University of Leicester

DEPARTMENT OF ECONOMICS

BAYESIAN INFERENCE IN COINTEGRATED I (2) SYSTEMS: A GENERALISATION OF THE TRIANGULAR MODEL

Rodney W. Strachan, University of Leicester, UK

Working Paper No. 05/14
July 2005

Bayesian Inference in Cointegrated $I(2)$ Systems: a Generalisation of the Triangular Model.

Rodney W. Strachan
Department of Economics, University of Leicester,
Leicester, U.K.
email: rws7@le.ac.uk

July 22, 2005

ABSTRACT

This paper generalises the cointegrating model of Phillips (1991) to allow for $I(0)$, $I(1)$ and $I(2)$ processes. The model has a simple form that permits a wider range of $I(2)$ processes than are usually considered, including a more flexible form of polynomial cointegration. Further, the specification relaxes restrictions identified by Phillips (1991) on the $I(1)$ and $I(2)$ cointegrating vectors and restrictions on how the stochastic trends enter the system. To date there has been little work on Bayesian $I(2)$ analysis and so this paper attempts to address this gap in the literature. A method of Bayesian inference in potentially $I(2)$ processes is presented with application to Australian money demand using a Jeffreys prior and a shrinkage prior.

1 Introduction

This paper generalises Phillips' (1991) triangular model for cointegration in several directions and also simplifies the specification and analysis of a wide range of $I(2)$ processes of interest in cointegration studies. In addition, the identifying restrictions and the form by which the stochastic trends enter the system are relaxed, and the model is extended to allow for both a range of deterministic processes. The range of $I(2)$ processes permitted in this specification includes some that appear not to have been considered in the literature previously. Having established the model specification, this paper takes a Bayesian approach to inference.

The focus of this paper is cointegration analysis. Since the development of the concept of cointegration by Granger (1983) and Engle and Granger (1987), many applied papers have investigated equilibrium economic relations by treating the cointegrating relations among the variables as those equilibrium relations. That is, their focus was on the cointegrating space as defined by the cointegrating vectors. Examples of such relations that have been specified this way include money demand, purchasing power parity, term structures of interest rates, income-wealth relations, and balanced growth hypotheses. The extent of the literature in this area prevents us giving even a reasonable coverage, but a very small sample of specific examples include Campbell (1987), Galí (1990), Johansen and Juselius (1990), King, Plosser, Stock and Watson (1991), Lettau and Ludvingson (2004), and Garratt, Lee, Pesaran and Shin (2003).

For much of the past decade and a half the most popular model used for studying cointegrating relations has been the cointegrating vector error correction (VECM) form of the VAR model. A valid alternative model for cointegration analysis was proposed by Phillips (1991) to permit full information maximum likelihood analysis. This model has a triangular structure to introduce the stochastic trends and models the cointegrating relations directly as estimated equations. Short run dynamics are modelled with a flexible vector autoregressive form and permits the same range of deterministic processes as can be specified in the VECM. This model treats the cointegrating relations as the objects of interest, in accordance with the focus in the more applied papers already mentioned. Phillips (1994) compared the finite sample properties of estimators of the normalised cointegrating vectors from a triangular model with those from the reduced rank regression model. In that paper, he showed that while the density of the former is Gaussian with finite integer moments, little can be said of the latter except that it has Cauchy-like tails and no finite integer moments.

The triangular model has proven very useful for theoretical analysis and for such purposes as the development of data generating processes in Monte Carlo studies or in asymptotic theory. Stock and Watson (1993) point to a number of early classical applications using this model by Campbell (1987), Campbell and Shiller (1987, 1989) and Bewley (1979). More recently Gael Martin has used this model in Bayesian cointegration analysis in a series of papers (Martin & Martin 2000, Martin 2000, and Martin 2001).

Given the success this model has achieved in theoretical work and in these few applied papers, it would appear that this is an under-appreciated model worthy of further attention. While little guidance is given in the literature as to why this model is not employed or even considered more often, we speculate that it may be the very features of this model that permit a focus upon cointegration analysis that impose structures economists or econometricians find inflexible and so unattractive. For example, how the stochastic trends enter the system or the restrictions chosen to identify the cointegrating vectors may be perceived as overly inflexible. As discussed later, Phillips himself noted possible concerns with these restrictions.

A first contribution of this paper, then, is to provide a generalisation of the of the triangular model to permit investigation of a wide range of cointegrating processes within a relatively simple, less structured but unified framework. The approach uses the basic implications of cointegration that if an n-vector x_t is $I(1)$, there exists an $n \times r$ matrix β such that $\beta' x_t$ is $I(0)$ and the common stochastic trends may be regarded as $\beta'_{\perp} x_t \sim I(1)$ where $\beta' \beta_{\perp} = 0$. From a brief outline of the important features of the triangular model as first proposed by Phillips (1991), we use the results of Strachan and van Dijk (2003) to generalise the model so as to relax several of the more binding restrictions, while retaining the cointegrating relations as the focus of the analysis. This generalisation removes the triangular structure proposed by Phillips (1991), but the new specification is shown to encompass the model of Phillips and map to the standard VECM.

Another significant contribution of this paper is to extend the model to allow inference upon more general processes, particularly $I(2)$ processes. The model is developed to allow for $I(0)$, $I(1)$ or $I(2)$ processes and to allow for a wide range of deterministic processes within the stochastic trends and the cointegrating relations. There have been a number of classical applied studies that have investigated the support for possible $I(2)$ cointegration. Examples include Paruolo (1996), Rahbek, Kongsted, and Jørgensen (1999), Fiess and MacDonald (2001), Nielsen (2002), Kongsted (2003), and Georgoutsos and Kouretas (2004). We develop a simple method of considering the form of $I(2)$ relations used in the papers mentioned, but extend to other interesting possibilities. We also demonstrate how to investigate the support for restrictions upon the cointegrating space that may be implied by various economic theories.

After specifying the general form for the likelihood for the model, we are able to take either a Bayesian or classical approach in developing of inferential methods. This paper takes an explicitly Bayesian approach as we are not aware of any paper that presents a fully Bayesian method of inference on $I(2)$ processes.

Cointegration analysis involves investigating questions such as: What is the dimension of the cointegrating space? Are the variables $I(0)$, $I(1)$ or $I(2)$? or, Do a particular set of equilibrium relations form cointegrating relations? While these questions focus upon the cointegrating relations, a restriction encountered in Bayesian cointegration analysis using either the triangular model of Phillips (1991) or the VECM is that improper priors on the nuisance parameters 'outside the cointegrating space' cannot be used due to Barlett's paradox. Strachan and van Dijk (2003) demonstrate how to be uninformative about these nuisance parameters using flat improper priors by focusing the model upon the cointegrating relations. It is relatively straightforward to demonstrate that it is possible to use a range of improper or proper priors for this model that have attractive properties on grounds of information theoretic justifications, invariance, estimation performance, or simply familiarity. One prior that is commonly used for Bayesian analysis is the improper Jeffreys prior and another contribution of this paper is the development of the Jeffreys prior for these parameters outside the cointegrating space.

The plan of this paper is as follows. In Section 2, we present the triangular model of Phillips (1991) and the basic extensions of Strachan and van Dijk (2003) for the I(1) model. We build upon this basic model to demonstrate how to incorporate I(2) analysis, deterministic processes and overidentifying restrictions. The priors are outlined or, in the case of the Jeffreys prior, developed in Section 3. In Section 4, we present a brief application to investigate the evidence in support of a stable money demand relation in Australian data. Section 5 concludes with some remarks and suggestions for further directions for research.

2 The model

In this section we present the essential features of the model specification and then extend this to allow for a range of processes.

We begin with the assumption that the the $n \times 1$ vector x_t is $I(1)$ such that it can be given the representation

$$x_t = C \sum_{i=1}^{t} \nu_i + \mu_0 + z_t$$

where $z_t \sim I(0)$, ν_i is an IID zero mean process and C is of reduced rank. Later in the paper we will relax the assumption that x_t is $I(1)$ to allow for $I(0)$ and $I(2)$ processes. In fact we will relax the assumption that all elements of x_t be integrated of the same order. Let $x_{1,t}$ be the $r \times 1$ vector of the first r elements of x_t and $x_{2,t}$ be the $(n-r) \times 1$ vector of the remaining elements so $x_t' = \left(x_{1,t}'\ x_{2,t}'\right)$. If the elements of x_t cointegrate with an r-dimensional cointegrating space, then the triangular model of Phillips (1991) brings the $(n-r)$ stochastic trends into the system by assuming $w_{2,t} = \Delta x_{2,t}$ is $I(0)$ where $\Delta = 1 - L$ and L is the lag operator such that $Lx_t = x_{t-1}$. The cointegrating relations are given by $x_{1,t} = Bx_{2,t} + w_{1,t}$ where $w_{1,t}$ is $I(0)$. Thus we have the $n \times 1$, $I(0)$ vector $w_t = \left(w_{1,t}'\ w_{2,t}'\right)'$ described by the two equations

$$w_{1,t} = \left[I_r \ \ -B'\right] x_t = \beta' x_t \quad (1)$$

$$w_{2,t} = \Delta x_{2,t}. \quad (2)$$

The matrix of cointegrating vectors for the system x_t is given in the relations in $w_{1,t}$ as $\beta = \left[I_r \ \ -B'\right]'$. With this specification linear restrictions have been employed to uniquely identify the cointegrating vectors and so permit estimation of the cointegrating space. Phillips notes that the above specification of $w_{1,t}$ attaches a specific importance to the variables in $x_{1,t}$. The use of linear identifying restrictions requires the assumption that we have some minimal knowledge of the cointegrating space such that we know the appropriate restrictions to apply (Strachan, 2003). In various papers such as Boswijk (1996), Luukkonen *et al.* (1999) and Strachan (2003), however, examples of seemingly sensible restrictions of this form are shown, in fact, to be invalid. To relax this restriction, Strachan and van Dijk (2003) respecify (1) as $w_{1,t} = \beta' x_t$ where $\beta'\beta = I_r$ such that no structure is imposed *a priori* upon the cointegrating relations.

Given the uncertainty over whether certain macroeconomic time series have unit roots (see discussion in Bauwens, Lubrano & Richard, 1999 with reference to Sims and Uhlig 1991 and the 1991 special issue of the *Journal of*

Applied Econometrics), it would seem overly restrictive to assume we always know that all variables in a particular vector have unit roots and, if some do not, that we know which variables these are. Therefore to relax the assumptions on how the stochastic trends enter the system, we replace (2) with the specification $w_{2,t} = \beta_\perp' \Delta x_t$ where $\beta_\perp' \beta_\perp = I_{n-r}$ and $\beta' \beta_\perp = 0$[1]. If $r = n$, then $\beta \in O(n)$ and if $r = 0$, then $\beta_\perp \in O(n)$. This gives us the new form of equations (1) and (2) as

$$w_t = \begin{bmatrix} w_{1,t} \\ w_{2,t} \end{bmatrix} = \begin{bmatrix} \beta' \\ \beta_\perp' \Delta \end{bmatrix} x_t = \begin{bmatrix} \beta' x_t \\ \beta_\perp' \Delta x_t \end{bmatrix} \sim I(0). \tag{3}$$

In an obvious abuse of notation, we use notation similar to

$$\begin{bmatrix} \beta' \\ \beta_\perp' \Delta \end{bmatrix} x_t \text{ to represent } \begin{bmatrix} \beta' & 0 \\ 0 & \beta_\perp' \end{bmatrix} \begin{bmatrix} x_t \\ \Delta x_t \end{bmatrix}$$

throughout the paper.

As we will later extend the above model to define a number of $I(2)$ processes to allow $I(2)$ analysis, we will call the process in (3) Process 0.

The equations in (3) describe the most fundamental relations implied by $I(0) - I(1)$ cointegration and so describe long run behaviour or equilibrium relations with respect to x_t. Phillips does not, in general, place restrictions upon the short run behaviour of w_t beyond the requirement that it be a stationary process. However, to demonstrate how optimal inference can achieved with this model he assumes that w_t is iid Normal with zero mean and fixed covariance matrix. We take a specification between these two cases. A reasonably flexible form of short run dynamics that is commonly assumed in both theoretical and applied work, and which we adopt here, is a vector autoregressive form. That is, we assume

$$w_t = \sum_{i=1}^{l} \Pi_i w_{t-i} + \varepsilon_t \tag{4}$$

and $\varepsilon_t \sim iid\, N(0, \Sigma)$, such that $w_t = C(L)\varepsilon_t$ where the polynomial $C(L) = I + C_1 L + C_2 L^2 + C_3 L^3 + ...$ is such that $C^w = \sum_{i=1}^{\infty} C_i \neq 0$ and full rank

[1] $\beta_\perp$ is simply specified as any determinsitic function of β such that $\beta_\perp$ lies in the orthogonal complement of the space of β. For example, we may choose $\beta_\perp$ to be the $(n-r)$ eigenvectors associated with the unit eigenvalues of $I_n - \beta\beta'$.

This condition implies $\beta'\beta_\perp = 0$ and $\beta_\perp$ given β, is not random.

so $w_t \sim I(0)$. It can readily be shown using (3) that this model implies $\beta' x_t = [I_r \quad 0] C(L) \varepsilon_t \sim I(0)$ while $\beta'_\perp x_t = [0 \quad I_{n-r}] C(L) \sum_{i=1}^{\infty} \varepsilon_t \sim I(1)$ (see Appendix I).

The model in (4) is a generalisation of the triangular model of Phillips (1991) and we can also show how this specification can be obtained from a VECM and *vice versa*. The VECM representation of a cointegrating $VAR(l)$ model is

$$\Delta x_t = \alpha \beta' x_{t-1} + \sum_{i=1}^{l-1} \Gamma_i \Delta x_{t-i} + \nu_t. \tag{5}$$

If we premultiply the above by B' where $B = [\beta \quad \beta_\perp]$ and make use of the relation $I = \beta\beta' + \beta_\perp \beta'_\perp$, we can rearrange the above equation into (4) to show that a $VAR(l)$ in x_t will map to a $VAR(l)$ in w_t with simple restrictions on the dynamic structure. The reverse approach, beginning from (4) and mapping to (5), shows that a $VAR(l)$ in w_t will map to a $VAR(l+1)$ in x_t again with simple restrictions on the dynamic structure (see Appendix II for the exact forms).

The model above gives a representation for the process x_t with potential $I(0) - I(1)$ cointegration. We now need to allow for deterministic processes, over-identifying restrictions, and the possibility that some elements of x_t may be $I(2)$.

Restrictions upon the cointegrating space: It is not uncommon that an economist will have some apriori support for, or belief in, a particular cointegrating space as representative of a set of possible equilibrium relations. These relations may, for example, be suggested by theories on financial relations such as term structures of interest rates or price-dividend relations, money demand, or reaction functions of policy-makers. Therefore we need to consider how to obtain inference on potentially interesting cointegrating spaces. Denote the cointegrating space as $\mathfrak{p} = sp(\beta)$, and the support of this parameter is the Grassman manifold which we denote by $G_{r,n-r}$, such that $\mathfrak{p} \in G_{r,n-r}$. It is common to specify a likely cointegrating space as a matrix of coefficients. For example, if we believe that a bivariate system $x_t = (y_t \, z_t)'$ will have the cointegrating relation $y_t - z_t$, then we would identify this space with the matrix $H = \begin{pmatrix} 1 & -1 \end{pmatrix}'$. For a further example, if we believe that in a four variable system $x_t = (u_t \, v_t \, y_t \, z_t)'$ the combinations $u_t - v_t - y_t$, and $y_t - z_t$ are either stationary or the variables enter the cointegrating relations via these combinations, we would specify the matrix

$$H=\begin{bmatrix} 1 & 0 \\ -1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix}.$$

In each case we have specified a $n\times q$ $(q\geq r)$ matrix H and the space is $\mathfrak{p}^0 = sp(H)$. Since $sp(H) = sp\left(H\left(H'H\right)^{-1/2}\right)$, and it is mathematically simpler to work with H as a semiorthogonal matrix, it is suggested that after specifying H, this matrix be mapped to the Stiefel manifold by $H\rightarrow H\left(H'H\right)^{-1/2}$ before entering it into the model.

To incorporate this restriction on the cointegrating space into the model, we can specify $\beta = H\kappa$ and $\mathfrak{p} = sp(\beta) = sp(H\kappa) \subseteq sp(H)$ where $\kappa \in V_{r.q}$. This specification is rather dogmatic as it implies certainty about the location of the cointegrating space as $\mathfrak{p} = \mathfrak{p}^0 = sp(H)$ if $r = q$ and $\mathfrak{p} \subset \mathfrak{p}^0$ if $r < q$. If taking a Bayesian approach and a less dogmatic prior is preferred, then the informative prior developed in Strachan and Inder (2004) can be used to assign positive mass over all regions in the space $G_{r,n-r}$. Nonlinear restrictions can also be easily implemented as demonstrated in Koop, Potter and Strachan (2004).

Deterministic processes: To allow for deterministic processes such as nonzero means and trends in the stationary processes in w_t, and linear drifts in the nonstationary processes in x_t, we assume,

$$\begin{aligned} \underline{\beta}'x_t + \mu_1 + \delta_1 t &= w_{1,t} \sim I(0) \text{ and} \\ \underline{\beta}'_{\perp}x_t + \beta'_{\perp}x_0 + \mu_2 t &= \sum_{i=1}^{t} w_{2,i} \sim I(1) \text{ or} \\ \underline{\beta}'_{\perp}\Delta x_t + \mu_2 &= w_{2,i} \sim I(0) \end{aligned}$$

such that $w_t = \left(w'_{1,t}\ w'_{2,t}\right)' \sim I(0)$ with nonzero means (we will revisit this extension when we introduce the $I(2)$ specification). We can now re-specify the matrix of cointegrating vectors as $\beta' = \begin{pmatrix}\mu_1 & \delta_1 & \underline{\beta}'\end{pmatrix}$, and a matrix lying in the orthogonal complement of the cointegrating space as $\beta'_{\perp} = \begin{pmatrix}\gamma_2 & \mu_2 & \underline{\beta}'_{\perp}\end{pmatrix}$. Further, denote any $(n-r)$ columns of $\beta_{\perp}$ as $\beta^c_{\perp}$ and let $\underline{x}_t = (1,\, t,\, x_t')'$ such that

$$\begin{aligned} w_{1,t} &= \begin{pmatrix} \mu_1 & \delta_1 & \underline{\beta}' \end{pmatrix} \underline{x}_t = \beta' \underline{x}_t \\ w_{2,t} &= \begin{pmatrix} \gamma_2 & \mu_2 & \underline{\beta}_\perp' \end{pmatrix}^c \Delta \underline{x}_t = \beta_\perp^{c\prime} \Delta \underline{x}_t \end{aligned}$$

and again we have

$$w_t = \begin{bmatrix} w_{1,t} \\ w_{2,t} \end{bmatrix} = \begin{bmatrix} \beta' \underline{x}_t \\ \beta_\perp^{c\prime} \Delta \underline{x}_t \end{bmatrix} \sim I\,(0)\,.$$

Note that the first row of $\beta_\perp$ does not enter the model since $\Delta \underline{x}_t = (0,\, 1,\, \Delta x_t')'$ and so this vector γ_2 is not identified. This implies the likelihood is not a function of γ_2 and, under the usual specifications of parameters with unbounded support, this lack of identification would pose serious problems for inference. In classical methods this can imply nonuniqueness of the estimates while in Bayesian methods nonintegrable posteriors can result. However, we specify β and $\beta_\perp$ to be elements of Steifel manifolds such that the matrix $B = [\beta\, \beta_\perp]$ is also an element of a Steifel manifold[2]. As the support of β is compact, integrals over a constant function over this support will be finite and so posteriors will be proper.

The above specification encompasses the case were there are fewer deterministic terms. These may be simply specified as particular subspaces within the cointegrating space. For example, if there is no trend in the equilibrium relations $w_{1,t}$, this would imply $sp\,(\beta) \subseteq sp\,(H)$ where

$$H = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & I_n \end{bmatrix}$$

and this is imposed by setting $\beta = H\varphi$ and the $(n+1) \times r$ matrix φ is treated as the unkown parameter on which we specify the prior and estimate from the posterior. In practice, however, this is only necessary when the number of deterministic processes entering $w_{1,t}$ differs from the number entering $w_{2,t}$. If, for example, $\mu_2 = 0$ while μ_1, δ_1 and γ_2 are all non zero. Otherwise we simply restrict β to $\begin{pmatrix} \mu_1 & \underline{\beta}' \end{pmatrix}'$ or $\underline{\beta}$ as required and define $\beta_\perp$ accordingly.

On the general issue of identification, β is identified up to its orientation in the space it spans, however as the posterior is invariant to rotation within this

[2] If there are no deterministic processes such that β is $n \times r$, then $B = [\beta\, \beta_\perp]$ will be an element of the orthogonal group, $B \in O\,(n)$.

space, this simply implies integrals with respect to the posterior need only be adjusted by a known constant to accomodate the discrepancy between an integral over the space $V_{r,n}$ and the an integral over the space $G_{r,n-r}$ (Strachan and Inder 2004 and Strachan and van Dijk 2003). This lack of concern for the orientation of β within its space is in line with the principal argued in Strachan and Inder (2004) and Villani (2005) that it is the cointegrating space spanned by the cointegrating vectors that is the object of interest in cointegration analysis rather than the cointerating vectors themselves. Once an estimate of β is obtained, we only require that the vectors in the matrix $\beta_\perp$ lie in the orthogonal complement of the space spanned by β, and their orientation within this space does not matter (again due to invariance).

I(2) processes: There exist a range of different representations of cointegrating $I(2)$ processes and we develop simple models of four important cases of these processes to permit what we will call $I(0) - I(1) - I(2)$ analysis. While few applications exist, to date, of the first three cases, the fourth and most complicated case seems to have attracted a reasonable amount of attention in the literature (see for example Paruolo 1996, Haldrup and Salmon 1998, Rahbek, Kongsted, and Jørgensen 1999, Fiess and MacDonald 2001, Nielsen 2002, Kongsted 2003, and Georgoutsos and Kouretas 2004).

We will use the following general model (see Johansen, 1995) of $I(2)$ processes to demonstrate the essential features of these processes,

$$x_t = C_2 \sum_{j=1}^{t} \sum_{i=1}^{j} \nu_i + C_1 \sum_{i=1}^{t} \nu_i + \mu_0 + z_t$$

where $z_t \sim I(0)$ and ν_i is an IID zero mean process. The first, and simplest case (Process I) we consider assumes there exists an $n \times r$ matrix β such that $\beta' C_1 = \beta' C_2 = 0$ and therefore $\beta' x_t \sim I(0)$. This process can be readily accommodated within the model for w_t by simply redefining $w_{2,t}$ as $w_{2,t} = \beta'_\perp \Delta^2 x_t \sim I(0)$. That is,

$$w_t = \begin{bmatrix} \beta' \\ \beta'_\perp \Delta^2 \end{bmatrix} x_t = \begin{bmatrix} \beta' x_t \\ \beta'_\perp \Delta^2 x_t \end{bmatrix} \sim I(0).$$

In the second process (Process II), we assume there exists an $n \times r$ matrix β such that $\beta' C_2 = 0$ but $\beta' C_1 \neq 0$ and therefore $\beta' x_t \sim I(1)$, which defines x_t as cointegrated CI(2,1) in the notation of Engle and Granger (1987). To extend the model to allow for this case we define

$$w_t = \begin{bmatrix} \beta' \Delta \\ \beta'_\perp \Delta^2 \end{bmatrix} x_t = \begin{bmatrix} \beta' \Delta x_t \\ \beta'_\perp \Delta^2 x_t \end{bmatrix} \sim I(0).$$

Process III adopts the same assumption as Process II that $\beta' C_2 = 0$, but we further assume that there exists an $r \times s$ $(s \leq r)$ matrix η such that $\eta' \beta' C_1 = 0$ and so the process $\eta' \beta' x_t \sim I(0)$. We will refer to η as the $I(2)$ cointegrating vectors. A simple example of this case is given in Johansen (1995, p. 37). The model specification begins with the most basic relations of similar form to those in (3). That is, we know that $x_t \sim I(2)$, but there exists r linearly independent combinations of the elements of x_t that are $I(1)$, that is $\beta' x_t \sim I(1)$. Further, we know there exists s linearly independent combinations of the elements of $\beta' x_t$ that are $I(0)$, that is $\eta' \beta' x_t \sim I(0)$. These relations imply

$$w_t = \begin{bmatrix} \eta' & 0 \\ \eta'_\perp \Delta & 0 \\ 0 & I_{n-r}\Delta \end{bmatrix} \begin{bmatrix} \beta' \\ \beta'_\perp \Delta \end{bmatrix} x_t = \begin{bmatrix} \eta' \beta' x_t \\ \eta'_\perp \beta' \Delta x_t \\ \beta'_\perp \Delta^2 x_t \end{bmatrix} \sim I(0).$$

Again β and $\beta_\perp$ are $n \times r$ and $n \times (n-r)$ respectively, and the new matrices η and $\eta_\perp$ are $r \times s$ and $r \times (r-s)$ $(s \leq r)$ respectively, and we specify these matrices to be semiorthogonal such that $[\eta \, \eta_\perp] \in O(r)$.

Notice that at $s = 0$, then $\eta_\perp \in O(r)$ and since $\eta'_\perp \beta' C_1 \neq 0$, Process III collapses down to Process II. At $s = r$, then $\eta \in O(r)$ and since $\eta' \beta' C_1 = 0$, this case collapses down to Process I. Finally, we see that this process encompasses Process 0 which we obtain if $r = n$ so that $\beta \in O(n)$ and $\beta_\perp$ is not defined. In this case the matrix η plays the role of the $I(1)$ cointegrating vectors, not β.

In the fourth and most general process (Process IV) we begin again with the matrices β and $\beta_\perp$ which are $n \times r$ and $n \times (n-r)$ respectively. These matrices are assumed to satisfy the conditions $\beta' C_2 = 0$ and there exists an $n \times s$ matrix $\eta = [\eta'_1 \quad \eta'_2]'$ where η_1 is $r \times s$ and η_2 is $(n-r) \times s$, such that $\eta'_1 \beta' C_1 + \eta'_2 \beta'_\perp C_2 = 0$.

The implications of these specifications are that premultiplication of x_t by β reduces the process from $I(2)$ to $I(1)$, such that $\beta' x_t \sim I(1)$, $\beta'_\perp \Delta x_t \sim I(1)$ and that the vector

$$z_t = \begin{bmatrix} \beta' x_t \\ \beta'_\perp \Delta x_t \end{bmatrix} \sim I(1)$$

cointegrates with (multiple) cointegrating vector η such that $\eta' z_t \sim I(0)$. In this case x_t is called polynomially cointegrated and there exists an $n \times (n-s)$

matrix $\eta_\perp$ such that $\eta'_\perp \eta = 0$. The specification of $w_t \sim I(0)$ is then

$$\begin{aligned} w_t &= \begin{bmatrix} w_{1,t} \\ w_{2,t} \end{bmatrix} = \begin{bmatrix} \eta' \\ \eta'_\perp \Delta \end{bmatrix} \begin{bmatrix} \beta' \\ \beta'_\perp \Delta \end{bmatrix} x_t \\ &= \begin{bmatrix} \eta' \\ \eta'_\perp \Delta \end{bmatrix} \begin{bmatrix} \beta' x_t \\ \beta'_\perp \Delta x_t \end{bmatrix} = \begin{bmatrix} \eta' z_t \\ \eta'_\perp \Delta z_t \end{bmatrix}. \end{aligned}$$

To explore this process further, we recall the partition of η and partition $\eta_\perp$ as

$$\eta = \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} \text{ and } \eta_\perp = \begin{bmatrix} \eta_{11\perp} & \eta_{12\perp} \\ \eta_{21\perp} & \eta_{22\perp} \end{bmatrix}$$

such that η_1 is $r \times s$, and $\eta_{11\perp}$ is $r \times (r-s)$ which defines the dimensions of the remaining submatrices: η_2 is $(n-r) \times s$; $\eta_{21\perp}$ is $(n-r) \times (r-s)$; $\eta_{12\perp}$ is $r \times (n-r)$; and $\eta_{22\perp}$ is $(n-r) \times (n-r)$. Further, $[\eta\, \eta_\perp] \in O(n)$. Therefore

$$w_t = \begin{bmatrix} \eta'_1 \beta' x_t + \eta'_2 \beta'_\perp \Delta x_t \\ \eta'_{11\perp} \beta' \Delta x_t + \eta'_{21\perp} \beta'_\perp \Delta^2 x_t \\ \eta'_{12\perp} \beta' \Delta x_t + \eta'_{22\perp} \beta'_\perp \Delta^2 x_t \end{bmatrix}.$$

This specification encompasses Process III and collapses to that process where $\eta_2 = 0$, $\eta_{21\perp} = 0$, and $\eta_{12\perp} = 0$[3].

We can obtain the specification for polynomial cointegration in many previous studies if

$$\eta_1 = \begin{bmatrix} I_s \\ 0 \end{bmatrix}. \tag{6}$$

As most earlier $I(2)$ studies have focussed upon the number and form of stochastic trends entering the process, deterministic trends in $I(2)$ processes have recieved little attention relative to that for $I(1)$ processes. Useful exceptions include Rahbek, Kongsted, and Jørgensen (1999) who develop a VAR model for $I(2)$ processes with careful consideration of the specification of the deterministic processes, and the application by Kongsted (2003) to the analysis of import price determination. To generalise the model in this paper to permit deterministic process such as nonzero means, linear trends and drifts in an $I(2)$ process, we take the same approach as in the previous subsection. Using the specific restrictions in (6) of Process IV as an example,

[3] Note that the additional 'restriction' $\eta_{22\perp} = I_{n-r}$ is not a restriction since so long as $\eta_{22\perp}$ is any element of $O(n-r)$ the conditions of Process III are met and the likelihood and posterior will be invariant to choice of $\eta_{22\perp}$ from within $O(n-r)$.

redefine the β and $\beta_\perp$ as $\underline{\beta}$ and $\underline{\beta}_\perp$ respectively, partition $\underline{\beta}$ as $\underline{\beta} = \left[\underline{\beta}_1, \underline{\beta}_2\right]$ where $\underline{\beta}_1$ is $n \times s$ and we assume $\Delta x_0 = 0$ and

$$\begin{aligned}\underline{\beta}' x_t + \mu_1 + \delta_1 t &\sim I(1) \text{ and} \\ \underline{\beta}_\perp' x_t + \underline{\beta}_\perp' x_0 + \mu_0 t &= \sum_{j=1}^{t} \sum_{i=1}^{j} w_{2,i} \sim I(2).\end{aligned}$$

Next, partition μ_1 and δ_1 conformably with $\underline{\beta}$ as $\mu_1 = [\mu_{11}, \mu_{12}]$ and $\delta_1 = [\delta_{11}, \delta_{12}]$ such that

$$\begin{aligned}\underline{\beta}_1' x_t + \eta_2' \underline{\beta}_\perp' \Delta x_t + [\mu_{11} + \eta_2' \mu_0] + \delta_{11} t &\sim I(0) \\ \underline{\beta}' \Delta x_t + \delta_1 &\sim I(0) \text{ and} \\ \underline{\beta}_\perp' \Delta^2 x_t &\sim I(0).\end{aligned}$$

We may write the first equation as

$$\eta' \begin{bmatrix} \underline{\beta}' x_t + \mu_1 + \delta_1 t \\ \underline{\beta}_\perp' \Delta x_t + \mu_0 \end{bmatrix} = \eta' \begin{bmatrix} \left(\mu_1, \delta_1, \underline{\beta}'\right) \\ \left(0, \mu_0, \underline{\beta}_\perp'\right) \Delta \end{bmatrix} \begin{pmatrix} 1 \\ t \\ x_t \end{pmatrix}.$$

While there are many alternative specifications of the deterministic terms and even of the $I(2)$ processes, the above design allows us to specify the deterministic terms for each process of interest directly, and to restrict the parameters for these processes to a compact space without loss of generality.

3 The Likelihood and The Jeffreys prior

An important component in Bayesian analysis is the specification of the prior distribution for the parameters in the model. In this section we present the likelihood and the Jeffreys prior, and discuss another prior with attractive properties that might also be used for the Π_i with this model: the shrinkage prior. For this discussion, we will find it convenient to rewrite the model as follows. Beginning with the model in (4), we collect the lags of w_t into $\omega_t = \left[w_{t-1}', w_{t-2}', \ldots, w_{t-l}'\right]$ and define $\Pi = [\Pi_1\, \Pi_2\, \ldots\, \Pi_l]'$ such that (4) can be re-expressed as

$$w_t' = \omega_t \Pi + \varepsilon_t'.$$

Before we develop the priors, however, we bring to the reader's attention an attractive feature of the model specification used in this paper. That is, we are less restricted in the specification of the priors than we would be if we were using the cointegrating vector error correction model as we do not encounter the issue of Barlett's (1957) paradox. Bartlett's paradox essentially states that if we use improper priors on all parameters (many ignorance priors tend to be improper), the posterior probability attached to the smallest model will be one and all other (larger) models will be assigned a posterior probability of zero. Strachan and van Dijk (2004) demonstrate that this issue is linked directly to the relative dimensions of the parameter spaces in the different models and the subsequent rate of divergence of the prior normalising constants[4] for improper priors.

To explain, we separate the priors into those for parameters determining the long run behaviour, β and η, and those determining the short run dynamics, Π and Σ. Across the models of cointegrating processes, the dimensions of Π and Σ do not change. The dimensions of β and η do change, but these parameters have compact supports such that the most appropriate ignorance or diffuse priors will be proper. This feature is useful in a Bayesian approach as it implies we may use ignorance priors upon all parameters and the Bayes factors and posterior probabilities for the various models will still be well defined. There are several precedents for using improper priors only for the common (to all models) parameters and proper priors for the remaining parameters when computing posterior model probabilities. See for example, Fernándes, Ley and Steel (2001) and further examples listed in Kass and Raftery (1995).

Several prior specifications have been proposed in Bayesian cointegration literature, each having one or more attractive properties. Geweke (1996) proposed a proper Normal-Inverted Wishart prior which permitted an efficient Gibbs sampling scheme for the VECM. This prior avoids problems of local nonidentification in computation and implies a prior on the cointegrating space that is coherent with the use of linear identifying restrictions. Although the issue of local nonidentification does not arise in this model specification

[4] Clearly an improper prior cannot be normalised so by 'normalising constant for an improper prior' we mean the integral of the kernal of the prior over a support of given diamter.

By rate of divergence, we mean the rate at which the normalising constant diverges to infinity as the diameter of the support increases to infinity.

used in this paper, a Normal prior for Π and a Wishart for Σ^{-1} could be used that would provide analytical tractability and simplify computation.

Early work by Stein (1956, 1960, and 1962) established that the shrinkage prior can produce estimates with smallest frequentist loss (smaller than for the maximum likelihood estimator) and the resultant estimator is therefore admissible. This prior for $\pi = vec\,(\Pi)$ has the form $(\pi'\pi)^{-\left(n^2l-2\right)/2}$ and has been investigated and employed by several authors (see for example Stein 1956, 1960, 1962, Lindley 1962, Lindley and Smith 1972, Sclove 1968, 1971, Zellner and Vandaele 1974, Berger 1985, Judge *et al.* 1985, Mittelhammer *et al.* 2000, and Leonard and Hsu 2001). While the early work demonstrated admissibility in specific situations, in more general models it has also been shown to produce an estimator with smaller expected frequentist loss than standard estimators as may result from flat or proper informative priors (see for example, Zellner 2002 and Ni and Sun 2003). Of particular importance to this paper is work by Ni and Sun (2003) who provide evidence of improved performance with this prior for estimating the parameters of a VAR.

In an attempt to reconcile classical and Bayesian evidence on unit roots in macroeconometric time series, Phillips (1991a) proposed the Jefferys prior - defined as the square root of the determinant of the information matrix - as an appropriate objective reference prior to counter a bias introduced by flat priors. The arguments for this prior are based upon information theoretic justifications and its invariance properities. Generalising from univariate analysis of unit roots to cointegrating systems, Kleibergen and van Dijk (1994) propose the Jeffreys prior for an additional, more practical reason. That is, as the square root of the determinant of the information matrix, the Jeffreys prior is zero at points of local non-identification and so excludes these points from the support under the posterior. As a result, Kleibergen and van Dijk were able to develop a valid Markov chain sampling scheme to allow estimation and this approach was successfully adapted to the triangular model by Martin & Martin (2000), Martin(2000), and Martin (2001).

Here we develop the Jeffreys prior for the short run dynamics and a prior that retains many of the attractive properties of the Jeffreys prior for the long run parameters. As $\varepsilon_t \sim N\,(0, \Sigma)$, we obtain the contribution to the log of the likelihood for one observation, l_t, as

$$l_t = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma| - \frac{1}{2}tr\Sigma^{-1}\varepsilon_t\varepsilon_t'.$$

The log likelihood for a sample of size T is then $l = \Sigma_{t=1}^{T} l_t$. The information matrix, $\mathfrak{I}$, is defined as the minus the expectation of the Hessian of[5] l_t.

A problem arises in deriving the full Jeffreys prior when we attempt to take expectations of the second partial derivatives with respect to the space of β, $\mathfrak{p}$. As these partial derivatives perturb the from space of β, the expectations of terms such as $\frac{\partial^2 l_t}{\partial(vec\mathfrak{p})\partial(vec\mathfrak{p})'}$ will involve expectations of nonstationary variables with respect to their unconditional distributions. As these distributions do not in all cases exist or are not defined (as they are nonstationary), this component of the information matrix is not generally defined. To give a feel for the problem, the unconditional expectation $E(\beta' x_t)$ will be well defined as $\beta' x_t$ has a stationary distribution. However, the derivative with respect to the elements of β, $\frac{\partial\beta}{\partial\mathfrak{p}}$, will span a different space to β, and so $E\left(\frac{\partial\beta}{\partial\mathfrak{p}} x_t\right)$ will not be generally[6] defined.

Possible solutions to this problem are to condition upon the intial values or to take $E(x_t' x_t) = x_t' x_t$. In this paper, we use only the partial Jeffreys priors for Π and Σ and specify an alternative prior for $\mathfrak{p}$ which captures the attractive properties of the Jeffreys prior. Therefore the relevant block of the information matrix (for Π and Σ) is

$$\begin{aligned}
\mathfrak{I} &= \begin{pmatrix} -E\left(\frac{\partial^2 l_t}{\partial(vec\Pi)\partial(vec\Pi)'}\right) & 0 \\ 0 & -E\left(\frac{\partial^2 l_t}{\partial(vech\Sigma)\partial(vech\Sigma)'}\right) \end{pmatrix} \\
&= \begin{pmatrix} \Sigma^{-1} \otimes E(\omega_t'\omega_t) & 0 \\ 0 & \frac{1}{2} D_n' \left(\Sigma^{-1} \otimes \Sigma^{-1}\right) D_n \end{pmatrix}
\end{aligned}$$

As ω_t is a stationary, zero mean, VAR process, the expectation $\Gamma = E(\omega_t'\omega_t)$ is well defined. The Jeffreys prior for Π and Σ is then

$$|\mathfrak{I}|^{1/2} = |\Gamma|^{\frac{n}{2}} |\Sigma|^{-\frac{n(l+1)+1}{2}} 2^{-\frac{n}{2}}.$$

The $nl \times nl$ matrix Γ is a symmetric Toeplitz matrix made up of $n \times n$ blocks $\Gamma_i = E\left(w_t w_{t-i}'\right)$, $i = 0, \ldots, l-1$. For given values of Π and Σ, we can readily

[5] It is common to specify the information matrix in terms of the Hessian for the full log likelihood l as $\mathfrak{F}_n = -E\left(\frac{\partial^2 l}{\partial(vec\Pi)\partial(vec\Pi)'}\right)$. However, for forming the Jeffreys prior in this case $\mathfrak{F}_n = n\mathfrak{F}$, and the two are equivalent because $|\mathfrak{F}_n| = |n\mathfrak{F}| \propto |\mathfrak{F}|$.

[6] Cases in which this expectation is well defined and finite may be found. For example, if x_t were a vector of matingales all with bounded support, the martingale convergence theorem implies x_t will have a convergent distribution with well defined moments. However result this is not true for all $I(1)$ processes.

solve for the Γ_i using the multivariate Yule-Walker equations. As a simple example, let $l = 2$ then

$$\Gamma = E\left(\omega_t' \omega_t\right) = \begin{bmatrix} \Gamma_0 & \Gamma_1 \\ \Gamma_1 & \Gamma_0 \end{bmatrix}$$

and from (4) we have the equations

$$\begin{aligned} \Gamma_0 - \Pi_1\Gamma_1 - \Pi_2\Gamma_2 &= \Sigma, \\ \Gamma_1 - \Pi_1\Gamma_0 - \Pi_2\Gamma_1 &= 0, \\ \Gamma_2 - \Pi_1\Gamma_1 - \Pi_2\Gamma_0 &= 0. \end{aligned}$$

Vectorising the above equations and using the definitions[7] $\gamma_0 = vech\left(\Gamma_0\right)$, $\gamma_1 = vec\left(\Gamma_1\right)$ and $\gamma_2 = vec\left(\Gamma_2\right)$, we have the solution defined by

$$\begin{bmatrix} I_{n(n+1)/2} & -D_n^+\left(I_n \otimes \Pi_1\right) & -D_n^+\left(I_n \otimes \Pi_2\right) \\ -\left(I_n \otimes \Pi_1\right) D_n & I_{n^2} - \left(I_n \otimes \Pi_1\right) & 0 \\ -\left(I_n \otimes \Pi_2\right) D_n & -\left(I_n \otimes \Pi_1\right) & I_{n^2} \end{bmatrix} \begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \end{bmatrix} = \begin{bmatrix} vech\left(\Sigma\right) \\ 0 \\ 0 \end{bmatrix}.$$

Of course straightforward substitution for Γ_2 with $\Pi_1\Gamma_1 + \Pi_2\Gamma_0$ reduces the number of equations to be solved by n^2.

In specifying the prior for β and η, we begin with the principle that the object of interest in cointegration analysis is the cointegrating space, $\mathfrak{p}$ (following from arguments in Strachan and Inder 2004, Strachan and van Dijk 2004, and Villani 2005) which is an element of the Grassman manifold. The same arguments hold for η as for β, so we will only discuss β. As demonstrated in Strachan and Inder (2004), specifying β as semiorthogonal such that it is an element of the Stiefel manifold, and specifying a Uniform distribution on the Stiefel manifold implies a Uniform distribution on the Grassman manifold. As with the Jeffreys prior, this prior for $\mathfrak{p}$ is invariant to rescaling of the data or transformations of β.

To demonstrate this last point, consider two types of transformations. The first is $\beta^* = \beta\kappa$ and the second $\beta^+ = P\beta$ where κ and P are full rank. The first transformation (β^*) implies a rotation of the vector β within the cointegrating space $\mathfrak{p} = space\left(\beta\right) = space\left(\beta^*\right)$ such that the space is not affected. This transformation might arise if we wished to change the way we

[7] The matrix D_n is a section matrix that solves, for example, $vec\left(\Sigma\right) = D_n vech\left(\Sigma\right)$, and D_n^+ is the Moore-Penrose inverse of D_n. such that $D_n^+ D_n = I_{n(n+1)/2}$. See Magnus and Neudecker (1988) for further definitions and explanations of these terms.

identify the elements of β to enable estimation. For example, we might wish to employ linear identifying restrictions such that $\beta^* = [I_r \quad B']'$. In this case the Uniform prior for $\mathfrak{p}$ would be induced by a Cauchy prior for B and this Cauchy form results directly from the Jacobian for the transformation $\beta \to B$. Provided κ is of full rank, the posterior for $\mathfrak{p}$ will be invariant to such a transformation.

The second transformation (β^+) implies a rotation *of* the cointegrating space $\mathfrak{p}$. Mapping from β^+ to the Stiefel manifold by $\widetilde{\beta} = \beta^+ \left(\beta^{+\prime}\beta^+\right)^{-\frac{1}{2}}$ again implies a Uniform prior for $\widetilde{\beta}$ will induce a Uniform prior for the cointegrating space. Thus the appropriate ignorance prior remains the same – uniform on the Grassman manifold. In this case, however, the posterior distribution for the cointegrating space will not be invariant to this transformation. A space previously located at $\mathfrak{p} = space\,(\beta)$ will now be located at $\mathfrak{p}^+ = space\left(\widetilde{\beta}\right) = space\left(\beta^+\right) \neq \mathfrak{p} = space\,(\beta)$. Of course this result will hold no matter what identification method is used and is, in fact, desirable. The result is desirable since the cointegrating space must rotate to accommodate the new linear combinations that are stationary by $x_t\beta = x_tP^{-1}P\beta = z_t\beta^+$. Consider, for example, if $s_t = s_{t-1} + (WhiteNoise)$ is a stochastic trend, $v_t = as_t + (WhiteNoise)$ and $y_t = bs_t + (WhiteNoise)$ such that v_t and y_t are $I\,(1)$, then $\frac{b}{a}v_t - y_t \sim I(0)$. If however we have data on $z_t = 100y_t = 100bs_t + (WhiteNoise)$, which implies

$$\begin{aligned}
x_t\beta &= \frac{b}{a}v_t - y_t = \begin{bmatrix} v_t & y_t \end{bmatrix} \begin{bmatrix} b/a \\ -1 \end{bmatrix} = x_tP^{-1}P\beta \\
&= \begin{bmatrix} v_t & y_t \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 100 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & .01 \end{bmatrix} \begin{bmatrix} b/a \\ -1 \end{bmatrix} \\
&= \begin{bmatrix} v_t & z_t \end{bmatrix} \begin{bmatrix} b/a \\ -.01 \end{bmatrix}
\end{aligned}$$

then $\frac{b}{a}v_t - .01z_t \sim I(0)$ and the cointegrating space has rotated as required.

4 Application

In this section we use the above models to obtain inference upon the integration and cointegration properties of Australian M1 measure of money, prices and income to investigate support for a stable money demand relation. We have quarterly observations in logs of the measure of M1, m_t, the price level, p_t, such that $m_t - p_t$ measures real money, and real gross national income, y_t. We collect these variables into the vector $x_t = (m_t, p_t, y_t)$. The sample is from September 1976 to December 2002 sourced from the web site of The Australian Bureau of Statistics, specifically tables D03, G09 and G02.

As we are interested in the existence of some stable relation among these variables, a common approach to this question (see for example Johansen, 1995 and Funke, Hall and Beeby, 1997) is to regard the money demand relation as the cointegrating relation among the variables

$$z_t = \beta_1 m_t + \beta_2 p_t + \beta_3 y_t + \mu_1 + \delta_1 t = x_t \beta.$$

This implies quite a relaxed form of stability as it allows for trends and possibly unit roots and only requires that this relation be integrated of a lower level than the original variables. For a more complete investigation of this type of problem see Fiess and MacDonald (2001). A further question of interest is whether this stable relation is the velocity of money. From the money demand relation above we can see that stability of the (negative log) velocity of money, $\nu = m_t - p_t - y_t$, is implied by the identifying restrictions $\beta_1 = -\beta_2 = -\beta_3$ or $H = (1, -1, -1)'$ in the notation of Section 2.The data are plotted in Figure (1) and from this plot we can see there is little evidence of quadratic trends, but certainly there may be linear trends. We conduct our analysis using the shrinkage and Jeffreys priors described earlier for the short-run dynamics. Using Theorem 3 of Ni and Sun (2003), with $T = 101$, $n = 3$ and $l = 2$ in (4), we can show the posterior with the shrinkage prior will be proper. We estimate the marginal likelihoods using independent importance sampling. The details of this computation are given in Appendix III.

Using the Jeffreys prior we find a posterior probability of one for Process I implying that the variables are $I(2)$ but cointegrate to form a single $I(0)$ relationship with a nonzero mean. As there is one cointegrating relation, this suggests that there exists an identifiable money demand relation. However, the evidence is against the stability of the money demand relation.

The models with more than 1% posterior probabilities are reported for the shrinkage prior in Table 1 below, along with these probabilities.

Table 1: Posterior probabilities using the shrinkage prior of models. Probabilities are expressed as percentages.

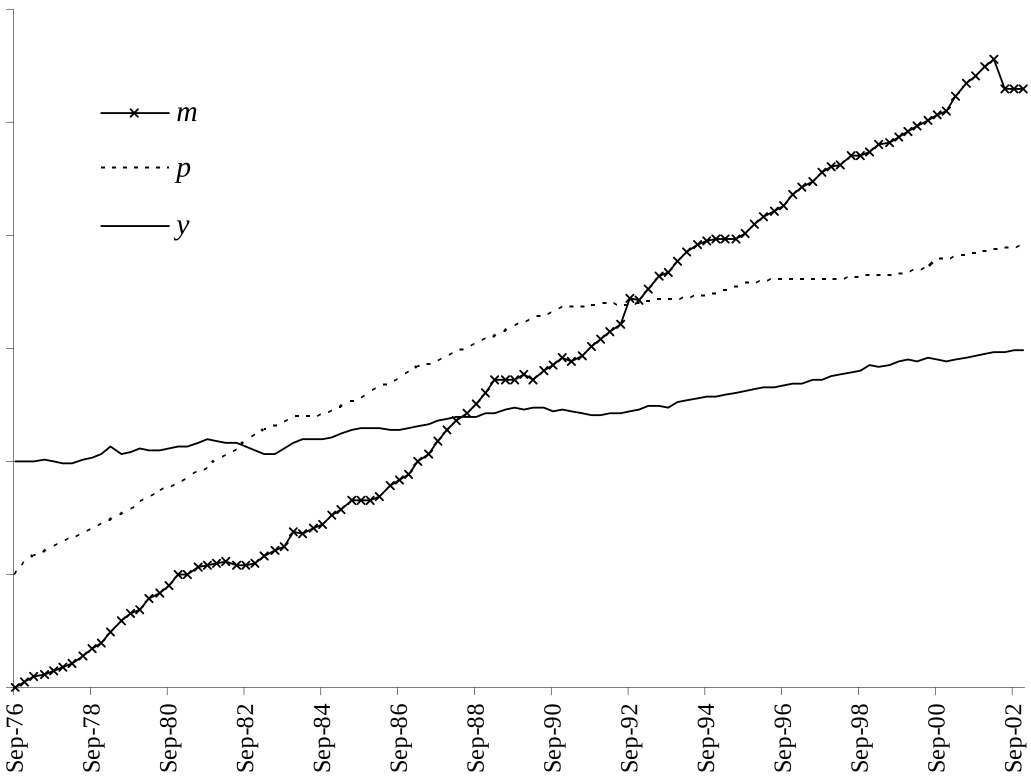

Figure 1: Plot of quarterly log M1, m, price level, p, and real gross national income, y. The sample is from September 1976 to December 2002.

Model Description				
Restricted (R)	Process	Deterministic	β Rank	$P(M\|y)$
R	II	$E(\beta x_t) = 0$, No trend in x_t	1	41.1
R	II	$E(\beta x_t) = \mu$, trend in x_t	1	23.9
R	0	$E(\beta x_t) = \mu$, No trend in x_t	1	15.9
R	0	$E(\beta x_t) = \mu$, trend in x_t	1	6.7
R	0	$E(\beta x_t) = 0$, No trend in x_t	1	3.6
R	0	$E(\beta x_t) = \mu + \delta t$, trend in x_t	1	1.9
U	II	$E(\beta x_t) = 0$, No trend in x_t	1	1.5

R in the first column denotes the restriction $sp(\beta) \subseteq sp(H)$.

The results above agree with those from the Jeffreys prior in that with posterior probability $P(r = 1|y) = 98.9\%$, there is strong evidence of a single stable relationship among money, prices and income. Conditional upon there being one cointegrating relation, the posterior probability that this relationship is the velocity of money is $P(sp(\beta) \subseteq sp(H)|r = 1, y) = 96.8\%$ (the marginal probability of this restriction is $P(sp(\beta) \subseteq sp(H)|y) = 95.9\%$). These two results provide strong evidence that the velocity of money is more stable than its components: money supply; prices; and income. However, this does not suggest that the velocity of money is necessarily stable in the sense of being an $I(0)$ process. The probability of this event, which is the joint probability of Process 0 and $r = 1$, is a moderately low 28%.

On the general statistical properties of x_t, the posterior probability that x_t is $I(d)$ and $\beta' x_t$ is $I(d-1)$ (that x_t is $CI(d, 1)$ in the notation of Engle and Granger (1987)) is quite strong with posterior probability of 96.6%. The evidence suggests it is most likely, with 67.7% probability, that this relationship occurs as the vector x_t is $I(2)$ and cointegrates as $CI(2, 1)$ (Process II). There is, however, a significant probability of 28.9% that x_t is $I(1)$ and cointegrates as $CI(1, 1)$ (Process 0). The evidence on which deterministic processes are evident is not clear. There is moderate evidence (52.1%) that there some deterministic process is present, and given this is the case, the most likely of these is that there is a linear trend in x_t with conditional probability of 61.5%. The conditional probability of a quadratic trend is only 4.4%.

5 Conclusion

In this paper we have generalised the triangular model of Phillips (1991) to permit a wide range of cointegrating processes to be investigated within a simple general specification. Importantly the specification in this paper incorporates a wide range of $I(2)$ processes some of which do not appear to have been previously considered. The specification of the cointegrating relations as variables of interest could lead to other applications outside those considered in this paper. We conclude with a brief discussion of one possible new direction.

The specification in (4) provides a simple means of investigating the responses to shocks to the stationary cointegrating relations and the stochastic trends. Recent work in applied economic analysis, particularly in macroeconometrics (see for example, Lettau and Ludvigson 2004, Koop, Potter and Strachan 2004 or Strachan and van Dijk 2004), has made use of decompositions of the variance of processes into permanent and temporary components in an effort to better understand the behaviour and relationships among variables. There are several valid ways to decompose a series into temporary and permanent components to permit the computation of variance decompositions. On the principle that the limit of the effect of a temporary shock will be zero whereas for a permanent shock will be nonzero, Gonzalos and Ng (2001) treat $\beta' x_t$ as the temporary component of x_t and $\alpha'_\perp x_t$ as the permanent component since premultiplication by $\alpha'_\perp$ eliminates the stationary relation $\alpha\beta' x_t$in (5). This specification implies the temporary and permanent shocks may be represented as $\beta'\nu_t$ and $\alpha'_\perp\nu_t$ respectively. Centoni and Cubadda (2002) use the Wold decomposition to demonstrate that $\alpha'\Sigma^{-1}\nu_t$ and $\alpha'_\perp\nu_t$ may also be regarded as temporary and permanent shocks.

Recall the decomposition $x_t = \beta\beta' x_t + \beta_\perp\beta'_\perp x_t$. Since premultiplication by β eliminates the common stochastic trends, βx_t may be regarded as the temporary component and $\beta'_\perp x_t$ can be regarded as the common trends (Johansen 1995, Corollary 4.4) or the permanent component. This implies that the temporary shocks will be $\varepsilon_{1,t} = \beta'\nu_t$, and $\varepsilon_{2,t} = \beta'_\perp\nu_t$ will be the permanent shocks. Thus the standard variance decomposition for a stationary VAR in w_t will give the proportions due to temporary and permanent shocks as those due to $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ respectively.

Similarly we could decompose the variances of $I(2)$ processes into components attributable to $I(0)$, $I(1)$ and $I(2)$ shocks. This depth of analysis does not yet appear to have been conducted.

6 References

Bartlett, M, S. (1957). A Comment on D. V. Lindley's Statistical Paradox. *Biometrica* 44:533–534.

Bauwens, L., Lubrano, M. and Richard, J.-F. (1999) *Bayesian Inference in Dynamic Econometric Models*. Oxford: Oxford University Press.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis* (2nd ed.).New York: Springer-Verlag.

Bewley, R. (1979). The Direct Estimation of the Equilibrium Response in a Linear Model. *Economic Letters* 3:357-361.

Boswijk, H.P. (1996). Testing Identifiability of Cointegrating Vectors. *Journal of Business and Economic Statistics* 14:153-160.

Campbell, J. Y. (1987). Does Savings Anticipate Labour Income? An Alternative Test of the Permanent Income Hypothesis. *Econometrica* 55:1249-1274.

Campbell J. Y. and R. J. Shiller (1987). Cointegration and tests of present value models. *The Journal of Political Economy* 95:1062-1088.

Campbell J. Y. and R. J. Shiller (1989). The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors. *The Review of Financial Studies* 1, 3:195-228.

Centoni, M., Cubadda, G. (2003). Measuring the business cycle effects of permanent and transitory shocks in cointegrated time series. *Economics Letters* 80:45-51.

Chikuse, Y. (1990). The matrix angular central Gaussian distribution, *Journal of Multivariate Analysis* 33, 265-274.

Engle, R.F. and C.W.J. Granger (1987). Co-Integration and Error Correction: Representation, Estimation and Testing. *Econometrica* 55:251-276.

Fernández, C., Ley, E. Steel, M. (2001). Model uncertainty in cross-country growth regressions. *Journal of Applied Econometrics* 16:563-576.

Fiess, N., MacDonald R. (2001). The Instability of the Money Demand Function: an $I(2)$ Interpretation. *Oxford Bulletin of Economics and Statistics* 63:475-495.

Galí, J. (1990). Finite Horizons, Life-Cycle Savings, and Time-Series Evidence on Consumption. *Journal of Monetary Economics* 26:433-452.

Garratt, A., Lee, K., Pesaran, M. H., Shin Y. (2003). A Long Run Structural Macroeconomic Model of the UK. *Economic Journal* 113 (April):412–455.

Georgoutsos, D. A., Kouretas, G. P. (2004). A Multivariate $I(2)$ Cointegration Analysis of German Hyperinflation. *Applied Financial Economics* 14:29-41.

Geweke, J. (1996). Bayesian reduced rank regression in econometrics. *Journal of Econometrics* 75:121-146.

Gonzalo, J., Ng, S. (2001). A systematic framework for analyzing the dynamic effects of permanent and transitory shocks. *Journal of Economic Dynamics and Control*, 25, 1527-1546.

Granger, C.W.J. (1983). Cointegrated Variables and Error Correction Models. Unpublished USCD Discussion Paper 83-13.

Haldrup N., Salmon M. (1998). Representations of $I(2)$ Cointegrated Systems Using the Smith-McMillan Form. *Journal of Econometrics* 84:303-325.

Johansen, S. (1990). Juselius, K. Maximum Likelihood Estimation and Inference on Cointegration: with Applications to the Demand for Money. *Oxford Bulletin of Economics and Statistics* 52:169-210.

Johansen, S. (1995). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models.* New York: Oxford University Press.

Judge G. G., Griffiths, W.E., Hill, R.C., Lutkepohl, H. and Lee, T. (1985). *The Theory and Practice of Econometrics* (2nd ed.). New York: Wiley.

Kass, R. and Raftery, A. (1995). Bayes Factors. *Journal of the American Statistical Association* 90:773-795.

King, R.G., Plosser, C.I., Stock, J.H., Watson M.W. (1991). Stochastic Trends and Economic Fluctuations. *The American Economic Review* 81:819-840.

Kleibergen, F.,van Dijk, H.K. (1994). On the Shape of the Likelihood/Posterior in Cointegration Models. *Econometric Theory* 10:514-551.

Kongsted, H. C. (2003). An $I(2)$ Cointegration Analysis of Small-Country Import Price Determination. *Econometrics Journal* 6:53-71.

Koop G., Potter, S., Strachan, R. (2004) Re-examining the Consumption-Wealth Relationship: The Role of Model Uncertainty. University of Leicester Working Paper XXXX.

Leonard, T. and Hsu, J. S. J. (2001). *Bayesian Methods.* Cambridge: Cambridge University Press.

Lettau, M.; Ludvigson, S. (2004). Understanding Trend and Cycle in Asset Values: Reevaluating the Wealth Effect on Consumption. *American Economic Review* 94:276-299.

Lindley, D.V. (1962). Discussion on Professor Stein's paper. *Journal of the Royal Statistical Society Series B* 24:285-287.

Lindley, D.V., Smith, A.F.M. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society Series B* 34, 1-41.

Lindley D. V. (1997). Discussion forum: Some comments on Bayes factors. *Journal of Statistical Planning and Inference* 61:181-189.

Luukkonen, R., Ripatti, A., Saikkonen, P. (1999). Testing for a Valid Normalisation of Cointegrating Vectors in Vector Autoregressive Processes. *Journal of Business and Economic Statistics* 17:195-204.

Martin, G. M. (2000). US Deficit Sustainability: a New Approach Based on Multiple Endogenous Breaks. *Journal of Applied Econometrics* 15:83-105.

Martin, G. M. (2001) Bayesian Analysis of a Fractional Cointegration Model. *Econometric Reviews* 20:217-234.

Martin, G. M., Martin V. L. (2000). Bayesian Inference in the Triangular Cointegration Model Using a Jeffreys Prior. *Communications in Statistics, Theory and Methods* 29:1759-1785.

Mittelhammer, R.C., Judge, G.G., and Miller, D.J. (2000). *Econometric Foundations.* Cambridge: Cambridge University Press.

Ni, S. X. and D. Sun (2003). Noninformative Priors and Frequentist Risks of Bayesian Estimators of Vector-Autoregressive Models. *Journal of Econometrics* 115:159-197.

Nielsen, H. B. (2002). An $I(2)$ Cointegration Analysis of Price and Quantity Formation in Danish Manufactured Exports. *Oxford Bulletin of Economics and Statistics* 64:449-472.

Paruolo, P. (1996). On the Determination of Integration Indices in $I(2)$ Systems. *Journal of Econometrics* 72:313-356.

Phillips, P. C. B. (1991). Optimal Inference in Cointegrated Systems. *Econometrica* 59:283-306.

Phillips, P. C. B. (1994). Some Exact Distribution Theory for Maximum Likelihood Estimators of Cointegrating Coefficients in Error Correction Models. *Econometrica* 62:73-93.

Rahbek, A., Kongsted, H. C., Jørgensen, C. (1999). Trend Stationarity in the $I(2)$ Cointegration Model. *Journal of Econometrics* 90:265–289.

Sclove, S. L. (1968). Improved Estimators for Coefficients in Linear Regression, *Journal of the American Statistical Association* 63:596-606.

Sclove, S.L. (1971). Improved Estimation of Parameters in Multivariate Regression, *Sankhya, Series A* 33:61-66.

Sims, C. A., Uhlig H. (1991). Understanding Unit Rooters: A Helicopter Tour. *Econometrica* 59:1591-1600.

Stein, C. (1956). Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution. In Proceedings of the *Third Berkeley Symposium on Mathematical Statistics and Probability.* Vol. 1 Berkeley, CA: University of California Press:197-206.

Stein, C. (1960). Multiple Regression, in I. Olkin (ed.), *Contributions to Probability and Statistics in Honor of Harold Hotelling.* Stanford: Stanford University Press.

Stein, C. (1962). Confidence Sets for the Mean of a Multivariate Normal Distribution. *Journal of the Royal Statistical Society, Series B* 24:265-296.

Stock, J. H., WatsonM. W. (1993). A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems. *Econometrica* 61:783-820.

Strachan, R. (2003). Valid Bayesian Estimation of the Cointegrating Error Correction Model. *Journal of Business and Economic Statistics* 21:185-195.

Strachan, R. W., Inder B. (2004). Bayesian Analysis of The Error Correction Model. *Journal of Econometrics* 123:307-325.

Strachan, R., van Dijk H. K. (2003). Bayesian Model Selection for a Sharp Null and a Diffuse Alternative with Econometric Applications. *Oxford Bulletin of Economics and Statistics* 65:681-688.

Strachan, R., van Dijk H. K. (2004). Improper Priors with Well Defined Bayes Factors. *Econometric Institute Report* EI 2004-18, Erasmus University.

Villani, M. (2005). Bayesian reference analysis of cointegration. *Econometric Theory* 21:326-357.

Zellner, A. (2002). 'Bayesian shrinkage estimates and forecasts of individual and total or aggregate outcomes' mimeo University of Chicago.

Zellner, A. and Vandaele, W.A. (1974). Bayes-Stein Estimators for k-means, Regression and Simultaneous Equation Models, in Fienberg, S.E. and Zellner, A., (eds.), *Studies in 21 Bayesian Econometrics and Statistics in Honor of Leonard J. Savage.* Amsterdam: North-Holland:627-653.

7 Appendix I

In this appendix we transform from w_t to the implied process generating x_t to show how the stochastic trends enter x_t and how premultiplication by β' removes those trends.

Under the assumption given in the paper that $C \neq 0$, the process

$$w_t = C(L)\,\varepsilon_t$$

is $I(0)$. Premultiply by $[\beta \quad \beta_\perp]$ where $\beta'\beta = I_r$ and $\beta_\perp'\beta_\perp = I_{n-r}$,

$$\begin{aligned} [\beta \quad \beta_\perp]\, w_t &= \beta\beta' x_t + \beta_\perp \beta_\perp' \Delta x_t = x_t - \beta_\perp \beta_\perp' x_{t-1} \\ &= [\beta \quad \beta_\perp]\, C(L)\,\varepsilon_t. \end{aligned}$$

By repeatedly substituting for the lagged value of x_t we obtain

$$x_t = \beta_\perp \beta_\perp' x_0 + [0 \quad \beta_\perp] \sum_{i=1}^{T-1} C(L)\,\varepsilon_{t-i} + [\beta \quad \beta_\perp]\, C(L)\,\varepsilon_t.$$

Next we make use of the well known decomposition $C(L) = C + C^*(L)(1-L)$ to show how the stochastic trend accumulates in x_t by

$$x_t = \beta_\perp \beta_\perp' x_0 + [0 \quad \beta_\perp]\, C \sum_{i=1}^{T-1} \varepsilon_{t-i} + [0 \quad \beta_\perp] \sum_{i=1}^{T-1} C^*(L)\,\Delta\varepsilon_{t-i} + [\beta \quad \beta_\perp]\, C(L)\,\varepsilon_t.$$

Clearly then, premultiplication be β' will annihilate the stochastic trend as

$$\begin{aligned}
\beta' x_t &= [I_r \quad 0]\, C(L)\, \varepsilon_t \text{ and} \\
\beta'_\perp x_t &= \beta'_\perp x_0 + [0 \quad I_{n-r}]\, C \sum_{i=1}^{T-1} \varepsilon_{t-i} + [0 \quad I_{n-r}] \sum_{i=1}^{T-1} C^*(L)\, \Delta\varepsilon_{t-i} + [0 \quad I_{n-r}]\, C(L)\, \varepsilon_t.
\end{aligned}$$

8 Appendix II

We provide the forms when we transform from a VAR in w_t to a VAR in x_t with the implied VECM for x_t. First, we provide the form of the transformation of the $VAR(l)$ in w_t to the $VAR(l+1)$ in x_t and VECM in x_t. We begin from (4) but add the decomposition

$$\begin{bmatrix} w_{1,t} \\ w_{2,t} \end{bmatrix} = \sum_{i=1}^{l} \begin{bmatrix} \Pi_{11,i} & \Pi_{12,i} \\ \Pi_{21,i} & \Pi_{22,i} \end{bmatrix} \begin{bmatrix} w_{1,t-i} \\ w_{2,t-i} \end{bmatrix} + \varepsilon_t.$$

Premultiplying by $B = [\beta \quad \beta_\perp]$ and rearrange, we obtain the $VAR(l+1)$ in x_t

$$\begin{aligned}
x_t &= [\beta \quad \beta_\perp] \begin{bmatrix} \Pi_{11,i} & \Pi_{12,1} \\ \Pi_{21,1} & \Pi_{22,1} + I_{n-r} \end{bmatrix} \begin{bmatrix} \beta' \\ \beta'_\perp \end{bmatrix} x_{t-1} + \\
&\quad \sum_{i=2}^{l} [\beta \quad \beta_\perp] \begin{bmatrix} \Pi_{11,i} & \Pi_{12,i} - \Pi_{12,i-1} \\ \Pi_{21,i} & \Pi_{22,i} - \Pi_{22,i-1} \end{bmatrix} \begin{bmatrix} \beta' \\ \beta'_\perp \end{bmatrix} x_{t-i} \\
&\quad + [\beta \quad \beta_\perp] \begin{bmatrix} 0 & -\Pi_{12,l} \\ 0 & -\Pi_{22,l} \end{bmatrix} \begin{bmatrix} \beta' \\ \beta'_\perp \end{bmatrix} x_{t-l-1} + \varepsilon_t.
\end{aligned}$$

The usual manipulations result in the form for the VECM with l lags of differences

$$\Delta x_t = \alpha\beta' x_{t-1} + \sum_{i=1}^{l} \Gamma_i \Delta x_{t-i} + \nu_t$$

where, using $\beta\beta' = I_n - \beta_\perp\beta'_\perp$,

$$\begin{aligned}
\alpha &= \beta\left(I_r + \Sigma_{i=1}^{l}\Pi_{11,i}\right) + \beta_\perp \Sigma_{i=1}^{l}\Pi_{21,i}, \\
\Gamma_i &= \beta\Pi_{12,i-1}\beta'_\perp + \beta_\perp\Pi_{22,i-1}\beta'_\perp - \beta\Pi_{11,i}\beta' - \beta_\perp\Pi_{21,i}\beta' \\
\text{for } i &= 1, ..., l-1 \\
\Gamma_l &= [0 \quad \beta\Pi_{12,i-1}\beta' + \beta_\perp\Pi_{22,i-1}\beta'_\perp]
\end{aligned}$$

Next, if we transform from the equation VECM in x_t in (5) to the VAR in w_t in (4) we obtain

$$\begin{aligned}
\Pi_1 &= \begin{bmatrix} I_r + \beta'\alpha + \beta'\Gamma_1\beta & \beta'\Gamma_1\beta_\perp \\ \beta'_\perp\left[\alpha + \Gamma_1\beta\right] & \beta'_\perp\Gamma_1\beta_\perp \end{bmatrix} \\
\Pi_i &= \begin{bmatrix} \beta'\left[\Gamma_i - \Gamma_{i-1}\right]\beta & \beta'\Gamma_{i-1}\beta_\perp \\ \beta'_\perp\left[\Gamma_i - \Gamma_{i-1}\right]\beta & \beta'_\perp\Gamma_{i-1}\beta_\perp \end{bmatrix} \text{ for } i = 2, ..., l-1 \\
\Pi_l &= \begin{bmatrix} -\beta'\Gamma_{l-1}\beta & 0 \\ -\beta'_\perp\Gamma_{l-1}\beta & 0 \end{bmatrix} \text{ and } \varepsilon_t = \begin{bmatrix} \beta'\nu_t \\ \beta'_\perp\nu_t \end{bmatrix}.
\end{aligned}$$

9 Appendix III

We estimate the marginal likelihoods using independent importance sampling. We first draw the cointegrating vectors β and η, then draw the covariance matrix from an inverted Wishart distribution that depends upon β and η. Finally we draw Π from a Normal distribution that depends upon β, η and Σ. The details are as follows.

Draws of β and η are taken independently from a matrix angular central Gaussian distribution (MACG) (see Chikuse, 1990) located approximately at the posterior modal values. These modal values of β and η were determined from the burn-in sample of $33,333$ draws. During the burn-in period, the location of the MACG distributions were taken as the first r (in the case of β) and the first s (in the case of η) co-ordinate vectors. Denote the respective modes as $\widetilde{\beta}$ and $\widetilde{\eta}$, and the MACG distributions as $p\left(\beta|\widetilde{\beta}\right)$ and $p\left(\eta|\widetilde{\eta}\right)$. We demonstrate how to obtain a draw from the MACG $p\left(\beta|\widetilde{\beta}\right)$ and the process will be similar for $p\left(\eta|\widetilde{\eta}\right)$.

Draw X from the matrix Normal with zero mean and covariance matrix $\widetilde{\beta}\widetilde{\beta}' + \widetilde{\beta}_\perp\widetilde{\beta}'_\perp\tau$ where we set τ between zero and one. A value of τ close to zero will tend to produce draws very close to the space of $\widetilde{\beta}$. At $\tau = 1$, the MACG collapses to the Uniform distribution on the Steifel manifold. Next decompose the matrix X as $X = V\kappa$ where V is semiorthogonal and κ is lower triangular. Then take $\beta = V$ (and discard κ) as a draw from $p\left(\beta|\widetilde{\beta}\right)$.

Having obtained β and η, construct w_t and ω_t. The covariance matrix is then drawn from the inverted Wishart conditional upon β and η proportional to

$$|\Sigma|^{-(T-k)/2} \exp\left\{-\frac{1}{2}tr\Sigma^{-1}S\right\}$$

where k is the number of columns in Π, $S = \sum_{t=1}^{T}\left(w_t' - \omega_t\widehat{\Pi}\right)'\left(w_t' - \omega_t\widehat{\Pi}\right)$ and $\widehat{\Pi} = \left(\sum_{t=1}^{T}\omega_t'\omega_t\right)^{-1}\sum_{t=1}^{T}\omega_t'w_t'$. Finally the Π matrix is drawn from the Normal distribution, conditional upon Σ, β and η, with mean $\widehat{\Pi}$ and covariance matrix $\Sigma \otimes \left(\sum_{t=1}^{T}\omega_t'\omega_t\right)^{-1}$.

We now have the draws from

$$g = g(\beta, \eta, \Sigma, \Pi) = p\left(\beta|\widetilde{\beta}\right)p(\eta|\widetilde{\eta})\,p(\Sigma|\beta, \eta)\,p(\Pi|\beta, \eta, \Sigma).$$

Collect the free parameters in $(\beta, \eta, \Sigma, \Pi)$ into the vector θ and denote the product of the likelihood and the priors as $f = f(\beta, \eta, \Sigma, \Pi|y)$[8]. Now we can write the marginal likelihood as $\int f d\theta = m$. Note the relation

$$\int f d\theta = \int g\frac{f}{g}d\theta = m$$

implies that the expectation of the ratio $\frac{f}{g}$ is m. That is $E_g\left(\frac{f}{g}\right) = m$. Therefore if we take draws of $\theta^{(i)}$ for $i = 1, \ldots, J$ of θ from g and compute the average of the ratio $\frac{f^{(i)}}{g^{(i)}}$, then we have the approximation to the marginal likelihood as

$$\widehat{E}\left(\frac{f}{g}\right) = \frac{1}{J}\sum_{i=1}^{J}\frac{f^{(i)}}{g^{(i)}} = \widehat{m}.$$

The above specification of g will match the term in the posterior proportional to the likelihood so the term f involves only the terms in the priors. Note that we must know g completely, including the normalising constants. The Normal and inverted Wisharts distributions are well known, while expressions for the MACG are given in Chikuse (1990) and Strachan and Inder (2004).

A total of 100,000 draws were taken with the first third used as the burn-in and the remaining draws were used to estimate the marginal likelihoods.

[8] This expression excludes the normalising constants for the priors for Σ and Π as they will cancel in any Bayes factor calculation.