statistical inference. As such, all information relevant for inference is contained in the likelihood. Sims observed that the likelihood for the $AR(1)$ model is normal, whether or not the data are stationary. This point has been powerfully illustrated by Sims and Uhlig (1991) and shown to be more generally applicable by Kim (1994, 1998). Kim demonstrated that subject to a fairly general set of conditions, the shape of the likelihood will be asymptotically normal.[8]

Kim's (1998) result of asymptotic posterior normality is useful in the present context because it carries the implication that the posterior probabilities involved in testing for non-linearity will be unchanged whether or not there is a unit root in the DGP. In contrast, the asymptotic properties of the alternative $Exp\text{-}LM_T$ and Sup-Wald statistics under similar conditions are unknown. When a unit root is present, the second derivative of the log of the likelihood function $L''(\theta)$ will be discontinuous. Furthermore, $L''(\theta)$ will no longer converge to a constant. These properties result in non-standard asymptotic distributions of both the $Exp\text{-}LM_T$ and the Sup-Wald test statistics. It is therefore unclear whether or not the $Exp\text{-}LM_T$ test for a threshold effect will retain its optimality properties under the null hypothesis of a unit root. In addition, a unit root in the DGP will violate several of the regularity conditions required for (10) to hold. The Bayesian

[7] Differences between the classical and Bayesian approaches have been emphasized at an applied level by DeJong and Whiteman (1991). These authors adopted a Bayesian flat prior methodology to revisit Nelson and Plosser's data set and found far less support for unit roots than standard classical tests had previously suggested. In particular, they rejected the unit root hypothesis for stock-price data, dividend data and U.S. real GNP.

As an aside, Phillips (1991) also noted that Bayesian unit root tests are able to avoid the complexity of the frequentist approach, but warns against the use of flat priors. He shows that flat priors are informative in the sense that they down-weight the non-stationary region of the posterior distribution. In order to develop an 'objective', or non-informative Bayesian unit test, Phillips formulated a framework for the application of Jeffreys' prior on the autoregressive parameters. Phillips found that the results from Bayesian unit root tests accord far more closely with the frequentist alternative under this prior.

[8] It is useful to note that one prior for which asymptotic normality will not hold is Phillips (1991) version of Jeffreys' prior. As discussed by Kim (1998), Phillips' prior is dependent on sample size and diverges to infinity as the sample size becomes large, resulting in a very non-normal asymptotic posterior distribution.

test, however, will retain its form when a nuisance parameter is unidentified under the null hypothesis, with no discontinuities, even when there may be a unit root in the DGP.

These results form a strong motivation for the adoption of a Bayesian testing strategy with priors that conform to $Q_\mu \sim N(0, c\Sigma)$ whenever a nuisance parameter is unidentified under the null hypothesis, particularly when the DGP may be non-stationary. The remainder of this paper is devoted to the development of such a strategy for Caner and Hansen's (2001) testing problem.

4 A Bayesian test

Assume that the threshold λ is known. Define $x_t = (1, t, y_{t-1}, \Delta y_{t-1}, ..., \Delta y_{t-k})$ and let $x_t^1 = x_t' I_t$ and $x_t^2 = x_t'(1 - I_t)$ where

$$I_t = \begin{cases} 1 \text{ if } Z_{t-1} < \lambda \\ 0 \text{ otherwise.} \end{cases}$$

Define X_1 and X_2 to be the vectors $x^1 = [x_1^1, x_2^1, ..., x_T^1]'$ and $x^2 = [x_1^2, x_2^2, ..., x_T^2]'$ respectively, with elements re-ordered according to the observed value of the threshold variable Z_t associated with each x_t. This implies that if there are g observations for which $Z_{t-1} \geq \lambda$, then the last g observations of X_1 will be zero and the first $T - g$ observations of X_2 will be zero. Finally, define $X(\lambda)$ to be the $T \times (4 + 2k)$ matrix $[X_1 \ X_2]$. Using this notation (1) can be expressed as

$$\Delta Y(\lambda) = X(\lambda)\theta + e.$$

Here $\theta = (\theta_1', \theta_2')$ and $\Delta Y(\lambda)$ and e are the vectors $(\Delta y_1, \Delta y_2, ..., \Delta y_T)'$ and $(e_1, e_2, ..., e_T)'$ respectively, also re-ordered by the observed value of the threshold variable Z_t at time t.

Based on the hypotheses described in section 2, there are seven competing models of interest defined by restrictions on θ and $\rho_1, \rho_2 \in \theta$:

$M_1 : \theta_1 = \theta_2$ and $\rho_1 = \rho_2 = 0$: a linear autoregression with a unit root

$M_2 : \theta_1 = \theta_2$ and $\rho_1 = \rho_2 < 0$: a stationary linear autoregression

$M_3 : \theta_1 \neq \theta_2$ and $\rho_1 = \rho_2 = 0$: a nonlinear autoregression with a unit root

$M_4 : \theta_1 \neq \theta_2$ and $\rho_1 = \rho_2 < 0$: a stationary nonlinear autoregression

$M_5 : \theta_1 \neq \theta_2$ and $\rho_1 \neq \rho_2 < 0$: a stationary nonlinear autoregression

$M_6 : \theta_1 \neq \theta_2$ and $\rho_1 = 0$ and $\rho_2 < 0$: a unit root in the lower regime

$M_7 : \theta_1 \neq \theta_2$ and $\rho_1 < 0$ and $\rho_2 = 0$: a unit root in the upper regime

The likelihood function associated with (1) conditional upon hypotheses M_j, $j = \{1, 2, \ldots 7\}$ is

$$f_T(\widetilde{Y}_T; \Psi, \lambda | M_j) = \frac{1}{(\sigma\sqrt{2\pi})^T} \exp\left\{ -\frac{(\Delta Y(\lambda) - X(\lambda)\theta)' (\Delta Y(\lambda) - X(\lambda)\theta)}{2\sigma^2} \right\}. \qquad (11)$$

The conditional marginal likelihood $f_T(\widetilde{Y}_T | M_j)$ is obtained by integrating the joint density $p(\theta, \sigma, \lambda)$ over θ_j, σ_j and λ:

$$\begin{aligned} f_T(\widetilde{Y}_T | M_j) &= \iiint p(\theta, \sigma, \lambda) d\theta_j d\sigma_j d\lambda \qquad (12) \\ &= \iiint f_T(\theta, \lambda, \sigma) p(\theta, \sigma | \lambda) p(\lambda) d\theta_j d\sigma_j d\lambda \\ &= \int f_T(\widetilde{Y}_T | M_j, \lambda) d\lambda. \end{aligned}$$

The analytical evaluation of integrals like (12) is often not possible for nonlinear time series models. It must therefore be approximated using some numerical integration technique. However, analytic solutions are available for (12), conditional on λ, if independent natural conjugate priors are adopted for ρ_1, ρ_2 and σ.

Convention dictates that the prior densities of the parameters of the model be proper and not too diffuse. This is because Bayesian estimation requires the integration of $f_T(\widetilde{Y}_T|M_j,\lambda)$ over the entire domain of the nuisance parameter λ. When non-informative or improper priors are adopted the Bayes factors will tend to favour the linear model.[9]

A standard normal-gamma prior is adopted for θ and σ:

$$
\begin{aligned}
p(\theta|\sigma) &= (2\pi)^{-n/2}\,\sigma^{-n}\,|A|^{1/2}\exp\left\{-\frac{1}{2\sigma^2}\left(\theta-\overline{\theta}\right)'A\left(\theta-\overline{\theta}\right)\right\} \\
p(\sigma) &= \frac{2}{\Gamma\left(\overline{v}/2\right)}\left(\frac{\overline{vs}^2}{2}\right)^{\overline{v}/2}\frac{1}{\sigma^{\overline{v}+1}}\exp\left\{-\frac{\overline{vs}^2}{2\sigma^2}\right\}. \\
p(\theta,\sigma) &= p(\theta|\sigma)p(\sigma) \\
&= C\sigma^{-n-\overline{v}-1}\exp\left\{-\frac{1}{2\sigma^2}\left[\overline{vs}^2+\left(\theta-\overline{\theta}\right)'A\left(\theta-\overline{\theta}\right)\right]\right\}
\end{aligned}
$$

where $\overline{\theta}$ is the prior mean of θ, $cov[\theta|\sigma]=\sigma^2A^{-1}$ and $E[\sigma^2]=\overline{vs}^2/(\overline{v}-2)$ and $C=(2\pi)^{-T/2}\,|A|^{1/2}\,\frac{2}{\Gamma(\overline{v}/2)}\left(\frac{\overline{vs}^2}{2}\right)^{\overline{v}/2}$. This prior fulfills the conditions described by Andrews (1994) necessary for $PO_T - Exp\text{-}LM_T \xrightarrow{p} 0$.

Under these priors the conditional marginal likelihood will be:

$$
f_T(\widetilde{Y}_T|M_j)=\int z\left(\frac{|A_j|}{\left|A_j+X_j\left(\lambda\right)'X_j\left(\lambda\right)\right|}\right)^{\frac{1}{2}}\left(\overline{vs}^2+v\widehat{\sigma}_j^2+Q_j\right)^{-(\overline{v}+T)/2}d\lambda \qquad (13)
$$

where

$$
\begin{aligned}
Q_j &= (\widehat{\theta}_j-\overline{\theta}_j)'[A_j^{-1}+(X_j'\left(\lambda\right)X_j\left(\lambda\right)^{-1})]^{-1}(\widehat{\theta}_j-\overline{\theta}_j) \\
v &= T-n \\
z &= \frac{\Gamma\left[\left(\overline{v}+T\right)/2\right]\left(\overline{vs}^2\right)^{\overline{v}/2}}{\Gamma\left(\overline{v}/2\right)\pi^{T/2}}.
\end{aligned}
$$

[9] For more information and references see Kass and Raftery (1994) section 5 and Koop and Potter (1999*a*).

See Judge, Hill, Griffiths, Lütkepoh and Lee (1985, p.129) for more details.

A difficulty arises here as we only have information about λ at discrete and unevenly spaced intervals as determined by observations on the threshold variable Z_t. However, Koop and Potter (2000) note that while λ is continuous, its effect on the likelihood (11) is the same as if it was discrete. This is because the possible values of λ can be restricted to the observed data points on Z_t. Thus the likelihood (11), marginal likelihood function (13) and the posterior distribution of θ will be flat between the observed data points for λ – in this case each observation on Z_t. This allows (13) to be calculated by obtaining the conditional marginal likelihood $f_T(\widetilde{Y}_T|M_j, \lambda)$ for each observation on Z_t and finding the weighted sum across the T observations, with weights determined by the relative size of the interval between each Z_t. Then one simply has to divide these by the height of the integrating constant, which in this case is the sum of the value of the marginal likelihoods calculated for each observed Z_t.

Following Koop and Potter a continuous uniform prior is adopted for the threshold parameter:

$$p(\lambda) = \frac{1}{\lambda_u - \lambda_l}. \tag{14}$$

Here λ_u is the upper allowable value for the prior and λ_l is the lower allowable value for the prior. It is necessary to choose λ_u and λ_l to be within the range of observed threshold variables Z_t to ensure that neither X_1 or X_2 are null vectors and that analytical solutions for (13) will exist. This is similar to Caner and Hansen's (2001) restriction of λ to the interval $[\lambda_l, \lambda_u]$ such that $\Pr(Z_t \leq \lambda_l) = \pi_1$ and $\Pr(Z_t \leq \lambda_u) = \pi_2$ where $0 > \pi_1, \pi_2 > 1$.

Caner and Hansen's (2001) and Koop and Potter's (1999*a*) recommendations are followed in treating π_1 and π_2 symmetrically so that $\pi_2 = 1 - \pi_1$. This equivalent to restricting λ so that neither regime accounts for more than $100\pi_1\%$ of observations on Z_t. Caner and Hansen recommend that π_1 be selected carefully so as to ensure that there will enough observations to identify the parameters. For the Bayesian test π_1 and π_2 need to be selected so that there are enough observations in each regime for the data to be reasonable informative about the posterior distributions of regime dependent parameters.

Let Λ be the vector $(Z_1, Z_2, ..., Z_T)'$ sorted into ascending order and let Λ_i denote the

i^{th} element of Λ. Under (14) the marginal conditional likelihood associated with (1) can be calculated as

$$f_T(\widetilde{Y}_T|M_j) = \sum_{i=\pi_1 T}^{\pi_2 T} \left\{ \frac{(\Lambda_{i+1} - \Lambda_i)}{\sum_{i=\pi_1 T}^{\pi_2 T}(\Lambda_{i+1} - \Lambda_i)} \right\} f_T(\widetilde{Y}_T|M_j, \lambda = \Lambda_i). \tag{15}$$

This is similar to Phillips and Zivot's (1994) Bayesian approach to inference on unit roots when break points in the DGP are unknown. Zivot and Phillips place a uniform prior on the position of the break point and produce posterior distributions for making unconditional inferences about the parameters of the model by averaging the marginal likelihood of the data over all possible break points.[10]

As X_1 and X_2 are linearly independent with a prior covariance of zero, the posterior distributions of ρ_1 and ρ_2 will be distributed as independent, univariate-t random variables. This property can be exploited to find weights for the conditional marginal likelihoods of model M_2 and models M_4 through to M_7 so that only the non-explosive regions of ρ_1 and ρ_2 are considered.

5 Empirical performance of the Bayesian test

This section compares the small sample performance of the Bayesian test to Caner and Hansen's (2001) bootstrap procedure. To do so the size and power of the bootstrap tests are compared with Bayesian size and power equivalents. The Bayesian 'power equivalent' is defined to be the percentage of Monte Carlo draws for which the posterior model probability associated with the null hypothesis is smaller than some constant c: $\Pr(H_0|\widetilde{Y}_T) \leq c$. The 'size equivalent' is defined analogously.

As the Bayesian paradigm places emphasis on the strength of evidence in favour of the null hypothesis, many Bayesian statisticians, such as Gelman, Carlin, Stern and Rubin (1995), would argue that such a strict decision rule imposes an artificial dichotomy between the null and alternative hypotheses. However if a comparison between the

[10] Marriot and Newbold (2000) adopt a similar strategy.

performance of Bayesian and frequentist inferential procedures is to be made, a decision rule needs to be put in place.

It is well known that the posterior model probability associated with a null hypothesis can be very different from the p-value of a frequentist test.[11] As discussed by Berger and Sellke (1987) the prior model probability attached to a point null hypothesis must be relatively small in order to produce a posterior model probability $\Pr(H_0|\widetilde{Y}_T) \leqq 0.05$.[12] Put another way, posterior odds ratios are often more conservative than p-values in their propensity to 'reject' the null hypothesis. As it is unclear precisely what decision rule should be employed when comparing the small sample properties of Bayesian and frequentist inferential procedures, the results of the experiments for both $c = 0.05$ and $c = 0.5$ are reported. $\Pr(H_0|\widetilde{Y}_T) = 0.05$ is equivalent to a Bayes factor of 19, which represents strong evidence against the null on Kass and Raftery's scale. Alternatively $\Pr(H_0|\widetilde{Y}_T) = 0.5$ corresponds to a Bayes factor of 1 and represents only weak evidence against the null.

Kass and Raftery (1994) stress that the results of posterior odds are quite sensitive to how the priors are specified for the parameters of interest. As emphasized by Koop and Potter (1999a), the specification of diffuse priors for any parameter in θ will tend to bias inference towards the linear model. Similarly, a very diffuse prior on the ρ_i parameters will tend to bias posterior evidence towards non-rejection of the unit root hypothesis.

The prior means of each θ_i are set to zero, which is equivalent to a prior of a linear process with a unit root. The prior variance of the ρ_i and α_i parameters are set to 1, while the priors for the intercepts, μ_i are more diffuse, with prior variances of 1/4. The prior probability assigned to each model $M_1, ..., M_7$ is distributed evenly. The priors on the upper and lower bound of the threshold parameter, λ_l and λ_u, are specified such that λ must lie within the middle 70% of the observed Z_t.

Monte Carlo experiments identical to those described in Caner and Hansen (2001) are conducted. Selected results of Caner and Hansen's Monte Carlo experiments are

[11] See, for example, Berger and Sellke (1987), Edwards, Lindman and Savage (1963) and Dickey (1977)

[12] Although Casella and Berger (1987) and DeGroot (1973) have shown that when comparing a one-sided null hypothesis on a scalar parameter with a one-sided alternative, p-values and posterior model probabilities may be directly comparable.

provided for comparison. Due to the computational expense incurred with Caner and Hansen's methods, size and power results for Caner and Hansen's procedures have been taken directly from their article and reproduced here. In each experiment the delay lag of the threshold parameter m is set to equal 1, $\lambda = 0$ and $k = 1$.

5.1 Power of the threshold tests

Caner and Hansen (2001) examine the power of their threshold test by investigating thresholds in each of the coefficients of (1) individually. They measure the empirical power of their test by the empirical rejection rates from 2000 replications of each experiment at the nominal 5% level. Following Caner and Hansen the sample size for the data generated in each experiment is $T = 100$ and 2000 replications of each experiment are performed.

The first Monte Carlo experiment allows for a switching intercept only, with data simulated from the DGP

$$\Delta y_t = \begin{cases} \mu_1 + \rho y_{t-1} + \alpha \Delta y_{t-1} + e_t \text{ if } Z_{t-1} < \lambda \\ \mu_2 + \rho y_{t-1} + \alpha \Delta y_{t-1} + e_t \text{ if } Z_{t-1} \geqq \lambda \end{cases} \tag{16}$$

$$e_t \sim i.i.d.N(0,1)$$

with $\alpha = 0.5$, $\lambda = 0$ and ρ taking on the values $\{0, -0.05\}$. The magnitude of the threshold effect is controlled by the difference $\Delta\mu = \mu_2 - \mu_1$ with $\mu_1 = -\mu_2$.

The second experiment allows for a threshold effect in the coefficient on y_{t-1}:

$$\Delta y_t = \begin{cases} \mu + \rho_1 y_{t-1} + \alpha \Delta y_{t-1} + e_t \text{ if } Z_{t-1} < \lambda \\ \mu + \rho_2 y_{t-1} + \alpha \Delta y_{t-1} + e_t \text{ if } Z_{t-1} \geqq \lambda \end{cases} \tag{17}$$

$$e_t \sim i.i.d.N(0,1)$$

with $\mu = 1$, $\lambda = 0$ and $\alpha = 0.5$. Here the magnitude of the threshold effect is controlled by $\Delta\rho = \rho_2 - \rho_1$.

In the third experiment there is a threshold effect in the autoregressive parameters α_1 and α_1:

$$\Delta y_t = \begin{cases} \mu + \rho y_{t-1} + \alpha_1 \Delta y_{t-1} + e_t \text{ if } Z_{t-1} < \lambda \\ \mu + \rho y_{t-1} + \alpha_2 \Delta y_{t-1} + e_t \text{ if } Z_{t-1} \geqq \lambda \end{cases} \tag{18}$$

$$e_t \sim i.i.d.N\,(0,1)\,.$$

The magnitude of the threshold effect is defined as $\Delta\alpha = \alpha_2 - \alpha_1$ and $\alpha_1 = -\Delta\alpha/2$. The intercept $\mu = 1$, $\lambda = 0$ and ρ takes on the values {0,-0.05}.

The results of these experiments are presented in Table 1. The power of the Bayesian test to reject the linear model in the presence of a threshold effect appears to compare quite favorably with Caner and Hansen's (2001) procedure. In each experiment the power of test is monotonically increasing in the size of the threshold effect and reasonably large, even against quite conservative alternatives for the Bayesian 5% test. The Bayesian 50% test dominates Caner and Hansen's test for all experiments except the one for which $\rho = 0$ and $\Delta\mu = 2$.[13]

Table 1: Empirical power of the threshold tests

	Bayesian test (5%)			Bayesian test (50%)			Sup-Wald test		
	$\Delta\mu$								
	0.2	1.0	2.0	0.2	1.0	2.0	0.2	1.0	2.0
$\rho = -0.05$	0.8	14.9	83.3	13.8	56.2	98.4	5.4	38.9	98.2
$\rho = 0$	0.5	12.1	70.7	9.9	49.5	96.3	5.2	35.7	97.9
	$\Delta\ \rho$								
	-0.05	-0.1	-0.2	-0.05	-0.1	-0.2	-0.05	-0.1	-0.2
$\rho_1 = -0.05$	4.4	19.0	51.9	33.6	64.4	88.1	16.5	43.0	73.8
$\rho_1 = 0$	65.1	80.7	88.0	92.8	97.3	98.8	82.1	93.1	95.5
	$\Delta\ \alpha$								
	0.5	1.0	1.9	0.5	1.0	1.9	0.5	1.0	1.9
$\rho = -0.05$	6.5	44.8	99.7	39.5	85.8	99.9	15.7	57.5	99.6
$\rho = 0$	1.7	26.4	99.5	16.7	67.2	99.7	4.7	10	34.4

[13] As a qualification it should be kept in mind that a 51% posterior model probability in favour of the threshold model represents only weak evidence against the null hypothesis

5.2 Size of the threshold tests

Caner and Hansen (2001) examine the size of their threshold test by simulating the data under the null hypothesis of a linear AR(1) process:

$$\Delta y_t = \mu + \rho y_{t-1} + \alpha \Delta y_{t-1} + e \tag{19}$$
$$e_t \sim i.i.d.N\,(0, 1)\,.$$

Here ρ is varied among the values $\{0, -0.05, -0.15, -0.25\}$ and α alternates between $\{0.5\}$ and $\{-0.5\}$. Following Caner and Hansen, the sample size of each experiment is set to $T = 100$ and 10000 replications of each experiment are performed. Results of this experiment are presented in Table 2. They are reported as the percentage of p-values which are smaller than the nominal size for Caner and Hansen's Sup-Wald test and the percentage of simulations for which posterior model probability of a linear model is less than 0.05 and 0.5 for the Bayesian test.[14]

The results presented in Table 2 illustrate the more conservative nature of the Bayesian test. In particular, the percentage of rejections of the (true) null hypothesis are much smaller for the 5% Bayesian test than for Caner and Hansen's (2001) Sup-Wald test.

Table 2: Emprical size of the threshold tests

	Bayesian test 5 %			
	$\rho = -0.25$	$\rho = -0.15$	$\rho = -0.05$	$\rho = 0$
$\alpha = -0.5$	0.9	0.7	0.5	0.4
$\alpha = 0.5$	0.6	0.6	0.6	0.3
	Bayesian test 50 %			
$\alpha = -0.5$	15.8	15.2	12.8	10.8
$\alpha = 0.5$	13.2	14.1	12.3	7.7
	Sup-Wald test			
$\alpha = -0.5$	3.8	5.5	5.1	4.8
$\alpha = 0.5$	4.0	5.0	4.9	4.2

[14] Caner and Hansen found their constrained bootstrap procedure to perform slightly better in terms of size than the unconstrained bootstrap. Therefore, Caner and Hansen's results for the constrained bootstrap are reproduced in Table 2.

5.3 Power of the unit root tests

Caner and Hansen (2001) examine the power of nominal 5% tests with data simulated from the DGP

$$\Delta y_t = \begin{cases} \mu_1 + \rho_1 y_{t-1} + e_t \text{ if } Z_{t-1} < \lambda \\ \mu_2 + \rho_2 y_{t-1} + e_t \text{ if } Z_{t-1} \geqq \lambda \end{cases} \tag{20}$$

$$e_t \sim i.i.d.N\,(0,1)\,.$$

Following Caner and Hansen the sample size of each experiment is set to $T = 100$ and 1000 replications of each experiment are performed.

In the first experiment the DGP is stationary and linear in ρ but subject to a threshold effect in the intercept. Specifically $\rho_1 = \rho_2$ is varied among {-0.05,-0.10,-0.15} and $\Delta\mu$ is varied among {0,1,2,3}. The second Monte Carlo experiment examines the case of a partial unit root. This time $\rho_1 = 0$ and ρ_2 is varied among {-0.05,-0.10,-0.15}. In the final experiment the experimental data is simulated under the condition that $\rho_1 = -0.05$. Results are presented in Table 3.

The Bayesian 50% test outperforms Caner and Hansen's (2001) test for all experiments and, strikingly, the more conservative Bayesian 5% test outperforms Caner and Hansen's test for half of the experiments. The advantage of the Bayesian test over the R_{1t} test increases in the magnitude of threshold effect.

5.4 Size of the unit root tests

Caner and Hansen (2001, p. 1574) examine the size of their unit root test by simulating the data under the null hypothesis of a unit root with a threshold effect in the intercept

Table 3: Empirical power of the unit root tests

		Bayesian 5% test			Bayesian 50% test			R_{1t} test		
	ρ_2 :	-0.05	-0.1	-0.15	-0.05	-0.1	-0.15	-0.05	-0.1	-0.15
						$\rho_1 = \rho_2$				
	0	4	12	53	36	68	90	8	15	28
$\Delta\mu$:	1	7	30	64	46	88	99	14	38	62
	2	26	81	98	79	100	100	29	76	95
	3	76	100	100	98	100	100	54	96	99
						$\rho_1 = 0$				
	0	3	6	11	26	40	63	6	12	23
$\Delta\mu$:	1	5	18	45	35	70	90	11	28	54
	2	21	73	94	64	97	100	17	64	92
	3	63	96	99	89	98	100	41	90	97
						$\rho_1 = -0.05$				
	0	4	8	13	33	51	69	8	12	20
$\Delta\mu$:	1	7	18	41	50	73	92	14	28	47
	2	25	64	90	79	97	100	29	58	87
	3	77	98	99	98	100	100	54	89	99

only:

$$\Delta y_t = \begin{cases} \mu_1 + \alpha \Delta y_{t-1} + e_t \text{ if } Z_{t-1} < \lambda \\ \mu_2 + \alpha \Delta y_{t-1} + e_t \text{ if } Z_{t-1} \geqq \lambda \end{cases} \tag{21}$$

$$e_t \sim i.i.d.N(0,1).$$

The autoregressive coefficient α is varied among $\{-0.5, -0.2, 0, 0.2, 0.5\}$ and $\Delta\mu = \mu_2 - \mu_1$ is varied among $\{0, 1, 2, 3\}$ with $\mu_1 = -\mu_2$. Following Caner and Hansen, the sample size for the data generated in each experiment is $T = 100$ and 1000 replications of each experiment are performed.

The results of experiment (21), reported as the percentage of posterior model probabilities smaller than c, are presented in Table 4. Also presented in this table are the percentage of p-values which are smaller than the nominal size reported in Table IV of Caner and Hansen (2001, p. 1574). Caner and Hansen find that while the t_1 test and ADF test have reasonable size, the R_{1T} and R_{2T} tests tend to over-reject the (true) null hypothesis. The R_{1T} test is found to slightly outperform the R_{2T} test. For this reason results from both the R_{1T} test and the t_1 test are reproduced here.

As expected, the more conservative Bayesian 5% procedure seems to do as well as the t_1 test in terms of size for all experiments. The 50% Bayesian test also appears to perform quite well. As would be expected, the number of false rejections by the Bayesian test decreases with the magnitude of the threshold effect. Interestingly, the degree by which Caner and Hansen's (2001) R_{1T} test is oversized seems to be increasing with the magnitude of the threshold effect.

Table 4: Empirical size of the unit root tests

		Bayesian 5 % test				Bayesian 50 % test			
	$\Delta\mu$:	0	1	2	3	0	1	2	3
	0.5	0.8	0.4	0.0	0.2	10.1	6.4	2.5	1.5
	0.2	2.1	0.4	0.3	0.2	13.9	8.9	6.5	2.3
α :	0.0	1.2	0.7	1.3	0.1	15.1	12.5	9.6	3.4
	-0.2	1.7	1.2	0.5	0.4	17.8	16.3	12.2	5.0
	-0.5	1.4	2.0	1.2	0.2	20.9	22.0	13.4	7.4
		t_1 test				R_{1T} test			
	$\Delta\mu$:	0	1	2	3	0	1	2	3
	0.5	4.2	7.0	7.0	0.4	8.5	10.8	24.7	21.8
	0.2	5.4	2.7	1.4	4.2	7.2	7.9	10.4	26.0
α :	0.0	5.2	4.7	3.2	2.4	7.0	5.6	8.2	13.3
	-0.2	4.5	5.3	3.7	2.0	5.7	7.0	6.9	9.6
	-0.5	3.1	4.6	4.5	4.1	5.9	6.7	7.9	8.7

5.5 Prior sensitivity analysis

As discussed by Koop and Potter (1999*a*) the posterior evidence in favour of nonlinear specifications is usually reduced when the prior variance on the regression coefficients θ becomes large. In order to investigate the sensitivity of the results of this section to the specification of the prior, the Monte Carlo experiments described above were repeated under two alternative prior specifications.[15] In the first alternative prior specification the prior standard deviations of the regression coefficients μ_i, α_i and ρ_i, $i = \{1, 2\}$, were doubled. In the second alternative prior specification the prior mean of the residual variance σ^2 was doubled.[16]

[15] These alternative specifications are similar to those adopted by Koop and Potter (1999*b*) for a similar investigation into the sensitivity of Bayesian inference for threshold models to prior specifications.

[16] For convenience, the results of Caner and Hansen's Monte Carlo experiments are also replicated in these tables. Obviously they are identical to those presented above.

The results of the Monte Carlo experiments under the first alternative prior specification are presented in tables A1 to A4 of Appendix A. While the power of the threshold test are generally reduced for this prior, they remain comparable with Caner and Hansen's (2001) Sup-Wald test. The size of the threshold tests are also reduced under this prior. The size of the 5% threshold test is very small, with very few false rejections, while the size of the 50% threshold test performs better than Caner and Hansen's Sup-Wald test for $\rho = 0$ and $\rho = -0.05$ under this prior.

When the prior standard deviations of the regression coefficients in θ are doubled the power of the Bayesian unit root test is generally reduced. The empirical power of the Bayesian 50% test, however, remains comparable with Caner and Hansen's (2001) R_{1T} test. The empirical power of the Bayesian test is generally superior to the R_{1T} test for values of ρ which are close to 0 and smaller threshold effects. Meanwhile the performance of the R_{1T} test is superior to the Bayesian 50% test for larger threshold effects. The size of the Bayesian unit root tests are reduced for $\Delta\mu = 0$ and $\Delta\mu = 1$ under the first alternative prior but increased for $\Delta\mu = 2$ and $\Delta\mu = 3$. The performance of the Bayesian 50% test is better than the R_{1T} test for large (positive) values of α but only better than the t_1 test when $\Delta\mu = 0$.

The results of the set of Monte Carlo experiments under the second alternative prior specification are presented in tables A5 to A8 of appendix A. Halving the prior mean of σ^2 generally increases the power of the Bayesian 50% threshold test, although the magnitude of this increase is not large. However, the size of the threshold tests are more or less unchanged. Where there is a difference in the size of the Bayesian threshold tests the number of false rejections is slightly increased.

The power of the Bayesian unit root test under the second alternative prior is increased for the experiments for which the DGP is linear $\Delta\mu = 0$ but decreased for the other Monte Carlo experiments. In general the power of the Bayesian 50% unit root test is greater than the power of the R_{1T} test for the experiments for which $\Delta\mu = 0$ and $\Delta\mu = 1$ while the R_{1T} test is the better performer for the larger threshold effects $\Delta\mu = 2$ and $\Delta\mu = 3$.

The size of the Bayesian unit root tests are increased under the second alternative prior specification. Except for the experiments for which $\alpha = 0.5$ the R_{1T} test outperforms the Bayesian 50% test in terms of the number of false rejection of the unit root hypothesis. However the t_1 test always outperforms the Bayesian 50% unit root test in terms of power while the Bayesian 5% test generally outperforms the t_1 test on this measure.

In summary, the two alternative prior specifications do not appear to affect the broad conclusion that the simpler Bayesian test performs comparably with the more complex set of tests developed by Caner and Hansen (2001).

6 Application to the male unemployment rate in the United States

Having established that the small sample performance of the Bayesian test compares favorably with Caner and Hansen's (2001) test, it is of interest to ascertain how the two approaches compare in practice.

Caner and Hansen (2001) estimate the parameters of the following self exciting threshold autoregression for the monthly unemployment rate for males in the United States over the period January 1956 through to August 1999[17]:

$$\begin{aligned} \Delta y_t &= \begin{cases} \mu_1 + \rho_1 y_{t-1} + \sum_{j=1}^{k} \alpha_{1j} \Delta y_{t-j} + e_t \text{ if } Z_{t-1} < \lambda \\ \mu_2 + \rho_2 y_{t-1} + \sum_{j=1}^{k} \alpha_{2j} \Delta y_{t-j} + e_t \text{ if } Z_{t-1} \geqq \lambda. \end{cases} \\ Z_t &= y_t - y_{t-m}. \end{aligned} \tag{22}$$

Conditioning on an autoregressive order of $k = 12$, Caner and Hansen (2001) reject a linear specification for the unemployment rate in favour of the threshold specification for delay lags $m = 1$ to $m = 12$. In the subsequent threshold unit root tests Caner and Hansen select m so as to minimize the residual sum of squares. This is reported to occur for $m = 12$. However as the value of the residual sum of squares is almost identical for $m = 9$ and favoring smaller values of m Caner and Hansen select $\widehat{m} = 9$. Based on this

[17] Caner and Hansen's (2001) data set was obtained from Bruce Hansen's web page: http://www.ssc.wisc.edu/~bhansen/progs/progs_paper.htm

specification, they report the point estimate of the threshold parameter to be $\widehat{\lambda} = 0.33$. While Caner and Hansen reject the null hypothesis $\rho_1 = 0$ in favour of $\rho_1 < 0$ at the 5% significance level, they are unable to reject the hypothesis that $\rho_2 = 0$.

In applying the Bayesian test to Caner and Hansen's unemployment data the prior means of μ_i, ρ_i and α_i $i \in \{1, 2\}$ are set to zero and the prior variances of ρ_i and α_i are set to 1. The priors for the intercept terms μ_i are set to be more diffuse with prior variances of 4. The prior degrees of freedom for the residual variance parameter is set to 3 while the prior mean of σ^2 is set to 1.4.[18] The prior probabilities attached to each model are distributed evenly.

For the purpose of making a meaningful comparison with Caner and Hansen's (2001) results, m is also chosen so as to maximize the marginal likelihood associated with (22). Conditional upon $k = 12$ this occurs for $m = 12$. However mirroring Caner and Hansen's results the marginal likelihood associated with (22) for $m = 9$ is nearly identical. Caner and Hansen are followed in selecting the smaller value for m.

The posterior distribution of the threshold variable λ is presented in Figure 1 and posterior probabilities for models M_1 to M_7 are reported in the first column of Table 5. Posterior evidence of a threshold effect is 'decisive' on Kass and Raftery's (1994) scale. It is apparent from Figure 1 that the posterior mode of the threshold occurs at $\lambda = 0.30$ which is quite similar to Caner and Hansen's estimate of $\widehat{\lambda} = 0.33$. However, the posterior model probabilities associated with each hypothesis lead to quite different conclusions about which model is favored by the data. The highest posterior model probability is associated with M_4 while Caner and Hansen's preferred specification M_7 is somewhat less preferred with a posterior model probability of 0.326.

Caner and Hansen's procedure conditions on m and k which are generally unknown. However an advantage of the Bayesian approach to inference is to the ability to accommodate specification uncertainty by integrating posterior model probabilities over plausible ranges of m and k.

The posterior model probabilities associated with models M_1 to M_7, integrated over

[18]This is diffuse relative to the variance of first difference of the unemployment rate: 0.043.

Figure 1: Posterior distribution of the threshold parameter $\Pr(\lambda|\widetilde{Y}_T, m = 9, k = 12)$ for the U.S. male unemployment rates January 1956 to August 1999

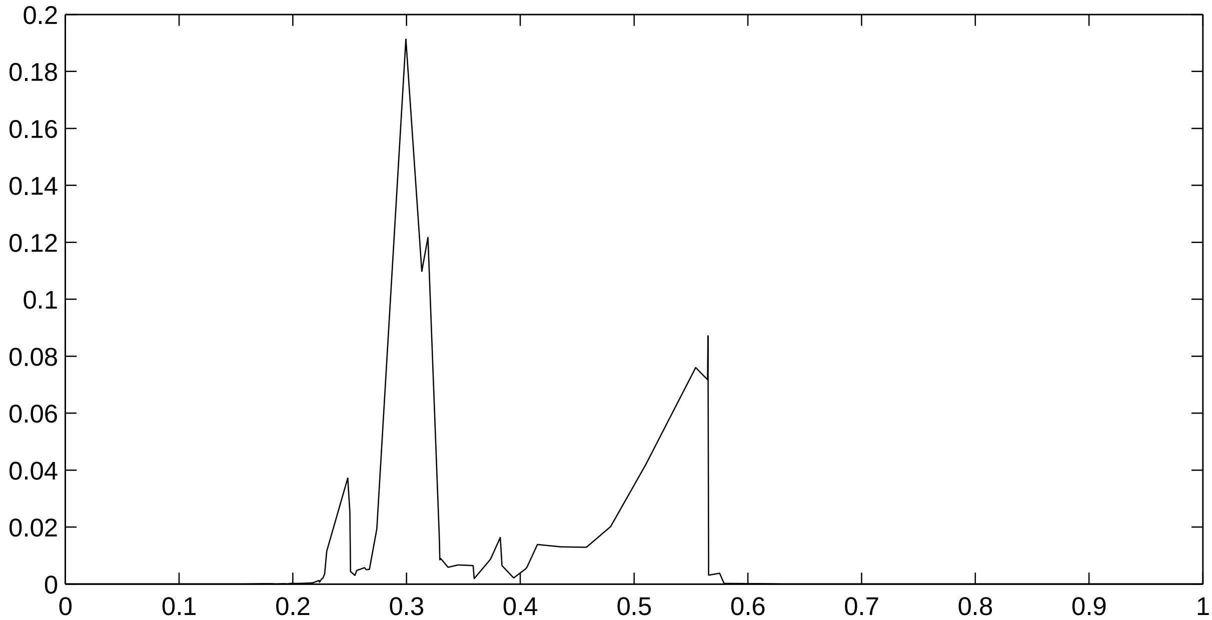
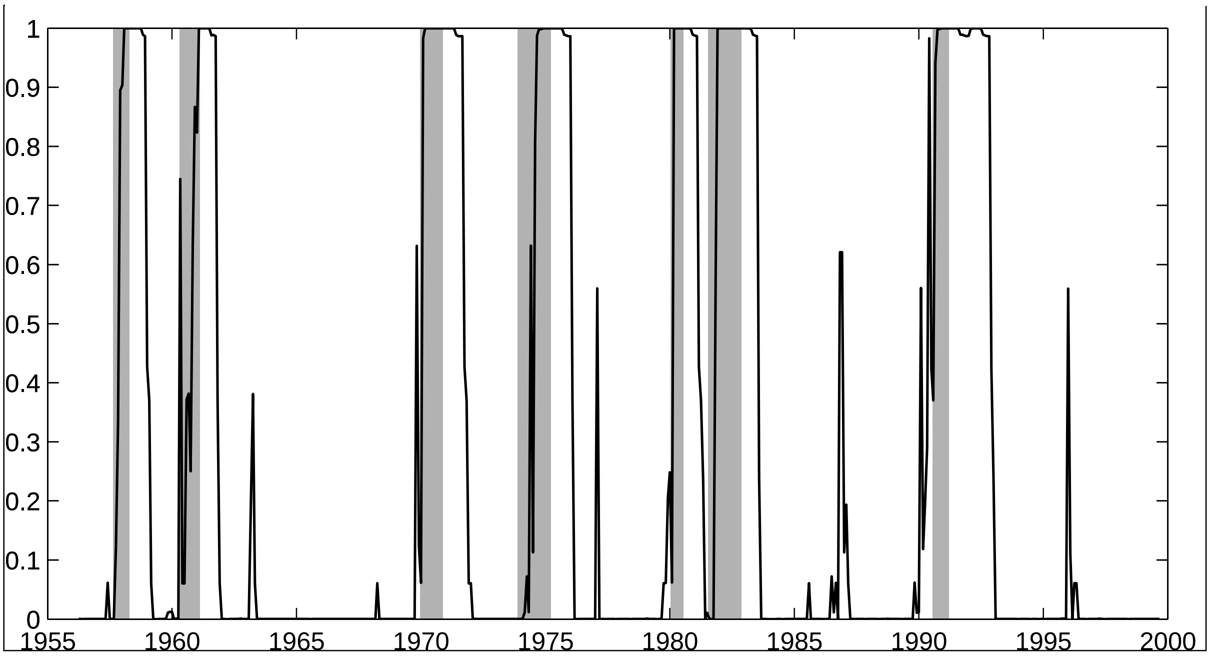


Table 5: Posterior model probabilities for the U.S. male unemployment rate January 1956 to August 1999

Model	$\Pr(M_j\|\widetilde{Y}_T, m = 9, k = 12)$	$\Pr(M_j\|\widetilde{Y}_T)$
$M_1 : \rho_1 = \rho_2 = 0,\ \theta_1 = \theta_2$	0.000	0.000
$M_2 : \rho_1 = \rho_2 < 0, \theta_1 = \theta_2$	0.000	0.000
$M_3 : \rho_1 = \rho_2 = 0, \theta_1 \neq \theta_2$	0.156	0.033
$M_4 : \rho_1 = \rho_2 < 0, \theta_1 \neq \theta_2$	0.480	0.842
$M_5 : \rho_1 < 0, \rho_2 < 0, \theta_1 \neq \theta_2$	0.032	0.052
$M_6 : \rho_1 = 0, \rho_2 < 0, \theta_1 \neq \theta_2$	0.017	0.011
$M_7 : \rho_1 < 0, \rho_2 = 0, \theta_1 \neq \theta_2$	0.314	0.061
$\rho_1 \neq \rho_2\ ,\ \theta_1 \neq \theta_2$	1.000	1.000
$\rho_1 = \rho_2 = 0$	0.156	0.033
Log of the marginal likelihood	65.769	

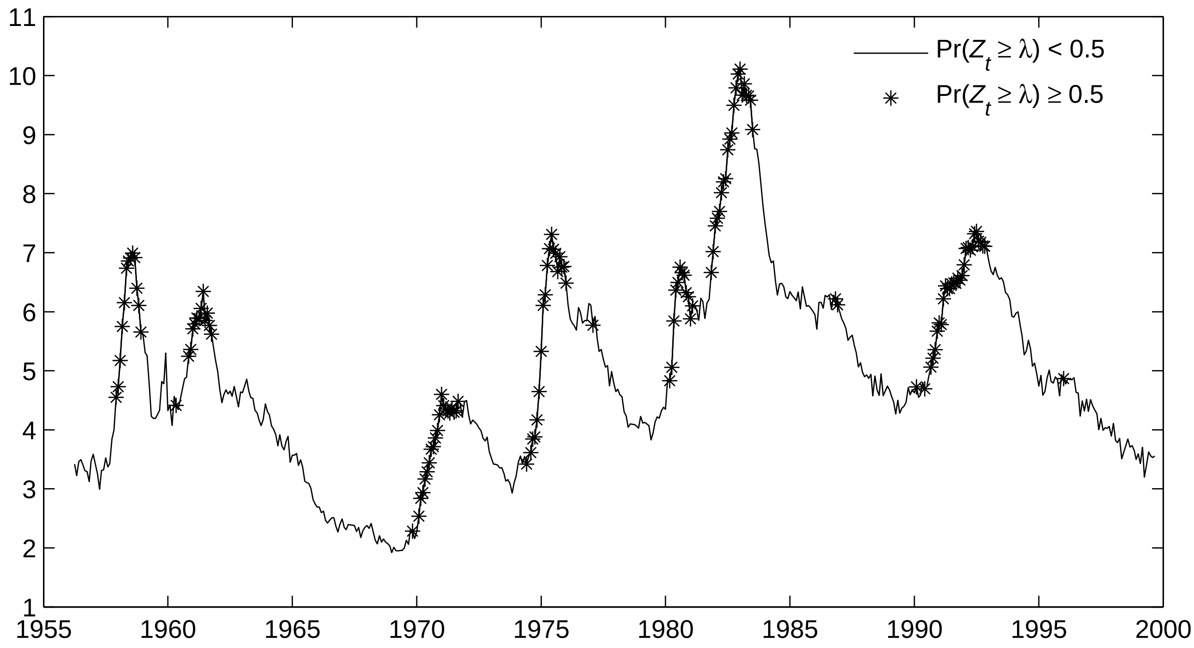
$m = 1:12$ and $k = 1:12$, are reported in the last column of Table 5. While the posterior evidence of a threshold effect remains 'decisive' on Kass and Raftery's (1994) scale, taking account of specification uncertainty in this way leads to a much greater posterior weight being attached to M_4, with only a small proportion of the posterior weight attached to M_7.

Figure 2: $\Pr(Z_t \geq \lambda | \widetilde{Y}_T)$ for the U.S. male unemployment rate January 1956 to August 1999



Caner and Hansen (2001) classify alternative regimes for the unemployment rate using the dating rule $Z_t \geq \widehat{\lambda}$. However the Bayesian approach to inference can be used to obtain $\Pr(Z_t \geq \lambda | \widetilde{Y}_T)$ − the posterior probability of being in the upper regime at any point in the sample − as an additional output. This is plotted in Figure 2 along with the periods during which the US economy was in recession as dated by the NBER. It is evident from this figure that the behavior of the unemployment rate in the United States is altered during the final few months of macroeconomic contraction and for several months after the end of the recession phase. The unemployment rate was classified by regime using the dating rule $\Pr(Z_t \geq \lambda | \widetilde{Y}_T) \geq 0.5$ and is plotted in Figure 3. This figure is almost identical to its equivalent Caner and Hansen (2001, Figure 1, p.1577).

Figure 3: U.S. male unemployment rates classified by regime January 1956 to August 1999



7 Conclusions

This paper has argued that a Bayesian approach to distinguishing between unit roots and threshold effects offers several practical advantages over an equivalent procedure developed by Caner and Hansen (2001). These advantages include increased computational efficiency and the avoidance of pre-testing biases by jointly testing for a unit root and a threshold effect rather than adopting a sequential testing strategy.

A new Bayesian test for unit roots and threshold effects has been developed and the small sample power and size properties of this have been compared with those reported by Caner and Hansen (2001). After differences between Bayesian and frequentist approaches to hypothesis testing have been taken into account the results from Monte Carlo experiments appear to indicate that the Bayesian test performs at least as well as Caner and Hansen's testing strategy.

There are several avenues for further research. These include a detailed study of the asymptotic properties of Bayesian posterior distributions for combined threshold and unit root effects and extending the test to more complex nonlinear forms and alternative prior specifications.

Appendix A: Prior sensitivity analysis

Table A 1: Empirical power of the threshold tests: prior standard deviation of θ doubled

	Bayesian test (5%)			Bayesian test (50%)			Sup-Wald test		
	$\Delta\mu$								
	0.2	1.0	2.0	0.2	1.0	2.0	0.2	1.0	2.0
$\rho = -0.05$	0.2	6.4	69.7	3.5	32.4	92.8	5.4	38.9	98.2
$\rho = 0$	0.1	9.4	91.2	1.7	34.2	98.7	5.2	35.7	97.9
	$\Delta\ \rho$								
	-0.05	-0.1	-0.2	-0.05	-0.1	-0.2	-0.05	-0.1	-0.2
$\rho_1 = -0.05$	4.1	18.3	51.7	23.4	53.2	81.2	16.5	43.0	73.8
$\rho_1 = 0$	71.0	82.4	87.4	91.1	96.3	98.3	82.1	93.1	95.5
	$\Delta\ \alpha$								
	0.5	1.0	1.9	0.5	1.0	1.9	0.5	1.0	1.9
$\rho = -0.05$	2.2	31.9	99.0	17.7	71.5	99.9	15.7	57.5	99.6
$\rho = 0$	0.8	20.7	96.8	8.8	54.2	99.4	4.7	10	34.4

Table A 2: Empirical size of the threshold tests: prior standard deviation of θ doubled

	Bayesian test 5 %			
	$\rho = -0.25$	$\rho = -0.15$	$\rho = -0.05$	$\rho = 0$
$\alpha = -0.5$	0.3	0.0	0.1	0.1
$\alpha = 0.5$	0.4	0.1	0.1	0.2
	Bayesian test 50 %			
$\alpha = -0.5$	6.8	5.0	3.3	3.6
$\alpha = 0.5$	4.3	5.6	2.6	2.3
	Sup-Wald test			
$\alpha = -0.5$	3.8	5.5	5.1	4.8
$\alpha = 0.5$	4.0	5.0	4.9	4.2

Table A 3: Empirical power of the unit root tests: prior standard deviation of θ doubled

		Bayesian 5% test			Bayesian 50% test			R_{1t} test		
	ρ_2 :	-0.05	-0.1	-0.15	-0.05	-0.1	-0.15	-0.05	-0.1	-0.15
						$\rho_1 = \rho_2$				
	0	2	7	21	22	49	78	8	15	28
$\Delta\mu$:	1	2	6	15	23	45	66	14	38	62
	2	4	9	20	31	57	81	29	76	95
	3	4	8	20	33	55	79	54	96	99
						$\rho_1 = 0$				
	0	1	2	6	14	24	41	6	12	23
$\Delta\mu$:	1	2	4	7	19	26	45	11	28	54
	2	3	6	10	26	40	59	17	64	92
	3	2	4	7	24	37	55	41	90	97
						$\rho_1 = -0.05$				
	0	2	3	8	19	33	50	8	12	20
$\Delta\mu$:	1	2	3	8	23	35	51	14	28	47
	2	4	7	10	33	46	61	29	58	87
	3	4	6	10	34	44	62	54	89	99

Table A 4: Empirical size of the unit root tests: prior standard deviation of θ doubled

		Bayesian 5 % test				Bayesian 50 % test			
	$\Delta\mu$:	0	1	2	3	0	1	2	3
	0.5	0.2	0.1	1.3	1.9	2.6	7.6	12.2	14.8
	0.2	0.3	0.3	1.8	2.6	6.3	9.4	15.8	16.8
α :	0.0	0.5	0.5	1.4	2.9	5.5	10.1	15.5	16.7
	-0.2	0.2	1.0	1.5	2.6	7.8	10.7	16.5	17.4
	-0.5	0.3	0.9	2.5	2.6	7.2	11.4	19.0	21.3
		t_1 test				R_{1T} test			
	$\Delta\mu$:	0	1	2	3	0	1	2	3
	0.5	4.2	7.0	7.0	0.4	8.5	10.8	24.7	21.8
	0.2	5.4	2.7	1.4	4.2	7.2	7.9	10.4	26.0
α :	0.0	5.2	4.7	3.2	2.4	7.0	5.6	8.2	13.3
	-0.2	4.5	5.3	3.7	2.0	5.7	7.0	6.9	9.6
	-0.5	3.1	4.6	4.5	4.1	5.9	6.7	7.9	8.7

Table A 5: Empirical power of the threshold tests: prior mean of σ^2 halved

	Bayesian test (5%)			Bayesian test (50%)			Sup-Wald test		
	$\Delta\mu$								
	0.2	1.0	2.0	0.2	1.0	2.0	0.2	1.0	2.0
$\rho = -0.05$	0.9	14.3	81.6	14.2	57.3	98.1	5.4	38.9	98.2
$\rho = 0$	0.3	14.1	95.4	7.4	20.7	99.7	5.2	35.7	97.9
	$\Delta\ \rho$								
	-0.05	-0.1	-0.2	-0.05	-0.1	-0.2	-0.05	-0.1	-0.2
$\rho_1 = -0.05$	9.7	30.0	64.8	43.2	71.8	92.0	16.5	43.0	73.8
$\rho_1 = 0$	82.0	90.9	94.9	96.8	99.0	99.7	82.1	93.1	95.5
	$\Delta\ \alpha$								
	0.5	1.0	1.9	0.5	1.0	1.9	0.5	1.0	1.9
$\rho = -0.05$	5.7	47.4	99.2	36.5	86.4	100.0	15.7	57.5	99.6
$\rho = 0$	0.8	26.1	98.4	14.7	66.5	99.9	4.7	10	34.4

Table A 6: Empirical size of the threshold tests: prior mean of σ^2 halved

	Bayesian test 5 %			
	$\rho = -0.25$	$\rho = -0.15$	$\rho = -0.05$	$\rho = 0$
$\alpha = -0.5$	1.3	1.4	0.6	0.9
$\alpha = 0.5$	1.1	0.6	0.8	0.3
	Bayesian test 50 %			
$\alpha = -0.5$	16.5	15.8	12.3	10.3
$\alpha = 0.5$	11.6	15.5	11.6	7.1
	Sup-Wald test			
$\alpha = -0.5$	3.8	5.5	5.1	4.8
$\alpha = 0.5$	4.0	5.0	4.9	4.2

Table A 7: Empirical power of the unit root tests: prior mean of σ^2 halved

		Bayesian 5% test			Bayesian 50% test			R_{1t} test		
	ρ_2 :	-0.05	-0.1	-0.15	-0.05	-0.1	-0.15	-0.05	-0.1	-0.15
					$\rho_1 = \rho_2$					
	0	4	14	38	35	69	92	8	15	28
$\Delta\mu$:	1	6	15	28	43	72	87	14	38	62
	2	7	16	35	50	78	94	29	76	95
	3	8	14	31	51	79	95	54	96	99
					$\rho_1 = 0$					
	0	3	5	15	27	44	66	6	12	23
$\Delta\mu$:	1	3	6	15	35	52	70	11	28	54
	2	5	9	17	42	60	78	17	64	92
	3	7	7	14	38	57	77	41	90	97
					$\rho_1 = -0.05$					
	0	5	8	15	39	53	75	8	12	20
$\Delta\mu$:	1	5	9	16	42	60	75	14	28	47
	2	6	10	17	50	69	81	29	58	87
	3	6	10	15	51	70	81	54	89	99

Table A 8: Empirical size of the unit root tests: prior mean of σ^2 halved

		Bayesian 5 % test				Bayesian 50 % test			
	$\Delta\mu$:	0	1	2	3	0	1	2	3
	0.5	0.6	1.3	2.7	3.1	6.7	15.7	21.3	20.4
	0.2	0.8	2.1	3.3	4.2	10.2	20.0	25.4	25.8
α :	0.0	1.2	1.1	3.6	3.3	14.2	19.6	28.4	26.6
	-0.2	0.8	1.3	3.2	3.5	11.8	23.1	25.8	26.3
	-0.5	1.1	2.2	3.2	3.8	15.7	25.4	28.5	30.5
		t_1 test				R_{1T} test			
	$\Delta\mu$:	0	1	2	3	0	1	2	3
	0.5	4.2	7.0	7.0	0.4	8.5	10.8	24.7	21.8
	0.2	5.4	2.7	1.4	4.2	7.2	7.9	10.4	26.0
α :	0.0	5.2	4.7	3.2	2.4	7.0	5.6	8.2	13.3
	-0.2	4.5	5.3	3.7	2.0	5.7	7.0	6.9	9.6
	-0.5	3.1	4.6	4.5	4.1	5.9	6.7	7.9	8.7

References

Andrews, D. (1994), 'The large sample correspondence between classical hypothesis tests and Bayesian posterior odds tests', *Econometrica* **62**(5), 1207–1232.

Andrews, D. and Ploberger, W. (1994), 'Optimal tests when a nuisance parameter is present only under the alternative', *Econometrica* **62**(6), 1383–1414.

Aretis, P., Cipollini, A. and Fattouh, B. (2002), 'Threshold effects in the U.S. budget deficit', The Levy Economics Institute Working Paper 358.

Basci, E. and Caner, M. (2005), 'Are real exchange rates nonlinear or nonstationary? evidence from a new threshold unit root test', *Studies in Nonlinear Dynamics and Econometrics* **4**(2).

Bec, F., Salem, M. B. and MacDonald, R. (2002), 'Real exchange rates and real interest rates: A nonlinear perspective', THEMA Working Paper 99-17.

Berger, J. O. and Sellke, T. (1987), 'Testing a point null hypothesis: The irreconcilability of p values and evidence', *Journal of the American Statistical Association* **82**(397), 112–122.

Caner, M. and Hansen, B. (2001), 'Threshold autoregression with a unit root', *Econometrica* **69**(6), 1555–1596.

Casella, G. and Berger, R. (1987), 'Reconciling Bayesian and frequentist evidence in the one-sided testing problem', *Journal of the American Statistical Association* **82**, 106–111.

Chan, K. and Tong, H. (1985), 'On the use of the deterministic Lyapunov function for the ergodicity of stochastic dfference equations', *Advances in Applied Probability* **17**, 667–678.

Chen, R. and Tsay, R. S. (1991), 'On the ergodicity of TAR(1) processes', *The Annals of Applied Probability* **1**, 613–634.

Davies, R. B. (1977), 'Hypothesis testing when a nuisance parameter is present only under the alternative', *Biometrika* **64**, 247–254.

DeGroot, M. H. (1973), 'Doing what comes naturally: Interpreting a tail area as a posterior probability or as a likelihood ratio', *Journal of the American Statistical Association* **68**(344), 966–969.

DeJong, D. and Whiteman, C. (1991), 'Reconsidering trends and random walks in macroeconomic series', *Journal of Monetary Economics* **28**(2), 221–54.

Dickey, D. A. and Fuller, W. A. (1981), 'Likelihood ratio statistics for autorregresive time series with a unit root', *Econometrica* pp. 1057–1072.

Dickey, J. M. (1977), 'Is the tail area useful as an approximate Bayes factor?', *Journal of the American Statistical Association* **72**, 138–142.

Edwards, W., Lindman, H. and Savage, L. (1963), 'Bayesian statistical inference for psychological research', *Psychological Review* **70**, 193–242.

Elliot, G., Rothenberg, T. and Stock, J. (1996), 'Efficient tests for an autoregressive unit root', *Econometrica* **64**(4), 813–836.

Gelman, A., Carlin, J., Stern, H. and Rubin, D. (1995), *Bayesian Data Analysis*, Chapman and Hall, New York.

Gouveia, P. and Rodrigues, P. (2004), 'Threshold cointegration and the PPP hypothesis', *Journal of Applied Statistics* **31**, 115–127.

Hansen, B. E. (1996), 'Inference when a nuisance parameter is not identified under the null hypothesis', *Econometrica* **64**, 413–430.

Henry, Ó. T. and Shields, K. (2004), 'Is there a unit root in inflation?', *Journal of Macroeconomics* p. forthcoming.

Judge, G., Hill, R., Griffiths, W., Lütkepoh, H. and Lee, T.-C. (1985), *The Theory and Practice of Econometrics*, Wiley, New York.

Kass, R. E. and Raftery, A. E. (1994), Bayes factors, Technical report, Department of Statistics, the Univeristy of Washington.

Kim, J. Y. (1994), 'Bayesian asymptotic theory in a time series model with a possible nonstationary process', *Econometric Theory* **10**(3-4), 764–773.

Kim, J. Y. (1998), 'Large sample properties of posterior densities, Bayesian information criterion and the likelihood principle in nonstationary time series models', *Econometrica* **66**(2), 359–380.

Koop, G. and Potter, S. (2000), Nonlinearity, structural breaks or outliers in economic time series?, *in* W. A. Barnett, D. F. Hendry, S. Hylleberg, T. Teräsvirta, D. Tjøstheim and A. Würtz, eds, 'Nonlinearity and Econometrics', Cambridge University Press, pp. 61–78.

Koop, G. and Potter, S. M. (1999*a*), 'Bayes factors and nonlinearity: Evidence from economic time series', *Journal of Econometrics* **88**(2), 251–81.

Koop, G. and Potter, S. M. (1999*b*), 'Dynamic asymmetries in U.S. unemployment', *Journal of Business and Economic Statistics* **17**(3), 298–312.

Marriot, J. and Newbold, P. (2000), 'The strength of evidence for unit autoregressive roots and structural breaks: A Bayesian perspective', *Journal of Econometrics* **98**, 1–25.

Perron, P. (1989), 'The great crash, the oil price shock, and the unit root hypothesis', *Econometrica* **57**(6), 1361–1401.

Phillips, P. C. (1991), 'To criticize the critics: An objective Bayesian analysis of stochastic trends', *Journal of Applied Econometrics* **6**(4), 333–64.

Phillips, P. C. and Perron, P. (1988), 'Testing for a unit root in time series regression', *Biometrika* **75**, 335–346.

Phillips, P. C. and Zivot, E. (1994), 'A Bayesian analysis of trend determination in economic time series', *Econometric Reviews* **13**(3), 291–336.

Sims, C. A. (1988), 'Bayesian skepticism on unit root econometrics', *Journal of Economic Dynamics and Control* **12**, 463–74.

Sims, C. A. and Uhlig, H. (1991), 'Understanding unit rooters: A helicopter tour', *Econometrica* **59**(6), 1591–99.

Tong, H. (1978), On a threshold model, *in* C. Chen, ed., 'Pattern Recognition and Signal Processing', Sijhoff and Noordhoff.