ISSN 1032-3813
ISBN 0 7326 0778 7

MONASH UNIVERSITY

AUSTRALIA

FRACTIONAL COINTEGRATION: A BAYESIAN APROACH

Gael Martin

Working Paper 17/95
October 1995

DEPARTMENT OF ECONOMETRICS

Fractional Cointegration: a Bayesian Approach

Gael Martin*
Department of Econometrics,
Monash University,
Clayton, Victoria 3168,
Australia.
Gael.Martin@BusEco.monash.edu.au

September, 1995

Abstract

The concept of fractional cointegration, whereby deviations from an equilibrium relationship are allowed to follow a fractionally integrated process, has attracted some attention in the literature of late. The long memory aspect of the fractional process is seen as an appropriate characterization of slow reversion to an equilibrium relationship. This paper presents a Bayesian method for conducting inference within the context of a fractional cointegration model. The analysis is based on an approximate likelihood function, which is motivated by the need both to solve a fundamental identification problem and to produce a posterior density with a relatively simple algebraic form. Inferences are based on the associated marginal posterior densities, estimated by a hybrid of the Gibbs and Metropolis Markov Chain Monte Carlo methods.

*I would like to thank Brett Inder and Vance Martin for helpful comments during the preparation of this paper.

1. Introduction

The concept of cointegration applied in the literature is almost exclusively that of sets of I(1) variables producing I(0) linear combinations, with emphasis given as a consequence to the stationarity of the cointegrating errors. There has been some work, for example Stock and Watson (1993), Kitamura (1995) and Johansen (1995), in which variables of higher integration orders are incorporated in cointegration models. However, all variables are assumed to have integer orders of integration and the error processes an integration order of zero.

Recently, certain authors have extended the definition of cointegration to allow for non-integer orders of integration. Cheung and Lai (1993) reassess the doctrine of purchasing power parity (PPP) by applying the Geweke and Porter-Hudak (GPH) test for fractional integration[1] to the residuals of an estimated PPP relationship.[2] Silvapulle (1995) develops a score test for seasonal fractional cointegration, again applied to the residuals of an estimated model. Dueker and Startz (1993) apply a Generalized Method of Moments procedure to a model which allows both the variables and the error term to be fractionally integrated. Their method produces simultaneous estimates of all (fractional) orders of integration, in additional to estimates of the cointegrating parameters.

One important outcome of the extension of cointegration to incorporate fractionally integrated error processes is the highlighting of the mean reversion aspect of cointegration. Specifically, the idea that an equilibrium relationship holds in the long run can implicitly be

1 See Geweke and Porter-Hudak (1983).
2 Related work by Diebold, Husted and Rush (1991) examines the PPP issue via a univariate fractional integration approach, whereby the regression coefficient in the PPP relationship is assigned a value of one, rather than estimated.

equated with the tendency of the error term to ultimately revert to its mean after a shock to the relationship. The stationarity of the error process is of secondary importance. Rather than imposing *both* mean reversion and stationarity by equating cointegration with an I(0) error process, the allowance of a fractional value for the relevant differencing parameter enables a conclusion in favour of cointegration to be reached on the basis of a finding of mean reversion alone. The hope is that the existence of a long-run economic relation is more easily discerned via this more appropriate characterization of the relation.

The present paper presents a Bayesian approach to inference in a model which allows for fractional cointegration. It outlines a hybrid Markov Chain Monte Carlo (MCMC) numerical strategy for estimating marginal posterior densities for both the parameter controlling the presence of fractional cointegration and the parameter(s) of the cointegrating relation. Inferences concerning these aspects of the model are then based on the estimated densities.

We show that the exact likelihood function for the assumed cointegration model is problematic in two ways. First, it renders the regression parameter(s) unidentified when there is a lack of cointegration. Second, it is a highly complicated function of the parameters. In contrast, a simple approximation to the exact likelihood produces a form for which the associated identification problem is solvable via a judicious specification of a so-called Jeffreys' prior density. In addition, the parameters enter the resultant joint posterior in such a way that the relevant marginals are easily estimated via a combination of Gibbs and Metropolis MCMC algorithms.

We also demonstrate that when the fractionally cointegrated error term is allowed a short-run autoregressive (AR) component, inferences become ambiguous. In particular, the marginal density for the fractional differencing parameter can be bimodal, with the second mode reflecting the parameter which summarizes the long-term memory of the AR component. Whether this ambiguity is truly specific to our parameterization and/or inferential method, or is a manifestation of a more fundamental problem associated with the estimation

of short-run dynamics in conjunction with a long-memory fractional component, is unclear at this stage.

The paper is organized as follows. Section 2 provides an outline of the concepts of fractional integration and cointegration, including the correspondence between a certain range of values for the fractional differencing parameter and mean reversion. Section 3 begins by deriving the exact likelihood for the fractional cointegration model, and demonstrating its attendant difficulties. The approximate likelihood, along with the solution of its associated identification problem via a Jeffreys' prior, is then presented. Some instances of the posterior bimodality mentioned above are presented. The proposed hybrid Gibbs/Metropolis sampling method is described very briefly in Section 4. In Section 5 we provide the results of a Monte Carlo study, in which the Bayesian inferences are compared with a variety of Classical alternatives, including Maximum Likelihood applied to the residuals of the estimated cointegrating equation. Although preliminary, these results tend to suggest that, at least for a parameterization which avoids the short-run/long-run problem mentioned above, the Bayesian method provides a very viable alternative to the Classical procedures. Section 6 describes the application of the Bayesian method to the problem of both testing for and estimating PPP relations for various countries. Some comparison is made with Classical inferences. Once again with the qualification made as to the allowable parameterization, the results produce fairly convincing evidence that PPP exists, but with reversion to parity after a shock being very slow. The paper gives some conclusions in Section 7.

2. Fractional Cointegration and Mean Reversion

Consider the following bivariate model for the generation, at time t, of observations on the variables y and x respectively:

$$y_t = \beta x_t + u_{1t}, \tag{1}$$

$$x_t = x_{t-1} + u_{2t}. \tag{2}$$

We assume that u_{1t} is an autoregressive fractionally integrated moving average (ARFIMA) $(p, d, 0)$ process represented by:

$$\Phi(L)(1-L)^d u_{1t} = e_{1t}, \tag{3}$$

where $\Phi(L)$ is a p-dimensional polynomial in the lag operator L, denoted by $\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$, all roots of $\Phi(L)$ lie outside the unit circle and e_{1t} is as defined below. Crucially, d is allowed to assume any real number value greater than -1. The error term in (2) is assumed to be an AR(q) process of the form

$$\Theta(L)u_{2t} = e_{2t}, \tag{4}$$

where the roots of the polynomial $\Theta(L) = 1-\theta_1 L-\theta_2 L^2 -\cdots-\theta_q L^q$ also lie in the stationary region. It is assumed that $\mathbf{e}_t = (e_{1t}, e_{2t})'$ has a bivariate Normal distribution of the form:

$$\begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} \sim NID(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}), \qquad \sigma_{12} = \sigma_{21}. \tag{5}$$

The allowance of a non-zero value for σ_{12} allows for the endogeneity of x_t.

It is the interaction between $\Phi(L)$ and $(1-L)^d$ which leads to the inferential problems regarding d alluded to earlier. We shall demonstrate our inferential procedure with $\Phi(L)$ incorporated, including examples of its impact on the marginal density of d. The reason for omitting moving average (MA) components in the assumed processes for u_{1t} and u_{2t} is that their incorporation leads to an added complication for the MCMC numerical procedure, which we have chosen to avoid in the present paper.[3] We have assumed that the use of an MA rather than an AR representation for the short-run component in (3) is not sufficient to avoid the inferential problems regarding d. However, this deserves further attention.

Given the $I(1)$ nature of the regressor, (1) potentially represents a cointegrating relationship between two $I(1)$ variables. It is of interest both to test for this possibility and to

[2] See below.
[3] See Chib and Greenburg (1994) for an example of the application of an MCMC method to a regression model with ARMA errors.

estimate the value of β in the event that the latter is concluded to be the parameter of a cointegrating relationship. The existence of cointegration depends upon the value assumed by d.

The fractional differencing operator $(1-L)^d$ in (3) is defined through the binomial expansion:

$$(1-L)^d = \sum_{j=0}^{\infty} d_j L^j, \tag{6}$$

where:

$$d_j = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)}, \qquad j = 0, 1, 2, \ldots,$$

and $\Gamma(.)$ denotes the gamma function, which is, in turn, defined by:

$$\Gamma(z) \equiv \begin{cases} \int_0^\infty s^{z-1}e^{-s}ds & \text{if } z > 0, \\ \infty & \text{if } z = 0, \\ \Gamma(z+1)/z & \text{if } z < 0. \end{cases}$$

The expansion given by (6) is valid for any value of d, but has declining coefficients if and only if $d > -1$. The coefficients are square summable if and only if $d > -0.5$.[4]

Three further ranges of values for d need to be distinguished for our purposes:

1. For values of $d < 0.5$, u_{1t} is a covariance stationary process with a valid Wold representation. However, for $d \neq 0$, both the coefficients of the infinite moving average representation and the autocorrelations of u_{1t} decline at a hyperbolic rate, in contrast to the faster exponential rate of decay associated with the stationary AR process corresponding to $d = 0$. (See Hosking (1981)). For $d < 0.5$, but not equal to zero then, u_{1t} is said to be a *long-memory stationary* process.

2. For $0.5 < d < 1$, the coefficients of the moving average representation are no longer square summable and u_{1t} is *nonstationary* as a consequence. However, the individual

[4] See Hamilton (1994, pp.448-52) for a demonstration of the relationship between the value of d and the coefficients in (6). By square summable, it is meant that the d_j decline at a fast enough rate so that $\sum_j d_j^2 < \infty$.

coefficients still decline to zero with an increase in the lag length. In such a situation, the long-run impact of an innovation on u_{1t} is zero and the error process is still *mean reverting* . In the language of Diebold, Husted and Rush (1991) and Cheung and Lai (1993), amongst others, the *infinite cumulative impulse response* is zero.

3. For $d \geq 1$, u_{1t} is both *nonstationary* and *non mean reverting*, with the infinite cumulative impulse response being finite and non-zero for $d = 1$, and infinite for $d > 1$.

Our proposition is that mean reversion in u_{1t} is the crucial condition for (1) to represent a cointegrating regression, rather than u_{1t} having to be both mean reverting and stationary. As such, we are interested in calculating the posterior probability that $d < 1$. It is also important to produce posterior point estimates of d, in order to estimate the rate at which mean reversion occurs.

Conditional on $d < 1$, β is the parameter of a cointegrating relationship. Inferences regarding β are to be based on an estimate of its marginal posterior density. Since the essence of the method used to produce inferences remains the same when β is a vector of parameters, we are justified in simplifying the exposition by concentrating on the bivariate model.[5]

3. An Approximate Likelihood

3.1 Motivation for the Approximation

The Bayesian approach to inference requires specification of both a likelihood and prior density function. As already alluded to, the exact likelihood function associated with the assumed model does not appear to be a sensible basis for inference. It is of interest to view the form of this likelihood in order to fully motivate our approximation to it.

For a univariate fractionally integrated process as described in (3), with d explicitly

[5] As noted in previous work (see Martin (1995)), the inclusion of a constant term in the model has a non-trivial impact on the convergence of the Gibbs Sampler. We demonstrate our procedure without a constant term, applying it to de-meaned data in Section 6, in order to indirectly allow for a non-zero intercept.

constrained to the range: $-0.5 < d < 0.5$, Sowell (1992) has specified the full matrix of autocovariances for the n-dimensional vector $\mathbf{u}_1 = (u_{11}, u_{12}, \ldots, u_{1n})'$ as having a Toeplitz form:

$$\sigma_{11}\Omega_{11} = \sigma_{11}\{\gamma(k)\} \qquad for \quad k, = 0, 1, 2, \ldots, n-1,$$

where the $\gamma(k)$ are given by:

$$\gamma(k) = \sum_{j=1}^{p} \zeta_j C_j(d, p-k, \rho_j),$$

with the ρ_j representing the p roots of $\Phi(L)$ and the ζ_j and C_j given respectively by:

$$\zeta_j = [\rho_j \prod_{i=1}^{p}(1-\rho_i\rho_j) \prod_{m\neq j}(\rho_j - \rho_m)]^{-1}$$

and:

$$\begin{aligned} C_j(d, p-k, \rho_j) &= \tfrac{1}{2\pi}\textstyle\int_0^{2\pi}[\frac{\rho_j^{2p}}{(1-\rho_j e^{-i\lambda})} - \frac{1}{(1-\rho_j^{-1}e^{-i\lambda})}] \\ &\times (1-e^{-i\lambda})^{-d}(1-e^{i\lambda})^{-d}e^{-i\lambda(p-k)}d\lambda. \end{aligned}$$

If an MA component were assumed to appear in the u_{1t} process, in addition to the AR component, then the parameters of the former would further complicate the form of the $\gamma(k)$. If u_{1t} were assumed to be fractional white noise, (i.e. $p = 0$ above), then the form of the autocovariances would simplify to:

$$\gamma(k) = \frac{\Gamma(1-2d)\Gamma(d+k)}{\Gamma(d)\Gamma(1-d)\Gamma(1-d-k)}. \tag{7}$$

In the case of our cointegration model, we need to specify the covariance matrix for $\mathbf{u} = (\mathbf{u}_1', \mathbf{u}_2')'$, where $\mathbf{u}_2$ denotes the n-dimensional vector $(u_{21}, u_{22}, \ldots, u_{2n})'$. We denote this matrix by Ω and partition it as follows:

$$\Omega = \begin{bmatrix} \sigma_{11}\Omega_{11} & \sigma_{12}\Omega_{12} \\ \sigma_{12}\Omega_{12}' & \sigma_{22}I \end{bmatrix}.$$

It can be shown that the off-diagonal matrix is given by:

$$\sigma_{12}\Omega_{12} = \sigma_{12}\begin{bmatrix} \omega_0 & \omega_1 & \omega_2 & \cdots & \omega_{n-1} \\ \delta_1 & \omega_0 & \omega_1 & \omega_2 & \\ \delta_2 & \delta_1 & \ddots & \ddots & \vdots \\ \vdots & \delta_2 & \ddots & \ddots & \omega_1 \\ \delta_{n-1} & & \cdots & \delta_1 & \omega_0 \end{bmatrix},$$

where:

$$\omega_k = \sum_{j=k}^{\infty} \alpha_j \psi_{j-k}, \quad k = 0, 1, 2, \ldots, n-1,$$

$$\delta_k = \sum_{j=k}^{\infty} \psi_j \alpha_{j-k}, \quad k = 1, 2, \ldots n-1,$$

and the ψ_j and α_j are the coefficients of the infinite MA representations of u_{1t} and u_{2t} respectively. The α_j are given by the expansion of $\Theta(L)^{-1}$ and are thus indirect functions of the θ_j via the roots of $\Theta(L)$, whilst the ψ_j are functions of d and the roots of $\Phi(L)$. For $p = 0$, the ψ_j simplify to:

$$\psi_j = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)}.$$

Hosking (1981) gives the form of the ψ_j when $\Phi(L)$ is a first order polynomial in L.

In total then, the exact likelihood function for the entire parameter set is given by:

$$L(\beta, \sigma_{11}, \sigma_{12}, \sigma_{22}, d, \phi, \theta | \mathbf{y}, \mathbf{x}) \propto |\Omega|^{-1/2} \exp\{\mathbf{u}'\Omega^{-1}\mathbf{u}\}, \tag{8}$$

where $\phi = (\phi_1, \phi_2, \ldots, \phi_p)'$, $\theta = (\theta_1, \theta_2, \ldots, \theta_q)'$ and $\mathbf{y}$ and $\mathbf{x}$ denote the observation vectors for y_t and x_t respectively. If the true value of d were deemed to lie in the nonstationary region of the parameter space, then u_{1t} would need to be differenced prior to basing inferences on (8). Cointegration would then be associated with that part of the parameter space in which $d - 1 = d^*(\text{say}) < 0$, and a lack of cointegration with $d^* \geq 0$.

It is obvious that the truncation of the parameter space at ± 0.5 may well impact on the relative probabilities of cointegration/non-cointegration as calculated from (8). However,

there is an even more fundamental problem associated with the use of the latter. In the subspace of the parameter space in which $d^* = 0$, $\Phi(L) = 1$ and $\Theta(L) = 1$, β is *unidentified.* To see this, we need to note that, when evaluated at these parameter values, the matrices Ω_{11} and Ω_{12} equal the identity matrix I_n, in which case (8) (as applied to Δu_{1t} and u_{2t}) is proportional to:

$$|\Sigma \otimes I|^{-1/2} \exp\{(\Delta \mathbf{u}_1', \mathbf{u}_2')(\Sigma^{-1} \otimes I)(\Delta \mathbf{u}_1', \mathbf{u}_2')'\}, \tag{9}$$

which can, in turn, be expressed as:

$$|\Sigma|^{-n/2} \exp\{-1/2 tr \Sigma^{-1} S\}, \tag{10}$$

where $S = \sum_t [(\Delta y_t - \beta \Delta x_t), \Delta x_t]'[(\Delta y_t - \beta \Delta x_t), \Delta x_t]$. Decomposing (10) into the product of conditional and marginal likelihoods as follows:

$$\begin{aligned} &|\Sigma|^{-n/2} \exp\{-1/2 tr \Sigma^{-1} S\} \\ &= \sigma_{11.2}^{-n/2} \exp\{\tfrac{-1}{2\sigma_{11.2}} \textstyle\sum_t [(\Delta y_t - \beta \Delta x_t) - \tfrac{\sigma_{12}}{\sigma_{22}} \Delta x_t]^2\} . \sigma_{22}^{-n/2} \exp\{\tfrac{-1}{2\sigma_{22}} \textstyle\sum_t [\Delta x_t]^2\}, \end{aligned} \tag{11}$$

where $\sigma_{11.2} = \sigma_{11} - \sigma_{12}^2/\sigma_{22}$, we see that β enters the likelihood via the sum $\beta + (\sigma_{12}/\sigma_{22})$ and, as such, cannot be individually identified. Any flat prior Bayesian inference based on (9) would be invalid, since the posterior density would be improper.

We argue below that, within a Bayesian context, any exact identification problem can potentially be eliminated via a redefinition of the likelihood/posterior as an equivalent (almost everywhere with respect to Lebesque measure) density in which the identification problem is eliminated. However, previous work (see Kleibergen and Van Dijk (1994a and b) and Martin (1995)) has highlighted the fact that a *near* lack of identification in the surrounding region can cause sufficient distortion to preclude sensible inferences. Since, in this case, the region in question is that which incorporates the cointegration/non-cointegration dichotomy, such distortion would be a significant problem.

In all three studies, it was found that an appropriately defined *Jeffreys' prior* was sufficient to offset both the exact and near identification problem. In the present context, we

have been unable to derive such a prior, with the likelihood as defined in (8). However, the approximation to (8) which we present in the following section enables an easy derivation of the requisite prior.[6]

Just as important, with regard to the application of a Gibbs-based MCMC strategy, even if an appropriate identifying prior could be found for (8), the resultant posterior must also be such that the induced conditional posteriors are, at least in the main, of a form from which simulation is both possible (i.e. the conditional densities have a known form) and efficient (i.e. the parameters can be handled in sets). Given the very complicated way in which the parameters enter (8), it is unlikely that it and any prior would combine in such a way to satisfy either of these requirements. In contrast, the approximate likelihood below produces (in combination with the Jeffreys' prior to be defined) a posterior which *is* amenable to an efficient MCMC algorithm.

3.2 The Approximation

The approximation we use is based on the truncation of the infinite expansion of:

$$(1-L)^d u_{1t}$$

to a finite expansion, denoted by:

$$u_{1t} + d_1 u_{1t-1} + d_2 u_{1t-2} + \cdots + d_l u_{1t-l} = D(L)u_{1t}. \tag{12}$$

The d_j are defined as in (6) and are thus different polynomial functions of the single parameter d. We reiterate that the coefficients of the infinite expansion which $(1-L)^d$ represents, approach zero as the lag length increases for any $d > -1$. As such, the truncation of the expansion is valid for any $d > -1$.[7]

[6] We note that Maximum Likelihood applied to (8) would be fraught with convergence problems if the likelihood had a large amount of mass in the region of near lack of identification.

[7] The truncation in (12) has been used in previous Classical work. For example, Chung and Baillie (1993) propose a conditional sum of squares estimator based on this form of truncation as an alternative to the more computationally burdensome exact MLE method of Sowell for the univariate ARFIMA case.

Implementing the truncation in (12), the likelihood function is now proportional to

$$\begin{aligned} &|\Sigma|^{-n/2} \exp\{-1/2 \textstyle\sum_{t=1}^{n} (\Phi(L)D(L)u_{1t}, \Theta(L)u_{2t})\Sigma^{-1}(\Phi(L)D(L)u_{1t}, \Theta(L)u_{2t})'\} \\ &= |\Sigma|^{-n/2} \exp\{-1/2 tr \Sigma^{-1} S\}, \end{aligned} \tag{13}$$

where S is now defined in terms of the data as:

$$S = \sum_{t} (\Phi(L)D(L)(y_t - \beta x_t), \Theta(L)\Delta x_t)'(\Phi(L)D(L)(y_t - \beta x_t), \Theta(L)\Delta x_t). \tag{14}$$

The contemporaneous covariance parameters now enter the likelihood via Σ, separately from the parameters controlling the dynamics, which enter via S. Moreover, if we expand the exponent in (13) as

$$-1/2 * |\Sigma|^{-1} (\sigma_{22} \textstyle\sum [\Phi(L)D(L)u_{1t}]^2 + \sigma_{11} \sum [\Theta(L)u_{2t}]^2 - 2\sigma_{12} \sum [\Phi(L)D(L)u_{1t}\Theta(L)u_{2t}]),$$

we see that each set of time series parameters essentially enters separately from the others. These features of the approximate likelihood function are readily exploited by the MCMC method to be outlined in Section 4.

However, the approximate likelihood in (13) still contains an identification problem. Decomposing the full likelihood in (13) as:

$$\begin{aligned} \sigma_{11.2}^{-n/2} . \exp\{-1/(2\sigma_{11.2}) \textstyle\sum_{t} [(\Phi(L)D(L)y_t - \beta\Phi(L)D(L)x_t) - (\sigma_{12}/\sigma_{22})\Theta(L)\Delta x_t]^2\} . \\ \sigma_{22}^{-n/2} \exp\{-1/(2\sigma_{22}) \textstyle\sum_{t} [\Theta(L)\Delta x_t]^2\}, \end{aligned} \tag{15}$$

β is seen to be unidentified by the approximate likelihood when $\Phi(L) = \Theta(L) = 1$ and $d = 1$, which corresponds once again to the subspace in which there is an lack of cointegration between x_t and y_t. In contrast to the identification problem in the exact likelihood, however, the present problem is solvable via the specification of a particular Jeffreys' prior.

3.3 A Jeffreys' Prior.

In order to motivate the required form for the prior, we need to look a little more closely at the way in which the identification problem manifests itself in the joint posterior density based

on (13). Let us concentrate, for ease of exposition, on the case where $\Phi(L) = \Theta(L) = 1$ is specified from the outset, in which case the parameter set is simply (β, Σ, d). We shall assume prior independence between Σ and the remaining parameters, thereby decomposing the joint prior for all parameters as:

$$p(\beta, \Sigma, d) \propto p(\Sigma).p(\beta, d). \tag{16}$$

For Σ we shall utilize the noninformative Jeffreys' prior, $|I_\Sigma|^{1/2} \propto |\Sigma|^{-3/2}$, where I_Σ denotes the submatrix of the information matrix which relates to the elements of Σ; i.e. $I_\Sigma = E(-\partial^2 \ln L/\partial\Sigma\partial\Sigma')$. For the time being, we shall allow the second component in (16) to be a uniform density. The resultant joint prior enables standard Bayesian analysis to be performed, and the precise nature of the identification problem highlighted.[8]

Given the approximate likelihood function given by (13) and the prior function given by (16), the form of the joint posterior density is:

$$p(\beta, \Sigma, d) \quad \propto \quad |\Sigma|^{-(n+3)/2} . \exp\{-1/2tr(\Sigma^{-1}S)\}, \tag{17}$$

with (β, Σ, d) defined on $D = \mathbb{R}^1 \times \mathbb{S}^{pds} \times \mathbb{R}^1(> -1)$, where $\mathbb{S}^{pds}$ denotes the space of (2×2) positive definite symmetric matrices and $\mathbb{R}^1(> -1)$ denotes the real number line beyond -1.

Standard techniques enable integration with respect to Σ, yielding the following form for the joint density of β and d:

$$p(\beta, d|\, \mathbf{y}, \mathbf{x}) \propto \{C_1 + C_2\beta^2 - 2C_3\beta\}^{\frac{-n}{2}}, \tag{18}$$

where:

$$\begin{aligned}
C_1 &= \sum_t [D(L)y_t]^2 \sum_t [\Delta x_t]^2 - [\sum_t D(L)y_t \Delta x_t]^2, \\
C_2 &= \sum_t [D(L)x_t]^2 \sum_t [\Delta x_t]^2 - [\sum_t D(L)x_t \Delta x_t]^2, \\
C_3 &= \sum_t D(L)y_t D(L)x_t \sum_t [\Delta x_t]^2 - \sum_t D(L)y_t \Delta x_t \sum_t D(L)x_t \Delta x_t
\end{aligned}$$

[8] See Box and Tiao (1973, Chp.2) for an accessible discussion of the sense in which a Jeffreys' prior is "noninformative".

and we further define:

$$C_4 = C_1 - C_3^2/C_2.$$

It is easy to show that when $d = 1$, and $D(L) = 1 - L$ as a consequence, both C_2 and $C_3 = 0$. This in turn implies that the joint density for β and d is a constant function of and, hence, fails to identify β in this region of the parameter space. The integral, with respect to β, of this slice of the joint density at $d = 1$, being unbounded, ascribes an infinite value to the marginal density of d at the point $d = 1$.

For $d \neq 1$, standard integration techniques can again be used to produce respectively the conditional and marginal densities:

$$p(\beta|d, \mathbf{y}, \mathbf{x}) \propto [s_\beta^2]^{-1/2}\{1 + (\beta - \bar{\beta})^2/[(n-1)s_\beta^2]\}^{-n/2} \tag{19}$$

and:

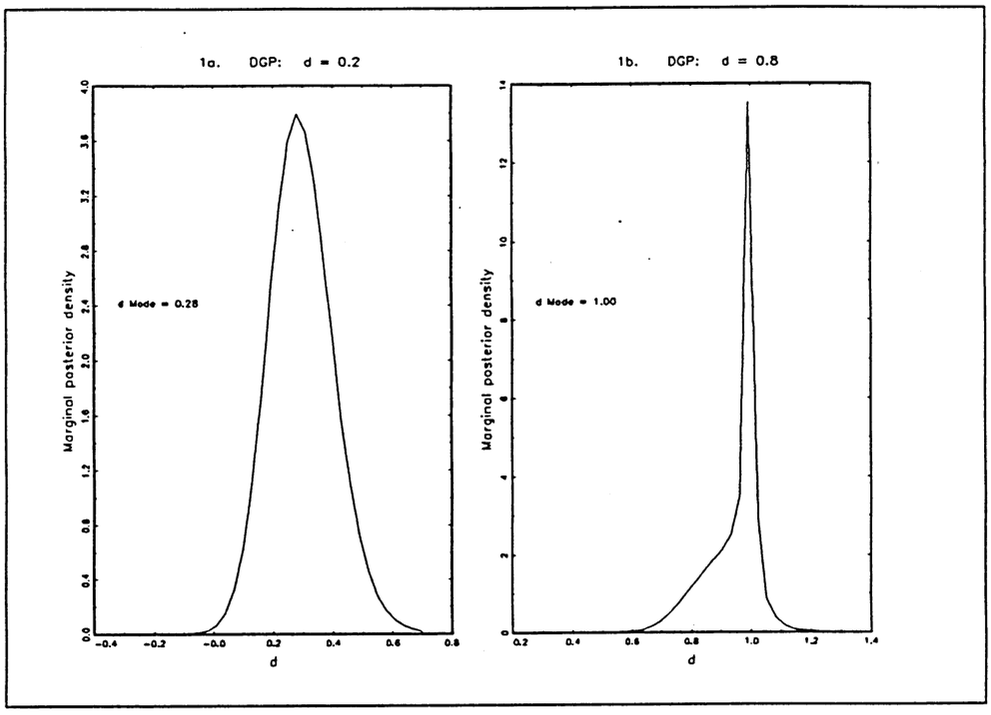
$$p(d|\mathbf{y}, \mathbf{x}) \propto C_2^{-1/2} C_4^{-(n-1)/2}, \tag{20}$$

where $\bar{\beta} = C_3/C_2$ and $(n-1)s_\beta^2 = C_4/C_2$.

If the impact of the lack of identification of β at $d = 1$ were to be felt only at that single point, then the solution would be to simply redefine the joint density function as (17) with support $D^* = D \cap \{(\beta, d, \Sigma); d \neq 1\}$. This amounts to finding a joint density which is equivalent almost everywhere (with respect to Lebesque measure) to the original density, but which does not incorporate the identification problem. Our inferences regarding both β and d would be unaffected by such a change in the definition of the joint posterior.

However, as found in the previously cited work, the impact of a "near" lack of identification can be significant for a wide range of d values around 1, depending on the nature of the true underlying data generating process (dgp). Figures 1a and 1b provide examples of the marginal density of d when the data has been generated from processes with true values of 0.2 and 0.8 respectively for d. As is quite evident, when the bulk of the density is situated

well below $d = 1$, as in Figure 1a, the density appears to be a well-behaved function of d, centred near the true value. In the case of Figure 1b however, the density has a distorted appearance, with an artificial global mode being produced at a point arbitrarily close to 1, despite the true d being well into the mean reversion region. This distortion is typical of densities produced from dgp's with high d values.[9]



Figures 1a and 1b. Marginal posterior of d. (Flat joint prior for β and d)

We can shed more light on the impact of this near lack of identification on the marginal density of d specifically, by analyzing more closely the two quantities C_2 and C_4, of which $p(d|\mathbf{y}, \mathbf{x})$ is comprised. Graphical analysis suggests that, in the range of interest (i.e. for $-1 < d < 1.5$), C_2 behaves like a quadratic function of d which assumes a minimum value of zero at $d = 1$, irrespective of the true value of d in the underlying dgp. The only data

[9] Numerical integration was used to produce these densities and those in Figures 2, 3 and 4. The MCMC algorithm is not applied until Section 6. The value $d = 1$ has been eliminated from the support of the density for the purpose of producing the graphs. The values for β, σ_{12}, σ_{11} and σ_{22} in the dgp underlying these and all other graphs produced in Sections 3 and 5 are 3, 0.5, 1 and 1 respectively.

dependent aspect of C_2 (and, hence $C_2^{-1/2}$) is the degree of concentration of the function around its minimum (maximum) value, and that is, in turn, affected by $\mathbf{x}$ only. C_4 on the other hand, behaves like a quadratic function of d whose minimum value appears to be appropriately influenced by the true value of d.

In summary, the marginal density of d is the product of an apparently well-behaved function of d, $C_4^{-(n-1)/2}$, which seems to possess sensible inferential content regarding d, and a function, $C_2^{-1/2}$, which possesses no such content. The latter, moreover, serves to dominate the former function, for certain dgp's, producing a density with a large amount of probability content around $d = 1$, even when the true d is well into the cointegration region. It would appear to be desirable, therefore, to somehow offset the $C_2^{-1/2}$ factor, in order to produce sensible inferences.

As has been mentioned several times now, the elimination of the impact of the identification problem, of which the distortion induced by $C_2^{-1/2}$ is the manifestation, may be achieved via the use of a particular Jeffreys' prior. The latter is proportional to the determinant of the information matrix and, hence, related to the inverse of the covariance matrix of the relevant posterior density. In the case of potentially unidentified parameters, it should tend to offset the infinite conditional variances which occur at the points at which the parameters become unidentified, as well as eliminating any associated irregularity in the marginal densities in regions of a near lack of identification. As pointed out by Kleibergen and Van Dijk, the success of the Jeffreys' prior in this regard depends crucially on the way in which the expectations within it are evaluated.

In order to derive the appropriate form of the Jeffreys' prior, it is necessary to allow the identification problem to reveal itself in the full likelihood function. That is, the relevant parameter must fail to be identified by the full likelihood function in order for the Jeffreys' prior, as derived from that function, to be operational in terms of offsetting the lack of identification. In our case then, we need to view the likelihood in terms of the decomposed

form in (15), reproduced here for the case where $\Phi(L) = \Theta(L) = 1$:

$$\begin{aligned} L(\beta, \Sigma, d|\mathbf{y}, \mathbf{x}) &\propto |\Sigma|^{-n/2} \exp\{-1/2 tr \Sigma^{-1} S\} \\ &= \sigma_{11.2}^{-n/2}.\exp\{-1/(2\sigma_{11.2}) \sum_t [(D(L)y - \beta D(L)x_t) - (\sigma_{12}/\sigma_{22})\Delta x_t]^2\}. \\ &\quad \sigma_{22}^{-n/2} \exp\{-1/(2\sigma_{22}) \sum_t [\Delta x_t]^2\}, \end{aligned} \tag{21}$$

Maintaining the assumption of prior independence of Σ and the remaining parameters, the first line of (21) can be used to derive the Jeffreys' prior for Σ, namely, $|\Sigma|^{-3/2}$. With this independence assumed, the element σ_{12}/σ_{22} appearing in the second line of (21) can be replaced by the artificial parameter α, and the Jeffreys' prior for β and α conditional on d derived from this first part of the decomposition, being the only part of (21) in which these parameters appear.

We need therefore to derive the determinant of the (2×2) information matrix:

$$I_{\beta,\alpha} = E \begin{bmatrix} -\partial^2 \ln L/\partial\beta^2 & -\partial \ln L/\partial\beta\partial\alpha \\ -\partial^2 \ln L/\partial\alpha\partial\beta & -\partial^2 \ln L/\partial\alpha^2 \end{bmatrix}, \qquad -\partial^2 \ln L/\partial\beta\partial\alpha = \partial^2 \ln L/\partial\alpha\partial\beta,$$

where it is implicit that all differentiation is conditional on d. The elements of this matrix reduce to:

$$\begin{aligned} E(-\partial^2 \ln L/\partial\beta^2) &= (1/\sigma_{11.2}) E \sum_t [D(L)x_t]^2 &&= (1/\sigma_{11.2}) E(\mathbf{x}^{*\prime}\mathbf{x}^*), \\ E(-\partial^2 \ln L/\partial\beta\partial\alpha) &= (1/\sigma_{11.2}) E \sum_t D(L)x_t \Delta x_t &&= (1/\sigma_{11.2}) E(\mathbf{x}^{*\prime}\mathbf{x}^{**}) \qquad \textit{and} \\ E(-\partial^2 \ln L/\partial\alpha^2) &= (1/\sigma_{11.2}) E \sum_t [\Delta x_t]^2 &&= (1/\sigma_{11.2}) E(\mathbf{x}^{**\prime}\mathbf{x}^{**}), \end{aligned}$$

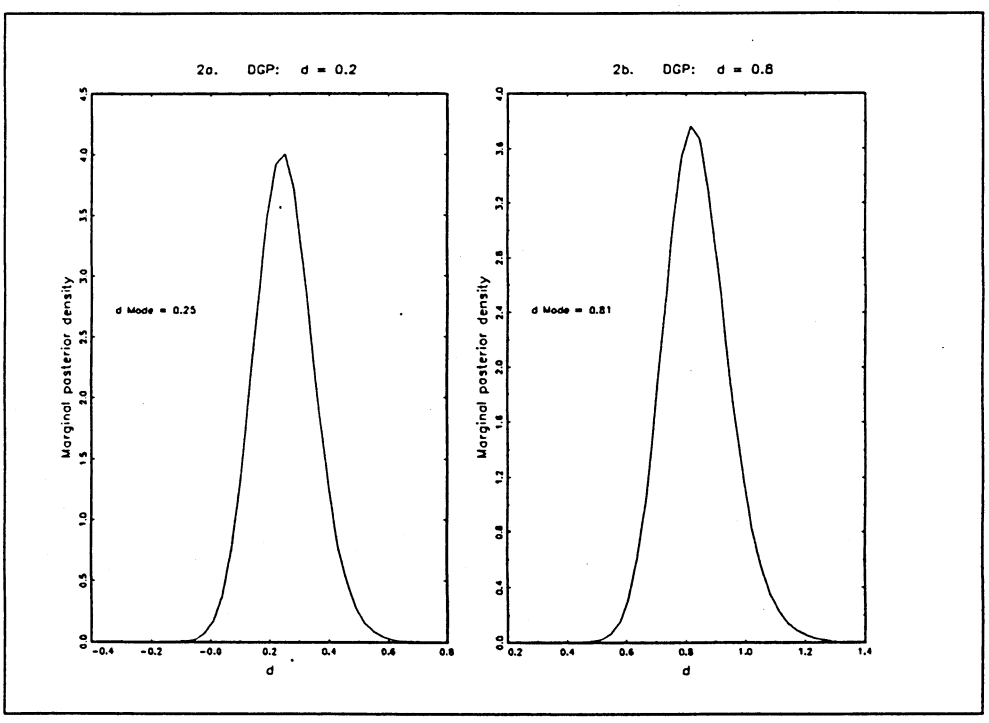
where $\mathbf{x}^*$ and $\mathbf{x}^{**}$ are the observation vectors for $D(L)x_t$ and Δx_t respectively. As such, the Jeffreys' prior, which is proportional to the square root of the determinant, is defined by:

$$|I_{\beta,\alpha}|^{1/2} \propto \{E(\mathbf{x}^{*\prime}\mathbf{x}^*).E(\mathbf{x}^{**\prime}\mathbf{x}^{**}) - [\mathbf{E}(\mathbf{x}^{*\prime}\mathbf{x}^{**})]^2\}^{1/2}. \tag{22}$$

Given that α is only an artificial parameter introduced for the purposes of the derivation, we can effectively view (22) as the prior for β conditional on d.

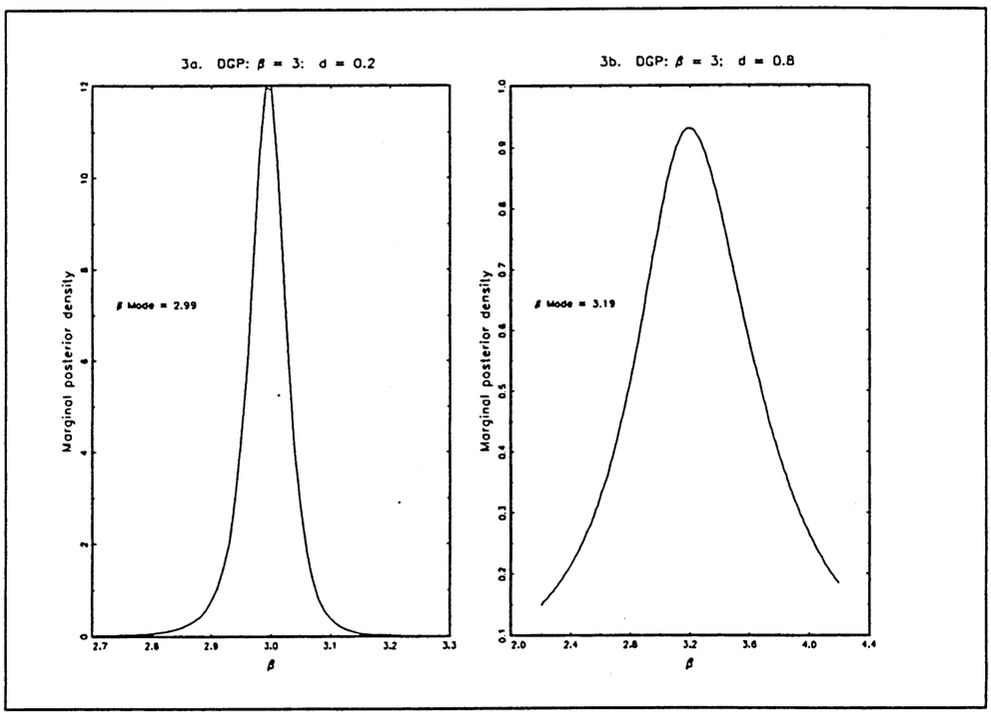
The source of the distortion in the d density, namely the quantity $C_2^{-1/2}$ is equivalent to the inverse of (22) , *so long as the expectations of the functions of* x_t *which appear are replaced by their realized values.* Since no expectation with respect to the x_t process is a function of either of the parameters of interest, namely β and d, this form of evaluation of the expectations implies no loss of information with regard to these parameters. As such, we are justified in using this particular version of the Jeffreys' prior which serves to exactly offset the impact of the identification problem.

Figures 2a and 2b present the "smoothed" versions of the densities in Figures 1a and 1b respectively. As would be anticipated, given the small impact of the $C_2^{-1/2}$ factor on the d density given in Figure 1a, the eradication of it in Figure 2a affects the density only slightly. The difference between the d densities in Figures 1b and 2b, however, is much more marked, with the latter displaying nothing of the irregularity present in the former, and having a modal value very close to the true value.



Figures 2a and 2b. Marginal posterior of d. (Jeffreys' prior for β conditional on d)

It is of interest at this point to also present the marginal densities of β produced by the same dgp and conditional prior as in Figures 2a and 2b respectively. As is evident in Figure 3a, when the degree of fractional cointegration is low, the marginal density of β provides an accurate basis for inference. When the degree of fractional cointegration is high, however, as in Figure 3b, the density of β is highly dispersed, with the modal estimate rather inaccurate. The tail behaviour of the density is such that moments may well not exist. Since this density is typical of those resulting from a high underlying d value, we are wary of using the mean to estimate β. In Sections 5 and 6, we use the mode of the density as the point estimator of β.



Figures 3a and 3b. Marginal posterior of β. (Jeffreys' prior for β conditional on d)

In total then, our prior specification is:

$$p(\beta, \Sigma, d) \propto |\Sigma|^{-3/2} .\{(\mathbf{x}^{*\prime}\mathbf{x}^{*}).(\mathbf{x}^{**\prime}\mathbf{x}^{**}) - (\mathbf{x}^{*\prime}\mathbf{x}^{**})^2\}^{1/2}.p(d). \tag{23}$$

For general $\Phi(L)$ and $\Theta(L)$, we simply redefine $\mathbf{x}^*$ and $\mathbf{x}^{**}$ appropriately and specify the full prior as:

$$p(\beta, \Sigma, d, \phi, \theta) \propto |\Sigma|^{-3/2} .\{(\mathbf{x}^{*\prime}\mathbf{x}^*).(\mathbf{x}^{**\prime}\mathbf{x}^{**}) - (\mathbf{x}^{*\prime}\mathbf{x}^{**})^2\}^{1/2}.p(d, \phi, \theta). \tag{24}$$

To limit the scope of the paper, we shall leave the third component of (24) as a uniform prior.[10] The resultant joint posterior is thus given by:

$$\begin{aligned} p(\beta, \Sigma, d, \phi, \theta | \mathbf{y}, \mathbf{x}) \propto \; & |\Sigma|^{-(n+3)/2} . \exp\{(-1/2) tr \Sigma^{-1} S\}. \\ & \{(\mathbf{x}^{*\prime}\mathbf{x}^*).(\mathbf{x}^{**\prime}\mathbf{x}^{**}) - (\mathbf{x}^{*\prime}\mathbf{x}^{**})^2\}^{1/2}. \end{aligned} \tag{25}$$

It is this density from which the marginal densities of β and d are to be derived via the MCMC procedure to be outlined in Section 4..[11]

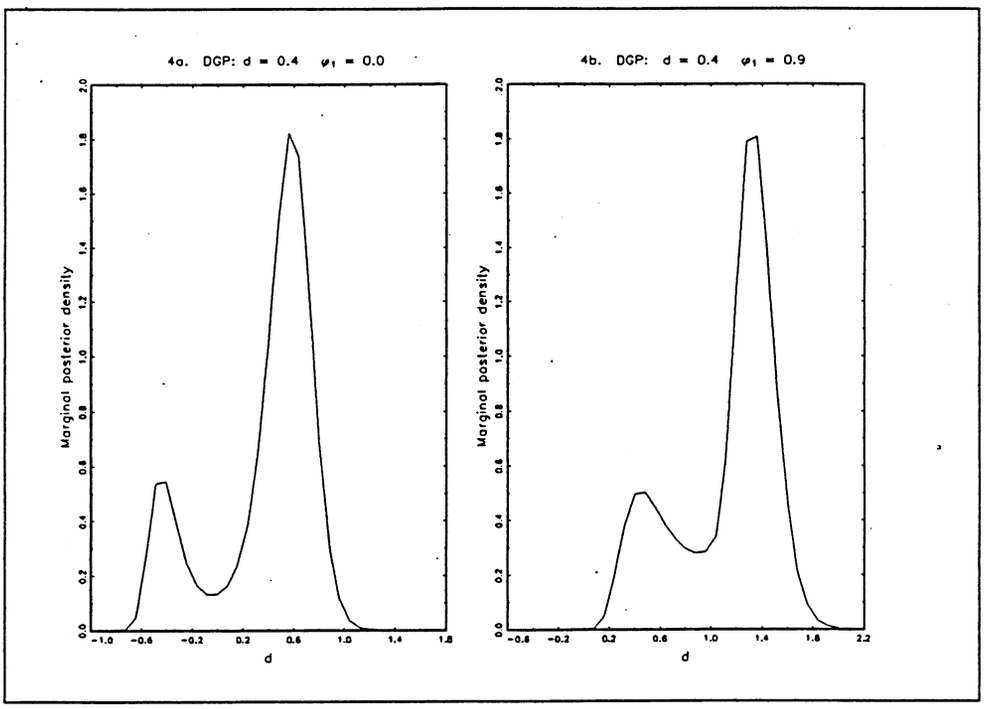
3.4 Bimodality in the marginal posterior for d.

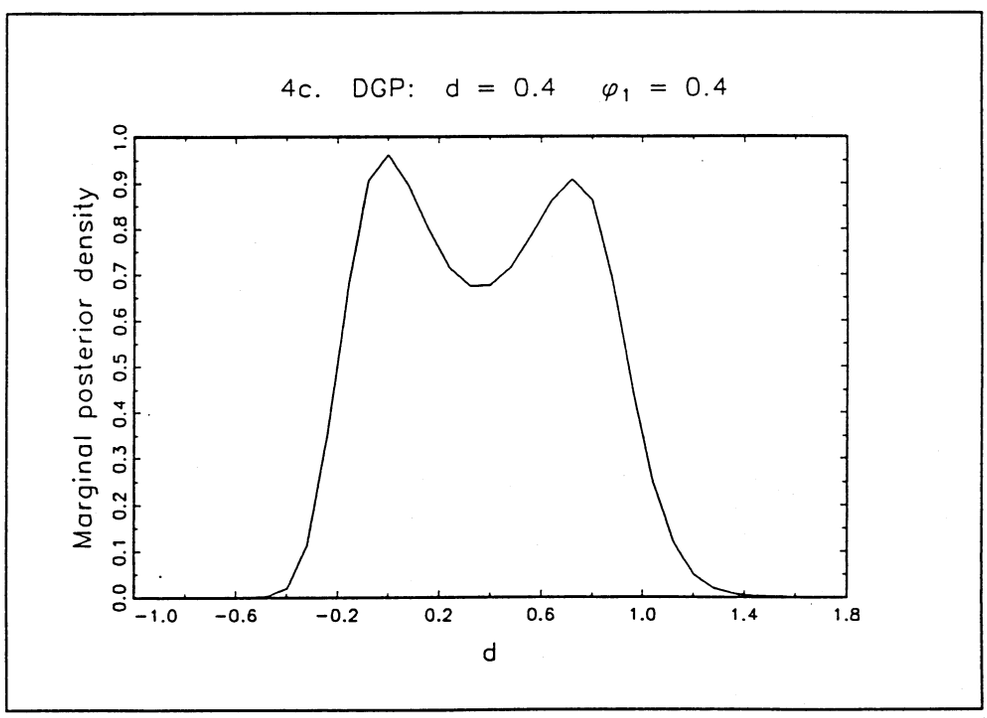
Although the incorporation general $\Phi(L)$ and $\Theta(L)$ polynomials in the model will be shown to be easily catered for by the proposed MCMC algorithm, the inclusion of $\Phi(L)$ specifically leads to confusing inferences regarding d. In Figures 4a b and c respectively, we present examples of marginal d densities for the following three parameter settings in an artificial dgp:

1. $d = 0.4$, $\phi_1 = 0.0$,
2. $d = 0.4$, $\phi_1 = 0.9$ and
3. $d = 0.4$, $\phi_1 = 0.4$.

[10] See Martin (1995) for some discussion of the appropriate marginal prior for time series parameters in this type of model.

[11] We note that (25) can be used for inference in any part of the parameter space for d beyond -1. In practice, this lower bound is unlikely to exert any influence on inferences. Most importantly, the stationary/nonstationary division of the space of d is irrelevant to inference based on (25). In applying the formal convergence criteria from Markov chain theory, it shall be convenient for us to eliminate from the support the sub-space on which the joint posterior density is rendered equal to zero by the conditional Jeffreys' prior. As such, from this point on we shall define the almost everywhere (with respect to Lebesgue measure) equivalent joint posterior as (25) defined on the support $D^* = D \cap \{(\beta, d, \phi, \theta, \Sigma); \mathbf{x}^* \neq \mathbf{x}^{**}\}$, where $D = \mathbb{R}^1 \times \mathbb{R}^1(> -1) \times \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{S}^{pds}$.





Figures 4a, 4b and 4c. Marginal posterior of d. (Jeffreys' conditional prior for β given d and ϕ_1.)

In all settings $\phi_j = 0$, $j = 2, 3, \ldots p$, and $\Theta(L) = 1$. In the parameterization of the model, we allow for the same parameterization as in the dgp; i.e. a one-dimensional polynomial $\Phi(L)$ and $\Theta(L) = 1$.

As is evident in Figures 4a and b, the bimodal d density places one mode in a position which corresponds reasonably closely to the true value of d. However a second mode occurs in either the low or high region of the d parameter space, according to whether the true value of ϕ_1 is low or high. For p-dimensional $\Phi(L)$, the position of the second mode appears to be determined by the true value of the sum of the ϕ_j coefficients, which represents the long-term memory aspect of $\Phi(L)$. Most importantly, depending on the particular data set, the *higher* mode is not necessarily the one pinpointing the underlying true d. Obviously, with such a density, no measure of central tendency, including the higher mode, is an accurate point estimate of d. Further, the relative sizes of $\Pr(d < 1)$ and $\Pr(d \geq 1)$ will be distorted by the second mode, leading to inaccurate probabilities of cointegration and non-cointegration respectively. In Figure 4c, in which the true d and ϕ_1 have the same value, we see the antimode in the d density being the most accurate estimate of that value.

It would appear that, as it stands, the Bayesian method is unable to discriminate between the long-memory fractional component and the long-memory part of the stationary component $\Phi(L)$. It may be that the problem can be offset by some form of prior specification. Or it may be that the Bayesian posterior bimodality is simply a manifestation of a more general problem associated with the fractional integration/cointegration inference when short-run and long-run dynamics enter simultaneously.

4. The Gibbs/Metropolis MCMC Algorithm

In this section we provide a very brief explanation of the hybrid MCMC algorithm used to estimate the marginals of interest. The algorithm is very similar to that used in Martin (1995). As such, we refer readers to that paper for a more detailed description of the method,

including discussion of the required convergence conditions. For recent papers discussing both the theory and implementation of MCMC procedures in general see Tierney (1991), Smith and Roberts (1993) and Roberts and Smith (1994). The book by Tanner (1994) also provides informative and comprehensive discussion of the methods within the broader context of Bayesian computational methods.

4.1 The Gibbs Sampling Algorithm

As applied in a Bayesian context, Gibbs sampling involves sampling from the joint posterior density *indirectly* via an iterative generation of random drawings from all of the conditional (posterior) densities induced by the joint density. Demonstrating the procedure for the case of our specific parameter groupings: β, d, ϕ, θ and Σ, the steps of the algorithm are as follows:[12]

Step 1 Specify initial values for d, ϕ, θ and Σ, $d^{(0)}, \phi^{(0)}, \theta^{(0)}$ and $\Sigma^{(0)}$.

Step 2 Cycle iteratively through the five conditional densities, drawing respectively:

1. $\beta^{(i)}$ from $p_1(\beta^{(i)}|d^{(i-1)}, \phi^{(i-1)}, \theta^{(i-1)}, \Sigma^{(i-1)}, \mathbf{y}, \mathbf{x})$,
2. $d^{(i)}$ from $p_2(d^{(i)}|\beta^{(i)}, \phi^{(i-1)}, \theta^{(i-1)}, \Sigma^{(i-1)}, \mathbf{y}, \mathbf{x})$,
3. $\phi^{(i)}$ from $p_3(\phi^{(i)}|\beta^{(i)}, d^{(i)}, \theta^{(i-1)}, \Sigma^{(i-1)}, \mathbf{y}, \mathbf{x})$,
4. $\theta^{(i)}$ from $p_4(\theta^{(i)}|\beta^{(i)}, d^{(i)}, \phi^{(i)}, \Sigma^{(i-1)}, \mathbf{y}, \mathbf{x})$ and
5. $\Sigma^{(i)}$ from $p_5(\Sigma^{(i)}|\beta^{(i)}, d^{(i)}, \phi^{(i)}, \theta^{(i)}, \mathbf{y}, \mathbf{x})$ until $i = M$.

Given the satisfaction of certain convergence conditions, the realized values, viewed as random variables, converge in distribution to the joint posterior distribution as $M \rightarrow \infty$. Alternatively, with M being large enough for convergence to have occurred, the continued application of the algorithm for a further N iterations produces both a sample of N $(\beta, d, \phi, \theta, \Sigma)$ values from the joint posterior density and a sample of N values of any individual parameter

[12] We demonstrate the MCMC method with $\Phi(L)$ incorporated, despite the demonstrated difficulty it causes for inference.

(parameter set) from its marginal (joint) density.

Obviously, in order for the Gibbs Sampler to be operational, one needs to be able to sample from the conditional densities. The conditionals induced by (25) are given by:

1.

$$p_1(\beta|d,\phi,\theta,\Sigma,\mathbf{y},\mathbf{x}) \propto \exp\{\frac{-1}{2Var(\beta)}(\beta-\bar{\beta})^2\},$$

where $\bar{\beta} = B_1/B_2$, $Var(\beta) = \sigma_{11.2}B_2^{-1}$, $B_1 = \sum_t(x_t^*[y_t^* - (\sigma_{12}/\sigma_{22})x_t^{**}])$, $B_2 = \sum_t(x_t^*)^2$, $y_t^* = \Phi(L)D(L)y_t$ and $x_t^{**} = \Theta(L)\Delta x_t$.

2.

$$\begin{aligned} p_2(d|\beta,\phi,\theta,\Sigma,\mathbf{y},\mathbf{x}) \quad \propto \quad & \exp\{(-1/2)\,|\Sigma|^{-1}\,(\sigma_{22}\textstyle\sum[\Phi(L)D(L)u_{1t}]^2 \\ & +\sigma_{11}\textstyle\sum[\Theta(L)u_{2t}]^2 - 2\sigma_{12}\textstyle\sum[\Phi(L)D(L)u_{1t}\Theta(L)u_{2t}])\}. \\ & \{\textstyle\sum_t(D(L)x_t - \phi'\mathbf{x}_{t-1}^+)^2.\,\sum_t(x_t^{**})^2 - [\sum_t(D(L)x_t - \phi'\mathbf{x}_{t-1}^+)(x_t^{**})]^2\}^{1/2}, \end{aligned}$$

where $\mathbf{x}_{t-1}^+$ denotes the p-dimensional vector $(D(L)x_{t-1}, D(L)x_{t-2}, \ldots, D(L)x_{t-p})'$ and d, the argument of interest, enters the function via the polynomial functions of d which the coefficients of $D(L)$ represent.

3.

$$\begin{aligned} p_3(\phi|\beta,d,\theta,\Sigma,\mathbf{y},\mathbf{x}) \quad \propto \quad & \exp\{-\frac{1}{2\sigma_{11.2}}(\phi-\bar{\phi})'(U_1'U_1)(\phi-\bar{\phi})\}. \\ & \{\sum_t(x_t^+ - \phi'\mathbf{x}_{t-1}^+)^2.\,\sum_t(x_t^{**})^2 - [\sum_t(x_t^+ - \phi'\mathbf{x}_{t-1}^+)(x_t^{**})]^2\}^{1/2}, \end{aligned}$$

where $\bar{\phi} = (U_1'U_1)^{-1}(U_1'[\mathbf{u}_1^+ - (\sigma_{12}/\sigma_{22})\mathbf{x}^{**}]), U_1 = (\mathbf{u}_{10}^+, \mathbf{u}_{11}^+, \ldots, \mathbf{u}_{1n-1}^+)'$, $\mathbf{u}_1^+$ denotes the $n-$dimensional vector with elements $u_{1t}^+ = D(L)u_{1t}$, $\mathbf{u}_{1t}^+$ denotes the $p-$dimensional vector $(u_{1t-1}^+, u_{1t-2}^+, \ldots, u_{1t-p}^+)'$, $x_t^+ = D(L)x_t$ and $\mathbf{x}^{**}$ is the observation vector for the x_t^{**}.

4.

$$\begin{aligned} p_4(\theta|\beta,d,\phi,\Sigma) \propto \exp\{-1/(2\sigma_{22.1})(\theta-\bar{\theta})'(U_2'U_2)(\theta-\bar{\theta})\}. \\ \{\textstyle\sum_t(x_t^*)^2.\,\sum_t(\Delta x_t - \theta'\Delta\mathbf{x}_{t-1})^2 - [\sum_t(x_t^*)(\Delta x_t - \theta'\Delta\mathbf{x}_{t-1})]^2\}^{1/2}, \end{aligned}$$

where $\bar{\theta} = (U_2'U_2)^{-1}(U_2'[\mathbf{u}_2 - (\sigma_{12}/\sigma_{11})\mathbf{u}_1^+])$, $U_2 = (\mathbf{u}_{20}, \mathbf{u}_{21}, \ldots, \mathbf{u}_{2n-1})'$, $\mathbf{u}_{2t}$ denotes the q-dimensional vector $(u_{2t-1}, u_{2t-2}, \ldots, u_{2t-q})'$ and $\sigma_{22.1} = \sigma_{22} - \sigma_{12}^2/\sigma_{11}$.

5.

$$p_5(\Sigma|\beta, d, \phi, \theta) \propto |\Sigma|^{-(n+3)/2} \exp(-1/2tr\Sigma^{-1}S).$$

The densities of β and Σ are Normal and Inverted Wishart respectively. As such they can be simulated from directly, via any Normal variate generator.[13] The way in which the parameter d enters the coefficients of the $D(L)$ polynomial and, in turn, the joint posterior, is such that its conditional density of d is very non-standard in form. However, d is a one-dimensional parameter. As such, simulation from its conditional density is easily performed via a numerical approximation to its inverse cumulative density; i.e. via so-called Griddy Gibbs.[14]

4.2 The Metropolis Algorithm

The presence of the conditional Jeffreys' prior for β given d, ϕ and θ also renders the conditional densities of the parameter vectors ϕ and θ non-standard in form, both densities being proportional to the product of a Normal kernel and the prior density viewed as a function of ϕ and θ respectively. Since we wish to allow these parameter vectors to be multidimensional, a grid-based simulation method is not a desirable option. Several other options are available in such a circumstance, all of which are variants on the idea of drawing from the unattainable (multidimensional) distributions indirectly, via another distribution. We choose to use another Markov chain algorithm, the so-called *Metropolis* algorithm.

[13]To generate values from the Normal density we use the GAUSS command RNDN. Generation of values from the Inverted Wishart density is achieved by taking the inverse of matrix values generated from the associated (non-inverted) Wishart density. The latter simulation is performed by generating (via RNDN) n two-dimensional normal deviates $\mathbf{z}_i = (z_{i1}, z_{i2})'$, with mean zero and variance covariance matrix S^{-1}. The random matrix $\Sigma^{-1} = \sum_i \mathbf{z}_i\mathbf{z}_i'$ then represents a realization of a Wishart variable with degrees of freedom n and covariance matrix S^{-1}. The inverse of this realization is the required realization of Σ.
[14]See Tanner (1994).

The basic thrust of the Metropolis algorithm is to simulate a value of ϕ (θ) indirectly, via a so-called *candidate* density q_3 (q_4), with the latter having the properties of being both a good match for p_3 (p_4) and easy to simulate from. Having a non-standard form, the integrating constant of p_3 (p_4) is obviously unknown. Fortunately, the Metropolis algorithm uses p_3 (p_4) in its unnormalized form only. We have argued previously (see Martin (1995)), that as a consequence of fact that this form of identification problem does not impact at the level of the full conditional densities, that part of p_3 (p_4) which would be present if no adjustment were made for the identification problem, namely the multivariate Normal kernel, represents a good approximation to the overall conditional density. As such, the relevant Normal kernel is used a candidate density in Metropolis generations of $\phi^{(i)}$ and $\theta^{(i)}$, within the Gibbs Sampler.[15]

Once the simulated values have been produced via the hybrid algorithm, estimates of the marginal densities of interest need to be produced for the purpose of inference. Via a Rao-Blackwell type argument, (see Gelfand and Smith (1990) for discussion on this point), it can be shown that the most accurate estimate of either marginal posterior of interest, $p(\beta|\mathbf{y},\mathbf{x})$ or $p(d|\mathbf{y},\mathbf{x})$, is a finite mixture density estimate. Demonstrated for the β marginal, this estimate is given by:

$$\hat{p(\beta|\mathbf{y})} = (1/N)\sum_{i=1}^{N} p_1(\beta|d^{(i)},\phi^{(i)},\theta^{(i)},\Sigma^{(i)},\mathbf{y},\mathbf{x}), \tag{26}$$

where N is the number of simulated sets of parameter values. (26) is, of course, simply the sample estimate of the expectation implicit in the relationship between a conditional and a marginal density.[16]

[15]See Martin (1995) for a more detailed discussion of the use of the Metropolis algorithm in this type of context, including a description of the algorithm itself.

[16]Since the conditional density for β is Normal in form, the component densities to be used in the mixture density estimate are known in their entirety, ie. including their integrating constant. In the case of d however, the relevant conditional density has a non-standard form. As a consequence, one-dimensional numerical integration needs to be performed on each of the N components in the mixture density estimate. This requirement obviously has implications for the speed with which results can be produced. However, the impact does not appear to be burdensome.

4.3 Convergence of the Markov Chain

There are two points to consider in relation to the question of convergence of the hybrid Markov chain to the joint posterior distribution. First, the structure of the Markov chain must be such that this joint posterior represents the so-called stationary or *invariant* distribution of the chain. Second, the structure of the chain must also be such that convergence towards the invariant distribution does indeed occur; i.e. that the chain is *ergodic*. If a straight Gibbs sampler were being used, then we would just need to consider these two points as they pertain to it alone. However, with a Metropolis algorithm embedded within the Gibbs Sampler at two points, we need to also ensure convergence to the relevant conditional distributions, so that the overall algorithm converges to the joint distribution. In Martin (1995), it was argued that the nature of the application was such that the Metropolis subchains were *uniformly ergodic* for the relevant conditional posteriors, whilst the overall outer Gibbs chain was *simply ergodic* for the joint posterior. The application considered in the present paper is sufficiently similar for us to draw the same conclusions regarding convergence.

5. A Monte Carlo Experiment

In this section we present the results of some Monte Carlo simulations in which the repeated sampling performance of the Bayesian inferences are compared with that of various Classical methods. Such a comparison is, of course, irrelevant to a "pure" Bayesian adhering to the likelihood principle. We feel, however, that whatever its inherent advantages, for any Bayesian method to be used in practice, the method must be seen to be acceptable according to standard Classical criteria.

The model underlying the simulations is the simplest version of that discussed in the paper, whereby $\Phi(L) = \Theta(L) = 1$. Since we know that any inference regarding d is problematic when $\Phi(L)$ is incorporated in the model, there seems to be little point in assessing

the repeated sampling performance of the Bayesian estimator in that context. Omitting $\Theta(L)$ as well means that the marginal posterior of d can be produced via one-dimensional numerical normalization of (20) multiplied by the conditional Jeffreys' prior $C_2^{1/2}$, which is a function of d only. The marginal density of β can be produced via two- and one-dimensional integration respectively applied to (18), also multiplied by the prior factor $C_2^{1/2}$. Since very low-dimensional numerical integration is faster than the MCMC method, its use has enabled us to produce repeated sampling results based on a reasonable number of replications. The disadvantage is that we cannot draw any conclusions from these results regarding the impact on inferences about d and β of the short-run dynamics in the x_t process.

In all simulations, the true value of β in (1) is 3, whilst the true value of both variance parameters σ_{11} and σ_{22} in (5) is 1. The latter values mean that the value of 0.5 for σ_{12} quoted at the top of each table, represents the correlation between the underlying error terms e_{1t} and e_{2t}. The number of replications underlying each result is 500, and results for sample sizes of both 50 and 100 are reproduced.[17]

The results reported fall into four categories:

1. The Bias and Root Mean Square Error (RMSE) of the modal estimator of β compared with that of the OLS estimator and the Fully Modified OLS (FMOLS) estimator of Phillips and Hansen (see Phillips and Hansen (1990)). (Table 1)

2. The Bias and RMSE of the modal estimator of d compared with that of the exact Maximum Likelihood estimator (MLE) applied to the OLS residuals from the augmented regression equation $y_t = \beta x_t + \lambda \Delta x_t + \eta_{1t}$. (Table 2)

3. The power of the Dickey Fuller (DF), augmented Dickey Fuller (ADF) and MLE-based

[17]The fractional white noise process for u_{1t} was generated via the finite truncation given in (12), with the truncation made at 30 lags. This method was used for time saving purposes only. The fractionally integrated process can be generated exactly via a Cholesky decomposition of the exact covariance matrix , previously denoted by $\sigma_{11}\Omega_{11}$. The finite expansion $D(L)u_{1t}$ which enters the algebraic expressions for the various densities used in the Bayesian procedure was also truncated at 30 lags. Some experimentation indicated very little change in the results as a result of changing either lag length.

tests against fractional alternatives. (Table 3)

4. The average probability in repeated samples of fractional cointegration, as calculated from the marginal posterior of d. (Table 4)

With reference to Classical estimation of β in the presence of a fractionally integrated error term, Cheung and Lai (1993) prove that the OLS estimator is consistent of $O(n^b)$, with b denoting the reduction in the order of x_t and y_t which the cointegration effects. This result nests the original Stock (1987) convergence result, which applies for $b = 1$. However, with no account taken of either endogeneity or the autocorrelation in the error, the performance of the OLS estimator is expected to be deficient in comparison with an explicit cointegration estimator. The results in Table 1 confirm this expectation to some extent, especially for the larger sample size.

The FMOLS results in Table 1 are based on an estimated long-run variance with a lag length of 10. We conjecture that the number of estimated autocorrelations required to sufficiently account for the long memory error process would adversely affect the small sample performance of the FMOLS procedure, so we have chosen 10 as a compromise value.[18] For the larger sample size, for all values of d in the dgp, the cointegration estimator performs better than (or equivalent to) OLS in terms of bias. However, as d increases, its RMSE exceeds that of OLS. For the smaller sample size, FMOLS is superior to OLS only for small values of d. Neither method provides accurate inference, in terms of either bias or RMSE, for values of d close to 1, for either sample size. The results for $d = 1$ reflect the sort of imprecision which we would anticipate for estimators of a non-cointegrated regression. (See Phillips (1986)).

[18]The results are virtually identical to those based on a lag length of 5, which were not reproduced here as a consequence.

Table 1. The Bias and RMSE of β estimates

$\beta = 3; \quad (1-L)^d u_{1t} = e_{1t}; \quad \sigma_{12} = 0.5.$

			d				
		Estimator	0.2	0.4	0.8	0.9	1.0
$n = 50$	*Bias*	β *mode*	0.006	0.015	0.146	0.286	0.426
		FMOLS	0.013	0.037	0.260	0.421	0.675
		OLS	0.023	0.048	0.258	0.406	0.640
	RMSE	β *mode*	0.051	0.098	0.362	0.519	0.634
		FMOLS	0.053	0.112	0.648	1.061	1.755
		OLS	0.057	0.109	0.572	0.930	1.532
$n = 100$	*Bias*	β *mode*	0.003	0.010	0.096	0.234	0.472
		FMOLS	0.007	0.024	0.199	0.332	0.548
		OLS	0.016	0.035	0.208	0.337	0.548
	RMSE	β *mode*	0.028	0.055	0.251	0.418	0.616
		FMOLS	0.029	0.067	0.446	0.765	1.334
		OLS	0.035	0.071	0.426	0.726	1.262

The modal estimator of β has a uniformly superior performance compared with the Classical estimators. Being a parametric estimator, it has a comparative advantage over a non-parametric method such as FMOLS when the dgp tallies with the assumed parameterization. However, it is still encouraging to see such good repeated sampling behaviour for the

Bayesian estimator, at least for small d. As d moves into the non-stationary, but mean reverting region, the Bayesian estimator, whilst still maintaining its relative superiority, exhibits large bias and RMSE in the same way that the Classical estimators do.

Table 2. The Bias and RMSE of d estimates

$\beta = 3; \quad (1-L)^d u_{1t} = e_{1t}; \quad \sigma_{12} = 0.5.$

			d				
		Estimator	0.2	0.4	0.8	0.9	1.0
$n = 50$	*Bias*	*d mode*	−0.055	−0.055	−0.051	−0.049	−0.045
		MLE	−0.083	−0.099	−0.001	−0.004	−0.019
	RMSE	*d mode*	0.136	0.139	0.146	0.145	0.142
		MLE	0.152	0.145	0.146	0.156	0.170
$n = 100$	*Bias*	*d mode*	−0.031	−0.032	−0.031	−0.030	−0.028
		MLE	−0.042	−0.050	−0.010	−0.016	−0.032
	RMSE	*d mode*	0.082	0.085	0.089	0.088	0.085
		MLE	0.092	0.085	0.103	0.124	0.156

In Table 2, the MLE figures result from the maximization of the exact likelihood function for a univariate fractionally integrated process; i.e.:

$$L(d, s_{11}) \propto |s_{11}\Omega_{11}|^{-1/2} \exp\{\eta'\Omega_{11}^{-1}\eta\},$$

where η denotes the $(n \times 1)$ vector of residuals from estimation of the augmented regression described earlier and s_{11} the variance of the assumed underlying white noise process.[19] The

[19]The augmentation of the cointegrating regression with the regressor Δx_t occurs in order to purge the error term u_{1t} of its correlation with $u_{2t} = \Delta x_t$.

parameter d in the likelihood function is restricted to lie within the range: $-0.5 < d < 0.5$. When the value of d in the true dgp is greater than 0.5, the exact likelihood is applied to the first differences of the residuals, and the value of d estimated indirectly via $d^* = d - 1$. The restriction: $-0.5 < d^* < 0.5$ implies the restriction $0.5 < d < 1.5$.

The Bayesian modal estimator of d is seen to be less biased than the MLE in the stationary region, but more biased (in all but one case) in the non-stationary region. It would appear that the downward bias likely to be present in the sampling distribution of the MLE is offset by the truncation at $d = 0.5$, which comes into effect when the true d exceeds 0.5. Since we have observed, for the simulation size of 500, a tendency for the empirical distribution of the MLE , given a non-stationary dgp, to become more negatively skewed as the sample size increases, the deterioration in the bias of the MLE as the sample size increases is to be anticipated. Whether this is a feature of the theoretical distribution is unknown.

The Bayesian estimator has consistent bias over different d values, an improvement in bias with an increase in n and a RMSE which is less than (or equal to) that of the MLE for all settings of d and n. That the Bayesian estimator mimics the MLE in having a negative bias is to be expected, given that, with the flat marginal prior on d, the posterior mode is closely linked to an MLE.[20]

In Table 3 we present results on the power of the MLE-based test in testing $d = 1$ against $d < 1$. The test statistic is simply the ratio of the MLE to the asymptotic Hessian-based standard error as produced by the Gauss Maximum Likelihood routine.[21] The MLE-based test is seen to largely mimic the power performance of the DF test across the parameter

[20] In Martin (1995) it is shown that a marginal Jeffreys' prior on a time series parameter can alleviate this negative bias to some extent. In the present case, however, the marginal Jeffreys' prior for d is a constant function of d, no matter how the relevant expectations within the prior are evaluated, and, as such, can play no such role.

[21] The critical values used are those associated with an empirical size of 5%, given respectively by -2.551 ($n = 50$) and -3.653 ($n = 100$). The procedure was too time consuming to allow for a more accurate estimate of the true 5% critical value. The 5% critical values for the DF and ADF tests were generated from 10,000 replications of the null model. They are given by: $n = 50$: -2.908 (DF); -2.710 (ADF, 5 lags); -2.537 (ADF, 10 lags); $n = 100$: -2.822 (DF); -2.748 (ADF, 5 lags); -2.663 (ADF, 10 lags).

space. Both this fact, plus the fact that the power of the latter is better for large values of d, is very surprising, given the fact that the Maximum Likelihood method is expressly designed to cater for fractional as opposed to $I(0)$ alternatives. Both tests are able to discern the presence of a stationary fractionally cointegrated equation. However, neither seems able to detect mean reversion alone in the error. The behaviour of the ADF tests illustrates the analytical result in Hassler and Wolters (1994), whereby an increase in the number of lags in an ADF test is shown to cause a decrease in the power of the test against fractional alternatives.

Whilst the DF and ADF tests exhibit appropriate behaviour as the sample size increases, the MLE-based test shows signs of inconsistency for high d values. We are reluctant to comment further on this since, as implied earlier, it may be that 500 replications is not enough to produce an accurate estimate of the true sampling distribution.

Given the apparent weakness of the Classical methods to discriminate between $I(1)$ and slow mean reverting error processes, it is of interest to note the accuracy of the corresponding Bayesian inferences. Table 4 gives the average probabilities (over the 500 replications) of mean reversion, as calculated from the marginal posterior of d. As is quite evident, for both sample sizes, high average probabilities obtain even for large values of d. The average probability of no cointegration, when the true d equals 1 is also high. Note that this probability would be higher, the further is the true d beyond 1.

Table 3. The power of ADF and MLE tests for

Cointegration against Fractional Alternatives

Size = 0.05

$\beta = 3; \quad (1-L)^d u_{1t} = e_{1t}; \quad \sigma_{12} = 0.5$

		d			
	Test	0.2	0.4	0.8	0.9
$n = 50$	*DF*	0.998	0.960	0.158	0.078
	ADF (5)	0.448	0.262	0.076	0.036
	ADF(10)	0.138	0.090	0.060	0.058
	MLE	1.000	1.000	0.210	0.078
$n = 100$	*DF*	1.000	1.000	0.282	0.130
	ADF (5)	0.918	0.560	0.110	0.078
	ADF (10)	0.462	0.250	0.078	0.062
	MLE	1.000	1.000	0.184	0.068

The average Bayesian probabilities are, of course, not directly comparable with the Classical powers. However, as a summary measure of the ability of the procedure to pinpoint the nature of the generating process, they suggest an accuracy that the Classical tests, whose abilities are summarized in the power measurements, lack.

Table 4. The Average Probability of Drawing the Correct Bayesian Inference About Cointegration

$\beta = 3; \quad (1-L)^d u_{1t} = e_{1t}; \quad \sigma_{12} = 0.5$

		d				
		0.2	0.4	0.8	0.9	1.0
$n = 50$	$\Pr(d < 1)$	1.000	1.000	0.911	0.801	
	$\Pr(d \geq 1)$					0.508
$n = 100$	$\Pr(d < 1)$	1.000	1.000	0.976	0.890	
	$\Pr(d \geq 1)$					0.531

6. Testing for Long-run PPP

In this section we use the Bayesian procedure to assess the validity of the PPP doctrine for certain countries. We have obtained data for the same time period and for three of the same countries as did Cheung and Lai (1993), in order to allow for some comparison between our Bayesian inferences and their Classical results. The comparison is, however, limited by the fact that 1) our data derives from a different source than does theirs for some of the sample period; and 2) our sample size is reduced to 50 compared with their sample size of 76 as a result of our using 26 lags in the approximation of $(1-L)^d$. We supplement our results with FMOLS estimates of the cointegrating parameter, estimates of d based on the Maximum Likelihood procedure described in Section 5, and DF and MLE-based unit root tests applied to the OLS residuals.[22]

[22]We do not apply any ADF tests as a result of their apparent lack of power to detect fractional alternatives.

Specifically, we consider the PPP relation linking the price levels of three countries, France, Japan and the United Kingdom, to the price level of the United States. For each country, we estimate the relationship:

$$fp_t = \beta p_t + u_{1t}$$

where fp_t denotes the log of the foreign price index (CPI) measured in U.S. currency and p_t is the log of the U.S. price index (CPI). These variables correspond to y_t and x_t in our previous notation, with their joint determination assumed to be described by the general fractional cointegration model given by (1) to (5) in Section 2, but with $\Phi(L)$ set to 1. As in the previous section, we have decided to avoid the problems associated with the incorporation of $\Phi(L)$ in the model. We present results based on values of 0, 2 and 4 respectively for the degree, q, of the $\Theta(L)$ polynomial. A more exhaustive study would choose q via some sort of model selection criterion. It is worth reiterating at this point that the MCMC method copes well with large q (and indeed p) values, where such values may well be necessary if an MA component were to be captured.[23]

The exchange rate and CPI data used in the analysis is annual data for the period 1914−1989. The data for the period 1949 −1989 is taken from the Reserve Bank International Supplement, and is extended back to 1914 via the data published in Lee (1978).

We begin by presenting in Table 5, modal estimates of d, along with the associated probabilities of mean reversion in the error. The latter probabilities, all exceeding 80% as they do, are very supportive of the existence of PPP between the U.S. and all countries considered. However, the high modal estimates of d suggest that the rate of reversion to parity

[23]Prior to conducting the cointegration analysis, x_t and all five y_t variables were accepted as unit root processes by ADF tests with 5 lags. Also, all data was demeaned in order to cater for the lack of intercept term. The MCMC density estimates are produced using the following iteration strategy. After a "burn-out" period of M, we take into the sample the output of every rth iteration, the intermittent sampling of the chain aimed at speeding up convergence via a breaking of the Markovian dependence. With N denoting the final number of sample values from which density estimates are constructed, the MCMC densities are produced from a total of $(r + [N + (M/r)]) - M$ iterations for the outer Gibbs chain. For all examples, we use $N = 1000$, $M = 300$ and $r = 10$. In all instances, we perform 20 iterations of the Metropolis sub-chain before taking a value as a realization from the relevant conditional density.

is very slow. Cumulative impulse responses based on these modal estimates are presented in Table 6.[24] The column for the U.K., for example, indicates that a one-unit shock to the parity relation, although having zero effect in the infinite future, still has an impact of almost 0.5 units after 50 years.

Table 5. Modal Estimates of d and the Probability of Fractional Cointegration

		Country		
		France	*Japan*	*U.K.*
d Mode	$q = 0$	0.601	0.640	0.840
	$q = 2$	0.601	0.640	0.800
	$q = 4$	0.561	0.561	0.760
Pr(d<1)	$q = 0$	0.988	0.970	0.860
	$q = 2$	0.988	0.985	0.887
	$q = 4$	0.992	0.987	0.911

Given the negative bias exhibited by the modal estimator in the Monte Carlo experiments, we could view our estimates as underestimating the true value of d, and, as a consequence, the sluggishness of the mean reversion. However, this fact should be counterbalanced by the fact that the d estimate is being forced to soak up all of the dynamic behaviour in u_{1t}, and may assume a larger value as a consequence. We note that the impact on inferences about d of an increase in the degree of autocorrelation in u_{2t} (i.e. q), is not very large.

[24]The impulse responses are calculated from the formula: $u_{1t} = (1-L)^{-d}e_{1t}$, which is appropriate given the specification of $\Phi(L) = 1$. The value of d in the formula is replaced by the modal estimate, based on $q = 0$.

Table 6. Impulse Responses

Based on the Modal

Estimates of d

	Country		
	France	*Japan*	*U.K.*
10 Years	0.265	0.307	0.612
30 Years	0.172	0.209	0.516
50 Years	0.141	0.174	0.476
∞	0.000	0.000	0.000

The Maximum Likelihood estimates of d presented in Table 7 are of a similar magnitude to the Bayesian modes. On the basis of such high estimates, however, the MLE-based test is able to reject the null of a unit root in the error in one case only, and then only at the 10% level. The DF test on the other hand rejects at the 5% level in favour of a stationary error in all three cases.[25]

Table 8 provides modal point estimates and 95% HPD interval estimates of β, based on the three assumed models.[26] The point estimates for France and the U.K. are reasonably close to the value of 1 which would obtain if the homogeneity condition were imposed. The estimates for Japan, however, are well in excess of 1, indicating that an assessment of PPP based on the real exchange rate may be inappropriate. For France and Japan, in particular, the impact on the modal estimates of β of an increase in q is very small. However, for all three countries, it is evident from the HPD estimates that allowance for higher degrees of autocorrelation in the x_t process produces correspondingly more concentrated β marginals.

[25]The critical values used for the DF and MLE-based tests are the $n = 50$ critical values produced numerically for the Monte Carlo experiments.

[26]A 95% HPD (Highest Probability Density) estimate is an interval with 95% probability coverage, whose inner density ordinates are not exceeded by any density ordinates outside the interval.

Table 7. Classical Inference

Regarding

Fractional Cointegration

((**) denotes rejection at the 10% (5%) level)*

	Country		
	France	*Japan*	*U.K.*
ML Estimate	0.667	0.715	0.839
ML t Stat.	-1.982*	-1.773	-0.944
DF t Stat.	-3.528**	-2.924**	-3.041**

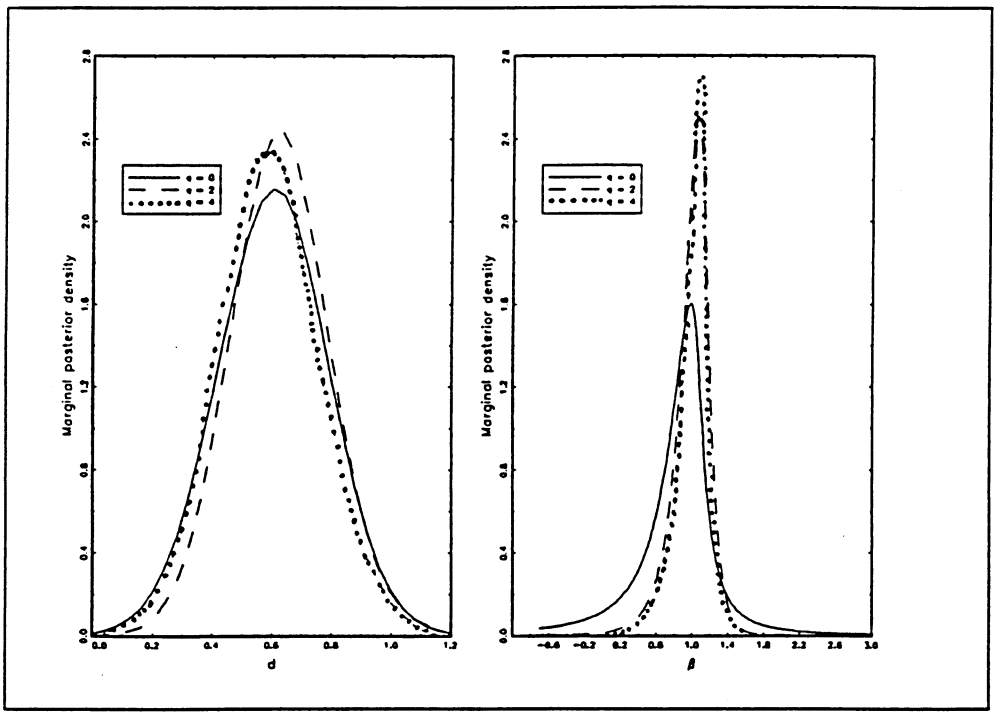


Figure 5. Marginal d and β densities for the French/U.S. relation.

Table 8. Estimates of β

		Country		
		France	*Japan*	*U.K.*
β Mode	*$q = 0$*	0.980	1.280	1.070
	$q = 2$	1.060	1.410	1.070
	$q = 4$	1.070	1.460	1.080
95% HPD	*$q = 0$*	(-0.27,1.68)	(-0.45,2.04)	(-0.46,2.37)
	$q = 2$	(0.50,1.39)	(0.53,1.84)	(0.46,1.36)
	$q = 4$	(0.54,1.38)	(0.74,1.88)	(0.56,1.38)
FMOLS		1.092	1.537	1.135
(s.e.)		(0.045)	(0.092)	(0.034)

The French densities reproduced in Figure 5 illustrate this fact. The d densities, on the other hand, illustrate the afore-mentioned relative lack of impact of q on the d marginals.

The FMOLS estimates tally most closely with the modal estimates based on $q = 4$. Asymptotically, the associated t ratios are Standard Normal. Based on the standard errors quoted, the 95% confidence intervals are much more narrow than the corresponding Bayesian interval estimates. However, since the former intervals have asymptotic justification only, they may be very misleading for a sample size of 50.

Finally, our results are roughly in accordance with those of Cheung and Lai. Their GPH estimates of d for the three countries we have considered all fall within the range $4.5 - 6.0$. Since their sample is larger than ours, we would anticipate these smaller d estimates, since the longer data span allows more time for mean reversion to occur. Interestingly, in the light of our comments about the lack of power of the ADF test against fractional alternatives,

their ADF tests do fail to detect cointegration in the case of both Japan and the U.K.

7. Conclusions

In this paper, we have argued that the appropriate characterization of a cointegrating relation is one in which the error term is mean reverting. In order to model mean reversion, the traditional procedure of equating cointegration with an $I(0)$ error term must be extended to one in which the error term is allowed to be fractionally integrated of order d, with d any real number between -1 and 1. We have presented a procedure for estimating the probability of mean reversion, or fractional cointegration, in addition to the values of the fractional differencing parameter d and the cointegrating parameter(s). Inferences are based on marginal posterior densities, which are able to be estimated in a straight forward manner via a combination of MCMC methods.

We have produced results which suggest that the Bayesian inferences have a repeated sampling performance which is very competitive with that of Classical alternatives. This finding, in conjunction with what we see as the intrinsic benefits of the Bayesian method, namely the production of a full density function as a basis for inference, and the opportunity of formally incorporating a prior density, lend strong support to the application of the method in this context. The issue of prior density specification has been particularly crucial. If we were not able to incorporate the conditional Jeffreys' prior, the identification problem, which is fundamental to the form of fractional cointegration model used, and, therefore, relevant to both Bayesian *and* Classical inference, may need to be offset via some more ad-hoc method. Moreover, it is the Bayesian paradigm, in which analysis of posterior density functions is valid, which has allowed the identification problem to manifest itself in such a way that the exact prior factor required to offset it has been made evident.

REFERENCES

Box, G.E.P. and G.C. Tiao, 1973, *Bayesian inference in statistical analysis,* (Addison-Wesley Publishing Co., Reading).

Cheung, Y-W. and K.S.Lai, 1993, A fractional cointegration analysis of purchasing power Parity, *Journal of Business and Economic Statistics* 11, No.1, 103-112.

Chung, C-F. and R.T. Baillie, 1993, Small sample bias in conditional sum-of-squares estimators of fractionally integrated ARMA models, *Empirical Economics,* 18, 791-806.

Chib, S. and E. Greenburg, 1994, Bayes inference in regression models with ARMA(p,q) errors, *Journal of Econometrics,* 64, 183-203.

Diebold, F.X., Husted, S. and M. Rush, 1991, Real exchange rates under the Gold Standard, *Journal of Political Economy* 99, No. 6, 1252-1271.

Dueker, M. and R. Startz, 1993, Fractional integration and cointegration, *Working Paper, Federal Reserve Bank of St. Louis.*

Gelfand, A.E. and A.F.M. Smith, 1990, Sampling based approaches to calculating marginal densities, *Journal of the American Statistical Association* 85, 398-409.

Geweke, J. and S. Porter-Hudak, 1983, The estimation and application of long memory time series models, *Journal of Time Series Analysis,* 4, 221-238.

Hamilton, J.D., 1994, *Time series analysis,* (Princeton University Press, Princeton).

Hassler, U. and J. Wolters, 1994, On the power of unit root tests against fractional alternatives, *Economic Letters,* 45, 1-5.

Hosking, J.R.M., 1981, Fractional differencing, *Biometrika,* 68, No. 1, 165-176.

Johansen, S., 1995, A statistical analysis of cointegration for I(2) variables, *Econometric Theory,* 11, 25-59.

Kitamura, Y., 1995, Estimation of cointegrated systems with I(2) processes, *Econometric Theory,* 11, 1-24.

Kleibergen, F., and H.K. van Dijk, 1994a, On the shape of the likelihood/posterior in cointegration models, *Econometric Theory* 10, 514- 551.

Kleibergen, F., and H.K. van Dijk, 1994b, Bayesian analysis of simultaneous equation models using noninformative priors, *Working paper, Econometric Institute and Tinbergen Institute, Rotterdam, The Netherlands.*

Lee, M.K., 1978, *Purchasing power parity,* (Marcel Dekker, New York).

Martin G.M., 1995, Bayesian analysis of a cointegration model using Markov chain Monte Carlo, *Manuscript, Monash University.*

Phillips, P.C.B., 1986, Understanding spurious regressions in Econometrics, *Journal of Econometrics* 33, 311-340.

Phillips, P.C.B., and B.E. Hansen, 1990, Statistical inference in instrumental variables regression with I(1) processes, *Review of Economic Studies* 57, 99-125.

Roberts, G.O. and A.F.M. Smith, 1994, Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms, *Stochastic Processes and their Applications* 49, 207-216.

Schotman, P. and H.K. van Dijk, 1991, On Bayesian routes to unit roots, *Journal of Applied Econometrics* 6, 387-401.

Silvapulle, P., 1995, A score test for seasonal fractional integration and cointegration, *Working Paper, University of Iowa.*

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, 3-23.

Sowell, F., 1992, Maximum likelihood estimation of stationary univariate fractionally integrated models, *Journal of Econometrics,* 53, 165-188.

Stock, J.H., 1987, Asymptotic properties of least squares estimators of cointegrating vectors, *Econometrica,* 55, 1035-1056.

Stock, J.H. and M.W. Watson, 1993, A simple estimator of cointegrating vectors in higher order integrated systems, *Econometrica,* 61, No. 4, 783-820.

Tanner, M.A., 1994, *Tools for statistical inference,* second edition (Springer-Verlag, New York, NY).

Tierney, L., 1991, Markov chains for exploring posterior distributions, *Technical Report No. 560., University of Minnesota School of Statistics.*