Bayesian Inference in a Cointegrating Panel Data Model[*]

Gary Koop
Department of Economics
University of Strathclyde
Gary.Koop@strath.ac.uk

Roberto Leon-Gonzalez
Department of Economics
University of Leicester
rlg19@le.ac.uk

and

Rodney Strachan
Department of Economics
University of Leicester
rws7@le.ac.uk

January 2006

Abstract: This paper develops methods of Bayesian inference in a cointegrating panel data model. This model involves each cross-sectional unit having a vector error correction representation. It is flexible in the sense that different cross-sectional units can have different cointegration ranks and cointegration spaces. Furthermore, the parameters which characterize short-run dynamics and deterministic components are allowed to vary over cross-sectional units. In addition to a noninformative prior, we introduce an informative prior which allows for information about the likely location of the cointegration space and about the degree of similarity in coefficients in different cross-sectional units. A collapsed Gibbs sampling algorithm is developed which allows for efficient posterior inference. Our methods are illustrated using real and artificial data.

Keywords: Bayesian, panel data cointegration, error correction model, reduced rank regression, Markov Chain Monte Carlo.

JEL Classification: C11, C32, C33

1   Introduction

The growing availability of panel data with large T dimension (i.e. where the number of time series observations is large) has stimulated a growth in research, both empirical and theoretical, which discusses time series issues in panel data models. Of particular interest are issues relating to nonstationarity and cointegration. In this paper, we develop a Bayesian approach to the analysis of cointegration in panels. We use a modelling framework which allows for great flexibility in the way heterogeneity across cross-sectional units is incorporated. In particular, we allow for both cointegrating vectors and ranks to vary over N. Our use of Bayesian methods allows for the cointegrating ranks to be treated as random variables. Thus, our methods can either be used to select a particular model with specified cointegrating ranks or to average across different cointegrating ranks. We also consider restricted models of interest (e.g. where all cross-sectional units have the same cointegrating rank). The use of Bayesian methods requires elicitation of a prior. We develop two priors, a noninformative and an informative one. The latter allows for the incorporation of prior beliefs that the same cointegrating relationship exists for all cross-sectional units. Furthermore, it allows for what we call "soft homogeneity" restrictions (i.e. that comparable parameters in different cross-sectional units are likely to be similar to one another). We derive efficient methods of posterior analysis in our class of models and illustrate our methods using artificial data and an application involving a monetary exchange rate model (see Groen, 2000 and Groen and Kleibergen, 2003).

The importance of this area of research is evidenced by the increasing tendency for researchers to employ panels of nonstationary processes in empirical studies in macroeconomics and international economics. For instance, the survey paper by Baltagi and Kai (2000) identifies many areas of application, including purchasing power parity (PPP), growth convergence and international R&D spillovers. To give one example which illustrates the issues which can be addressed through the use of panel data consider Jacobson, Lyhagen, Larsson and Nessén (2002). These authors use a multivariate panel cointegration model and demonstrate that, although strong purchasing power parity restrictions are rejected, the location of the cointegrating space is similar for all countries considered. This provides some evidence in support of PPP.

In terms of the frequentist econometric literature, there have been a range of methods proposed to obtain inference relating to cointegration in panel data models. Among many others, we note that residual-based, LM and likelihood based tests have been proposed by Kao (1999), McCoskey and Kao (1998), Pedroni (2004), Larsson, Lyhagen and Löthgren (2001) and Groen and Kleibergen (2003). The estimation methods used in these papers vary from OLS through maximum likelihood and generalized method of moments. The extent of this literature prevents us giving even a reasonable coverage here and so we refer the reader to the surveys by Phillips and Moon (2000) and Baltagi and Kao (2000).

While the bulk of the work to date has used frequentist methods, there have been a number of Bayesian approaches to obtain inference in dynamic panels. Pesaran, Hsiao and Tahmiscioglu (1999) present frequentist and Bayesian approaches to estimating the mean of the cross-sectional distribution for the coefficients in a dynamic panel data model. They show that the Bayesian approach performs

reasonably well in finite samples, and is even preferable to some consistent estimators when finite sampling performance is considered. Although they impose a stability condition, thus precluding discussion of issues relating to unit roots and cointegration, this assumption could be relaxed (see also Hsiao and Pesaran, 2004). Li (1999) investigates PPP by considering support for symmetry and proportionality restrictions in the PPP relationship. She allows for stationary AR(1) errors in the relationship between log exchange rates and prices. Interestingly, while this paper does not explicitly consider cointegration, with one small change the model of Li could - using a triangular setup as proposed by Phillips (1991) - be easily extended to allow investigation of whether or not cointegration between log exchange rates and prices occurs.

We are aware of only one paper explicitly proposing a Bayesian approach to estimation of a cointegrating system in panel data models. Carmeci (2005) presents a state space model which implies cointegration by directly modeling the common stochastic trends. Under the assumption that the cointegrating rank is known and assumed equal in every cross-sectional unit, the author develops Bayesian methods for estimation. We are not aware of any paper that presents a fully Bayesian method of inference on cointegration in panels, when the cointegrating rank is unknown and may differ across cross-sectional units. The present paper attempts to address this gap in the literature.

The remainder of the paper is organized as follows. Section 2 introduces the model and describes the elements of Bayesian analysis: likelihood, priors and methods of posterior simulation. Section 3 illustrates our methods using artificial data and Section 4 demonstrates the flexibility of inference in the application used in Groen and Kleibergen (2003), which involves an interesting set of restrictions implied by economic theory. Section 5 concludes.

2 The Models

In a standard time series framework, cointegration is typically investigated using a vector error correction model (VECM). To establish notation, to investigate cointegrating relationships involving an n-vector, y_t, we write the VECM for $t = 1, .., T$ as:

$$\Delta y_t = \Pi y_{t-1} + \sum_{h=1}^{l} \Gamma_h \Delta y_{t-h} + \Phi d_t + \varepsilon_t \tag{1}$$

where the $n \times n$ matrix $\Pi = \alpha\beta'$, where α and β are $n \times r$ full rank matrices and d_t denotes deterministic terms.[1] The value of r determines the number of cointegrating relationships. ε_t is a Normal mean zero error with positive definite covariance matrix.

Before extending (1) to the panel data case, it is important to digress briefly to motivate an important issue in Bayesian analysis of cointegrated models. The VECM suffers from both local and global identification problems. The local identification problem occurs since, if $\alpha = 0$, β does not appear in the likelihood function. The global identification problem can be seen by noting that

---

[1] The exact form of the deterministic terms is not crucial to our derivations so we leave these unspecified. See Johansen (1995), pages 81-84 for a commonly-used set of choices.

$\Pi = \alpha\beta'$ and $\Pi = \alpha GG^{-1}\beta'$ are identical for any nonsingular G. This indeterminacy is commonly surmounted by imposing the so-called *linear normalization* where $\beta = [I_r \quad B']'$. However, there are some serious drawbacks to this linear normalization (see Strachan and Inder, 2004 and Strachan and van Dijk, 2004b). Researchers in this field (see Strachan and Inder, 2004, Strachan and van Dijk, 2004b and Villani, 2005a,b) point out that it is only the cointegration space that is identified (not particular cointegrating vectors) and that, for most purposes (including prior elicitation), it is preferable to think in terms of the cointegration space. Accordingly, we introduce notation for the space spanned by β, $\mathfrak{p} = sp(\beta)$.

We can generalize (1) to the panel data case by including i subscripts to denote the cross-sectional unit which we refer to as the "individual" hereafter (where $i = 1,..,N$). That is, $y_{i,t}$ is an n vector of observations on the dependent variables for individual i at time t^2 and the panel VECM is written as:

$$\Delta y_{i,t} = \Pi_i y_{i,t-1} + \sum_{h=1}^{l_i} \Gamma_{i,h} \Delta y_{i,t-h} + \Phi_i d_{i,t} + \varepsilon_{i,t} \qquad (2)$$

where now $\Pi_i = \alpha_i \beta_i'$ where α_i and β_i are $n \times r_i$ full rank matrices. Our model allows for the number of cointegrating relationships to vary across individuals and thus, we extend our previous notation such that the cointegration spaces are now $\mathfrak{p}_i = sp(\beta_i)$. The covariance matrices for vectors $\varepsilon_{i,t}$ are assumed to be

$$E\left(\varepsilon_{i,t}\varepsilon_{j,s}'\right) = \left\{ \begin{array}{ll} \Sigma_{ij} & \text{for } t = s \\ 0 & \text{for } t \neq s \end{array} \right. . \qquad (3)$$

In other words, we are assuming the errors to be uncorrelated over time, but correlated across equations for a given individual and correlated across individuals. Note that the last assumption differs from much of the previous literature. For instance, Larsson, Lyhagen and Löthgren (2001) use a more restrictive model assuming $E\left(\varepsilon_{i,t}\varepsilon_{j,s}'\right) = 0$ if $i \neq j$ for all t and s. Although allowing for a correlation between errors for different individuals is not usually done with microeconomic survey data, with macroeconomic panels where the "individuals" are countries such a correlation is potentially important.We are therefore following the more general model of Groen and Kleibergen (2003) which does allow for such a correlation. Note also that our model is more flexible than the one of Groen and Kleibergen (2003) in that we relax the assumption of a common cointegrating rank.

There are many features of (2) that the researcher might be interested in. For each individual, we would naturally be interested in the dimension of the cointegrating space, r_i, and whether $r_i = r$ for all i. Other questions of interest relate to the cointegrating spaces, $\mathfrak{p}_i = sp(\beta_i)$. A restricted version of (2) would have the same cointegrating relationships (i.e. the same r_i and β_i) for every individual and, thus, $\mathfrak{p}_i = \mathfrak{p}$. Alternatively, if different individuals have different numbers of cointegrating vectors, then we might be interested in whether all of the cointegrating spaces lie within some more general one. That is, if $r_i \leq r$ for $i = 1,..,N$ and $\mathfrak{p}$ is a cointegration space with dimension r, then we might be interested in investigating whether $\mathfrak{p}_i \subseteq \mathfrak{p}$ for $i = 1,..,N$.

---

[2]It is not complicated to allow for $y_{i,t}$ to be of dimension n_i, but we will assume $n_i = n$, for simplicity. Similarly it is straightforward to extend our results to the unbalanced panel case.

3

As a simple illustration of how these questions might arise, consider the balanced growth hypothesis in the real business cycle model presented by King, Plosser, Stock and Watson (1991). Assume $y_{i,t} = (c_{i,t}, a_{i,t}, g_{i,t})'$ where $c_{i,t}$ is log consumption for country i, $a_{i,t}$ is log investment for that country and $g_{i,t}$ is log income. If the elements of the vector $y_{i,t}$ are $I(1)$ and are cointegrated then $0 < r_i < 3$. If there are two cointegrating relationships ($r_i = 2$) and the logs of the great ratios of consumption to income and investment to income are stable such that $c_{i,t} - g_{i,t}$ and $a_{i,t} - g_{i,t}$ are $I(0)$ in every country, then the cointegrating space, $\mathfrak{p}_i$, is

$$\mathfrak{p}_i = \mathfrak{p} = sp \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix}.$$

In an empirical analysis using panel data, it would be of interest to investigate this restriction (i.e. whether two cointegrating vectors exist for each country and whether their values are consistent with the great ratios). However, it is possible that some countries might only have one cointegrating relationship, so that $r_i = 2$ for some countries and $r_i = 1$ for others. In this case, investigating whether $\mathfrak{p}_i = \mathfrak{p}$ for all $i = 1, .., N$ is not reasonable. Instead, the researcher may be interested in investigating whether the cointegrating relationships either involve the great ratios individually (for $r_i = 2$) or involve a linear combination of the (logs of) the great ratios. In terms of our notation, this involves investigating whether $\mathfrak{p}_i \subseteq \mathfrak{p}$ for $i = 1, .., N$.

In most empirical applications, the cointegrating spaces will be of most interest and, hence, the researcher will be most interested in a set of models defined by restrictions on these. However, it is also common for the set of models to be broadened by considering different forms of the deterministic processes, $d_{i,t}$, and the number of lags l_i and it might be desirable to allow these to vary across individuals. Thus, in empirical work, the researcher might want to consider a very wide range of models indeed. However, in order to focus on the central issues relating to cointegration, we will assume a common lag length for all individuals (i.e. $l_i = l$ for all i) and common deterministic process (i.e. $d_{i,t} = d_t$ for all i) and develop methods of inference for r_i and $\mathfrak{p}$.

2.1 The Likelihood Function

In this section we show two representations of the likelihood, involving different parameterizations, which we draw on in our discussion of posterior simulation. Note that the matrix of long run multipliers can be written as:

$$\beta_i \alpha_i' = [\beta_i \kappa_i] \left[\alpha_i \kappa_i^{-1} \right]' \equiv \beta_i^* A_i' \tag{4}$$

where β_i is restricted to be semi-orthogonal (for reasons described in the next section) and κ_i is positive definite and defined so that A_i is semi-orthogonal. Here we have used $\beta_i^* = \beta_i \kappa_i$ and $\alpha_i = A_i \kappa_i'$. There are many choices for κ_i which satisfy these properties, but a convenient one we will use here is:

$$\kappa_i = \left(\alpha_i' \alpha_i \right)^{\frac{1}{2}} = \left(\beta_i^{*\prime} \beta_i^* \right)^{\frac{1}{2}}. \tag{5}$$

For reasons explained below, our posterior simulator will involve switching between the parameterizations in (4).

To establish notation, we collect the $n \times n$ blocks Σ_{ij} into the $Nn \times Nn$ matrix $\Sigma = \{\Sigma_{ij}\}$. Collecting the $(n \times 1)$ vectors $\varepsilon_{i,t}$ into $(T \times n)$ matrices $\varepsilon_i = (\varepsilon_{i,1}, \ldots, \varepsilon_{i,T})'$, then collecting these matrices into the $(T \times Nn)$ matrix $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_N)$, we obtain $e = vec(\varepsilon)$ being the vector of errors. This vector has covariance matrix

$$E(ee') = V_e = (\Sigma \otimes I_T). \tag{6}$$

The density of the errors, a key building block in forming the likelihood for this model, is then

$$|\Sigma|^{-T/2} \exp\left\{-\frac{1}{2}e'\left(\Sigma^{-1} \otimes I_T\right)e\right\} = |\Sigma|^{-T/2} \exp\left\{-\frac{1}{2}tr\Sigma^{-1}\varepsilon'\varepsilon\right\}.$$

We next give two representations for e that will prove useful in developing a sampling scheme for the parameters.

We rewrite (2) by defining $z_{i,t} = \beta_i' y_{i,t-1}$, the $1 \times (k + r_i)$ vector $X_{i,t} = \left(z_{i,t}', \Delta y_{i,t-1}', \ldots, \Delta y_{i,t-l}', d_t'\right)$, where k is the number of deterministic terms plus n times the number of lags (assumed to be common to all individuals and, hence, we have not included an i subscript), and the $(k + r_i) \times n$ matrix $B_i = (\alpha_i, \Gamma_{i,1}, \ldots, \Gamma_{i,l}, \Phi_i)'$ and, thus,

$$\Delta y_{i,t}' = X_{i,t}B_i + \varepsilon_{i,t}'. \tag{7}$$

If we stack the vectors in (7) over t as $\Delta y_i = (\Delta y_{i,1}, ..., \Delta y_{i,T})'$ and $X_i = \left(X_{i.1}', ..., X_{i.T}'\right)'$ then, we can write $\Delta y_i = X_i B_i + \varepsilon_i$. Vectorizing this equation gives us the form

$$
\begin{aligned}
vec\left(\Delta y_i\right) &= \left(I_n \otimes X_i\right)vec\left(B_i\right) + e_i \\
\text{or } y_i &= x_i b_i + e_i
\end{aligned}
$$

where $y_i = vec\left(\Delta y_i\right)$, $x_i = \left(I_n \otimes X_i\right)$, $b_i = vec\left(B_i\right)$ and $e_i = vec\left(\varepsilon_i\right)$ such that $E\left(e_i e_j'\right) = \Sigma_{ij} \otimes I_T$. We collect the vectors y_i and b_i into the vectors $y = (y_1', \ldots, y_N')'$, and $b = (b_1', \ldots, b_N')'$, and define the matrix x as the $TNn \times Nn(k + \overline{r})$ (where $\overline{r} = \frac{\sum_{i=1}^{N} r_i}{N}$) block diagonal matrix with diagonal equal to $(x_1, ..., x_N)$. Using these definitions, we can express the full system of equations as $y - xb = e$. The likelihood can now be expressed as

$$
\begin{aligned}
L(b, \Sigma, \beta) &= |\Sigma|^{-T/2} \exp\left\{-\frac{1}{2}(y - xb)'V_e^{-1}(y - xb)\right\} \\
&= |\Sigma|^{-T/2} \exp\left\{-\frac{1}{2}\left[s^2 + \left(b - \widehat{b}\right)'V^{-1}\left(b - \widehat{b}\right)\right]\right\}
\end{aligned}
\tag{8}
$$

where $s^2 = y'M_V y$, $M_V = V_e^{-1} - V_e^{-1}xVx'V_e^{-1}$, $\widehat{b} = Vx'V_e^{-1}y$, $V_e = (\Sigma \otimes I_T)$ and $V = \left(x'V_e^{-1}x\right)^{-1}$. Thus from our first representation of the likelihood, we can see that the form of the posterior for b conditional upon the β_i and Σ is Normal if the (conditional) prior for b is flat or Normal.

Our next representation of the likelihood demonstrates that we can obtain a Normal form for the cointegrating vectors (conditional on the other parameters of the model). That is, the conditional posterior density of the vector $b_{\beta^*} = \left(b'_{\beta^*,1}, \ldots, b'_{\beta^*,N}\right)'$, where $b_{\beta^*,i} = vec\left(\beta_i^*\right)$, can be shown to be Normal. To do this let us again rewrite (2) but this time define the $1 \times k$ vector $w_{i,t} = \left(\Delta y'_{i,t-1}, \ldots, \Delta y'_{i,t-l}, d'_t\right)$, and the $k \times n$ matrix $C_i = \left(\Gamma_{i,1}, \ldots, \Gamma_{i,l}, \Phi_i\right)'$ and, thus,

$$\Delta y'_{i,t} = y'_{i,t-1}\beta_i^* A_i' + w_{i,t}C_i + \varepsilon'_{i,t}. \tag{9}$$

If we stack the vectors over t as $\Delta y_i = (\Delta y_{i,1}, ..., \Delta y_{i.T})'$, $y_{i,-1} = (y_{i,0}, ..., y_{i,T-1})'$ and $w_i = \left(w'_{i,1}, ..., w'_{i,T}\right)'$, then we can write $\Delta y_i = y_{i,-1}\beta_i^* A_i' + w_i C_i + \varepsilon_i$. Vectorizing this equation we obtain

$$
\begin{aligned}
vec\left(\Delta y_i - w_i C_i\right) &= \left(A_i \otimes y_{i,-1}\right) vec\left(\beta_i^*\right) + vec\left(\varepsilon_i\right) \\
\text{or } \widehat{y}_i &= \widehat{x}_i b_{\beta^*,i} + e_i
\end{aligned}
$$

where $\widehat{y}_i = vec\left(\Delta y_i - w_i C_i\right)$, and $\widehat{x}_i = \left(A_i \otimes y_{i,-1}\right)$. Now stack the vectors $\widehat{y}_i$ into $\widehat{y} = \left(\widehat{y}_1', \ldots, \widehat{y}_N'\right)'$ and define $\widehat{x}$ as the $TNn \times Nn\overline{r}$ block diagonal matrix with diagonal equal to $\left(\widehat{x}_1, ..., \widehat{x}_N\right)$ so we can express the system of equations as $\widehat{y} - \widehat{x}b_{\beta^*} = e$. The likelihood can now be expressed as

$$
\begin{aligned}
L\left(b, \Sigma, \beta\right) &= |\Sigma|^{-T/2}\exp\left\{-\frac{1}{2}\left(\widehat{y} - \widehat{x}b_{\beta^*}\right)' V_e^{-1}\left(\widehat{y} - \widehat{x}b_{\beta^*}\right)\right\} \tag{10} \\
&= |\Sigma|^{-T/2}\exp\left\{-\frac{1}{2}\left[s_{\beta^*}^2 + \left(b_{\beta^*} - \widehat{b}_{\beta^*}\right)' V_{\beta^*}^{-1}\left(b_{\beta^*} - \widehat{b}_{\beta^*}\right)\right]\right\}
\end{aligned}
$$

where $s_{\beta^*}^2 = \widehat{y}' M_{V_{\beta^*}}\widehat{y}$, $M_{V_{\beta^*}} = V_e^{-1} - V_e^{-1}\widehat{x}V_{\beta^*}\widehat{x}'V_e^{-1}$, $\widehat{b}_{\beta^*} = V_{\beta^*}\widehat{x}'V_e^{-1}\widehat{y}$, and $V_{\beta^*} = \left(\widehat{x}'V_e^{-1}\widehat{x}\right)^{-1}$. This representation of the likelihood shows that the form of the posterior for b_{β^*} (conditional upon the C_i and Σ) is Normal if the (conditional) prior for b_{β^*} is Normal.

2.2 Priors

In this section, we describe two classes of priors which may be useful for empirical research. The first of these is a noninformative prior, suitable for reference or benchmark purposes. The second is an informative prior which contains what we call "soft homogeneity" restrictions. That is, in many cases economic theory suggests that the cointegration space should be the same for different individuals and of a particular form. In an empirical analysis, the researcher might not want to impose this sort of homogeneity in a strong sense, but, through the use of priors, we can do so in a soft sense. That is, rather than setting parameters to have the same values for all individuals, we specify common informative priors that favour parameter values which are similar for different individuals. This is likely to be of particular benefit since our model contains many parameters and, thus, issues relating to possible over-parameterization and efficiency of estimation are likely to be important.

Before describing the priors, we should highlight some important issues that have arisen in Bayesian analyses of cointegration in non-panel data contexts. As discussed previously, the VECM suffers from a lack of identification (both locally and globally). A large literature has grown which develops priors

which attempt to surmount the problems this causes (see the survey paper by Koop, Strachan, van Dijk and Villani, 2005). We will not recreate the arguments of this literature in detail here. Suffice it to note that it is unsatisfactory to use some apparently sensible approaches. For instance, at first sight it seems sensible just to use a standard prior (e.g. a flat prior or a Normal one) on B after imposing the linear normalization $\beta = [I_r\ B']'$. As discussed in Kleibergen and van Dijk (1994), the local non-identification of the model causes problems for this sort of Bayesian approach. The issue here is that when α has reduced rank (e.g., $\alpha = 0$) the conditional posterior distribution for $B|\alpha$ is equal to its prior (i.e. since B does not enter the likelihood function at the point $\alpha = 0$ there is no data-based learning about B and, thus, its posterior equals its prior at this point). If the prior for $B|\alpha = 0$ is improper (as it is in the common "noninformative" case), then the posterior will also be improper. Formally, Kleibergen and van Dijk (1994) associate the local non-identification problem with nonexistence of posterior moments and non-integrability of the posterior (under a common noninformative prior). Kleibergen and van Dijk (1998) additionally point out that local non-identification implies an absorbing state in a Gibbs sampler, thereby violating the convergence conditions for the sampler.

With regards to global identification, Strachan and Inder (2004) show how the use of linear identifying restrictions places a restriction on the estimable region of the cointegrating space. This paper also provides an extensive discussion of further problems associated with the use of linear identifying restrictions. Strachan and van Dijk (2004b) show that a flat and apparently "noninformative" prior on B in the linear normalization favors regions of the cointegration space near where the linear normalization is invalid. Hence, the linear normalization is used under the assumption that it is valid while at the same time the prior says that the normalization is likely to be invalid.

Coming out of this literature is the strong message that prior elicitation should be made directly off the cointegration space itself (which is all that is identified). Several papers, including Strachan (2003), Strachan and Inder (2004) and Villani (2005a,b) propose various approaches which involve such a focus. In this paper, we extend the general framework outlined in Strachan (2003) and Strachan and Inder (2004) to our panel cointegration model. The advantages of this approach are that they allow us to avoid identification restrictions that may restrict the estimable cointegration space, allow for priors which are, in a sense, noninformative (but are proper and, hence, allow for calculation of posterior odds ratios) and offer a convenient framework for incorporating prior information (if the researcher wishes to incorporate it).

To briefly motivate this approach (in the non-panel case), note that a cointegrating space $\mathfrak{p}$ is an r-dimensional hyperplane in a n-dimensional space and its relation to the cointegrating vectors β is that these vectors lie in and thereby identify that plane. Consider the case where $n = 2$ and a single cointegration vector exists which is parameterized in polar coordinates $\beta = (\cos\theta\ \sin\theta)'$, where $\theta \in [-\pi/2, \pi/2)$. It is only θ which determines the cointegration space and, thus, as shown in Strachan and Inder (2004), the length of β can be restricted to be unity for identification. Crucially, and in contrast to the linear normalization, this identifying restriction does not restrict the estimable cointegration space or distort the weight on the space of the parameter of interest, $\mathfrak{p}$. A natural candidate for a noninformative distribution on $\mathfrak{p}$ is the Uniform distribution on θ, the parameter

7

governing the direction of β and therefore $\mathfrak{p}$.

To extend these intuitive concepts to general n and r, some additional definitions are required. Our aim is to provide a rigorous definition of the intuitive idea of assigning equal prior probability to every possible cointegration space of dimension r. As described in Strachan and Inder (2004), having β being semi-orthogonal such that $\beta'\beta = I_r$ identifies the cointegrating vectors without placing any restrictions on the cointegrating space. The set of all $n \times r$ semi-orthogonal matrices is called the *Stiefel manifold* $\mathbb{V}_{n,r}$. The Stiefel manifold is a compact space and admits a Uniform distribution. In the case where $r = 1$, one might conceptualize the collection of directions of all n-dimensional unit vectors, $\beta \in \mathbb{V}_{n,1}$, as describing an n-dimensional unit sphere centered at the origin. Thus, we may visualize a Uniform distribution on the n-dimensional unit sphere as characterizing a Uniform distribution on $\mathbb{V}_{n,1}$. For $r > 1$, we can think of each vector in β as describing a unit sphere with the additional restriction that the vectors are all orthogonal to each other.

The *Grassman manifold* $\mathbb{G}_{n,r}$ is the abstract space of all possible r-dimensional planes of $\mathbb{R}^n$. The cointegration space is an element of the Grassman manifold, that is $\mathfrak{p} \in \mathbb{G}_{n,r}$. In the VECM only the space spanned by the columns of β is identified, such that we only have information on $\mathfrak{p} = sp(\beta) \in \mathbb{G}_{n,r}$. A Uniform prior for the cointegration space is therefore given by the Uniform distribution on $\mathbb{G}_{n,r}$.

For calculating posterior odds ratios, proper priors are required to avoid Bartlett's paradox (see Bartlett, 1957). But, since β now has a compact support, the prior over the cointegration space is proper. Formally, this approach uses the natural relationship between the Grassman manifold and the Stiefel manifold and the development of measures on these spaces presented in James (1954). In particular, a key result is that the Uniform distribution on $\mathbb{V}_{n,r}$ induces a Uniform distribution on $\mathbb{G}_{n,r}$ (see James, 1954, and Strachan and Inder, 2004). Thus, it is possible to work with the semi-orthogonal matrices, i.e. $\beta \in \mathbb{V}_{n,r}$, and adjust all integrals to account for the fact that $\mathbb{V}_{n,r}$ is a larger space than $\mathbb{G}_{n,r}$.

In this paper, we have only sketched out the basic ideas relating to prior elicitation in the cointegration models, and refer the reader to the literature we cite above for further details. Suffice it to note here that, in this paper, we extend these ideas to work with the panel cointegration model.

2.2.1 A Noninformative Prior

Let $b_\beta = \left(b'_{\beta,1}, \ldots, b'_{\beta,N}\right)'$, where $b_{\beta,i} = vec(\beta_i)$, contain all the parameters which determine the cointegration spaces. The remaining parameters of the model are Σ and b, where b is defined between (7) and (8). Noting that, conditional upon b_β, the model reduces to a linear one (see equation 7), a plausible candidate is the standard noninformative prior for multivariate linear models:

$$p(b, \Sigma, b_\beta) \propto |\Sigma|^{-(Nn+1)/2}, \tag{11}$$

where we add the additional restriction, arising from our wish to be noninformative about the cointegrating space and have an identifying restriction which does not limit that space, that β_i is semi-orthogonal such that $\beta'_i\beta_i = I_{r_i}$.

8

Note, however, that although the marginal prior for b_β is proper, the joint prior for all the parameters is improper. The impropriety relating to the prior for Σ is not a problem, since it is common to all models.[3] However, a proper prior would be required for the remaining parameters should we wish to calculate posterior odds ratios comparing different cointegrating ranks. That is, if we wish to estimate a single model for specified values for r_i (and specified values for l and d_t) the prior given in (11) will be appropriate. However, if we are wishing to compare this single model to another with different values for r_i (and/or different values for l and d_t), then an informative prior for b would be required. It is to such an informative prior to which we turn. However, it is worth noting in passing that a researcher who is interested in model comparison, but would prefer to avoid informative priors, could use information criteria to approximate marginal likelihoods or could adopt a training sample approach. That is, (11) could be used as a noninformative prior which is then combined with a training sample (e.g. the initial 10% of the data) to yield a "posterior". This "posterior" can then be used as an informative prior in a posterior analysis involving the rest of the data. See O'Hagan (1995) for a discussion of such an approach.

2.2.2 An Informative Prior (including soft homogeneity restrictions)

In many cases the researcher may wish to specify an informative prior on the cointegrating space. For instance, in our previous example arising from King, Plosser, Stock and Watson (1991), the researcher may wish to elicit a prior which implies that the cointegrating space lies in the region implied by the great ratios. Alternatively, the researcher may wish to elicit a prior which implies that the cointegration spaces (or other parameters) for different individuals are similar. We refer to the latter as soft homogeneity restrictions. In addition, in order to avoid the issues relating to Bartlett's paradox discussed in the previous section, the researcher may wish to elicit an informative prior for b. Here we describe an approach to prior elicitation which incorporates these aspects.

Some aspects of our prior are best motivated in the context of our posterior simulation algorithm. Hence, we digress briefly to informally discuss computation. Posterior computation is greatly complicated by the fact that β_i is semi-orthogonal which precludes use of the simple Gibbs sampling methods described, e.g., in Geweke (1996). For the non-panel cointegration model, Koop, Leon-Gonzalez and Strachan (2005) develop an efficient method of posterior simulation based on the idea of a collapsed Gibbs sampler developed in Liu (1994) and Liu, Wong and Kong (1994). To give some preliminary intuition for this, consider the relationships defined in (4). For prior elicitation or posterior computation, we may consider either (β_i, α_i) or (β_i^*, A_i). Crucially, in the first of these parameterizations, β_i is semi-orthogonal while α_i is unrestricted, whereas in the second it is β_i^* which is unrestricted whereas A_i is semi-orthogonal. In the next section we develop a collapsed Gibbs sampler which alternates between these two parameterizations. Arguments made in Liu (1994) and Liu, Wong and Kong (1994)

---

[3]When calculating posterior odds ratios, it is common to make use of improper, noninformative priors over parameters which are common to all models (see, e.g., Kass and Raftery, 1995). Fernández, Ley and Steel (2001) employ such a prior for an error variance in a model averaging exercise. As they point out, the prior in (11) is invariant to scale transformations and, although it is not strictly Jeffreys' prior, it is that part of Jeffreys' prior related to Σ and widely accepted as a noninformative prior for Σ.

prove that this will be more efficient than a Gibbs sampler which works only with (β_i, α_i) or (β_i^*, A_i). Even more crucially, with the priors developed in this section, the collapsed Gibbs sampler will only involve draws from the Normal distribution (and inverted Wishart[4] for Σ), enormously simplifying the computational burden.

We now turn to our informative prior and begin by discussing b and Σ. Typically, these parameters will be of less importance in an empirical exercise than the prior on the cointegrating space. For Σ we maintain the noninformative prior given in (11), although an inverted-Wishart form could also easily be handled. For b we assume:[5]

$$b \sim N\left(0, \underline{V}\frac{1}{\nu}\right) \tag{12}$$

where ν is a scalar which controls the degrees of informativeness or precision of the prior. Note that ν can be interpreted as a shrinkage parameter and, thus, (12) shares some similarities with shrinkage priors commonly used in the VAR literature (see, e.g., Litterman, 1986). Note, however, that we treat ν as a parameter (rather than a hyperparameter selected subjectively by the researcher).

Now consider the prior covariance matrix (up to the scalar shrinkage parameter) $\underline{V}$ in (12). Of course, any choice for $\underline{V}$ can be made. Here we motivate a particular form for the elements of $\underline{V}$ which relate to α_i or, equivalently, A_i. Considering the relationships in (4) and surrounding discussion, it makes sense, analogous to our noninformative prior for the semi-orthogonal β_i, to assume that the $n \times r_i$ semi-orthogonal matrices $(A_i, ..., A_N)$ are a priori independent and that:

$$p\left(A_i, ..., A_N | \tau, v, b_{\beta^*}\right) \propto 1 \tag{13}$$

as this implies a Uniform but proper density for each of subspaces defined by the A_i for $i = 1, .., N$. Given the relationships in (4) we can derive a prior for (β_i^*, A_i) from a prior for (β_i, α_i) or vice versa. The prior (13), along with the prior for β_i (to be defined shortly), implies that

$$vec\left(\alpha_i\right)|\tau, \beta_i, \nu \sim N\left(0, \frac{1}{\nu}\left(\beta_i' P_\tau^{-1}\beta_i\right)^{-1} \otimes I_n\right), \tag{14}$$

and, thus, that the diagonal blocks of $\underline{V}$ that correspond to α_i are equal to $\left(\beta_i' P_\tau^{-1}\beta_i\right)^{-1} \otimes I_n$ (where P_τ will be defined shortly). The remaining elements of $\underline{V}$, corresponding to the parameters $(C_1, ..., C_N)$, can be specified using either informative or noninformative choices and will be further discussed below.

For the cointegration spaces, $\mathfrak{p}_i$ (and therefore for β_i) it is often desirable to have a prior which allows for a common location across individuals. If an economist believes a parameter is likely to have a particular value, she will often place more prior mass around this likely point. To extend this idea from parameters to spaces, some new ideas are required. To provide some intuition, consider the case where we have a prior belief that the space of β_i should be approximately the space of H where H is semi-orthogonal and is of the same dimension as β_i (we will extend this to allow H to have a different number of columns from β_i below). To obtain the semi-orthogonal matrix H the researcher

---

[4]See, e.g., Bauwens, Lubrano and Richard (1999), page 305 for a definition of the inverted Wishart distribution.

[5]In the following material, we use notation where lower bars (e.g. as in $\underline{V}$) denote prior hyperparameters which must be selected by the researcher.

can first specify the matrix H^g containing desired coefficient values and then use the transformation $H = H^g \left(H^{g\prime} H^g \right)^{-1/2}$. The matrix H constructed in this way will span the same space as H^g but will be semi-orthogonal.

For instance, if, motivated by King, Plosser, Stock and Watson (1991), we wanted a prior reflecting the fact that the great ratios are probably cointegrating relationships, we would set:

$$H^g = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix}.$$

H^g is not semi-orthogonal but $H = H^g \left(H^{g\prime} H^g \right)^{-1/2}$ will be (and will span the same space).

A dogmatic prior would be obtained by setting $\beta_i = H$ which places all of the prior mass for $\mathfrak{p}_i$ at $\mathfrak{p}^H = sp\left(H\right)$. Strachan and Inder (2004) develop an informative, but non-dogmatic prior, for the cointegration space and we adopt a similar approach here. Intuitively, we want a prior which says the cointegration spaces, $\mathfrak{p}_i$, are likely to be close to $\mathfrak{p}^H = sp\left(H\right)$ and, thus, farthest from $\mathfrak{p}^{H\perp} = sp\left(H_\perp\right)$ where $H_\perp$ is the orthogonal complement of H. The $\mathfrak{p}_i$s are weighted averages of $\mathfrak{p}^H$ and $\mathfrak{p}^{H\perp}$ and we can elicit a prior about these weights.

One way to motivate our informative prior is through its implications for β_i^*. To this end, suppose we have an $n \times r_i$ matrix Z_i with all elements being i.i.d. $N\left(0, \nu^{-1}\right)$. A standard result tells us that the space of Z_i will be Uniformly distributed over the Grassman manifold. If we simply set $\beta_i^* = Z_i$ and used this as a prior for β_i^* then it would be noninformative over the cointegrating space. To get a dogmatic informative prior for β_i^* (and, thus, the cointegrating space), we can project Z_i into $\mathfrak{p}^H$. Another standard result in matrix algebra says this projection is given by $\beta_i^* = HH'Z_i$. At the opposite extreme, we could project Z_i into $\mathfrak{p}^{H\perp}$ as $\beta_i^* = H_\perp H_\perp' Z_i$ if we wanted a cointegration space as far away from $\mathfrak{p}^H$ as possible. A non-dogmatic informative prior can be introduced by introducing the random variable η (with distribution centered at 0) which centers the prior over $\mathfrak{p}^H$, but attaches weight to other spaces as:

$$\begin{aligned} \beta_i^* &= HH'Z_i + \eta H_\perp H_\perp' Z_i \\ &= P_\eta Z_i \end{aligned}$$

where $P_\eta = HH' + \eta H_\perp H_\perp'$.

Using the properties of the Normal distribution, it follows that

$$vec\left(\beta_i^*\right) | \eta, \nu \sim N\left(0, I_{r_i} \otimes \frac{1}{\nu} P_\eta P_\eta'\right).$$

But, given the structure of P_η, it follows that $P_\eta P_\eta' = HH' + \eta^2 H_\perp H_\perp' = P_{\eta^2}$. Thus, η enters the prior only through η^2 and, accordingly, we introduce the notation $\tau = \eta^2$ and use the following conditional Normal prior for β_i^*:

$$vec\left(\beta_i^*\right) | \tau, \nu \sim N\left(0, I_{r_i} \otimes \frac{1}{\nu} P_\tau\right) \tag{15}$$

or, equivalently,

$$b_{\beta^*}|\tau, \nu \sim N\left(0, \frac{1}{\nu}\underline{V}_{\beta^*}\right) \tag{16}$$

where $\underline{V}_{\beta^*} = diag\left(I_{r_i} \otimes P_\tau\right)$.[6]

This prior will more strongly weight the cointegrating space towards H the closer τ is to zero. At $\tau = 1$, this prior is Uniform over the Grassman manifold (since $P_{\tau=1} = I_n$) and $\tau > 1$ implies more weight towards the space of $H_\perp$. Therefore, it is sensible to either truncate the distribution of τ to the region $(0, 1]$ or to choose the hyperparameters in the prior for τ so that $\tau > 1$ is a very unlikely event.

Note that our informal motivation implicitly assumed H to be of the same dimension as β. However, if we define $H \in V_{s,n}$ to be a known $n \times s$ $(s \geq r_i)$ matrix and $H_\perp \in V_{n-s,n}$ its orthogonal complement, then our prior expresses the belief that the cointegration space $\mathfrak{p}_i$ is likely to be included in the higher dimensional space $\mathfrak{p}^H$.[7]

For any $p(\tau)$ and $p(\nu)$, we can write the joint prior for β_i^* and (ν, τ) as

$$
\begin{aligned}
& p(\nu) p(\tau) \left(\frac{2\pi}{\nu}\right)^{-Nn\overline{r}/2} \tau^{-N(n\overline{r}-r^2)/2} \exp\left\{-\frac{\nu}{2}\Sigma_{i=1}^N tr\beta_i^{*\prime} P_{\tau^{-1}} \beta_i^*\right\} \\
= \ & p(\nu) p(\tau) \left(\frac{2\pi}{\nu}\right)^{-Nn\overline{r}/2} \tau^{-N(n\overline{r}-r^2)/2} \exp\left\{-\frac{\nu}{2}\Sigma_{i=1}^N tr\beta_i^{*\prime} HH' \beta_i^*\right\} \\
& \times \exp\left\{-\frac{\nu}{2\tau}\Sigma_{i=1}^N tr\beta_i^{*\prime} H_\perp H_\perp' \beta_i^*\right\},
\end{aligned}
\tag{17}
$$

where $r^2 = \frac{\sum_{i=1}^N r_i^2}{N}$. In our empirical work, we select for $p(\nu)$ the form:

$$\nu \sim G\left(\underline{\mu}_\nu, \underline{\nu}_\nu - nN\overline{r}\right) \tag{18}$$

where $G\left(\underline{\mu}_\nu, \underline{\nu}_\nu - nN\overline{r}\right)$ denotes the Gamma distribution with mean $\underline{\mu}_\nu$ and degrees of freedom $\underline{\nu}_\nu - nN\overline{r}$. Note that the degrees of freedom depends on $nN\overline{r}$. This arises out of our wish to have the prior $p(\nu|\beta_i^* = 0)$ the same for every model we consider in our model comparison exercise. Such a condition is necessary for using the Savage-Dickey density ratio as we do below. For brevity, we will not provide details, but it turns out that if $p(\nu)$ has the form given in (18) then the resulting prior for ν satisfies the (reasonable and commonly-used) conditions for the Savage-Dickey density ratio to be used.

Using the transformations $\beta_i^* = \beta_i \kappa_i$ and $\kappa_i \kappa_i' = \varpi$ with change of measure $(d\beta_i^*) = 2^{-r_i}$ $|\varpi|^{(n-r_i-1)/2} (d\varpi)(d\beta_i)$, and using (18) to integrate out ν, we can obtain

$$p(\tau, b_\beta) = p(\tau) \tau^{-N(n\overline{r}-r^2)/2} \Pi_i^N \left|\beta_i' P_{\tau^{-1}} \beta_i\right|^{-n/2} \mathfrak{c}_r. \tag{19}$$

---

[6] Note that b and b_{β^*} share elements in common (i.e. κ_i) and therefore, the prior specification on b has implications on the prior of b_{β^*}. This is the reason why the shrinkage parameter, ν, appears in 16. Note that ν does not affect the marginal prior for cointegrating spaces.

[7] If we have the case that $s < n - 1$, then we will have models with $r > s$ and the above prior distribution is not applicable. In this case, in the absence of economic theory to guide us, we would assume a Uniform prior distribution for $\mathfrak{p}_i$.

where $\mathfrak{c}_r = 2^{-N\overline{r}}\pi^{N(r^2-\overline{r})/4-Nn\overline{r}/2}\Pi_i^N\Pi_j^{r_i}\Gamma\left[(n+1-j)/2\right].$ Since the cointegrating space $\mathfrak{p}_i$ is identified given a value for β_i^*, the expression in (17) or (19) can be regarded as the joint prior for $(\mathfrak{p}_i, \tau)$ conditional upon r_i.

From the form in (17), a convenient form of prior for τ^{-1} that suggests itself is Gamma

$$\tau^{-1} \sim G\left(\underline{\mu}_\tau, \underline{\nu}_\tau\right)$$

possibly truncated to $[1, \infty)$ to ensure $\tau < 1$. Alternatively, we could choose values, as we do in our application, such as $\underline{\mu}_\tau = 5$ and $\underline{\nu}_\tau = 15$ which will ensure $P\left(\tau < 1\right) \approx 1$. In the truncated case the normalizing constant $\mathfrak{c}_r$ in (19) is adjusted by dividing by $P\left(\tau^{-1} > 1\right)$.

Note that, if we use appropriate common values for $\underline{\mu}_\tau$ and $\underline{\nu}_\tau$ for every individual, we will ensure that each $\mathfrak{p}_i$ has its prior mass near to $\mathfrak{p}^H = sp\left(H\right)$. This is an example of what we refer to as a soft homogeneity restriction. That is, we are not restricting, a priori, each individual to have the same cointegration space, but we are expressing the view that different individuals are likely to have similar cointegration spaces. In general, such soft homogeneity restrictions can be imposed in two ways with this prior. First, priors (such as the prior for τ) can be the same or can share common locations. Second, we can choose $\underline{V}$ defined in (12) to have a structure which implies correlation between the same parameters for different individuals. Here we briefly describe one strategy for specifying $\underline{V}$. The $Nn\left(k+\overline{r}\right) \times Nn\left(k+\overline{r}\right)$ matrix $\underline{V}$ can be partitioned into $n\left(k+r_i\right) \times n\left(k+r_i\right)$ blocks on the diagonal $\underline{V}_{ii}$, which can be chosen to have various forms (see equation 14 for the form relating to the $\alpha_i's$). On the off-diagonals, it would often make sense for the $n\left(k+r_i\right) \times n\left(k+r_j\right)$ matrices $\underline{V}_{ij}$ to have zeros in the rows and columns relating to the $\alpha_i's$. Thus, no a priori correlation[8] is assumed between the $\alpha_i's$. However, it will usually be sensible to assume that $vec\left(C_i\right)$ and $vec\left(C_j\right)$ are positively correlated with one another, a priori. This can be done by specifying the $nk \times nk$ matrix of prior covariances between the elements of $vec\left(C_i\right)$ and $vec\left(C_j\right)$ to be equal to $\underline{\rho}_b I_{nk}$, where $0 < \underline{\rho}_b < 1$.

This completes our specification of an informative prior which has three key properties: i) It allows for prior information about the likely location of the cointegration space to be incorporated; ii) It allows for prior information about the degree of similarity in coefficients across individuals (which we refer to as soft homogeneity restrictions); iii) It contains a parameter ν which allows for shrinkage of coefficients on short run dynamics and deterministic terms.

2.3 Posterior Computation

Using the priors specified above and the likelihood in (8) and (10), we can derive various posterior conditional densities of use in our posterior simulation algorithm. Using standard results (e.g., Bauwens, Lubrano and Richard, 1999, Chapter 9), the conditional posterior of Σ can be confirmed to be inverted Wishart with degrees of freedom parameter T and scale matrix $\varepsilon'\varepsilon$, where ε is defined just after 6.

---

[8]If the cointegrating relations are exactly identified, all individuals share the same cointegrating rank and the same cointegrating relationship holds for all equations, then it would make sense to assume the adjustment coefficients (α_i) are a priori correlated. However, without these restrictions, it does not make sense to assume the columns of α_i and α_j will be correlated.

Similar standard results can be used to obtain the posterior distribution for b conditional upon (Σ, b_β) which is Normal with mean $\overline{b}_b = \overline{V}V^{-1}\widehat{b}$ and covariance $\overline{V}_b = \left[V^{-1} + \nu \underline{V}^{-1}\right]^{-1}$.

The final block in a standard Gibbs sampler would involve the cointegrating vectors, $b_\beta = (b'_{\beta,1}, \ldots, b'_{\beta,N})'$, where $b_{\beta,i} = vec(\beta_i)$. Because of the semi-orthogonality of β_i, this posterior conditional is difficult to draw from directly. However, the conditional posterior distribution of b_{β^*} turns out to be Normal (we remind the reader that $b_{\beta^*} = \left(b'_{\beta^*,1}, \ldots, b'_{\beta^*,N}\right)'$, where $b_{\beta^*,i} = vec(\beta_i^*)$). To be precise, recalling the definition of C_i before equation (9), and defining $c_i = \left(vec(A_i)', vec(C_i)'\right)'$ and $c = (c'_1, \ldots, c'_N)'$, the posterior distribution for b_{β^*} conditional upon (Σ, c) is Normal with mean $\overline{b}_{\beta^*} = \overline{V}_{\beta^*}V_{\beta^*}^{-1}\widehat{b}_{\beta^*}$ and covariance $\overline{V}_{\beta^*} = \left[V_{\beta^*}^{-1} + \nu \underline{V}_{\beta^*}^{-1}\right]^{-1}$.

In the process of drawing the parameters $\left(\Sigma, b, b_{\beta^*}\right)$, we need to draw ν and τ^{-1}. The conditional posterior distribution for ν is Gamma with mean

$$\overline{\mu}_\nu = \overline{\nu}_\nu \left[\left(\underline{\nu}_\nu - nN\overline{r}\right)/\underline{\mu}_\nu + b'\underline{V}^{-1}b\right]^{-1}$$

and degrees of freedom $\overline{\nu}_\nu = Nnk + \underline{\nu}_\nu$. The conditional distribution for τ^{-1} is Gamma with degrees of freedom $\overline{\nu}_\tau = \underline{\nu}_\tau + N\left(n\overline{r} - r^2\right)$ and mean $\overline{\mu}_\tau = \overline{\nu}_\tau \left[\underline{\nu}_\tau/\underline{\mu}_\tau + \nu\Sigma_{i=1}^{N}tr\beta_i^{*'}H_\perp H'_\perp \beta_i^*\right]^{-1}$.

From these conditional distributions we summarize the following sampling scheme using a collapsed Gibbs sampling method:

1. Initialize $(b, \Sigma, b_\beta, \nu, \tau) = \left(b^{(0)}, \Sigma^{(0)}, \beta^{(0)}, \nu^{(0)}, \tau^{(0)}\right)$.

2. Draw $\Sigma|b, b_\beta, \nu, \tau$ from $IW\left(\sum_{i=1}^{N}\varepsilon'_i\varepsilon_i, T\right)$

3. Draw $b|\Sigma, b_\beta, v, \tau$ from $N\left(\overline{b}_b, \overline{V}_b\right)$

4. Calculate $A_i = (\alpha'_i\alpha_i)^{-\frac{1}{2}}\alpha_i$ and create c.

5. Draw $b_{\beta^*}|c, \Sigma, v, \tau$ from $N\left(\overline{b}_{\beta^*}, \overline{V}_{\beta^*}\right)$.

6. Decompose each β_i^* as $\beta_i^* = \beta_i\kappa_i$ using $\kappa_i = \left(\beta_i^{*'}\beta_i^*\right)^{\frac{1}{2}}$ and $\beta_i = \beta_i^*\kappa^{-1}$. Construct $\alpha_i = A_i\kappa_i$.

7. Draw $\nu|b, b_\beta, \Sigma, \tau$ from $G\left(\overline{\mu}_\nu, \overline{\nu}_\nu\right)$.

8. Draw $\tau^{-1}|b, b_\beta, \Sigma, v$ from $G\left(\overline{\mu}_\tau, \overline{\nu}_\tau\right)$.

9. Repeat steps 2 to 8 for a suitable number of replications.

Note that, in this sampler, the transformations involving the long run multipliers are based on (5). To see why these steps suffice to set up a posterior simulator, we first show that, conditional on (v, τ, Σ), steps 3 to 6 define a collapsed Gibbs sampler (Liu, 1994). To show this, note from (4) that α_i can be decomposed into (A_i, κ_i), and that therefore the draw of b in step 3 is a draw of $(c, \kappa_1, ..., \kappa_N)$. Similarly, β_i^* can be decomposed into (β_i, κ_i) and a draw of b_{β^*} in step 5 is a draw of $(b_\beta, \kappa_1, ..., \kappa_N)$.

Therefore, dropping for simplicity the conditioning arguments (v, τ, Σ), the value of c obtained in step 3 is a draw from $c|b_\beta$, that is obtained marginally on $(\kappa_1, ..., \kappa_N)$. Similarly, the value of b_β obtained in step 5 is a draw from $b_\beta|c$, (i.e. obtained again marginally on $(\kappa_1, ..., \kappa_N)$). Therefore, steps 3 to 6 define the collapsed Gibbs sampler suggested by Liu (1994) and Liu, Wong and Kong (1994), who show that this algorithm is more efficient than a standard Gibbs sampling algorithm (i.e. one which simply draws from the conditional posteriors of b and b_β).

Finally, we extend the collapsed Gibbs sampler with steps that generate $(\kappa_1, ..., \kappa_N)$, Σ, v, and τ from their corresponding conditional posterior densities and it is trivial to show that the posterior density continues to be the invariant distribution of the sampler. For a more detailed explanation of this algorithm in the context of a standard (non-panel) cointegration model see Koop, Leon-Gonzalez and Strachan (2005).

We will usually be interested in comparing different models nested within the general model defined above. For instance, we might wish to compare the unrestricted model with one where the same cointegrating rank holds for all individuals. We also might wish to calculate the posterior for r_i for $i = 1, ..N$. The Savage-Dickey density ratio (see, e.g., Verdinelli and Wasserman, 1995) proves to be a simple and efficient way of doing so. That is, it allows us to compute the Bayes factor comparing every model to a base model (e.g. the model where cointegration does not occur for any individual). This information can then be used to compare any two models, build up the posterior for r_i for $i = 1, ..N$, do Bayesian model averaging or select a single model. To compute the Bayes factor for the model M_r with a particular set of cointegrating ranks: $r = (r_1, r_2, \ldots, r_N)$ against model M_0 with $r = (0, \ldots, 0)$, we note that the restricted case occurs when $\alpha_i = 0$. As α_i and Π_i have the same singular values (which determine the rank of a matrix, e.g. Golub and van Loan, 1996), $\Pi_i = 0$ occurs if and only if $\alpha_i = 0$. If we define $\alpha = (vec(\alpha_1)', ..., vec(\alpha_N)')'$, we can use the conditional posterior distribution and (marginal) prior for α to compute the Savage-Dickey density ratio (SDDR):

$$B_{0,r} = \frac{p(\alpha|M_r, y)|_{\alpha=0}}{p(\alpha|M_r)|_{\alpha=0}} \tag{20}$$

Thus we can use output from our Gibbs sampler and the prior to estimate the required ratio:

$$\widehat{B}_{0,r} = \frac{\frac{1}{M}\Sigma_{m=1}^{M} p\left(\alpha|M_r, \Sigma^{(m)}, C_1^{(m)}, ..., C_N^{(m)}, \tau^{(m)}, \nu^{(m)}, b_\beta, y\right)\Big|_{\alpha=0}}{p(\alpha|M_r)|_{\alpha=0}}, \tag{21}$$

where $m = 1, .., M$ denote the (post burn-in) Gibbs sampler replications and (m) superscripts denote the replications themselves (or, as below, functions of these replications).

We begin by deriving the analytical expression for $p(\alpha|M_r)|_{\alpha=0}$. Using the properties of the Gamma distribution and the MACG distribution (Chikuse, 1990), it can be shown that the marginal prior for α evaluated at $\alpha = 0$ is

$$p(\alpha|M_r)|_{\alpha=0} = \left(\frac{2\underline{\nu}_\tau}{\underline{\mu}_\tau}\right)^{N(n\overline{r}-r^2)/2} \frac{\Gamma\left(\frac{N(n\overline{r}-r^2)+\underline{\nu}_\tau}{2}\right)}{\Gamma(\underline{\nu}_\tau/2)} \frac{\Gamma(\underline{\nu}_\nu/2)}{\Gamma((\underline{\nu}_\nu - Nn\overline{r})/2)} \left(\frac{\underline{\nu}_\nu - Nn\overline{r}}{\underline{\mu}_\nu \pi}\right)^{Nn\overline{r}/2}$$

This expression gives us the denominator of the SDDR. The numerator of the SDDR is the marginal posterior for α evaluated at zero. Using the fact that the posterior for b conditional upon (Σ, b_β) is $N(\bar{b}_b, \overline{V}_b)$, it follows that α is $N(\bar{b}_\alpha, \overline{V}_\alpha)$, where $\bar{b}_\alpha$ and $\overline{V}_\alpha$ are given by the elements of $\bar{b}_b$ and $\overline{V}_b$ that correspond to α. Therefore, the Gibbs sampler can be used to estimate the numerator of the SDDR as:

$$\frac{(2\pi)^{-Nn\bar{r}/2}}{M} \sum_{m=1}^{M} \left| \overline{V}_\alpha^{(m)} \right|^{-1/2} \exp\left\{ -\frac{1}{2} \bar{b}_\alpha^{(m)\prime} \overline{V}_\alpha^{(m)-1} \bar{b}_\alpha^{(m)} \right\}.$$

There are other restricted versions of our general model in which the researcher may be interested. The Appendix describes how variants on the methods described above can be used to calculate Bayes factors relating to these models. Here we just list the restrictions of interest. Firstly, in practice it is often the case that there is interest in testing overidentifying restrictions of the form $\mathfrak{p}_i \subseteq \mathfrak{p}^H$ for a subset of the countries $i = 1, ..., N$. This restriction can be imposed by writing $\beta_i = H\varphi_i$, where $\varphi_i \in V_{r_i,s}$ is an unknown $s \times r_i$ full rank matrix. Our empirical example in the next section shows how such a restriction can arise. Secondly, we would also like to obtain the probability that all countries have the same unknown cointegrating space $\mathfrak{p} = sp(\beta)$. Finally, the Appendix also describes how to calculate the probability that $sp(\beta) \subseteq sp(H)$ in the case in which all countries share the same unknown β.

3   Illustration Using Simulated Data

This section uses simulated data to illustrate the properties of the proposed methodology and its robustness to the specification of the prior. Instead of a conventional Monte Carlo experiment, we draw on ideas outlined in Selke, Bayarri, and Berger (2001) to develop a simulation experiment which, as we explain below, better reveals the performance of our approach.

We consider seven data generating processes (DGPs) and one prior specification: $H^g = \begin{pmatrix} 1 & 1 \end{pmatrix}'$, $\underline{\nu}_\tau = 15$, $\underline{\mu}_\tau = 5$, $\underline{\nu}_\nu = 42$, $\underline{\mu}_\nu = 21$, and $\underline{\rho}_b = 0.4$ (we remind the reader that $H = H^g (H^{g\prime}H^g)^{-1/2}$). Except for H, this is the same prior that we use in the empirical application in the next section. We consider $N = n = 2$, $T = 85^9$, $l = 0$, an intercept in all models ($d_t = 1$) and, in each DGP, we fix the error covariance matrix equal to the value used by Groen and Kleibergen (2003) in their Monte Carlo experiment:

$$\Sigma_{ii} = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix} \qquad \text{and} \qquad \Sigma_{ij} = \begin{pmatrix} 0.70 & 0.60 \\ 0.60 & 0.85 \end{pmatrix} \qquad \text{with } i \neq j.$$

We assume that there are only 4 possible models: M_1:($r_1 = r_2 = 0$), M_2:($r_1 = 0, r_2 = 1$), M_3:($r_1 = 1, r_2 = 0$) and M_4:($r_1 = r_2 = 1$).

In a conventional Monte Carlo experiment draws from a DGP would involve simply drawing from a single model (with parameters set to particular values). This is consistent with the hypothesis

---

[9] These 85 observations were the last 85 observations of 135. That is, the initial 50 were discarded.

Table 1: Specification of the (hyper) parameters for the distributions from which the parameters are drawn in the simulation experiment.

DGP	Parameter (brief description of the distribution)	
	$\mathfrak{p}_i$	b
DGP_1	$H^g = \begin{pmatrix} 1 & 1 \end{pmatrix}', \underline{\nu}_\tau = 15, \underline{\mu}_\tau = 5$	$\underline{\nu}_\nu = 42, \underline{\mu}_\nu = 21$
DGP_2	$\tau = 1$ (Uniform)	$\underline{\nu}_\nu = 42, \underline{\mu}_\nu = 21$
DGP_3	$\tau = 1$ (Uniform)	$\nu^{-1} = 0.05$ (ν^{-1} equal to its prior mean)
DGP_4	$\tau = 1$ (Uniform)	$\nu^{-1} = 0.2$ (large $var(\alpha_i)$)
DGP_5	$\tau = 1$ (Uniform)	$\nu^{-1} = 0.5$ (very large $var(\alpha_i)$)
DGP_6	$\tau = 1$ (Uniform)	$\nu^{-1} = 0.02$ (small $var(\alpha_i)$)
DGP_7	$\tau = 1$ (Uniform)	$\nu^{-1} = 0.002$ (very small $var(\alpha_i)$)

testing ideas underlying frequentist econometrics (e.g. the idea of null hypothesis and the importance allocated to frequentist concepts such as the size of a test). However, as argued in Selke, Bayarri, and Berger (2001) and Berger and Selke (1987), the ideas underlying Bayesian model comparison are very different. Accordingly, following their arguments, in our simulation experiment we repeatedly draw data sets from different distributions. In particular, we set up distributions over our model space and parameter spaces and draw from these. For each draw of a model and parameter values, we then draw an artificial data set. All our DGPs involve the same distributions over the model space and, accordingly, each of our seven DGPs arise from different distributions over the parameters. Note that these distributions have the same functional form as our priors, but the hyperparameters selected to create our DGPs do not have to coincide with the prior hyperparameters we use to estimate our models.

To be precise, in each of our DGPs data is drawn from each model with probability 1/4, which is equal to the prior probability of each model. Conditional on model M_i, the parameters are drawn from distributions that are of the same form as the prior, but with different hyperparameters. In all cases we use $\underline{\rho}_b = 0.4$. The specification of the remaining hyperparameter values for each of these distributions is given in Table 1.

Note that DGP_1 involves the same informative distribution over the cointegrating space as we use in our prior, but the remaining DGPs are less informative. For the remaining parameters, we have a wide variety of specifications. The specifications in each DGP imply we draw A_i (defined in (4)) from a Uniform distribution on the Stiefel manifold. For DGP_2 to DGP_7 we fix τ equal to 1, which implies that β_i is also drawn from a Uniform distribution. This contrasts with the prior we use for $\mathfrak{p}_i = sp(\beta_i)$, which gives more weight to the space defined by H. In addition, DGP_3 to DGP_7 vary in the expected value of κ_i. Higher values of ν^{-1} imply higher expected values for κ_i and therefore higher expected values for the singular values of Π_i. Note that there is 95% prior probability that ν^{-1} lies in the interval (0.032, 0.077). Therefore, DGP_4 and DGP_5 specify a value of ν^{-1} that is large compared to the prior information, whereas DGP_6 and DGP_7 specify a value that is small.

For each DGP, 2500 artificial data sets were generated. For each data set, the posterior probability of each model (i.e. each rank combination) was calculated. In order to analyse the performance of posterior probabilities in this setup, let us define the following concepts (see Selke, Bayarri, and Berger, 2001, for the development of these concepts). Let $C_i(0.5)$ be the set of data sets in which model M_i had posterior probability higher than 0.5. Assume that a model is selected whenever its posterior probability is higher than 0.5 and let $R_i(0.5)$ be defined as an error rate that gives the proportion of the samples in $C_i(0.5)$ that were *not* generated from model M_i.

To motivate why these are interesting metrics, we digress briefly to provide a bit of the theory from Selke, Bayarri and Bayarri (2001). Consider the ideal case where the distribution used to generate the datasets is the same as the prior. For this case, suppose M_i is chosen whenever its posterior model probability (p_i) is equal to a particular value p_i^*. From the definition of posterior model probability, the error rate that results (i.e. the proportion of samples that were classified as M_i but were in fact generated from another model) is equal to $1 - p_i^*$. Thus, posterior model probabilities, unlike p-values, are constructed to reflect true error rates (see also Berger and Selke, 1987, for discussion). However, it is unlikely that we will ever simulate a dataset that results in posterior probability of M_i being *exactly* p_i^* so this approach is hard to implement. Therefore, one possibility would be to accept those draws with posterior model probability lying in the interval $(p_i^* - \varepsilon, p_i^* + \varepsilon)$, where ε is a small number. This is what Selke, Bayarri and Berger (2001) do. Alternatively, a simple rule of thumb such as "select M_i if $p_i > p_i^*$" can be used (as we have done with $p_i^* = 0.5$) and the average value of p_i $(\overline{p_i})$ among the datasets in $C_i(0.5)$ can be reported and the previous reasoning implies this will also be informative about the error rate $R_i(0.5)$. In particular, if the number of datasets is large and are generated from the prior, $R_i(0.5)$ will be equal to $1 - \overline{p_i}$.

Table 2 shows the values of $\overline{p_i}$, $R_i(0.5)$ and the number of data sets in $C_i(0.5)$. Overall, the strategy of choosing M_i when $p_i > 0.5$ seems to work very well, selecting the correct model much of the time. Recall that DGP_1 draws all model parameters, except for Σ, from the prior. Not surprisingly, therefore, Table 2 shows that for DGP$_1$, $R_i(0.5)$ is very close to $(1 - \overline{p_i})$ for every $i = 1, ..., 4$. These two quantities are still close for every i for DGP$_2$ and DGP$_3$, which indicates that posterior model probabilities are still a reliable measure of error when the prior of β_i is misspecified and/or ν^{-1} is fixed to a particular value instead of being random. When $\nu^{-1} = 0.2$ (DGP$_4$), which is far outside the prior 95% credible interval of $(0.032, 0.077)$, $(1 - \overline{p_i})$ is still close to $R_i(0.5)$ for every i. Similarly, when $\nu^{-1} = 0.02$ (DGP$_6$), which is small compared to prior information, posterior model probabilities continue to be a reliable measure of error for every i. However, when $\nu^{-1} = 0.5$ (DGP$_5$), posterior model probabilities are not reliable when model M_4 is chosen $((1 - \overline{p_i}) < R_i(0.5))$, although they still seem to be reasonable when models M_1 to M_3 are selected. Something similar, but in the opposite direction, happens when ν^{-1} is very small (DGP$_7$). In this case, the posterior model probability is only a reliable measure of error when M_4 is chosen.

Thus, the simulation illustrates that posterior model probabilities, unlike p-values (e.g., Selke, Bayarri, and Berger, 2001), are reliable measures of error unless the prior for ν^{-1} is seriously misspecified. Therefore, some amount of careful prior elicitation for ν^{-1} is desirable to avoid such misspecification.

In our case, for example, it should be noted that DGP$_5$ tends to generate very explosive processes whenever $r_i = 1$, resulting in data that would be extremely unreasonable (at least for standard applications with macroeconomic data such as the one considered in the next section). For example, it can be shown that DGP$_5$ implies that about 45% of the datasets would have $(|y_{1,t}| > 1000)$ for every $t = 1, .., T$ when $r_1 = r_2 = 1$, which is not sensible for macroeconomic data such as that which we use in our application.

		M_1	M_2	M_3	M_4
DGP_1	$R_i(0.5)$	0.07	0.05	0.06	0.03
	$1 - \overline{p_i}$	0.06	0.05	0.05	0.02
	$\widetilde{N_i}$	842	527	545	330
DGP_2	$R_i(0.5)$	0.05	0.03	0.04	0.03
	$1 - \overline{p_i}$	0.05	0.04	0.05	0.02
	$\widetilde{N_i}$	844	541	519	344
DGP_3	$R_i(0.5)$	0.04	0.05	0.03	0.03
	$1 - \overline{p_i}$	0.05	0.04	0.04	0.02
	$\widetilde{N_i}$	830	545	518	354
DGP_4	$R_i(0.5)$	0.01	0.03	0.07	0.09
	$1 - \overline{p_i}$	0.07	0.04	0.05	0.03
	$\widetilde{N_i}$	855	540	531	307
DGP_5	$R_i(0.5)$	0.004	0.105	0.107	0.255
	$1 - \overline{p_i}$	0.079	0.067	0.061	0.046
	$\widetilde{N_i}$	765	531	542	392
DGP_6	$R_i(0.5)$	0.11	0.06	0.07	0.02
	$1 - \overline{p_i}$	0.06	0.05	0.05	0.03
	$\widetilde{N_i}$	840	558	498	341
DGP_7	$R_i(0.5)$	0.36	0.22	0.24	0.05
	$1 - \overline{p_i}$	0.10	0.10	0.10	0.07
	$\widetilde{N_i}$	868	514	542	289

Table 2: Error rates $(R_i(0.5))$, one minus the average posterior probabilities $(1 - \overline{p_i})$ and number of samples in $C_i(0.5)$ $(\widetilde{N_i})$ for each DGP.

Table 3 shows other measures that illustrate the performance of Bayesian model selection in this context. For each DGP and model from which the data was generated, it gives the proportion of times (denoted $\%_i$) that model M_i had the largest posterior probability. In addition, it shows the average posterior model probability (denoted P_i) of M_i for each DGP and each generating model. Note that the proportion of times that the correct model has largest posterior model probability is almost always near or above 90%, and that on average posterior model probabilities are accordingly large. The exception is DGP$_7$, where the detection rate of the true model is lower, as are average posterior model probabilities. This is to be expected, as lower values of ν^{-1} mean that data generated when $r > 0$ will

be more similar to data generated when $r = 0$, and hence model selection becomes more difficult and we see slightly larger values of $\%_1$ and P_1.

		$\%_1$	$\%_2$	$\%_3$	$\%_4$	P_1	P_2	P_3	P_4
DGP_1	M_1	**0.99**	0.00	0.01	0.00	**0.94**	0.03	0.03	0.00
	M_2	0.05	**0.94**	0.01	0.01	0.05	**0.91**	0.01	0.03
	M_3	0.05	0.00	**0.94**	0.01	0.05	0.00	**0.91**	0.04
	M_4	0.03	0.07	0.07	**0.84**	0.02	0.06	0.07	**0.84**
DGP_2	M_1	**0.99**	0.00	0.00	0.00	**0.94**	0.03	0.03	0.00
	M_2	0.04	**0.95**	0.00	0.01	0.04	**0.92**	0.00	0.03
	M_3	0.03	0.00	**0.95**	0.01	0.04	0.00	**0.92**	0.04
	M_4	0.01	0.03	0.04	**0.92**	0.01	0.03	0.05	**0.91**
DGP_3	M_1	**0.98**	0.01	0.00	0.00	**0.94**	0.03	0.03	0.00
	M_2	0.03	**0.96**	0.00	0.01	0.03	**0.93**	0.00	0.04
	M_3	0.03	0.00	**0.96**	0.01	0.03	0.00	**0.93**	0.03
	M_4	0.00	0.04	0.04	**0.92**	0.00	0.04	0.04	**0.91**
DGP_4	M_1	**0.95**	0.02	0.03	0.00	**0.90**	0.05	0.05	0.00
	M_2	0.01	**0.96**	0.00	0.03	0.01	**0.94**	0.00	0.05
	M_3	0.01	0.00	**0.97**	0.02	0.01	0.00	**0.94**	0.06
	M_4	0.01	0.01	0.04	**0.94**	0.01	0.01	0.04	**0.94**
DGP_5	M_1	**0.85**	0.07	0.06	0.02	**0.79**	0.08	0.09	0.03
	M_2	0.00	**0.93**	0.00	0.06	0.00	**0.90**	0.00	0.10
	M_3	0.00	0.00	**0.91**	0.08	0.00	0.00	**0.88**	0.11
	M_4	0.00	0.00	0.01	**0.99**	0.00	0.00	0.01	**0.99**
DGP_6	M_1	**0.98**	0.01	0.01	0.00	**0.94**	0.03	0.03	0.00
	M_2	0.07	**0.92**	0.00	0.01	0.07	**0.89**	0.01	0.04
	M_3	0.09	0.01	**0.90**	0.01	0.08	0.01	**0.87**	0.04
	M_4	0.03	0.07	0.08	**0.82**	0.03	0.07	0.08	**0.82**
DGP_7	M_1	**0.98**	0.01	0.01	0.00	**0.91**	0.04	0.04	0.00
	M_2	0.24	**0.73**	0.02	0.01	0.21	**0.70**	0.03	0.05
	M_3	0.23	0.02	**0.74**	0.01	0.22	0.03	**0.70**	0.05
	M_4	0.10	0.19	0.21	**0.49**	0.09	0.19	0.20	**0.51**

Table 3: Two summaries ($\%_i$ and P_i) for each DGP. $\%_i$ is the percentage of times that M_i has largest posterior model probability. P_i is the average posterior model probability of M_i.

4   Empirical Work

In this section we investigate support for the monetary model of the exchange rate commonly employed in international finance. We focus upon the specification proposed by Groen (2000) which implies a particular testable relationship among the following variables: $e_{i,t}$, the log of the exchange rate for country i at time t; $m_{i,t}$, the log of the ratio of the quantity of domestic to foreign money supply; and $x_{i,t}$, the log of the relative real income. Groen (2000) shows that in a long-run model for bilateral

exchange rates, the theory implies the relation

$$e_{i,t} - \beta_1 m_{i,t} - \beta_2 x_{i,t} = \beta_0 + z_{i,t}$$

will be stationary (i.e., $z_{i,t}$ should be an $I(0)$ process) with $\beta_1 = 1$ and $\beta_2 < 0$. If the variables in the vector $y_{i,t} = (e_{i,t}, m_{i,t}, x_{i,t})$ are $I(1)$, this model implies they cointegrate with a particular cointegrating space. The data are quarterly and consist of U.S. dollar exchange rates and the log ratio of money (m) and income (x) for France ($i = 1$), Germany ($i = 2$), and the United Kingdom ($i = 3$) to the U.S. equivalents. The dat runs from the first quarter of 1973 to the last quarter of 1994. The data were those used in Groen and Kleibergen (2003) and are described in detail in Groen (2000).

We have chosen this application because the economic model implies a varied and clear set of testable restrictions on the cointegrating space. That is, we have a requirement that the cointegrating rank be one for all countries, a linear restriction on β_1, as well as an inequality restriction upon β_2. We note that it is often the case that the economic model of interest implies such a set of joint restrictions, some of which are linear and some are nonlinear. In such a case, classical inference usually proceeds with a mixture of sequential testing and informal inference to gather evidence for or against the model, with no single statistic with known power to indicate the degree of support in favor of the model. Therefore, the classical work of Groen, which tested sequentially the rank restriction and the other restrictions, provided only informal evidence about the degree of support for the model. An advantage of using the Bayesian approach is that we are able to provide a formal summary of the evidence for the model via posterior model probabilities. We are also able to assess the evidence, if desired, for components of the model. For example, we may be interested in whether the variables cointegrate or whether the cointegrating ranks are common to all countries, or whether the $\beta's$ are common across all countries.

Within the specification of the statistical model we use, the monetary exchange rate model implies that $r_i = 1$ for each country and that the cointegrating spaces are restricted. In particular, if we define the orthogonal matrix H as:

$$H = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{2}} & 0 \\ 0 & 1 \end{bmatrix},$$

and introduce the semi-orthogonal vector $\varphi_i = \begin{bmatrix} \varphi_{1,i} \\ \varphi_{2,i} \end{bmatrix}$, we can write these restrictions as:

$$\beta_i = H\varphi_i = \begin{bmatrix} \frac{1}{\sqrt{2}}\varphi_{1,i} \\ -\frac{1}{\sqrt{2}}\varphi_{1,i} \\ \varphi_{2,i} \end{bmatrix}, \text{ with } \frac{\varphi_{2,i}}{\varphi_{1,i}} > 0. \tag{22}$$

Note that, for $r_i = 1$, this set of restrictions does not actually require that all panels share the same cointegrating space, since the inequality restriction allows a different value of $\phi_{2,i}$ for each panel, provided that it has the same sign as $\phi_{1,i}$. However, all of these spaces will be subspaces of the space defined by H. Other restrictions of more general interest in the cointegrating panel data model are

equality of the ranks for all panels ($r_i = r$ for all i), and equality of the cointegrating spaces for all panels, $\mathfrak{p}_i = \mathfrak{p}$ such that $r_i = r$ and $\beta_i = \beta$ for all i.

We compute posterior probabilities distributions for the cointegrating ranks from both unrestricted and restricted models. We consider two types of restrictions. The first imposes the same unknown cointegrating space: $\beta_i = \beta$ and $r_i = r$ for all i. The second restricts the cointegrating space of at least one country, such that $sp(\beta_i) \subseteq sp(H)$ for some i with $r_i = 1, 2$[10]. This makes a total of 221 models. Following Groen and Kleibergen (2003), all models include an intercept and 3 seasonal dummies and we fix the number of lags equal to 3. As in the artificial data experiment in the previous section, we choose our prior hyperparameters as: $\underline{\nu}_\nu = 42$, $\underline{\mu}_\nu = 21$, $\underline{\nu}_\tau = 15$, $\underline{\mu}_\tau = 5$, and $\underline{\rho}_b = 0.4$. We use 15000 replications of the sampling algorithm presented in Section 2.3. For the sake of comparison, we also calculate the Bayesian Information Criterion (BIC) for each of these models[11].

Recall we let $i = 1, 2, 3$ correspond to France, Germany and UK, respectively. The BIC selects the model with ($r_1 = r_2 = r_3 = 0$) as the best model, followed by the model with ($r_1 = 1, r_2 = r_3 = 0$) and no other restrictions. If posterior model probabilities are calculated using the BIC approximation, these two models would get 90% of the probability. However, the actual posterior model probabilities calculated using our approach are spread over a wide range of models: no less than 28 models would be required to contain 98% of the probability. Table 4 presents the details of the 5 most likely models, which get 71.4% of the probability mass. All these models assign rank equal to one to France and Germany and restrict $sp(\beta_i) \subseteq sp(H)$ in at least one country. In particular, the model with $r_i = 1$ and $sp(\beta_i) \subseteq sp(H)$ for every i, which gives support to the monetary exchange model, has a non negligible probability that is equal to 0.05. Conditional on this model, $Pr(\phi_{2i}/\phi_{1i} > 0$ for $i = 1, 2, 3)$[12]$= 0.12$, which means that the probability of all the restrictions implied by the the monetary exchange model holding in every country is $0.12 * 0.05 = 0.006$. The probability of many other restrictions of interest can be evaluated by simply adding up the posterior model probabilities of models in which the restriction is true. For example, $Pr(r_1 = r_2 = r_3) = 0.09$, $Pr(sp(\beta_1) = sp(\beta_2) = sp(\beta_3)) = 0.004$, $Pr(r_2 = 1) = 0.86$, $Pr(sp(\beta_1) \subseteq sp(H), r_1 > 0) = 0.79$. Finally, the probability that $(sp(\beta_i) \subset sp(H), r_i = 1)$ for at least one country is 0.94, which again gives support to the monetary exchange model holding in at least one country.

---

[10]If $r_i = 1$, then $\mathfrak{p}_i \subset \mathfrak{p}^H$, while if $r_i = 2$, then $\mathfrak{p}_i = \mathfrak{p}^H$.

[11]In order to calculate the penalty for the number of parameters in the BIC, we count the parameters in the semi-orthogonal but otherwise unrestricted β_i matrix as $nr_i - r_i^2$, which is the dimension of the Grassman manifold $\mathbb{G}_{n,r_i}$ defined above (Strachan and Inder, 2004). Similarly, when β_i is restricted such that $\mathfrak{p}_i \subseteq \mathfrak{p}^H$, we fix the penalty corresponding to the semi-orthogonal but otherwise unrestricted ϕ_i matrix to be $2r_i - r_i^2$. We use our algorithm to search for the maximum value of the actual likelihood by using 1000 draws from a modified posterior density. This modification increases the accuracy of the obtained maximum likelihood values and consists in analysing the posterior that results when the sample size is increased by a factor of 600 and the additional data is just a replication of the real data. Therefore, the maximum value of the log likelihood function in this modified dataset is 600 times the value of the log likelihood in our real data. And most importantly, the dispersion of the posterior around the mode will be much smaller and therefore the accuracy of the maximized likelihood will be much larger.

[12]This probability was approximated by the proportion of draws from the posterior of this model in which the restriction was verified.

r_1	r_2	r_3	o_1	o_2	o_3	E	Prob
1	1	0	1	1	0	0	0.35
1	1	0	1	0	0	0	0.12
1	1	0	0	1	0	0	0.10
1	1	2	1	1	1	0	0.10
1	1	1	1	1	1	0	0.05

Table 4: Posterior probabilities for the 5 most likely models. The first 3 columns indicate the rank of each country in a particular model. $i = 1, 2, 3$ corresponds to France, Germany and UK, respectively. In the following three columns o_i takes value 1 when the restriction $sp(\beta_i) \subseteq sp(H)$ is imposed and 0 otherwise. E takes value 1 if the restriction $sp(\beta_1) = sp(\beta_2) = sp(\beta_3)$ is imposed and zero otherwise. The last column indicates the probability of each model.

5   Conclusion

In this paper, we have discussed Bayesian inference in cointegrated panel data models. We adopt a very general specification where each individual is characterized by its own vector error correction model. Special cases of this model allow for individuals to have common cointegrating rank and/or common cointegrating spaces. We develop a noninformative prior as well as an informative prior which allows for sensible priors on the cointegration spaces. The latter prior also allows for prior information about the degree of common structure across individuals to be used. Efficient posterior simulation is carried out using a collapsed Gibbs sampler.

While we consider this a useful start to employing Bayesian methods in this area of models, there are a number of directions for future development. For instance, in a PPP study, Li (1999) argues that estimating relationships of interest individually for each country results in overly noisy estimates. On the other hand, imposing strict homogeneity by assuming these relationships are the same for all countries tends to be overly severe due to the differences in macroeconomic policies in each country. Such severe restriction are often rejected. Li suggests specifying an unknown hierarchical prior and conducting inference upon the distribution from which the parameters for the PPP relations come, not upon the actual PPP parameters themselves. In this paper we have assumed that the cointegrating spaces came from a common known prior distribution and investigated support for common cointegrating spaces ($\mathfrak{p}_i = \mathfrak{p}$ for all i). To adopt the Li approach, a hierarchical prior could be placed upon the prior distribution for the cointegrating spaces, rather than assuming a known prior distribution. That is, a prior could be placed upon $\mathfrak{p}^H$ in Section 2.2.2.

Further, while we have provided a method of conducting inference upon a class of models, we have only alluded to the conduct of policy advice. The application of this class of models to policy via, say, forecasts or cross impulse responses could provide useful information on, e.g., the forms of international linkages. Finally, Bayesian model averaging over combinations of $\mathfrak{p}_i$, l_i and d_i could be used to provide inference using an even wider set of models.

References

Baltagi, B. & C. Kao (2000) Nonstationary panels, cointegration in panels and dynamic panels: A survey. *Advances in Econometrics* 15, 7-51.

Bartlett, M. (1957) A Comment on D. V. Lindley's statistical paradox. *Biometrica* 44, 533–534.

Berger, J. & T. Selke (1987) Testing of a point null hypothesis: the irreconcilability of significance levels and evidence (with discussion), *Journal of the American Statistical Association*, 82, 112-139.

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

Carmeci, G. (2005) A Bayesian state space approach to cointegration in panel data models. Working paper available at http://www.cide.info/conf/papers/1128.pdf.

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

Fernández, C., E. Ley & M.F.J. Steel (2001) Benchmark priors for Bayesian model averaging. *Journal of Econometrics* 100, 381-427.

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

Golub, G.H. & van Loan (1996) *Matrix Computations, Third edition* Baltimore: John Hopkins Universty Press

Groen, J. (2000) The monetary exchange rate model as a long-run phenomenon. *Journal of International Economics* 52, 299-319.

Groen, J. & F. Kleibergen (2003) Likelihood-based cointegration analysis in panels of vector error-correction models. *Journal of Business and Economic Statistics* 21, 295-318.

Hsiao, C. & M.H. Pesaran (2004) Random coefficient panel data models. Working paper available at http://www.econ.cam.ac.uk/faculty/pesaran/rcm17june04.pdf.

Jacobson, T., J. Lyhagen, R. Larsson & M. Nessén (2002) Inflation, exchange rates and PPP in a multivariate panel cointegration model. Working paper available at http://www.riksbank.se/upload/7405/wp_145.pdf.

James, A.T. (1954) Normal multivariate analysis and the orthogonal group. *Annals of Mathematical Statistics* 25, 40-75.

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

Kao, C. (1999) Spurious regression and residual-based tests for cointegration in panel data. *Journal of Econometrics* 90, 1-44.

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

King, R., C. Plosser, J., Stock & M.Watson (1991) Stochastic trends and economic fluctuations. *American Economic Review* 81, 819-840.

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

Kleibergen, F. & H. van Dijk (1998) Bayesian simultaneous equations analysis using reduced rank structures. *Econometric Theory* 14, 701-743.

Koop, G., R. Leon-Gonzalez & R. Strachan (2005) Efficient posterior simulation for cointegrated models with priors on the cointegration space, manuscript.

Koop, G., R. Strachan, H. van Dijk and M. Villani (2005) Bayesian approaches to cointegration. To appear as Chapter 25 in T.C. Mills and K. Patterson (eds.). *Palgrave Handbook of Theoretical Econometrics*, manuscript available at http://www.le.ac.uk/economics/research/RePEc/lec/leecon/dp04-27.pdf.

Larsson, R., J. Lyhagen and M. Löthgren (2001) Likelihood-based cointegration tests in heterogeneous panels. *Econometrics Journal* 4, 109-142.

Li, K. (1999) Testing symmetry and proportionality in PPP: A panel data approach. *Journal of Business and Economic Statistics* 17, 409-418.

Litterman, R. (1986) Forecasting with Bayesian vector autoregressions: Five years of experience. *Journal of Business and Economic Statistics* 4, 25-38.

Liu, J.S. (1994) The collapsed Gibbs sampler with applications to a gene regulation problem. *Journal of the American Statistical Association* 89, 958-966.

Liu, J.S., W. Wong & A. Kong (1994) Covariance structure of the Gibbs sampler with applications to comparisons of estimators and augmentation schemes. *Biometrika* 81, 27-40.

McCoskey, S. & C. Kao (1998) A residual-based test of the null of cointegration in panel data. *Econometric Review*s 17, 57-84.

O'Hagan, A. (1995) Fractional Bayes factors for model comparison. *Journal of the Royal Statistical Society, Series B* 57, 99-138.

Pedroni, P. (2004) Panel cointegration; Asymptotic and finite sample properties of pooled time series tests with an application to the Purchasing Power Parity Hypothesis. *Econometric Theory* 20, 597-625.

Pesaran, M. H., C. Hsiao & K. Tahmiscioglu (1999) Bayes estimation of short-run coefficients in dynamic panel data models. In C. Hsiao, K. Lahiri, L-F Lee & M.H. Pesaran (eds), *Analysis of Panels and Limited Dependent Variables: A Volume in Honour of G S Maddala.* Cambridge: Cambridge University Press, pp .268-296.

Phillips, P. C. B. (1991) Optimal inference in cointegrated systems. *Econometrica* 59, 283-306.

Phillips, P.C.B. & H. Moon (2000) Nonstationary panel data analysis: An overview of some recent developments. *Econometric Reviews* 19, 263-286.

Selke, T., M.J. Bayarri, & J. Berger (2001) Calibration of P-values for testing precise null hypotheses. *The American Statistician* 55, 62-71.

Strachan, R. (2003) Valid Bayesian estimation of the cointegrating error correction model. *Journal of Business and Economic Statistics* 21, 185-195.

Strachan, R. & B. Inder (2004) Bayesian analysis of the error correction model. *Journal of Econometrics* 123, 307-325.

Strachan, R & H. van Dijk (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 and H. van Dijk (2004a) Improper priors with well defined Bayes factors. Econometric Institute Report EI 2004-18, Erasmus University.

Strachan, R. and H. van Dijk (2004b) Valuing structure, model uncertainty and model averaging in vector autoregressive processes. Econometric Institute Report EI 2004-23, Erasmus University Rotterdam.

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

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

Villani, M. (2005b) Bayesian point estimation of the cointegration space, forthcoming in *Journal of Econometrics.*

Appendix: Over-identifying Restrictions and Other Restricted Models

Probability of over-identifying restrictions

In practice it is often the case that there is interest in testing the overidentifying restriction $\mathfrak{p}_i \subseteq \mathfrak{p}^H$ for a subset of countries $i = 1, ..., N$. Our empirical example showed how such a restriction can arise. This restriction can be imposed by writing $\beta_i = H\varphi_i$, where H is a $n \times s$ semi-orthogonal matrix and $\varphi_i \in V_{r_i,s}$ is an unknown $s \times r_i$ full rank matrix, with $s \geq r_i$. This specification[13] requires a prior for φ_i and in this paper we use a Uniform prior for φ_i on $V_{r_i,s}$, although an informative MACG prior (Chikuse, 1990) could also be used. Within this framework, it is also possible to use the SDDR to obtain the Bayes factor, which in this case is defined as the probability of the restricted (overidentified) model over the probability of the unrestricted model. Note that one can write $\beta_i^* = H\varphi_i^* + H_\perp \lambda_i$, where $\varphi_i^* = H'\beta_i^*$ is a $s \times r_i$ full rank matrix and $\lambda_i = H_\perp' \beta_i^*$ is a $(n-s) \times r_i$ full rank matrix. Therefore, the overidentifying restriction can be imposed by setting $\lambda_i = 0$. The numerator in the SDDR can be easily estimated by noting that $(\lambda_1, ..., \lambda_N)$ is a linear transformation of b_{β^*} and therefore the conditional posterior of $(\lambda_1, ..., \lambda_N)$ given (c, Σ, v, τ) is a Normal density (e.g. Bauwens, Lubrano and Richard, 1999). In particular, let $\widetilde{H}$ be a block diagonal matrix with diagonal blocks equal to: $(I_{r_1} \otimes H_\perp, ..., I_{r_N} \otimes H_\perp)$. The conditional posterior density of $\lambda = (vec(\lambda_1)', ..., vec(\lambda_N)')'$ is Normal with variance $\widetilde{H}' \overline{V}_{\beta^*} \widetilde{H}$ and mean $\widetilde{H}' \overline{b}_{\beta^*}$, where $\overline{V}_{\beta^*}$ and $\overline{b}_{\beta^*}$ were defined in Section 2.3. The numerator of the SDDR for this restriction is estimated as:

$$\frac{(2\pi)^{-(n-s)N\overline{r}/2}}{M} \sum_{m=1}^{M} \left| \widetilde{H}' \overline{V}_{\beta^*}^{(m)} \widetilde{H} \right|^{-1/2} \exp\left\{ -\frac{1}{2} \overline{b}_{\beta^*}^{(m)'} \widetilde{H} \left(\widetilde{H}' \overline{V}_{\beta^*}^{(m)} \widetilde{H} \right)^{-1} \widetilde{H}' \overline{b}_{\beta^*}^{(m)} \right\}$$

where $N\overline{r} = \sum_{i=1}^{N} r_i$. The denominator, which is given by the prior of λ evaluated at zero, is equal to:

$$p(\lambda)|_{\lambda=0} = \left(\frac{2\underline{\nu}_\tau (\underline{\nu}_\nu - nN\overline{r})}{\underline{\mu}_\tau \underline{\mu}_\nu \pi} \right)^{(n-s)N\overline{r}/2} \frac{\Gamma\left(\frac{\underline{\nu}_\tau + (n-s)N\overline{r}}{2} \right) \Gamma\left(\frac{(n-s)N\overline{r} + (\underline{\nu}_\nu - nN\overline{r})}{2} \right)}{\Gamma(\underline{\nu}_\tau/2) \Gamma((\underline{\nu}_\nu - nN\overline{r})/2)}$$

Estimation with equal cointegrating spaces: $\{sp(\beta_i) = sp(\beta), \ i = 1, ..., N\}$

We would also like to obtain the probability that all countries have the same cointegrating space. For this purpose, we first discuss how to set up a posterior simulator for the model that restricts all cointegrating spaces to be the same. Let β be a $n \times r$ semi-orthogonal matrix that represents the cointegrating space common to all countries. We rewrite the matrix of long-run multipliers as:

$$\beta \alpha_i' = \beta_i DD^{-1}\alpha_i' \equiv \widetilde{\beta} \widetilde{\alpha}_i'$$

where D is a $r \times r$ is a symmetric positive definite matrix. We stress that unlike κ_i, which was defined as one of the components of the polar decomposition of α_i (Golub and van Loan, 1996, p. 149), the matrix D is not identified. However, the introduction of D facilitates posterior computations because neither $\widetilde{\beta}$ nor $\widetilde{\alpha}_i$ are subject to restrictions. Our strategy is to specify a proper prior on D and to use a simple Gibbs sampling algorithm. We note that this strategy could also be used for the case

---

[13]Further motivation is given in Strachan and Inder (2004). Strachan and van Dijk (2004b) shows how this specification can be implemented in a macroeconomic example.

in which cointegrating spaces differ between countries, but stress that it would be less efficient than the 'κ–algorithm' that we specify in Section 2.3 above for two reasons. Firstly, the κ–algorithm (implicitly) integrates out the parameter D, and it thereby achieves a comparative advantage (Liu, 1994). Secondly, the κ–algorithm draws b_β and c marginally on $(\kappa_1, ..., \kappa_N)$, which is likely to result in smaller autocorrelations in the Markov Chain.

We specify the same prior on the cointegrating space as in Section 2.2.2. We define $b_{\widetilde\beta} = vec(\widetilde\beta)$ and assume that $b_{\widetilde\beta}$ follows a priori a $N(0, n^{-1}I_r \otimes P_\tau)$, where P_τ is defined as above. Note that this implicitly specifies a proper prior for D and that the marginal prior for β continues to be the same as the one presented in Section 2.2.2. As a prior for $vec(\widetilde\alpha_i)$ we choose a $N(0, \nu^{-1}I_{nr})$. Note that this prior specification implies that the prior mean of α_i is zero and the variance-covariance matrix of $vec(\alpha_i)$ conditional on (ν, β, τ) is $\left(\nu^{-1}\left(\beta'P_\tau^{-1}\beta\right)^{-1} \otimes I_n\right)$. Therefore, the prior mean and variance for $(\alpha_1, ..., \alpha_N)$ are the same as in Section 2.2.2.

Let us define $\widetilde{b} = \left(\widetilde{b}_1', ..., \widetilde{b}_N'\right)'$, where $\widetilde{b}_i = vec\left((\widetilde\alpha_i, \Gamma_{i,1}, ..., \Gamma_{i,l}, \Phi_i)'\right)$. The prior for $\widetilde{b}$ is a $N(0, \nu^{-1}\underline{\widetilde{V}})$, where $\underline{\widetilde{V}}$ is specified in the same way as $\underline{V}$ except for the diagonal blocks corresponding to $\widetilde\alpha_i$, which are now equal to the identity matrix.

In order to obtain the posterior conditional of $\widetilde\beta$, let $\ddot{x}_i = (\widetilde\alpha_i \otimes y_{i,-1})$, $\ddot{x} = (\ddot{x}_1', ..., \ddot{x}_N')'$, $V_{\widetilde\beta} = \left(\ddot{x}'V_e^{-1}\ddot{x}\right)^{-1}$, $\widehat{b}_{\widetilde\beta} = V_{\widetilde\beta}\ddot{x}V_e^{-1}\widetilde{y}$ and $\underline{V}_{\widetilde\beta} = n^{-1}I_r \otimes P_\tau$. The conditional posterior of $\widetilde\beta$ given (b, Σ, τ, ν) is Normal with covariance $\overline{V}_{\widetilde\beta} = \left[V_{\widetilde\beta}^{-1} + \underline{V}_{\widetilde\beta}^{-1}\right]^{-1}$ and mean $\overline{b}_{\widetilde\beta} = \overline{V}_{\widetilde\beta}V_{\widetilde\beta}^{-1}\widehat{b}_{\widetilde\beta}$.

The conditional posterior of $\widetilde{b}$ given $\left(\widetilde\beta, \Sigma, \tau, \nu\right)$ is a Normal with mean $\overline{b}_{\widetilde{b}}$ and variance $\overline{V}_{\widetilde{b}}$. The expressions for $\overline{b}_{\widetilde{b}}$ and $\overline{V}_{\widetilde{b}}$ are obtained in the same way as $\overline{b}_b$ and variance $\overline{V}_b$ respectively in Section 2.3, except we replace all β_i everywhere with $\widetilde\beta$. Similarly the posterior for ν has the same form as in Section 2.3, except we replace b with $\widetilde{b}$ and $\underline{V}$ with $\underline{\widetilde{V}}$, such that ν has a Gamma posterior distribution with mean $\overline{\nu}_\nu\left[\left(\underline{\nu}_\nu - Nnr\right)/\underline{\mu}_\nu + \widetilde{b}'\underline{\widetilde{V}}^{-1}\widetilde{b}\right]^{-1}$ and degrees of freedom $\overline{\nu}_\nu = Nnk + \underline{\nu}_\nu$. The conditional distribution for τ^{-1} is Gamma with degrees of freedom $\overline{\nu}_\tau = \underline{\nu}_\tau + (n-r)r$ and mean $\overline{\mu}_\tau = \overline{\nu}_\tau\left[\underline{\nu}_\tau/\underline{\mu}_\tau + n^{-1}tr\widetilde\beta'H_\perp H_\perp'\widetilde\beta\right]^{-1}$.

Therefore, a Gibbs sampling algorithm is defined by simply sampling iteratively from $\widetilde{b}|\left(b_{\widetilde\beta}, \Sigma, \tau, \nu\right)$, $b_{\widetilde\beta}|\left(\widetilde{b}, \Sigma, \tau, \nu\right)$, $\Sigma|\left(\widetilde{b}, b_{\widetilde\beta}, \tau, \nu\right)$, $\tau|\left(\widetilde{b}, b_{\widetilde\beta}, \Sigma, \nu\right)$ and $\nu|\left(\widetilde{b}, b_{\widetilde\beta}, \tau, \Sigma\right)$. A sample from the posterior of $(b, \beta, \Sigma, \tau, \nu, D)$ can be obtained using the following transformations:

$$D = \left(\widetilde\beta'\widetilde\beta\right)^{1/2} \qquad \beta = \widetilde\beta D^{-1} \qquad \alpha_i = \widetilde\alpha_i D$$

Finally, we note that although the κ-algorithm cannot be used to sample from the posterior when the restriction $sp(\beta_i) \subset sp(H)$ is imposed for some i, the D- algorithm just described can be easily adapted to this case.

Calculating the probability that $\{sp(\beta_i) = sp(\beta), \ i = 1, ..., N\}$

Let $\widetilde{M}_r$ be the model in which $sp(\beta_i) = sp(\beta)$, for every $i = 1, ..., N$ and $rank(\beta) = r$. To compute the Bayes factor for this model $\left(\widetilde{M}_r\right)$ against the model M_0 with $r = (0, ..., 0)$, note that M_0 arises when $\widetilde\alpha = (vec(\widetilde\alpha_1)', ..., vec(\widetilde\alpha_N))' = 0$. Therefore, the SDDR can be estimated as:

$$\widetilde{B}_{0,r} = \frac{\frac{1}{M}\Sigma_{m=1}^M p\left(\widetilde\alpha|\widetilde{M}_r, \Sigma^{(m)}, C_1^{(m)}, ..., C_N^{(m)}, \widetilde\beta^{(m)}, \nu^{(m)}, \tau^{(m)}, y\right)\Big|_{\widetilde\alpha=0}}{p\left(\widetilde\alpha|\widetilde{M}_r\right)\Big|_{\widetilde\alpha=0}}$$

where the numerator and denominator are given, respectively, by the following two expressions:

$$\frac{(2\pi)^{-Nnr/2}}{M} \sum_{m=1}^{M} \left|\overline{V}_{\widetilde{\alpha}}^{(m)}\right|^{-1/2} \exp\left\{-\frac{1}{2}\left(\overline{b}_{\widetilde{\alpha}}^{(m)}\right)' \overline{V}_{\widetilde{\alpha}}\left(\overline{b}_{\widetilde{\alpha}}^{(m)}\right)\right\} \text{ and}$$

$$\frac{\Gamma\left(\frac{\underline{\nu}_\nu}{2}\right)}{\Gamma\left((\underline{\nu}_\nu - Nnr)/2\right)} \left(\frac{\underline{\nu}_\nu - Nnr}{\underline{\mu}_\nu \pi}\right)^{Nnr/2}$$

where M is the number of (post burn-in) replications of the Gibbs sampler, (m) superscripts denote Gibbs sampler draws and $(\overline{b}_{\widetilde{\alpha}}, \overline{V}_{\widetilde{\alpha}})$ are the elements of $(\overline{b}_{\widetilde{b}}, \overline{V}_{\widetilde{b}})$ that correspond to $\widetilde{\alpha}$.

Calculating the probability that $sp(\beta) \subseteq sp(H)$.

Finally, we present the SDDR to obtain the posterior probability that $sp(\beta) \subseteq sp(H)$ in the case where $sp(\beta_i) = sp(\beta)$. To evaluate the hypothesis that $sp(\beta) \subseteq sp(H)$, we write $\widetilde{\beta} = H\widetilde{\varphi} + H_\perp \widetilde{\lambda}$, where $\widetilde{\varphi} = H'\widetilde{\beta}$ is a $s \times r$ full rank matrix and $\widetilde{\lambda} = H_\perp'\widetilde{\beta}$ is a $(n-s) \times r$ full rank matrix. Note that $vec\left(\widetilde{\lambda}\right)$ is Normally distributed with mean $\overline{\lambda} = (I_r \otimes H_\perp')\overline{b}_{\widetilde{\beta}}$ and covariance matrix $V_{\widetilde{\lambda}} = (I_r \otimes H_\perp')\overline{V}_{\widetilde{\beta}}(I_r \otimes H_\perp)$. Therefore, the overidentifying restriction can be imposed by setting $\widetilde{\lambda} = 0$. The SDDR can be estimated by:

$$\frac{(2\pi)^{-(n-s)r/2}}{M\left(\frac{\underline{\nu}_\tau}{\underline{\mu}_\tau \pi n}\right)^{(n-s)r/2} \frac{\Gamma\left(\frac{\underline{\nu}_\tau + (n-s)r}{2}\right)}{\Gamma\left(\underline{\nu}_\tau/2\right)}} \sum_{m=1}^{M} \left|V_{\widetilde{\lambda}}^{(m)}\right|^{-1/2} \exp\left\{-\frac{1}{2}\overline{\lambda}^{(m)'} V_{\widetilde{\lambda}} \overline{\lambda}^{(m)}\right\}.$$