STRATHCLYDE

DISCUSSION PAPERS IN ECONOMICS

BAYESIAN INFERENCE IN THE TIME VARYING COINTEGRATION MODEL

BY

GARY KOOP, ROBERT LEON-GONZALEZ AND RODNEY W. STRACHAN

NO. 11-21

DEPARTMENT OF ECONOMICS
UNIVERSITY OF STRATHCLYDE
GLASGOW

Bayesian Inference in the Time Varying Cointegration Model*

Gary Koop
University of Strathclyde

Roberto Leon-Gonzalez
National Graduate Institute for Policy Studies

Rodney W. Strachan†
University of Queensland

May 22, 2008

*All authors are Fellows of the Rimini Centre for Economic Analysis

†Corresponding author: School of Economics, University of Queensland, St Lucia Brisbane Qld 4072, Australia. Email: r.strachan@uq.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 time varying parameter models which permit cointegration. Time-varying parameter VARs (TVP-VARs) typically use state space representations to model the evolution of parameters. In this paper, we show that it is not sensible to use straightforward extensions of TVP-VARs when allowing for cointegration. Instead we develop 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 a permanent/transitory variance decomposition for inflation.

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

JEL Classification: C11, C32, C33

1 Introduction

Empirical macroeconomics increasingly relies on multivariate time series models where the parameters which characterize the conditional mean and/or conditional variance can change over time. These models are motivated by a realization that factors such as financial liberalization or changes in monetary policy (or many other things) can cause the relationships between variables to alter. There are many ways in which parameters can evolve over time, but it is increasingly popular to use state space modelling techniques and allow parameters to evolve according to an AR(1) process or a random walk. 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. Papers such as Cogley and Sargent (2005) or Primiceri (2005) 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.

However, many issues in empirical macroeconomics involve the concept of cointegration and, thus, there is a need for a time varying vector error correction model (VECM) comparable to the TVP-VAR. The obvious extension of the existing time varying VAR literature to allow for cointegration would be to redefine (1) appropriately so as to be a VECM and allow the identified cointegrating vectors to evolve according to (2). A contention of this paper is that this is not a sensible strategy. The reasons for this will be explained in detail in the paper. Basically, 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. Hence, ideally 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. We show in this paper that working with a state equation such as

(2) which allows for identified cointegrating vectors to evolve according to an AR(1) or random walk yields a model which does not have these properties. In fact, we show that it has some very undesirable properties (e.g. the cointegrating space will be drawn towards an absorbing state).

Having established that the obvious extension of (1) and (2) to allow for cointegration is not sensible, we develop an alternative approach. From a Bayesian perspective, (2) defines a hierarchical prior for the parameters. Our alternative approach involves developing a better hierarchical prior. 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 for such questions. Using these, we 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, 2008a,b) 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.S. macroeconomics variables (i.e. unemployment, inflation and interest rates) and show how the one cointegrating relationship between them varies over time. We then carry out a time-varying permanent-transitory variance decomposition for inflation. We find that the role of transitory shocks, although small, has increased over time.

2 Modelling Issues

2.1 The Time Varying Cointegration Model

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

$$\Delta y_t = \Pi y_{t-1} + \sum_{h=1}^{l} \Gamma_h \Delta y_{t-h} + \Phi d_t + \varepsilon_t \tag{3}$$

where the $n \times n$ matrix $\Pi = \alpha\beta'$, α and β are $n \times r$ full rank matrices 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, at this stage we will leave these unspecified. We assume ε_t to be i.i.d. $N(0, \Omega)$.

Before extending (3) to allow for time varying cointegration, it is important to digress briefly to motivate an important issue in the Bayesian analysis of cointegrated models. The VECM suffers from a global identification problem. This can be seen by noting that $\Pi = \alpha\beta'$ and $\Pi = \alpha A A^{-1} \beta'$ are identical for any nonsingular A. This indeterminacy is commonly surmounted by imposing the so-called *linear normalization* where $\beta = [I_r \quad B']'$. However, there are some serious drawbacks to this linear normalization (see Strachan and Inder, 2004 and Strachan and van Dijk, 2007). Researchers in this field (see 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, $\mathfrak{p} = sp(\beta)$. 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, unlike the linear normalization. Another key result from Strachan and Inder (2004) is that a Uniform prior on β will imply a Uniform prior on $\mathfrak{p}$.

We can generalize (3) for the time-varying cointegrating space case by including t subscripts on each of the parameters including the cointegrating space $\mathfrak{p}$. Thus we can replace Π in (3) with $\Pi_t = \alpha_t \beta_t'$ where $\mathfrak{p}_t = sp(\beta_t)$ where β_t is semi-orthogonal. We also replace $\alpha, \Gamma_1, \ldots, \Gamma_l, \Phi,$ and Ω by $\alpha_t, \Gamma_{1,t}, \ldots, \Gamma_{l,t}, \Phi_t,$ and Ω_t. 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.

We can write the measurement equation for our time varying cointegrating space model as:

$$\Delta y_t = \alpha_t \beta_t' y_{t-1} + \sum_{h=1}^{l} \Gamma_{h,t} \Delta y_{t-h} + \Phi_t d_t + \varepsilon_t \tag{4}$$

where ε_t are independent $N(0, \Omega_t)$ for $t = 1, .., T$. The parameters $(\alpha_t, \Gamma_{1,t}, \ldots, \Gamma_{l,t}, \Phi_t)$ follow a standard state equation. Details are given in Appendix A. With respect to the 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 are also 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 (2005) or Saikkonen and Choi (2004).

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 $\mathfrak{p}_{t-1}$". We will return to this shortly, before we do so we provide some intuitive motivation for the issues involved and a motivation for why some apparently sensible approaches are not sensible at all.

2.2 Some Problems with Modelling Time Variation of the Cointegration Space

To illustrate some of the issues involved in developing a sensible hierarchical prior for the time varying cointegrating space model, consider the case where $n = 2$ and $r = 1$. Hence, for the two variables, y_{1t} and y_{2t}, we have the following cointegrating relationship at time t:

$$B_{1t}y_{1t} + B_{2t}y_{2t}$$

and the cointegrating space is $\mathfrak{p}_t = sp\left[(B_{1t}, B_{2t})'\right]$ or, equivalently, $\mathfrak{p}_t = sp\left[c\,(B_{1t}, B_{2t})'\right]$ for any non-zero constant c. Following Strachan (2003), it is useful to provide some intuition in terms of basic geometry. In the $n = 2$ case, the cointegrating space is simply a straight line (which cuts through the origin and has slope given by $-\frac{B_{2t}}{B_{1t}}$) in the two dimensional real space, R^2. Any such straight line can be defined in polar coordinates (i.e. with the polar angle — the angle between the X-axis and the line — determining the slope of the line). This motivates our link with the directional statistics literature (which provides statistical tools for modelling empirical processes involving, e.g., wind directions which are defined by angles) which we will use and extend in our theoretical derivations. The key point to note is that cointegrating spaces can be thought of as angles defining a straight line (or higher dimensional extensions of the concept of an angle). In the $n = 2$ case, identifying the cointegrating vector using a semi-orthogonality restriction (such as we impose on β_t) is equivalent to working in polar coordinates (and identification is achieved through restricting the length of the radial coordinate). Thus, β_t determines the polar angle.

A popular alternative way of identifying a particular cointegrating vector is through the linear normalization which chooses c such that $\mathfrak{p}_t = sp\left[(1, B_t)'\right]$ and, thus, the cointegrating vector is $(1, B_t)'$. This normalization has the obvious disadvantage of ruling out the cointegrating vector $(0, 1)'$. In the $n = 2$ case, this may not seem a serious disadvantage, but when $n > 2$ it can be. However, in the context of the time-varying cointegration model, this normalization causes additional problems.

In the spirit of the TVP-VAR model of (1) and (2), it is tempting to model the time variation of the cointegration model by assuming that B_t follows a stationary AR(1) process:

$$B_t = \rho B_{t-1} + \eta_t \tag{5}$$

with $\eta_t \sim N(0, \sigma_\eta^2)$. Here we provide a simple example to illustrate why working with such a specification leads to a model with highly undesirable properties.

Suppose that $\rho = 1$, $B_{t-1} = 0$ and that a shock of size $\eta_t = -1$ occurs. This causes a large change in the cointegrating vector (i.e. from $(1,0)'$ to $(1,-1)'$). Thinking in terms of the polar angle which defines the cointegrating space, this is an enormous 45 degree change. Now suppose instead that $B_{t-1} = 50$ (but all else is the same including $\eta_t = -1$). Then the corresponding change in the cointegrating vector will be imperceptible, since $(1,50)'$ and $(1,49)'$ are virtually the same (they differ by about 0.02 of a degree). Hence, two identical values for the increment in the state equation ($\eta_t = -1$) lead to very different changes in the cointegrating space.

Note also that the lines defined by $(1,50)'$ and $(1,49)'$ are very close to that defined by $(0,1)'$ (which was excluded a priori by the linear normalization). An implication of this is that, when $sp[(1, B_{t-1})']$ is close to $sp[(0,1)']$, the distribution of the cointegrating vector at t will be highly concentrated on the location at $t-1$. A first negative consequence of this is that the dispersion of the process given by (5) at t (conditional on $t-1$) depends on how close the cointegrating vector at $t-1$ is to $(0,1)'$. An even more negative consequence is that this vector plays the role of an absorbing state. That is, once the process defined by (5) gets sufficiently close to $(0,1)'$, it will be very unlikely to move away from it. Put another way, random walks wander in an unbounded fashion. Under the linear normalization, this means they will always wander towards $(0,1)'$. Formally, assuming a non-stationary process for B_t implies a degenerate long-run distribution for the cointegrating space. That is, if $\rho = 1$ in (5), the variance of B_t increases with time. This means that the probability that B_t takes very high values also increases with time. Therefore, this process would imply the cointegrating space converges to $sp[(0,1)']$ with probability 1 (regardless of the data).

Finally, although $E_{t-1}(B_t) = B_{t-1}$ when $\rho = 1$, this does not imply $\mathfrak{p}_t$ is centred over $\mathfrak{p}_{t-1}$. The reason for this is that the transformation from B_t to $\mathfrak{p}_t$ is nonlinear. A simple example will demonstrate. We noted earlier that in this simple case, the angle defined by the semi-orthogonal vector β_t (call this angle θ_t), defines the space and we have $B_t = \tan(\theta_t)$. If in (5) we have $\rho = 1$ and $\sigma_\eta = 2$ and we observe $B_{t-1} = 1$ such that $\theta_{t-1} = 0.79$, then $E_{t-1}(B_t) = 1 = \tan(0.79)$. However, the angle which defines the space

$E_{t-1}(\mathfrak{p}_t)$ is[1] 1.10, which is such that $tan(1.10) = 1.96$. Thus, even though $E_{t-1}(B_t) = B_{t-1}$, we have $E_{t-1}(\mathfrak{p}_t) \neq \mathfrak{p}_{t-1}$.

These examples (which can be extended to higher dimensions in a conceptually straightforward manner) show clearly how using a standard state space formulation for cointegrating vectors identified using the linear normalization is not appropriate.

A second common strategy (which we adopt in this paper) is to achieve identification through restricting β_t to be semi-orthogonal. One might be tempted to have β_t evolve according to an AR(1) or random walk process. However, given that β_t has to always be semi-orthogonal, it is obvious that this cannot be done in a conventional Normal state space model format. More formally, 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). It is clear that using a standard state space formulation for the cointegrating vectors identified using the orthogonality restriction is not appropriate.

The previous discussion illustrates some problems with using a state equation such as (2) to model the evolution of identified cointegrating vectors using two popular identification schemes. Similar issues apply with other schemes. For instance, similar examples can be constructed for the identification method suggested in Johansen (1991).

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.

This discussion establishes that working with state space formulations

[1] Here the expected value of a space is calculated as proposed in Villani (2006). See also the discussion around (12) in Section 2.3.

such as (2) to model the evolution of identified cointegrating vectors is not sensible. What then do we propose? To answer this question, we begin with some additional definitions. We have so far used notation for identified cointegrating vectors: β_t is identified by imposing semi-orthogonality, that is $\beta_t'\beta_t = I_r$, and, under the linear normalization, we have B_t being (the identified part of) the cointegrating vectors. 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:

$$\beta_t = \beta_t^* (\kappa_t)^{-1} \tag{6}$$

where

$$\kappa_t = \left(\beta_t^{*\prime}\beta_t^*\right)^{1/2}. \tag{7}$$

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

To present our preferred state equation for the time-variation in the cointegrating spaces, consider first the case with one cointegrating relationship ($r = 1$), β_t^* is an $n \times 1$ vector and suppose we have (for $t = 2, .., T$):

$$\begin{aligned} \beta_t^* &= \rho \beta_{t-1}^* + \eta_t \\ \eta_t &\sim N(0, I_n) \\ \beta_1^* &\sim N(0, I_n \frac{1}{1-\rho^2}) \end{aligned} \tag{8}$$

where ρ is a scalar and $|\rho| < 1$. 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, (8) 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).

Allowing $r > 1$ and defining $b_t^* = vec\left(\beta_t^*\right)$, we can write our state equation for the matrix of cointegrating vectors as

$$\begin{aligned} b_t^* &= \rho b_{t-1}^* + \eta_t \\ \eta_t &\sim N(0, I_{nr}) \text{ for } t = 2, ..., T. \\ b_1^* &\sim N(0, I_{nr} \frac{1}{1-\rho^2}), \end{aligned} \tag{9}$$

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 in (5)).

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.

2.3 Properties of Proposed State Equation

In the previous section, we demonstrated that some apparently sensible ways of extending the VECM to allow for time-varying cointegration led to models with very poor properties. This led us to propose (9) as an alternative. However, we have not yet proven that (9) has attractive properties. In this section, we do so. In particular, (9) is written in terms of b_t^*, but we are interested in $\mathfrak{p}_t$. Accordingly, we work out the implications of (9) for $\mathfrak{p}_t$.

We collect each of the $nr \times 1$ vectors $b_t^* = vec\,(\beta_t^*)$ into a single $Tnr \times 1$

vector

$$b^* = \begin{bmatrix} b_1^* \\ b_2^* \\ \vdots \\ b_T^* \end{bmatrix}.$$

The conditional distribution in (9) implies that the joint distribution of b^* is Normal with zero mean and covariance matrix V where

$$V = \frac{1}{1-\rho^2} \begin{bmatrix} I & I\rho & I\rho^2 & & I\rho^{T-1} \\ I\rho & I & I\rho & \cdots & I\rho^{T-2} \\ I\rho^2 & I\rho & I & & I\rho^{T-3} \\ & \vdots & & \ddots & \vdots \\ I\rho^{T-1} & I\rho^{T-2} & I\rho^{T-3} & \cdots & I \end{bmatrix}.$$

We begin with discussion of the marginal prior distribution of $\mathfrak{p}_t = sp(\beta_t^*) = sp(\beta_t)$. 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 $b_t^* \sim N(0, cI_{nr})$ 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. Hence, this marginal prior is noninformative. 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 (9) 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 $\mathfrak{p}_t = sp(\beta_t)$ 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

$$f(P_t|F_t) = \exp\left(-\frac{1}{2}tr(F_t)\right) {}_1F_1\left(\frac{n}{2};\frac{r}{2};\frac{1}{2}F_tP_t\right) \tag{10}$$

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* $F_t = \beta_{t-1}\rho^2\kappa^2_{t-1}\beta'_{t-1}$*, 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 getting closer to 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

$$d\left(\beta_t,\beta_{t-1}\right) = tr\left(\beta'_t\beta_{t-1\perp}\beta'_{t-1\perp}\beta_t\right)^{1/2}. \tag{11}$$

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

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

then defining this location measure (which he refers to as the *mean cointegrating space*) as $\overline{\mathfrak{p}} = sp\left(\overline{\beta}\right)$. Villani proves that $\overline{\mathfrak{p}}$ 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 $E\left(\mathfrak{p}_t\right) \equiv \overline{\mathfrak{p}}$ 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* $E(\mathfrak{p}_t) = sp(\beta_{t-1})$.

Proof: See Appendix B.

This proposition formalizes our previous informal statements about our state equation (9). That is, it implies the expected cointegrating space at time t is the cointegrating space at $t-1$. That is, we have $E_{t-1}(\mathfrak{p}_t) = \mathfrak{p}_{t-1}$ where the expected value is defined using Villani (2006)'s location measure. Propositions 1 and 2 prove that there are two senses in which (9) 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 $t-1$. It is straightforward to show that, if we had used alternative state equations such as (5) to model the evolution of the cointegrating space, neither proposition would hold.

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 (9), 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.

A further understanding of the prior can be obtained through simulation methods. For the sake of brevity, we only present prior simulation results relating to the change in the cointegrating space over time. To facilitate interpretation, we present results for the $n = 2, r = 1$ case in terms of the angle defining the cointegrating space. For the case where $T = 250$ and $\rho = 0.999$, Figure 1 plots the median and 16th and 84th percentiles of the prior of the change in this angle between time periods. It can be seen that this prior density is constant over time (as is implied by Proposition 4) and, hence, our prior implies constant expected change in the cointegrating space. Furthermore, $\rho = 0.999$ allows for fairly substantive changes in the cointegrating space over time. That is, it implies that changes in the cointegrating space of one or two degrees each time period are common. For monthly or quarterly data, this would allow for the angle defining the cointegrating space to greatly change within a year or two. Figure 2, which plots the entire prior density for the change in this angle for $\rho = 0.99$, $\rho = 0.999$ and $\rho = 0.9999$, reinforces this point. Even if we restrict consideration of ρ to values quite near one we are able to allow for large variations in the cointegrating space over time. In fact, $\rho = 0.99$ allows for implausibly huge changes in the cointegrating space, while $\rho = 0.9999$ still allows for an appreciable degree of movement in the cointegrating space. Given that our desire is to find a state equation which allows for constant *gradual* evolution in the cointegrating space we accordingly focus on values of ρ which are near to one.[2]

2.4 An Informative Marginal Prior

The hierarchical prior in (9) is conditionally informative in that it says "this period's cointegrating space is centered over last period's", but marginally is noninformative (i.e. as shown previously, the marginal prior distribution for $\mathfrak{p}_t$ is Uniform on the Grassmann manifold). However, economic theory often provides us with prior information about the likely location of the cointegrating space. Suppose H is a semi-orthogonal matrix which summarizes such prior information (see Strachan and Inder (2004) or Koop, Strachan, van Dijk and Villani (2006) for examples of how such an H can be constructed and used in standard cointegration models). In this sub-section, we describe how such prior information can be incorporated in our approach. In particular, we want a prior which combines our previous prior which said "$\mathfrak{p}_t$ is centered over $\mathfrak{p}_{t-1}$" with the prior information in H and is marginally informative (i.e. $\mathfrak{p}_t$ has a marginal distribution which is centered over H). We refer to this as an informative marginal prior.

It turns out that such an informative marginal prior can easily be devel-

[2] This is analogous to the apparently tight priors commonly used for the error covariance matrix in the state equation of TVP-VAR models. For instance, Primiceri (2005) uses a prior with mean 0.0001 times the OLS covariance of the VAR coefficients calculated using a training sample. This looks very small but in reality allows for substantial evolution of the VAR coefficients.

oped as a small extension of (9). If we define $P_\tau = HH' + \tau H_\perp H_\perp'$, where $0 \leq \tau \leq 1$ and use a state equation (now written in terms of the $n \times r$ matrices β_t^*).

$$\begin{aligned} \beta_t^* &= P_\tau \beta_{t-1}^* \rho + \epsilon_t \\ \epsilon_t &\sim N(0, I_{nr}) \text{ for } t = 2, ..., T. \\ \beta_1^* &\sim N(0, \frac{1}{1-\rho^2} I_r \otimes P_{\tau^*}), \end{aligned} \tag{12}$$

where $\tau^* = \frac{1-\rho^2}{1-\rho^2\tau^2}$, $P_{\tau^*} = HH' + \tau^* H_\perp H_\perp'$, then we have a prior with attractive properties as formalized in the following proposition.

Proposition 5 *If β_t^* follows the process defined by (12), with $|\rho| < 1$, then:*

1. *b^* is $N(0, V)$, where V is defined by it variances and covariances (for $t = 1, .., T$):*

$$\begin{aligned} var(b_t^*) &= \frac{1}{1-\rho^2} I_r \otimes P_{\tau^*} \\ E(b_t^* b_{t-h}^{*\prime}) &= \frac{\rho^h}{1-\rho^2} I_r \otimes P_{\tau^*\tau^h} \text{ for } h = 2, .., t-1. \end{aligned}$$

2. *$E(\mathfrak{p}_t | \mathfrak{p}_{t-1}) = sp(P_\tau \beta_{t-1})$.*

3. *Let $\bar{\mu}$ be a semi-orthogonal matrix such that $sp(\bar{\mu}) = E(\mathfrak{p}_t | \mathfrak{p}_{t-1})$. The following two inequalities hold:*

$$\begin{aligned} d(\bar{\mu}, H) &\leq d(\beta_{t-1}, H) \\ d(\bar{\mu}, \beta_{t-1}) &\leq d(\beta_{t-1}, H) \end{aligned}$$

4. *The mode of the marginal distribution of $sp(\beta_t)$ is $sp(H)$.*

Proof: See Appendix B.

The fourth property establishes that our prior is an informative marginal prior in the sense defined above. To help further interpret this proposition, note that we can write β_{t-1} as the sum of two components: $\beta_{t-1} = H(H'\beta_{t-1}) + H_\perp(H_\perp'\beta_{t-1})$. The first component is the projection of β_{t-1} on H, and the second is the projection of β_{t-1} on $H_\perp$. Note that $P_\tau \beta_{t-1} =$

$H(H'\beta_{t-1})+\tau H_{\perp}(H'_{\perp}\beta_{t-1})$. Thus, $(P_{\tau}\beta_{t-1})$ results from adding the two components of β_{t-1} while giving less weight to the projection of β_{t-1} on $H_{\perp}$. As τ approaches 1, $E(\mathfrak{p}_t|\mathfrak{p}_{t-1})$ will approach $\mathfrak{p}_{t-1}$ (and the state equation used previously in (9) is obtained), and τ approaching 0 implies that $E(\mathfrak{p}_t|\mathfrak{p}_{t-1})$ is approaching $sp(H)$. In addition, the distance between $sp(H)$ and $sp(\beta_{t-1})$ is greater than both $d(\bar{\mu}, H)$ and $d(\bar{\mu}, \beta_{t-1})$ and, in this sense, $E(\mathfrak{p}_t|\mathfrak{p}_{t-1})$ will be pulled away from $\mathfrak{p}_{t-1}$ in the direction of $sp(H)$.

In words, the conditional prior of $sp(\beta_t)$ given $sp(\beta_{t-1})$ introduced in this sub-section is a weighted average of last period's cointegrating space $(sp(\beta_{t-1}))$ and the subjective prior belief about the cointegrating space $(sp(H))$ and the weights are controlled by τ.

2.5 Summary

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 (9) 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 together with the prior simulation results establish that (9) 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 (9) implies that the marginal prior distribution of the cointegrating space is noninformative, but showed how, using the informative marginal prior in (12) the researcher can incorporate subjective prior beliefs about the cointegrating space if desired.

2.6 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 (9). The extension for the informative marginal prior is a trivial one (simply plug the P_{τ} into the state equation as in (12)). Note that the informative marginal prior depends on the hyperparameter τ. The researcher can, if desired, treat τ as an unknown parameter. The extra block to the Gibbs sampler required by this extension is given in Koop, León-González and Strachan (2008b).

We have specified a state space model for 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). The 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). For the third of 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 (2008a) 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 (4) 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[3] and, thus, is identified under the posterior. Even though β_t is semi-orthogonal, Koop, León-González and Strachan (2008a) 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 in draws of β_t or any feature of the cointegrating space. In the traditional VECM, Koop, León-González and Strachan (2008a) 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 (9) 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 Griddy-Gibbs sampler for this

[3] 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.

parameter.

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

3 Application: A Permanent-Transitory Decomposition for Inflation

In this section, we illustrate our methods using US data from 1953Q1 through 2006Q2 on the unemployment rate (seasonally adjusted civilian unemployment rate, all workers over age 16), u_t, interest rate (yield on three month Treasury bill rate), r_t, and inflation rate (the annual percentage change in a chain-weighted GDP price index), π_t.[4] These variables are commonly-used to investigate issues relating to monetary policy (e.g. Primiceri, 2005) and we contribute to this discussion but focus on issues relating to cointegration.[5] In particular, with cointegration it is common to estimate permanent-transitory variance decompositions which shed insight on the relative roles of permanent and transitory shocks in driving each of the variables. Such issues cannot be investigated with the TVP-VAR methods which are commonly-used with these variables. However, the appropriate policy response to a movement in inflation depends upon whether it is permanent or transitory. Thus, our time-varying cointegration methods are potentially useful for policymakers.

An important concern in the day to day practice of monetary policy is whether shocks to inflation are transitory or permanent. This can be an important consideration in forecasting and in determining the appropriate policy response. For example, the ratio of transitory volatility to total volatility can be used to weight the recent history of inflation for forecasting (see Lansing, 2006a). Further, policymakers often focus attention upon movements in core measures of inflation as an approximation to permanent changes to inflation[6] (see Mishkin, 2007). Recent evidence has suggested that the role of

[4] The data were obtained from the Federal Reserve Bank of St. Louis website, http://research.stlouisfed.org/fred2/.

[5] Note that, by allowing for random walk intercepts, TVP-VARs allow for unit root behavior in the variables (as does our model). Thus, both approaches allow for permanent shocks. Our cointegration-based approach has the additional benefit that standard methods can be used to carry out a permanent-transitory decomposition.

[6] It is important to distinguish between permanent movements in inflation and permanent movements in prices. The latter may be due to transitory deviations in inflation from

permanent shocks in driving US inflation has been decreasing (see Lansing, 2006b). In the application in this section, we shed light on this issue using our time-varying cointegration model. In particular, we show that there has been an increase in the transitory proportion of inflation, particularly after 1980. Furthermore, the uncertainty associated with this transitory component has increased also.

We use the time varying VECM in (4) with one lag of differences $(l = 1)$ and a single cointegrating relationship $(r = 1)$. Using the time-invariant version VECM, the BIC selects $l = 1$ and $r = 1$, which provides support for these choices. We use the state equation for β_t^* in (9). Further modelling details and prior choices are given in Appendix A.

We begin by considering measures of the location of the cointegrating space. These are based upon the distance measure $d(.,.)$ in (11), except rather than measure the distance from the last observed space, we measure the distance from a fixed space so as to give an idea of movement in the estimated cointegrating space. That is, we compute $d_{1,t} = d(\beta_t, H_1)$ where $sp(H_1)$ is a known space with a fixed location. Defining it in this way, $d_{1,t}$ is bounded between zero and one and is a measure of the distance between $\mathfrak{p}_t = sp(\beta_t)$ and $\mathfrak{p}_1 = sp(H_1)$ at time t. A change in the value of $d_{1,t}$ implies movement in the location of $\mathfrak{p}_t$.

A limitation of measuring movements in this way is that, while the space $\mathfrak{p}_t$ may move significantly, it can stay the same distance from $\mathfrak{p}_1$ such that $d_{1,t}$ does not change. Thus, if we observe no change in $d_{1,t}$, this does not imply that $\mathfrak{p}_t$ did not move. This is an unlikely (measure zero) event, but certainly small changes in $d_{1,t}$ do not necessarily indicate the location of $\mathfrak{p}_t$ has not changed very much. A simple solution to this issue is to simultaneously measure the distance from another space $\mathfrak{p}_2 = sp(H_2) \neq \mathfrak{p}_1$. That is, we compute both $d_{1,t} = d(\beta_t, H_1)$ and $d_{2,t} = d(\beta_t, H_2)$. The rationale here is that if $\mathfrak{p}_t$ moves in such a way that $d_{1,t}$ does not change, then $d_{2,t}$ must change provided $\mathfrak{p}_1 \neq \mathfrak{p}_2$. The rules for interpretation then are quite simple. If neither $d_{1,t}$ nor $d_{2,t}$ have changed, then $\mathfrak{p}_t$ has not moved. If either $d_{1,t}$ or $d_{2,t}$ or both have changed, then $\mathfrak{p}_t$ has moved. The choices of $\mathfrak{p}_1$ and $\mathfrak{p}_2$ are arbitrary. The only necessary condition is that they are not the same. However, movements are likely to be more evident if $\mathfrak{p}_1$ and $\mathfrak{p}_2$ are orthogonal to each other. We choose to set $\mathfrak{p}_1$ to be the space that implies the real interest rate is stationary and $\mathfrak{p}_2$ to be the space that implies the unemployment rate

its trend. The former implies an acceleration in the growth of prices.

is stationary. That is, for the ordering $y_t = (\pi_t, u_t, r_t)$ then

$$H_1 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \text{ and } H_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

Figure 3 presents plots of the posterior median and the 25^{th} and 75^{th} percentiles of $d_{1,t}$ and $d_{2,t}$. It is evident that there has been some movement in the cointegrating space. In particular, $\mathfrak{p}_t$ is moving towards $\mathfrak{p}_1$ over the full sample, while $\mathfrak{p}_t$ has moved away from $\mathfrak{p}_2$ by 1970 and stayed far from it for the remainder of the period shown. This suggests the evidence in support of stationary real interest rates has gradually strengthened over time, while the evidence that unemployment is stationary has weakened.

Next, we compute the permanent-transitory variance decomposition for inflation using the approach of Centoni and Cubadda (2003). This variance decomposition is a function of the parameters in the VECM. As, in our approach, these parameters vary over time, the computed transitory proportion of inflation will vary over time. Figure 4 presents a plot of the posterior median and the 25^{th} and 75^{th} percentiles of the time varying proportion of the variance of inflation that is transitory. This figure shows that the transitory component of inflation has historically been low - the median, for example, has always remained below 7%. Permanent shocks, thus, are predominant. But there does appear to have been a slight increase in the size of the transitory component around the middle of the 1980s (as the entire posterior distribution has shifted upwards). Another important observation is the increase in the uncertainty associated with the size of the transitory component. That is, the increase in the interquartile range of the posterior shows how it is becoming more dispersed. At all points in time, the 25^{th} percentile is very close to zero (the value which implies that it is only permanent shocks which are driving inflation). However the 75^{th} percentile increases noticeably after 1980 as the distribution becomes more dispersed. This rise in the size of the transitory component corresponds roughly with the often-reported fall in total volatility of inflation (e.g., Primiceri, 2005).

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 demonstrated 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. Our focus is upon how the proportion of inflation variability that is transitory has evolved. We find that, although permanent shocks are predominant, there has been a slight increase in the transitory component of inflation after 1980 and an increase in the uncertainty associated with the estimate of this component.

References

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

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

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

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.

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.

Johansen, S., 1991, Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, *Econometrica*, 59, 1551-1580.

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.

Koop, G., León-González, R. and Strachan R.W., 2007, On the evolution of monetary policy, working paper available at

http://personal.strath.ac.uk/gary.koop/koop_leongonzalez_strachan_kls5.pdf.

Koop, G., León-González, R. and Strachan, R., 2008a, Efficient posterior simulation for cointegrated models with priors on the cointegration space, forthcoming in *Econometric Reviews*,

manuscript available at http://personal.strath.ac.uk/gary.koop/.

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

manuscript available at http://personal.strath.ac.uk/gary.koop/.

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.

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.

Martins, L. and Bierens, H., 2005, Time varying cointegration, manuscript available at http://econ.la.psu.edu/~hbierens/TVCOINT.PDF.

Mishkin, F., 2007, Headline versus core inflation in the conduct of monetary policy, *Business Cycles, International Transmission and Macroeconomic Policies Conference*, HEC Montreal, Montreal, Canada.

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

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.

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.

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.

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:

$$a_t = a_{t-1} + \zeta_t \tag{13}$$

where $\zeta_t \sim N(0, Q)$.[7] We can rewrite (4) 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,

$$\Delta y_t = A_t Z_t + \varepsilon_t. \tag{14}$$

Vectorizing this equation gives us the form

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

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 (13), 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 (2008b), 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 (4) 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 ε_t

[7] One attractive property of this state equation is that, when combined with (9), it implies $E(\Pi_t|\Pi_{t-1}) = \rho \Pi_{t-1}$. 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.

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

$$\Omega_t = \Lambda_t^{-1} \Sigma_t \Sigma_t' \left(\Lambda_t^{-1}\right)', \tag{15}$$

where Σ_t is a diagonal matrix with diagonal elements $\sigma_{j,t}$ for $j = 1, .., p$ and Λ_t is the lower triangular matrix:

$$\Lambda_t = \begin{bmatrix} 1 & 0 & ... & . & 0 \\ \lambda_{21,t} & 1 & ... & . & . \\ . & . & ... & . & . \\ . & . & ... & 1 & 0 \\ \lambda_{p1,t} & . & ... & \lambda_{p(p-1),t} & 1 \end{bmatrix}.$$

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 $\sigma_t = (\sigma_{1,t}, .., \sigma_{p,t})'$, $h_{i,t} = \ln(\sigma_{i,t})$, $h_t = (h_{1,t}, .., h_{p,t})'$ then Primiceri uses:

$$h_t = h_{t-1} + u_t, \tag{16}$$

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 $\frac{p(p-1)}{2}$ vector as $\lambda_t = \left(\lambda_{21,t}, \lambda_{31,t}, \lambda_{32,t}, .., \lambda_{p(p-1),t}\right)'$. These are allowed to evolve according to the state equation:

$$\lambda_t = \lambda_{t-1} + \xi_t, \tag{17}$$

where ξ_t is $N(0, C)$ and is independent over t and of u_t, ε_t, ζ_t and η_t. We assume the same block diagonal structure for C as in Primiceri (2005).

Prior Distributions

Our model involves four sets of state equations: two associated with the measurement error covariance matrix ((16) and (17)), one for the cointegrating space given in (9) and one for the other conditional mean coefficients (13). The prior for the initial condition for the cointegrating space is already given in (9), 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: $\rho \in [0.999, 1)$.

With TVP-VARs it is common to use training sample priors as in, e.g., Cogley and Sargent, 2005, and Primiceri (2005). In this paper we adapt this strategy to the TVP-VECM. Note that for Ω_t our prior is identical to Primiceri (2005). Our training sample prior estimates a time-invariant VAR using the first ten years of data to choose many of the key prior hyperparameters. To be precise, we calculate OLS estimates of the VAR coefficients, $\widehat{A}$ including $\widehat{\Pi}$, and the error covariance matrix, $\widehat{\Omega}$ and decompose the latter as in (15) to produce $\widehat{\lambda}_0$ and $\widehat{h}_0$ (where these are both vectors stacking the free elements as we did with λ_t and h_t). We also obtain OLS estimates of the covariance matrices of $\widehat{A}$ and $\widehat{\lambda}_0$ which we label $\widehat{V}_A$ and $\widehat{V}_\lambda$. Using the singular value decomposition $\widehat{\Pi} = USV'$, we reorder the elements of U, S, and V such that the diagonal elements of S are in descending order. We then set the elements of $\widehat{a}_0$ corresponding to $\widehat{\alpha}$ to be the first r columns of US. Next, define V_r to be the first r columns of V. We pre-multiply the rows of $\widehat{V}_A$ corresponding to $\widehat{\Pi}$ by $(V_r' \otimes I)$ and post-multiply the corresponding columns of $\widehat{V}_A$ by $(V_r \otimes I)$.

Using these, we construct the priors for the initial conditions in each of our state equations as:

$$a_0 \sim N\left(\widehat{a}_0, 4\widehat{V}_\Pi\right),$$

$$\lambda_0 \sim N\left(\widehat{\lambda}_0, 4\widehat{V}_\lambda\right)$$

and

$$\log(h_0) \sim N\left(\log\left(\widehat{h}_0\right), I_n\right).$$

Next we describe the priors for the error variances in the state equations. We follow the common practice of using Wishart priors for the error precision matrices in the state equations:

$$Q^{-1} \sim W\left(\underline{\nu}_Q, \underline{Q}^{-1}\right) \tag{18}$$

For (18) we set $\underline{\nu}_Q = 40$ and $\underline{Q} = 0.0001\widehat{V}_{\Pi}$.

For W^{-1}:

$$W^{-1} \sim W\left(\underline{\nu}_W, \underline{W}^{-1}\right) \tag{19}$$

with $\underline{\nu}_W = 4$ and $\underline{W} = 0.0001I_n$.

We adopt the block-diagonal structure for C given in Primiceri (2005). Call the two blocks C_1 and C_2. We use a Wishart prior for C_j^{-1}:

$$C_j^{-1} \sim W\left(\underline{\nu}_{Cj}, \underline{C}_j^{-1}\right).$$

with $\underline{\nu}_{C1} = 2, \underline{\nu}_{C2} = 3$ and $\underline{C}_j = 0.01\widehat{V}_{\lambda j}$ for $j = 1, 2$ and $\widehat{V}_{\lambda j}$ is the block corresponding to C_j taken from $\widehat{V}_{\lambda}$. See Primiceri (2005) for further discussion and motivation for these choices.

These values are either identical or (where not directly comparable due to our having a VECM) similar in spirit to those used in Primiceri (2005) and Cogley and Sargent (2005).

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

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

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

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

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

$$C_j^{-1}|Data \sim W\left(\overline{\nu}_{Cj}, \overline{C}_j^{-1}\right)$$

where

$$\overline{\nu}_{Cj} = T + \underline{\nu}_{Cj}$$

and

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

and $a_t^{(j)}$ are the elements of a_t corresponding to C_j.

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 (9). We therefore draw this scalar using a Griddy-Gibbs algorithm based on our Uniform prior.

In our empirical work, we run our sampler for 8000 burn-in replications and then include the next 16000 replications.

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 (9) 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:

$$ {}_0F_1(\frac{1}{2}r; \frac{1}{4}G_t) = \int_{\tilde{O}_r} \prod_{i=1}^{r} (exp(\sqrt{g_i} q_{ii}) + exp(-\sqrt{g_i} q_{ii}))[dQ] \qquad (20) $$

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 (20). Note that the function $\exp(cx) + \exp(-cx)$ is an increasing function of x when both x and c are positive. Thus, from expression (20), ${}_0F_1(\frac{1}{2}r; \frac{1}{4}G_t)$ 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 $Z=U'\beta_tO$, 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:

$$\begin{aligned}&A_L\int_{S_1}\left(\sum_{l=1}^{r}(\tilde{w}_{jl})^2 - \sum_{l=1}^{r}(\hat{w}_{pl})^2\right)\left(\exp\left\{tr(M\tilde{W}N)\right\} - \exp\left\{tr(M\bar{W}N)\right\}\right)[dW][dN] = \\ &A_L\int_{S_1}\left(\sum_{l=1}^{r}(\tilde{w}_{jl})^2 - \sum_{l=1}^{r}(\hat{w}_{pl})^2\right)\left({}_0F_1(\tfrac{1}{2}r;\tfrac{1}{4}M\tilde{W}\tilde{W}'M) - {}_0F_1(\tfrac{1}{2}r;\tfrac{1}{4}M\bar{W}\bar{W}'M)\right)[dW]\end{aligned} \tag{21}$$

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 (21) 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 (21) is not negative.

Proof of Proposition 3: The density function of $P_t = \beta_t\beta_t'$ given by expression (10) 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 (9) 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:

$$K_{t+1} = \rho^2 K_t + \rho^2\epsilon_t'\epsilon_t + \rho^3\beta_{t-1}^{*'}\epsilon_t + \rho^3\epsilon_t'\beta_{t-1}^* \tag{22}$$

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 (22), and noting that $E(\epsilon_t'\epsilon_t|K_t,...,K_2) = nI_r$ we get $E(K_{t+1}|K_t,...,K_2) = \rho^2K_t + \rho^2 nI_r$. 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:

$$E(k_{tij}|K_{(t-h)},...,K_2) = \rho^{2h}k_{(t-h)ij} + n\delta_{ij}\sum_{c=1}^{h}\rho^{2c} \tag{23}$$

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 (23) times $k_{(t-h)kl}$, subtracting $E(k_{tij})E(k_{(t-h)kl})$ and taking expectations:

$$cov(k_{tij}, k_{(t-h)kl}) = (\rho)^{2h}E(k_{(t-h)ij}k_{(t-h)kl}) + \delta_{ij}E(k_{(t-h)kl})n\sum_{c=1}^{h}(\rho)^{2c} - E(k_{(t-h)kl})E(k_{tij}) \tag{24}$$

From the properties of the Wishart distribution, all expectations in the right side of equation (24) 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:

$$\begin{aligned} cov(k_{tij}, k_{(t-h)kl}) = (\rho)^{2h}E(k_{(t-h)ij}k_{(t-h)kl}) - E(k_{(t-h)kl})E(k_{tij}) = \\ (\rho)^{2h}E(k_{(t-h)ij}k_{(t-h)kl}) = (\rho)^{2h}cov(k_{(t-h)ij}, k_{(t-h)kl}) \end{aligned} \tag{25}$$

From the properties of the Wishart distribution with diagonal parameter matrix, (25) 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 $i = j = k = l$, (24) can be written as:

$$\begin{aligned} cov(k_{tii}, k_{(t-h)ii}) = \rho^{2h}[E(k^2_{(t-h)ii}) - (E(k_{(t-h)ii}))^2] + \\ (\rho^{2h} - 1)[E(k_{(t-h)ii})]^2 + E(k_{(t-h)ii})n\sum_{c=1}^{h}\rho^{2c} \end{aligned} \tag{26}$$

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:

$$(\rho^{2h} - 1)(E(k_{(t-h)ii}))^2 + E(k_{(t-h)ii})n\sum_{c=1}^{h}\rho^{2c} = 0 \tag{27}$$

Thus, (26) 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 (27) and noting that $E(k_{(t-h)ii}) = E(k_{(t-h)kk})$, it can be shown that (24) is equal to zero.

Proof of Proposition 5:

First note that $P_\tau P_\tau = P_{\tau^2}$ and so $(P_\tau)^n = P_{\tau^n}$. Using the properties of the *vec* operator, (12) can be written as $b_t^* = \rho(I_r \otimes P_\tau)b_{t-1}^* + vec(\epsilon_t)$, and using this equation recursively we get:

$$b_t^* = \rho^{t-1}(I_r \otimes P_{\tau^{t-1}})b_1^* + vec(\epsilon_t) + \sum_{i=1}^{t-2}(\rho^i(I_r \otimes P_{\tau^i}))vec(\epsilon_{t-i})$$

which implies $E(b_t^*) = 0$ and:

$$var(b_t^*) = \rho^{2(t-1)}(I_r \otimes P_{\tau^{t-1}})var(b_1^*)(I_r \otimes P_{\tau^{t-1}}) + I_{nr} + \sum_{i=1}^{t-2}\rho^{2i}(I_r \otimes P_{\tau^{2i}}) \quad (28)$$

where we have used that P_τ is a symmetric matrix and that ϵ_t have zero serial correlation. Note that $I_{nr} = I_r \otimes I_n = I_r \otimes (HH' + H_\perp H_\perp')$ and recall that we have assumed $var(b_1^*) = \frac{1}{1-\rho^2} I_r \otimes P_{\tau^*}$. Thus, (28) can be written as:

$$var(b_t^*) = I_r \otimes \left[HH' \left(\frac{\rho^{2(t-1)}}{1-\rho^2} + 1 + \sum_{i=1}^{t-2}\rho^{2i} \right) + H_\perp H_\perp' \left(\frac{\rho^{2(t-1)}}{1-\rho^2}\tau^{2(t-1)}\tau^* + 1 + \sum_{i=1}^{t-2}(\rho\tau)^{2i} \right) \right]$$
$$= \frac{1}{1-\rho^2} I_r \otimes P_{\tau^*}$$

where we have used that $\tau^* = (1-\rho^2)/(1-(\rho\tau)^2)$. To calculate the covariance, use (12) recursively to obtain $b_t^* = \rho^h(I_r \otimes P_{\tau^h})b_{t-h}^* + vec(\epsilon_t) + \sum_{i=1}^{h-1}(\rho^i(I_r \otimes P_{\tau^i}))vec(\epsilon_{t-i})$. Thus,

$$E(b_t^* b_{t-h}^{*\prime}) = \rho^h(I_r \otimes P_{\tau^h})E\left(b_{t-h}^* b_{t-h}^{*\prime}\right) = \frac{\rho^h}{1-\rho^2} I_r \otimes P_{\tau^*\tau^h}$$

The second property follows from Proposition 2.
If either $\tau = 0$ or $\tau = 1$, it is clear that the third property holds. For example, $\tau = 0$ implies $sp(\bar{\mu}) = sp(H)$ and so $d(\mu, H) = 0 < d(\beta_{t-1}, H)$ and $d(\bar{\mu}, \beta_{t-1}) = d(\beta_{t-1}, H)$. In order to prove the third property when $0 < \tau < 1$, let $\bar{\mu} = P_\tau\beta_{t-1}(\beta_{t-1}' P_{\tau^2}\beta_{t-1})^{-1/2}$, which is semi-orthogonal. To simplify notation, from now on in this proof let $\beta \equiv \beta_{t-1}$ (i.e. we are dropping the t-1). Note that: $P_\tau(H_\perp H_\perp')P_\tau = \tau^2 H_\perp H_\perp'$ and define $K = \beta' P_{\tau^2}\beta$. Thus:

$$d^2(\bar{\mu}, H) = tr(\bar{\mu}' H_\perp H_\perp' \bar{\mu}) = tr(\tau^2 K^{-1/2}(\beta' H_\perp H_\perp' \beta)K^{-1/2}) = tr(\tau^2(\beta' H_\perp H_\perp' \beta)K^{-1})$$

Let USU' be the singular value decomposition of $\beta' H_\perp H_\perp' \beta$, with $S = diag(s_1, ..., s_r)$. From now we will show that $d^2(\bar{\mu}, H)$ can be written as in (29). Since $(H, H_\perp)(H, H_\perp)' = HH' + H_\perp H_\perp' = I_r$, we have that $\beta HH'\beta' + \beta H_\perp H_\perp' \beta' = I_r$ and so $\beta HH'\beta' = I_r - \beta H_\perp H_\perp' \beta'$. Using the singular value decomposition of $\beta H_\perp H_\perp' \beta'$ we have that $\beta HH'\beta' = I_r - USU' = UU' - USU' = U(I_r - S)U'$. Note that $(\beta HH'\beta')U = U(I_r - S)$, thus U is a matrix with the eigenvectors of $\beta HH'\beta'$ and $(I_r - S)$ are the eigenvalues. Since $\beta HH'\beta'$ is positive semidefinite, we have $(I_r - S) \geq 0$, which implies $0 \leq S \leq I_r$, that is, the singular values of $\beta H_\perp H_\perp' \beta'$ are between 0 and 1. Because $K = \beta' P_{\tau^2} \beta = \beta HH'\beta' + \tau^2 \beta H_\perp H_\perp' \beta'$ and $\beta HH'\beta' = I_r - \beta H_\perp H_\perp' \beta'$, we can write K as $I_r - (1-\tau^2)\beta' H_\perp H_\perp' \beta = UU' - (1-\tau^2)USU' = U\tilde{S}U'$, with $\tilde{S}$ being a diagonal matrix $\tilde{S} = I_r - (1-\tau^2)S$ with strictly positive elements $(0 < \tau)$. Hence $K^{-1} = U\tilde{S}^{-1}U'$ and $(\tau^2 \beta' H_\perp H_\perp' \beta)K^{-1} = \tau^2 US\tilde{S}^{-1}U'$. Because the trace of a matrix is the sum of its eigenvalues, $d^2(\bar{\mu}, H)$ is the sum of the elements of $\tau_2 S\tilde{S}^{-1}$. Recall $S = diag(s_1, ..., s_r)$ so that:

$$d^2(\bar{\mu}, H) = tr(\tau^2(\beta' H_\perp H_\perp' \beta)K^{-1}) = \sum_{i=1}^{r} \frac{s_i \tau^2}{1 - (1-\tau^2)s_i} \tag{29}$$

Recall that $d^2(\beta, H) = tr(\beta' H_\perp H_\perp' \beta) = \sum_{i=1}^{r} s_i$. Because $0 \leq s_i \leq 1, 0 < \tau < 1$, it can be verified that $\tau^2/(1-(1-\tau^2)s_i) \leq 1$. Hence, $d^2(\bar{\mu}, H) \leq d^2(\beta, H)$.

In order to prove that $d(\bar{\mu}, \beta) \leq d(\beta, H)$, we will show that $d^2(\bar{\mu}, \beta)$ can be written as in (30). First note that:

$$d^2(\bar{\mu}, \beta) = tr(\beta' P_\tau \beta_\perp \beta_\perp' P_\tau \beta K^{-1})$$

Note also that $P_\tau = HH' + \tau H_\perp H_\perp' = (I_r - H_\perp H_\perp') + \tau H_\perp H_\perp' = I_r - (1-\tau)H_\perp H_\perp'$, and so $\beta' P_\tau \beta_\perp = -(1-\tau)\beta' H_\perp H_\perp' \beta_\perp$. Thus,

$$d^2(\bar{\mu}, \beta) = tr((1-\tau)^2 \beta' H_\perp H_\perp' \beta_\perp \beta_\perp' H_\perp H_\perp' \beta K^{-1})$$

Note that $\beta_\perp \beta_\perp' = (I_r - \beta\beta')$ and thus:

$$d^2(\bar{\mu}, \beta) = tr((1-\tau)^2 \beta' H_\perp H_\perp' \beta K^{-1}) - tr((1-\tau)^2 (\beta' H_\perp H_\perp' \beta)(\beta' H_\perp H_\perp' \beta)K^{-1})$$

As before, let USU' be the singular value decomposition of $\beta' H_\perp H_\perp' \beta$ and recall that $K^{-1} = U\tilde{S}^{-1}U'$ and that $(\beta' H_\perp H_\perp' \beta)K^{-1} = US\tilde{S}^{-1}U'$. Thus:

$$d^2(\bar{\mu}, \beta) = tr((1-\tau)^2 US\tilde{S}^{-1}U') - tr((1-\tau)^2 (US^2\tilde{S}^{-1}U') = \sum_{i=1}^{r} \frac{s_i(1-s_i)(1-\tau)^2}{1-(1-\tau^2)s_i} \tag{30}$$

Because $\frac{(1-s_i)(1-\tau)^2}{1-(1-\tau^2)s_i} < 1$ when $0 < \tau \leq 1$, $0 \leq s_i \leq 1$, then $d^2(\bar{\mu}, \beta) \leq d^2(\beta, H)$.

To prove the fourth property, note that because the marginal distribution of β_t^* is a matric normal distribution with zero mean and covariance matrix $1/(1-\rho^2) I_r \otimes P_{\tau^*}$, it follows that the marginal distribution of β_t on the Stiefel manifold is the matrix angular Gaussian distribution (Chikuse, 2003, p.40) with parameter matrix P_{τ^*}. The density function of $sp(\beta_t)$ can be written in terms of β_t (Chikuse (2003), p. 15), if understood to be written with respect to the appropriate measure in the Grassmann manifold (which can be found in Chikuse, 2003, p. 15, expression 1.4.3). This density can be obtained using Theorem 2.4.8 in Chikuse (2003, p. 46) and is proportional to:

$$|\beta_t' P_{(\tau^*)^{-1}} \beta_t|^{-n/2}$$

where $P_{(\tau^*)^{-1}} = (P_{\tau^*})^{-1} = HH' + (\tau^*)^{-1} H_\perp H_\perp'$. Let us use the singular value decomposition to write $\beta_t' H_\perp H_\perp' \beta_t = USU'$, where S is diagonal containing the singular values. Reasoning in the same manner as we did in the paragraph just before equation (29), we get $\beta_t' P_{(\tau^*)^{-1}} \beta_t = U\hat{S}U'$, where $\hat{S} = I_r - (1 - (\tau^*)^{-1})S = I_r + ((\tau^*)^{-1} - 1)S$. Thus, the density function is proportional to $|\hat{S}|^{-n/2}$. Since $(\tau^*)^{-1} > 1$, the density is maximized when $S = 0$, which happens when $sp(\beta_t) = sp(H)$.

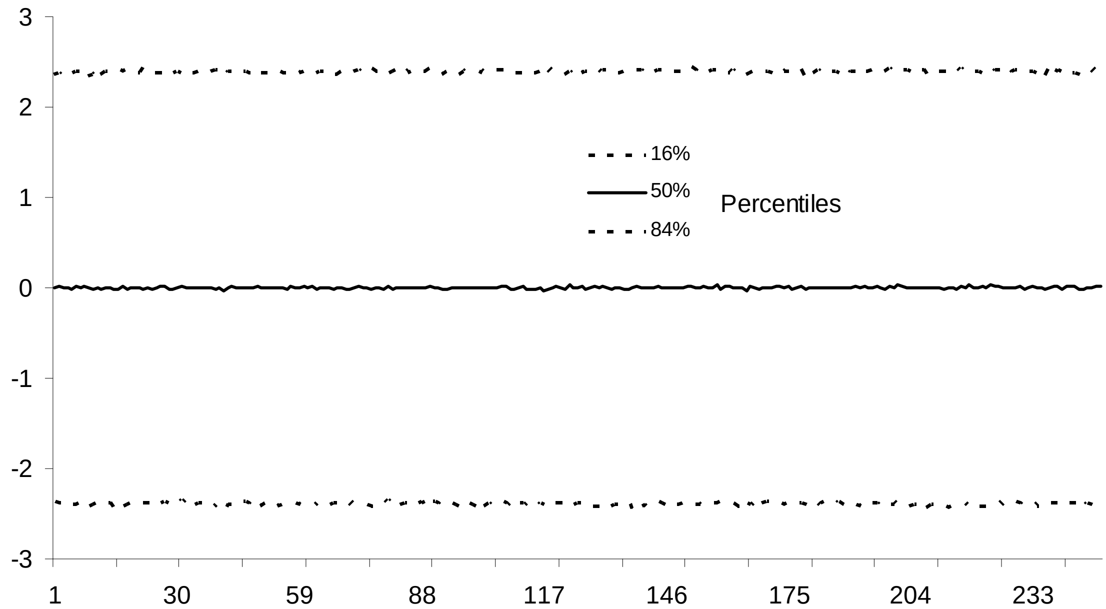

Figure 1: Properties of Change in angle defining cointegration space between $t-1$ and t ($T = 250$ and $\rho = 0.999$).

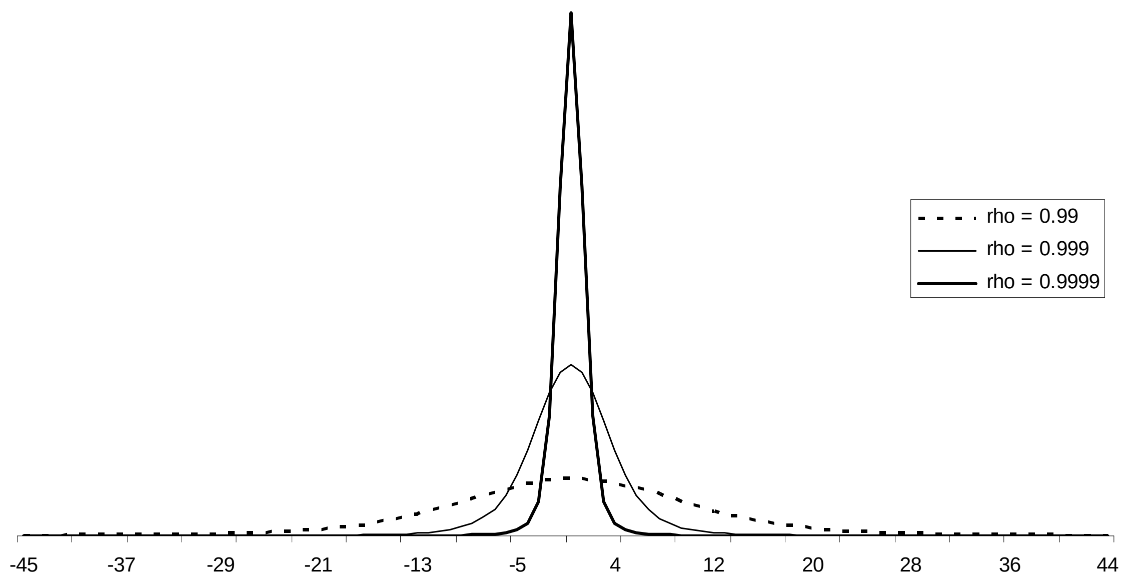

Figure 2: Prior p.d.f. of the change in the angle defining the cointegrating space between $T-1$ and T.

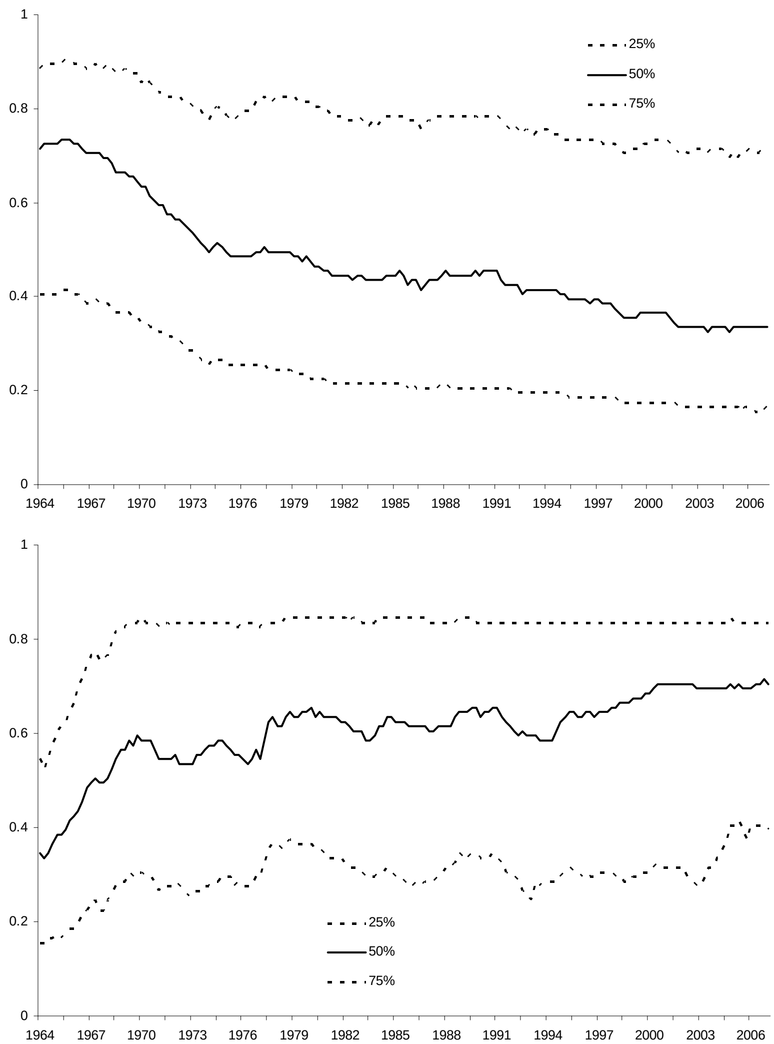

Figure 3: The top panel plots the 25th, 50th and 75th percentiles for $d_{1,t}$. The bottom panel plots the same percentiles for $d_{2,t}$.

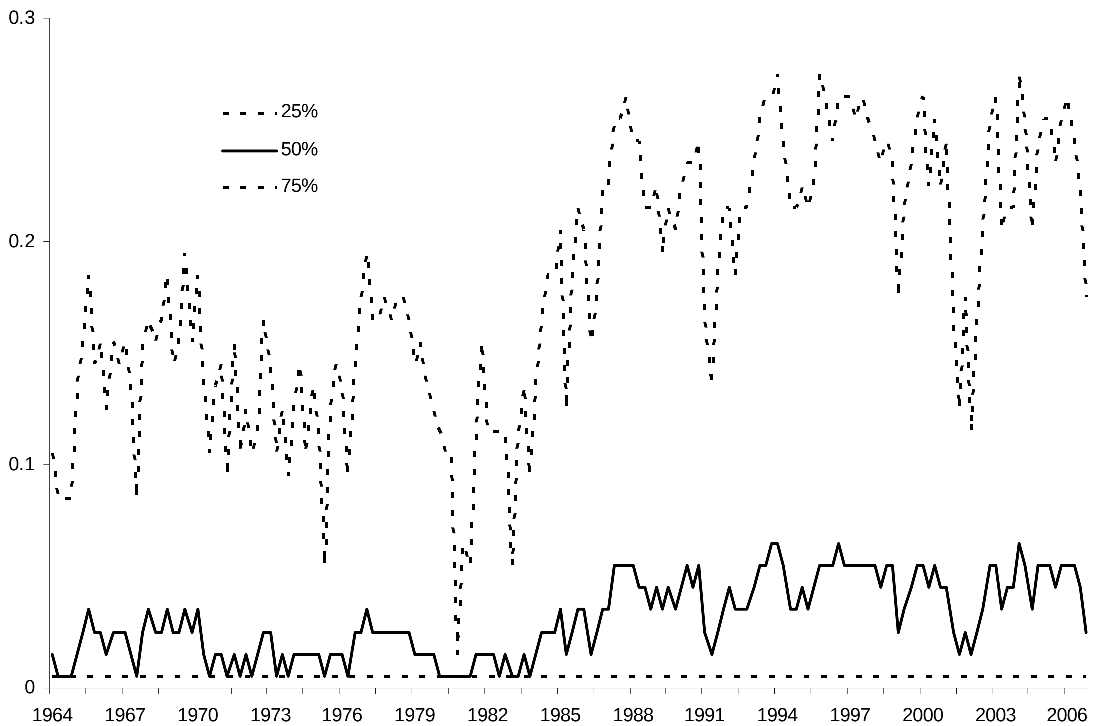

Figure 4: This figure plots the 25th, 50th and 75th percentiles for the distribution of the proportion of the variance of inflation that is transitory.