Testing for Integration using Evolving Trend and Seasonals Models: A Bayesian Approach[*]

GARY KOOP [†]

Department of Economics, University of Edinburgh,
Edinburgh, U.K.
E-mail: G.Koop@ed.ac.uk

HERMAN K. VAN DIJK[‡]

Econometric Institute, Erasmus University,
Rotterdam, The Netherlands
E-mail: hkvandijk@few.eur.nl

August 1999

ABSTRACT: In this paper, we make use of state space models to investigate the presence of stochastic trends in economic time series. A model is specified where such a trend can enter either in the autoregressive representation or in a separate state equation. Tests based on the former are analogous to Dickey-Fuller tests of unit roots, while the latter are analogous to KPSS tests of trend-stationarity. We use Bayesian methods to survey the properties of the likelihood function in such models and to calculate posterior odds ratios comparing models with and without stochastic trends. We extend these ideas to the problem of testing for integration at seasonal frequencies and show how our techniques can be used to carry out Bayesian variants of either the HEGY or Canova-Hansen test. Stochastic integration rules, based on Markov Chain Monte Carlo, as well as deterministic integration rules are used. Strengths and weaknesses of each approach are indicated.

Keywords: State space models, Bayes Factor, Gibbs sampler,
unit root, seasonality

1   Introduction

State space models have been widely used for the analysis of time series in many fields in the physical and social sciences. The literature on state space modelling is extensive. Influential references include Harvey (1989), Aoki (1990), Nerlove, Grether and Carvalho (1979) and West and Harrison (1997). Such time series models can also be used to analyze so-called stochastic trends in macroeconomic and financial data. Stock and Watson (1988) offer an expository survey of stochastic trend behavior in economic time series. One of the models they focus on is a type of state space model.

In this paper, we use state space models and Bayesian methods to investigate whether stochastic trends are present in economic time series. In classical econometrics, a large number of tests have been developed which test for stochastic trends (see the survey by Stock, 1994 or see Dickey and Fuller, 1979). The vast majority of these tests have the unit root as the null hypothesis. In light of the low power of unit root tests, Kwiatkowski, Phillips, Schmidt and Shin (1992) developed a test for trend-stationarity, hereafter the KPSS test (i.e. the null is trend stationarity and the alternative is the unit root, see also Leybourne and McCabe, 1994, Nyblom and Makelainen, 1983, Harvey and Streibel, 1996 and Tanaka, 1996 and the references cited therein).

The two types of classical tests can be illustrated in the following models. Dickey-Fuller type unit root tests use:

$$y_t = \rho y_{t-1} + e_t, \tag{1}$$

where e_t is a stationary error term and the null hypothesis is $\rho = 1$. A simple version of the KPSS test for stationarity makes use of a state space representation:

$$\begin{aligned} y_t &= \tau_t + e_t \\ \tau_t &= \tau_{t-1} + u_t, \end{aligned} \tag{2}$$

where u_t is white noise with variance σ_u^2, e_t is white noise with variance σ_e^2 and u_t and e_s are independent for all s and t. The null hypothesis is $\sigma_u^2 = 0$, in which case the series is stationary.

Bayesian analysis of nonstationarity (see, among many others, DeJong and Whiteman, 1991, Koop, 1992, Phillips, 1991, and Schotman and van Dijk, 1991a,b) has focussed almost exclusively on generalizations of (1). Hence, one purpose of this paper is to develop Bayesian tests based on

1

extensions of (2) which can be used to test for stochastic trends by looking at σ_u^2 (as in the KPSS test) or by looking at the autoregressive coefficients (as in the Dicky-Fuller test) or both. The first part of this paper is devoted to analyzing evolving trends models (i.e. investigating roots at the zero frequency). We begin by focussing on (2) to provide intuition into this class of models. For empirical relevance, however, it is important to allow for deterministic components and more general stationary dynamics. These are added as we generalize the model. The proposed model is very flexible and allows for stationary and integrated process which may be I(1) or I(2). Using Bayesian methods we can, unlike classical approaches, compare several hypotheses on stationarity and nonstationarity in a single analysis. The second part of the paper focuses on testing for integration at the seasonal frequency using the extension of (2) referred to as the evolving seasonals model (Hylleberg and Pagan 1997). In the context of seasonal models one can test for roots by looking at the autoregressive coefficients (see Hylleberg, Engle, Granger and Yoo, 1990 — hereafter HEGY) or at parameters similar to σ_u^2 (see Canova and Hansen, 1995). We show how the evolving seasonals model can be used to nest both these approaches and, hence, Bayesian tests for seasonal integration analogous to HEGY or Canova-Hansen can be developed.

Related Bayesian literature on models with time varying structure include (amongst others) West and Harrison (1997) and the references cited therein, Shively and Kohn (1997), Kato, Naniwa and Ishiguro (1996), Carter and Kohn (1994), Shephard (1994), De Jong and Shephard (1995), Fruhwirth-Schnatter (1994, 1995), Kim, Shephard and Chib (1998), Min (1992) and Min and Zellner (1993). West and Harrison (1997) is the standard Bayesian reference on dynamic linear models with time varying parameters, but these authors do not discuss the issues of prior elicitation and testing involving σ_u^2. Fruhwirth-Schnatter (1994), Carter and Kohn (1994) and the papers involving Shephard focus on simulation methods for carrying out Bayesian inference in very general (e.g. non-Normal) state space models. Kato, Naniwa and Ishiguro (1996) estimate a multivariate nonstationary system, but do not test for nonstationarity. Shively and Kohn (1997) use Bayesian state space methods and Gauss-Legendre quadrature to investigate whether regression parameters are time-varying. Fruhwirth-Schnatter (1995) is a theoretically-oriented paper developing methods for Bayesian inference and model selection in state space models. Although the focus of these latter two papers is different from ours, some of the basic issues are similar. In particular, they are interested in questions analogous to our testing $\sigma_u^2 = 0$. It is worth noting that Shively and Kohn use truncated uniform priors for their error variance parameters, while Fruhwirth-Schnatter uses training sample methods to elicit informative priors for these parameters.

A further purpose of this paper is to develop computational tools for analyzing state space models from a Bayesian perspective. We want to emphasize, however, at the outset that as far as numerical methods for the evaluation of integrals is concerned there is, in our opinion, no single best approach which is relevant for all applications. Accordingly, this paper illustrates how different computational methods can be used and outlines the strengths and weaknesses of each.

The effectiveness and efficiency of a computational procedure depends, of course, on the complexity of the model. For instance, one may be able to integrate a posterior analytically with respect to a subset of the parameters. This happens, in particular, when part of the model is linear and/or the prior is conjugate. This has the additional advantage of obtaining analytical insight into part of the model. If the analytical methods can be used to reduce the dimensionality of the problem sufficiently, deterministic integration[4] rules can be used efficiently for the resulting low dimensional problem, (see, e.g., Schotman and van Dijk, 1991a). Stochastic integration has truly revolutionized Bayesian analysis of state space models (see the references cited before, in particular the works involving Shephard). The best known methods are Markov Chain Monte Carlo (MCMC) and Metropolis-Hastings (see, e.g., Casella and George, 1992, and Chib and Greenberg, 1995, for clear expositions and Geweke, 1999, for a recent survey).

In this paper we make use of both deterministic and stochastic integration[5] methods and indicate the strengths and weaknesses of each approach. Typically, for the Normal state space model one can use analytical methods to integrate out all but one or two of the parameters of interest. The resulting marginal posterior can be handled more efficiently by deterministic integration rules than by stochastic integration methods. Furthermore, the use of analytical methods allows us to derive formulae for the marginal posterior of the parameter of interest and for the Bayes factor for testing for unit root behavior. With deterministic integration methods, it proves convenient to calculate this Bayes factor using the Savage Dickey density ratio (SDDR, see Verdinelli and Wasserman, 1995). As shall be stressed below, this combination of deterministic integration plus SDDR is perfectly suited for handling the relatively simple evolving trends model with any sort of prior.

The great advantage of Markov Chain Monte Carlo (MCMC) methods is that they are very general and can be used for all the models in this paper and the many extensions discussed in the conclusion to this paper. With MCMC methods, Chib (1995) provides an excellent method

---

[4] These are frequently called "numerical" integration techniques. However, we find this terminology misleading and prefer the more precise term "deterministic".

[5] In order to minimize possible confusion to the reader, note that we use the word "integration" in two ways in this paper. Whether it refers to calculating an integral or unit root behavior should be clear from the context.

for calculating the marginal likelihoods which are used to construct the Bayes factor. We shall refer to this as the "Chib method". As shall be stressed below, this combination of MCMC plus Chib method is perfectly suited for handling high-dimensional state space models. However, this approach may be somewhat complicated and computationally inefficient when we move away from a restricted class of priors.[6] Furthermore, the routine use of MCMC methods without fully understanding the analytical properties of the posterior can be misleading in some cases (e.g. the posterior or its moments may not exist, yet MCMC methods may incorrectly yield posterior results, see Fernandez, Osiewalski and Steel, 1997).[7]

The outline of this paper is as follows. In section 2 we start with the local level model as a canonical case. As a next step, we add autoregressive dynamics to the model. This gives a flexible structure so that we can analyze four hypotheses of interest: stationarity, nonstationarity through the state equation, nonstationarity through the autoregressive part, and nonstationarity through both parts (i.e. I(2) behavior). We note that nonstationarity of the state equation is an indication of a strong moving average component in the series. We present results using both deterministic and MCMC methods. We also investigate the sensitivity of the posterior results with respect to the parameterization and to the choice of the prior. In section 3 we introduce the evolving trend model and investigate the presence of stochastic trends in the extended Nelson Plosser data sets (see Schotman and Van Dijk, 1991b). In section 4, our modeling approach is extended to analyze the case of unit roots at seasonal frequencies. Some illustrative results are presented using several seasonal series from the United Kingdom. In section 5, we summarize our conclusions and discuss extensions for further work. The appendices contain some analytical results, a description of our MCMC methods, and a discussion of the choice of the parameterization.

2 Canonical Times Series Models

2.1 The Local Level Model

We begin with the simplest state space model given in (2) with the further assumptions that the errors, u_t and e_t are Normally distributed and that $\tau_0 = 0$. This model is referred to by Harvey (1989) as the local level model. There exist several different ways of interpreting this model.

---

[6] Of course, we are not saying that the MCMC plus Chib method cannot be used in every case. However, with non-standard priors a Metropolis-Hastings step may have to be added (see the evolving seasonals model in this paper). With truncated priors (such as we have in our evolving trend model), additional prior simulation may be required. Furthermore, when we have many different hypotheses to compare, the Chib method requires simulation from each model to be done. These issues are discussed in Appendix B.

[7] Since the state space models used in this paper have moving average representations, a third computational approach would be to use the algorithm in Chib and Greenberg (1994) combined with either the Chib method or the SDDR for Bayes factor calculation. We do not consider such an approach in this paper.

First, it can be interpreted as saying that the observed series is decomposed into a local level plus error where the local level contains a unit root. Secondly, it can be interpreted as a time-varying parameter model (i.e. τ_t is the mean which varies over time). Thirdly, by substituting the state equation into the measurement equation, the observed series can be seen to have an ARIMA(0,1,1) representation. Fourthly, by successively substituting the state equation into the measurement equation we obtain:

$$y_t = e_t + \sum_{i=1}^{t} u_i. \tag{3}$$

The dependent variable y_t is, thus, the sum of a random walk and a white noise component with a weight for each component which depends on the ratio $\lambda = \frac{\sigma_u^2}{\sigma_e^2}$. This ratio is commonly used by state space modellers (e.g., Harvey, 1989).

It may be convenient to map the parameter λ from the interval $[0,\infty)$ to the interval $[0,1)$ through the transformation $\theta = \frac{\lambda}{1+\lambda} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$. This parameterization also has a simple interpretation: θ is the share of the variance of y_1 accounted for by the random walk component. Alternatively, θ is the share of the variance of y_t *conditional on* y_{t-1} accounted for by the random walk component. Thus, there are three common parameterizations for the local level model: i) in terms of σ_e^2 and σ_u^2, ii) in terms of σ_e^2 and λ, and iii) in terms of σ_e^2 and θ. The choice of parameterization is crucial in Bayesian analysis since it is much easier to elicit priors on parameters which have an intuitive interpretation. In the present paper, we focus largely on θ, but the basic methods of the paper can be used for any parameterization. The consequences of our prior specification on θ and σ_e^2 for the other parameterizations are discussed in Appendix C.

It is well-known that proper, informative priors are required when calculating the Bayes factor in favor of a point hypothesis (e.g. $\theta = 0$) against an unrestricted alternative. Noninformative priors defined on an unbounded region typically lead to the case where the point hypothesis is always supported. This is known as Bartlett's paradox (see Poirier, 1995, page 390). However, following Jeffreys (1961), it is common to use noninformative priors on nuisance parameters appearing in both hypotheses (e.g. σ_e^2 appears in both the unrestricted model and the one with $\theta = 0$ imposed). Kass and Raftery (1995, page 783) provides a discussion of this issue along with numerous citations. With these considerations in mind, in this paper we pay close attention to prior elicitation of parameters involved in the tests (e.g. θ), but are relatively noninformative on the other parameters.

Since θ lies in the bounded interval $[0,1)$, a plausible prior is $p(\theta) = 1$, which is proper. In a

prior sensitivity analysis, we consider a more general prior for θ. In particular, we use a Beta prior which contains the uniform as a special case. The formulae derived below assume the uniform prior, but can be extended in the obvious way to include the Beta prior. In the Normal linear regression model, a Gamma prior for the error precision, $h_e = \sigma_e^{-2}$ is natural conjugate. We maintain this common choice and assume, *a priori*, that h_e is independent of θ. Formally, we assume the following prior:

$$p(h_e, \theta) = f_G(\nu_e, s_e^{-2})$$

for $0 \leq \theta < 1$ and $0 < h_e < \infty$, where $f_G(a, b)$ indicates the Gamma distribution with mean b and a degrees of freedom (see Poirier, 1995, page 100). However, h_e is a nuisance parameter which we will integrate out shortly, so its prior will have little effect on the Bayes factors we calculate (assuming the prior is reasonably flat).[8] In practice, we set $\nu_e = 10^{-300}$ and hence use a prior that is proper but is extremely close to the usual improper noninformative prior for the precision. For this choice of ν_e, the value of s_e^{-2} is essentially irrelevant and we just set it to 1.

To develop a Bayesian version of the KPSS test, consider the Bayes factor (B_{01}) comparing $H_0 : \theta = 0$ to $H_1 : 0 < \theta < 1$, which can be calculated using the Savage-Dickey density ratio[9] (see Verdinelli and Wasserman 1995). The Bayes factor can be written as:

$$B_{01} = \frac{p(\theta = 0|Data)}{p(\theta = 0)},$$

where the numerator of the Bayes factor is the marginal posterior of θ for the unrestricted model (or the alternative hypothesis) and the denominator is the marginal prior for θ evaluated at the point of interest $\theta = 0$ (or the null hypothesis).

For the case of the local level model with our prior for θ and h_e,:

$$B_{01} = \frac{(y'y)^{-\frac{T}{2}}}{\int_0^1 |V|^{-\frac{1}{2}} (y'V^{-1}y)^{-\frac{T}{2}} d\theta}. \tag{4}$$

---

[8] We have also experimented with more informative priors for this parameter and found that they have little effect on Bayes factors. The argument for using noninonformative priors on nuisance parameters is further strengthened if they are not strongly correlated with the parameter being tested. In many cases, it makes sense to assume that magnitude of the error in the measurement equation is independent of the relative contributions of the random walk and stationary components to the overall variance. In such cases, it is reasonable to assume that σ_e^2 is *a priori* independent of either θ or λ . See Appendix C for further details.

[9] The Savage-Dickey density ratio is a very general way of calculating Bayes factors for sharp null hypotheses. It is valid provided two conditions hold: i) $0 < p(\theta = 0|Data) < \infty$, ii) $0 < p_1(\psi, \theta = 0) < \infty$, and iii) $p_1(\psi|\theta = 0) = p_0(\psi)$. In the previous formulae ψ contains all the parameters in the model other than θ and $p_i(.)$ is the prior under H_i. These conditions hold in the present paper. If the third condition is violated, a slightly more complicated expression can be used (see Verdinelli and Wasserman, 1995).

For details, including a definition of V, see Appendix A. The use of a different prior for θ will cause only minor changes in this formula. In particular, if $p(\theta)$ is the prior for θ then it will appear inside the integral sign in the denominator. Since one dimensional deterministic integration is a simple procedure, virtually any form for $p(\theta)$ can easily be accommodated.

Note that the Bayes factor with the uniform prior on [0,1) reduces to something similar to a likelihood ratio (with σ_e^2 integrated out), except the denominator of the likelihood ratio is an average over the parameter space under the alternative hypothesis.

To illustrate our test procedure, we simulated two data sets from the local level model. In all cases, $T = 100$ and $\sigma_e^2 = 1$. For the first data we set $\theta = 0$ and for the second $\theta = 0.5$. Using simple deterministic integration, we calculated the integrating constant for $p(\theta|Data)$ used in the Bayes factor. The Bayes factors comparing the stationary to the unit root model for the two data sets are 90.82 and 2.86×10^{-86}, respectively, indicating that they distinguish well between the two hypotheses. A third data set is simulated from the standard AR(1) unit root model: $\Delta y_t = \varepsilon_t$, where ε_t is $i.i.N(0,1)$. Note that this model can be obtained from the local level by setting $\sigma_e^2 = 0$ and, hence, $\theta = 1$.[10] The Bayes factor in favor of stationarity is 9.85×10^{-146}. This suggests that if there is an AR unit root in the data generating process, our methods will be good at detecting nonstationarity.

We note that these results can also be calculated using the MCMC plus Chib method. Details on how to do this are given in Appendix B. However, in order to achieve the same accuracy as the deterministic method, MCMC requires considerably more compututational effort in this simple case.

2.2   Adding an AR(1) Component

The Bayes factor above compares a white noise model to one with a random walk plus noise. With macroeconomic series, we are usually interested in testing whether a series can be characterized by stationary fluctuations around a deterministic trend, or whether it is better characterized by a stochastic trend. As a step in this direction, and as a way of illustrating the connections between the Dickey-Fuller and KPSS tests, consider:

---

[10] Note that when $\theta = 1$, the matrix V becomes infinite. Hence, formally speaking, the pure random walk model is not nested in the local level model, although the latter can come arbitrarily close to the former. This is why we restrict θ to lie in the interval $[0, 1)$. When doing deterministic integration we use a grid over the interval $[0, 0.9999]$.

$$y_t \;=\; \tau_t + \rho y_{t-1} + e_t \tag{5}$$

$$\tau_t \;=\; \tau_{t-1} + u_{t,}$$

where the assumptions about the errors are as in the previous section. If $\theta > 0$ and $|\rho| < 1$, then y_t has a random walk component plus a stationary component. If $\theta = 0$, then we get the AR(1) model: $y_t = \rho y_{t-1} + e_t$.

In (5), a unit root is present if either $\theta > 0$ or $\rho = 1$. Our specification is very flexible and allows us to consider four hypotheses:

$H_1 : \theta = 0$ and $|\rho| < 1$. The series is stationary.

$H_2 : 0 < \theta < 1$ and $|\rho| < 1$. The series is I(1) plus a stationary component.

$H_3 : \theta = 0$ and $|\rho| = 1$. The series is I(1) and a random walk.

$H_4 : 0 < \theta < 1$ and $|\rho| = 1$. The series is I(2).

We use the same prior on θ and σ_e^2 as before and add the assumption that $p(\rho)$ is uniform over the interval $[-1, 1]$ and ρ is *a priori* independent of the other parameters. If we condition on the initial observation, set presample values of u_t to zero, multiply likelihood function by prior and integrate out σ_e^2 analytically, we obtain:

$$p(\theta, \rho | Data) \propto |V|^{-\frac{1}{2}} [(y - \rho y_{-1})' V^{-1} (y - \rho y_{-1})]^{-\frac{T}{2}}, \tag{6}$$

where $y = (y_2, ..., y_T)'$ and $y_{-1} = (y_1, ..., y_{T-1})'$.

We label B_θ, B_ρ and $B_{\theta\rho}$ as the Bayes factors for testing $\theta = 0$, $|\rho| = 1$ and $(\theta = 0, |\rho| = 1)$, respectively. The Savage-Dickey density ratio can be used to calculate any of these Bayes factors. In particular, any such Bayes factor will involve only the two-dimensional unrestricted posterior in (6) and the prior for θ and ρ. Although the setup here is more general than the simple Dickey-Fuller or Schotman and van Dijk (1991a,b) setup, the similarities between B_ρ and these tests are apparent. The similarity between B_θ and the KPSS test is also apparent. However, our setup allows for more general comparisons. In fact, the posterior probability of any of the four hypotheses listed above can be calculated using B_θ, B_ρ and $B_{\theta\rho}$.

To investigate posterior properties and the performance of Bayesian model comparison procedures, we simulate data assuming $T = 100$ and $\sigma_e^2 = 1$. Table 1 presents posterior probabilities for the four hypothesis listed above for different values of θ and ρ.

Table 1: Posterior Model Probabilities for Simulated Data Sets

8

	$p(H_1|Data)$	$p(H_2|Data)$	$p(H_3|Data)$	$p(H_4|Data)$
$\theta = 0, \rho = 0$	0.975	0.025	0.000	0.000
$\theta = .5, \rho = 0$	0.000	0.998	0.002	0.000
$\theta = 0, \rho = 1$	0.154	0.040	0.798	0.009
$\theta = .5, \rho = 1$	0.000	0.010	0.000	0.990
$\theta = .5, \rho = .5$	0.008	0.199	0.772	0.020
$\theta = 0, \rho = .5$	0.987	0.013	0.000	0.000

Given that our simulated data sets exhibit a wide variety of behavior: from white noise, through stationary but persistent, to I(1), to I(2) series, it can be seen from Table 1 that the Bayes factors, as reflected in the posterior model probabilities, do detect the appropriate degree of integration with high probability. In general, they also seem to detect whether nonstationarity is entering through an AR unit root or through a non-degenerate random walk state equation. The only exception is the case $\theta = .5, \rho = .5$ where more weight is put on the AR unit root than we would expect. This result may be explained as follows. When we compare (5) with a general ARIMA specification, it can easily be shown that, in the case of $\theta = .5, \rho = .5$, the implied ARIMA nearly has a common factor. It is well-known that the posterior (with a relatively noninformative prior) is ill-behaved in such a case. We have used this pathological case to show the flexibility of model selection in a Bayesian setup. Of course, in practice, an applied time series researcher may use prior information to surmount such difficulties. For instance, a tight prior on ρ (e.g. $\rho \sim N(0, 0.10)$) would force all the persistence in the series into the state equation, leaving the AR component to pick up only the temporary component.

One may question the robustness of the results in Table 1 to the choice of prior. In this respect, we make the following comments. In Bayesian analysis, a desirable strategy is to specify the model so that its parameters can be easily interpreted. The researcher can then elicit informative priors about them in a straightforward way. In time series models, the parameters rarely have a structural interpretation and, hence, it is often difficult to follow this strategy. So far, we have responded to this problem by working with a parameterization which is rather natural. Furthermore, we have made a particular choice for the prior on this parameter. We acknowledge, however, that some of our readership might prefer other parameterizations (e.g. in terms of moving average coefficients) and other priors. Since the purpose of the present paper is to develop Bayesian methods for testing for integration, we note that other readers can use different parameterizations and priors with only minor alterations to the techniques introduced here (see Appendix C). Furthermore, we perform a prior sensitivity analysis on θ to test the robustness of the results reported.

Since θ is bounded in the unit interval, a sensible class of prior distribution is the Beta, which can take on a myriad of different shapes (see Poirier, 1995, pp. 104-105). In the table below, we

9

assume $\theta \sim f_B(\theta_0, \theta_1)$ for different choices of θ_0 and θ_1. For the rest of the parameters we retain the priors used in the body of the paper. To aid in interpretation note that the mean and variance of the Beta are $\frac{\theta_0}{\theta_0+\theta_1}$ and $\frac{\theta_0\theta_1}{(\theta_0+\theta_1+1)(\theta_0+\theta_1)^2}$, respectively. The distribution is symmetric around $\theta = \frac{1}{2}$ if $\theta_0 = \theta_1$, positively skewed if $\theta_0 < \theta_1$ and negatively skewed otherwise. Special cases worth noting are: i) The uniform which implies $\theta_0 = \theta_1 = 1$, ii) If θ_0 and θ_1 are both greater than one then the distribution has an interior mode and becomes roughly bell-shaped as θ_0 and θ_1 increase, and iii) If θ_0 and θ_1 are both less than one then the distribution is U-shaped. Using these facts, it can be seen that the prior sensitivity analysis below covers an enormous range of priors.[11]

For brevity, we present only the Bayes factor in favor of the hypothesis that $\theta = 0$. Table 2 reports results from a new artificial data set from (3) with T=100, $\theta = \frac{1}{3}$, $\sigma_e^2 = 1$ and $\rho = 0$.

Table 2: Bayes Factors in Favor of $\theta = 0$ for Artificial Data Set

	$\theta_0 = .1$	$\theta_0 = \frac{1}{2}$	$\theta_0 = 1$	$\theta_0 = 2$	$\theta_0 = 10$
$\theta_1 = .1$	3.5×10^{-4}	2.0×10^{-6}	3.0×10^{-9}	5.7×10^{-15}	2.6×10^{-62}
$\theta_1 = \frac{1}{2}$	4.5×10^{-4}	2.7×10^{-6}	4.4×10^{-9}	9.5×10^{-15}	1.2×10^{-61}
$\theta_1 = 1$	5.8×10^{-4}	3.6×10^{-6}	6.1×10^{-9}	1.5×10^{-14}	4.2×10^{-61}
$\theta_1 = 2$	8.9×10^{-4}	5.8×10^{-6}	1.0×10^{-8}	2.8×10^{-14}	2.4×10^{-60}
$\theta_1 = 10$	8.4×10^{-3}	6.6×10^{-5}	1.5×10^{-7}	6.9×10^{-13}	3.5×10^{-57}

We note that the table provides clear evidence in favour of the unit root hypothesis, despite the fact that we have considered an enormously wide range of priors. It can be seen that most evidence for a unit root is found when the prior has an interior mode and allocates less weight to the region near zero (see the right hand side of the table). Priors that allocate substantial weight near $\theta = 0$ yield less evidence for the unit root (see the left hand side of the table). In summary, for a clear cut case like the present one (i.e. $\theta = \frac{1}{3}$) the Bayes factor can vary a lot, but in all cases strong evidence of a unit root appears. This prior sensitivity analysis will be continued below and in the following section.

The previous tables were calculated using deterministic integration methods plus the SDDR. It is useful to also consider the MCMC plus Chib method. This depends on the precise parameterization and priors used. Appendix B develops this method for the case where either the θ or λ parameterization is used and the prior for ρ, σ_e^{-2} is either noninformative or Normal-Gamma. In order to continue our investigation of the sensitivity of results to different priors and parameterizations, we use the methods of Appendix B along with the artificial data set used to make Table 2. We work with the λ parameterization and try different priors for λ in the inverted-Gamma

---

[11] The Beta distribution is defined on the interval (0,1) and not our desired interval of [0,1). So formally speaking, what we are using in this paper is not the Beta distribution but the Beta distribution plus the assumption that the density evaluated at the point zero is some finite constant. Since zero is a point of measure zero it can easily be verified that the precise choice of constant does not matter.

class. For ρ we use a Normal prior with mean 0 and standard deviation 1. Note that the inverted-Gamma distribution can be parameterized in terms of a degrees of freedom parameter, v_λ and the mean, μ_λ. We set $v_\lambda = 2$ and $\mu_\lambda = 0.005, 0.01, 0.1$ or 1.0. In other words, we are expressing a wide range of prior means reflecting a range of beliefs from σ_u being very small relative to σ_e through a case where they are roughly equal. The Bayes factors for testing $\theta = 0$ for these four priors are $5.8 \times 10^{-8}, 3.2 \times 10^{-8}, 1.4 \times 10^{-9}$ and 2.7×10^{-9}, respectively. These results are similar to those given in Table 2, indicating that the MCMC plus Chib method is giving reliable results.[12] As before, the sensitivity analysis indicates that priors which place more weight near the trend-stationary hypothesis (here $\lambda = 0$) give it more support. The degree of prior sensitivity in the λ-paramterization appears less than was found in Table 2. This is due to the fact that all of the inverted-Gamma priors set $v_\lambda = 2$, a relatively noninformative value. However, some of the priors in Table 2 are very informative and differ enormously from one another. Hence, the greater prior sensitivity found in Table 2 is not surprising.

3 Testing For Integration in the Evolving Trend Model

Economic time series typically have more dynamic and deterministic terms than (5) allows for. These considerations suggest that the following specification is more appropriate for empirical research:

$$
\begin{aligned}
\phi(L)y_t &= \tau_t + e_t &\quad (7)\\
\tau_t &= \alpha + \tau_{t-1} + u_t
\end{aligned}
$$

where $\phi(L)$ is a polynomial in the lag operator of order p and the assumptions about the errors are the same as for the previous models, but here we no longer assume $\tau_0 = 0$. It is worthwhile to motivate briefly this particular extension as opposed to one which puts the deterministic component directly in the measurement equation or puts the AR component in the state equation. If we assume that $\phi(L)$ satisfies the stationarity conditions and difference y_t, we can write:

$$
\phi(L)\Delta y_t = \alpha + u_t + \Delta e_t. \qquad (8)
$$

---

[12] However, at least for this simple model, MCMC methods are much more computationally demanding than deterministic ones. Furthermore, the calculation of posterior probabilities of all four hypotheses would have required MCMC simulation from four different models. Note that the use of the SDDR requires only that the researcher work with the unrestricted model.

That is, if $\theta > 0$ the model becomes an ARIMA(p,1,1) plus drift. If $\theta = 0$, then the model can be written in terms of stationary fluctuations around a deterministic trend:

$$\phi(L)y_t = \tau_0 + \alpha t + e_t. \tag{9}$$

Hence, if we test $\theta = 0$ we are testing a null of trend-stationarity against an alternative of a unit root with drift. We feel that these are the sensible hypotheses to be considering in practice. An alternative way of extending (5) is to add the AR component to the state equation. Then, under $\theta = 0$, the model would reduce to white noise fluctuations around a deterministic trend which is not a reasonable null hypothesis for most macroeconomic data. We note that the present specification is identical to the one presented in Leybourne and McCabe (1994).[13]

Since this specification is now suitable for working with macroeconomic time series, in this section we investigate the properties of the extended Nelson-Plosser data in an empirical illustration. Schotman and van Dijk (1991b) use this data set to carry out Bayesian tests for a unit root in an AR process (allowing for deterministic time trend). The reader is referred to this paper for a description of the data. In an attempt to make our results comparable to Schotman and van Dijk (1991b), we set $p = 3$ for all series except the unemployment rate for which we set $p = 4$. Table 3 presents posterior model probabilities for these series, the last column of this table presents the probability of a unit root calculated by Schotman and van Dijk.[14]

---

[13] Another interesting specification is used in Harvey and Streibel (1996) which forces θ to zero as ρ approaches 1.

[14] The last column of Table 2 is taken from Hoek (1997), who made some corrections to Schotman and van Dijk's original calculations.

Table 3: Posterior Model Probabilities for Nelson-Plosser Data

	$p(H_1\|Data)$	$p(H_2\|Data)$	$p(H_3\|Data)$	$p(H_4\|Data)$	$S.v.D.$ $p(\rho=1)$
Real GNP	0.169	0.819	0.012	0.000	0.300
Nominal GNP	0.010	0.931	0.055	0.004	0.619
GNP per capita	0.247	0.740	0.013	0.000	0.290
Industrial Production	0.293	0.686	0.021	0.000	0.316
Employment	0.002	0.998	0.001	0.000	0.313
Unemployment	0.463	0.533	0.004	0.000	0.217
GNP Deflator	0.011	0.866	0.110	0.014	0.678
Consumer Prices	0.000	0.996	0.003	0.001	0.697
Nominal Wages	0.026	0.887	0.078	0.010	0.602
Real Wages	0.006	0.948	0.042	0.004	0.642
Money	0.036	0.897	0.055	0.012	0.397
Velocity	0.001	0.983	0.015	0.000	0.666
Interest Rate	0.001	0.973	0.011	0.015	0.641
Stock Prices	0.021	0.898	0.079	0.001	0.653

The results in Table 3 accord reasonably well with the results of Schotman and van Dijk (1991b), despite differences in specification (and slight differences in the prior). In particular, most evidence for stationarity is found for series like real GNP, GNP per capita, unemployment and industrial production. Other series provide much stronger evidence of integration. The present approach, however, finds more evidence of evolving trends. Given the results reported in Hoek (1997, p. 91) on the strong presence of MA terms in the Nelson-Plosser data, we conclude that the implicit MA component added in our state-space approach is an important extension for macro data. For most series, H_2 receives much more probability than H_3 indicating that the data prefer the state space unit root (which implicitly adds a moving average component) to the autoregressive unit root. To see why this might increase the probability of integration, suppose that a true data generating process exists and it is ARIMA(3,1,1) and that the MA coefficient is substantial and negative. This series, of course, is I(1) and we would hope a test would indicate this. The Schotman and van Dijk approach would approximate the ARIMA(3,1,1) by an AR(3) model. The presence of a negative MA coefficient would tend to pull the AR coefficients into the stationary region, reducing the probability of the unit root relative to the present approach which would correctly model the ARIMA(3,1,1).

13

The following table continues our prior sensitivity analysis, using the Beta family of priors for θ for one of the Nelson-Plosser series.

Table 4: Bayes Factors in Favor of $\theta = 0$ for Real GNP

	$\theta_0 = .1$	$\theta_0 = \frac{1}{2}$	$\theta_0 = 1$	$\theta_0 = 2$	$\theta_0 = 10$
$\theta_1 = .1$	243.55	15.85	0.02	2.0×10^{-8}	2.4×10^{-56}
$\theta_1 = \frac{1}{2}$	244.84	52.40	0.08	1.1×10^{-7}	2.2×10^{-55}
$\theta_1 = 1$	246.90	88.40	0.21	3.4×10^{-7}	1.4×10^{-54}
$\theta_1 = 2$	248.00	130.03	0.52	1.2×10^{-6}	1.6×10^{-53}
$\theta_1 = 10$	249.17	206.12	4.68	4.0×10^{-5}	1.9×10^{-49}

Table 4 indicates a greater degree of prior sensitivity than Table 2. It is worthwhile to discuss this result. The uniform prior for θ indicates moderate support for the hypothesis that $\theta > 0$. If we use a prior which allocates more weight to the region $\theta > \frac{1}{2}$ or keeps the prior mean greater than $\frac{1}{3}$ and tightens the prior variance, the support for the hypothesis that $\theta > 0$ is strengthened (i.e. if we look in the right and upper right hand parts of the table we see strong support for integration). However, if the prior allocates significant weight near the region $\theta = 0$, we find support for $\theta = 0$. This lack of robustness is due to the strong correlation between θ and ρ. If the prior for θ places a great deal of weight near $\theta = 0$, then the marginal posterior for θ also gets pulled towards zero and ρ becomes larger. Since the posterior for θ is located near zero, the hypothesis that $\theta = 0$ gains support. However, if the prior for θ is more spread out, then the opposite happens. Loosely speaking, in our model there are two ways that integrated behavior can enter. For real GNP, the data are happy with either of them and the prior can determine whether persistence enters through θ or through ρ. Our conclusion is that this macroeconomic time series is only weakly informative about the presence of a stochastic trend. This corresponds with other Bayesian studies in the literature and appears to be much more sensible than the mechanical classical failure to reject the unit root hypothesis for U.S. real GNP. It is worth noting, however, that with other data sets (either artificial or real) that this lack of prior robustness is usually not observed.

4 Testing for Integration in the Evolving Seasonals Model

4.1 Theory

The evolving seasonals model has recently been reintroduced to the econometrics literature in Hylleberg and Pagan (1997). Originally developed in Hannan, Terrell and Tuckman (1970), this model is a very flexible specification which allows the seasonal pattern to vary over time. A simple variant of this model is given by:

$$y_t = \tau_{0t} \cos(\alpha_0 t) + \tau_{1t} \cos(\alpha_1 t) + 2\tau_{2t} \cos(\alpha_2 t) + 2\tau_{3t} \sin(\alpha_2 t) + e_t, \qquad (10)$$

where $\alpha_0 = 0, \alpha_1 = \pi$ and $\alpha_2 = \frac{\pi}{2}$ capture behavior at the relevant 0 and seasonal frequencies, respectively. The $\tau'_{it}s$ capture the evolution of the trend and seasonal patterns over time. Hylleberg and Pagan (1997) shows how this specification nests most common seasonal models. Note that there are other ways of modelling seasonality (see, for instance, Franses, 1996, West and Harrison, 1997, chapter 8 or Harvey, 1889, chapters 2 and 6). The evolving seasonals model is a particularly flexible specification.

In this paper we focus on testing for seasonal unit roots from a Bayesian perspective. It is worthwhile to briefly digress and describe the two chief classical approaches. The most common of these is outlined in Hylleberg, Engle, Granger and Yoo (1990) — HEGY — and is based on the fact that an AR(p) specification: $\phi(L)y_t = e_t$ can be written as

$$\phi^*(L)y_{4,t} = \delta_0 y_{1,t-1} + \delta_1 y_{2,t-1} + \delta_2 y_{3,t-2} + \delta_3 y_{3,t-1} + e_t,$$

where $y_{1,t} = (1 + L + L^2 + L^3)y_t, y_{2,t} = -(1-L)(1+L)y_t, y_{3,t} = -(1-L^2)y_t$ and $y_{4,t} = (1-L^4)y_t$. A nonseasonal unit root is present if $\delta_0 = 0$, while if $\delta_1 = 0$ a seasonal unit root at frequency π is present. δ_2 and δ_3 relate to possible seasonal unit roots at frequency $\frac{\pi}{2}$ and HEGY suggests a joint test of $\delta_2 = \delta_3 = 0$. An alternative test is given by Canova and Hansen (1995) and is based on a specification similar to (10) under the assumption that, for $i = 0, 1, 2, 3$:

$$\tau_{it} = \tau_{i,t-1} + u_{it},$$

and $var(u_{it}) = \sigma_i^2$. If $\sigma_1^2 = 0$ then a seasonal unit root at frequency π is present while if $\sigma_2^2 = \sigma_3^2 = 0$ then a seasonal unit root at frequency $\frac{\pi}{2}$ is present. The nonseasonal unit root occurs if $\sigma_0^2 = 0$.

Given the evolving seasonals model, it is apparent that we can derive a specification that nests both these approaches in the same way that our specification in the previous section nested both Dickey-Fuller and KPSS tests. As before, it is important to allow for deterministic terms and hence we work with the following specification:

$$\phi^*(L)y_{4,t} \quad = \quad \tau_{0t} + \tau_{1t}\cos(\pi t) + 2\tau_{2t}\cos(\frac{\pi t}{2}) + 2\tau_{3t}\sin(\frac{\pi t}{2}) \tag{11}$$

$$+\delta_0 y_{1,t-1} + \delta_1 y_{2,t-1} + \delta_2 y_{3,t-2} + \delta_3 y_{3,t-1} + e_t$$

$$\tau_{it} \quad = \quad \alpha_i + \tau_{i,t-1} + u_{it,}$$

where the $e_t's$ are $i.i.N(0,\sigma_e^2)$, the $u_{it}'s$ are $i.i.N(0,\sigma_i^2)$ and all error terms are independent of one another. As in the previous section, we can test for unit roots either through the AR coefficients or through the error variances in the state equations (e.g. testing $\delta_0 = 0$ or $\sigma_0 = 0$ for the nonseasonal unit root). If the state equations are substituted into the measurement equation it can be seen that the $\tau_{i0}'s$ enter as a deterministic seasonal pattern and the inclusion of drift terms in the state equations (i.e. the $\alpha_i's$) allows for a deterministic trend in the seasonal patterns. In our empirical work, we rule out the latter and set $\alpha_1 = \alpha_2 = \alpha_3 = 0$, but leave α_0 unrestricted. Assuming the AR coefficients satisfy the stationarity condition, then if $\sigma_i = 0$ for $i = 0, 1, 2, 3$ the model is characterized by stationary fluctuations around a deterministic seasonal pattern. Hence, equation (11) is an extremely flexible specification which nests most common seasonal models, and our Bayesian counterpart to the Canova-Hansen test has as its null hypothesis a reasonable model for macroeconomic time series.

As before, we reparameterize in terms of

$$\theta_i = \frac{\sigma_i^2}{\sigma_i^2 + \sigma_e^2}.$$

This parameterization is less intuitive than we obtained for the evolving trends model. Nevertheless, it seems as intuitive as other alternatives. Tests of the various sorts of seasonal integration reduce to testing for zero restrictions on the $\theta_i's$.

Note, however, that there are eight parameters of interest (i.e. δ_i and θ_i for $i = 0, 1, 2, 3$), so that, even if we analytically integrate out all nuisance parameters, deterministic integration is extremely difficult given current computational power. However, it is possible to set up an MCMC algorithm to analyze this model (for details, see Appendix B). To calculate Bayes factors, it is necessary to specify priors for the $\theta_i's$. To do this, we extend the strategy of the previous section, assume prior independence between these parameters, and obtain: $p(\theta_i) = 1$ if $0 \le \theta_i < 1$.[15] For all other parameters, we use traditional, flat, noninformative priors. Hence, the Bayes factors

---

[15] Note that we are using an improper prior for the $\delta_i's$ and, hence, do not calculate Bayes factors for these parameters. The methodology outlined in this section could be used do this, but proper priors would be needed. Such priors could either be elicited subjectively or we could use a flat prior over the stationary region. The necessary restriction for imposing the latter is complicated (see Franses, 1996, pp. 64-66). Hence, for reasons of simplicity and to keep the empirical illustration focussed on the $\theta_i's$, we do not consider proper priors for the AR parameters.

calculated here have the same "weighted likelihood ratio" form as in the previous section. Of course, subjective informative priors can be used if so desired.

4.2 Empirical Illustration

The techniques described above are here illustrated using several U.K. seasonal series: GDP, total consumption (TOTCON), consumption of nondurables (NONDUR), total investment (TOTINV), exports (EXPORTS) and imports (IMPORTS). All data are quarterly, logged and run from 1955:1 to 1988:4. These series have been analysed extensively by many authors (see Franses, 1996, chapter 5 for a list of citations). Franses, 1996, Table 5.2 presents results from the HEGY test on these series (and others), concluding that the nonseasonal unit root seems to be present in all series, and TOTCON and NONDUR have in addition roots at both seasonal frequencies. Table 5 presents Bayes factors for testing $\theta_i = 0$, which we call B_{θ_i} for $i = 0, 1, 2, 3$. Small values of B_{θ_i} indicate evidence in favor of seasonal integration. The last four rows present posterior means of the $\delta_i's$, with posterior standard deviations in parentheses.

Table 5: Posterior Information on UK Seasonal Series

	GDP	$TOTCON$	$NONDUR$	$EXPORTS$	$IMPORTS$	$TOTINV$
B_{θ_o}	$4.9x10^{-114}$	$1.5x10^{-13}$	$3.2x10^{-41}$	$7.0x10^{-142}$	$8.4x10^{-3}$	$2.1x10^{-35}$
B_{θ_1}	0.10	$5.8x10^{-3}$	$6.3x10^{-4}$	0.26	$5.7x10^{-2}$	0.27
B_{θ_2}	0.14	$2.4x10^{-2}$	$2.7x10^{-3}$	0.66	$2.8x10^{-2}$	$3.1x10^{-2}$
B_{θ_3}	0.31	$4.8x10^{-3}$	$7.4x10^{-4}$	0.18	0.15	0.29
δ_0	-0.19	-0.11	-0.09	-0.35	-0.20	-0.16
	(0.07)	(0.05)	(0.07)	(0.07)	(0.06)	(0.05)
δ_1	-0.51	-0.75	-0.82	-0.27	-0.41	-0.34
	(0.21)	(0.26)	(0.34)	(0.06)	(0.10)	(0.18)
δ_2	-0.53	-0.98	-0.77	-0.63	-0.69	-0.62
	(0.11)	(0.23)	(0.19)	(0.09)	(0.10)	(0.12)
δ_3	-0.21	-0.34	-0.40	0.08	-0.21	-0.02
	(0.12)	(0.27)	(0.25)	(0.08)	(0.13)	(0.14)

A standard Bayesian rule of thumb (see, e.g., Poirier, 1995, page 380) is to say that there is slight evidence against $\theta_i = 0$ if $0.10 < B_{\theta_i} < 1.0$, strong evidence if $0.01 \leq B_{\theta_i} \leq 0.10$, and decisive evidence if $B_{\theta_i} < 0.01$. Using this rule of thumb, all series provide decisive evidence in favor of a unit root at the nonseasonal frequency. TOTCON and NONDUR provide decisive evidence in favor of roots at both seasonal frequencies. These results accord with those provided by the HEGY test. The Bayes factors for the seasonal unit roots for the other series do not provide decisive evidence, but nevertheless some evidence for seasonal unit roots is found.

Our specification allows for seasonal and nonseasonal unit roots to enter through either the AR coefficients or the state equation. Although we do not calculate Bayes factors for the former, the posterior moments for the $\delta_i's$ indicate that the data chooses to put unit roots (if they exist) in

17

the state equations. This finding is analogous to that noted in Section 3, where the Nelson-Plosser data tended to favor H_2 over H_3.

It is also worth noting that we test each of the $\theta'_i s$ individually. Given the aliasing problem, one may be interested in doing a joint test of $\theta_2 = \theta_3 = 0$. This can, of course, be easily done using our present framework.

5 Conclusion

In this paper, we develop Bayesian tests of stochastic trends in economic time series using state space representations. We consider both trend and seasonal models, and AR unit roots and unit roots arising in the state equation(s). Our general framework nests most of the common approaches to testing for integration in the literature. We construct computational methods involving either deterministic integration or posterior simulation to calculate the probability associated with each type of unit root. Empirical evidence using simulated and real data indicate that the approach advocated in this paper is both simple to use and yields reasonable results. The added flexibility of state space modelling and the allowance for the test of stationarity to be a point hypothesis (in contrast to the usual setup where the unit root is the point hypothesis) heighten the advantages of our approach.

The basic ideas in this paper can be extended in a conceptually straightforward manner. For instance, state space modelling of financial time series involving fat-tailed distributions and stochastic volatility is studied by Kim, Shephard and Chib (1998) and Bos, Mahieu and van Dijk (1999). Model comparison involving nonlinear models, outliers and models with structural instability is taken up by Koop and Potter (1999b, 2000). Issues relating to lag length selection are discussed in Koop and Potter (1999a). In all of these areas, Bayesian state space methods have a potentially important role to play (see, in particular, Koop and Potter, 1999b). Furthermore, multivariate models, including those for panel data, can be easily handled. For instance, testing for common trends (i.e. cointegration) in multivariate systems and unit roots in panels with Bayesian state space methods is a topic of our present research.

We end this paper with a remark. MCMC algorithms for all of these extensions are available in the literature (see our list of references). Deterministic integration methods would be difficult to use with these extensions due to the large number of parameters in the model than cannot be integrated out analytically. Hence, we recommend the MCMC plus Chib method for Bayes factor calculation as a very general approach for Bayesian analysis.

Appendix A: Further Analytical Results

For the local level model of Section 2.1, we calculate the Savage-Dickey density ratio by integrating out the nuisance parameter σ_e^2. We set presample values of u_t to zero. Using (3) and defining $y = (y_1, ..., y_T)'$, we obtain:

$$y \sim N(0, \sigma_e^2 V),$$

where $V = I_T + \frac{\theta}{1-\theta} CC'$ and

$$C = \begin{pmatrix} 1 & 0 & . & . & . & . & 0 \\ 1 & 1 & 0 & . & . & . & 0 \\ 1 & 1 & 1 & 0 & . & . & 0 \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

C is known as the random walk generating matrix. Multiplying prior by likelihood and integrating out σ_e^2 yields the marginal posterior for θ :

$$p(\theta|Data) \propto |V|^{-\frac{1}{2}} (y'V^{-1}y + \nu_e s_e^2)^{-\frac{T+\nu_e}{2}}. \tag{A.1}$$

Since we are setting $\nu_e = 10^{-300}$ and $s_e^2 = 1$, terms involving these hyperparameters are extremely small (at least for the data sets used in this paper) and can be ignored in equation (A.1). In our empirical work, they are included (although they are numerically irrelevant). However, to make our expressions for posterior and Bayes factors easy to interpret and compare to classical likelihood ratio statistics, we omit them in the formulae in Section 2, which should be considered as providing (extremely good) approximations to the true posteriors and Bayes factors.

The integrating constant of posterior (A.1) is, to our knowledge, not known in terms of elementary functions (such as the Gamma function). However, one-dimensional integration suffices to calculate it and the Bayes factor in (4).

For the local level model with AR(1) component, one starts from (5) and the prior discussed in Section 2.2. Proceeding in a similar way as for the simple local level model, one obtains the marginal posterior of (θ, ρ) given in (6). Note that, if we had assumed an untruncated uniform prior for ρ, we could also have integrated out ρ analytically, using the properties of the Student-t

density. Details are omitted here. If we were to integrate out ρ, we could derive an expression for the Bayes factor analogous to that given in Section 2.1:

$$B_{01} = \frac{(y'_{-1}y_{-1})^{-\frac{1}{2}} (y'My)^{-\frac{T-1}{2}}}{\int_0^1 |V|^{-\frac{1}{2}} (y'_{-1}V^{-1}y_{-1})^{-\frac{1}{2}} (s^2)^{-\frac{T-1}{2}} d\theta},$$

where $M = I - y'_{-1}(y'_{-1}y_{-1})^{-1}y_{-1}$ and $s^2 = (y - \widehat{\rho}y_{-1})'V^{-1}(y - \widehat{\rho}y_{-1})$. Furthermore, $\widehat{\rho} = (y'_{-1}V^{-1}y_{-1})^{-1}y'_{-1}V^{-1}y$.

For the case of the evolving trend model given in Section 3, it is convenient to rewrite the measurement equation in (7) as:

$$y_t = \tau_t + \rho y_{t-1} + \sum_{i=1}^{p-1} \pi_i \Delta y_{t-i} + e_t.$$

With this specification, we can focus on the bivariate posterior for θ and ρ in order to make inferences about the presence of stochastic trends.

By repeatedly substituting the state equation into the measurement equation in (7) we can write:

$$y_t = x_t \beta + v_t,$$

where $x_t = (y_{t-1}, 1, t, \Delta y_{t-1}, ..., \Delta y_{t-p+1}), \beta = (\rho, \gamma')', \gamma = (\tau_0, \alpha, \pi_1, ..., \pi_{p-1})', k = p + 2$ and

$$v_t = e_t + \sum_{i=1}^{t} u_i.$$

Defining $y = (y_1, ..., y_T)'$, $X = (x'_1, ..., x'_T)'$ and treating p initial values of y_t as fixed[16] we obtain:

$$y \sim N(X\beta, \sigma_e^2 V).$$

Using the same prior as in previous cases plus untruncated uniform priors for the new parameters added, and integrating out σ_e^2, we obtain an expression for the joint posterior of θ and β:

---

[16] Note that, when we condition on p initial values, we are implicitly redefining T so that it is now equal to the old $T - p$. That is, we are treating our observed data as running from period $1 - p$ through T instead of as running from 1 through T as before. We maintain this convention throughout the remainder of this Appendix.

$$p_1(\theta, \beta | Data) \propto |V|^{-\frac{1}{2}} [(y - X\beta)'V^{-1}(y - X\beta)]^{-\frac{T}{2}}, \tag{A.2}$$

which is similar to (6).

To get the bivariate posterior for θ and ρ, we can integrate out γ using the presence of a Student-t kernel in (A.2), yielding:

$$p_1(\theta, \rho | Data) \propto |V|^{-\frac{1}{2}} |X^{*\prime}V^{-1}X^*|^{-\frac{1}{2}} s^{2^{-\frac{v}{2}}}, \tag{A.3}$$

where $v = T - k + 1$, X^* has $t'th$ row given by $x_t^* = (1, t, \Delta y_{t-1}, ..., \Delta y_{t-p+1})$,

$$s^2 = \frac{(y^* - X^*\widehat{\gamma})'V^{-1}(y^* - X^*\widehat{\gamma})}{\nu},$$

y^* has $t'th$ element given by $y_t^* = y_t - \rho y_{t-1}$ and $\widehat{\gamma} = (X^{*\prime}V^{-1}X^*)^{-1}X^{*\prime}V^{-1}y^*$. Using two-dimensional numerical integration we can calculate posterior properties of θ and ρ using equation (A.3). Bayes factors for the various hypothesis listed in Section 3 can be calculated using the Savage-Dickey density ratio.

Appendix B: MCMC Methods

In this Appendix, we describe MCMC methods for posterior inference in the evolving trend model of section 3 and the evolving seasonals model of section 4. The formulae below assume standard noninformative priors for any regression coefficients and σ_e^{-2}. However, adding a Normal prior for the regression coefficients and a Gamma prior for σ_e^{-2} can be easily done in the standard way or see de Jong and Shephard, 1995, section 5).

For the evolving trend model, conditional on knowing σ_e^2 and θ, the Gibbs sampler can be set up exactly as in de Jong and Shephard (1995).[17] In particular, our evolving trend model is exactly in the form as the model in Section 3 of de Jong and Shephard if we condition on p initial observations. Using their equations (2) and (4) modified for the inclusion of regression effects as in their Section 5, we can sample jointly from all the states and all regression parameters (conditional on σ_e^2 and the θ). In our experience, the de Jong-Shephard algorithm is highly efficient. Of particular value is the fact that it reduces the Gibbs sampler to three blocks. For the sake of brevity, we do not

---

[17] Fruhwirth-Schnatter (1994, 1995) and Carter and Kohn (1994) provide alternative methods for Gibbs sampling with state space models.

repeat the exact form of the algorithm here, but refer the reader to de Jong and Shephard (1995). Hence, if we can sample from $p(\sigma_e^{-2}|Data, \beta, \theta, \tau)$ and $p(\theta|Data, \beta, \sigma_e^2, \tau)$ we can complete our MCMC algorithm The conditional density of σ_e^{-2} is:

$$p(\sigma_e^{-2}|Data, \beta, \theta, \tau) = f_G(\sigma_e^{-2}|T, \frac{T}{\sum_{i=1}^{T} e_t^2}). \tag{B.1}$$

The conditional posterior for θ can be obtained by noting that θ is closely related to the variance of the state equation and $u_t = \Delta\tau_t$. The resulting conditional posterior is:

$$p(\theta|Data, \sigma_e^2, \tau) \propto \left(\frac{1-\theta}{\theta}\right)^{\frac{T}{2}} \exp\left(-\frac{1-\theta}{\theta}SSE\right), \tag{B.2}$$

where

$$SSE = \sum_{i=1}^{T} \frac{u_i^2}{2\sigma_e^2}.$$

This distribution is non-standard and, hence, we do not draw from it directly, but instead add a Metropolis-Hastings step to our MCMC algorithm, which is described below. Note that the use of the θ parameterization implies a complication to the MCMC algorithm, one reason for prefering deterministic integration rules for the evolving trend model.

If we had parameterized with $\lambda = \frac{\sigma_u^2}{\sigma_e^2}$ and used a flat prior for λ, then the resulting conditional posterior for λ^{-1} would be Gamma and, hence, λ is inverted-Gamma:[18]

$$p(\lambda|Data, \sigma_e^2, \tau) \propto \lambda^{-\frac{T}{2}} \exp\left(-\frac{SSE}{\lambda}\right). \tag{B.3}$$

The uniform prior for θ, which is truncated to ensure $0 \leq \theta < 1$, is proper and implies a prior for λ which is proportional to $\frac{1}{(1+\lambda)^2}$. This suggests as simple strategy for drawing from θ using a Metropolis-Hastings algorithm (see, for instance, Chib and Greenberg, 1995). Suppose the current draw of λ is called λ^{Old}. First take a candidate draw of λ from (B.3) using the inverted-Gamma distribution (call it λ^{New}). This draw is accepted with probability:

$$\frac{\frac{1}{(1+\lambda^{New})^2}}{\frac{1}{(1+\lambda^{Old})^2}},$$

---

[18]In the body of the paper, we include some MCMC results using the λ parameterization. These are obtained by combining this formula with a inverted Gamma prior for λ in the standard way.

where probabilities greater than one are rounded down to one. If the candidate draw is not accepted then the draw for λ remains λ^{Old}. Draws from λ can be converted into draws from θ using the fact that $\theta = \frac{\lambda}{1+\lambda}$.

The MCMC algorithm for the evolving seasonals model is developed along similar lines, except that θ is replaced by θ_i for $i = 0, 1, 2, 3$. In particular, the conditional distribution of σ_e^{-2} is:

$$p(\sigma_e^{-2}|Data, \phi^*, \delta_0, \delta_1, \delta_2, \delta_3, \tau) = f_G(\sigma_e^{-2}|T, \frac{T}{\sum_{i=1}^{T} e_t^2}). \tag{B.4}$$

The conditional posteriors for the $\theta_i's$ (for $i = 0, 1, 2, 3$) are:

$$p(\theta_i|Data, \sigma_e^2, \tau) \propto \left(\frac{1-\theta_i}{\theta_i}\right)^{\frac{T}{2}} \exp\left(-\frac{1-\theta_i}{\theta_i} SSE_i\right), \tag{B.5}$$

where

$$SSE_i = \sum_{j=1}^{T} \frac{u_{ij}^2}{2\sigma_e^2}.$$

Since these conditional posteriors are nonstandard, we use a similar Metropolis-Hastings step as described above. If we had parameterized with $\lambda_i = \frac{\sigma_i^2}{\sigma_e^2}$ and used a flat prior for λ_i, then resulting conditional posterior for λ_i would be inverted-Gamma:

$$p(\lambda_i|Data, \sigma_e^2, \tau) \propto \lambda_i^{-\frac{T}{2}} \exp\left(-\frac{SSE_i}{\lambda_i}\right). \tag{B.6}$$

We use the conditionals for λ_i as candidate generating densities in a Metropolis-Hastings algorithm. Suppose the current draw of λ_i is called λ_i^{Old}. First take a candidate draw of λ_i from (B.6) using the inverted-Gamma distribution (call it λ_i^{New}). This draw is accepted with probability:

$$\frac{\frac{1}{(1+\lambda_i^{New})^2}}{\frac{1}{(1+\lambda_i^{Old})^2}},$$

where probabilities greater than one are rounded down to one. If the candidate draw is not accepted then the draw for λ_i remains λ_i^{Old}. Draws from λ_i can be converted into draws from θ_i using the fact that $\theta_i = \frac{\lambda_i}{1+\lambda_i}$.

Output from these posterior simulators can be used to calculate posterior features of interest as well as the Bayes factor using the Savage-Dickey density ratio (see, for instance, Verdinelli and Wasserman, 1995, section 2.2)[19] or Chib's method. The results in the body of the paper indicate that Chib's method is a very reliable way of calculating the marginal likelihood. Note, however, that it requires the user to know the posterior and prior densities and the likelihood functions precisely. Knowing the kernels of these densities is not enough. With nonstandard priors (especially if they are truncated), figuring out the integrating constants of densities is difficult to do analytically. Of course, it is usually possible to figure out these integrating constants using prior simulation methods, but this adds to the computational and programming burden. Hence, when we use Chib's method in the body of the paper, we do not impose stationarity on the autoregressive coefficients. Furthermore, the use of Chib's method with the evolving seasonals model would require simulation from several different models (e.g. the unrestricted model, the model with a unit root at frequency π imposed, the model with a unit root at frequency $\frac{\pi}{2}$ imposed, etc.). Hence, we used the SDDR for the evolving seasonals model, which requires only posterior simulation from the unrestricted model.

We take 11,000 replications from our MCMC algorithm and discard the initial 1,000. Experimentation with different starting values (and the experience of other Bayesian state space modellers) indicates that our algorithm is well-behaved.

Appendix C: Priors and Parameterizations

In this Appendix, we discuss the issue of prior and parameterization choice for the case of state space models. We note that these issues are well-known in autoregressive models (e.g. Schotman, 1994). In the local level model, we parameterize the variance of the state equation in terms of the parameter:

$$\theta = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2} \tag{C.1}$$

which, as stressed in Section 2.1, has a natural interpretation relating to the variance of y_t conditional on y_{t-1}. Formally, we work with (θ, h_e), where $h_e = \sigma_e^{-2}$. Proper priors on both these parameters ensure that the posterior is proper and that meaningful Bayes factors can be calcu-

[19] Due to the difficulties of evaluating (B.5) at the point 0 due to division by zero, we evaluate it at a point close to zero. Formally speaking, this means we are testing the hypothesis that $\theta_i = 0.0001$ rather than $\theta_i = 0$. In practical applications the differences between these two hypotheses are negligeable.

lated.[20] We discuss the connection between the prior for (θ, h_e) and priors implied for two other commonly-used parameterizations, *viz.* (h_e, h_u) and (h_e, λ) where $h_u = \sigma_u^{-2}$ and $\lambda = \frac{\sigma_u^2}{\sigma_e^2}$.

Here, and in the material below, h_e is $f_G(\nu_e, s_e^{-2})$. For θ, we use a flat prior over the interval $[0, 1)$ and the stationary case corresponds to $\theta = 0$.

We begin by asking what this prior implies in the other parameterizations. Using the change of variable theorem, it can be seen that our prior implies:

$$p(h_e, h_u) \propto \frac{h_e^{\frac{\nu_e}{2}}}{(h_e + h_u)^2} \exp\{-\frac{h_e \nu_e s_e^2}{2}\}.$$

Note that our prior implies that h_e and h_u are not independent. If we condition on a value for h_e, it can be seen that this prior goes to zero as h_u goes to ∞ or zero, indicating inverted-U behavior in h_u space. This behavior is, of course, very different from the Jeffreys'-type prior[21] often considered: $p(h_u) \propto \frac{1}{h_u}$.

In the λ parameterization, we find that our prior implies:

$$p(h_e, \lambda) \propto \frac{h_e^{\frac{\nu_e - 2}{2}}}{(1 + \lambda)^2} \exp\{-\frac{h_e \nu_e s_e^2}{2}\},$$

which implies prior independence between h_e and λ. The marginal for the latter parameter is finite at $\lambda = 0$ and monotonically decreases to zero. Note that a flat prior for θ implies a prior for λ which has a Cauchy tail.

It is also interesting to begin in an alternative parameterization, elicit a sensible prior, and see what prior for θ is implied. In the alternative parameterizations, we assume informative Gamma priors. That is, h_u is $f_G(\nu_u, s_u^{-2})$ and λ is $f_G(\nu_\lambda, s_\lambda^{-2})$. The limiting cases with $\nu_u = 0$ or $\nu_\lambda = 0$ yield standard Jeffreys'-type priors, which are improper and will yield improper posteriors. These cases are to be avoided, but by setting ν_e or ν_λ to small but positive values one obtains a relatively noninformative, but proper, prior. Alternatively, one can work with these limiting case priors truncated to lie in some large but finite region.

If we had assumed that the prior for h_u was $f_G(\nu_u, s_u^{-2})$, we would have obtained the following prior for (h_e, θ):

[20]See Fernandez, Osiewalski and Steel (1997), which provides proofs on the existence of the posterior in a wide class of models, including state space models.

[21]Note that we refer to this as a "Jeffreys'-type prior" rather than a "Jeffreys' prior" since, in the local level model, the latter is quite complicated and is improper. Since we cannot use improper priors for parameters restricted under the null hypothesis for Bayes factor calculation, we do not investigate the Jeffreys' prior in this paper. For some background relevant for use of this prior see Shephard (1993).

$$p(h_e, \theta) \propto \frac{(\frac{1-\theta}{\theta})^{\frac{\nu_u-2}{2}} h_e^{\frac{\nu_e+\nu_u-2}{2}}}{\theta^2} \exp\{-\frac{h_e \nu_e s_e^2}{2}\} \exp\{-\frac{(\frac{1-\theta}{\theta}) h_e \nu_u s_u^2}{2}\},$$

a complicated prior which does not exhibit independence between its parameters. To better understand its behavior in "noninformative" cases, note that if we set $\nu_e = \nu_u = 0$, we obtain:

$$p(h_e, \theta) \propto \frac{1}{h_e(1-\theta)\theta}.$$

This limiting case is composed of the standard Jeffreys'-type prior for h_e and a U-shaped prior for θ which goes to infinity at 0 and 1.

If we had begun directly eliciting a prior for λ of the form $f_G(\nu_\lambda, s_\lambda^{-2})$, we would have obtained:

$$p(h_e, \theta) \propto \frac{(\frac{\theta}{1-\theta})^{\frac{\nu_\lambda-2}{2}} h_e^{\frac{\nu_e-2}{2}}}{(1-\theta)^2} \exp\{-\frac{h_e \nu_e s_e^2}{2}\} \exp\{-\frac{\frac{\theta}{1-\theta} \nu_\lambda s_\lambda^2}{2}\},$$

a complicated form with exhibits prior independence between the two parameters. The noninformative limiting case, $\nu_e = \nu_e = 0$, implies:

$$p(h_e, \theta) \propto \frac{1}{h_e\theta(1-\theta)}.$$

This limiting case is identical to that given above. That is, it is composed of the standard Jeffreys' type prior for h_e and a U-shaped prior for θ which goes to infinity at 0 and 1.

Hence, we have different "noninformative" priors which imply very different prior views about θ (i.e. uniform or U-shaped). This illustrates the great care that must be taken in prior elicitation, even when the researcher is striving to be noninformative. However, we have found that, for reasonably large sample sizes (e.g. $T > 100$) that the choice of prior has little effect on posterior inference. In a more serious empirical exercise, the researcher would likely have prior information which could be used to guide construction of a suitable informative prior.

References

[1] Aoki, M. (1990). *State Space Modeling of Time Series,* Second Edition. Berlin: Springer-Verlag.

[2] Bos, C., Mahieu, R.J. and van Dijk, H.K. (1999). "Relevance of alternative modelling decisions for hedging currency risk," manuscript, Econometric Institute, Erasmus University, Rotterdam.

[3] Canova, F. and Hansen, B. (1995). "Are seasonal patterns constant over time? A test for seasonal stability," *Journal of Business and Economic Statistics,* 13, 237-252.

[4] Carter, C. and Kohn, R. (1994). "On Gibbs sampling for state space models," *Biometrika,* 81, 541-553.

[5] Casella, G. and George, E. (1992). "Explaining the Gibbs sampler," *The American Statistician,* 46, 167-174.

[6] Chib, S. (1995). "Marginal likelihood from the Gibbs output," *Journal of the American Statistical Association,* 90, 1313-1321.

[7] Chib, S. and Greenberg, E. (1994). "Bayes inference for regression models with ARMA(p,q) errors," *Journal of Econometrics,* 64, 183-206.

[8] Chib, S. and Greenberg, E. (1995). "Understanding the Metropolis-Hastings algorithm," *The American Statistician,* 49, 327-335.

[9] DeJong, D. and Whiteman, C. (1991). "The temporal stability of dividends and stock prices: Evidence from the likelihood function," *American Economic Review,* 81, 600-617.

[10] Dickey, D. and Fuller, W. (1979). "Distribution of the estimators for autoregressive series with a unit root," *Journal of the American Statistical Association,* 74, 427-431.

[11] Fernandez, C., Osiewalski, J. and Steel, M. (1997). "On the use of panel data in stochastic frontier models with improper priors," *Journal of Econometrics,* 79, 169-193.

[12] Franses, P.H. (1996). *Periodicity and Stochastic Trends in Economic Time Series,* Oxford: Oxford University Press.

[13] Fruhwirth-Schnatter, S. (1994). "Data augmentation and dynamic linear models," *Journal of Time Series Analysis,* 15, 183-202.

[14] Fruhwirth-Schnatter, S. (1995). "Bayesian model discrimination and Bayes factors for linear Gaussian state space models," *Journal of the Royal Statistical Society, Series B*, 57, 237-246.

[15] Geweke, J. (1999). "Using simulation methods for Bayesian econometric models: Inference, development, and communication" (with discussion and rejoinder), *Econometric Reviews*, 18, 1-126.

[16] Hannan, E.J., Terrell, R. and Tuckwell, N. (1970). "The seasonal adjustment of economic time series," *International Economic Review*, 11, 24-52.

[17] Harvey, A. (1989). *Forecasting, Structural Time Series Models, and the Kalman Filter.* Cambridge: Cambridge University Press.

[18] Harvey, A. and Streibel, M. (1996). "Testing for a slowly changing level with special reference to stochastic volatility," manuscript.

[19] Hoek, H. (1997). "Variable trends: A Bayesian perspective," *PhD dissertation, Tinbergen Institute, Erasmus University.*

[20] Hylleberg, S., Engle, R., Granger, C., and Yoo, B. (1990). "Seasonal integration and cointegration," *Journal of Econometrics*, 44, 215-238.

[21] Hylleberg, S. and Pagan, A. (1997). "Seasonal integration and the evolving seasonals model," *International Journal of Forecasting*, 13, 329-340.

[22] Jeffreys, H. (1961). *Theory of Probability* (3rd edition), Oxford: Oxford University Press.

[23] de Jong, P. and Shephard, N. (1995). "The simulation smoother for time series models," *Biometrika*, 82, 339-350.

[24] Kass, R. and Raftery, A. (1995). "Bayes factors," *Journal of the American Statistical Association*, 90, 773-795.

[25] Kato, H., Naniwa, S. and Ishiguro, M. (1996). "A Bayesian multivariate nonstationary time series model for estimating mutual relationships among variables," *Journal of Econometrics*, 75, 147-161.

[26] Kim, S., Shephard, N. and Chib, S. (1998). "Stochastic volatility: Likelihood inference and comparison with ARCH models," *Review of Economic Studies*, 64, 361-393.

[27] Koop, G. (1992). "'Objective' Bayesian unit root tests," *Journal of Applied Econometrics*, 7, 65-82.

[28] Koop, G. and Potter, S. (2000). "Nonlinearity, structural breaks or outliers in economic time series?" in *Nonlinear Econometric Modelling in Time Series*, W. Barnett and S. Johansen (eds.). Cambridge University Press: Cambridge, forthcoming.

[29] Koop, G. and Potter, S. (1999a). "Dynamic asymmetries is U.S. unemployment," *Journal of Business and Economic Statistics*, 17, 298-312.

[30] Koop, G. and Potter, S. (1999b). "Are apparent findings of nonlinearity due to structural instability in economic time series," manuscript available at http://www.ed.ac.uk/~gkoop/.

[31] Kwiatkowski, D., Phillips, P.C.B, Schmidt, P. and Shin, Y. (1992). "Testing the null hypothesis of stationarity against the alternative of a unit root," *Journal of Econometrics*, 54, 159-178.

[32] Leybourne, S. and McCabe, B. (1994). "A consistent test for a unit root," *Journal of Business and Economic Statistics*, 12, 157-186.

[33] Min, C. (1992). "Unconditional estimation of time-varying parameter models: A Gibbs sampling approach," manuscript.

[34] Min, C. and Zellner, A. (1993). "Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates," *Journal of Econometrics,* 56, 89-118.

[35] Nerlove, M., Grether, M. and Carvalho, J. (1979). *Analysis of Economic Time Series: A Synthesis,* New York: Academic Press.

[36] Nyblom, J. and Makelainen, T. (1983). "Comparison of tests for the presence of random walk coefficients in a simple linear model," *Journal of the American Statistical Association*, 78, 856-864.

[37] Phillips, P.C.B. (1991). "To criticize the critics: An objective Bayesian analysis of stochastic trends," *Journal of Applied Econometrics*, 6, 333-364.

[38] Poirier, D. (1995). *Intermediate Statistics and Econometrics*, Cambridge: The MIT Press.

[39] Schotman, P. (1994). "Priors for the AR(1) model: Parameterization issues and time series considerations," *Econometric Theory*, 10, 579-595.

[40] Schotman, P. and van Dijk, H. K. (1991a). "A Bayesian analysis of the unit root in real exchange rates," *Journal of Econometrics*, 49, 195-238.

[41] Schotman, P. and van Dijk, H. K. (1991b). "On Bayesian routes to unit roots," *Journal of Applied Econometrics*, 6, 387-402.

[42] Shephard, N. (1993). "Distribution of the ML estimator of an MA(1) and a local level model," *Econometric Theory*, 9, 377-401.

[43] Shephard, N. (1994). "Partial non-Gaussian state space," *Biometrika*, 81, 115-131.

[44] Shively, T. and Kohn, R. (1997). "A Bayesian approach to model selection in stochastic coefficient regression models and structural time series models," *Journal of Econometrics*, 76, 39-52.

[45] Stock, J. (1994). "Unit roots and trend breaks in econometrics," in *Handbook of Econometrics*, vol. 4, R. Engle and D. McFadden (eds.). New York: North-Holland, 274-2781.

[46] Stock, J. and Watson, M. (1988). "Variable trends in economic time series," *Journal of Economic Perspectives*, 2, 147-174.

[47] Tanaka, K. (1996). *Time Series Analysis: Nonstationary and Noninvertible Distribution Theory*. New York: John Wiley and Sons.

[48] Verdinelli, I. and Wasserman, L. (1995). "Computing Bayes factors using a generalization of the Savage-Dickey density ratio," *Journal of the American Statistical Association*, 90, 614-618.

[49] West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models,* Second Edition. Berlin: Springer Verlag.