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†
The Australian National University

January 30, 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: 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 against structural change. 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. To balance these considerations, the empirical macroeconomic literature is increasingly turning to time-varying parameter (TVP) models which use a particular class of hierarchical priors to model variation in parameters. The use of hierarchical priors can greatly 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:

\[ y_t = Z_t \gamma_t + \left( [\sigma] \right) \varepsilon_t \]  

and a state equation

\[ \gamma_t = \rho \gamma_{t-1} + \eta_t, \]

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 \( \gamma_t \) 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 \( \rho \) is set to one. From a Bayesian perspective, (2) defines a hierarchical prior for the parameters.

[[Mis-specification can have a significant effect on the results of an empirical study; which parameters are allowed to vary matters. Cogley and Sargent (2005) and Primiceri (2005) investigate the roles of ‘good luck’ and ‘good management’ in the effectiveness of monetary policy over time. These papers also allow for time varying volatility, replacing \( \sigma \) with \( \sigma_t \), and Sims and Zha (2006) argue that doing so mitigates evidence from earlier work that the mean coefficients, \( \gamma_t \), vary over time. This change in specification shifted]
the resulting evidence from support for the ‘good management’ story to support for the ‘good luck’ story. Sims and Zha (2006) provide clear evidence that models in which all parameters vary have little support against models in which a few important parameters are allowed to vary. We provide further evidence to support this conclusion, but that equilibria may vary over time.]

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.

[[We find evidence that a model that allows for time varying cointegration and heteroskedasticity, but keeps adjustment and lag coefficients constant has strong support relative to more relaxed models. This result is in line with the argument of Sims and Zha (2006) that allowing variation in few parameters gives a better model specification.]]

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. 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

\[1\text{In this paper, we focus on the noninformative prior. The working paper version, available at http://personal.strath.ac.uk/gary.koop/, discusses informative priors.}\]
methods can accurately estimate time-variation in the cointegration space and shows the importance of allowing for such time-variation.

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 \( t = 1, \ldots, T \):

\[
\Delta y_t = \alpha_t \beta_t' y_{t-1} + \sum_{h=1}^{t} \Gamma_{h,t} \Delta y_{t-h} + \Phi_t d_t + \varepsilon_t
\]  

(3)

where \( \varepsilon_t \) are independent \( N(0, \Omega_t) \), the \( n \times r \) matrices \( \alpha_t \) and \( \beta_t \) 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 \( \beta \), \( p = sp(\beta) \) and present a method for estimating the space. In this paper, we follow Strachan and Inder (2004) by achieving identification by specifying \( \beta \) 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 \( \beta \) will imply a uniform prior on \( p \).

The TVP-VECM in (3) includes \( t \) subscripts on each of the parameters, including the cointegrating space. Thus \( p_t = sp(\beta_t) \) where \( \beta_t \) is semi-orthogonal. In modelling the evolution of \( p_t \) 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 \( p_t \) from \( p_{t-1} \) 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 $\Omega_t$. 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 “$p_t$ is centered over $p_{t-1}$”. The fact that we are achieving identification through restricting $\beta_t$ to be semi-orthogonal means that we cannot have $\beta_t$ evolving according to an AR(1) or random walk process in a conventional normal state space model. 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 (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 orthogo-

---

The working paper version of this paper, discusses time-varying cointegration using other identification schemes.
nality 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: $\beta_t$ is identified by imposing $\beta'_t \beta_t = I_r$. We will let $\beta^*_t$ be the unrestricted matrix of cointegrating vectors (without identification imposed). These will be related to the semi-orthogonal $\beta_t$ as:

$$\beta_t = \beta^*_t (\kappa_t)^{-1} \quad \text{(4)}$$

where

$$\kappa_t = (\beta^*_t \beta_t^*)^{1/2}. \quad \text{(5)}$$

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 $b^*_t = \text{vec} \left( \beta^*_t \right)$ for $t = 2, \ldots, T$ as

$$b^*_t = \rho b^*_{t-1} + \eta_t \quad \text{(6)}$$

$$\eta_t \sim N(0, I_{nr}) \text{ for } t = 2, \ldots, T.$$

$$b^*_1 \sim N(0, I_{nr} \frac{1}{1 - \rho^2}),$$

where $\rho$ 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 $\beta_t$. 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 $\mathbf{p}_t$ is centered over $\mathbf{p}_{t-1}$ (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 $\mathbf{p}_t$ would shrink until, at the limit, it imposes $\mathbf{p}_t = \mathbf{p}_{t-1}$. Note also that we have normalized the error covariance matrix in the state equation to the identity. As we shall see, it is $\rho$ which controls the dispersion of the state equation (and, thus, plays a role similar to that played by $\sigma^2_\nu$, 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 $\beta_t$ do not always translate directly to comparable locations and dispersions on the space $\mathbf{p}_t$. For example, it is possible to construct simple cases where a distribution on $\beta_t$ has its mode and mean at $\hat{\beta}_t$, while the mode or mean of the distribution on $\mathbf{p}_t$ is in fact located upon the space orthogonal to the space of $\hat{\beta}_t$. The distributions we use avoid such inconsistencies.

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

Our hierarchical prior in (6) is written in terms of $b^*_t$, but we are interested in $\mathbf{p}_t$. Accordingly, we work out the implications of (6) for $\mathbf{p}_t$. We collect each of the $nr \times 1$ vectors $b^*_t = \text{vec}(\beta^*_t)$ into a single $Tnr \times 1$ vector $b^* = (b^*_1', ..., b^*_T')'$.

---

\(^3\)In this more general case Propositions 1, 2 and 3 continue to hold if we simply write $(\kappa_{t-1} \hat{\rho}^2 \kappa_{t-1})$ 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 $\rho$ as a scalar.
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 \( p_t = sp(\beta^*_t) = sp(\beta_t) \). To do so, we use some results from Strachan and Innder (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 \( b^*_t \sim N(0, cI_{nr}) \) implies a uniform distribution for \( \beta_t \) on the Stiefel manifold and a uniform distribution for \( p_t \) on the Grassmann manifold (for any \( c > 0 \)). It can immediately be seen from the joint distribution of \( b^*_t \) that the marginal distribution of any \( b^*_t \) has this form and, thus, the marginal prior distribution on \( p_t \) 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 \( p_t \) conditionally on \( p_{t-1} \) 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 \( \beta^*_t \) given \( \beta^*_{t-1} \) 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 \( p_t \) (conditional on \( p_{t-1} \)) 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 \( p_t = sp(\beta_t) \) first note that the space \( p_t \) 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 \( p_t \) as the density of \( P_t \). The form of the density function for \( p_t \) is given by

\[
 f(P_t|F_t) = \exp \left( -\frac{1}{2} tr(F_t) \right) _1F_1 \left( \frac{n}{2}; \frac{r}{2}; \frac{1}{2} F_t P_t \right) \tag{7}
\]

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

**Proposition 1** Since \( p_t = sp(\beta_t) \) 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 \( p_t \) 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 $p_t | p_{t-1}$ is $p_{t-1}$). 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

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

(8)

Using this measure, Villani (2006) defines a location measure for spaces such as $p_t = sp(\beta_t)$ by first defining

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

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

**Proposition 2** Since $p_t$ follows an $OPG(F_t)$ distribution with $F_t = \beta_{t-1} \rho^2 \kappa_{t-1} \beta_{t-1}'$, it follows that $E_{t-1}(p_t) = sp(\beta_{t-1}) = p_{t-1}$.

**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 $E_{t-1}(p_t) = p_{t-1}$ 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 den-
sity 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
\( \rho^2 \kappa_t^2 \):

**Proposition 3** Assume \( p_t \) 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 \( p_t \) at the mode depends only on the
eigenvalues of \( K_t = \rho^2 \kappa_t^2 \).

2. The value of the density function of \( p_t \) 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 \(K_i\) 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 \(K_i\) and the \((i, j)\) element of \(K_{t-h}\) is \(\rho^{2h}\).

**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 \(\rho\) 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 \(p_t\) from \(p_{t-1}\) 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 the time varying VECM. Our parameters break into three main blocks: the

\(^4\)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 \(p_t\). 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.
error covariance matrices \((\Omega_t)\) 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. \(\beta_t^*\) 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 \(\Omega_t\) and use his algorithm for producing MCMC draws from the posterior of \(\Omega_t\) 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 \(\alpha_t^*\beta''_t\) where \(\alpha_t^* = \alpha_t\kappa_t^{-1}\) and \(\beta''_t = \beta_t\kappa_t\) where \(\kappa_t\) is a \(r \times r\) symmetric positive definite matrix. Note that \(\kappa_t\) is not identified in the likelihood function but the prior we use for \(\beta''_t\) implies that \(\kappa_t\) has a proper prior distribution and, thus, is identified under the posterior. Even though \(\beta'_t\) is semi-orthogonal, Koop, León-González and Strachan (2010) show that the posterior for \(\beta''_t\) has a normal distribution (conditional on the other parameters). Thus, \(\beta''_t\) 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 \(\beta''_t\) can be transformed into draws of \(\beta'_t\) 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 \(\rho\). However, if it is a hierarchical prior and \(\rho\) 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 for this parameter. Further details on the prior distribution and posterior distribution of \(\kappa_t\) are easily derived from those of the prior for \(K_t = \rho^2\kappa_{t-1}\), which are described in Proposition 4.
computations are provided in Appendix A.

3 Application: The Fisher Effect

We illustrate our methods using a bivariate example \( n = 2 \) involving a short term interest rate, \( r_t \), and inflation rate, \( \pi_t \), 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 \( a_t = \text{vec}(A_t) \)) evolve according to random walks. Appendix A provides complete details.

We consider variants of this model that restrict one of \( \Omega_t \), \( \beta_t \) or \( a_t \) to be constant. To denote our models, we introduce three indicators \( (d_\Omega, d_\beta, d_a) \) which take value one whenever the corresponding element of \( (\Omega_t, \beta_t, a_t) \) is time-varying, and zero when it is time-invariant. We consider models with different cointegrating ranks \( (r = 0, \ldots, 2) \). 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 \( \rho \) 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 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) \(\alpha_t\) coefficients is in the range \((0.054 – 0.15)\). In the case of \(\rho\), 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 \((d_\Omega, d_\beta, d_\alpha) = (1, 1, 0)\) with \(r = 1\), which implies that there is time-varying cointegration and stochastic volatility (i.e. only \(\alpha_t\) 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 \(\beta_t^* = (\beta_{11}^*, \beta_{12}^*)'\) and define \(B_t = \beta_{12}^*/\beta_{11}^*\). Figure 1 plots the posterior median of \(B_t\) and a 90% credible interval. The interval contains \(-1\), the value implied by the Fisher hypothesis, 34% of the time. 60% of the time the

\(^{11}\)This parameter is constructed ex post. By that we mean for \(n = 2, r = 1\) we draw \(\beta_t^* = (\beta_{11}^*, \beta_{12}^*)'\) from the posterior, and then construct a draw of \(B_t\) as \(B_t = \beta_{12}^*/\beta_{11}^*\).

\(^{12}\)In our case this refers to \(\sum_{t = T - 100 + 1}^T \ln(p(\Delta y_t|y_{t-1}, \ldots, y_1)).\) Each component of this sum was calculated by approximating \(p(\Delta y_t|y_{t-1}, \ldots, y_1)\) with an average of \(p(\Delta y_t|\Omega_t, \beta_t, \alpha_t, y_{t-1}, \ldots, y_1)\) over 7000 draws of the posterior \((\Omega_t, \beta_t, \alpha_t|y_{t-1}, \ldots, y_1)\). We used our algorithm to obtain 7000 draws from \((\Omega_t, \beta_t, \alpha_t|y_{t-1}, \ldots, y_1)\). For each of these we generated \((\Omega_t, \beta_t, \alpha_t)\) using the state equation \((\Omega_t, \beta_t, \alpha_t)\|(\Omega_{t-1}, \beta_{t-1}, \alpha_{t-1})\). 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).
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.\textsuperscript{13}

<table>
<thead>
<tr>
<th>( (d_\Omega, d_\beta, d_\alpha) )</th>
<th>( n = 2 )</th>
<th>( r = 1 )</th>
<th>( r = 2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1,1,1)</td>
<td>-397.6 (0.66)</td>
<td>-423.8 (0.93)</td>
<td></td>
</tr>
<tr>
<td>(1,1,0)</td>
<td>-381.6 (0.61)</td>
<td>-384.1 (0.58)</td>
<td></td>
</tr>
<tr>
<td>(0,0,1)</td>
<td>-427.5 (0.14)</td>
<td>-431.0 (0.97)</td>
<td></td>
</tr>
<tr>
<td>(0,1,1)</td>
<td>-424.9 (0.12)</td>
<td>-434.1 (1.10)</td>
<td></td>
</tr>
<tr>
<td>(1,0,0)</td>
<td>-390.1 (0.86)</td>
<td>-386.5 (1.11)</td>
<td></td>
</tr>
<tr>
<td>(1,0,1)</td>
<td>-392.2 (0.72)</td>
<td>-412.8 (1.21)</td>
<td></td>
</tr>
<tr>
<td>(0,1,0)</td>
<td>-421.4 (0.08)</td>
<td>-423.3 (0.12)</td>
<td></td>
</tr>
<tr>
<td>(0,0,0)</td>
<td>-437.1 (0.05)</td>
<td>-436.9 (0.05)</td>
<td></td>
</tr>
</tbody>
</table>

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.

4 Conclusion

TVP-VARs have become very popular in empirical macroeconomics. In this paper, we have extended such models to allow for cointegration. However,\textsuperscript{13}

\textsuperscript{13}The working paper version of this paper contains additional empirical results, in particular for \( \alpha_t \) and \( \rho \).
Figure 1: Posterior median and 90% credible interval for $B_t$ when $n = 2, r = 1$ and $(d_\Omega, d_\beta, d_\alpha) = (1, 1, 0)$. A horizontal line at -1 has been added to aid visualization.
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.

<table>
<thead>
<tr>
<th>$(d_Q, d_H)$</th>
<th>$n = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0,0)</td>
<td>-442.5 (0.04)</td>
</tr>
<tr>
<td>(0,1)</td>
<td>-443.2 (0.05)</td>
</tr>
<tr>
<td>(1,0)</td>
<td>-398.0 (0.62)</td>
</tr>
<tr>
<td>(1,1)</td>
<td>-403.2 (0.59)</td>
</tr>
</tbody>
</table>

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


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


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


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:

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

where $\zeta_t \sim N(0, Q)$.\(^\text{14}\) We can rewrite (3) by defining $z_t = \beta^{*'}_t y_{t-1}$ and $Z_t = (z'_t, \Delta y'_{t-1}, \ldots, \Delta y'_{t-l}, d'_t)$. $Z_t$ is a $1 \times (k + r)$ vector where $k$ is the number of deterministic terms plus $n$ times the number of lags. Thus,

$$\Delta y_t = A_t Z_t + \varepsilon_t.$$  \hspace{1cm} (10)

Vectorizing this equation gives us the form

$$\Delta y_t = (Z'_t \otimes I_n) vec(A_t) + \varepsilon_t$$

or $\Delta y_t = x_t a_t + \varepsilon_t$

where $x_t = (Z'_t \otimes I_n)$. As we have assumed $\varepsilon_t$ is normally distributed, the above expression gives us the linear normal form for the measurement equation for $a_t$ (conditional on $\beta^{*}_t$). 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 $\kappa_t$ is a $r \times r$ symmetric positive definite matrix. To show how $\beta^{*}_t$ can be drawn, we rewrite (3) by defining

$$\tilde{y}_t = \Delta y_t - \sum_{h=1}^{l} \Gamma_{h,t} \Delta y_{t-h} - \Phi_t d_t = \alpha^{*}_t \beta^{*'}_t y_{t-1} + \varepsilon_t$$

or $\tilde{y}_t = \tilde{x}_t b^{*}_t + \varepsilon_t$

where we have used the relation $\alpha^{*}_t \beta^{*'}_t y_{t-1} = (y'_{t-1} \otimes \alpha^{*}_t) b^{*}_t$ where $b^{*}_t = vec(\beta^{*}_t)$ and the definition $\tilde{x}_t = (y'_{t-1} \otimes \alpha^{*}_t)$. 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.

24
\( \varepsilon_t \) 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, \( \Omega_t \), such that:

\[
\Lambda_t \Omega_t \Lambda_t' = \Sigma_t \Sigma_t'
\]

or

\[
\Omega_t = \Lambda_t^{-1} \Sigma_t \Sigma_t' (\Lambda_t^{-1})',
\]

where \( \Sigma_t \) is a diagonal matrix with diagonal elements \( \sigma_j,t \) for \( j = 1, \ldots, n \) and \( \Lambda_t \) is a lower triangular matrix with ones on the diagonal and lower diagonal elements \( \lambda_{ij,t} \). To model evolution in \( \Sigma_t \) and \( \lambda_t \) we must specify additional state equations. For \( \Sigma_t \) a stochastic volatility framework can be used. In particular, if \( \sigma_t = (\sigma_{1,t}, \ldots, \sigma_{n,t})' \), \( h_{i,t} = \ln (\sigma_{i,t}) \), \( h^t = (h_{1,t}, \ldots, h_{n,t})' \) 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 \( \varepsilon_t, \eta_t \) and \( \zeta_t \).

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

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

where \( \xi_t \) is \( N(0,C) \) and is independent over \( t \) and of \( u_t, \varepsilon_t, \zeta_t \) and \( \eta_t \).

**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 \( p_1 \). We now describe the prior for initial
conditions \((h_0, \lambda_0 \text{ and } a_0)\) and the variances of the errors in the other three state equations \((W, C \text{ and } Q)\). We also require a prior for \(\rho\) 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 \(a_0 \sim N\left(0, 2V_a\right)\), where \(V_a\) is the identity matrix except for those diagonal elements that correspond to \(\alpha_t^*\), 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: \(Q^{-1} \sim W\left(\nu_Q, Q^{-1}\right)\), \(W^{-1} \sim W\left(\nu_W, W^{-1}\right)\) and \(C^{-1} \sim W\left(\nu_C, C^{-1}\right)\). We choose a small number for the degrees of freedom, which is equal to the dimension of the matrix plus 2 (i.e. \(\nu_Q = \text{dim}(a_t) + 2\), \(\nu_W = n + 2\), \(\nu_C = n(n-1)/2 + 2\)) and we make each of the matrices \((Q, W, C)\) 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 \(\rho\). Since most of these involve standard algorithms, we do not provide much detail. As in Primiceri (2005), draws of \(\lambda_t\) 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:

\[
Q^{-1}|\text{Data} \sim W\left(\bar{\nu}_Q, \bar{Q}^{-1}\right)
\]

where

\[
\bar{\nu}_Q = T + \nu_Q
\]

and

\[
\bar{Q}^{-1} = \left[Q + \sum_{t=1}^{T} (a_t - a_{t-1}) (a_t - a_{t-1})\right]^{-1}.
\]
Next we have:

\[ W^{-1} | Data \sim W \left( \mathbf{v}_W, W^{-1} \right) \]

where

\[ \mathbf{v}_W = T + \mathbf{v}_W \]

and

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

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

\[ C^{-1} | Data \sim W \left( \mathbf{v}_C, C^{-1} \right) \]

where

\[ \mathbf{v}_C = T + \mathbf{v}_C \]

and

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

The posterior for \( \rho \) 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 \( p_t \) conditional on \( (\kappa_t, \beta_{t-1}, \kappa_t-1) \) is maximized at \( p_t = sp(\beta_{t-1}) \) for any value of \( (\kappa_t, \kappa_t-1) \). This proves that the density of \( p_t \) conditional on \( (\beta_{t-1}, \kappa_t-1) \) is also maximized at \( p_t = sp(\beta_{t-1}) \). Clearly, the mode is also the same if we do not condition on \( \kappa_{t-1} \).

The state equation in (6) implies that the conditional density of \( \beta_t^* \) given \( \beta_{t-1}^* \) 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 \( L \left( n, r; \hat{F} \right) \) (Chikuse, 2003, p. 31), where

\[ \hat{F} = \beta_{t-1}^* \rho \kappa_t = \beta_{t-1} \kappa_{t-1} \rho \kappa_t \]
The form of the density function for $L(n, r; \tilde{F})$ 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 $\beta_t$ using Theorem 2.4.8 in Chikuse (2003, p. 46):

$$
 f_{P_t}(P_t) = A_L \int_{O_r} \exp(tr\tilde{F}'\beta_tQ)[dQ] = A_L \cdot 0F_1\left( \frac{1}{2}r; \frac{1}{4}\tilde{F}'P_t\tilde{F} \right)
$$

where $A_L$ is a constant not depending on $P_t$ ($A_L^{-1} = 0F_1\left( \frac{1}{2}n; \frac{1}{4}\tilde{F}'\tilde{F} \right)$) 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 $0F_1\left( \frac{1}{2}r; \frac{1}{4}\tilde{F}'P_t\tilde{F} \right)$ is equal to $0F_1\left( \frac{1}{2}r; \frac{1}{4}G_t \right)$, 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 $0F_1\left( \frac{1}{2}r; \frac{1}{4}G_t \right)$ 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 $p_t = sp(\beta_{t-1})$.

We first show that the following standard expression for $0F_1\left( \frac{1}{2}r; \frac{1}{4}G_t \right)$ (e.g. Muirhead, 1982, p. 262):

$$
 0F_1\left( \frac{1}{2}r; \frac{1}{4}G_t \right) = \int_{O(r)} \exp\left( \sum_{i=1}^{r} \sqrt{g_i}q_{ii} \right)[dQ]
$$

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

$$
 0F_1\left( \frac{1}{2}r; \frac{1}{4}G_t \right) = \int_{\tilde{O}_r} \prod_{i=1}^{r} (\exp(\sqrt{g_i}q_{ii}) + \exp(-\sqrt{g_i}q_{ii}))[dQ] \quad (14)
$$

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:

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

28
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:

\[
\int_{\{O(r); q_{i1} < 0\}} \exp\left(\sum_{i=1}^{r} \sqrt{g_i} q_{i1}\right) [dQ] = \int_{\{O(r); z_{i1} \geq 0\}} \exp(-\sqrt{g_i} z_{i1}) \exp\left(\sum_{i=2}^{r} \sqrt{g_i} z_{ii}\right) [dZ]
\]

Thus:

\[
\int_{O(r)} \exp\left(\sum_{i=1}^{r} \sqrt{g_i} q_{i1}\right) [dQ] = \int_{\{O(r); q_{i1} \geq 0\}} (\exp(\sqrt{g_i} q_{i1}) + \exp(-\sqrt{g_i} q_{i1})) \exp\left(\sum_{i=2}^{r} \sqrt{g_i} q_{ii}\right) [dQ]
\]

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), \( qF_1(\frac{1}{2}r; \frac{1}{4}G_t) \) 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 \( C = \kappa_{t-1}P_t \kappa_{t} \) is a \( r \times r \) matrix. Let \( \beta_{t\perp} \) be the orthogonal complement of \( \beta_t \) (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 \( P_t + P_{t\perp} = (\beta_t, \beta_{t\perp})(\beta_t, \beta_{t\perp})' = I_r \). 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 prove that \( E(p_t|(\kappa_t, \beta_{t-1}, \kappa_{t-1})) = sp(\beta_{t-1}) \). Note that this proves that \( E(p_t|(\beta_{t-1}, \kappa_{t-1})) = sp(\beta_{t-1}) \), because if \( \beta = \beta_{t-1} \) minimizes
$$E(d^2(\beta_t, \beta_t' | (\kappa_t, \beta_{t-1}, \kappa_{t-1})))$$ for every $\kappa_t$, it will also minimize $E(d^2(\beta_t, \beta_t' | (\beta_{t-1}, \kappa_{t-1})))$. Similarly, it also proves that $E(p_t| \beta_{t-1}) = sp(\beta_{t-1})$.

Recall that $\beta_t | (\kappa_t, \beta_{t-1}, \kappa_{t-1})$ follows a Langevin distribution $L\left(n, r; \bar{F}\right)$, with $\bar{F} = \beta_{t-1}' \rho \kappa_t = \beta_{t-1}' \kappa_{t-1} P \kappa_t$. Let $G = \kappa_t P \kappa_{t-1}$ 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, $\bar{F} = \beta_{t-1} P M O'$. Write $\bar{F} = \beta_{t-1} P M O' = \Gamma M O'$, 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 $E(p_t) = sp(\Gamma)$.

In order to prove this, we will prove that $E(\beta_t \beta_t') = UDU'$, with $U = (\Gamma, \Gamma')$, where $\Gamma'$ is the orthogonal complement of $\Gamma$, and $D = \{d_{ij}\}$ is a diagonal matrix with $d_{11} \geq d_{22} \geq \ldots \geq d_{nn}$. Define the $n \times r$ matrix $Z = U' \beta_t O$, so that $E(\beta_t \beta_t') = U E(Z O' O' Z' U') = U E(Z Z') U'$. 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) $i = k = j = l$, $i = 1, 2, \ldots, r$; (b) $i = k$, $j = l$ ($i \neq j$); (c) $i = j$, $k = l$ ($i \neq k$); (d) $i = l$, $j = k$ ($i \neq j$), $i, j = 1, 2, \ldots, r$. Note that the $(i, k)$ element of $Z Z'$ is $\sum_{h=1}^{n} z_{ih} z_{kh}$. Thus, $E(Z Z')$ is a diagonal matrix and we can write $E(\beta_t \beta_t') = U D U'$, where $D = E(Z Z')$.

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

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

where $\tilde{Z} = \{\tilde{z}_{ij}\}$ consists of the first $r$ rows of $Z$, $M = diag(m_1, \ldots, m_r)$ and $A_L$ is a normalizing constant. If we let $\tilde{Z} = \{\tilde{z}_{ij}\}$ be the other $n - r$ rows, what needs to be proved can be written as: $E(\sum_{l=1}^{r}(\tilde{z}_{jl})^2) \geq E(\sum_{l=1}^{r}(\tilde{z}_{pl})^2)$ 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 $\sum_{l=1}^{r}(\tilde{z}_{pl})^2$ 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:
\[ A_L \int_{S_1} \left( \sum_{l=1}^{r} (\tilde{z}_{jl})^2 \right) \exp \left\{ \sum_{l=1}^{r} m_l \tilde{z}_{ll} \right\} [dZ] + A_L \int_{S_2} \left( \sum_{l=1}^{r} (\tilde{z}_{jl})^2 \right) \exp \left\{ \sum_{l=1}^{r} m_l \tilde{z}_{ll} \right\} [dZ] \]

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

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

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:

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

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

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

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

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

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. \( \sum_{i=1}^{r} (\tilde{z}_{ij})^2 = \sum_{i=1}^{r} (\hat{w}_{ij})^2 \), which is a consequence of \( ZZ' = WW' \)). Define \( \tilde{W} \) as a matrix that is equal to \( \hat{W} \) for all rows except for the \( j^{th} \) one. Let the \( j^{th} \) row of \( \tilde{W} \) be equal to the \((r + p)^{th}\) row of \( W \). Note that \( m_j \tilde{z}_{pj} + \sum_{i=1, i \neq j}^{r} m_i \tilde{z}_{pi} = tr(M\tilde{W}N) \). Thus, \( E(\sum_{i=1}^{r} (\tilde{z}_{ij})^2) - E(\sum_{i=1}^{r} (\tilde{z}_{pi})^2) \) can be written as:

\[
A_{L} \int_{S_{1}} (\sum_{i=1}^{r} (\tilde{w}_{ij})^2 - \sum_{i=1}^{r} (\hat{w}_{ij})^2) \left( \exp \{ tr(M\tilde{W}N) \} - \exp \{ tr(M\hat{W}N) \} \right) [dW][dN] = A_{L} \int_{S_{1}} (\sum_{i=1}^{r} (\tilde{w}_{ij})^2 - \sum_{i=1}^{r} (\hat{w}_{ij})^2) \left( _0F_1(\frac{1}{2}r; \frac{1}{4}M\tilde{W}W'M) - _0F_1(\frac{1}{2}r; \frac{1}{4}M\hat{W}W'M) \right) [dW]
\]  

(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), \( _0F_1(\frac{1}{2}r; \frac{1}{4}M\tilde{W}W'M) \) is a function only of the singular values of \( M\tilde{W}W'M \). In addition, as we argued when we found the mode of the distribution, \( _0F_1(\frac{1}{2}r; \frac{1}{4}M\tilde{W}W'M) \) increases with each of the singular values of \( M\tilde{W}W'M \).

Let \( A = M\tilde{W}W'M \) and \( B = M\hat{W}W'M \). Let the singular values of \( A \) be \((a_1, \ldots, a_r)\) with \( a_i \geq a_{i+1} \) and let the singular values of \( B \) be \((b_1, \ldots, b_r)\), with \( b_i \geq b_{i+1} \). From now on we will show that in the region \( S_1, a_i \geq b_i, i = 1, \ldots, r \). 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 \( E(\sum_{i=1}^{r} (\tilde{z}_{ij})^2) \geq E(\sum_{i=1}^{r} (\tilde{z}_{pi})^2) \).

Define the matrix \( C = A - B = M(WW' - \hat{W}W')M \). 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 \( c_i = m_i^2 (\sum_{i=1}^{r} (\tilde{w}_{ji})^2 - \sum_{i=1}^{r} (\hat{w}_{ji})^2) \) which is positive in \( S_1 \). Let the singular values of \( C \), ordered from high to low, be \((c_1, \ldots, c_r)\), 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: \( b_1 + c_r \leq a_1, b_2 + c_r \leq a_2, \ldots, b_r + c_r \leq a_r \). Since \( c_r = 0 \) this implies \( b_1 \leq a_1, b_2 \leq a_2, b_3 \leq a_3, \ldots, b_r \leq a_r \). 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 \( P_t = \beta_i, \beta'_i \) given by ex-
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):

$$1F_1\left(\frac{n}{2}; \frac{r}{2}; \frac{1}{2}F_tP_t\right) = \sum_{k=0}^{\infty} \sum_{\kappa} \frac{(n/2)_\kappa C_\kappa(1/2)F_tP_t}{(r/2)_\kappa k!}$$

where $\kappa$ 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 $\kappa$ of $k$. $C_\kappa$ 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: $\exp\left(-\frac{1}{2}tr(D)\right) 1F_1(n/2; r/2; 1/2D)$. 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:

\[
\frac{\Gamma(r/2)}{\Gamma(n/2)} \exp\left(\frac{1}{2}d_i\right) \left(\frac{d_i}{2}\right)^{(n-r)/2} \frac{1}{1F_1\left(\frac{(n-1)}{2}; \frac{(r-1)}{2}; \frac{1}{2}D_{-i}\right)} \left[1 + O(d_{ii}^{-1}) + O(d_{ii}^{-2})\right]
\]

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:

\[
\frac{\Gamma(r/2)}{\Gamma(n/2)} \exp\left(-\frac{1}{2}tr(D_{-i})\right) \left(\frac{d_i}{2}\right)^{(n-r)/2} \frac{1}{1F_1\left(\frac{(n-1)}{2}; \frac{(r-1)}{2}; \frac{1}{2}D_{-i}\right)} \left[1 + O(d_{ii}^{-1}) + O(d_{ii}^{-2})\right]
\]

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:

\[
K_{t+1} = \rho(\beta^*_{t})'(\beta^*_{t})\rho = \rho(\beta^*_{t-1}\rho + \epsilon_t)'(\beta^*_{t-1}\rho + \epsilon_t)\rho
\]

And thus:

\[
K_{t+1} = \rho^2 K_t + \rho^2 \epsilon_t \epsilon_t' + \rho^3 \beta^*_{t-1}' \epsilon_t + \rho^3 \epsilon_t' \beta^*_{t-1}
\]

(16)

By the law of iterated expectations, \(E(\rho^3 \beta^*_{t-1}' \epsilon_t | \kappa_{t-1}) = E(E(\rho^3 \beta^*_{t-1}' \epsilon_t | \kappa_{t-1}) | \kappa_{t-1})\).

Since \(E(\rho^3 \beta^*_{t-1}' \epsilon_t | \kappa_{t-1}) = 0\) 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 \(E(K_{t+1} | K_t, ..., K_2) = \rho^2 K_t + \rho^2 nI_r\). 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 \(E(k_{tij} | K_{(t-1)}, ..., K_2) = \rho^2 k_{(t-1)ij} + n\rho^2 \delta_{ij}\), where \(\delta_{ij} = 1\) if \(i = j\) and is 0 otherwise. By the law of iterated expectations, this implies:

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

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

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

(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 $E(k_{ij}) = 0$ for $i \neq j$. Thus, for $i \neq j$ we have:

$$\text{cov}(k_{ij}, k_{(t-h)kl}) = (\rho)^{2h}E(k_{(t-h)ij}k_{(t-h)kl}) - E(k_{(t-h)kl})E(k_{ij}) =$$

$$= (\rho)^{2h}E(k_{(t-h)ij}k_{(t-h)kl}) = (\rho)^{2h}\text{cov}(k_{(t-h)ij}, k_{(t-h)kl})$$

(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 $j = l$, $\text{cov}(k_{(t-h)ij}, k_{(t-h)kl}) = \text{var}(k_{(t-h)ij})$, so that the correlation (i.e. covariance over square root of product of variances) between $k_{ij}$ and $k_{ij}$ is $\rho^{2h}$, for $i \neq j$. When $i = j = k = l$, (18) can be written as:

$$\text{cov}(k_{iii}, k_{(t-h)iii}) = \rho^{2h}[E(k_{(t-h)iii}) - (E(k_{(t-h)iii}))^2] +$$

$$+ (\rho^{2h} - 1)[E(k_{(t-h)iii})]^2 + E(k_{(t-h)iii})n \sum_{c=1}^{h} \rho^{2c}$$

(20)

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

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

(21)

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