CENTRE FOR APPLIED MACROECONOMIC ANALYSIS

The Australian National University

CAMA Working Paper Series August, 2011

BAYESIAN INFERENCE IN A TIME VARYING COINTEGRATION MODEL

Gary Koop
University of Strathclyde

Roberto Leon-Gonzalez
National Graduate Institute for Policy Studies

Rodney W. Strachan
Centre for Applied Macroeconomics (CAMA), ANU

CAMA Working Paper 25/2011
http://cama.anu.edu.au

Bayesian Inference in a Time Varying Cointegration Model*

Gary Koop
University of Strathclyde

Roberto Leon-Gonzalez
National Graduate Institute for Policy Studies

Rodney W. Strachan†
Centre for Applied Macroeconomic Analysis (CAMA),
Australian National University

July 27, 2011

*All authors are Fellows of the Rimini Centre for Economic Analysis. We acknowledge financial support from the Leverhulme Trust under Grant F/00 273/J. We would like to thank two anonymous referees and seminar participants at the following conferences and institutions: the 2009 Far East and South Asia Meeting of the Econometric Society, the 2009 Econometric Society European Meeting, the 2010 International Workshop on Bayesian Econometrics and Statistics (University of Tokyo), Center for Operations Research and Econometrics, University of California - Irvine, Humboldt University, Monash University, Australian National University, Keele University and the University of Iowa.

†Corresponding author: Research School of Economics, The Australian National University, Canberra, ACT, Australia. Email: rodney.strachan@anu.edu.au.

ABSTRACT

There are both theoretical and empirical reasons for believing that the parameters of macroeconomic models may vary over time. However, work with time-varying parameter models has largely involved Vector autoregressions (VARs), ignoring cointegration. This is despite the fact that cointegration plays an important role in informing macroeconomists on a range of issues. In this paper we develop a new time varying parameter model which permits cointegration. We use a specification which allows for the cointegrating space to evolve over time in a manner comparable to the random walk variation used with TVP-VARs. The properties of our approach are investigated before developing a method of posterior simulation. We use our methods in an empirical investigation involving the Fisher effect.

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

JEL Classification: C11, C32, C33

1 Introduction

There is a large amount of empirical evidence of parameter change in many macroeconomic time series (e.g. Ang and Bekaert, 2002 and Stock and Watson, 1996). When doing econometric modelling, it is important to allow for such change in order to avoid mis-specification. This raises the issue of how to appropriately model time-variation in parameters in macroeconomic models. Especially when dealing with parameter-rich multivariate time series models, such as VARs, worries about over-parameterization can arise. So the researcher faces a trade-off. If a constant parameter model is used, then mis-specification may occur. If the model is too flexible in its treatment of parameter change, then over-fitting and/or imprecise inferences can occur.

The empirical macroeconomic literature is increasingly using time-varying parameter (TVP) models which use a particular class of hierarchical priors to model variation in parameters. The use of hierarchical priors can mitigate over-parameterization worries. Consider, for instance, Cogley and Sargent (2005) and Primiceri (2005). These papers use a state space representation involving a measurement equation:

$$y_t = Z_t \gamma_t + \varepsilon_t \tag{1}$$

and a state equation

$$\gamma_t = \rho \gamma_{t-1} + \eta_t, \tag{2}$$

where y_t is an $n \times 1$ vector of observations on dependent variables, Z_t is an $n \times m$ vector of explanatory variables and γ_t an $m \times 1$ vector of states. These papers use time varying vector autoregression (TVP-VAR) methods and, thus, Z_t contains lags of the dependent variables (and appropriate deterministic terms such as intercepts). Often ρ is set to one. From a Bayesian perspective, (2) defines a hierarchical prior for the parameters.

TVP-VARs still have a large number of parameters to estimate and some researchers have argued that there is little evidence that all the VAR parameters are changing in common empirical contexts. Instead these papers argue that most evidence of parameter change relates to the error covariance matrix (e.g., Sims and Zha, 2006 and Sims, Waggoner and Zha, 2008). Accordingly, such papers have stressed the importance of trying to restrict time-variation to only some parameters. The trouble is that the researcher rarely knows, a priori, which parameters are changing and which are not.

This suggests a strategy where a model's parameters are divided into blocks and the statistical methodology decides which, if any, of these blocks of parameters exhibit time-variation. In the case of multivariate time series models where cointegration may be present, it seems natural to divide the parameters into blocks determining the cointegrating relationships, the coefficients controlling short-run dynamics and the error covariance matrix. This requires a method for modelling time-varying cointegrating relationships and helps motivate the present paper.

In addition to the strong empirical motivation for allowing for parameter change in multivariate time series models, there are also theoretical motivations. However, with TVP-VARs these tend to be fairly informal (e.g. it is common to argue informally that financial liberalization or changes in monetary policy can cause the relationships between macroeconomic variables to alter and, thus, coefficients in a VAR should change). Many macroeconomic theories relate more formally to the concept of cointegration. For instance, Garratt, Lee, Pesaran and Shin (2003) use the purchasing power parity relationship, an interest rate parity condition, a neoclassical growth model, the Fisher hypothesis and a theory of portfolio balance to build a macroeconometric model involving five cointegrating relationships. Many macroeconomists find such approaches attractive since they infuse the empirical modelling process with economic theory. Combining this desire for macroeconomic models influenced by economic theory with the empirical reality of parameter change suggests the need for a time varying parameter vector error correction model (TVP-VECM) comparable to the TVP-VAR. After all, it is possible that cointegrating relationships change over time in a comparable manner to VAR coefficients in a TVP-VAR. Furthermore, a finding of a time-varying cointegrating relationship will typically shed much more insight on the underlying economics than a finding that reduced form VAR coefficients have changed.

There are a large number of theoretical and empirical papers that model breaks or other forms of nonlinearity in cointegrating relationships, do cointegration work with subsamples of the data or attribute failures of cointegration tests to parameter change (see, among many others, Michael, Nobay and Peel, 1997, Quintos, 1997, Park and Hahn, 1999, Lettau and Ludvigson, 2004, Saikkonen and Choi, 2004, Andrade, Bruneau and Gregoir, 2005, Beyer, Haug and Dewald, 2009 and Bierens and Martins, 2010). All this work provides evidence of widespread empirical and theoretical interest in changing cointegrating spaces in a variety of empirical applications. However, with few exceptions (e.g. Martin, 2000, Paap and van Dijk, 2003 and Sugita, 2006), this work is non-Bayesian. And none of the existing Bayesian work involves a TVP hierarchical prior, despite the popularity of such approaches when working with TVP-VARs. The purpose of the present paper is to fill this gap in the literature and develop Bayesian methods for a TVP-VECM.

With cointegrated models there is a lack of identification. Without further restrictions, it is only the cointegrating space (i.e. the space spanned by the cointegrating vectors) that is identified. This consideration suggests that we want a model where the cointegrating space evolves over time in a manner such that the cointegrating space at time t is centered over the cointegrating space at time $t-1$ and is allowed to evolve gradually over time. Furthermore, we want a specification which allows for noninformative and informative priors with sensible properties.[1] In this paper we develop such a model.

From a statistical point of view, the issues involved in allowing for cointegrating spaces to evolve over time are closely related to those considered in the field of directional statistics (see, e.g., Mardia and Jupp, 2000). That is, in the two dimensional case, a space can be defined by an angle indicating a direction (in polar coordinates). By extending these ideas to the higher dimensional case of relevance for cointegration, we can derive analytical properties of our approach. For instance, we have said that we want the cointegrating space at time t to be centered over the cointegrating space at time $t-1$. But what does it mean for a space to be "centered over" another space? The directional statistics literature provides us formal answers to questions such as this. Thus, we can show analytically that our proposed hierarchical prior has attractive properties.

Next we derive a Markov Chain Monte Carlo (MCMC) algorithm which allows for Bayesian inference in our time varying cointegration model. This algorithm combines the Gibbs sampler for the time-invariant VECM derived in our previous work (Koop, León-González and Strachan, 2008, 2010) with a standard algorithm for state space models (Durbin and Koopman, 2002).

We then apply our methods in an empirical application involving a standard set of U.K. macroeconomic variables. This application shows how our methods can accurately estimate time-variation in the cointegration space and shows the importance of allowing for such time-variation. We relate our

[1] In this paper, we focus on the noninformative prior. The working paper version, available at http://personal.strath.ac.uk/gary.koop/, discusses informative priors.

application to debates in the literature on the degree of parameter change in empirical macroeconomic models. We find that a model which allows for time-variation in all of the model coefficients does not perform well. The best model is the one which allows for time-variation only in the error covariance matrix and the cointegrating relationships.

2 Modelling Issues

2.1 The Time Varying Cointegration Model

In a standard time series framework, cointegration is typically investigated using a VECM. To investigate cointegration relationships involving an n-vector, y_t, we write the measurement equation for our time varying cointegrating space model as a TVP-VECM for $t = 1, .., T$:

$$\Delta y_t = \alpha_t \beta_t' y_{t-1} + \sum_{h=1}^{l} \Gamma_{h,t} \Delta y_{t-h} + \Phi_t d_t + \varepsilon_t \tag{3}$$

where ε_t are independent $N(0, \Omega_t)$, the $n \times r$ matrices α_t and β_t are full rank and d_t denotes deterministic terms. The value of r determines the number of cointegrating relationships. The role of the deterministic terms are not the main focus of the theoretical derivations in this paper and, hence, we will leave these unspecified. Our empirical illustration uses just an intercept (i.e. $d_t = 1$).

Researchers in this field (see Koop, Strachan, van Dijk and Villani, 2006, Strachan, 2003, Strachan and Inder, 2004, Strachan and van Dijk, 2007 and Villani, 2000, 2005, 2006) point out that it is only the cointegrating space that is identified (not particular cointegrating vectors). Accordingly, we introduce notation for the space spanned by β, $\mathfrak{p} = sp(\beta)$ and present a method for estimating the space. In this paper, we follow Strachan and Inder (2004) by achieving identification by specifying β to be semi-orthogonal (i.e. $\beta'\beta = I$). Note that such an identifying restriction does not restrict the estimable cointegrating space. Another key result from Strachan and Inder (2004) is that a uniform prior on β will imply a uniform prior on $\mathfrak{p}$.

The TVP-VECM in (3) includes t subscripts on each of the parameters, including the cointegrating space. Thus $\mathfrak{p}_t = sp(\beta_t)$ where β_t is semi-orthogonal. In modelling the evolution of $\mathfrak{p}_t$ we adopt some simple principles.

First, the cointegrating space at time t should have a distribution which is centered over the cointegrating space at time $t-1$. Second, the change in location of $\mathfrak{p}_t$ from $\mathfrak{p}_{t-1}$ should be small, allowing for a gradual evolution of the space comparable to the gradual evolution of parameters which occurs with TVP-VAR models. Third, we should be able to express prior beliefs (including total ignorance) about the marginal distribution of the cointegrating space at time t.

The parameters $(\alpha_t, \Gamma_{1,t}, \ldots, \Gamma_{l,t}, \Phi_t)$ follow a standard state equation such as (2). No new theoretical or computational issues arise in relation to them and we will not discuss them in detail in the body of the paper. With respect to the error covariance matrix, many empirical macroeconomic papers have found this to be time-varying. Any sort of multivariate stochastic volatility model can be used for Ω_t. In this paper, we use the same specification as Primiceri (2005). Details on all these parameters are given in Appendix A.

As a digression, we note that cointegration is typically thought of as a long-term property, which might suggest a permanence which is not relevant when the cointegrating space is changing in every period. Time-varying cointegration relationships are better thought of as equilibria toward which the variables are attracted at any particular point in time but not necessarily at all points in time. These relations are slowly changing. Further details and motivation can be found in any of the classical econometric papers on time-varying cointegration such as Martins and Bierens (2010) or Saikkonen and Choi (2004).

2.2 A Hierarchical Prior for the Cointegrating Space

The question arises as to how we can derive a sensible hierarchical prior with our desired properties such as "$\mathfrak{p}_t$ is centered over $\mathfrak{p}_{t-1}$". The fact that we are achieving identification through restricting β_t to be semi-orthogonal means that we cannot have β_t evolving according to an AR(1) or random walk process in a conventional normal state space model.[2] In the directional statistics literature, strong justifications are provided for not working with regression-type models (such as the AR(1)) directly involving the polar angle as the dependent variable. See, for instance, Presnell, Morrison and Littell

[2]The working paper version of this paper, discusses time-varying cointegration using other identification schemes.

(1998) and their criticism of such models leading them to conclude they are "untenable in most situations" (page 1069). Thus, using a standard state space formulation for the cointegrating vectors identified using the orthogonality restriction is not appropriate.

In general, what we want is a state equation which permits smooth variation in the cointegrating space, not in the cointegrating vectors. This issue is important because, while any matrix of cointegrating vectors defines one unique cointegrating space, any one cointegrating space can be spanned by an infinite set of cointegrating vectors. Thus it is conceivable that the vectors could change markedly while the cointegrating space has not moved. In this case, the vectors have simply rotated within the cointegrating space. It is more likely, though, that the vectors could move significantly while the space moves very little. This provides further motivation for our approach in which we explicitly focus upon the implications for the cointegrating space when constructing the state equation.

We have so far used notation for identified cointegrating vectors: β_t is identified by imposing $\beta_t' \beta_t = I_r$. We will let β_t^* be the unrestricted matrix of cointegrating vectors (without identification imposed). These will be related to the semi-orthogonal β_t as:

$$\beta_t = \beta_t^* (\kappa_t)^{-1} \tag{4}$$

where

$$\kappa_t = (\beta_t^{*\prime} \beta_t^*)^{1/2} . \tag{5}$$

We shall show how this is a convenient parameterization to express our state equation for the cointegrating space.

Our preferred state equation for the time-variation in the cointegrating space is written in terms of $b_t^* = vec\,(\beta_t^*)$ for $t = 2, .., T$ as

$$\begin{aligned} b_t^* &= \rho b_{t-1}^* + \eta_t \\ \eta_t &\sim N(0, I_{nr}) \text{ for } t = 2, ..., T. \\ b_1^* &\sim N(0, I_{nr} \frac{1}{1-\rho^2}), \end{aligned} \tag{6}$$

where ρ is a scalar and $|\rho| < 1$. In the case where $r = 1$, Breckling (1989), Fisher (1993, Section 7.2) and Fisher and Lee (1994) have proposed this process to analyze times series of directions when $n = 2$ and Accardi, Cabrera and Watson (1987) looked at the case $n > 2$ (illustrating the properties of the process using simulation methods). The directions are given by the projected vectors β_t. As we shall see in the next section, (6) has some highly desirable properties and it is this framework (extended to allow for $r > 1$) that we will use. In particular, we can formally prove that it implies that $\mathfrak{p}_t$ is centered over $\mathfrak{p}_{t-1}$ (as well as having other attractive properties).

It is worth mentioning the importance of the restriction $|\rho| < 1$. In the TVP-VAR model it is common to specify random walk evolution for VAR parameters since this captures the idea that "the coefficients today have a distribution that is centered over last period's coefficients". This intuition does not go through to the present case where we want a state equation with the property: "the cointegrating space today has a distribution that is centered over last period's cointegrating space". As we shall see in the next section, the restriction $|\rho| < 1$ is necessary to ensure this property holds. In fact, the case where $\rho = 1$ has some undesirable properties in our case and, hence, we rule it out. To be precise, if $\rho = 1$, then b_t^* could wander far from the origin. This implies that the variation in $\mathfrak{p}_t$ would shrink until, at the limit, it imposes $\mathfrak{p}_t = \mathfrak{p}_{t-1}$. Note also that we have normalized the error covariance matrix in the state equation to the identity. As we shall see, it is ρ which controls the dispersion of the state equation (and, thus, plays a role similar to that played by σ_η^2, the variance of the error in equation (2)).

The preceding discussion shows how caution must be used when deriving statistical results when our objective is inference on spaces spanned by matrices. The locations and dispersions of β_t do not always translate directly to comparable locations and dispersions on the space $\mathfrak{p}_t$. For example, it is possible to construct simple cases where a distribution on β_t has its mode and mean at $\widetilde{\beta}_t$, while the mode or mean of the distribution on $\mathfrak{p}_t$ is in fact located upon the space orthogonal to the space of $\widetilde{\beta}_t$. The distributions we use avoid such inconsistencies.

It is also worth noting that, if we believe that certain vectors in β_t^* (or directions in $\mathfrak{p}_t$) evolve more quickly than others, we can readily accommodate this by replacing the scalar ρ with a diagonal matrix $(\widetilde{\rho} \otimes I_n)$ where $\widetilde{\rho} = diag\{\rho_1, \rho_2, ..., \rho_r\}$. Allowing $\rho_i \neq \rho_j$ will allow the different vectors to move at different speeds[3].

[3] In this more general case Propositions 1, 2 and 3 continue to hold if we simply write $(\kappa_{t-1}\widetilde{\rho}^2\kappa_{t-1})$ instead of $\rho^2\kappa_{t-1}^2$. The exact formulas in Proposition 4 would need to be adapted in a slightly different way, but the qualitative properties would remain the same. For the sake of parsimony and computational simplicity, in our empirical application we use ρ as a scalar.

Our hierarchical prior in (6) is written in terms of b_t^*, but we are interested in $\mathfrak{p}_t$. Accordingly, we work out the implications of (6) for $\mathfrak{p}_t$. We collect each of the $nr \times 1$ vectors $b_t^* = vec\left(\beta_t^*\right)$ into a single $Tnr \times 1$ vector $b^* = \left(b_1^{*\prime}, .., b_T^{*\prime}\right)'$. The conditional distribution in (6) implies that the joint distribution of b^* is normal with zero mean and the standard covariance matrix of a stationary AR(1) process.

We begin with discussion of the marginal prior distribution of $\mathfrak{p}_t = sp\left(\beta_t^*\right) = sp\left(\beta_t\right)$. To do so, we use some results from Strachan and Inder (2004), based on derivations in James (1954), on specifying priors on the cointegrating space. These were derived for the time-invariant VECM, but are useful here if we treat them as applying to a single point in time. A key result is that $b_t^* \sim N\left(0, cI_{nr}\right)$ implies a uniform distribution for β_t on the Stiefel manifold and a uniform distribution for $\mathfrak{p}_t$ on the Grassmann manifold (for any $c > 0$). It can immediately be seen from the joint distribution of b^* that the marginal distribution of any b_t^* has this form and, thus, the marginal prior distribution on $\mathfrak{p}_t$ is uniform. The previous literature emphasizes that this is a sensible noninformative prior for the cointegrating space. Note also that this prior has a compact support and, hence, even though it is uniform it is a proper prior. However, we are more interested in the properties of the distribution of $\mathfrak{p}_t$ conditionally on $\mathfrak{p}_{t-1}$ and it is to this we now turn.

Our state equation in (6) implies that b_t^* given b_{t-1}^* is multivariate normal. Thus, the conditional density of β_t^* given β_{t-1}^* is matric normal with mean $\beta_{t-1}^*\rho$ and covariance matrix I_{nr}. From the results in Chikuse (2003, Theorem 2.4.9), it follows that the distribution for $\mathfrak{p}_t$ (conditional on $\mathfrak{p}_{t-1}$) is the orthogonal projective Gaussian distribution with parameter $F_t = \beta_{t-1}\rho^2\kappa_{t-1}^2\beta_{t-1}'$, denoted by $OPG(F_t)$.

To write the density function of $\mathfrak{p}_t = sp\left(\beta_t\right)$ first note that the space $\mathfrak{p}_t$ can be represented with the orthogonal idempotent matrix $P_t = \beta_t\beta_t'$ of rank r (Chikuse 2003, p. 9). Thus, we can think of the density of $\mathfrak{p}_t$ as the density of P_t. The form of the density function for $\mathfrak{p}_t$ is given by

$$f(P_t|F_t) = \exp\left(-\frac{1}{2}tr(F_t)\right) {}_1F_1\left(\frac{n}{2}; \frac{r}{2}; \frac{1}{2}F_tP_t\right) \qquad (7)$$

where ${}_pF_q$ is a hypergeometric function of matrix argument (see Muirhead, 1982, p. 258).

Proposition 1 *Since $\mathfrak{p}_t = sp\left(\beta_t\right)$ follows an $OPG(F_t)$ distribution with $F_t = \beta_{t-1}\rho^2\kappa_{t-1}^2\beta_{t-1}'$, the density function of $\mathfrak{p}_t$ is maximized at $sp(\beta_{t-1})$.*

Proof: See Appendix B.

We have said we want a hierarchical prior which implies that the cointegrating space at time t is centered over the cointegrating space at time $t-1$. Proposition 1 establishes that our hierarchical prior has this property, in a modal sense (i.e. the mode of the conditional distribution of $\mathfrak{p}_t|\mathfrak{p}_{t-1}$ is $\mathfrak{p}_{t-1}$). In the directional statistics literature, results are often presented as relating to modes, rather than means since it is hard to define the "expected value of a space". But one way of defining this concept is given in Villani (2006). Larsson and Villani (2001) provide a strong case that the Frobenius norm should be used (as opposed to the Euclidean norm) to measure the distance between cointegrating spaces. Adopting our notation and using $\perp$ to denote the orthogonal complement, Larsson and Villani (2001)'s distance between $sp(\beta_t)$ and $sp(\beta_{t-1})$ is

$$d\left(\beta_t, \beta_{t-1}\right) = tr\left(\beta_t' \beta_{t-1\perp} \beta_{t-1\perp}' \beta_t\right)^{1/2}. \tag{8}$$

Using this measure, Villani (2006) defines a location measure for spaces such as $\mathfrak{p}_t = sp\left(\beta_t\right)$ by first defining

$$\overline{\beta}_t = \arg\min_{\overline{\beta}_t} E\left[d^2\left(\beta_t, \overline{\beta}_t\right)\right]$$

then defining this location measure (which he refers to as the *mean cointegrating space*) as $\overline{\mathfrak{p}}_t = sp\left(\overline{\beta}_t\right)$. Villani proves that $\overline{\mathfrak{p}}_t$ is the space spanned by the r eigenvectors associated with the r largest eigenvalues of $E\left(\beta_t\beta_t'\right)$. See Villani (2006) and Larsson and Villani (2001) for further properties, explanation and justification. Using the notation $E\left(\mathfrak{p}_t\right) \equiv \overline{\mathfrak{p}}_t$ to denote the mean cointegrating space, we have the following proposition.

Proposition 2 *Since $\mathfrak{p}_t$ follows an $OPG(F_t)$ distribution with $F_t = \beta_{t-1}\rho^2\kappa_{t-1}^2\beta_{t-1}'$, it follows that $E_{t-1}\left(\mathfrak{p}_t\right) = sp\left(\beta_{t-1}\right) = \mathfrak{p}_{t-1}$.*

Proof: See Appendix B.

This proposition shows that the expected cointegrating space at time t is the cointegrating space at $t-1$. That is, we have $E_{t-1}\left(\mathfrak{p}_t\right) = \mathfrak{p}_{t-1}$ where the expected value is defined using Villani (2006)'s location measure. Propositions 1 and 2 prove that there are two senses in which (6) satisfies the first of our desirable principles, that the cointegrating space at time t should have a distribution which is centered over the cointegrating space at time $t-1$.

The role of the matrix $\rho^2\kappa_{t-1}^2$ is to control the concentration of the distribution of $sp(\beta_t)$ around the location $sp(\beta_{t-1})$. In line with the literature on directional statistics (e.g. Mardia and Jupp, 2000, p. 169), we say that one distribution has a higher concentration than another if the value of the density function at its mode is higher. As the next proposition shows, the value of the density function at the mode is controlled solely by the eigenvalues of $\rho^2\kappa_{t-1}^2$:

Proposition 3 *Assume* $\mathfrak{p}_t$ *follows an* $OPG(F_t)$ *distribution with* $F_t = \beta_{t-1}\rho^2\kappa_{t-1}^2\beta_{t-1}'$. *Then:*

1. *The value of the density function of* $\mathfrak{p}_t$ *at the mode depends only on the eigenvalues of* $K_t = \rho^2\kappa_{t-1}^2$.

2. *The value of the density function of* $\mathfrak{p}_t$ *at the mode tends to infinity if any of the eigenvalues of* K_t *tends to infinity.*

Proof: See Appendix B.

The eigenvalues of K_t are called concentration parameters because they alone determine the value of the density at the mode but do not affect where the mode is. If all of them are zero, which can only happen when $\rho = 0$, the distribution of $sp(\beta_t)$ conditional on $sp(\beta_{t-1})$ is uniform over the Grassmann manifold. This is the purely noninformative case. In contrast, if any of the concentration parameters tends to infinity, then the density value at the mode also goes to infinity (in the same way as the multivariate normal density modal value goes to infinity when any of the variances goes to zero).

Thus, K_t plays the role of a time-varying concentration parameter. In the case $r = 1$ the prior distribution for $K_2, ..., K_T$ is the multivariate Gamma distribution analyzed by Krishnaiah and Rao (1961). The following proposition summarizes the properties of the prior of $(K_2, ..., K_T)$ in the more general case $r \geq 1$.

Proposition 4 *Suppose* $\{\beta_t^* : t = 1, ..., T\}$ *follows the process described by (6), with* $|\rho| < 1$. *Then:*

1. *The marginal distribution of* K_t *is a Wishart distribution of dimension* r *with* n *degrees of freedom and scale matrix* $I_r\frac{\rho^2}{1-\rho^2}$.

2. $E(K_t) = I_r\frac{n\rho^2}{1-\rho^2}$

3. $E(K_t|K_{t-1},...,K_2) = \rho^2 K_{t-1} + (1-\rho^2)E(K_t)$

4. *The correlation between the* (i,j) *element of* K_t *and the* (k,l) *element of* K_{t-h} *is* 0 *unless* $i = k$ *and* $j = l$.

5. *The correlation between the* (i,j) *element of* K_t *and the* (i,j) *element of* K_{t-h} *is* ρ^{2h}.

Proof See Appendix B

In TVP-VAR models researchers typically use a constant variance for the error in the state equation. This means that, a priori, the expected change in the parameters is the same in every time period. This allows for the kind of constant, gradual evolution of parameters which often occurs in practice. Proposition 4 implies that such a property holds for our model as well. In addition, it shows that when ρ approaches one, the expected value of the concentration parameters will approach infinity.[4]

Early on in this section, we set out three desirable qualities that state equations for the time varying cointegrating space model should have. We have now established that our proposed state equations do have these properties. Propositions 1 and 2 establish that (6) implies that the cointegrating space at time t has a distribution which is centered over the cointegrating space at time $t-1$. Propositions 3 and 4 establish that (6) allows for the change in location of $\mathfrak{p}_t$ from $\mathfrak{p}_{t-1}$ to be small, thus allowing for a gradual evolution of the space comparable to the gradual evolution of parameters which occurs with TVP-VAR models. We have proved that (6) implies that the marginal prior distribution of the cointegrating space is noninformative.

2.3 Bayesian Inference in the Time Varying Cointegration Model

In this section we outline our MCMC algorithm for the time varying cointegrating space model based on (6). We have specified a state space model for

[4]Our prior is not invariant to scale. However, this limitation applies to much of the existing literature on priors in cointegration models. There do exist invariant priors in the literature (e.g., Kleibergen and van Dijk, 1994, and Strachan, 2003), however these are data dependent. Furthermore, they are not in a form that could readily be incorporated into a state space framework and do not represent the desired prior beliefs for $\mathfrak{p}_t$. We considered a range of other priors - specifically those of Geweke (1996), Kleibergen and Paap (2002), Strachan and Inder (2004), and Villani (2005) - but we found none of these are invariant to scaling.

the time varying VECM. Our parameters break into three main blocks: the error covariance matrices (Ω_t for all t), the VECM coefficients apart from the cointegrating space (i.e. $(\alpha_t, \Gamma_{1,t}, \ldots, \Gamma_{l,t}, \Phi_t)$ for all t) and the parameters characterizing the cointegrating space (i.e. β_t^* for all t). Our algorithm draws all parameters in each block jointly from the conditional posterior density given the other blocks. Standard algorithms exist for providing MCMC draws from all of the blocks and, hence, we will only briefly describe them here. We adopt the specification of Primiceri (2005) for Ω_t and use his algorithm for producing MCMC draws from the posterior of Ω_t conditional on the other parameters. For $(\alpha_t, \Gamma_{1,t}, \ldots, \Gamma_{l,t}, \Phi_t)$ standard algorithms for linear normal state space models exist which can be used to produce MCMC draws from its conditional posterior. We use the algorithm of Durbin and Koopman (2002) which is a multi-move sampler. For the third block of parameters relating to the cointegrating space, we use the parameter augmented Gibbs sampler (see van Dyk and Meng, 2001) developed in Koop, León-González and Strachan (2010) and the reader is referred to that paper for further details. The structure of this algorithm can be explained by noting that we can replace $\alpha_t \beta_t'$ in (3) by $\alpha_t^* \beta_t^{*\prime}$ where $\alpha_t^* = \alpha_t \kappa_t^{-1}$ and $\beta_t^* = \beta_t \kappa_t$ where κ_t is a $r \times r$ symmetric positive definite matrix. Note that κ_t is not identified in the likelihood function but the prior we use for β_t^* implies that κ_t has a proper prior distribution[5] and, thus, is identified under the posterior. Even though β_t is semi-orthogonal, Koop, León-González and Strachan (2010) show that the posterior for β_t^* has a normal distribution (conditional on the other parameters). Thus, β_t^* can be drawn using any of the standard algorithms for linear normal state space models, and we use the algorithm of Durbin and Koopman (2002). Then, if desired, the draws of β_t^* can be transformed into draws of β_t or any feature of the cointegrating space. In the traditional VECM, Koop, León-González and Strachan (2010) provide evidence that this algorithm is very efficient relative to other methods (e.g. Metropolis-Hastings algorithms) and significantly simplifies the implementation of Bayesian cointegration analysis.

Finally, if (6) is treated as a prior then the researcher can simply select a value for ρ. However, if it is a hierarchical prior and ρ is treated as an unknown parameter, it is simple to add one block to the MCMC algorithm and draw it. In our empirical work, we use a Metropolis–within-Gibbs step

[5] The properties of the prior distribution of κ_t are easily derived from those of the prior for $K_t = \rho^2 \kappa_{t-1}$, which are described in Proposition 4.

for this parameter. Further details on the prior distribution and posterior computations are provided in Appendix A.

3 Application: The Fisher Effect

We illustrate our methods using a bivariate example[6] ($n = 2$) involving a short term interest rate, r_t, and inflation rate, π_t, in an investigation of the Fisher effect. In this case, $y_t = (r_t, \pi_t)'$. We use quarterly UK data.[7]

The unrestricted TVP-VECM is given in (3). We allow for multivariate stochastic volatility and all the parameters in a_t (where $A_t = (\alpha_t^*, \Gamma_{1,t}, \ldots, \Gamma_{l,t}, \Phi_t)$ and $a_t = vec(A_t)$) evolve according to random walks. Appendix A provides complete details.

We consider variants of this model that restrict one of Ω_t, β_t or a_t to be constant. To denote our models, we introduce three indicators (d_Ω, d_β, d_a) which take value one whenever the corresponding element of (Ω_t,β_t,a_t) is time-varying, and zero when it is time-invariant. We consider models with different cointegrating ranks ($r = 0, .., 2$). The case $r = 0$ leads to a TVP-VAR (or VAR) with differenced data and, in this case, we use notation where $d_\Pi = 0/1$ if the VAR coefficients are constant/time-varying.

We fix the number of lags at 1 (i.e. $l = 1$) and d_t includes just a constant (i.e. $d_t = 1$)[8]. The prior for ρ is uniform over the interval $(0.999, 1)$.[9] Details of the priors for other parameters can be found in Appendix A.

We begin with some evidence on the efficiency of our posterior simulation algorithm. For the model with $(d_\Omega, d_\beta, d_a) = (1, 1, 1)$ our computer[10] produces 1000 iterations of the MCMC algorithm in about 45 seconds. As we

[6] In the working paper version we also provide an example with $n = 3$.

[7] The raw data covers the period 1957Q4 through 2009Q3. The data were obtained from the IMF database on International Financial Statistics. The inflation rate is calculated as the quarterly change in the log CPI, and is annualized by multiplying the change by 400. The short term rate is a Treasury bill rate (tender rate at which 91-day bills are allotted, IFS 11260C).

[8] The BIC criterion applied to time-invariant models of various rank values favored models with either $l = 1$ or $l = 3$. For simplicity we report only results corresponding to $l = 1$.

[9] The working paper version of this paper contains a prior simulation results which show how this interval covers the reasonable range of values (e.g. we showed that $\rho = 0.99$ allows for unrealistically huge changes in the cointegrating space).

[10] We are using GAUSS software in Windows XP with a standard Hewlett- Packard desktop (E8600 @ 3.33 GHz, 1.96 GHz and 3.46 GB of RAM).

shall see below, the model with $(d_\Omega, d_\beta, d_a) = (1, 1, 0)$ is our preferred choice and we use this model to measure the efficiency of our algorithm, using effective sample size (ESS). ESS is the number of *independent* draws from the posterior that would give the same amount of information as one iteration from the algorithm (e.g. Liu (2001, p. 126)). After a burn-in of a 1000 iterations, we use 36000 consecutive iterations. We find ESS to take a value of one (maximum efficiency) for most of the linearly normalized coefficients of the cointegrating vectors (B_t)[11], regardless of whether t is at the beginning, middle or end of the sample. The ESS value for $\ln(\det(\Omega_t))$ takes values in the range 0.23 - 0.28. The ESS for (linearly normalized) α_t coefficients is in the range $(0.054 - 0.15)$. In the case of ρ, ESS is 0.0046. In summary, our posterior sample is equivalent to at least 166 and at most 36000 independent draws.

Tables 1 and 2 show for each model the log predictive likelihood for the last 100 observations[12]. A predictive likelihood is a predictive density evaluated at the realized outcome (see Geweke, 1996 or Geweke and Amisano 2010). The Fisher hypothesis implies there should be one cointegrating relationship and, under the linear normalization of the cointegrating vectors, the single cointegrating coefficient is -1 (i.e. if we normalize as $(1, B_t)'$ then $B_t = -1$). The model with the highest predictive likelihood is (d_Ω, d_β, d_a) $= (1, 1, 0)$ with $r = 1$, which implies that there is time-varying cointegration and stochastic volatility (i.e. only α_t is restricted to be time-invariant). This is consistent with one implication of the Fisher hypothesis. The following results are based on this model, although we could also have done Bayesian model averaging (i.e. average across models using predictive likelihoods to construct weights).

Let $\beta_t^* = (\beta_{t1}^*, \beta_{t2}^*)'$ and define $B_t = \beta_{t2}^*/\beta_{t1}^*$. Figure 1 plots the posterior median of B_t and a 90% credible interval. The interval contains -1, the

[11] This parameter is constructed ex post. By that we mean for $n = 2$, $r = 1$ we draw $\beta_t^* = (\beta_{t1}^*, \beta_{t2}^*)'$ from the posterior, and then construct a draw of B_t as $B_t = \beta_{t2}^*/\beta_{t1}^*$.

[12] In our case this refers to $\sum_{t=T-100+1}^{T} \ln(p(\Delta y_t | y_{t-1}, ..., y_1))$. Each component of this sum was calculated by approximating $p(\Delta y_t | y_{t-1}, ..., y_1)$ with an average of $p(\Delta y_t | \Omega_t, \beta_t, a_t, y_{t-1}, ..., y_1)$ over 7000 draws of the posterior $p(\Omega_t, \beta_t, a_t | y_{t-1}, ..., y_1)$. We used our algorithm to obtain 7000 draws from $(\Omega_{t-1}, \beta_{t-1}, a_{t-1} | y_{t-1}, ..., y_1)$. For each of these we generated (Ω_t, β_t, a_t) using the state equation $(\Omega_t, \beta_t, a_t) | (\Omega_{t-1}, \beta_{t-1}, a_{t-1})$. We used a burn-in of 1000 iterations each time. Numerical standard errors for each component of the sum were calculated using the method of Chib (1995).

value implied by the Fisher hypothesis, 34% of the time. 60% of the time the 90% credible interval lies entirely below -1. This illustrates one important feature of a time-varying cointegration model such as ours: instead of finding support for or against an economic hypothesis, we can conclude that it holds at some points in time, but not others.[13]

	$n = 2$	
(d_Ω, d_β, d_a)	$r = 1$	$r = 2$
(1,1,1)	-397.6	-423.8
	(0.66)	(0.93)
(1,1,0)	-381.6	-384.1
	(0.61)	(0.58)
(0,0,1)	-427.5	-431.0
	(0.14)	(0.97)
(0,1,1)	-424.9	-434.1
	(0.12)	(1.10)
(1,0,0)	-390.1	-386.5
	(0.86)	(1.11)
(1,0,1)	-392.2	-412.8
	(0.72)	(1.21)
(0,1,0)	-421.4	-423.3
	(0.08)	(0.12)
(0,0,0)	-437.1	-436.9
	(0.05)	(0.05)

Table 1: Predictive likelihoods for $r \geq 1$. Numerical standard errors in brackets.

Finally, note that $r = 0$ and $r = 2$ both lead to TVP-VARs. We never find evidence in favor of either of these cases. For $r = 1$ there is always strong evidence that the cointegration space is varying over time. Thus the extension to the TVP-VECM is empirically warranted.

Note also that one of our main findings (i.e. that the best model allows for time-variation only in the cointegrating space and error covariance matrix), is consistent with the findings of papers such as Sims and Zha (2006) and Sims, Waggoner and Zha (2008). Although the empirical applications are

[13]The working paper version of this paper contains additional empirical results, in particular for α_t and ρ.

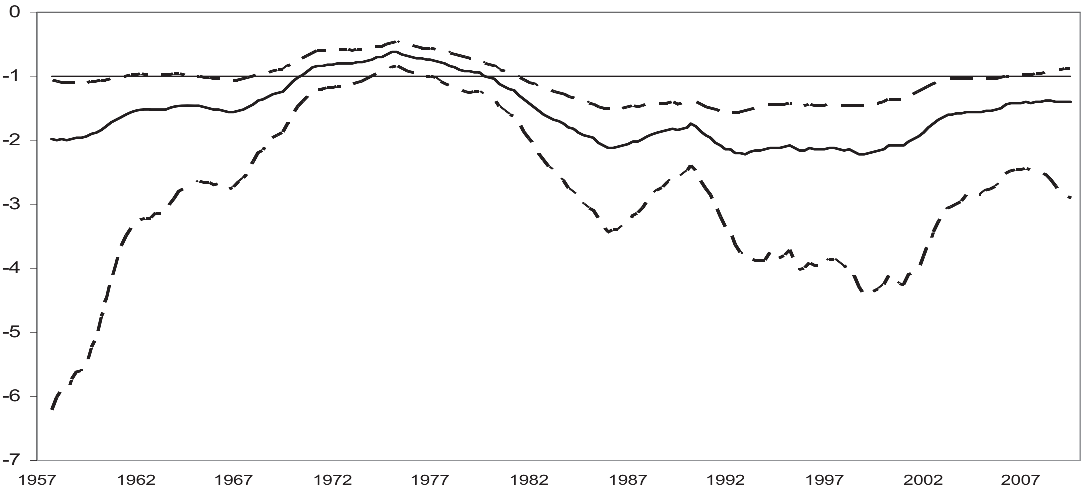

Figure 1: Posterior median and 90% credible interval for B_t when $n = 2$,$r = 1$ and $(d_\Omega,\ d_\beta,\ d_a) = (1, 1, 0)$. A horizontal line at -1 has been added to aid visualization.

(d_Ω, d_Π)	$n = 2$
(0,0)	-442.5 (0.04)
(0,1)	-443.2 (0.05)
(1,0)	-398.0 (0.62)
(1,1)	-403.2 (0.59)

Table 2: Predictive likelihoods for rank equal 0. Numerical standard errors in brackets.

different, all of these papers suggest the importance of a statistical methodology which allows for time-variation only in some important parameters of the model. In all of these empirical exercises, allowing for time-variation in the many coefficients controlling the short-run dynamics of the model seems unimportant.

4 Conclusion

TVP-VARs have become very popular in empirical macroeconomics. In this paper, we have extended such models to allow for cointegration. However, we have argued that such an extension cannot simply involve adding an extra set of random walk or AR(1) state equations for identified cointegrating vectors. Instead, we have developed a model where the cointegrating space itself evolves over time in a manner which is analogous to the random walk variation used with TVP-VARs. That is, we have developed a state space model which implies that the expected value of the cointegrating space at time t equals the cointegrating space at time $t-1$. Using methods from the directional statistics literature, we prove this property and other desirable properties of our time varying cointegrating space model.

Posterior simulation can be carried out in the time varying cointegrating space model by combining standard state space algorithms with an algorithm adapted from our previous work with standard (time invariant) VECMs. We also carry out an empirical investigation on a small system of variables commonly used in studies of inflation and monetary policy. We find strong evidence of time-varying cointegration and illustrate the benefit of our approach relative to conventional approaches such as the TVP-VAR or the constant-coefficient VECM.

References

Accardi, L., Cabrera, J. and Watson, G., 1987, Some stationary Markov processes in discrete time for unit vectors, *Metron*, XLV, 115-133.

Andrade, P., Bruneau, C. and Gregoir, S., 2005, Testing for the cointegration rank when some cointegrating directions are changing, *Journal of Econometrics*, 124, 269-310.

Ang, A. and Bekaert, G., 2002, Regime switches in interest rates, *Journal of Business and Economic Statistics*, 20, 163-182.

Beyer, A., Haug, A. and Dewald, W., 2009, Structural breaks, cointegration and the Fisher effect, ECB working paper 1013.

Bierens, H. and Martins, L., 2010, Time varying cointegration, *Econometric Theory*, 26, 1453-1490.

Breckling, J., 1989, *The Analysis of Directional Time Series: Applications to Wind Speed and Direction*, London: Springer.

Chib, S. 1995, Marginal likelihood from the Gibbs Output, *Journal of the American Statistical Association*, 90, 1313-1321.

Chikuse, Y., 2003, Statistics on special manifolds, volume 174 of *Lecture Notes in Statistics*, Springer-Verlag, New York.

Chikuse, Y. 2006, State space models on special manifolds, *Journal of Multivariate Analysis* 97, 1284-1294.

Cogley, T. and Sargent, T., 2005, Drifts and volatilities: Monetary policies and outcomes in the post WWII U.S, *Review of Economic Dynamics*, 8, 262-302.

Downs, T.D., 1972, Orientation statistics, *Biometrika* 59, 665–676.

Durbin, J. and Koopman, S., 2002, A simple and efficient simulation smoother for state space time series analysis, *Biometrika*, 89, 603-616.

Fisher, N., 1993, *Statistical Analysis of Circular Data*, Cambridge: Cambridge University Press.

Fisher, N. and Lee, A., 1992, Regression models for an angular response, *Biometrics*, 48, 665-677.

Fisher, N. and Lee, A., 1994, Time series analysis of circular data, *Journal of the Royal Statistical Society, Series B*, 56, 327–339.

Garratt, A., Lee, K., Pesaran, M.H. and Shin, Y., 2003, A Long-run structural macroeconometric model of the UK economy, *Economic Journal*, 113, 412-455.

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

Geweke, J. and Amisano, G., 2010, Hierarchical Markov normal mixture models with applications to financial asset returns, *Journal of Applied Econometrics,* forthcoming.

Godsil, C.D., and Royle, G., 2004, *Algebraic Graph Theory,* New York: Springer

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

James, A.T., 1964, Distributions of matrix variates and latent roots derived from normal samples, *Annals of Mathematical Statistics*, 35, 475-501.

Khatri, C.G. and Mardia K.V., 1977, The Von Mises-Fisher matrix distribution in orientation statistics, *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 39, No. 1., pp. 95-106.

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

Kim, S., Shephard, N. and Chib, S., 1998, Stochastic volatility: likelihood inference and comparison with ARCH models, *Review of Economic Studies*, 65, 361-93.

Kleibergen, F. and Paap, R., 2002, Prior, posteriors and bayes factors for a Bayesian analysis of cointegration, *Journal of Econometrics*, 111, 223-249.

Koop, G., León-González, R. and Strachan, R., 2010, Efficient posterior simulation for cointegrated models with priors on the cointegration space, *Econometric Reviews*, 29, 224-242.

Koop G., León-González, R. and Strachan R., 2008, Bayesian inference in a cointegrating panel data model, *Advances in Econometrics*, 23, 433-469.

Koop, G., Strachan, R., van Dijk, H. and Villani, M., 2006, Bayesian approaches to cointegration, Chapter 25 in *The Palgrave Handbook of Econometrics, Volume 1: Theoretical Econometrics* edited by T. Mills and K. Patterson. Basingstoke: Palgrave-Macmillan.

Krishnaiah, P. and Rao, M., 1961, Remarks on a multivariate Gamma distribution, *The American Mathematical Monthly*, 68, 342-346.

Lansing, K.J. 2006a, Will moderating growth reduce inflation? *FRBSF Economic Letter* 2006-37 (December 22).

Lansing, K.J. 2006b, Time-varying U.S. inflation dynamics and the new Keynesian Phillips curve, *FRBSF Working Paper* 2006-15.

Lettau, M. and Ludvigson, S., 2004, Understanding trend and cycle in asset values: Reevaluating the wealth effect on consumption, *American Economic Review*, 94, 276-299.

Liu, J.S., 2001, *Monte Carlo Strategies in Scientific Computing*, Springer. New York.

Liu, J. and Wu, Y., 1999, Parameter expansion for data augmentation, *Journal of the American Statistical Association*, 94, 1264-1274.

Mardia, K. and Jupp, P., 2000, *Directional Statistics*, Chichester: Wiley.

Martin, G., 2000, US deficit sustainability: A new approach based on multiple endogenous breaks, *Journal of Applied Econometrics*, 15, 83-105.

Michael, P., Nobay, A. and Peel, D., 1997, Transactions costs and nonlinear adjustment in real exchange rates: An empirical investigation, *Journal of Political Economy*, 105, 862-879.

Muirhead, R.J., 1982, *Aspects of Multivariate Statistical Theory.* New York: Wiley.

Paap, R., and van Dijk. H., 2003, Bayes estimates of Markov trends in possibly cointegrated series: An application to U.S. consumption and income, *Journal of Business Economics and Statistics*, 21, 547-563.

Park, J., and Hahn, H., 1999, Cointegrating regressions with time varying coefficients, *Econometric Theory*, 15, 664-703.

Presnell, B., Morrison, S. and Littell, R., 1998, Projected multivariate linear models for directional data, *Journal of the American Statistical Association*, 93, 1068-1077.

Primiceri. G., 2005, Time varying structural vector autoregressions and monetary policy, *Review of Economic Studies*, 72, 821-852.

Quintos, C.E., 1997, Stability tests in error correction models, *Journal of Econometrics*, 82, 289-315.

Rao, C.R. and Rao, M.B. 1998, *Matrix Algebra and its Applications to Statistics and Econometrics.* Singapore: World Scientific.

Saikkonen, P. and Choi, I., 2004, Cointegrating smooth transition regressions, *Econometric Theory*, 20, 301—340.

Sims, C., Waggoner, D. and Zha, T., 2008, Methods for inference in large multiple-equation Markov-switching models, *Journal of Econometrics*, 146, 255–247.

Sims, C. and Zha, T., 2006. Were there regime switches in macroeconomic policy? American Economic Review 96, 54-81.

Stock, J. and Watson, M., 1996, Evidence on structural instability in macroeconomic time series relations, *Journal of Business and Economic Statistics*, 14, 11-30.

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

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

Strachan, R. and van Dijk, H., 2007, Valuing structure, model uncertainty and model averaging in vector autoregressive processes, Econometric Institute Report EI 2007-11, Erasmus University Rotterdam.

Sugita, K., 2006, Bayesian analysis of dynamic multivariate models with multiple structural breaks, Discussion paper No 2006-14, Graduate School of Economics, Hitotsubashi University.

van Dyk, D.A., and Meng, X.L., 2001, The art of data augmentation, *Journal of Computational and Graphical Statistics*, 10, 1-111, (with discussion).

Villani, M., 2000, *Aspects of Bayesian Cointegration.* PhD thesis, University of Stockholm.

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

Villani, M., 2006, Bayesian point estimation of the cointegration space, *Journal of Econometrics*, 127, 645-664.

Appendix A: Posterior Computation and Prior Distributions

Drawing from the Conditional Mean Parameters Other Than Those Determining the Cointegration Space

Let us define $A_t = (\alpha_t^*, \Gamma_{1,t}, \ldots, \Gamma_{l,t}, \Phi_t)$ where $\alpha_t^* = \alpha_t \kappa_t^{-1}$ and $a_t = vec(A_t)$ and assume:

$$a_t = a_{t-1} + \zeta_t \tag{9}$$

where $\zeta_t \sim N(0, Q)$.[14] We can rewrite (3) by defining $z_t = \beta_t^{*\prime} y_{t-1}$ and $Z_t = \left(z_t', \Delta y_{t-1}', \ldots, \Delta y_{t-l}', d_t'\right)'$. Z_t is a $1 \times (k+r)$ vector where k is the number of deterministic terms plus n times the number of lags. Thus,

$$\Delta y_t = A_t Z_t + \varepsilon_t. \tag{10}$$

Vectorizing this equation gives us the form

$$\begin{aligned} \Delta y_t &= \left(Z_t' \otimes I_n\right) vec\left(A_t\right) + \varepsilon_t \\ \text{or } \Delta y_t &= x_t a_t + \varepsilon_t \end{aligned}$$

where $x_t = (Z_t' \otimes I_n)$. As we have assumed ε_t is normally distributed, the above expression gives us the linear normal form for the measurement equation for a_t (conditional on β_t^*). This measurement equation along with the state equation (9), specify a standard state space model and the method of Durbin and Koopman (2002) can be used to draw a_t.

Drawing the Parameters which Determine the Cointegration Space

As in Koop, León-González, and Strachan (2008), we use the transformations $\alpha_t^* = \alpha_t(\kappa_t)^{-1}$ and $\beta_t^* = \beta_t \kappa_t$ where κ_t is a $r \times r$ symmetric positive definite matrix. To show how β_t^* can be drawn, we rewrite (3) by defining

$$\begin{aligned} \widetilde{y}_t &= \Delta y_t - \sum_{h=1}^{l} \Gamma_{h,t} \Delta y_{t-h} - \Phi_t d_t = \alpha_t^* \beta_t^{*\prime} y_{t-1} + \varepsilon_t \\ \text{or } \widetilde{y}_t &= \widetilde{x}_t b_t^* + \varepsilon_t \end{aligned}$$

where we have used the relation $\alpha_t^* \beta_t^{*\prime} y_{t-1} = \left(y_{t-1}' \otimes \alpha_t^*\right) b_t^*$ where $b_t^* = vec\left(\beta_t^*\right)$ and the definition $\widetilde{x}_t = \left(y_{t-1}' \otimes \alpha_t^*\right)$. Again the assumption that

[14] One attractive property of this state equation is that, when combined with (6), it implies $E(\Pi_t|\Pi_{t-1}) = \rho\Pi_{t-1}$ where $\Pi_t = \alpha_t\beta_t'$. Moreover, if desired, it is straightforward to adapt this prior in such a way that $E(\Pi_t|\Pi_{t-1}) = \Pi_{t-1}$, while all calculations would remain virtually the same.

ε_t is normally distributed gives us a linear normal form for the measurement equation, this time for b_t^*. This measurement equation along with the state equation, specify a standard state space model and the method of Durbin and Koopman (2002) can be used to draw b_t^* (conditional on the other parameters in the model).

Treatment of Multivariate Stochastic Volatility

In the body of the paper, we did not fully explain our treatment of the measurement error covariance matrix, since it is unimportant for the main theoretical derivations in the paper. Here we provide details on how this is modelled.

We follow Primiceri (2005) and use a triangular reduction of the measurement error covariance, Ω_t, such that:

$$\Lambda_t \Omega_t \Lambda_t' = \Sigma_t \Sigma_t'$$

or

$$\Omega_t = \Lambda_t^{-1} \Sigma_t \Sigma_t' \left(\Lambda_t^{-1}\right)', \tag{11}$$

where Σ_t is a diagonal matrix with diagonal elements $\sigma_{j,t}$ for $j = 1, .., n$ and Λ_t is a lower triangular matrix with ones on the diagonal and lower diagonal elements $\lambda_{ij,t}$. To model evolution in Σ_t and λ_t we must specify additional state equations. For Σ_t a stochastic volatility framework can be used. In particular, if $\sigma_t = (\sigma_{1,t}, .., \sigma_{n,t})'$, $h_{i,t} = \ln(\sigma_{i,t})$, $h_t = (h_{1,t}, .., h_{n,t})'$ then Primiceri uses:

$$h_t = h_{t-1} + u_t, \tag{12}$$

where u_t is $N(0, W)$ and is independent over t and of ε_t, η_t and ζ_t.

To describe the manner in which Λ_t evolves, we first stack the unrestricted elements by rows into a $\frac{n(n-1)}{2}$ vector as $\lambda_t = \left(\lambda_{21,t}, \lambda_{31,t}, \lambda_{32,t}, .., \lambda_{n(n-1),t}\right)'$. These are allowed to evolve according to the state equation:

$$\lambda_t = \lambda_{t-1} + \xi_t, \tag{13}$$

where ξ_t is $N(0, C)$ and is independent over t and of u_t, ε_t, ζ_t and η_t.

Prior Distributions

Our model involves four sets of state equations: two associated with the measurement error covariance matrix ((12) and (13)), one for the cointegrating space given in (6) and one for the other conditional mean coefficients (9). The prior for the initial condition for the cointegrating space is already given in (6), and implies a uniform for $\mathfrak{p}_1$. We now describe the prior for initial conditions (h_0, λ_0 and a_0) and the variances of the errors in the other three state equations (W, C and Q). We also require a prior for ρ which, inspired by the prior simulation results, we set to being uniform over a range close to one: $\rho \in [0.999, 1)$.

The priors for the initial conditions are $\lambda_0 \sim N\left(0, 2I_{n(n-1)/2}\right)$, $h_0 \sim N\left(0, 2I_n\right)$ and $a_0 \sim N\left(0, 2V_{a_0}\right)$, where V_{a_0} is the identity matrix except for those diagonal elements that correspond to α_t^*, which are set to be equal to $(1-\rho^2)$. By doing this the prior variance of each of the elements of the product $\alpha_0^* \beta_0^{*\prime}$ has the desired value of 2.

We select Wishart priors for the inverse of error variances in the state equations: $Q^{-1} \sim W\left(\underline{\nu}_Q, \underline{Q}^{-1}\right)$, $W^{-1} \sim W\left(\underline{\nu}_W, \underline{W}^{-1}\right)$ and $C^{-1} \sim W\left(\underline{\nu}_C, \underline{C}^{-1}\right)$. We choose a small number for the degrees of freedom, which is equal to the dimension of the matrix plus 2 (i.e. $\underline{\nu}_Q = \dim(a_t) + 2$, $\underline{\nu}_W = n + 2$, $\underline{\nu}_C = n(n-1)/2 + 2$) and we make each of the matrices $(\underline{Q}, \underline{W}, \underline{C})$ equal to the identity matrix times 0.0001. Hence the prior mean of these matrices is small (reflecting that parameters are expected to change slowly), but the moderately large value of the prior variance (due to small degrees of freedom and small mean matrix) allows for substantially bigger values.

Remaining Details of Posterior Simulation

The blocks in our algorithm for producing draws of b_t^*, a_t have already been provided. Here we discuss the other blocks of our MCMC algorithm. In particular, we describe how to draw from the full posterior conditionals for the remaining two sets of state equations, the covariance matrices of the errors in the state equations and ρ. Since most of these involve standard algorithms, we do not provide much detail. As in Primiceri (2005), draws of λ_t can be obtained using the algorithm of Durbin and Koopman (2002) and draws of h_t using the algorithm of Kim, Shephard and Chib (1998).

The conditional posteriors for the state equation error variances begin with:

$$Q^{-1}|Data \sim W\left(\overline{\nu}_Q, \overline{Q}^{-1}\right)$$

where

$$\overline{\nu}_Q = T + \underline{\nu}_Q$$

and

$$\overline{Q}^{-1} = \left[\underline{Q} + \sum_{t=1}^{T} \left(a_t - a_{t-1}\right)\left(a_t - a_{t-1}\right)'\right]^{-1}.$$

Next we have:

$$W^{-1}|Data \sim W\left(\overline{\nu}_W, \overline{W}^{-1}\right)$$

where

$$\overline{\nu}_W = T + \underline{\nu}_W$$

and

$$\overline{W}^{-1} = \left[\underline{W} + \sum_{t=1}^{T} \left(h_t - h_{t-1}\right)\left(h_t - h_{t-1}\right)'\right]^{-1}.$$

The posterior for C^{-1} (conditional on the states) is then Wishart:

$$C^{-1}|Data \sim W\left(\overline{\nu}_C, \overline{C}^{-1}\right)$$

where

$$\overline{\nu}_C = T + \underline{\nu}_C$$

and

$$\overline{C}^{-1} = \left[\underline{C} + \sum_{t=1}^{T} \left(a_t - a_{t-1}\right)\left(a_t - a_{t-1}\right)'\right]^{-1}$$

The posterior for ρ is non-standard due to the nonlinear way in which it enters the distribution for the initial condition for b_1^* in (6). We therefore draw this scalar using a Metropolis-within-Gibbs step.

Appendix B: Proofs

Proof of Proposition 1: We will show that the density of $\mathfrak{p}_t$ conditional on $(\kappa_t, \beta_{t-1}, \kappa_{t-1})$ is maximized at $\mathfrak{p}_t = sp(\beta_{t-1})$ for any value of (κ_t, κ_{t-1}). This proves that the density of $\mathfrak{p}_t$ conditional on $(\beta_{t-1}, \kappa_{t-1})$ is also maximized at $\mathfrak{p}_t = sp(\beta_{t-1})$. Clearly, the mode is also the same if we do not condition on κ_{t-1}.

The state equation in (6) implies that the conditional density of β_t^* given β_{t-1}^* is matric normal with mean $\beta_{t-1}^*\rho$ and covariance matrix I_{nr}. Thus, using Lemma 1.5.2 in Chikuse (2003), it can be shown that the implied distribution for $\beta_t|(\kappa_t, \beta_{t-1}, \kappa_{t-1})$ is the matrix Langevin (or von Mises–Fisher) distribution denoted by $L\left(n, r; \tilde{F}\right)$ (Chikuse, 2003, p. 31), where

$$\tilde{F} = \beta_{t-1}^*\rho\kappa_t = \beta_{t-1}\kappa_{t-1}\rho\kappa_t$$

The form of the density function for $L\left(n, r; \tilde{F}\right)$ is given by

$$f_{\beta_t}(\beta_t|\tilde{F}) = \frac{\exp\left\{tr(\tilde{F}'\beta_t)\right\}}{{}_0F_1\left(\frac{n}{2}, \frac{1}{4}\tilde{F}'\tilde{F}\right)}$$

Recall that $P_t = \beta_t\beta_t'$. The density function $f_{P_t}(P_t)$ of P_t, conditional on $\tilde{F}$, can be derived from the density function $f_{\beta_t}(\beta_t)$ of β_t using Theorem 2.4.8 in Chikuse (2003, p. 46):

$$f_{P_t}(P_t) = A_L \int_{O_r} \exp(tr\tilde{F}'\beta_t Q)[dQ] = A_L \ {}_0F_1(\frac{1}{2}r; \frac{1}{4}\tilde{F}'P_t\tilde{F})$$

where A_L is a constant not depending on P_t ($A_L^{-1} = {}_0F_1(\frac{1}{2}n; \frac{1}{4}\tilde{F}'\tilde{F})$) and O_r is the orthogonal group of $r \times r$ orthogonal matrices (Chikuse (2003), p. 8). Note that we have used the integral representation of the ${}_0F_1$ hypergeometric function (Muirhead, 1982, p. 262). Khatri and Mardia (1976, p. 96) show that ${}_0F_1(\frac{1}{2}r; \frac{1}{4}\tilde{F}'P_t\tilde{F})$ is equal to ${}_0F_1(\frac{1}{2}r; \frac{1}{4}G_t)$, where $G_t = diag(g_1, ..., g_r)$ is an $r \times r$ diagonal matrix containing the singular values of $\tilde{F}'P_t\tilde{F}$. We first show that ${}_0F_1(\frac{1}{2}r; \frac{1}{4}G_t)$ is an increasing function of each of the singular values g_i, for each $i = 1, ..., r$. We then show that each of these singular values is maximized when $\beta_{t-1}'P_t\beta_{t-1} = I_r$. Note that $\beta_{t-1}'P_t\beta_{t-1} = I_r$ implies that the distance between $sp(\beta_t)$ and $sp(\beta_{t-1})$, as defined in Larsson and Villani (2001), is zero and thus $\mathfrak{p}_t = sp(\beta_{t-1})$.

We first show that the following standard expression for ${}_0F_1(\frac{1}{2}r; \frac{1}{4}G_t)$ (e.g. Muirhead, 1982, p. 262):

$${}_0F_1(\frac{1}{2}r; \frac{1}{4}G_t) = \int_{O(r)} \exp(\sum_{i=1}^{r} \sqrt{g_i}q_{ii})[dQ]$$

with $Q = \{q_{ij}\}$, is equivalent to:

$${}_0F_1(\frac{1}{2}r; \frac{1}{4}G_t) = \int_{\tilde{O}_r} \prod_{i=1}^{r}(exp(\sqrt{g_i}q_{ii}) + exp(-\sqrt{g_i}q_{ii}))[dQ] \qquad (14)$$

where $\tilde{O}(r)$ is a subset of $O(r)$ consisting of matrices $Q \in O(r)$ whose diagonal elements are positive. This equivalence can be noted by writing:

$$\int_{O(r)} \exp(\sum_{i=1}^{r} \sqrt{g_i}q_{ii})[dQ] = \int_{\{O(r):q_{11}\geq 0\}} \exp(\sum_{i=1}^{r} \sqrt{g_i}q_{ii})[dQ] + \int_{\{O(r):q_{11}<0\}} \exp(\sum_{i=1}^{r} \sqrt{g_i}q_{ii})[dQ]$$

The second integral in the sum can be rewritten by making a change of variables from Q to Z, where Z results from multiplying the first row of Q by (-1). Note that Z results from pre-multiplying Q by an orthogonal matrix and thus Z still belongs to $O(r)$ and the Jacobian is one (Muirhead, 1982, Theorem 2.1.4). Thus, the second integral in the sum can be written as:

$$\begin{aligned}\int_{\{O(r):q_{11}<0\}} \exp(\sum_{i=1}^{r}\sqrt{g_i}q_{ii})[dQ] &= \int_{\{O(r):z_{11}\geq 0\}} \exp(-\sqrt{g_i}z_{11})\exp(\sum_{i=2}^{r}\sqrt{g_i}z_{ii})[dZ] \\ &= \int_{\{O(r):q_{11}\geq 0\}} \exp(-\sqrt{g_i}q_{11})\exp(\sum_{i=2}^{r}\sqrt{g_i}q_{ii})[dQ]\end{aligned}$$

Thus:

$$\int_{O(r)} \exp(\sum_{i=1}^{r}\sqrt{g_i}q_{ii})[dQ] = \int_{\{O(r):q_{11}\geq 0\}} (\exp(\sqrt{g_i}q_{11}) + \exp(-\sqrt{g_i}q_{11}))\exp(\sum_{i=2}^{r}\sqrt{g_i}q_{ii})[dQ]$$

Doing analogous changes of variables for the other rows, we arrive at equation (14). Note that the function $\exp(cx)+\exp(-cx)$ is an increasing function of x when both x and c are positive. Thus, from expression (14), ${}_0F_1(\frac{1}{2}r;\frac{1}{4}G_t)$ is an increasing function of each of the singular values g_i, for each $i=1,...,r$.

Let us now see that each of the singular values of $\tilde{F}'P_t\tilde{F}$ is maximized when $\beta'_{t-1}P_t\beta_{t-1}=I_r$. Write $\tilde{F}=\beta_{t-1}C$, where $C=\kappa_{t-1}\rho\kappa_t$ is a $r\times r$ matrix. Let $\beta_{t\perp}$ be the orthogonal complement of β_t (i.e. $(\beta_t,\beta_{t\perp})$ is an $(n\times n)$ orthogonal matrix) and $P_{t\perp}=\beta_{t\perp}\beta'_{t\perp}$. Note that $P_t+P_{t\perp}=I_n$ (because $P_t+P_{t\perp}=(\beta_t,\beta_{t\perp})(\beta_t,\beta_{t\perp})'=I_r$). Thus, $C'\beta'_{t-1}P_t\beta_{t-1}C+C'\beta'_{t-1}P_{t\perp}\beta_{t-1}C=C'C$. Let $(a_1,..,a_r)$ be the singular values of $A=C'\beta'_{t-1}P_t\beta_{t-1}C$, with $(a_1\geq a_2\geq ...\geq a_r\geq 0)$. Similarly, let $(b_1,..,b_r)$ be the singular values of $B=C'\beta'_{t-1}P_{t\perp}\beta_{t-1}C$ (ordered also from high to low). Similarly, let $(c_1,..,c_r)$ be the singular values of $(C'C)$. Because $A,B,(C'C)$ are positive semidefinite and symmetric, eigenvalues and singular values coincide. Thus, Proposition 10.1.1 in Rao and Rao (1998, p. 322) applies, which implies that: $a_1+b_r\leq c_1, a_2+b_r\leq c_2, a_3+b_r\leq c_3$,...,$a_r+b_r\leq c_r$. Since $b_r\geq 0$ this implies $a_1\leq c_1, a_2\leq c_2, a_3\leq c_3$,...,$a_r\leq c_r$. Note that if $\beta'_{t-1}P_t\beta_{t-1}=I_r$ then $A=C'C$ and so $a_1=c_1, a_2=c_2, a_3=c_3$,...,$a_r=c_r$. Thus, each of the singular values of $\tilde{F}'P_t\tilde{F}$ is maximized when $\beta'_{t-1}P_t\beta_{t-1}=I_r$.

Proof of Proposition 2:

We will proof that $E(\mathfrak{p}_t|(\kappa_t,\beta_{t-1},\kappa_{t-1}))=sp(\beta_{t-1})$. Note that this proves that $E(\mathfrak{p}_t|(\beta_{t-1},\kappa_{t-1}))=sp(\beta_{t-1})$, because if $\bar{\beta}=\beta_{t-1}$ minimizes

$E(d^2(\beta_t, \bar{\beta})|(\kappa_t, \beta_{t-1}, \kappa_{t-1}))$ for every κ_t, it will also minimize $E(d^2(\beta_t, \bar{\beta})|(\beta_{t-1}, \kappa_{t-1}))$. Similarly, it also proves that $E(\mathfrak{p}_t|\beta_{t-1}) = sp(\beta_{t-1})$.

Recall that $\beta_t|(\kappa_t, \beta_{t-1}, \kappa_{t-1})$ follows a Langevin distribution $L\left(n, r; \tilde{F}\right)$, with $\tilde{F} = \beta^*_{t-1}\rho\kappa_t = \beta_{t-1}\kappa_{t-1}\rho\kappa_t$. Let $G = \kappa_t\rho\kappa_{t-1}$ and write G using its singular value decomposition as $G = OMP'$, where O and P are $r \times r$ orthogonal matrices, and M is an $r \times r$ diagonal matrix. Hence, $\tilde{F} = \beta_{t-1}PMO'$. Write $\tilde{F} = \beta_{t-1}PMO' = \Gamma MO'$, where $\Gamma = \beta_{t-1}P$. Since P is a $r \times r$ orthogonal matrix, $sp(\beta_{t-1}) = sp(\Gamma)$. We will prove that $E(\mathfrak{p}_t) = sp(\Gamma)$. In order to prove this, we will prove that $E(\beta_t\beta_t') = UDU'$, with $U = (\Gamma, \Gamma_\perp)$, where $\Gamma_\perp$ is the orthogonal complement of Γ, and $D = \{d_{ij}\}$ is a diagonal matrix with $d_{11} \geq d_{22} \geq ... \geq d_{nn}$. Define the $n \times r$ matrix $Z = U'\beta_t O$, so that $E(\beta_t\beta_t') = UE(ZO'OZ')U' = UE(ZZ')U'$. Let $Z = \{z_{ij}\}$.

The distribution of Z' is the same as the distribution of the matrix that Khatri and Mardia (1977) denote as Y at the bottom of page 97 of their paper. They show, in page 98, that $E(z_{ij}z_{kl}) = 0$ for all i, j, k, l except when (a) $i = k = j = l$, $i = 1, 2, ..., r$; (b) $i = k$, $j = l$ $(i \neq j)$; (c) $i = j$, $k = l$ $(i \neq k)$; (d) $i = l$, $j = k$ $(i \neq j)$, $i, j = 1, 2, ..., r$. Note that the (i, k) element of ZZ' is $\sum_{h=1}^r z_{ih}z_{kh}$. Thus, $E(ZZ')$ is a diagonal matrix and we can write $E(\beta\beta') = UDU'$, where $D = E(ZZ')$.

To finish the proof we need to show that each of the first r values in the diagonal of $D = E(ZZ')$ is at least as large as any of the other $n - r$ values. The Jacobian from β_t to Z is one (Muirhead, 1982, Theorem 2.1.4), and hence the density function of Z is:

$$A_L \exp(tr(M\tilde{Z})) = A_L \exp(\sum_{l=1}^{r}(m_l\tilde{z}_{ll}))$$

where $\tilde{Z} = \{\tilde{z}_{ij}\}$ consists of the first r rows of Z, $M = diag(m_1, ..., m_r)$ and A_L is a normalizing constant. If we let $\hat{Z} = \{\hat{z}_{ij}\}$ be the other $n - r$ rows, what needs to be proved can be written as: $E(\sum_{l=1}^r(\tilde{z}_{jl})^2) \geq E(\sum_{l=1}^r(\hat{z}_{pl})^2)$ for any j, p such that $1 \leq j \leq r$, $1 \leq p \leq n - r$.

Note that $(\sum_{l=1}^r(\tilde{z}_{jl})^2)$ is the euclidean norm of the j^{th} row of Z and similarly $\sum_{l=1}^r(\hat{z}_{pl})^2$ is the norm of the $(r+p)^{th}$ row of Z. Let S_1 be defined as the set of $n \times r$ semi-orthogonal matrices whose j^{th} row has bigger norm than the $(r+p)^{th}$ row. Let S_2 be the set of semi-orthogonal matrices where the opposite happens. Thus, $E(\sum_{l=1}^r(\tilde{z}_{jl})^2)$ can be written as the following sum of integrals:

$$A_L \int_{S_1} (\sum_{l=1}^{r} (\tilde{z}_{jl})^2) \exp\left\{\sum_{l=1}^{r} m_l \tilde{z}_{ll}\right\} [dZ] + A_L \int_{S_2} (\sum_{l=1}^{r} (\tilde{z}_{jl})^2) \exp\left\{\sum_{l=1}^{r} m_l \tilde{z}_{ll}\right\} [dZ]$$

where $[dZ]$ is the normalized invariant measure on the Stiefel manifold (e.g. Chikuse, 2003, p. 18). Now note that:

$$A_L \int_{S_2} (\sum_{l=1}^{r} (\tilde{z}_{jl})^2) \exp\left\{\sum_{l=1}^{r} m_l \tilde{z}_{ll}\right\} [dZ] = A_L \int_{S_1} (\sum_{l=1}^{r} (\hat{z}_{pl})^2) \exp\left\{m_j \hat{z}_{pj} + \sum_{l=1, l\neq j}^{r} m_l \tilde{z}_{ll}\right\} [dZ]$$

This equality can be obtained by making a change of variables from Z to Q where Q results from swapping the j^{th} and $(r+p)^{th}$ rows of Z. Note that Q is also semi-orthogonal, and that because the transformation involves simply swapping the position of variables, the Jacobian is one. Thus, $E(\sum_{l=1}^{r} (\tilde{z}_{jl})^2)$ can be written as:

$$A_L \int_{S_1} (\sum_{l=1}^{r} (\tilde{z}_{jl})^2) \exp\left\{\sum_{l=1}^{r} m_l \tilde{z}_{ll}\right\} [dZ] + A_L \int_{S_1} (\sum_{l=1}^{r} (\hat{z}_{pl})^2) \exp\left\{m_j \hat{z}_{pj} + \sum_{l=1, l\neq j}^{r} m_l \tilde{z}_{ll}\right\} [dZ]$$

Similarly, $E(\sum_{l=1}^{r} (\hat{z}_{pl})^2)$ can be written as:

$$A_L \int_{S_1} (\sum_{l=1}^{r} (\hat{z}_{pl})^2) \exp\left\{\sum_{l=1}^{r} m_l \tilde{z}_{ll}\right\} [dZ] + A_L \int_{S_1} (\sum_{l=1}^{r} (\tilde{z}_{jl})^2) \exp\left\{m_j \hat{z}_{pj} + \sum_{l=1, l\neq j}^{r} m_l \tilde{z}_{ll}\right\} [dZ]$$

Thus, $E(\sum_{l=1}^{r} (\tilde{z}_{jl})^2) - E(\sum_{l=1}^{r} (\hat{z}_{pl})^2)$ is equal to:

$$A_L \int_{S_1} (\sum_{l=1}^{r} (\tilde{z}_{jl})^2 - \sum_{l=1}^{r} (\hat{z}_{jl})^2) \left(\exp\left\{\sum_{l=1}^{r} m_l \tilde{z}_{ll}\right\} - \exp\left\{m_j \hat{z}_{pj} + \sum_{l=1, l\neq j}^{r} m_l \tilde{z}_{ll}\right\}\right) [dZ]$$

Following Chikuse (2003, p. 17), we can make a change of variables $Z = WN$, where W is a $n \times r$ semi-orthogonal matrix that represents an element in the Grassmann manifold, and N is an $r \times r$ orthogonal matrix. That is, W is seen as an element of the Grassmann manifold of planes ($G_{r,n-r}$) and N is an element of the orthogonal group of $r \times r$ orthogonal matrices ($O(r)$). The measure $[dZ]$ can be written as $[dZ] = [dW][dN]$, where $[dN]$ is the normalized invariant measure in $O(r)$ and $[dW]$ is another normalized measure whose expression can be found in Chikuse (2003, p. 15). Let the first r rows of W be denoted as $\tilde{W} = \{\tilde{w}_{ij}\}$ and the other rows as $\hat{W} = \{\hat{w}_{ij}\}$. Note that the norm of a row of Z is equal to the norm of the corresponding row of W, because N is orthogonal (e.g. $\sum_{l=1}^{r}(\tilde{z}_{jl})^2 = \sum_{l=1}^{r}(\tilde{w}_{jl})^2$, which is a consequence of $ZZ' = WW'$). Define $\bar{W}$ as a matrix that is equal to $\tilde{W}$ for all rows except for the j^{th} one. Let the j^{th} row of $\bar{W}$ be equal to the $(r+p)^{th}$ row of W. Note that $m_j\hat{z}_{pj} + \sum_{l=1,l\neq j}^{r} m_l\tilde{z}_{ll} = tr(M\bar{W}N)$. Thus, $E(\sum_{l=1}^{r}(\tilde{z}_{jl})^2) - E(\sum_{l=1}^{r}(\hat{z}_{pl})^2)$ can be written as:

$$\begin{aligned}&A_L \int_{S_1} \left(\sum_{l=1}^{r}(\tilde{w}_{jl})^2 - \sum_{l=1}^{r}(\hat{w}_{pl})^2\right)\left(\exp\left\{tr(M\tilde{W}N)\right\} - \exp\left\{tr(M\bar{W}N)\right\}\right)[dW][dN] = \\ &A_L \int_{S_1} \left(\sum_{l=1}^{r}(\tilde{w}_{jl})^2 - \sum_{l=1}^{r}(\hat{w}_{pl})^2\right)\left({}_0F_1(\tfrac{1}{2}r; \tfrac{1}{4}M\tilde{W}\tilde{W}'M) - {}_0F_1(\tfrac{1}{2}r; \tfrac{1}{4}M\bar{W}\bar{W}'M)\right)[dW]\end{aligned} \tag{15}$$

where we have used the integral representation of the hypergeometric function (e.g. Muirhead, 1982, p. 262). As noted by Khatri and Mardia (1979, p. 96), ${}_0F_1(\frac{1}{2}r; M\tilde{W}\tilde{W}'M)$ is a function only of the singular values of $M\tilde{W}\tilde{W}'M$. In addition, as we argued when we found the mode of the distribution, ${}_0F_1(\frac{1}{2}r; \frac{1}{4}M\tilde{W}\tilde{W}'M)$ increases with each of the singular values of $M\tilde{W}\tilde{W}'M$. Let $A = M\tilde{W}\tilde{W}'M$ and $B = M\bar{W}\bar{W}'M$. Let the singular values of A be $(a_1, ..., a_r)$ with $(a_i \geq a_{i+1})$ and let the singular values of B be $(b_1, ..., b_r)$, with $b_i \geq b_{i+1}$. From now we will show that in the region S_1, $a_i \geq b_i$, $i = 1, ..., r$. Note that this implies ${}_0F_1(\frac{1}{2}r; \frac{1}{4}A) \geq {}_0F_1(\frac{1}{2}r; \frac{1}{4}B)$ in S_1, and thus the integral in (15) is not negative, so that $E(\sum_{l=1}^{r}(\tilde{z}_{jl})^2) \geq E(\sum_{l=1}^{r}(\hat{z}_{pl})^2)$.

Define the matrix $C = A - B = M(\tilde{W}\tilde{W}' - \bar{W}\bar{W}')M$. Note that all elements in C are zero except for those that are either in the j^{th} row or in the j^{th} column. Thus, C has rank equal to one. Hence, all its singular values are zero, except for one. Because C is symmetric, the sum of its singular values is equal to the trace. Because C has only one non-zero diagonal element, we get that the only non-zero singular value of C is equal to the (j, j) element of C. Note that this element is equal to $c_1 = m_{jj}^2(\sum_{l=1}^{r}(\tilde{w}_{jl})^2 - \sum_{l=1}^{r}(\hat{w}_{pl})^2)$ which is positive in S_1. Let the singular values of C, ordered from high to low, be $(c_1, ..., c_r)$, with ($c_i = 0$ for $2 \leq i \leq r$). Note that A and B are positive definite and symmetric, and thus their singular values are equal to their eigenvalues. Note also that C is positive semidefinite in S_1. Thus, we can write $B + C = A$ and apply Proposition 10.1.1 in Rao and Rao (1998, p. 322), which implies that: $b_1 + c_r \leq a_1, b_2 + c_r \leq a_2, b_3 + c_r \leq a_3, ..., b_r + c_r \leq a_r$. Since $c_r = 0$ this implies $b_1 \leq a_1, b_2 \leq a_2, b_3 \leq a_3, ..., b_r \leq a_r$. Thus, ${}_0F_1(\frac{1}{2}r; \frac{1}{4}A) \geq {}_0F_1(\frac{1}{2}r; \frac{1}{4}B)$ in S_1, and integral (15) is not negative.

Proof of Proposition 3: The density function of $P_t = \beta_t\beta_t'$ given by expression (7) depends on ${}_1F_1\left(\frac{n}{2};\frac{r}{2};\frac{1}{2}F_tP_t\right)$, which is equal to: ${}_1F_1\left(\frac{n}{2};\frac{r}{2};\frac{1}{2}\rho\kappa_{t-1}\beta_{t-1}'P_t\beta_{t-1}\kappa_{t-1}\rho\right)$. To see that these two hypergeometric functions are equal, we can write ${}_1F_1\left(\frac{n}{2};\frac{r}{2};\frac{1}{2}F_tP_t\right)$ in terms of zonal polynomials as (e.g. Muirhead, 1982, p. 258):

$$
{}_1F_1\left(\frac{n}{2};\frac{r}{2};\frac{1}{2}F_tP_t\right)=\sum_{k=0}^{\infty}\sum_{\kappa}\frac{(n/2)_\kappa}{(r/2)_\kappa}\frac{C_\kappa(\frac{1}{2}F_tP_t)}{k!}
$$

where κ is a partition of k into as many terms as the dimension of F_tP_t. That is, $\kappa=(k_1,...,k_n)$, $k=k_1+...+k_n$, $k_1\geq...\geq k_n\geq 0$. $\sum_\kappa$ denotes summation over all possible partitions κ of k. C_κ is a zonal polynomial and $(n/2)_\kappa,(r/2)_\kappa$ are generalized hypergeometric coefficients whose definition can be found in Muirhead (1982, p. 258, expression 2). The zonal polynomial $C_\kappa(F_tP_t)$ depends on F_tP_t only through its nonzero eigenvalues (James, 1964, pp. 478-479). Zonal polynomials are usually expressed in terms of symmetric matrices, so that $C_\kappa(F_tP_t)$ can be written as $C_\kappa(S)$, where S is a $n\times n$ symmetric matrix with the same eigenvalues as F_tP_t. Note that for any two matrices $A: r\times n$, $B: n\times r$, (AB) and (BA) have the same nonzero eigenvalues with the same multiplicities (e.g. Godsil and Royle, 2001, Lemma 8.2.4). Thus, $C_\kappa(F_tP_t)=C_\kappa(\rho\kappa_{t-1}\beta_{t-1}'P_t\beta_{t-1}\kappa_{t-1}\rho)$ for $\kappa=(k_1,...,k_n)$. Note that because the matrix $(\rho\kappa_{t-1}\beta_{t-1}'P_t\beta_{t-1}\kappa_{t-1}\rho)$ has dimension r and full rank, $C_\kappa(\rho\kappa_{t-1}\beta_{t-1}'P_t\beta_{t-1}\kappa_{t-1}\rho)=0$ if $k_{r+1}\neq 0$ (James (1964), p. 478). This shows:

$$
{}_1F_1\left(\frac{n}{2};\frac{r}{2};\frac{1}{2}F_tP_t\right)={}_1F_1\left(\frac{n}{2};\frac{r}{2};\frac{1}{2}\rho\kappa_{t-1}\beta_{t-1}'P_t\beta_{t-1}\kappa_{t-1}\rho\right)
$$

Thus, the density function of $P_t=\beta_t\beta_t'$ evaluated at the mode $P_t=\beta_{t-1}\beta_{t-1}'$ is:

$$
\exp\left(-\frac{1}{2}tr(F_t)\right){}_1F_1\left(\frac{n}{2};\frac{r}{2};\frac{1}{2}K_t\right)
$$

Since $F_t=\beta_{t-1}\rho^2\kappa_{t-1}^2\beta_{t-1}'$, from the properties of the trace function, $tr(F_t)=tr(\rho^2\kappa_{t-1}^2\beta_{t-1}'\beta_{t-1})=tr(\rho^2\kappa_{t-1}^2)=tr(K_t)$, which depends only on the eigenvalues of K_t. In addition, as argued before, a hypergeometric function of matrix argument K_t depends on K_t only via its eigenvalues. Thus, the value of the density at the mode depends on K_t only through its eigenvalues.

Let $D=diag(d_1,...,d_r)$ be a diagonal matrix containing the eigenvalues of K_t, so the value of the mode can be written as: $\exp(-\frac{1}{2}tr(D))\,{}_1F_1(n/2;r/2;1/2D)$. Following the result in Chikuse (2003, p. 317), when d_i is large, ${}_1F_1(n/2;r/2;1/2D)$

can be written as:

$$\frac{\Gamma(r/2)}{\Gamma(n/2)}\exp(\frac{1}{2}d_i)\left(\frac{d_i}{2}\right)^{\frac{(n-r)}{2}} {}_1F_1(\frac{(n-1)}{2};\frac{(r-1)}{2};\frac{1}{2}D_{-i})\left[1+O(d_{ii}^{-1})+O(d_{ii}^{-2})\right]$$

where $\Gamma(.)$ is the Gamma function and D_{-i} is a $(r-1)\times(r-1)$ diagonal matrix containing the diagonal elements of D except for d_i. Thus, the limit of the mode when d_i tends to infinity is the same as the limit of the following expression:

$$\frac{\Gamma(r/2)}{\Gamma(n/2)}\exp(-\frac{1}{2}tr(D_{-i}))\left(\frac{d_i}{2}\right)^{\frac{(n-r)}{2}} {}_1F_1(\frac{(n-1)}{2};\frac{(r-1)}{2};\frac{1}{2}D_{-i})\left[1+O(d_{ii}^{-1})+O(d_{ii}^{-2})\right]$$

This expression tends to infinity as d_i tends to infinity.

Proof of Proposition 4: Noting that $K_t=(\beta_{t-1}^*\rho)'(\beta_{t-1}^*\rho)$ and that $vec(\beta_{t-1}^*\rho)\sim N(0,\frac{\rho^2}{1-\rho^2}I_r\otimes I_n)$, the first property follows from the definition of Wishart distribution (e.g. Muirhead, 1982, p. 82). The second follows from the properties of the Wishart distribution (e.g. Muirhead, 1982, p. 90). To prove the third property, let us first write (6) in matrix form as: $\beta_t^*=\beta_{t-1}^*\rho+\epsilon_t$, where $vec(\epsilon_t)=\eta_t$.

$$K_{t+1}=\rho(\beta_t^*)'(\beta_t^*)\rho=\rho(\beta_{t-1}^*\rho+\epsilon_t)'(\beta_{t-1}^*\rho+\epsilon_t)\rho$$

And thus:

$$K_{t+1}=\rho^2K_t+\rho^2\epsilon_t'\epsilon_t+\rho^3\beta_{t-1}^{*'}\epsilon_t+\rho^3\epsilon_t'\beta_{t-1}^* \tag{16}$$

By the law of iterated expectations, $E(\rho^3\beta_{t-1}^{*'}\epsilon_t|\kappa_{t-1})=E(E(\rho^3\kappa_{t-1}\beta_{t-1}'\epsilon_t|\beta_{t-1},\kappa_{t-1}))$. Since $E(\rho^3\kappa_{t-1}\beta_{t-1}'\epsilon_t|\beta_{t-1},\kappa_{t-1})=0$ we obtain $E(\rho^3\beta_{t-1}^{*'}\epsilon_t|\kappa_{t-1})=0$. Thus, taking conditional expectations on both sides of (16), and noting that $E(\epsilon_t'\epsilon_t|K_t,...,K_2)=nI_r$ we get $E(K_{t+1}|K_t,...,K_2)=\rho^2K_t+\rho^2nI_r$. Combining this with the second property, the third property is obtained.

Let k_{tij} be the (i,j) element of K_t. Note that the third property implies $E(k_{tij}|K_{(t-1)},...,K_2)=\rho^2k_{(t-1)ij}+n\rho^2\delta_{ij}$, where $\delta_{ij}=1$ if $i=j$ and is 0 otherwise. By the law of iterated expectations, this implies:

$$E(k_{tij}|K_{(t-h)},...,K_2)=\rho^{2h}k_{(t-h)ij}+n\delta_{ij}\sum_{c=1}^{h}\rho^{2c} \tag{17}$$

Note that $cov(k_{tij}, k_{(t-h)kl}) = E[(k_{tij} - E(k_{tij}))k_{(t-h)kl}] = E((k_{tij}k_{(t-h)kl})) - E(k_{tij})E(k_{(t-h)kl})$. Thus, $cov(k_{tij}, k_{(t-h)kl})$ can be obtained by multiplying both sides of (17) times $k_{(t-h)kl}$, subtracting $E(k_{tij})E(k_{(t-h)kl})$ and taking expectations:

$$cov(k_{tij}, k_{(t-h)kl}) = (\rho)^{2h}E(k_{(t-h)ij}k_{(t-h)kl}) + \delta_{ij}E(k_{(t-h)kl})n\sum_{c=1}^{h}(\rho)^{2c} - E(k_{(t-h)kl})E(k_{tij}) \tag{18}$$

From the properties of the Wishart distribution, all expectations in the right side of equation (18) are known. In particular, since K_t follows a Wishart with diagonal parameter matrix, it follows that $E(k_{tij}) = 0$ for $i \neq j$. Thus, for $i \neq j$ we have:

$$\begin{aligned} cov(k_{tij}, k_{(t-h)kl}) &= (\rho)^{2h}E(k_{(t-h)ij}k_{(t-h)kl}) - E(k_{(t-h)kl})E(k_{tij}) = \\ &(\rho)^{2h}E(k_{(t-h)ij}k_{(t-h)kl}) = (\rho)^{2h}cov(k_{(t-h)ij}, k_{(t-h)kl}) \end{aligned} \tag{19}$$

From the properties of the Wishart distribution with diagonal parameter matrix, (19) is zero unless $i = k$ and $j = l$. When $i = k$ and $j = l$, $cov(k_{(t-h)ij}, k_{(t-h)kl}) = var(k_{(t-h)ij})$, so that the correlation (i.e. covariance over square root of product of variances) between k_{tij} and $k_{(t-h)ij}$ is ρ^{2h}, for $i \neq j$. When $i = j = k = l$, (18) can be written as:

$$\begin{aligned} cov(k_{tii}, k_{(t-h)ii}) &= \rho^{2h}[E(k^2_{(t-h)ii}) - (E(k_{(t-h)ii}))^2] + \\ &(\rho^{2h} - 1)[E(k_{(t-h)ii})]^2 + E(k_{(t-h)ii})n\sum_{c=1}^{h}\rho^{2c} \end{aligned} \tag{20}$$

Noting that $E(k_{(t-h)ii}) = n\rho^2/(1-\rho^2)$ and $\sum_{c=1}^{h}\rho^{2c} = \rho^2(1-\rho^{2h})/(1-\rho^2)$, we get that:

$$(\rho^{2h} - 1)(E(k_{(t-h)ii}))^2 + E(k_{(t-h)ii})n\sum_{c=1}^{h}\rho^{2c} = 0 \tag{21}$$

Thus, (20) implies $cov(k_{tii}, k_{(t-h)ii}) = \rho^{2h}var(k^2_{(t-h)ii})$, and hence the correlation between k_{tii} and $k_{(t-h)ii}$ is ρ^{2h}. Finally, in the case $(i = j, k = l, i \neq k)$, using (21) and noting that $E(k_{(t-h)ii}) = E(k_{(t-h)kk})$, it can be shown that (18) is equal to zero.