BAYESIAN ANALYSIS OF INTEGRATION AT DIFFERENT FREQUENCIES IN QUARTERLY DATA

Gianni Amisano

No.426

WARWICK ECONOMIC RESEARCH PAPERS

DEPARTMENT OF ECONOMICS

UNIVERSITY OF WARWICK
COVENTRY

BAYESIAN ANALYSIS OF INTEGRATION AT DIFFERENT FREQUENCIES IN QUARTERLY DATA

Gianni Amisano

No.426

August 1994

* Dipartimento di Scienze
Economiche
Università di Brescia
Via Porcellaga 21
25121 Brescia, Italy

** Department of Economics
University of Warwick
Coventry CV4 7AL
UK

BAYESIAN ANALYSIS OF INTEGRATION AT DIFFERENT FREQUENCIES IN QUARTERLY DATA

Gianni Amisano

Dipartimento di Scienze Economiche
Università di Brescia, Via Porcellaga 21
25121 Brescia, Italy

and

Dept. of Economics
University of Warwick
CV4 7AL Coventry
England

Tel: +39 30 2988307
Fax: +39 30 3770474
E-mail: amisano@bsvax.unibs.it

Abstract

The paper provides a unifying framework for conducting Bayesian inference on the presence of seasonal and zero frequency unit roots in quarterly data. The main technique used is the analysis of posterior odds ratios. A new parameterization is provided for the model and the prior distributions implemented are discusses and justified. The analysis relies heavily on the application of a Gibbs sampling algorithm. Such techniques render the Bayesian approach more flexible and implementable, giving the applied researcher the possibility of specifying a vast range of prior distributions. The methods are applied to a set of UK quarterly series. Compared to previous studies, less evidence is found to support seasonal integration hypotheses.

Keywords: Seasonal unit roots, Bayesian analysis, Posterior odds ratios, Monte Carlo Integration, Gibbs sampling.

This draft: July 1994.
I gratefully acknowledge precious suggestions from Ken Wallis and Jeremy Smith.
Denise R. Osborn provided me with the data. The usual disclaimer obviously applies.

Introduction

Interest about the non-stationary characteristics of macroeconomic time series is widespread. The debate among applied researchers was sparked by the work by Nelson and Plosser (1982), whose aim was to use sound statistical criteria in order to discriminate between competing non-stationary hypotheses, i.e. trend and difference stationarity. For univariate analysis under a classical statistical approach, the available statistical tools are the integration tests developed in Fuller (1976), Dickey (1976) Dickey and Fuller (1979, 1981), Phillips and Perron (1988), Schmidt and Phillips (1992). These tests aim at detecting the presence of a unit root in the univariate AR representation. A comprehensive exposition of this topic is provided by Hamilton (1994, chapter 17). The most difficult aspect related to the empirical application of these tests is the evident difficulty of discriminating, in small samples, between alternative models of non-stationarity which replicate sufficiently well the correlation properties of the series under analysis. This is the so-called "near observational equivalence" described in Sims (1989) and Campbell and Perron (1991).

It is well known that three characteristics of the classical inference approach present difficulties. First, it is known that the unit root test statistics have peculiar distributions under the null and the associated critical values have to be obtained numerically. Thus there exist a distributional asymmetry between the two hypotheses considered. Second, the Neyman-Pearson inferential apparatus assigns asymmetrical roles to the two hypotheses. Third, unit root tests share the alarming feature of having unsatisfactory power properties. DeJong, Narkervis, Savin and Whiteman (1992) conducted an interesting simulation exercise and report type II errors comparable to ones attained in a coin tossing game.

The Bayesian inference framework has been applied to the unit root testing problem relatively recently: see for example, Sims (1988), Schotman and van Dijk (1991a, 1991b, 1992), Phillips (1991, 1992), Phillips and Ploberger (1992), Geweke (1993). In the Bayesian framework it is possible to devise inferential strategies with properties that diverge substantially from those of the classical techniques. In the posterior odds ratio inference setting (see Zellner 1971), the hypotheses being compared are treated in a symmetric fashion, their relative plausibility being gauged on the basis of the corresponding posterior probability. The posterior distributions under both the hypotheses are not asymmetric, and the testing is fully "consistent", in that the probability of picking the wrong model goes to zero as the sample size increases. Moreover, since the relevance of the unit root inferential problem is often related to some decisions the applied researcher has to take in the model building process, i.e. difference the series, insert a trend, apply a seasonal frequencies filter, it seems conceptually appealing to be able to provide Bayesian techniques that have consistently proved to be a valid support to decision making in other fields of human activity.

The more general advantages and disadvantages of the Bayesian techniques also apply to this problem. The advantages, together with the ones already described, are essentially related to the fact that the Bayesian approach is simple, and it constitutes the only logical formalization of the process of learning. The disadvantages are mainly of a computational kind. Bayesian methods are "labour intensive techniques". When the attention of the researcher is confined to a subset of the parameters entering the model, one has to take the joint posterior distribution and marginalise it accordingly. This entails the use of analytical techniques, exact or approximated, and of numerical techniques, quadratures and Monte Carlo integration. Another important aspect is related to the specification of sound prior distributions. They have to reflect the state of a priori knowledge of the applied researcher, and not to be devised exclusively in order to minimise the computational burden.

In the present paper a Bayesian procedure is applied to testing the seasonal features of quarterly data. The effects of blind reliance on the published seasonally adjusted series are well known (see Wallis, 1974). In particular the adjustment of data prior to modelling might result in biased inference. We therefore consider it worthy to analyse the seasonal features of quarterly time series on a univariate basis. Given that the classical seasonal unit roots tests (like Hylleberg, Engle, Granger and Yoo (1990), henceforth HEGY) appear to share the poor properties of the zero frequency unit root test, we decided to take the less travelled road, and explore the application of Bayesian inference techniques.

We begin by devising a parameterization which seems to be particularly well suited for the inferential setting being proposed. Section 1 is devoted to describing the general characteristics of the AR model used, together with the particular parameterization chosen. The specification allows easy discrimination between deterministic and stochastic seasonality (in the form of the occurrence of seasonal unit roots), and between trend and difference stationarity. In section 2 the structure of the prior pdf's is presented, while section 3 describes how the joint posterior is dealt with via application of a Gibbs sampling algorithm. Section 4 is ancillary to this, since it presents the descriptions of the conditional posterior distribution for subsets of the parameter vector. Section 5 provides a description of the posterior odds ratio which seems to ease the computing burden. Section 6 contains the results of the application conducted on a set of UK quarterly series, and section 7 concludes. Appendices A and B deal with the strict technicalities of the analysis. Appendix A contains the detailed description of the conditional posterior distributions of different groups of parameters. Appendix B illustrates the key aspects of the rejection sampling algorithms being implemented.

[1] General features of the model

We consider an autoregressive model for quarterly data. Seasonality can be either deterministic or stochastic. Deterministic seasonality can be accounted for by introduction of dummy variables. Stochastic seasonality requires the application of the adequate filter to induce stationarity and raises the issue of the occurrence of seasonal cointegration (see HEGY, 1990 and Engle et al., 1993). We have stochastic seasonality when the AR polynomial contains some unit modulus roots at seasonal frequencies. In the quarterly case, the seasonal roots are -1, for the biannual cycle, and +/- i, for the annual cycle.

The model considered for the observable variable z_t is the following:

$\phi(L)\, y_t = e_t$, $e_t \sim N.I.D.(0, \sigma^2)$,

$$y_t = z_t - S_t - \gamma \cdot t ,$$

$$S_t = \alpha_0 + \alpha_1 \cos(\pi/2)t + \beta_1 \sin(\pi/2)t + \alpha_2 \cos \pi t .$$

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - ... - \phi_k L^k .$$

The hypotheses of interest concern the roots of the equation $\phi(L) = 0$, and are as follows:

1) $\phi(-1) = 0$, (integration at frequency $\lambda=\pi$: semi-annual cycle)
2) $\phi(i) = \phi(-i) = 0$ (integration at frequency $\lambda=\pi/2$: annual cycle)
3) $\phi(1) = 0$ (integration at frequency $\lambda=0$: the series is difference stationary)
4) Any combination of the above hypotheses.

Each one of the "null" hypotheses considered is compared to a parallel "alternative", in which the envisaged non-stationary feature is modelled with an appropriate deterministic component. The last part of this section is devoted to setting out the chosen parameterization. In the next section the structure of the prior used is outlined and justified.

The model is cast in terms of the parameterization used by HEGY, which makes use of the Laplace expansion of $\phi(L)$ around the roots +/-1, and +/-i:

$$\phi^*(L)y_{4t} = \psi_1 y_{1t-1} + \psi_2 y_{2t-1} + \psi_3 y_{3t-2} + \psi_4 y_{3t-1} + e_t , \qquad (1)$$

in which $\phi^*(L)$ is a polynomial in L with degree $k^*=k-4$, ψ_1, ψ_2, ψ_3 and ψ_4 are linear functions of the parameters in $\phi(L)$, and the variables y_{1t} , y_{2t}, y_{3t} and y_{4t} are defined in the following terms:

$$y_{1t} = (1+L+L^2+L^3)\, y_t ,$$
$$y_{2t} = -(1-L+L^2-L^3)\, y_t ,$$
$$y_{3t} = -(1-L^2)\, y_t ,$$
$$y_{4t} = (1-L^4)\, y_t = (1-L)\, S(L) y_t , \quad S(L) = 1+L+L^2+L^3.$$

In HEGY's setting the hypotheses of integration at different frequencies are represented as the following restrictions on the representation (1) :

integration at frequency $\lambda = 0$: $\psi_1 = 0$;
integration at frequency $\lambda = \pi/2$: $\psi_3 = \psi_4 = 0$;
integration at frequency $\lambda = \pi$: $\psi_2 = 0$.

We consider now how to use Bayesian inference techniques for the analysis of such hypotheses.In order to ease the implementation, a variant of the parameterization (1) is used, so as to represent the $\pi/2$ integration hypothesis as a restriction on a single parameter. In fact defining :

$\psi_3 = -2r \cos\theta$, $\psi_4 = 2r \sin\theta$,
we can write the model as:

$$\phi^*(L)y_{4t} = \psi_1 y_{1t-1} + \psi_2 y_{2t-1} + 2r(\sin\theta\, y_{3t-1} - \cos\theta\, y_{3t-2}) = e_t , \qquad (2)$$

or equivalently:

$$\phi^*(L)z_{4t} - \psi_1 z_{1t-1} - \psi_2 z_{2t-1} - 2r(\sin\theta\, z_{3t-1} - \cos\theta\, z_{3t-2}) =$$
$$= [4\phi^*(1) + 10\psi_1 + 2\psi_2 + 4r(\sin\theta - \cos\theta) - 4\psi_1 t]\gamma - 4\psi_1 \alpha_0$$
$$+ 4r\cos((\pi/2)t - \theta)\alpha_1 + 4r\sin((\pi/2)t - \theta)\beta_1 - 4\psi_2 \cos(\pi t)\alpha_2 + e_t , \qquad (3)$$

where:

$$z_{1t} = (1+L+L^2+L^3) z_t ,$$
$$z_{2t} = -(1-L+L^2-L^3) z_t ,$$
$$z_{3t} = -(1-L^2) z_t ,$$
$$z_{4t} = (1-L^4) z_t = (1-L) S(L) z_t .$$

We have integration at frequency $\pi/2$ when $r=0$; in such an occurrence, the parameter θ is indeterminate, and the parameters α_1 ,β_1 and θ disappear. When there is integration at frequency zero, the parameter α_0 disappears, and so does the trend term $-4\psi_1\gamma$ t. The model is difference stationary. Under the hypothesis of integration at frequency π, the parameter α_2 disappears.
It is evident that the model is not linear in the parameters involved. Nevertheless, we believe that it may provides the most correct framework to conduct inference, because it is based on a "structural" parameterization (see Barghava, 1986), Schmidt and Phillips (1992)). No parameter is redundant under any of the hypotheses considered.
The specification of prior distributions is bound to yield analytically intractable posterior distributions. For this reason, we resort to simulation methods, such as Monte Carlo integration with Gibbs Sampling, drawing on Geweke (1993) and Chib (1993).

[2] The specification of the priors

In order to describe the application of the simulation techniques used in this paper, the parameter of the model can be distinguished in 7 different groups: $\eta = [\eta_1' \ \eta_2' \eta_3 \eta_4 \eta_5 \eta_6 \eta_7]'$:

$$\eta_1 = \beta^* = [\beta' \gamma]', \ \beta = [\alpha_0\ \alpha_2\ \alpha_1\ \beta_1]'$$

$$\eta_2 = [\phi^{*\prime}]', \ \phi^* = [\phi^*_1 \ldots \phi^*_{K^*}]',$$

$$\eta_3 = \sigma, \ \eta_4 = \psi_1, \ \eta_5 = \psi_2, \ \eta_6 = r, \ \eta_7 = \theta.$$

As it will become clear in the next Sections, this division is done in order to associate these subsets of parameters with tractable conditional posterior distributions.
We also adopt the notation $\overline{\eta_i}$, i=1, ...7, to indicate that subset of parameters in η such that $\overline{\eta_i} \cup \eta_i = \eta$.

The following prior distribution structure is put forward:

$$p(\beta \; \gamma \mid \overline{\eta_1}) \sim N(b^*, \sigma^2 V^*) \qquad \beta \in R^4,$$

$$V^* = \text{diag}(-\psi_1^{-1}, -\psi_2^{-1}, r^{-1}, r^{-1}, (\sigma_\gamma/\sigma)^2],$$

$$\mathbf{b}^* = [\mathbf{b}' \quad \mu\gamma]'$$

$$p(\phi^*) \propto 1, \qquad \phi^* \in \mathbf{R}^{k^*}$$

$$p(\sigma) \propto \sigma^{-1}, \qquad \sigma \in \mathbf{R}_+,$$

$$p(\psi_i) = \lambda_i \exp(\lambda_i \psi_i), \qquad \psi_i \in \mathbf{R}_-, \quad i=1,2;$$

$$p(r) = \lambda_r \exp(-\lambda_r r), \qquad r \in \mathbf{R}_+;$$

$$p(\theta) \sim N[\mu_\theta, \sigma_\theta], \qquad \theta \in [-\pi, +\pi]. \qquad (4)$$

The choice of the priors is justified in the following way:

1)The prior on β, the parameters of the deterministic seasonal structure, is 4-variate normal, around a location vector **b** which is determined on the basis of the initial observations of the process. The prior variances of the single elements of β are designed to go to infinity as the model approaches the corresponding frequency integration setting. We have that:

$$\lim_{\psi_1 \to 0} V_{11} = \infty, \qquad \lim_{\psi_2 \to 0} V_{22} = \infty, \qquad \lim_{r \to 0} V_{ii} = \infty, \; i=3, 4,$$

but we have also:

$$\lim_{\psi_1 \to 0} \psi_1^2 V_{11} = 0, \qquad \lim_{\psi_1 \to 0} \psi_2^2 V_{22} = 0, \qquad \lim_{\psi_1 \to 0} r^2 V_{ii} = 0, \; i=3, 4.$$

i.e. the prior precisions go to zero, but slower than ψ_1^2, ψ_2^2 and r^2. This property is particularly important because it ensures that the deterministic component of the "reduced form" model has a logically sound prior distribution, and that the posterior distribution of the parameters under the stationary alternative passes smoothly to the

posterior distribution under the different integration hypotheses being considered. The analytical proofs of the smooth transition properties are available on request and are contained in Amisano (1994).
The linear trend parameter γ is given a normal prior, with position and variance specified by means of the two corresponding hyperparameters (μ_γ, σ_γ). Of course the choice of such hyperparameters is entirely subjective. The specification of a flat prior for γ would induce only marginal modifications to our analysis, and can be seen as a particular case, when the prior precision goes to zero.

2) The parameters in the transient AR dynamics, i.e. on the lags of y_{4t}, are given a flat prior, just to ease the computations. Of course the specification of more articulated priors is possible. For instance one follows Geweke (1993), and put a prior on each of them along the lines of Doan, Litterman and Sims (1984): normal prior with zero mean and prior variance that shrinks to zero as the lag order increases. This would not modify our analysis very much. It is nevertheless believed that we do not have such problems as overparameterization here, and we can focus on models with not too many lags.

3) The prior on the variance parameter σ is customarily a Jeffreys prior.

4) and (5) As for the parameters ψ_1 and ψ_2, which are associated with the hypotheses of zero frequency and π frequency integration, we choose to specify negative-exponential priors with hyperparameters λ_1 and λ_2. It is believed that such a functional form is quite appropriate because:
i) it does not force any restriction on the support of the parameters, other than the legitimate one of stationarity.
ii) it is a proper non- flat prior which therefore can be used in a posterior odds ratio testing framework involving sharp point nulls. The problems outlined in Schotman and van Dijck (1991a, 1991b, 1992) and regarded as occurrences of the "Lindley paradox" (see Lindley, 1957) in the specification of flat priors under a composite hypothesis are therefore solved.
The researcher has to provide a choice for the hyperparameters involved. Since these two parameters are the inverse of the respective prior means, in the absence of any extra-sample observation, one might choose them to be equal to the inverse of the unrestricted OLS estimates of the HEGY parameterization.
6) For r exactly the same kind of prior is chosen,except that the parameter r is defined to be positive.
7) The prior on θ, the phase angle in the deterministic $\pi/2$ seasonal, poses some problems. Although the parameter θ ceases to be identified under the $\pi/2$ frequency unit root hypothesis, it is not possible to assign it a prior dispersion determined on the basis of r. This would put obstacles in the way of the smooth transition results, which form the basis of the posterior odds ratio evaluation. Therefore we use a truncated normal distribution. The functional form chosen is normal because the domain of θ is not restricted: being an angle, it will be bound to lie between $-\pi$ and π. For this reason, we choose a symmetric prior. Other choices, such as Cauchy or Student-t, are equally legitimate. The choice of the hyperparameters is based on the OLS estimates of ψ_3 and

ψ_4 in the standard HEGY parameterization. Of course it is necessary to provide a check for sensitivity with respect to the specification of all the prior distributions. This comes with the results of the application contained in section 6.

[3] The joint posterior distribution

The likelihood of the model can be written as follows:

$$p(\text{data} \mid \eta\, \mathbf{D}_0) \propto \sigma^{-T} \exp\{-\frac{1}{2\sigma^2}\, \mathbf{e}'\mathbf{e}\}, \qquad (5)$$

$$\mathbf{e} = \{e_t\}_{t=1}^{T}$$

In the text the notation " $\mathbf{D}_t$ " means conditional on the data evidence up to period t, and therefore when we condition upon $\mathbf{D}_0$ or $\mathbf{D}_T$, we respectively indicate "conditional on initial conditions" or "conditional on the whole sample information" (a posteriori).
Combining the information provided by the prior distribution with the likelihood function, we obtain the joint posterior:

$$p(\eta \mid \mathbf{D}_t) \propto \sigma^{-(T+5)} (-\psi_1)^{1/2} (-\psi_2)^{1/2}\, r\, \exp\{-\frac{1}{2\sigma^2}[\varepsilon'\varepsilon + \sigma_\theta^{2}(\theta-\mu_\theta)^2 + \\ +(\sigma/\sigma_\gamma)^2 (\gamma-\mu_\gamma)^2] + \lambda_1\psi_1 + \lambda_2\psi_2 - \lambda_r r\},$$

$$\mathbf{e} = \begin{bmatrix} \mathbf{e} \\ \mathbf{V}^{*-1/2}(\beta^* - \mathbf{b}^*) \end{bmatrix} \qquad (6)$$

When it comes to conducting inference on a subset of parameters of interest, on the basis of the posterior pdf, it is clear that we have to be able to marginalise it with respect to the parameters we are not interested in, i.e. the nuisance parameters. Expression (6) does not allow the possibility of easily obtaining marginal distributions or posterior moments on the basis of available analytical results. Another possibility is to resort to an approximations, such as that described by Phillips (1983) and used in a Bayesian setting by Tierney and Kadane (1986). But it is still difficult to obtain manageable results. On the other hand, numerical integration is not feasible, given the high dimensionality of the parameter space (we have in total k+6 parameters).One has then to rely on fast, efficient and precise numerical simulation techniques.
Suppose that our interest focused on the posterior mean of some function of the parameters, say $g(\eta)$:$E(g(\eta)|D_T) = \int g(\eta)\, p(\eta|D_T)\, d\eta$. We show in section 4 that the posterior odds ratio (henceforth POR) can be thought of as the posterior expectation of a particular function of certain parameters. The posterior moments can be computed numerically to an arbitrary degree of accuracy on the basis of the Monte Carlo integration principle; see Hammersley and Handscomb (1964) or Geweke (1989). If it were possible to obtain m draws from the joint posterior pdf, say $\eta^{(i)}$, i=1,2,...m, then the posterior expectation of $g(\eta)$ could be easily estimated as the sample mean:

$$\overline{g}_m = m^{-1} \sum_{i=1}^{m} g(\eta^{(i)}). \qquad (7)$$

Given an i.i.d. assumption on the draws, the law of large numbers ensures convergence of the above expression to the posterior expectation of $g(\eta)$. Of course as m increases, so does the accuracy of the estimate.

If it is not possible to provide i.i.d. draws from the joint posterior distribution, as in our case where it is of no known analytical form, then some other methods have to be adopted. Following the suggestions of Hammersley and Handscomb (1964) one could choose an "importance function", to sample from. In any case that choice is not easy, and it might yield very poor estimates. In fact, as it is stressed in Koop (1994), it is necessary that the tails of the importance distribution be fatter than those of the posterior distribution, otherwise the draws from the tails of the importance function dominate the behaviour of the Monte Carlo estimate. For this reason, one should know exactly the shape of the posterior distribution, in order to choose correctly the importance function. We do not know the form of the joint posterior in our context, and therefore we follow Chib (1993) and Geweke (1993), in adopting a Gibbs Sampling Algorithm (henceforth GSA). This algorithm is being increasingly applied in the Bayesian literature since it provides a feasible way of computing posterior moments. Geman and Geman (1984) introduced the technique, and Gelfand and Smith (1990) and Smith and Roberts (1993) give interesting discussions about the interrelation between different numerical methods. Tierney (1991) contains a thorough discussion of the GSA properties.

The idea behind the GSA is quite simple and intuitive. Suppose that all we know is the mathematical expression for the joint posterior, but this is of no known analytical form. Suppose further that the conditional posterior distributions of a class of mutually exclusive exhaustive subsets of the parameters $p(\eta_i|\overline{\eta_i}\ \mathbf{D}_T)$ are "available" in the sense that each of them can be easily simulated. If these conditions are met the GSA works as follows. We start from an arbitrary initialisation of the parameter vector:

$$\eta^{(0)} = [\eta_1^{(0)'}, \eta_2^{(0)'}, \ldots\ldots, \eta_k^{(0)'}]'.$$

At each pass of the algorithm, a random draw from each of the $p(\eta_i|\overline{\eta_i}\ \mathbf{D}_T)$ distributions is obtained, and the results from the draw are used to condition the posterior distributions in the next pass.

What we have is hence a Markovian updating scheme, in which the draws are not independent, nor identically distributed. Geman and Geman (1984) show that mild conditions ensure that: 1) the continuous state Markov chain induced by the GSA converges in distribution to the true joint posterior distribution at a rate which is geometric in the number of passes used in the algorithm. 2) The numerical approximation of the posterior mean of any function of the parameters (if exists) converges a.s. to its true value. Geweke (1993, section 3.1) proves corresponding results for the GSA as applied to an AR model plus trend with fat-tailed disturbances. What has to be shown is how to characterise the conditional posterior pdf's in our context, and how it is possible to obtain random drawings from them. This forms the object of the next section.

[4] The conditional posterior distributions

Even if the joint posterior distribution hasno known analytical form, it turns out to be possible to characterize the conditional posterior distributions of some subsets of parameters. Some of these pdf's are of known analytical form, whereas others are not; in appendix B we will describe how it is possible to obtain synthetic draws from whichever univariate distribution through the method of "rejection sampling" (see Devroye 1986). Starting from expression (6) we can readily obtain some results for the 7 different groups of parameters in the model. These results are presented as lemmas, whose proofs are contained in appendix A.

Lemma 4.1

$p(\eta_1|\overline{\eta_1}\, D_T)$, where $\eta_1 = [\beta' \ \gamma]' = \beta^*$, is 5-variate normal, from which independent random draws are readily obtained.

Lemma 4.2

$p(\eta_2|\overline{\eta_2}\, D_T)$ is k^*-variate normal, and again independent random draws are readily obtained. Remember that $\eta_2 = [\phi^*]$ is a (k^*x1) vector containing the parameters on the lags of $y_{4t} = z_{4t} - 4\gamma$.

Lemma 4.3

$p(\eta_3 \mid \overline{\eta_3}\, D_T)$, where $\eta_3 = \sigma$, allows for indirect drawing through a χ^2 distribution.

Lemma 4.4

$p(\eta_4|\overline{\eta_4}\, D_T)$, with $\eta_4 = \psi_1$, $p(\eta_5|\overline{\eta_5}\, D_T)$, with $\eta_5 = \psi_2$, and $p(\eta_6 \mid \overline{\eta_6}\, D_T)$, with $\eta_6 = r$, have no standard form. Their simulations require "rejection sampling" (see Devroye, 1986, and Geweke, 1993). In appendix B the method is explained, with a complete account of the choices made in terms of functional form for the reference distributions and their parameters.
To summarize, the last two sections show how the resulting joint posterior distribution is not of any analytically known kind. Nevertheless, given that we are able to draw independently from an exhaustive set of conditional posterior pdf's. Putting these draws into a Markov chain sequence, we can apply a Gibbs sampling algorithm to obtain

synthetic draws from the joint posterior pdf. These draws form the basis for the evaluation of the posterior moments of any function of the parameters. We next show that the posterior odds ratios of hypotheses can be seen as posterior means of certain functions of the parameters.

[5] A convenient description of the posterior odds ratio

We can now describe the POR for some hypotheses of interest. It can be thought of as representing the posterior expectation of an aptly defined function of the parameters in the model. Suppose we are faced with the problem of deciding between two competing hypotheses H_A and H_B which concern η, the parameter vector of our model. Associated to the two hypotheses we have, as usual, two families of priors, $p_A(\eta)$ and $p_B(\eta)$.

The POR is defined as:

$$POR = \frac{p(H_A|data)}{p(H_B|data)} = \frac{\int p_A(\eta)p(data|\eta\ H_A)d\eta}{\int p_B(\eta)p(data|\eta\ H_B)d\eta}. \tag{8}$$

We conjecture that the two hypotheses involve the same set of parameters, and so we are allowed to write quite easily:

$$POR = \int \frac{p_A(\eta)}{p_B(\eta)}\ p(\eta|\mathbf{D}_T\ H_B)\ d\eta, \tag{9}$$

i.e. the POR can be seen as the posterior expectation, under hypothesis B of the function: $g(\eta)= p_A(\eta)/p_B(\eta)$.

Therefore one might obtain a set of synthetic draws from the posterior distribution of the parameters in the model under H_B, by application of the Gibbs sampling algorithm. For each draw the function $g(\eta^{(i)})$ can be computed, and its posterior expectation can be estimated as the sample mean (7).

The context is only slightly more complicated when we consider comparing a point hypothesis against a composite competing hypothesis. We illustrate the case by referring it directly to our model, where we are interested in gauging the posterior evidence in support of the presence of different unit moduli roots. The sharp point hypothesis can consider just one unit root at a time or more, and it compares directly to the stationary alternative specification.

For explanatory purposes, we restrict attention to the $\pi/2$ frequency integration hypothesis, and consider a comparison between H_A: $r=0$ and H_B: $r>0$. Comparisons involving different frequencies integration hypotheses (even between joint hypotheses) can be conceptually dealt with on the basis of exactly the same framework.

We consider H_A as the limiting expression, for ε approaching zero, of the following hypothesis:

$$H_A(\eta) : r \in (0, \varepsilon), \tag{11}$$

$$p_A(\eta) = \varepsilon^{-1} I_{(0;\varepsilon)}(r) , r \in (0, \varepsilon),$$

with $I_{(0;\varepsilon)}(r)$ the indicator function, which takes unit values within the $(0, \varepsilon)$ interval and is equal to zero elsewhere.

For homogeneity we restore the notation used in section 2 and indicate r as η_6; all the other parameters in the model are collected in the set $\overline{\eta_6}$, and the set η^*indicates the parameter vector in the likelihood function of the model under the $\pi/2$ frequency integration hypothesis, i.e. all the parameters in η except α_1, β_1, θ and r.

As for the parameters in $\overline{\eta_6}$, we adopt the prior pdf specification discussed in section 2, i.e. we have $p_A(\overline{\eta_6}|\eta_6) = p_B(\overline{\eta_6}|\eta_6)$.

As for η_6, we specify the same prior as in section 2 under hypothesis B; under H_A, a flat prior is adopted as described in expression (11).

We can therefore write the POR as follows:

$$POR= \int [\int \frac{p_A(\eta_6)}{p_B(\eta_6)} p(\eta_6| \overline{\eta_6} \mathbf{D}_T H_B) d\eta_6] \; p(\overline{\eta_6} | \mathbf{D}_T H_B) d\overline{\eta_6} \; . \tag{12}$$

The posterior expectation of the function $g(\eta_6) = p_A(\eta_6)/p_B(\eta_6)$, conditional on the other parameters in $\overline{\eta_6}$, has to be averaged by using the posterior pdf of $\overline{\eta_6}$ as a weighting function. From the discussion in section 2, we already know the form of the conditional posterior distribution of η_6 (see also appendix B). In what follows, we show what happens when we contemplate $\varepsilon \rightarrow 0$. We make use of the smooth transition results, as described in Amisano (1994).

It is easy to see that the function of interest, i.e. $g(\eta_6) = \varepsilon^{-1} I_{(0,\varepsilon)}(r) \lambda_T^{-1} \exp(\lambda_T r)$, only depends on η_6. Theoretically one could marginalise out all the parameters but η_6, and evaluate the POR as posterior expectation of $g(\eta_6)$ on the basis of the uni-dimensional posterior pdf of η_6. That is analytically impossible. What we can do, in order to make efficient use of the numerical evaluation techniques being used, is to marginalise with respect to the parameters of the deterministic representation that disappear when r=0,namely α_1 and β_1.

We define $\overline{\eta}_6^*$ such that :

$$\overline{\eta}_6^* = \overline{\alpha_1 \; \beta_1 \; r}'$$

i.e. $\overline{\eta}_6^*$ is the vector of all the parameters bar α_1, β_1, and r. We can write the POR as follows:

$$POR = \int\left[\int \frac{p_A(\eta_6)}{p_B(\eta_6)}\; p\left(\eta_6 | \overline{\eta}_6^*\, \mathbf{D}_T\, H_B\right) d\eta_6\right] p\left(\overline{\eta}_6^* |\, \mathbf{D}_T\, H_B\right) d\overline{\eta}_6^* \qquad (13)$$

The smooth transition results then allow us to write:

$$p(\eta_6 |\, \overline{\eta}_6^*\, \mathbf{D}_T\, H_B) = [1+8\,T\,r]^{-1} \exp\{(-1/(2\sigma^2))\, \mathbf{w}^{*'} M(\mathbf{X}^*)\mathbf{w}^* - \lambda_r\, r\}/\, k(\overline{\eta}_6^*),$$

$$k(\overline{\eta}_6^*) = \int_0^\infty [1+8\,T\,r]^{-1} \exp\{(-1/(2\sigma^2))\, \mathbf{w}^{*'} M(\mathbf{X}^*)\mathbf{w}^* - \lambda_r\, r\}\, d\eta_6$$

where the [(T+2)×1] vector $\mathbf{w}^*$ and the [(T+2)×2] matrix $\mathbf{X}^*$ are defined as:

$\mathbf{w}^* = [r^{1/2} a_1\, ,\, r^{1/2} b_1,\ \mathbf{w}']$, $\quad \mathbf{X}^{*\,'} = [r^{1/2}\, I_2\ ,\ \mathbf{X}]'$,

$\mathbf{w} = \{w_t\}_{t=1}^{T}$, $\qquad \mathbf{X} = \{x_t'\}_{t=1}^{T}$,

$w_t = \phi^*(L) y_{4t} - \psi_1\, y_{1t-1} - \psi_2\, y_{2t-1} - 2\, r\, [\sin\theta(z_{3t-1} + 2\,\gamma) - \cos\theta\, (z_{3t-2} + 2\gamma)]$,

$x_t' \;=\; 4\, r\, [\cos((\pi/2)t-\theta)\ |\ \sin((\pi/2)t-\theta)]$,

and $M(\mathbf{X}^*) = I - \mathbf{X}^*(\mathbf{X}^{*'}\mathbf{X}^*)^{-1}\mathbf{X}^{*'}$.

Therefore the POR is:

$$POR = \int [[\int_0^\varepsilon \frac{\exp\{-1/(2\sigma^2)\mathbf{w}^{*'} M(\mathbf{X}^*)\mathbf{w}^*\}}{\varepsilon\, \lambda_r [1+8\ \ T\, r]}\, dr]/k(\overline{\eta}_6^*)]\ \ p(\overline{\eta}_6^* |\, \mathbf{D}_T\, H_B)\, d\overline{\eta}_6^*\ .$$

Given the smooth transition results, we have that:

$$\lim_{r\to 0}\ [1+8\,T\,r]^{-1} = 1, \qquad \lim_{r\to 0}\ \mathbf{w}^{*'} M(\mathbf{X}^*)\mathbf{w}^* = \mathbf{e}_A'\, \mathbf{e}_A\, ,$$

where $\mathbf{e}_A$ is the vector of the error terms in the model under the $\pi/2$ frequency integration hypothesis:

$e_{At} = \phi^*(L)y_{4t} - \psi_1\, y_{1t-1} - \psi_2\, y_{2t-1}$.

Now we are ready to consider what happens to the above quantity as ε shrinks to zero, and the prior distribution of r under H_A attains unit probability mass at $\eta_6 = r = 0$. The limiting expression is obtained as:

$$\mathrm{POR} = \int_{\overline{\eta^*_6}} [\lambda_r^{-1}\exp\{(-1/(2\sigma^2))\mathbf{e}_A'\,\mathbf{e}_A\} / k(\overline{\eta^*_6})]\ p(\overline{\eta^*_6} \mid \mathbf{D}_T\, H_B)d\,\overline{\eta^*_6}\ . \qquad (14)$$

This means that the function of interest :

$$g(\overline{\eta^*_6}) = \lambda_r^{-1}\exp\{(-1/(2\sigma^2))e_0'\,e_0\} / k(\overline{\eta^*_6}),$$

which depends on all parameters of the model but α_1, β_1 and r, is evaluated for any draw from the posterior distribution of $\overline{\eta_6}^*$, and then averaged numerically, on the basis of the draws obtained from the posterior distribution under H_B.The application of the GSA renders this approach feasible.

The above framework applies to any unit root hypothesis comparison with the associated stationary alternative, and the corresponding POR can be obtained exactly in the same conceptual way. For this reason the inferential strategy for the problem under study is as follows: we make use of the Gibbs sampling algorithm to generate draws from the joint posterior distribution. At each pass of the sampler we keep track of the relevant functions of the parameters. These functions are averaged to yield the posterior odds ratios.

[6] An application

This section presents an application to five of the UK macroeconomic series studied in Osborn (1990), where details of the source of the data can be found. We consider real GDP, total real consumption (including durables and non-durables), real investment (total gross fixed capital formation), employment, and real narrow money (M0). All the series are in natural logarithms. Data run from 1955, first quarter, to 1988, last quarter, except for M0, for which a shorter sample period is available (1969:3-1988:4).
For these variables, the application of HEGY' s testing procedures led to the conclusion that GDP and consumption possess unit roots at all frequencies, whereas investment, employment and real money have only a zero frequency unit root (Osborn, 1990, Table 2). The application of the Bayesian technique only partly confirms these results, as shown below.

Before presenting the results, some explanation of how the univariate models were specified and how prior distribution hyperparamenters were chosen is required.
The model lag order was chosen on the basis of the application of a series of different criteria: informative criteria (Akaike, Hannan and Quinn and Schwartz), variable deletion tests on an over-paramerized general model, and Godfrey portmanteau test to check the validity of the resulting model. It emerged that all the series being analysed required an autoregressive representation of the 5th order.
As for the deterministic parameters α_0, α_2 α_1 and β_1, their location hyperparameters a_0, a_2 a_1 and b_1, are determined on the basis of the pre-sample observations, treated as initial conditions of the underlying processes. The parameters in the prior distribution of γ, μ_γ and σ_γ are determined such that μ_γ matches the average in-sample growth rate of the series, and such that the prior distribution assigns 95% of the whole probability mass to the interval $\mu_\gamma \pm 2\%$. The hyperparameters of the prior distributions of ψ_1, ψ_2, and r, i.e. respectively λ_1, λ_2 and λ_r were determined on the basis of the following procedure. A unrestricted AR process was fitted to the data:

$$\phi(L)z_t = \sum_{i=1}^{4} \delta_i \, D_{it} + \varsigma \, t + e_t,$$

and estimated by means of the OLS estimator. On the basis of these estimates, indirect estimates for ψ_1, ψ_2, and r were provided for all the series under study. The reliability of these estimates has been previously gauged on the basis of a Monte Carlo esperiment. This experiment points out that these indirect estimates have a well behaved, bell-shaped distribution around the true values. The indirect estimates, $\tilde{\psi}_1$, $\tilde{\psi}_2$, $\tilde{r}$,form the basis of the choice of the hyperparameters. In the absence of any a priori information, recall that the parameter of a negative exponential distribution is the reciprocal of its expected value. On the basis of this consideration the hyperparameters were determined as:

$$\lambda_1 = \frac{1}{\tilde{\psi}_1}, \ \lambda_2 = \frac{1}{\tilde{\psi}_2}, \ \lambda_r = \frac{1}{\tilde{r}}. \qquad (15)$$

While this choice seem plausible and sound, it may have an important bearing on the analysis. Consequently sensitivity analysis of the results with respect to different choices of such hyperparameters is carried out, and the results are discussed at the end of this section. Finally for the hyperparameters of the prior distribution of θ, μ_θ is chosen equal to zero and σ_θ is determined as that value that gives 95% of the Gaussian probability mass to the $(-\pi,+\pi)$ interval.
The hyperparameters used in the application are summarized in Table 1, presented here.

Table 1: Hyperparameters

	GDP	Consumpt.	Investm.	Employm.	M0
a_0	10.557	10.190	8.712	10.105	8.208
a_2	0.004	0.023	0.046	-0.001	0.016
a_1	-0.008	-0.023	0.001	-0.003	-0.020
b_1	0.021	0.023	0.019	-0.000	0.003
μ_γ	0.006	0.007	0.009	0.001	0.009
σ_γ	1.020	1.020	1.020	1.020	1.020
λ_1	49.03	39.781	75.803	57.250	69.905
λ_2	5.915	6.463	5.107	2.749	2.028
λ_r	5.851	5.436	4.848	3.082	2.859
μ_θ	0.000	0.000	0.000	0.000	0.000
σ_θ	1.603	1.603	1.603	1.603	1.603

On the basis of these hyperparameters, the resulting joint posterior distributions were simulated via application of a Gibbs sampling algorithm [1]. The number of iterations used was 2,000, plus a batch of 100 unretained iterations used to "warm up" the sampler. The results obtained include not only the posterior odds ratios (see Table 2), but also the posterior mean of the parameters, collected in Table 3, and the marginal posterior distributions which are graphed in Figures 1 to 5.

Table 2: Posterior odds ratios

	GDP	Consumpt.	Investm.	Employm.	M0
zero freq.	1.753	3.629	2.449	0.510	2.113
$\pi/2$ freq.	0.049	0.130	0.003	0.001	0.001
π freq.	0.744	1.162	0.433	0.022	0.055

Table 3: Posterior means

	GDP	Consumpt.	Investm.	Employm.	M0
α_0	10.646	10.058	8.774	10.076	2.475
α_2	-0.014	0.013	-0.002	0.001	0.002
α_1	-0.014	0.027	0.045	0.001	-0.003
β_1	-0.019	-0.024	0.009	-0.003	0.001
γ	0.006	0.007	0.009	0.001	0.013
ϕ_1	0.292	0.457	0.446	0.301	0.264
σ	0.022	0.016	0.036	0.005	0.008
ψ_1	-0.012	-0.005	-0.005	-0.008	-0.003
ψ_2	-0.148	-0.120	-0.160	-0.034	-0.618
r	0.175	0.151	0.186	0.032	0.309
θ	-0.012	-0.354	-0.292	-0.988	-0.855

[1] All the Monte Carlo and Bayesian analysis computations were performed with software developed by the author and written in GAUSS 2.1. The preliminary analysis for the choice of the lag length was done by means of RATS 4.02 routines.

The results can be summarised as follows.

1) GDP: the posterior odds ratio analysis seems to clearly favour the hypothesis of zero frequency integration (POR=2.479). The posterior mean of ψ_1 is very low (-0.012), and its posterior distribution assigns high probability mass to the immediate neighbourhood of zero (see Figure 1). The posterior odds ratio instead soundly rejects the hypothesis of $\pi/2$ frequency integration (POR=0.049). This is confirmed by the value of the posterio mean of r (0.175) and by the shape of its posterior distribution, which assigns almost no weight to values near to zero.The possible presence of a π frequency unit root is more controversial (POR=0.74). The posterior distribution of ψ_1 assigns a non-negligible probability mass to values near to zero, although the mode of the distribution is well distant from zero. Considering all these results together, one might cautiously assume that the series has a non-seasonal unit root, but that its seasonality might be dealt with by seasonal dummies. This contrasts with Osborn's results.

2) Consumption: the posterior odds ratio leads to clear acceptance of the long run unit root hypothesis (POR=3.629). The posterior mean of ψ_1 is close to zero (0.005), and the whole posterior distribution is concentrated near zero (see Figure 2). As for the $\pi/2$ frequency integration hypothesis, (POR=0.134), it is squarely rejected by the data, and the marginal posterior distribution of r gives all its weight to values well away from zero (posterior mean=0.151). Data are not conclusive on the issue of the presence of a π frequency unit root (POR=1.16): the posterior distribution of ψ_2 has mean equal to -1.120, mode equal to .0.06, but gives high weight to values near to zero. Varying the values of hyperparameters did not help to resolve uncertainty: the POR remained close to 1 for all the prior configurations being specified. Data are simply not very informative in this respect. Therefore one could weakly favour the presence of a bi-annual stochastic cycle in the data, but not the presence of an annual cycle. This again contrasts with Osborn's findings.

3) Investment: again for this series the presence of a zero frequency unit root seems unquestionable (POR=2.449): the posterior distribution of ψ_1 (see Figure 3) is squeezed to the immediate left of zero, with a posterior expectation of -0.012. The results contradicts the presence of a $\pi/2$ frequency unit root, given that the POR is 0.003, and the posterior distribution of r does not assign any weight to the neighbourhood of zero; its posterior mean is 0.186. Likewise the model rejects the hypothesis of π frequency unit root (POR=0.433), and the posterior distribution of ψ_2 does not give the neighbourhood of zero substantial probability mass. For this series, one could thus conclude that, first, the series is I(1) in the conventional sense, and second, that non-stationary stochastic seasonality can be ruled out, receiving no support from the posterior analysis. Deterministic seasonals account for the seasonal pattern. This is in perfect accordance with Osborn's results.

4) Employment: the presence of a zero frequency unit root is rejected by the data, the alternative hypothesis being preferred in the light of the POR (0.510). This is in sharp contrast with Osborn's findings concerning this series. The posterior distribution of ψ_1

(see Figure 4) has mean -0.008, mode -0.009, and it does not give much weight to values near zero. Similarly, but more neatly, the posterior analysis reject the hypotheses of π and $\pi/2$ frequencies integration (POR = 0.001 and 0.022 respectively). Also the posterior distributions of ψ_1 and ψ_2 are both clearly distant from zero. The series is therefore taken to be stationary around a deterministic linear trend with seasonal intercept shifts.

5) Real M0: the zero frequency integration hypothesis is clearly accepted on the basis of a POR of 2.113. On the contrary, both the hypotheses of seasonal integration are rejected on the basis of the posterior odds ratios (0.001 and 0.055, respectively). Also the examination of the posterior distributions of ψ_1 and ψ_2 are consistent with these findings (see Figure 5). This is consistent with OSborn's results.

As a partial corroboration of these findings, a sensitivity analysis experiment has been carried out. For the sake of brevity, we consider only the GDP series. Clearly, given the high dimensionality of the hyperparameter space, it is not feasible to monitor the effects of changes on all hyperparameters, and we focus only on the most crucial ones, that is, those controlling the prior distributions of ψ_1, ψ_2 and r. The prior hyperparameter specification (15), which produces the benchmark prior 1, is modified to generate another 4 priors along the following lines:

prior 2: $\lambda_1 = \frac{2}{\tilde{\psi}_1}$, $\lambda_2 = \frac{2}{\tilde{\psi}_2}$, $\lambda_r = \frac{2}{\tilde{r}}$,

prior 3: $\lambda_1 = \frac{4}{\tilde{\psi}_1}$, $\lambda_2 = \frac{4}{\tilde{\psi}_2}$, $\lambda_r = \frac{4}{\tilde{r}}$,

prior 4: $\lambda_1 = \frac{0.5}{\tilde{\psi}_1}$, $\lambda_2 = \frac{0.5}{\tilde{\psi}_2}$, $\lambda_r = \frac{0.5}{\tilde{r}}$,

prior 5: $\lambda_1 = \frac{025}{\tilde{\psi}_1}$, $\lambda_2 = \frac{0.25}{\tilde{\psi}_2}$, $\lambda_r = \frac{0.25}{\tilde{r}}$.

Prior distributions 2 and 3 are more and more squeezed near zero values, whereas priors 4 and 5 assign greater weight to values distant from zero. For each one of these prior specifications, the posterior analysis described above was repeated out in exactly the same terms. The results in terms of the associated posterior odds ratios are presented in Table 4.

Table 4: Sensitivity analysis, GDP series

	prior 1	prior 2	prior 3	prior 4	prior 5
zero freq.	1.753	2.201	2.485	1.551	1.547
$\pi/2$ freq.	0.049	0.079	0.401	0.038	0.025
π freq.	0.744	0.811	0.899	0.626	0.555

As one can easily see, these changes to the prior specification do not radically alter the nature of the results. As would be expected, priors 2 and 3 tend to give a higher

posterior probability to the integration hypotheses, whereas priors 4 and 5 tend to favour the alternative hypotheses. These results seem encouraging and are interpreted as giving strength to the findings of this paper.
Summing up, the Bayesian approach we propose is helpful in shedding new light on the inferential problem connected to the presence of unit roots at different frequencies. It is a sensible approach because it is base on a sensible parameterisation, and it allows to treat symmetrically all the hypotheses being tested. No use of asymptotics is made, and all the relevant posterior distributions are exact. In the particular application run here, the procedure seems to work well enough, giving in most of the cases a clear response to the issue of the presence of unit roots. The results seem to be robust with respect to alternative sensible prior specifications.

[7] Conclusion

The paper presents a new testing procedure to ascertain the presence of unit roots at different frequencies in quarterly data. Given the weaknesses and logical inconsistencies of the classical inference setting, the proposed procedure is Bayesian, and relies on posterior odds ratio computations. Special emphasis is placed ondevising a sensible prior distribution specification. The resulting joint posterior distribution is treated by means of a Gibbs sampling algorithm.
The procedure is applied to a set of UK series, previously analysed by Osborn (1990).In contrast to her results, less evidence was found in favour of non stationary stochastic seasonality, which seems to occur only for the consumption series. For the employment series it was found that the trend stationary alternative is preferred to the hypothesis of zero frequency integration: this series seem stationary around a deterministic time trend. All the other series are found I(1) in the traditional sense, that is they possess a zero frequency unit root, as in Osborn (1990).
The Bayesian approach described in this paper is particularly well suited to cope with situations where the classical inference techniques present difficulties, as they do in the presence of non-stationary variables. In a multivariate context, the inference problem in cointegrated settings is complicated by the fact that different deterministic components originate different asymptotic distributions for the cointegrating rank test statistics (see Johansen, 1992). In addition, only asymptotic and potentially misleading results are available, as it happens for the distribution of the identified cointegrating vectors (see Bauwens and Lubrano, 1993). We believe that a Bayesian approach could help overcoming these difficulties in a sensible way, and our research agenda is orientated in that direction.

Appendix A: Proofs of distributional results

Proof of lemma 4.1

The exponential term in (6) can be written as:
$(-1/(2\sigma^2))[(\mathbf{z}-\mathbf{C}\eta_1)'(\mathbf{z}-\mathbf{C}\eta_1) + (\beta^*-\mathbf{b}^*)'\ \mathbf{V}^{*-1}(\beta^*-\mathbf{b}^*),$

where $\mathbf{z}$ is a (Tx1) vector with t^{th} element:

$\phi^*(L)z_{4t} - \psi_1 z_{1t-1} - \psi_2 z_{2t-1} - 2r(\sin\theta\, z_{3t-1} - \cos\theta\, z_{3t-2})$,

$\mathbf{C}$ is a (Tx5) matrix with t^{th} row:

$[\,-4\psi_1,\ -4\psi_2\cos(\pi t),\ 4r\cos((\pi/2)t-\theta),\ 4r\sin((\pi/2)t-\theta),\ 4\phi^*(1)+$
$+10\psi_1+2\psi_2+4r(\sin\theta-\cos\theta)-4\psi_1 t]$.

Defining:

$\mathbf{z}^* = [\mathbf{V}^{*-1/2}\mathbf{b}^{*\prime}, \mathbf{z}']'$,

$\mathbf{C}^* = [\mathbf{V}^{*-1/2}, \mathbf{C}']'$,

the exponential term in the joint posterior can be written as:

$(-1/(2\sigma^2))[(\mathbf{z}^*-\mathbf{C}^*\eta_1)'(\mathbf{z}^*-\mathbf{C}^*\eta_1)]$.

Therefore we have:

$$p(\eta_1 | \overline{\eta_1}\ D_T) \sim N[(\mathbf{C}^{*\prime}\mathbf{C}^*)^{-1}\mathbf{C}^{*\prime}\mathbf{z}^*,\quad \sigma^2(\mathbf{C}^{*\prime}\mathbf{C}^*)^{-1}],$$

i.e. a 5-variate normal distribution, whose position vector and dispersion matrix depend on the other parameters of the system.

Proof of lemma 4.2

Consider the joint posterior pdf (6). It is evident that this depends on η_2 only through the term $\mathbf{e}'\mathbf{e}$ in the exponential part, which can be written as:

$\mathbf{e}'\mathbf{e}=(\mathbf{y}-\mathbf{X}\eta_2)'(\mathbf{y}-\mathbf{X}\eta_2)$,

where $\mathbf{y}$ is (Tx1) with t^{th} element given by:

$y_{4t} - \psi_1 y_{1t-1} - \psi_2 y_{2t-1} - 2r(\sin\theta\, y_{3t-1} - \cos\theta\, y_{3t-2})$,

and $\mathbf{X}$ is a (Txk^*) matrix with k^* lags of y_{4t} in its t^{th} row.

Thus, conditionally on the other parameters of the system, η_4 has the following posterior pdf:

$$p(\eta_2 | \overline{\eta_2}\ D_T) \sim N[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y},\quad \sigma^2(\mathbf{X}'\mathbf{X})^{-1}].$$

Proof of lemma 4.3

From expression (6) we have that :

$$p(\eta_3 \mid \overline{\eta_3}\, D_T) \propto \sigma^{-T-5} \exp\{ -c/(2\sigma^2)\},\ c = \mathbf{e'e} + (\boldsymbol{\beta}^* - \mathbf{b}^*)'\, \mathbf{V}^{*-1}(\boldsymbol{\beta}^* - \mathbf{b}^*).$$

where c depends on the data and on the other parameters of the model.
The expression above has the appearance of the inverted-Gamma distribution (see Zellner, 1971, p.371):

$$p(y|v\ \alpha) = 2[\Gamma(\alpha)\gamma^{\alpha}\, y^{2\alpha-1}]^{-1} \exp\{-1/(2\gamma\, y^2)\},$$

the connections between inverse Gamma, Gamma and χ^2 distributions can be exploited. We define $\sigma' = c/\sigma^2$, and since σ' is a monotone function of σ, we conclude that:

$$p(\sigma' | \overline{\eta_3}\, D_T) \propto \sigma'^{\,(T+2)/2} \exp\{-\sigma'/2\},$$

i.e. that the conditional posterior pdf of σ' given all the other parameters is $\chi^2(T+4)$. This is intuitive, since:

$$c/\sigma^2 = (\varepsilon'\varepsilon + \sigma_\theta^2(\theta-\mu_\theta)^2)/\sigma^2 = \sigma^{-2}\, [\sum_{t=1}^{T} e_t^2 + (\boldsymbol{\beta}-\mathbf{b})'\, \mathbf{V}^{-1}(\boldsymbol{\beta}-\mathbf{b}) + \sigma_\theta^2(\theta-\mu_\theta)^2],$$

is just the sum of the squares of T+4 independent standard normal variates.

Proof of lemma 4.4

Starting from expression (6), we realise that the parameter ψ_1 appears both in the exponential term, via **e'e**, $(\alpha_0 - a_0)^2\psi_1$, and $\lambda_1\psi_1$, and outside, via $(-\psi_1)^{-1/2}$.
The term **e'e** in the exponential part can be represented as:

$$\mathbf{e'e} = (\mathbf{y} - \mathbf{x}\eta_4)'\, (\mathbf{y} - \mathbf{x}\eta_4),$$

where the vector (Tx1) **y** has t^{th} element equal to $\phi^*(L)y_{4t} - y_{2t-1} - 2\, r\, (\sin\theta\ y_{3t-1} - \cos\theta\ y_{3t-2})$, and **x** is another (Tx1) vector with corresponding element equal to y_{1t-1}.

On the basis of this representation, and using a notation consistent with expression above, we can represent the whole relevant exponential term as quadratic in η_4:

$$p(\eta_4 | \overline{\eta_4}\, D_T) \propto (-\eta_4)^{1/2} \exp\{-1/(2\tau_1^2)(\eta_4 - \mu_1)^2\},$$

where:

$\mu_1 = [\mathbf{x'y} + (1/2)(\alpha_0\text{-}a_0)^2 + \lambda_1 \sigma^2]/ (\mathbf{x'x}), \qquad \tau_1 = \sigma^2/(\mathbf{x'x}).$

Although Gaussian-looking in the exponential part, the above distribution is not unfortunately of any analytically known form.

As for η_5 , we have:

$p(\eta_5 | \overline{\eta_5}\ \mathbf{D}_T) \propto (-\eta_5)^{1/2} \exp\{(-1/(2\tau_2^2))(\eta_5-\mu_2)^2\},$

with:

$\mu_2 = [\mathbf{x'y} + (1/2)(\alpha_2\text{-}a_2)^2 + \lambda_2 \sigma^2]/ (\mathbf{x'x}), \qquad \tau_2 = \sigma^2/(\mathbf{x'x}),$

the vectors **y** and **x** have been conveniently defined to decompose:

$\mathbf{e'e} = (\mathbf{y}\text{-}\mathbf{x}\eta_5)' (\mathbf{y}\text{-}\mathbf{x}\eta_5).$

For the parameter η_5 (ψ_2) we have results that coincide with those seen for η_4 (ψ_1). The parameters τ_2 and μ_2 derive from a similar sort of decomposition of the exponential part as seen above.

The conditional posterior pdf of η_6 (that is r) is likewise complicated:

$p(\eta_6 | \overline{\eta_6}\ \mathbf{D}_T) \propto \ r \exp\{(-1/(2\tau_r^2))(r-\mu_r)^2\}$, where $\eta_6 = r$,

$\mu_r = \{\mathbf{x'y} - [(\alpha_1\text{-}a_1)^2 + (\beta_1\text{-}b_1)^2]/2 + \lambda_r \sigma^2\}/ (\mathbf{x'x}), \ \tau_2 = \sigma^2/(\mathbf{x'x});$

$\mathbf{e'e} = (\mathbf{y}\text{-}\mathbf{x}\eta_6)' (\mathbf{y}\text{-}\mathbf{x}\eta_6).$

The conditional posterior of η_7 (θ, the phase angle) is even more complicated:

$p(\eta_7 | \overline{\eta_7}\ \mathbf{D}_T) \propto \exp\{(-1/(2\sigma^2))[\mathbf{e'e} + (\sigma/\sigma_\theta)^2 (\theta - \mu_\theta)^2]\},$

$$\mathbf{e'e} = \sum_{t=1}^{T} [\phi^*(L)y_{4t} - \psi_1 y_{1t-1} - \psi_2 y_{2t-1} - \sin\theta(2r\, y_{3t-1}) + \cos\theta\, (2r\, y_{3t-2})\,]^2.$$

We obtain draws from these conditional posterior distribution using the algorithms described in Appendix B.

Appendix B : Rejection sampling from the conditional posterior distributions

We followthe approach of Geweke (1993). First, a brief description of the method used is given, then the solutions adopted to the particular problem being treated are developed with particular attention their capability of providing efficient random drawings from the conditional posterior pdf's.

Suppose that $f(x|\theta)$ is the kernel of the pdf of the random variable x, which depends on a certain vector of parameters. We furthermore suppose that no analytic results that allow direct random variate generation from f() are available. We then choose a comparison function, with kernel $g(x|\phi)$. The function g() must be such that it is possible to obtain random draws from it directly.

A single result draw, say X_i is retained or rejected on the basis of the outcome of another independent random drawing from the uniform distribution defined over the support :

$$\mathbf{S} = \left[0, \max_x\left(\frac{f(x|\theta)}{g(x|\phi)}\right)\right].$$

If the result from such drawing, say Y_i , is less than $f(X_i|\theta)/g(X_i|\phi)$, then the draw X_i is accepted, and rejected otherwise. This implies that for any subset A of the support of x, the probability of getting retained draws is given by:

$$\int_{x\in A} g(x|\theta)\left[\frac{\frac{f(x|\theta)}{g(x|\phi)}}{\max_x\left(\frac{f(x|\theta)}{g(x|\phi)}\right)}\right]dx \propto \int_{x\in A} f(x|\theta)dx,$$

i.e. the algorithm generates synthetic draws from the target distribution $f(x|\theta)$, via the comparison function.

Therefore we face the the problem of optimally choosing the comparison function. We suppose that the aim is to maximise computational efficiency, i.e. to maximise the unconditional probability of retaining draws from the comparison function.

We have thus to solve:

$$\min_\phi\left[\max_x\left(\frac{f(x|\theta)}{g(x|\phi)}\right)\right].$$

We emphasise that the choice of the parameters in ϕ, together with the choice of the functional form of g(.|.) does not affect the correctness of the results from the synthetic draws, but only the efficiency of the procedure, i.e. the rejection rate of the draws from g(.|.), and hence the computational time.

In the remainder of the present appendix, the choices made in this respect are discussed for each of the 4 synthetically replicated conditional posterior pdf 's, namely those of ψ_1, ψ_2, r and θ.

1) Conditional posterior density of ψ_1.

The distribution is:

$$f(x) \propto (-x)^{1/2} \exp\{(-1/(2\sigma^2))(x-\mu)^2\}, \qquad x \in \mathbf{R}_-$$

with the quantities μ and σ defined as in Section 4.
The comparison function chosen is:

$$g(x) \propto \exp\{(-1/(2\sigma^2))(x-\nu)^2\}\, I_{(-0,0)}(x),$$

a negative truncated normal distribution. In order to avoid further complications, we confine our choice to the location parameter ν.
The differentiation of log f() -log g() with respect to x gives:

$$x^* = -\sigma^2/(2(\mu-\nu)).$$

Provided $\mu-\nu > 0$, x^* belongs to the support of f(x). The second order condition for a maximum holds.
The expression log g() - log f() evaluated in x^* is maximised with respect to ν, yielding:

$$\nu = (\mu - (\mu^2+2\sigma^2)^{1/2})/2.$$

This is the only admissible solution. The second order condition is satisfied.

2) Conditional posterior pdf of ψ_2.

Exactly the same computations as above apply.

3) Conditional posterior pdf of r.

The distribution is:

$$f(x) \propto x \exp\{(-1/(2\sigma^2))(x-\mu)^2\}, \qquad x \in \mathbf{R}_+$$

with the quantities μ and σ defined as in section [4].

The comparison function chosen is:

$$g(x) \propto \exp\{(-1/(2\sigma^2))(x-\nu)^2\}\, I_{(0,\,0)}(x),$$

a positive truncated normal distribution. The same kind of computations as previously described yield:

$$x^* = \sigma^2/(\nu-\mu),$$

$$\nu = (\mu + (\mu^2+4\sigma^2)^{1/2})/2.$$

4) Posterior distribution of θ.
Recalling the analysis contained in Section 4, we can write:

$$f(\theta) \propto \exp\{(-1/(2\sigma^2))[\mathbf{e'e} + (\sigma/\sigma_\theta)^2 (\theta - \mu_\theta)^2]\}, \quad \theta \in [-\pi, \pi],$$

where:

$$\mathbf{e'e} = \sum_{t=1}^{T} [\phi^*(L)y_{4t} - \psi_1 y_{1t-1} - \psi_2 y_{2t-1} - \sin\theta(2r\, y_{3t-1}) + \cos\theta\,(2r\, y_{3t-2})]^2.$$

$$= (\mathbf{y}-\mathbf{x}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{x}\boldsymbol{\beta}),\ \boldsymbol{\beta} = [\sin\theta\ ,\ \cos\theta]'\ ,\ \mathbf{x}=\{x_t\}_{t=1}^{T},\ x_t = [2r\, y_{3t-1}|-2r\, y_{3t-2}]$$

The problem of obtaining draws looks quite cumbersome, given that the maximisation of log f() -log g() involves trigonometric expressions. An easy way out is to choose:

$$g(\theta) \propto \exp\{(-1/(2\sigma_\theta^2))(\theta - \mu_\theta)^2\} I_{[-\pi,\pi]}(\theta),$$

so that the function to be maximised is:

$$\log f() - \log g() = -1/(2\sigma^2)\ (\mathbf{y}-\mathbf{x}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{x}\boldsymbol{\beta}).$$

We maximise this expression with respect to β, under the non- linear constraint: $\beta'\beta = 1$. This can be done numerically.Once the maximum value is obtained, say M, one then draws y from U(0,M), and applies the rejection sampling technique.

References

AMISANO, G. (1994): A Bayesian Perspective on Unit Root Testing,, Ph.D. thesis in preparation, Department of Economics, University of Warwick, U.K.

BARGHAVA, A. (1986): On the Theory of Testing for Unit Roots in Observed Time Series, *Review of Economic Studies*, 52, 369-384.

BAUWENS, L., and M. LUBRANO (1993): Identification Restrictions and Posterior Densities in Cointegrated Gaussian VAR Systems, CORE discussion paper, Universite' Catholique de Louvain.

BEAULIEU, J.J., and J.A. MIRON (1993): Seasonal Unit Roots in Aggregate U.S. Data, *Journal of Econometrics*, 55, 305-348.

CAMPBELL, J.Y., and P. PERRON (1991), Pitfalls and Opprtunities: What Economists Should Know about Unit Roots, *NBER Macroeconomic Annual.*

CHIB, S. (1993): Bayes Regression with Autoregressive Errors, *Journal of Econometrics*, 58, 275-294.

DEJONG D.N. and C.H. WHITEMAN (1989): Trends and Cycles as Unobserved Components in US Real GNP: a Bayesian Perspective, *Proceeding of the Business and Economic Statistics Section of the American Statistical Association.*

DEJONG, D.N., J.C. NANKERVIS, N.E. SAVIN and C.H. WHITEMAN (1992): Integration vs. Stationarity in Time Series, *Econometrica*, 60, 423-433.

DEJONG D.N. and C.H. WHITEMAN (1991): Reconsidering Trends and Random Walks in Macroeconomic Time Series, *Journal of Monetary Economics*, forthcoming.

DEVROYE, L. (1986): *Non-Uniform Random Variate Generation*, New York, Springer-Verlag.

DICKEY, D.A. (1976): Estimation and Testing of Non-Stationary Time Series, Ph.D. Thesis, Iowa State University.

DICKEY, D.A.and W.A. FULLER (1979): Distributions of the Estimators for Autoregressive Time Series with a Unit Root, *Journal of the American Statistical Association*, 74, 427-431.

DICKEY, D.A.and W.A. FULLER (1981): Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root, *Econometrica*, 49, 1057-1072.

ENGLE, R.F., C.W.J. GRANGER, S. HYLLEBERG and H.S. LEE (1993): Seasonal Cointegration: The Japanese Consumption Function, *Journal of Econometrics*, 55, 257-304.

FULLER, W.A. (1976) : *Introduction to Statistical Time Series*, New York, Wiley.

GEWEKE, J., (1989): Bayesian Inference in Econometric Models Using Monte Carlo Integration, *Econometrica*, 57, 6, 1317-1339.

GEWEKE, J., (1993): Priors for Macroeconomic Time Series and Their Application, discussion paper # 64, Institute for Empirical Macroeconomics, and Federal Reserve Bank of Minneapolis.

HAMILTON, J.D. (1994): *Time Series Analysis*, Princeton University Press, Princeton, New Jersey.

HAMMERSLEY, J.M., and D.C. HANDSCOMB (1964): *Monte Carlo Methods*, 1st Edition, London, Methuen.

HYLLEBERG, S., R.F. ENGLE , C.W.J. GRANGER and BS. YOO (1990) Seasonal Integration and Cointegration, *Journal of Econometrics*, 44, 215-238.

JOHANSEN, S. (1992) Determination of the Cointegrating rank in the Presence of a Linear Trend, *Oxford Bulletin of Economics and Statistics*, 54, 3, 383-397.

KLOEK, T. and H.K. VAN DICK (1978): Bayesian Estimates of Equation System Parameters: an Application of Integration by Monte Carlo, *Econometrica*, 46, 881-896.

KOOP, G. (1994): Recent Progress in Applied Bayesian Econometrics, *Journal of Economic Surveys*, 8, 1-34.

LINDSAY, D.V. (1957): A Statistical Paradox, *Biometrika*, 44, 187-192.

LINDSAY, D.V. (1961): The Use of Prior Probability Distributions in Statistical Inference and Decision, *Proceeding of the 4th Berkeley Symposium in Mathematical Statistics and Probability*, 1, 453-468.

NELSON, C.R and C.I. PLOSSER (1982): Trend and Random Walks in Macroeconomic Time Series. Some Evidence and Implications, *Journal of Monetary Economics*, 10, 139-162.

OSBORN, D.R. (1990): A Survey of Seasonality in U.K. Macroeconomic Variables, *International Journal of Forecasting*, 6, 327-336.

PHILLIPS, P.C.B. (1983): Marginal Densities of Instrumental Variable Estimators in the General Single Equation Case, *Advances in Econometrics*, 2, 1-24.

PHILLIPS, P.C.B. and P. PERRON (1988): Testing for a Unit Root in Time Series Regression, *Biometrika*, 75,2, 335-346.

RUBINSTEIN, R.Y. (1985): *Simulation and the Monte Carlo Method*, New York, Wiley.

SCHMIDT P. and P.C.B. PHILLIPS (1992): Testing for a Unit Root in the Presence of Deterministic Trends, *Oxford Bulletin of Economics and Statistics*, 54, 257-287.

SCHOTMAN, P. and H.K. VAN DICK (1991 a): A Bayesian Analysis of the Unit Root in Real Exchange Rates, *Journal of Econometrics*, 49, 195-238.

SCHOTMAN, P. and H.K. VAN DICK (1991 b): On Bayesian Routes to Unit Roots, *Journal of Applied Econometrics*, 6, 387-401.

SCHOTMAN, P. and H.K. VAN DICK (1992): Posterior Analysis of Possibly Integrated Series with an Application to Real GNP, in P. Caines, J. Geweke and M. Taqqu (eds.) *New Directions in Time Series Analysis*, Part II. Berlin: Springer-Verlag.

SIMS, C.A. (1974): Seasonality in Regression, *Journal of the American Statistical Association*, 69, 618-626.

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

SIMS, C.A. (1989): Modeling Trends, discussion paper, Institute for Empirical Macroeconomics, and Federal Reserve Bank, Minneapolis

SIMS, C.A. and H. UHLIG (1991): Understanding Unit Rooters: a Helicopter Tour, *Econometrica*, 59, 1591-1599.

SMITH, A.F.M., and G.O. ROBERTS (1993): Bayesian Computation Via the Gibbs Sampler and Related Markov-Chain Monte Carlo Methods, *Journal of The Royal Statistical Society*, B, 55,1, with discussions.

TIERNEY, L. (1991): Markov Chains for Exploring Posterior Distributions, University of Minnesota School of Statistics Technical Report No. 560.

TIERNEY, L. and J.B. KADANE (1986): Accurate Approximations for Posterior Moments and Marginal Densities, *Journal of the American Statistic Association*, 79, 916-921.

VAN DICK (1986): Some Advances in Bayesian Estimation Methods Using Monte Carlo Integration, *Advances in Econometrics*, 6, 215-261.

WALLIS, K.F. (1974): Seasonal Adjustment and Relations Between Variables, *Journal of the American Statistical Association*, 69, 18-32.

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

Figure (1.a): UK GDP, 1955:1–1988:4

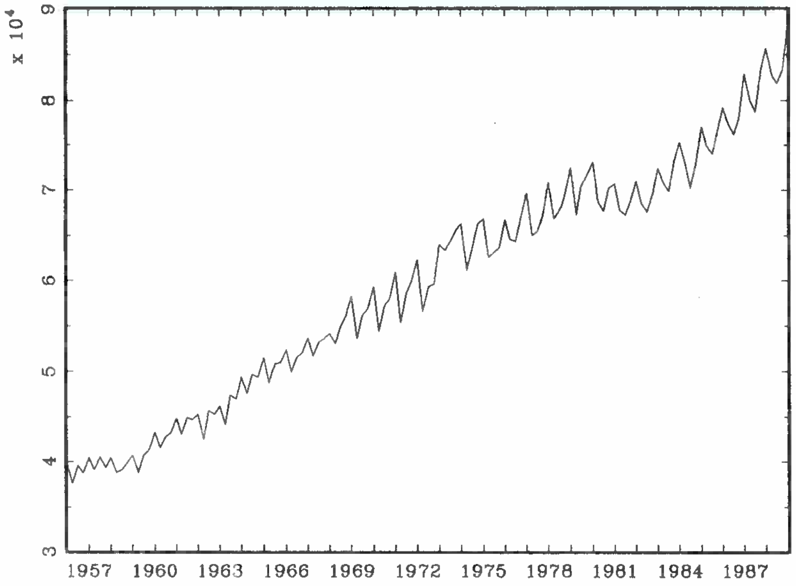

Figure (1.b): UK GDP posterior distribution of ψ_1

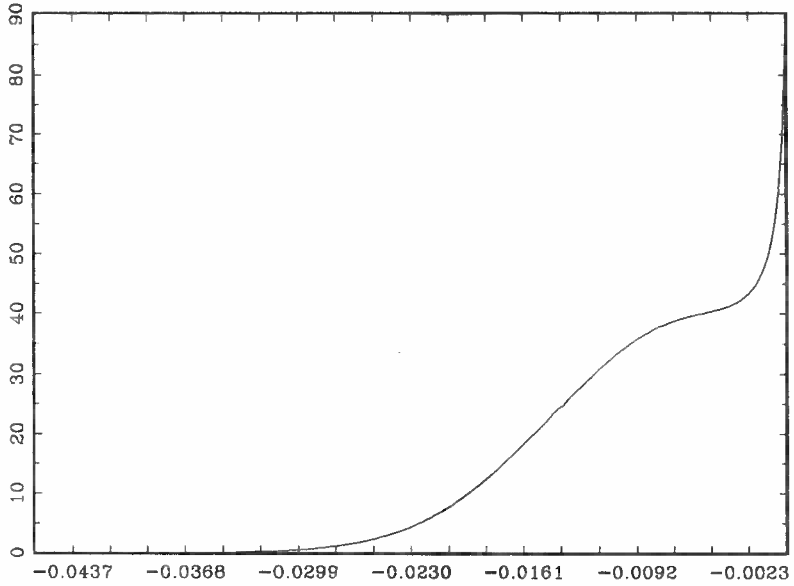

Figure (1.c): UK GDP, posterior distribution of ψ_2

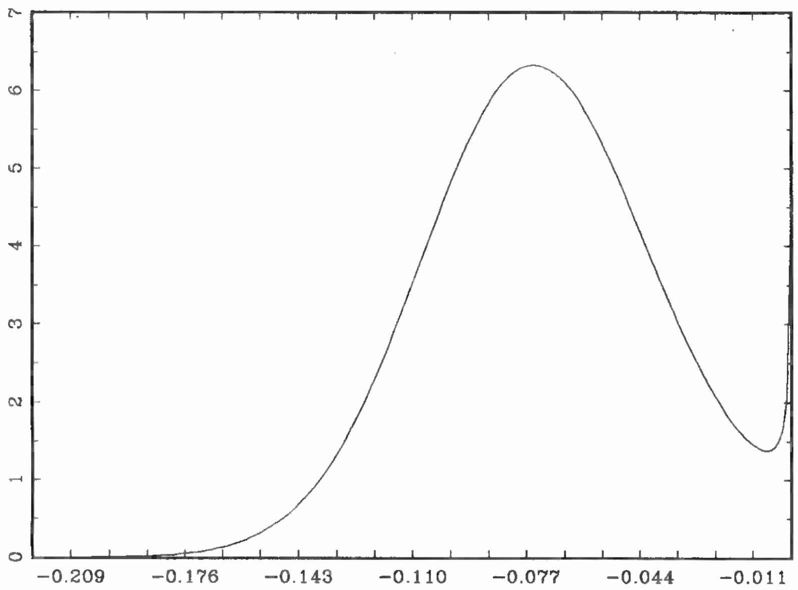

Figure (1.d): UK GDP, posterior distribution of r

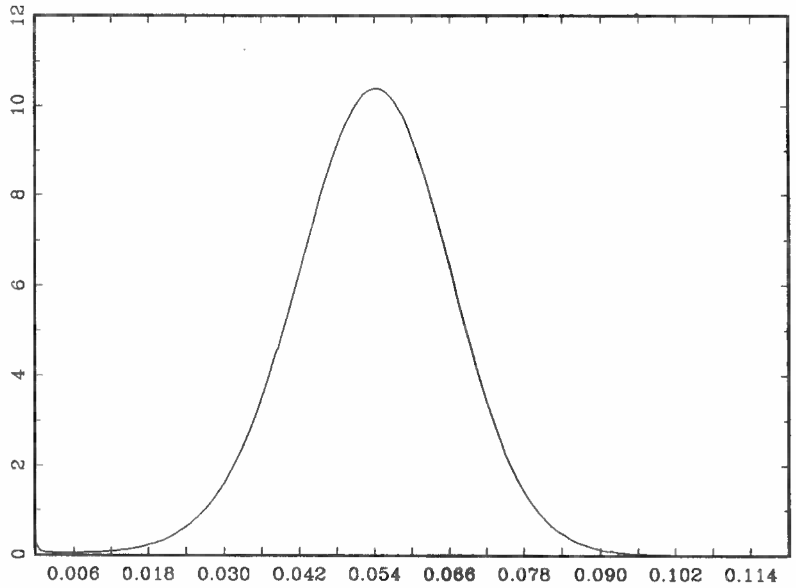

Figure (2.a): UK Consumption, 1955:1−1988:4

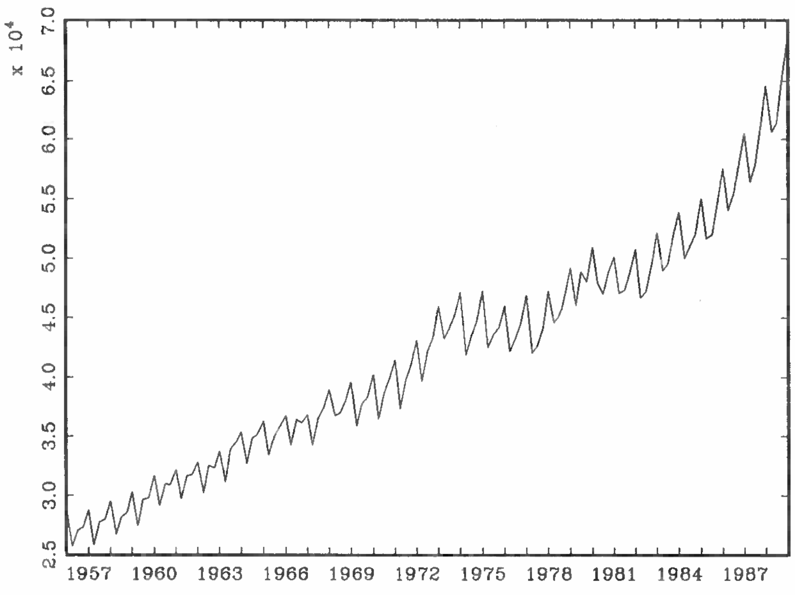

Figure (2.b): UK Consumption **posterior** distribution of ψ_1

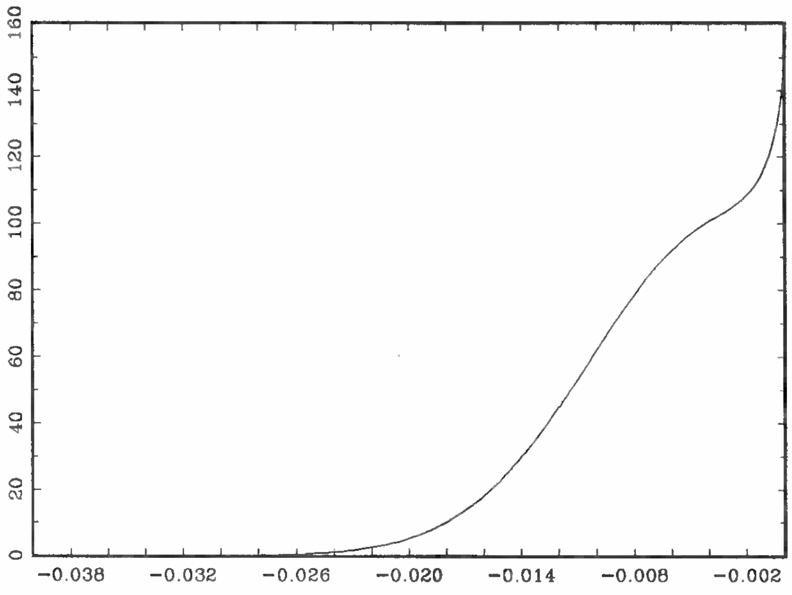

Figure (2.c): UK Consumption, posterior distribution of ψ_2

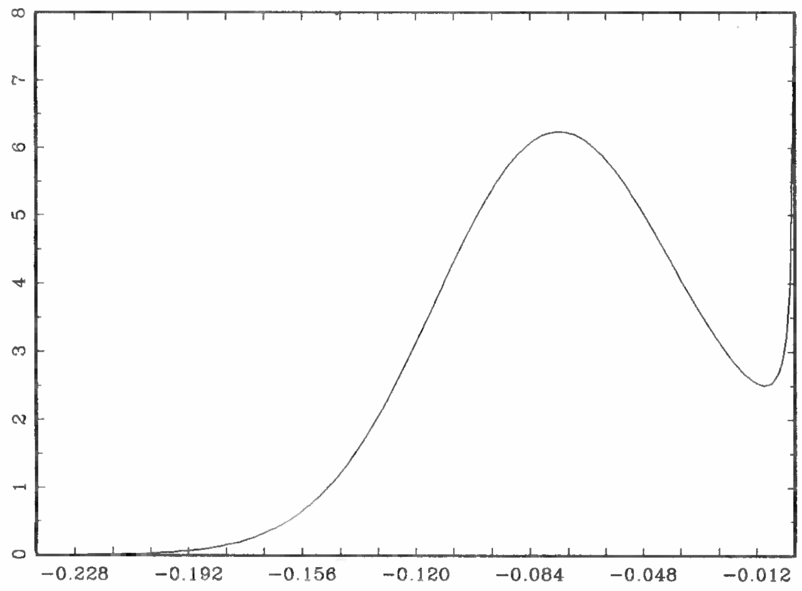

Figure (2.d): UK Consumption, **posterior** distribution of r

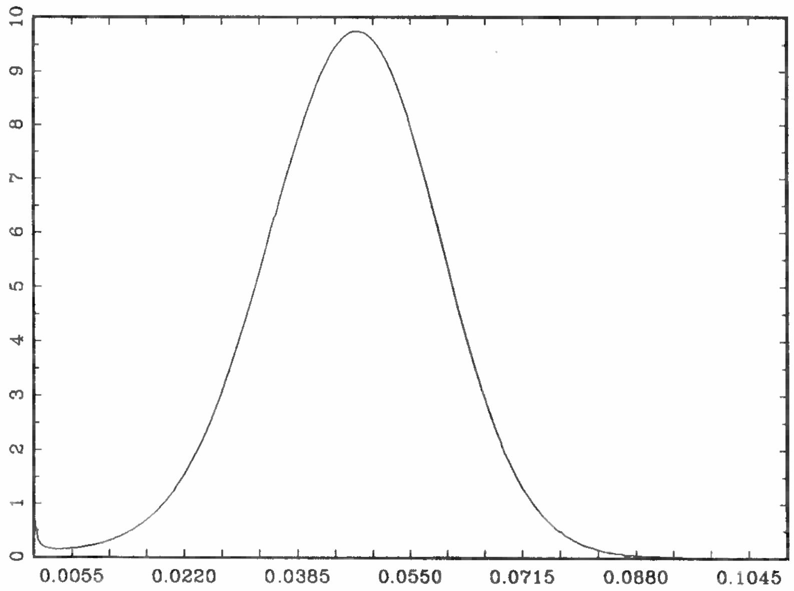

Figure (3.a): UK Investment, 1955:1−1988:4

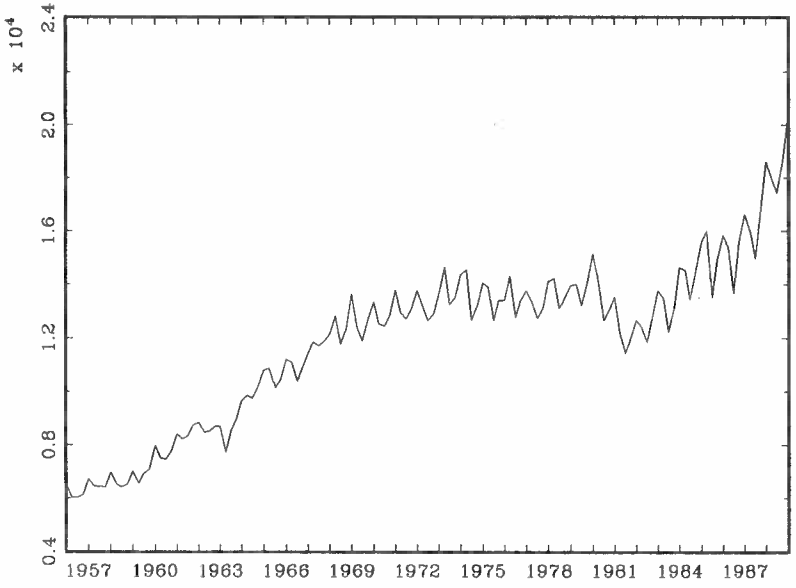

Figure (3.b): UK Investment posterior distribution of ψ_1

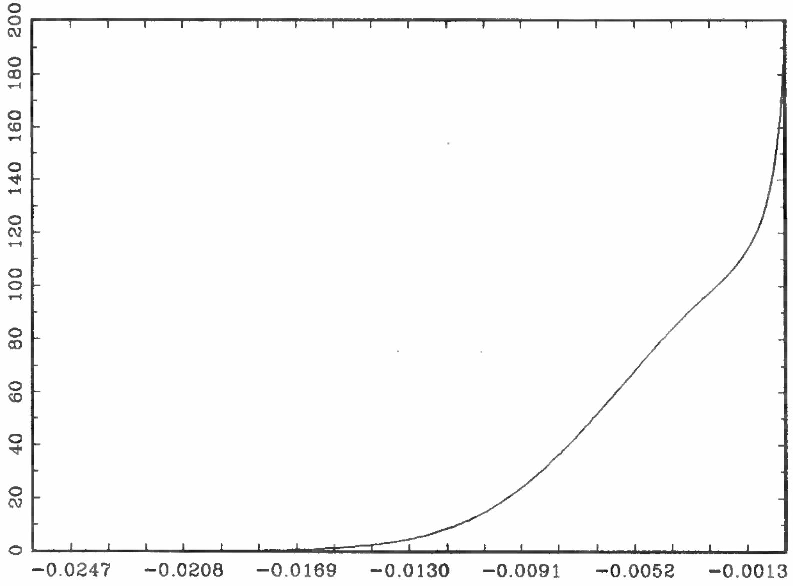

Figure (3.c): UK Investment, posterior distribution of ψ_2

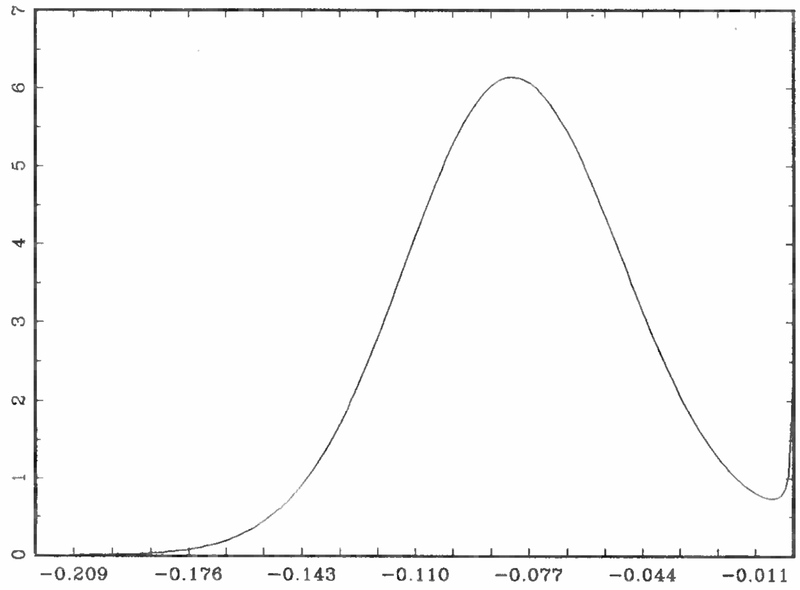

Figure (3.d): UK Investment, posterior distribution of r

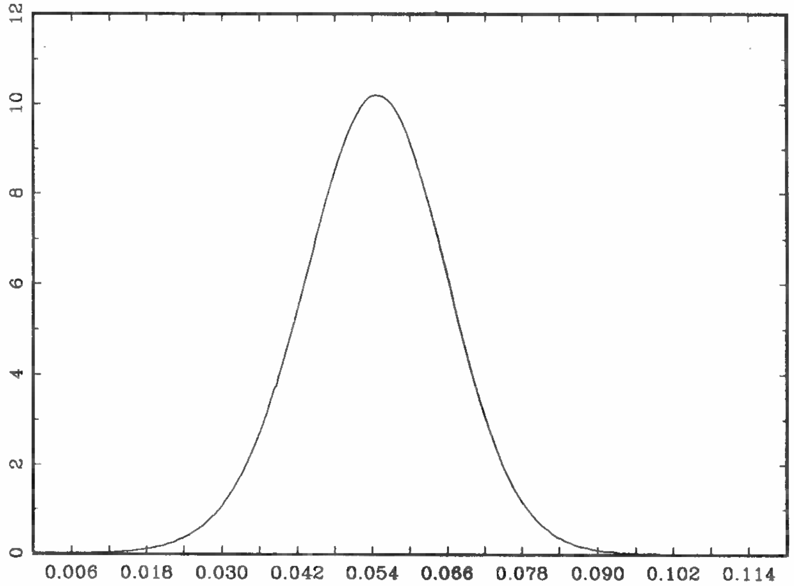

Figure (4.a): UK Employment, 1955:1−1988:4

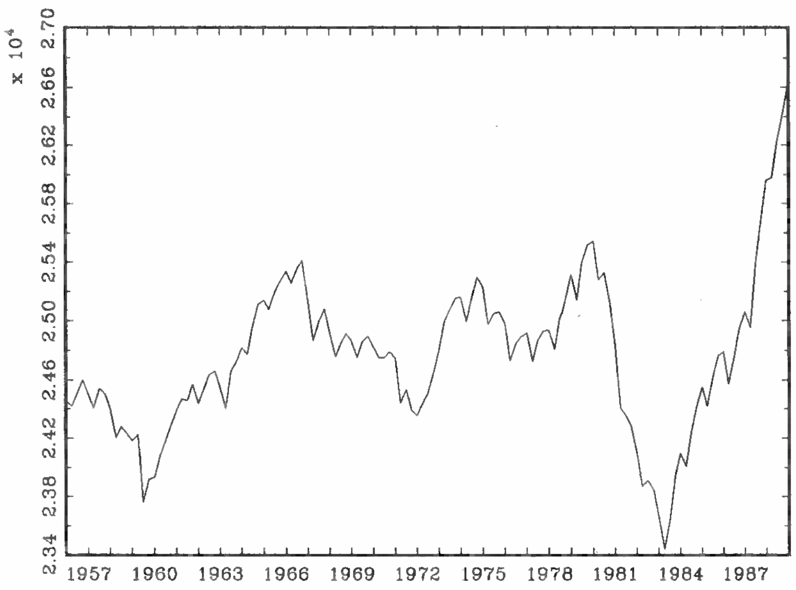

Figure (4.b): UK Employment **posterior** distribution of ψ_1

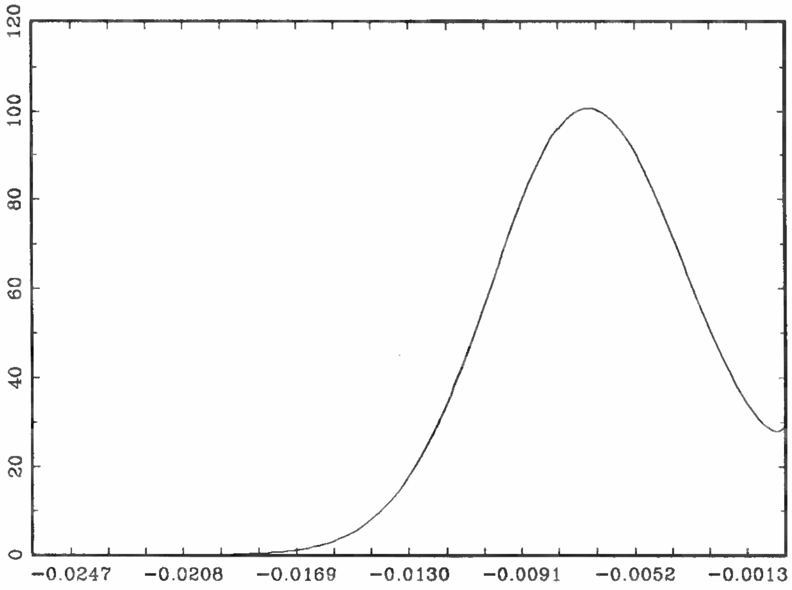

Figure (4.c): UK Employment, posterior distribution of ψ_2

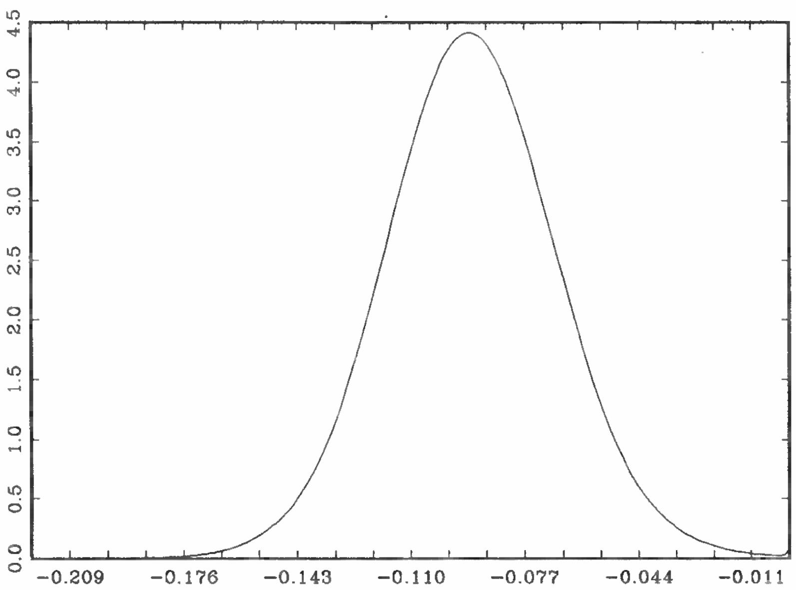

Figure (4.d): UK Employment, **posterior** distribution of r

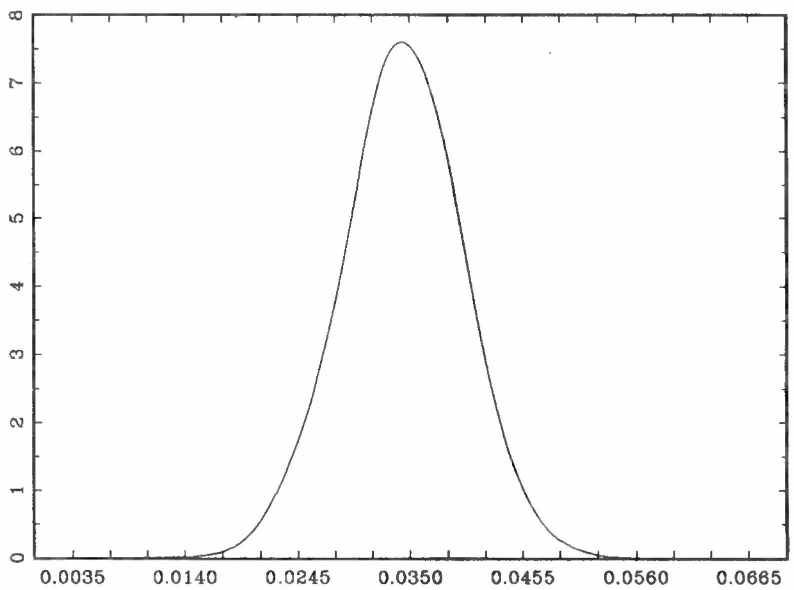

Figure (5.a): UK M0, 1955:1–1988:4

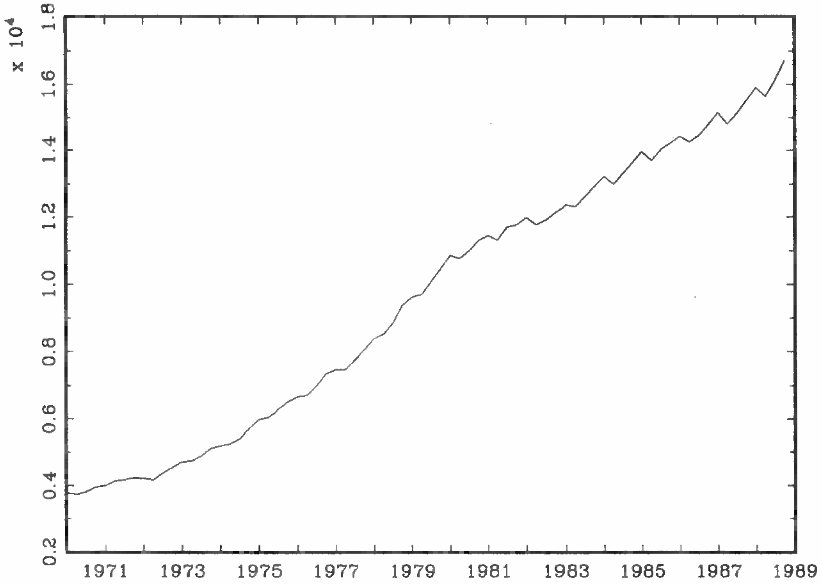

Figure (5.b): UK M0 posterior distribution of ψ_1

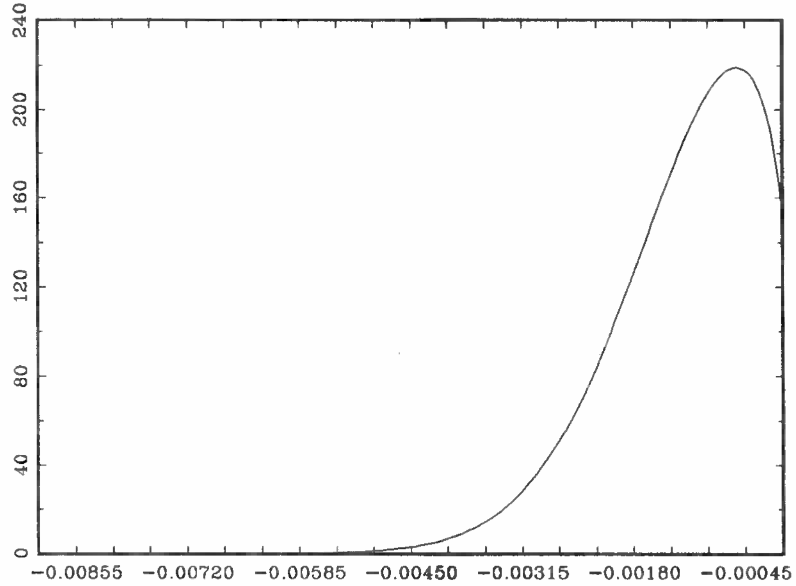

Figure (5.c): UK M0, posterior distribution of ψ_2

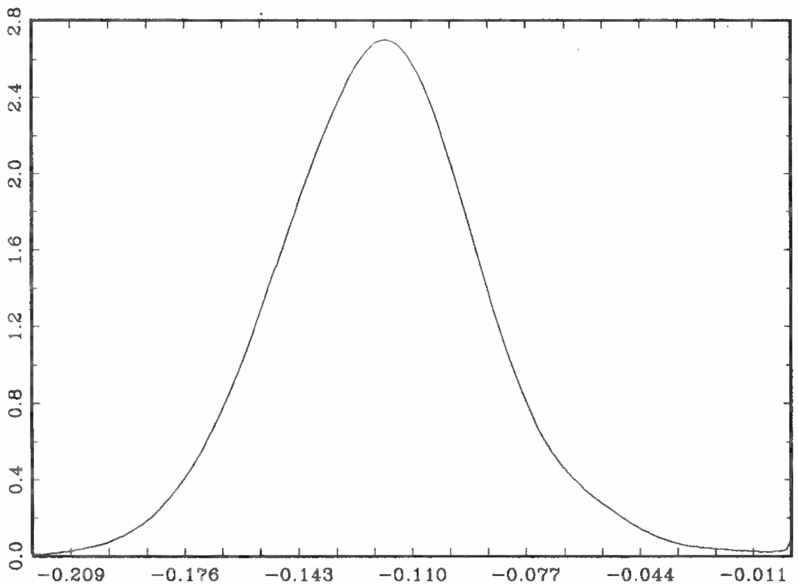

Figure (5.d): UK M0, posterior distribution of r