EUROPEAN CENTRAL BANK

WORKING PAPER SERIES

NO 692 / NOVEMBER 2006

BCE ECB EZB EKT EKP

BAYESIAN INFERENCE IN COINTEGRATED VAR MODELS

WITH APPLICATIONS TO THE DEMAND FOR EURO AREA M3

by Anders Warne

EUROPEAN CENTRAL BANK

WORKING PAPER SERIES

NO 692 / NOVEMBER 2006

BAYESIAN INFERENCE IN COINTEGRATED VAR MODELS

WITH APPLICATIONS TO THE DEMAND FOR EURO AREA M3 [1]

by Anders Warne [2]

In 2006 all ECB publications feature a motif taken from the €5 banknote.

1 I am grateful to Günter Coenen, Jaromír Hurník, Julian von Landesberger, Frank Smets, Ivars Tillers, Mattias Villani, participants at the ECB Workshop on Monetary Analysis, 13–14 June, 2005 in Frankfurt am Main, at the Swiss National Bank in Zürich, and at the ECB (March 2006) for helpful discussions. I am also grateful to an anonymous referee for suggestions that have helped improve the paper. The views expressed in this paper are those of the author and do not necessarily reflect those of the European Central Bank.

2 European Central Bank, Postfach 160319, 60311 Frankfurt am Main, Germany; e-mail: anders.warne@ecb.int; Web site: http://www.texlips.net/warne/.

Address
Kaiserstrasse 29
60311 Frankfurt am Main, Germany

Postal address
Postfach 16 03 19
60066 Frankfurt am Main, Germany

Telephone
+49 69 1344 0

Internet
http://www.ecb.int

Fax
+49 69 1344 6000

Telex
411 144 ecb d

ISSN 1561-0810 (print)
ISSN 1725-2806 (online)

CONTENTS

Abstract: The paper considers a Bayesian approach to the cointegrated VAR model with a uniform prior on the cointegration space. Building on earlier work by Villani (2005b), where the posterior probability of the cointegration rank can be calculated conditional on the lag order, the current paper also makes it possible to compute the joint posterior probability of these two parameters as well as the marginal posterior probabilities under the assumption of a known upper bound for the lag order. When the marginal likelihood identity is used for calculating these probabilities, a point estimator of the cointegration space and the weights is required. Analytical expressions are therefore derived of the mode of the joint posterior of these parameter matrices. The procedure is applied to a money demand system for the euro area and the results are compared to those obtained from a maximum likelihood analysis.

Keywords: Bayesian inference, cointegration, lag order, money demand, vector autoregression.

JEL Classification Numbers: C11, C15, C32, E41.

A non-stationary time series vector process for which there exists linear combinations that are stationary is said to be cointegrating. This concept was first considered by Granger (1983), who used it for modelling long-run economic relations. To date, the most frequently used method for cointegration analysis is probably the maximum likelihood approach first suggested by Johansen (1988); see also Johansen (1991, 1996) and Johansen and Juselius (1990, 1992). This procedure starts from a vector autoregressive (VAR) model for a set of variables in levels with normally distributed residuals. Without imposing any restrictions on its parameters, the VAR be can rewritten on error correction form with the long-run impact matrix of the VAR appearing as the coefficient matrix on the lagged levels, while the variables are otherwise expressed in first differences. If the variables are cointegrated, then the long-run impact matrix of the VAR has reduced rank, and its rank is equal to the number of cointegration relations. This general relationship between the concept of cointegration and the vector error correction or cointegrated VAR model was first established in the seminal paper by Engle and Granger (1987). The cointegration rank testing procedure developed by Johansen and co-authors therefore attempts to determine the rank of the long-run impact matrix.

During the last decade a number of Bayesian approaches to cointegration have also appeared in the literature. These include work by Kleibergen and van Dijk (1994), Bauwens and Lubrano (1996), Geweke (1996), Bauwens and Giot (1998), Kleibergen and Paap (2002), Strachan (2003), Strachan and Inder (2004), and Villani (2005b); see Koop, Strachan, van Dijk, and Villani (2006) for a review of Bayesian approaches to cointegration. All these studies share the assumption that the possibly cointegrated VAR model has normally distributed residuals and, hence, they have the same likelihood function as the classical Johansen method. Given that the posterior distribution of the cointegration rank exists, one Bayesian approach to cointegration rank determination is to select the rank whose posterior probability is the largest.

From a practical perspective, a Bayesian approach to cointegration is advantageous relative to a classical for several reasons. For example, it produces whole probability distributions for each parameter that are valid for any sample size. It also makes it possible to deal with inferences on the cointegration rank and other restrictions on the model parameters. However, the question of identification becomes more delicate within a Bayesian framework. Since the long-run impact matrix of the VAR has reduced rank under the assumption of cointegration it can be rewritten as a product of two matrices, where the first has fewer columns than rows and the second has fewer rows than columns. The second of these matrices represents the cointegration relations, while the first represents the weights on these relations. It is possible to premultiply the cointegration relations with a full rank matrix and postmultiply the weights by the inverse of this matrix without affecting the product of the two matrices. Since the full rank matrix used in these operations is arbitrary, the individual parameters of the cointegration relations and the weights are not uniquely determined without making some further assumptions. This identification problem also appears in the classical literature, where the Johansen procedure provides a beautiful treatment for ensuring that all the parameters of the cointegrated VAR are identified. However, the Johansen identification procedure makes use the observed data and can therefore not be applied under a pure Bayesian scheme.

The procedure suggested by Villani (2005b) considers a prior for the cointegration space rather than a prior for the exactly identified cointegration relations. Villani then shows that a uniform prior on the cointegration space implies that the prior for the exactly identified cointegration relations is Cauchy. The general idea of specifying the prior for the cointegration space rather then for the cointegration relations is also examined in papers by Strachan and Inder (2004), Strachan and van Dijk (2003, 2004, 2005), and Koop, Leon-Gonzalez, and Strachan (2006).

The study by Villani lists four properties that a reference prior for empirical cointegration analysis should meet. First, it should have relatively few "hyperparameters", each with a clear interpretation. Second, it should be transparent in the sense that a practioner can understand the type of information it conveys. Third, posterior calculations should be straightforward and it should be possible to perform them on a routine basis without the need for fine tuning in each application. Fourth, the computation of the posterior distribution of the cointegration rank should be feasible.

The current paper extends the Bayesian approach suggested by Villani (2005b) in two important dimensions. First, by allowing for a proper prior distribution on the parameters on lagged endogenous variables it is not necessary (as in Villani, 2005b) to determine the lag order of the VAR model prior to the cointegration rank analysis. Instead, the marginal posterior distributions of both the cointegration rank and the lag order (subject to an upper bound for the latter parameter) can be computed. Second, the method for estimating the posterior rank probabilities that Villani (2005b) advocates is based on the so called marginal likelihood identify; cf. Chib (1995). This identity states that the marginal likelihood is equal to the density of the data conditional on the parameters times the prior density divided by the posterior density, where these three densities need to be evaluated at a point in the support for the parameters. The identity holds for any such point, but a point where the posterior density is high is typically preferable. An analytical expression of the posterior mode is therefore derived in the paper.

As an application the euro area money demand system in Bruggeman, Donati, and Warne (2003) is re-examined using the Bayesian approach presented in the paper. The money demand model has 6 endogenous variables: real M3, real GDP, annualized quarterly GDP deflator inflation, a short-term interest rate, a long-term rate, and an own rate of return on M3, as well as an unrestricted constant. For this model Bruggeman, Donati, and Warne find evidence of two long-run relations. They interpret these as a long-run money demand and a long-run pricing relation for the own rate of return on M3. The data set in Bruggeman, Donati, and Warne is extended from 2001:Q4 to 2004:Q4, and the number of cointegration relations based on Bayesian methods are compared with the outcome of classical tests. Moreover, Bayesian posterior confidence bands of the income and interest rate semi-elasticities of long-run money demand are compared with both asymptotic and bootstrapped error bands.

1. Introduction

The general link between the concept of cointegration (Granger, 1983) and the error correction model — or cointegrated vector autoregressive (VAR) model — was first established in the highly influential paper by Engle and Granger (1987). In the wake of this study, an extensive literature has emerged within the *classical* domain on estimation and inference in such models. The most frequently used method in practise is probably the Gaussian maximum likelihood based approach advocated by Johansen and co-authors in a series of articles; see, e.g., Johansen (1988, 1991) and Johansen and Juselius (1990, 1992), or more recently Johansen (1996) for a comprehensive treatment of the cointegrated VAR. Other classical approaches, such as reduced rank regression (Ahn and Reinsel, 1990, and Reinsel and Ahn, 1992), quasi maximum likelihood (Pesaran and Shin, 2002), fully modified least squares (Phillips and Hansen, 1990), canonical cointegrating regression (Park, 1992), and dynamic least squares (Phillips and Loretan, 1991, Saikkonen, 1991, and Stock and Watson, 1993) have also been considered in empirical analyses of macroeconomic and financial time series.

During the last decade a number of *Bayesian* cointegration approaches have also appeared. These include work by Kleibergen and van Dijk (1994), Bauwens and Lubrano (1996), Geweke (1996), Bauwens and Giot (1998), Kleibergen and Paap (2002), Strachan (2003), Strachan and Inder (2004), and Villani (2005b). From a practical perspective, a Bayesian approach to cointegration is advantageous for several reasons. For example, it produces whole probability distributions for each parameter that are valid for any sample size. It also makes it possible to deal with inferences on the cointegration rank and other restrictions on the model parameters; see, e.g., Bauwens, Lubrano, and Richard (1999) and Koop, Strachan, van Dijk, and Villani (2006) and references therein.

A crucial step in a Bayesian analysis is the choice of prior distribution and in the literature on cointegration several priors have been suggested. The degree of motivation has varied, but most suggestions focus on vague priors that add only a small amount of information into the analysis. The study by Villani (2005b) is less concerned with whether or not the prior is "non-informative" and instead considers a sound prior which is intended to appeal to practitioners. Villani lists four properties that such a prior should have. First, the prior should have relatively few "hyperparameters", each with a clear interpretation. Second, it should be transparent in the sense that a practitioner can understand the type of information it conveys. Third, posterior calculations should be straightforward and be performed on a routine basis without the need for fine tuning in each application. And, finally, the computation of the posterior distribution of the cointegration rank should be feasible.

The main purpose of this paper is to extend the Bayesian approach suggested by Villani (2005b) in two important dimensions. First, I will allow for an informative prior distribution of the parameters on lagged endogenous variables, i.e., the parameters representing the short-run dynamics. Although the computation of posterior cointegration rank probabilities does not require an informative prior on these parameters given that the lag order is treated as known, this is a limitation in practise. One may, of course, condition on a long lag order, but then the number of degrees of freedom may be sharply reduced. By instead letting the prior on the short-run dynamics be informative this loss of degrees of freedom can be avoided. Moreover, it is then possible to calculate conditional, joint, and marginal posterior probabilities of the cointegration rank and of the lag order. Hence, the extension allows for a richer analysis of these parameters.

Second, the approach for computing the posterior rank probabilities that Villani (2005b) advocates is based on the so called basic marginal likelihood identity; cf. Chib (1995). This identity (Bayes rule) simply states that the marginal likelihood is equal to the density of the data conditional on the parameters times the prior density divided by the posterior density, where all densities need to be evaluated at a point in the support of these parameters. Within the current context, this approach requires the user to select a point for the free parameters in the identified cointegration vectors and for the parameters on these vectors. Such a point should preferably have high posterior density and the posterior mode of the joint density of these parameters is therefore the most natural candidate. Below, I will present a simple analytical procedure for determining this point for the reference prior in Villani (2005b) as well as for the extension to an informative prior on the short-run dynamics.

As an application I will re-examine the euro area money demand system in Bruggeman, Donati, and Warne (2003) — henceforth, BDW — using the Bayesian approach. The BDW model has 6 endogenous variables: real M3, real GDP, annualized quarterly GDP deflator inflation, a short-term interest rate, a long-term rate, and an own rate of return on M3, as well as an unrestricted constant. For this model they find evidence of two long-run relations that they interpret as money demand and a pricing relation for the own rate of return on M3.

The data set in BDW is extended from 2001:Q4 to 2004:Q4, and the number of cointegration relations based on Bayesian methods will be compared with the outcome of classical tests. Moreover, Bayesian posterior confidence bands of the income and interest rate semi-elasticities of long-run money demand will be compared with both asymptotic and bootstrapped error bands.

The remainder of the paper is organized as follows. The cointegrated VAR model is introduced in Section 2. The following Section gives a brief background to issues regarding the selection of a prior for the cointegrated VAR model before defining the priors used in this paper. Section 4 presents the full and marginal conditional posteriors needed for Gibbs sampling. In addition, expressions for computing the marginal likelihoods and the posterior probabilities of the possible cointegration ranks are given, along with analytic expressions for the posterior mode, as well as results on lag order determination. Section 5 turns to the empirical analyses of the euro area money demand system, while Section 6 summarizes the main conclusions. Proofs of some Propositions are collected in the Appendices.

2. The Cointegrated VAR Model

Let x_t be a p-dimensional process represented by a cointegrated VAR model with r stationary long-run relations

$$\Delta x_t = \Phi D_t + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} + \alpha\beta' x_{t-1} + \varepsilon_t, \quad t = 1, \ldots, T. \tag{1}$$

The errors ε_t are assumed to be i.i.d. $N_p(0,\Omega)$, where Ω is positive definite. The remaining parameters are α $(p \times r)$, β $(p \times r)$, $\Gamma_1, \ldots, \Gamma_{k-1}$ $(p \times p)$, and Φ $(p \times d)$ for some $r \in \{0, 1, \ldots, p\}$. It is assumed that α and β are full rank matrices with r being the cointegration rank. The Γ_i matrices govern the short-run dynamics of the process, and D_t $(d \times 1)$ is a vector of constant, trend, seasonal dummies, or other deterministic or exogenous variables; see, e.g., Johansen (1996) for a thorough treatment of the cointegrated VAR model. The lag order, k, is for now assumed to be known or at least determined before the Bayesian assessment of the cointegrated VAR, but I will relax this assumption in Section 4.4.

The representation in (1) can be reformulated as

$$Z_{0t} = \Phi D_t + \Gamma Z_{2t} + \alpha\beta' Z_{1t} + \varepsilon_t, \tag{2}$$

where $\Gamma = [\Gamma_1 \ \cdots \ \Gamma_{k-1}]$, $Z_{0t} = \Delta x_t$, $Z_{1t} = x_{t-1}$, and $Z_{2t} = [\Delta x'_{t-1} \ \cdots \ \Delta x'_{t-k+1}]'$. A more compact form of the model is often useful. Specifically, let

$$Z_0 = \Phi D + \Gamma Z_2 + \alpha\beta' Z_1 + \varepsilon. \tag{3}$$

Here, $Z_0 = [Z_{01} \ \cdots \ Z_{0T}]$ is a $p \times T$ matrix, while D, Z_1, Z_2, and ε are defined in a similar manner with all matrices having T columns. I shall also use the expression $\mathfrak{D} = \{Z_0, Z_1, Z_2, D\}$ for all available data.

3. Prior Distributions

3.1. *Background*

A crucial step in a Bayesian analysis is the choice of prior distribution of the parameters. Several priors have been suggested in the literature on the cointegrated VAR model; see, e.g., Koop et al. (2006) for a survey. Building on priors used for linear models, such as the VAR model, early work includes the studies by Kleibergen and van Dijk (1994), Geweke (1996), Bauwens and Lubrano (1996), and Bauwens and Giot (1998). In these studies both standard informative and non-informative priors are examined. By standard it should be understood that either a Gaussian or a flat prior is assumed for the parameters $(\alpha, \beta, \Phi, \Gamma)$, and an inverted Wishart or flat prior for Ω; see, e.g., Zellner (1971) for a treatment of the multivariate regression model.

However, there are two important features of the cointegrated VAR which make the standard priors unsuitable for cointegration analysis. First, the reduced rank restrictions on $\Pi = \alpha\beta'$ in (1) introduces a non-linearity in the otherwise linear model. Second, this non-linearity results in the (α, β) parameters being non-identified in the sense that, e.g., only the space spanned by the columns of β can be uniquely determined. This is termed the global non-identification issue by Koop et al. (2006). Accordingly, the cointegration vectors must be restricted for all parameters to be identified.

Let c be a known $p \times r$ matrix of rank r with $\bar{c} = c(c'c)^{-1}$. Linear identifying restrictions on β may then either be expressed as $\beta_c = \beta(\bar{c}'\beta)^{-1}$ or, equivalently, as

$$\beta_c = c + c_{\perp}\Psi, \tag{4}$$

where $\Psi = \bar{c}'_{\perp}\beta(\bar{c}'\beta)^{-1}$. For example, we may let c be given by the first r columns of I_p. When β is just identified as in (4), $\alpha_c = \alpha\beta'\bar{c}$ so that $\alpha_c\beta'_c = \alpha\beta' = \Pi$.

Notice that the linear restrictions in (4) involve the assumption that $c'\beta$ has full rank r. If this assumption is false, then the restrictions are not valid. However, even when the restrictions are valid, this does not settle the issue of selecting a prior for the cointegration vectors. As shown by Strachan and van Dijk (2004), a flat prior on the just identified parameters of β (Ψ) favors the cointegration spaces near the region where the linear normalization of β is invalid.

When α has reduced rank, such as $\alpha = 0$, the posterior distribution for the free parameters of β conditional on α is equal to its prior. This follows directly from noting that the free parameters of β do not enter the data density at $\alpha = 0$, thus, yielding a local non-identification issue; see Kleibergen and van Dijk (1994). Imposing weak exogeneity restrictions on α may, as shown by Strachan and van Dijk (2004), also lead to problems by resulting in an improper posterior.

To deal with (some of) these issues, three approaches have recently been suggested; see Koop et al. (2006) for a discussion. The most promising and natural of these considers a prior on the cointegration space. This approach has been used by, e.g., Villani (2000, 2005b), Strachan and Inder (2004), Strachan and van Dijk (2003, 2004, 2005), and Koop, Leon-Gonzalez, and Strachan (2006).

The support for the cointegration space is based on the set of all $p \times r$ orthonormal matrices. This set is called the Stiefel manifold and is a compact space which admits a uniform distribution. Since β can always be transformed into a $p \times r$ orthonormal matrix through $\beta_o = \beta(\beta'\beta)^{-1/2}$ and $\alpha_o = \alpha(\beta'\beta)^{1/2}$ without affecting Π, it follows that β_o is an element of the Stiefel manifold. The Grassman manifold is an analytic manifold of dimension $(p-r)r$ of all possible r-dimensional subspaces of $\mathbb{R}^p$ and it defines the support for the cointegration space. A uniform prior on the cointegration space is therefore given by a uniform distribution on the Grassman manifold.

The linear normalization in (4) may be viewed as convenient, but as mentioned above, it also has a number of drawbacks. The papers by Strachan and Inder (2004) and Strachan and van Dijk (2003, 2004) analyse an alternative strategy that avoids linear identifying restrictions on β. Their idea makes use of the relationship between the Stiefel and the Grassman manifold. As originally shown by James (1954), a uniform distribution on the Stiefel manifold induces a uniform distribution of the Grassman manifold. It is therefore possible to consider orthonormal matrices β in the Stiefel manifold and adjust all integrals by dividing by the volume of the r-dimensional orthonormal group to account for the fact that the Stiefel manifold is a larger space than the Grassman manifold.

Despite the drawbacks with a linear normalization of β, I will use it in this paper both for the analysis of the cointegration rank and for the analysis of the money demand relation. The main reason for using such a parameterization relative to the parameterization considered by Strachan and Inder (2004) and Strachan and van Dijk (2003, 2004) is simplicity. Specifically, a linear normalization of β with a uniform distribution on the cointegration space admits the use of Gibbs sampling from the marginal and full conditional posteriors for α and Ψ (and thus for β_c). This is not the case for β_o, where the marginal density for β conditional on $(\mathfrak{D}, r)$ is not of a standard form; see, e.g., Strachan and Inder (2004) for details.

Regarding inference on the cointegration rank, Bartlett's (or Lindley's) paradox should be kept in mind.[1] Hence, the prior distribution of (α, β) should be proper in order for posterior probabilities on sharp nulls to be well defined.

Before I present the prior, the following definitions of the matrix t and inverted Wishart distributions are useful. Let $\Gamma_b(a) = \prod_{i=1}^{b} \Gamma([a-i+1]/2)$ for positive integers a and b, with $a \geq b$, and $\Gamma(\cdot)$ is the Gamma function. If $b = 0$, then $\Gamma_0(a) = 1$ for $a \geq 0$.

Definition 1: *An $m \times s$ random matrix B has a matrix t distribution with parameters $\mu \in \mathbb{R}^{m\times s}$, P $(m \times m)$ and Q $(s \times s)$ being positive definite, and $n \geq 0$, denoted by $B \sim t_{m\times s}(\mu, P, Q, n)$, if its density function is*

[1] Since Bartlett (1957) it has generally been accepted that improper priors on all of the parameters result in ill-defined Bayes factors and posterior probabilities that prefer a smaller to a larger model regardless of the information in the data; see also Lindley (1957). In the recent article by Strachan and van Dijk (2005) it is shown that the class of priors that may be used to obtain posterior probabilities includes some improper priors. See also the survey by Bernardo and Ramón (1998).

given by:

$$p(B) = \frac{\Gamma_s(n+m+s)|P|^{s/2}}{\Gamma_s(n+s)\pi^{ms/2}|Q|^{m/2}}\left|I_s + Q^{-1}\big(B-\mu\big)'P\big(B-\mu\big)\right|^{-(n+m+s)/2}.$$

Definition 2: *An $m \times m$ random positive definite matrix Σ has an inverted Wishart distribution with parameters S, being positive definite, and $n \geq m$, denoted by $\Sigma \sim IW_m(S,n)$ if its density function is given by:*

$$p(\Sigma) = \frac{|S|^{n/2}}{2^{nm/2}\pi^{m(m-1)/4}\Gamma_m(n)}|\Sigma|^{-(n+m+1)/2}\exp\left(-\frac{1}{2}\operatorname{tr}\left[\Sigma^{-1}S\right]\right).$$

Properties of the matrix t and inverted Wishart distributions can be found in, e.g., Bauwens et al. (1999, pp. 305–309), Box and Tiao (1973, Chapter 8), and Zellner (1971, pp. 395–399). For example, when B has a matrix t distribution the mode is equal to μ, and $E[B] = \mu$ if $n \geq 1$, while $E[(\mathrm{vec}(B-\mu))(\mathrm{vec}(B-\mu))'] = (1/(n-1))[Q \otimes P^{-1}]$ if $n \geq 2$. Furthermore, if $B \sim t_{m\times s}(\mu,P,Q,n)$ then $B' \sim t_{s\times m}(\mu',P^*,Q^*,n)$, where $P^* = Q^{-1}$ and $Q^* = P^{-1}$.[2]

The distribution of individual elements of B can also be determined directly from the parameters of the matrix t distribution. Let B_{ij} and μ_{ij} denote the element in row i and column j of B and μ, respectively. Furthermore, let q_j denote the j:th diagonal element of Q, while p_i is the i:th diagonal element of P^{-1}. It can now be shown that if B is distributed as matrix t, then $B_{ij} \sim t_{1\times 1}(\mu_{ij}, 1/p_i, q_j, n)$, i.e., a univariate Student t distribution.

Similarly, if Σ has an inverted Wishart distribution, then $E[\Sigma] = (1/(n-m-1))S$ if $n \geq m+2$, while the mode is given by $(1/(n+m+1))S$. Letting Σ_{ii} and S_{ii} denote the i:th diagonal element of Σ and S, respectively, it can be shown that when Σ has an inverted Wishart distribution, then $\Sigma_{ii} \sim IW_1(S_{ii},n)$, i.e., an inverted Gamma-2 distribution with n degrees of freedom.

3.2. *The Prior*

The prior distribution I shall use for the cointegration rank analysis is based on the reference prior developed by Villani (2005b). The joint distribution is decomposed as:

$$p(\alpha,\beta,\Phi,\Gamma,\Omega,r) = p(\alpha,\beta,\Phi,\Gamma,\Omega|r)p(r), \tag{5}$$

where a uniform prior for the cointegration rank may be used, i.e., $p(r) = 1/(p+1)$ for all $r \in \{0,1,\ldots,p\}$. The prior density of $(\alpha,\beta,\Phi,\Gamma,\Omega)$ conditional on the cointegration rank is given by

$$p(\alpha,\beta,\Phi,\Gamma,\Omega|r) = c_r|\Omega|^{-(p+q+r+1)/2}\exp\left(-\frac{1}{2}\operatorname{tr}\left[\Omega^{-1}\big(A + (1/\lambda_\alpha^2)\alpha\beta'\beta\alpha'\big)\right]\right)p\big(\Gamma|\Omega\big), \tag{6}$$

where $\lambda_\alpha > 0$, $q \geq p$, and A, a $p \times p$ positive definite matrix, are three hyperparameters to be specified by the investigator. The normalizing constant c_r is given by

$$c_r = \nu_r|A|^{q/2}\frac{\Gamma_r(p)}{\Gamma_p(q)\Gamma_r(r)}\frac{2^{-qp/2}\pi^{-p(p-1)/4}}{\big(2\pi\lambda_\alpha^2\big)^{pr/2}\pi^{(p-r)r/2}},$$

where ν_r depends on the chosen normalization of β when $r \in \{1,\ldots,p-1\}$, while $\nu_0 = \nu_p = 1$. Specifically, for $\beta = \beta_c$ I let $\nu_r = |c_\perp'c_\perp|^{r/2}|c'c|^{r/2}$.

---

[2] This follows from a well known result for the determinant; see, e.g., Magnus and Neudecker (1988), Chap. 1, Sect. 12, Eq. 13. That is, if A and B are $m \times s$ matrices, then $|I_m + AB'| = |I_s + B'A|$. Furthermore, from properties of the Gamma function we know that $\Gamma_s(n+m+s)/\Gamma_s(n+s) = \Gamma_m(n+m+s)/\Gamma_m(n+m)$ for all $m, s \geq 1$ and $n \geq 0$; see, e.g., Phillips (1989).

If $d > 0$ then the density in (6) is improper since it is constant regardless of the value taken on by Φ. Accordingly, it is not suitable for determining posterior probabilities of models with different Φ matrices. If we assume that $p(\Gamma|\Omega) = 1$, then the same is true for the Γ matrix, thereby making the model unsuitable for, e.g., lag order determination. Under this assumption about $\Gamma|\Omega$, Villani (2005b) shows that the density in (6) implies that the marginal distribution of Ω is inverted Wishart, i.e.,

$$\Omega \sim IW_p\big(A, q\big),$$

when $c = [I_r\ 0]'$ and $\beta = \beta_c$. For these linear identifying restriction on β, Villani also shows that:

$$\mathrm{vec}(\alpha)|\beta, \Omega, r \sim N_{pr}\big(0, [(\beta'\beta)^{-1} \otimes \lambda_\alpha^2 \Omega]\big),$$

while the marginal distribution of Ψ for fixed $r \in \{1, \ldots, p-1\}$ is matrix t, i.e.,

$$\Psi \sim t_{(p-r)\times r}\big(0, I_{p-r}, I_r, 0\big).$$

It is straightforward to show that the results for Ω and $\alpha|\beta, \Omega, r$ hold also for general choices of c, while the third result is slightly modified as:

$$\Psi \sim t_{(p-r)\times r}\big(0, c_\perp' c_\perp, c'c, 0\big).$$

In fact, by multiplying the inverted Wishart density of Ω by the matrix t of Ψ and the multivariate normal of $\mathrm{vec}(\alpha)|\beta, \Omega$ we obtain the expression on the right hand side of (6). Conditional on (Φ, Γ), the density of (α, Ψ, Ω) is therefore proper and this is all we need in order to compute meaningful posterior probabilities of r given that we are willing to condition on a certain k.

Regarding the cointegration space, $\mathrm{sp}(\beta)$, Lemma 3.4 in Villani (2005b) states that if $\tilde{\beta} = [I_r\ \tilde{\Psi}']'$ and $\tilde{\Psi} \sim t_{(p-r)\times r}(0, I_{p-r}, I_r, 0)$, then $\mathrm{sp}(\tilde{\beta})$ is uniformly distributed over the $(p-r)r$ dimensional Grassman manifold. The first assumption in this Lemma need not worry us as it only concerns an ordering of variables. The second assumption is fulfilled by β_c if c is orthonormal. The orthonormality requirement is always fulfilled when c contains r unique columns of I_p. More generally, since we can always postmultiply β_c by a known $r \times r$ matrix, $(c'c)^{-1/2}$, without changing $\mathrm{sp}(\beta_c)$ we can redefine Ψ such that the second assumption is satisfied. Hence, we may regard $\mathrm{sp}(\beta_c)$ to be marginally uniformly distributed over the Grassman manifold also when $\Psi \sim t_{(p-r)\times r}(0, c_\perp' c_\perp, c'c, 0)$.

The hyperparameters A and q determine the prior distribution of Ω. If we are *not* concerned with cointegration rank determination, we may let both these parameters be equal to 0. This corresponds to applying the usual improper prior $p(\Omega) \propto |\Omega|^{-(p+1)/2}$.

The values of A and q also have implications for the prior of α. Letting $\beta = \beta_o$ and $\alpha = \alpha_o$ the distribution of $\alpha_o|\beta_o, \Omega$ is matricvariate normal and independent of β_o. By multiplying its density by the density of Ω and integrating with respect to Ω, it follows that

$$\alpha_o \sim t_{p\times r}\big(0, A^{-1}, \lambda_\alpha^2 I_r, q-p\big);$$

see, e.g., Bauwens et al. (1999, Theorem A.19). Hence, $E[\alpha_o] = 0$ if $q \geq p+1$, while the covariance matrix is $E[\mathrm{vec}(\alpha_o)\mathrm{vec}(\alpha_o)'] = (1/(q-p-1))[I_r \otimes \lambda_\alpha^2 A]$ if $q \geq p+2$. Since $E(\Omega) = (1/(q-p-1))A$ if $q \geq p+2$, the covariance matrix for α_o can also be expressed as $[I_r \otimes \lambda_\alpha^2 E[\Omega]]$.

The marginal uncertainty of Ω is, by definition, increasing in q, while the marginal uncertainty of α_o is decreasing in this hyperparameter for given A and λ_α. Furthermore, the columns of α_o have the same prior covariance matrix, $\lambda_\alpha^2 E[\Omega]$, where a larger absolute value for the shrinkage hyperparameter λ_α implies larger uncertainty for α_o. For a uniform distribution on the cointegration space, it is natural that the columns of α_o have the same prior distribution as the ordering of the columns of β_o does not matter.

The discussion thus far concerns the case when $p(\Gamma|\Omega) = 1$, resulting essentially in the reference prior suggested by Villani (2005b). When $p(\Gamma|\Omega)$ is proper, a structured shrinkage prior of these parameters will be employed. Specifically, I assume that $\Gamma|\Omega$ is matricvariate normal:

$$p\big(\Gamma|\Omega\big) = (2\pi)^{-p^2(k-1)/2}|\Sigma_\Gamma|^{-p/2}|\Omega|^{-p(k-1)/2}\exp\left(-\frac{1}{2}\operatorname{tr}\left[\Omega^{-1}\Gamma\Sigma_\Gamma^{-1}\Gamma'\right]\right). \tag{7}$$

The $p(k-1)\times p(k-1)$ matrix Σ_Γ is positive definite and in the empirical analyses I restrict it to be block diagonal with $(k-1)$ blocks consisting of the $p\times p$ matrices

$$\Sigma_{\Gamma_i} = \frac{\lambda_b^2}{i^{2\lambda_l}}I_p, \quad i = 1,\ldots,k-1. \tag{8}$$

The hyperparameter $\lambda_b > 0$ measures the baseline shrinkage (overall tightness around zero), while $\lambda_l > 0$ quantifies the lag order shrinkage.

The prior distribution of $\Gamma|\Omega$ has zero mean and covariance matrix $[\Sigma_\Gamma\otimes\Omega]$ and is thus in the same spirit as the prior of $\alpha|\beta,\Omega,r$. Since $p(\Gamma|\alpha,\beta,\Phi,\Omega,r) = p(\Gamma|\Omega)$ we have, by assumption, that Γ is independent of α,β,Φ and the cointegration rank r when we condition on Ω. This assumption not only facilitates the derivation of posterior distributions, but also means that the properties of the prior for (α,β,Ω) conditional on Φ, Γ and r are not affected by the choice of prior of Γ.

With the marginal distribution of Ω being $IW_p(A,q)$, it follows by the same arguments as above for α_o that the marginal distribution of Γ is matrix t, i.e.,

$$\Gamma \sim t_{p\times p(k-1)}\big(0, A^{-1}, \Sigma_\Gamma, q-p\big).$$

Hence, for $q \geq p+1$ we find that $E[\Gamma] = 0$, while for $q \geq p+2$ it holds that $E[\text{vec}(\Gamma)\text{vec}(\Gamma)'] = (1/(q-p-1))[\Sigma_\Gamma\otimes A] = [\Sigma_\Gamma\otimes E(\Omega)]$.

The matrix Ω is included in (7), as well as in $p(\alpha|\beta,\Omega,r)$, to handle differences in variability of the endogenous variables. Hence, this prior shares some features with the so called Minnesota prior (see, e.g., Litterman, 1986), but instead of assuming that the variability parameters are known the dependence on the covariance matrix of the residuals is explicitly taken into account through conditioning and with a marginal prior of Ω; see, e.g., Kadiyala and Karlsson (1997), Sims and Zha (1998), and Robertson and Tallman (1999) for further discussions.

Notice that while Σ_Γ is assumed to be block diagonal such that $\Gamma_i|\Omega$ and $\Gamma_j|\Omega$ are independent for $i\neq j$, the properties for $\Gamma|\Omega$ discussed above do not rely on this assumption. Rather, the independence assumption is made in order to have a prior that can be viewed as consistent with the first two properties listed by Villani (2005b) and reiterated in Section 1 above. In particular, the number of hyperparameters is only increased by two and both parameters are straightforward to interpret. Furthermore, the theoretical results presented in the next section do not depend on how Σ_Γ is parameterized. Hence, alternative parameterizations that allow for dependence among the elements of $\Gamma_i|\Omega$ as well as dependence between $\Gamma_i|\Omega$ and $\Gamma_j|\Omega$ may therefore be used instead of (8). Such parameterizations will typically involve a greater number of hyperparameters and are therefore less likely to be transparent.

I shall also consider models with over-identifying linear restrictions on β. For that purpose let

$$\mathrm{vec}(\beta) = h + H\psi, \tag{9}$$

where h $(pr \times 1)$ and H $(pr \times s)$ are known. The prior density of ψ is here assumed to be proportional to $|\beta'\beta|^{-p/2}$, where β satisfies (9). In the event that β is exactly identified, $s = r(p-r)$, $h = \mathrm{vec}(c)$, while $H = [I_r \otimes c_{\perp}]$ and $\psi = \mathrm{vec}(\Psi)$.

The prior on ψ is similar to the prior used by Villani and Warne (2003). Moreover, it is proportional to the prior on Ψ when $s = r(p-r)$, i.e., a matrix t distribution with zero degrees of freedom. For the over-identification case, I let $\mathrm{vec}(\alpha)|\beta, \Omega, r \sim N_{pr}(0, [(\beta'\beta)^{-1} \otimes \Omega])$, $\Omega \sim IW_p(A, q)$, while the two priors on $\Gamma|\Omega$ discussed above as well as the non-informative and improper prior on Φ will be applied. For the over-identifying cases, the assumption that $p(\psi|r) \propto |\beta'\beta|^{-p/2}$ means that the prior is not suitable if we wish to discriminate between different ψ models. Such models will have different integration constants and these constants are ignored when this prior is applied. Rather, the assumption is made out of convenience since the impact that β has on the conditional prior density of α is fully offset by the prior on ψ; see Villani (2005a) for a prior that allows for the posterior evaluation of different ψ models.

4. Posterior Distributions

4.1. *Gibbs Samplers*

The joint posterior distribution of the parameters of the cointegrated VAR is intractable. In this section, the necessary results for a numerical evaluation of the posterior distribution via the Gibbs sampler are provided. This sampler simulates from the joint posterior distribution of the model parameters by iteratively generating draws from the full conditional posterior distributions; see, e.g., Casella and George (1992), Tierney (1994), and Geweke (1999).

The full conditional posterior distribution is the posterior distribution for a group of parameters conditional on the data and the remaining parameters. Let θ denote the full set of model parameters, while θ_j is the j:th group of parameters with $\theta = (\theta_1, \ldots, \theta_g)$, where g is the number of groups. The full conditional posterior for θ_j is denoted by $p(\theta_j|\theta_1, \ldots, \theta_{j-1}, \theta_{j+1}, \ldots, \theta_g, \mathfrak{D}, M_m)$, with M_m representing model m. The Gibbs sampler begins with $\theta^{(0)}$, some value in the support of the joint posterior distribution, while $\theta^{(i)}$ is the i:th draw from the Gibbs sampler. The i:th draw of θ is obtained by collecting the draws from the full conditional posteriors for θ_j, $j = 1, \ldots, g$, where

$$\theta_j^{(i)} \sim p(\theta_j|\theta_1^{(i)}, \ldots, \theta_{j-1}^{(i)}, \theta_{j+1}^{(i-1)}, \ldots, \theta_g^{(i-1)}, \mathfrak{D}, M_m), \quad j = 1, \ldots, g, \;\; i = 1, \ldots, G,$$

so that $\theta^{(i)} = (\theta_1^{(i)}, \ldots, \theta_g^{(i)})$. If $\theta^{(i-1)}$ is a draw from the joint posterior it follows that

$$(\theta_1^{(i)}, \ldots, \theta_{j-1}^{(i)}, \theta_j^{(i)}, \theta_{j+1}^{(i-1)}, \ldots, \theta_g^{(i-1)}) \sim p\big(\theta|\mathfrak{D}, M_m\big),$$

at each step j by definition of the conditional density. Hence, $\theta^{(i)}$ is also a draw from the joint posterior distribution.

The outcome of the Gibbs sampler is a sequence of draws $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(G)}$ from the joint posterior distribution of θ. Due to the Markov chain property of the Gibbs sampler, the draws are not independent, but under certain conditions (Tierney, 1994), which are satisfied here, the following can be shown to hold:

$$f\left(\theta^{(i)}\right) \xrightarrow{\mathrm{d}} p\left[f\left(\theta\right)|\mathfrak{D}, M_m\right],$$

and

$$\frac{1}{G}\sum_{i=1}^{G} f\left(\theta^{(i)}\right) \xrightarrow{\mathrm{a.s.}} E\left[f\left(\theta\right)|\mathfrak{D}, M_m\right],$$

where $\xrightarrow{\mathrm{d}}$ denotes convergence in distribution, $\xrightarrow{\mathrm{a.s.}}$ convergence almost surely, and $f(\cdot)$ is any well-behaved real valued function. Hence, given the draws from the posterior of θ, the posterior distribution of any function of the model parameters, such as impulse responses, are immediately available by applying the appropriate f-function to each posterior draw and using some density estimator.

With $\theta = (\alpha, \psi, \Phi, \Gamma, \Omega)$, the following Proposition gives the full conditional posteriors of the 5 groups of parameters in θ given r.

Proposition 1: *The full conditional posterior distributions are:*

$$\Omega|\alpha, \beta, \Phi, \Gamma, \mathfrak{D}, r \sim IW_p(\varepsilon\varepsilon' + A + (1/\lambda_\alpha^2)\alpha\beta'\beta\alpha' + \Gamma\Sigma_\Gamma^{-1}\Gamma', T + q + r + m), \tag{10}$$

where $\varepsilon = Z_0 - \Phi D - \Gamma Z_2 - \alpha\beta' Z_1$*. If* $p(\Gamma|\Omega) = 1$*, then* $m = 0$ *and* $\Sigma_\Gamma^{-1} = 0$*. If* $p(\Gamma|\Omega)$ *is given by equation* (7)*, then* $m = p(k-1)$*.*

$$\mathrm{vec}(\Phi)|\alpha, \beta, \Gamma, \Omega, \mathfrak{D}, r \sim N_{pd}(\bar{\mu}_\Phi, \bar{\Sigma}_\Phi), \tag{11}$$

where $\bar{\mu}_\Phi = \mathrm{vec}([Z_0 - \Gamma Z_2 - \alpha\beta' Z_1]D'(DD')^{-1})$*, and* $\bar{\Sigma}_\Phi = [(DD')^{-1} \otimes \Omega]$*.*

$$\mathrm{vec}(\Gamma)|\alpha, \beta, \Phi, \Omega, \mathfrak{D}, r \sim N_{p^2(k-1)}(\bar{\mu}_\Gamma, \bar{\Sigma}_\Gamma), \tag{12}$$

where $\bar{\mu}_\Gamma = \mathrm{vec}([Z_0 - \Phi D - \alpha\beta' Z_1]Z_2'[Z_2 Z_2' + \Sigma_\Gamma^{-1}]^{-1})$*, and* $\bar{\Sigma}_\Gamma = [(Z_2 Z_2' + \Sigma_\Gamma^{-1})^{-1} \otimes \Omega]$*.*

$$\mathrm{vec}(\alpha)|\beta, \Phi, \Gamma, \Omega, \mathfrak{D}, r \sim N_{pr}(\bar{\mu}_\alpha, \bar{\Sigma}_\alpha), \tag{13}$$

where $\bar{\mu}_\alpha = \mathrm{vec}([Z_0 - \Phi D - \Gamma Z_2]Z_1'\beta[\beta'(Z_1 Z_1' + (1/\lambda_\alpha^2)I_p)\beta]^{-1})$*, and* $\bar{\Sigma}_\alpha = [[\beta'(Z_1 Z_1' + (1/\lambda_\alpha^2)I_p)\beta]^{-1} \otimes \Omega]$*.*

$$\psi|\alpha, \Phi, \Gamma, \Omega, \mathfrak{D}, r \sim N_s(\bar{\mu}_\psi, \bar{\Sigma}_\psi), \tag{14}$$

where $\bar{\mu}_\psi = \bar{\Sigma}_\psi H'\bar{\Sigma}_\beta^{-1}(\bar{\mu}_\beta - h)$*, and* $\bar{\Sigma}_\psi = (H'\bar{\Sigma}_\beta^{-1}H)^{-1}$*. Furthermore,* $\bar{\mu}_\beta = \bar{\Sigma}_\beta \mathrm{vec}(Z_1[Z_0 - \Phi D - \Gamma Z_2]'\Omega^{-1}\alpha)$ *whereas* $\bar{\Sigma}_\beta^{-1} = [\alpha'\Omega^{-1}\alpha \otimes (Z_1 Z_1' + (1/\lambda_\alpha^2)I_p)]$*.*

Proof: The full conditional posteriors under $p(\Gamma|\Omega) = 1$ are proved by Villani (2005b, Theorem 4.6) for $\beta = [I_r\ \Psi']'$. The extensions to β satisfying (9) and to $p(\Gamma|\Omega)$ satisfying (7) are straightforward. □

When analysing the cointegration rank, the parameters of interest are α and β. Since the dimension of Φ, Γ, Ω can be substantial in practise and thus greatly affect the computation time of the Gibbs sampler, it is useful to have the marginal posteriors of α conditional on β and the rank, and of Ψ conditional on α and the rank.

The results for the non-informative prior $p(\Gamma|\Omega) = 1$ were originally given by Villani (2005b, Theorem 4.6) for $c = [I_r\ 0]'$, but are below extended to general choices of c. Moreover, the next Proposition also provides the marginal conditional posteriors under the informative prior in (7):

Proposition 2: *Suppose* $\beta = \beta_c$*, and* $\alpha = \alpha_c$*, then*

$$\alpha|\beta, \mathfrak{D}, r \sim t_{p\times r}\left[\hat{\alpha}, \left(A + Z_0 N[Z_0 - \hat{\alpha}\beta' Z_1]'\right)^{-1}, \left(\beta' C_1 \beta\right)^{-1}, T + q - (pn + d)\right], \tag{15}$$

where

$$\hat{\alpha} = Z_0 N Z_1' \beta \left(\beta' C_1 \beta\right)^{-1}, \quad C_1 = Z_1 N Z_1' + (1/\lambda_\alpha^2) I_p.$$

Furthermore,

$$\Psi | \alpha, \mathfrak{D}, r \sim t_{(p-r)\times r}\left[\hat{\Psi}, \left(G_3 - G_2' G_1^{-1} G_2\right)^{-1}, C_3, T+q+r-(pn+d)\right], \tag{16}$$

where

$$\begin{aligned}
&\hat{\Psi} = \hat{\beta}_2 + G_2' G_1^{-1}\left(I_r - \hat{\beta}_1\right), \quad && C_3 = \left(I_r - \hat{\beta}_1\right)' G_1^{-1}\left(I_r - \hat{\beta}_1\right) + \left(\alpha' S^{-1}\alpha\right)^{-1}, \\
&\hat{\beta} = K\hat{\Pi}' S^{-1}\alpha\left(\alpha' S^{-1}\alpha\right)^{-1}, && \hat{\Pi} = Z_0 N Z_1' C_1^{-1}, \\
&K = \left[\bar{c}\ \bar{c}_\perp\right]', && S = A + Z_0 N Z_0' - Z_0 N Z_1' C_1^{-1} Z_1 N Z_0',
\end{aligned}$$

and

$$K C_1^{-1} K' + K \hat{\Pi}' S^{-1} \hat{\Pi} K' - \hat{\beta}\alpha' S^{-1} \alpha \hat{\beta}' = \begin{bmatrix} \underset{r\times r}{G_1} & \underset{r\times(p-r)}{G_2} \\ \underset{(p-r)\times r}{G_2'} & \underset{(p-r)\times(p-r)}{G_3} \end{bmatrix},$$

while $\hat{\beta}_1$ *is* $r\times r$ *and* $\hat{\beta}_2$ *is* $(p-r)\times r$ *such that* $\hat{\beta} = [\hat{\beta}_1'\ \hat{\beta}_2']'$. *The* $T\times T$ *integration matrix* N *(with respect to* Φ *and* Γ*) and the integer* n *depends on the prior distribution of* Γ.

(i) *If* $p(\Gamma|\Omega) = 1$, *then*

$$N = M_z, \quad M_z = I_T - Z'\left(ZZ'\right)^{-1} Z, \quad Z = \left[D'\ \ Z_2'\right]', \quad n = k.$$

(ii) *If* $p(\Gamma|\Omega)$ *is given by equation* (7), *then*

$$N = M_D - M_D Z_2' \left(Z_2 M_D Z_2' + \Sigma_\Gamma^{-1}\right)^{-1} Z_2 M_D, \quad M_D = I_T - D'\left(DD'\right)^{-1} D, \quad n = 1.$$

PROOF: See Appendix A. □

It may also be of interest to consider sampling directly from the marginal posterior distribution of β. Villani (2005b, Theorem 4.3) provides the kernel of its density for the case when $p(\Gamma|\Omega) = 1$. This kernel is a ratio of two matrix t kernels and the resulting kernel of the marginal posterior of β is therefore a 1-1 poly-matrix-t density; see Bauwens and van Dijk (1990) for further details on this distribution. In contrast with multivariate poly-t densities (see, e.g., Drèze, 1977, and Bauwens et al., 1999) poly-matrix-t densities have remained largely unexplored. However, as shown by Bauwens and Lubrano (1996, Corollary 3.2) and Villani (2005b, Theorem 4.4), the marginal posterior of Ψ is integrable but does not have any finite integer moments. The nonexistence of such moments is not due to the choice of prior but rather to the linear normalization of β in (4); see also Phillips (1994) for similar properties of the maximum likelihood estimator.[3] These results can directly be extended to the case when the informative $\Gamma|\Omega$ prior in (7) is used, but this distribution is of minor interest for the current study since it is not needed for the computation of the cointegration rank posterior. Moreover, it is based on exactly identifying restrictions on β and cannot be extended to the over-identifying cases.

[3] Bauwens and Giot (1998) implement a version of the Gibbs sampler called the Griddy-Gibbs sampler for poly-matrix-t distributions. It is based on computing the density for each cointegration vector conditional on all the others. This density is a vector 1-1 poly-t density and can be sampled from using the algorithm in Bauwens and Richard (1985).

4.2. *Posterior Cointegration Rank Probabilities*

The posterior probability of the cointegration rank can be determined from Bayes rule

$$p\big(r|\mathfrak{D}\big) = \frac{p\big(\mathfrak{D}|r\big)p(r)}{\sum_{i=0}^{p} p\big(\mathfrak{D}|i\big)p(i)}. \tag{17}$$

Hence, given that the marginal likelihoods, $p(\mathfrak{D}|r)$, are available, the posterior probability of the cointegration rank follows directly from (17).[4]

The marginal likelihoods for $r = 0$ and $r = p$ when $p(\Gamma|\Omega) = 1$ are given in Villani (2005b, Theorem 5.1). For convenience his results are presented in the next Proposition along with the marginal likelihoods for these cointegration rank values when the prior density of $\Gamma|\Omega$ is given by equation (7):

PROPOSITION 3: *The marginal likelihoods for $r = 0$ and $r = p$ are given by:*

$$p\big(\mathfrak{D}|r = 0\big) = k_1\Gamma_p\big(T + q - p(n-1) - d\big)\Big|A + Z_0NZ_0'\Big|^{-(T+q-p(n-1)-d)/2}, \tag{18}$$

while

$$p\big(\mathfrak{D}|r = p\big) = k_1\Gamma_p\big(T + q - p(n-1) - d\big)\lambda_\alpha^{-p^2}|C_1|^{-p/2}|S|^{-(T+q-p(n-1)-d)/2}, \tag{19}$$

where C_1, S, N, and n are defined in Proposition 2. The multiplicative constant is:

$$k_1 = \frac{|A|^{q/2}}{\pi^{(T-p(n-1)-d)p/2}z^{p/2}\Gamma_p(q)},$$

where $z = |ZZ'|$ if $p(\Gamma|\Omega) = 1$, while $z = |\Sigma_\Gamma||DD'||Z_2M_DZ_2' + \Sigma_\Gamma^{-1}|$ if $p(\Gamma|\Omega)$ is given by equation (7).

PROOF: See Appendix B. □

The marginal likelihood for cointegration rank $r \in \{1,\ldots,p-1\}$ can also be computed using the reference prior in (6) for $p(\Gamma|\Omega) = 1$, and for $p(\Gamma|\Omega)$ given by equation (7). Villani (2005b) suggests three methods for determining $p(\mathfrak{D}|r)$: Monte Carlo integration, importance sampling, and from what Chib (1995) has termed the *basic marginal likelihood identity*; see also Geweke (1999). Among these approaches Villani advocates the marginal likelihood identity since it strongly outperforms the others in his Monte Carlo study.

The marginal likelihood identity can in the current context be represented by:

$$p\big(\mathfrak{D}|r\big) = \frac{p\big(\mathfrak{D}|\alpha,\Psi,r\big)p\big(\alpha,\Psi|r\big)}{p\big(\Psi|\alpha,\mathfrak{D},r\big)p\big(\alpha|\mathfrak{D},r\big)}. \tag{20}$$

Chib (1995) suggests using this identity in combination with a Gibbs sampler to calculate the marginal likelihood. Although the identity holds for any α and Ψ, Chib explains that the point $(\tilde{\alpha},\tilde{\Psi})$ should preferably have high posterior density, such as the mode or the median.

The numerator in equation (20) can, as shown in Villani (2005b, Lemma 5.3), be determined analytically. In our two cases,

$$\begin{aligned} p\big(\mathfrak{D}|\alpha,\Psi,r\big)p\big(\alpha,\Psi|r\big) = k_1v_r&\frac{\Gamma_p\big(T + q + r - p(n-1) - d\big)\Gamma_r(p)}{\Gamma_r(r)\pi^{(2p-r)r/2}\lambda_\alpha^{pr}} \\ &\times\Big|A + (1/\lambda_\alpha^2)\alpha\beta'\beta\alpha' + WNW'\Big|^{-(T+q+r-p(n-1)-d)/2}, \end{aligned} \tag{21}$$

[4] The fact that the analysis is conditioned on the initial conditions of the VAR model is, for notational convenience, not explicitly written out.

where $W = Z_0 - \alpha\beta' Z_1$, while N and n are defined in Proposition 2. Since the multiplicative constant k_1 appears in all marginal likelihoods it can be dispensed with.

For the denominator we note that $p(\Psi|\alpha, \mathfrak{D}, r)$ is given in Proposition 2, equation (16), and, hence, only the term $p(\alpha|\mathfrak{D}, r)$ in (20) is not available in closed form. Its value at a point $\alpha = \tilde{\alpha}$ can, however, be computed from a posterior sample $\Psi^{(1)}, \ldots, \Psi^{(G)}$ of Ψ by

$$\hat{p}\big(\tilde{\alpha}|\mathfrak{D}, r\big) = \frac{1}{G}\sum_{i=1}^{G} p\big(\tilde{\alpha}|\Psi^{(i)}, \mathfrak{D}, r\big),$$

where $p(\alpha|\Psi, \mathfrak{D}, r)$ is given in Proposition 2, equation (15). From the ergodic theorem (Tierney, 1994), $\hat{p}(\tilde{\alpha}|\mathfrak{D}, r) \overset{\text{a.s.}}{\to} p(\tilde{\alpha}|\mathfrak{D}, r)$ as $G \to \infty$. A posterior sample $\Psi^{(1)}, \ldots, \Psi^{(G)}$ can be drawn from either the full conditional distribution of ψ in Proposition 1 using the proper h and H matrices for β_c, or from the marginal conditional distribution of Ψ in Proposition 2. The precision of $\hat{p}(\tilde{\alpha}|\mathfrak{D}, r)$ can be evaluated through its numerical standard error; see Section 4.6 and Chib (1995) for details.

4.3. *Posterior Mode Estimation*

The determination the marginal likelihood for cointegration rank $r \in \{1, \ldots, p-1\}$ requires a point estimate of (α, Ψ). Since the selected point should preferably be one that has high posterior density a natural candidate is the posterior mode. Within the current framework it turns out to be straightforward to find an analytic expression of the posterior mode estimator of these parameters as well as of all the other model parameters conditional on r.

Let us first examine the case when β is exactly identified so that $\beta = \beta_c$. Define the matrices S_{ij} as follows:

$$\begin{aligned} S_{00} &= (T+p+q+r+m+1)^{-1}\big(Z_0 N Z_0' + A\big), \\ S_{01} &= (T+p+q+r+m+1)^{-1} Z_0 N Z_1', \\ S_{10} &= S_{01}', \\ S_{11} &= (T+p+q+r+m+1)^{-1} C_1, \end{aligned} \tag{22}$$

where N and C_1 are given in Proposition 2, while $m = 0$ under $p(\Gamma|\Omega) = 1$ and $m = p(k-1)$ under the informative $\Gamma|\Omega$ prior. The next Proposition gives the procedure for computing the posterior mode.

Proposition 4: *Suppose $\beta = \beta_c$, and $\alpha = \alpha_c$. The posterior mode estimator of the parameters of the joint posterior $p(\alpha, \Psi, \Phi, \Gamma, \Omega|\mathfrak{D}, r)$ and of the posterior $p(\alpha, \Psi|\mathfrak{D}, r)$ is found by the following procedure. Solve the equation*

$$\left|\lambda S_{11} - S_{10} S_{00}^{-1} S_{01}\right| = 0, \tag{23}$$

for the eigenvalues $1 > \hat{\lambda}_1 > \ldots > \hat{\lambda}_p > 0$ and eigenvectors $\hat{V} = [\hat{v}_1 \cdots \hat{v}_p]$ which are normalized by $\hat{V}' S_{11} \hat{V} = I_p$. The posterior mode cointegration space estimator is:

$$\bar{\beta}_{\text{PM}} = \begin{bmatrix} \hat{v}_1 & \cdots & \hat{v}_r \end{bmatrix}, \tag{24}$$

and the maximized posterior density is proportional to $|S_{00}|\prod_{i=1}^{r}(1-\hat{\lambda}_i)$*. The posterior mode estimator of* $(\alpha_c,\Psi,\beta_c,\Omega)$ *is obtained by replacing* β *by* $\bar{\beta}_{\mathrm{PM}}$ *in the following equations:*

$$\alpha_c(\beta) = S_{01}\beta\beta'\bar{c}, \tag{25}$$

$$\Psi(\beta) = \bar{c}_{\perp}'\beta\big(\bar{c}'\beta\big)^{-1}, \tag{26}$$

$$\beta_c(\beta) = c + c_{\perp}\Psi(\beta), \tag{27}$$

$$\Omega(\beta) = S_{00} - S_{01}\beta\big(\beta' S_{11}\beta\big)^{-1}\beta' S_{10}. \tag{28}$$

The posterior mode estimator of (Φ,Γ) *is obtained by replacing* (α_c,β_c) *by their posterior mode estimators in the following equations:*

$$\Phi(\alpha_c,\beta_c) = \big(Z_0-\alpha_c\beta_c' Z_1\big)M_{Z_2}D'\big(DM_{Z_2}D'\big)^{-1}, \tag{29}$$

$$\Gamma(\alpha_c,\beta_c) = \big(Z_0-\alpha_c\beta_c' Z_1\big)M_{D}Z_2'\big(Z_2M_DZ_2'+\Sigma_{\Gamma}^{-1}\big)^{-1}, \tag{30}$$

where $M_{Z_2} = I_T - Z_2'(Z_2Z_2'+\Sigma_\Gamma^{-1})^{-1}Z_2$*, and* $\Sigma_\Gamma^{-1}=0$ *if* $p(\Gamma|\Omega)=1$*.*

Proof: See Appendix C. □

A direct consequence of Proposition 4 is that we can compute the posterior mode of the parameters without having made any draws from the relevant posterior distribution for all $r\in\{0,1,\ldots,p\}$. Moreover, the approach for computing the posterior mode is the same as the maximum likelihood approach of Johansen (1996), but with the changes to the definitions of the product moment matrices S_{ij} given in equation (22). Hence, software that allow the user to compute the maximum likelihood estimator of β based on inputs of the S_{ij} matrices in Johansen can be used to calculate the posterior mode cointegration space estimator of β, as given by equation (24). For a given cointegration rank, the posterior mode estimator is identical to the maximum likelihood estimator for α,β,Φ,Γ when the improper prior $p(\alpha,\beta,\Phi,\Gamma,\Omega|r)\propto|\Omega|^{-(p+1)/2}$ is applied. This corresponds to selecting $A=0$, $\Sigma_\Gamma^{-1}=0$, and $q=\lambda_\alpha^{-1}=0$. In that case the posterior mode estimator of Ω is equal to the maximum likelihood estimator times $T/(T+p+1)$.

The posterior mode estimator of (α,Ψ) in Proposition 4 is a good candidate when evaluating the marginal likelihood for cointegration rank r in (20) since it maximizes both the value of the numerator and the denominator in this expression. Although this point is the posterior mode of $p(\alpha,\Psi|\mathfrak{D},r)$ it does not give us the posterior mode of the marginal posteriors $p(\alpha|\mathfrak{D},r)$ and $p(\Psi|\mathfrak{D},r)$, respectively. For example, the log posterior of β can, apart from a constant, be expressed as

$$\ln p\big(\beta|\mathfrak{D},r\big) = -\frac{T+q-p(n-1)-d}{2}\ln\Big|S_{00}-S_{01}\beta\big(\beta'S_{11}\beta\big)^{-1}\beta'S_{10}\Big| - \frac{p}{2}\ln\Big|\beta'S_{11}\beta\Big|. \tag{31}$$

The first term on the right hand side is proportional to the factor being maximized by $\bar{\beta}_{\mathrm{PM}}$ once the log posteriors in Proposition 4 have been concentrated around β, i.e., the log of the determinant of $\Omega(\beta)$ in equation (28). The second term on the right hand side, evaluated at $\bar{\beta}_{\mathrm{PM}}$, is equal to zero since $\bar{\beta}_{\mathrm{PM}}'S_{11}\bar{\beta}_{\mathrm{PM}}=I_r$. The value of the log marginal posterior of β, evaluated at $\bar{\beta}_{\mathrm{PM}}$, is therefore proportional to, e.g., the value of the log of the posterior $p(\alpha,\Psi|\mathfrak{D},r)$ when it is evaluated at its posterior mode. Furthermore, the second term in (31) vanishes relative to the first as $T\to\infty$ and, consequently, the estimator $\bar{\Psi}=\bar{c}_{\perp}'\bar{\beta}_{\mathrm{PM}}(\bar{c}'\bar{\beta}_{\mathrm{PM}})^{-1}$ is asymptotically the posterior mode of $p(\Psi|\mathfrak{D},r)$.

The analysis in Johansen (1996) can also be used for computing the posterior mode of ψ for the full posterior under the over-identifying restrictions in equation (9). Specifically, if

$$\beta_i = h_i + H_i\psi_i, \quad i = 1, \ldots, r,$$

then the switching algorithm in Johansen (1996, Chap. 7.2.3) can be applied, but with the S_{ij} matrices in (22) and the $\bar{\beta}_{\mathrm{PM}}$ estimator to initialize the algorithm. As mentioned by Johansen, there is no proof that the algorithm converges to the correct value, but it has the property that the posterior density is maximized in each step. If the algorithms converges, we may calculate the posterior mode of α from:

$$\alpha(\beta) = S_{01}\beta\big(\beta' S_{11}\beta\big)^{-1},$$

by replacing β with its the posterior mode estimate under the over-identifying restrictions. Similarly, the posterior mode of Ω, Φ, and Γ can be determined by replacing β by its posterior mode estimate in equations (28)–(30).

4.4. *Lag Order Determination*

The analysis thus far has been conducted under the assumption that the lag order k is known or at least determined before the cointegration rank analysis. Under the assumption that $p(\Gamma|\Omega)$ is given by equation (7), however, the results in Section 4.2 can be applied both for the determination of the lag order conditional on, e.g., a full cointegration rank, and for the joint determination of the rank and the lag order posterior probabilities as well as for their marginal probabilities. In this Section I shall briefly discuss these extensions, while the case $p(\Gamma|\Omega) = 1$ is not considered due to Bartlett's paradox.

Suppose first that we wish to determine a proper choice for k under the assumption that $r = p$. Let k^* be the maximum lag order under consideration and let T^* be the largest possible sample size, subject to $T^* \leq T$, under this choice for k. The data matrices Z_i and D now have T^* columns for all k, while the number of rows of Z_2 depends on k. Specifically, p rows of one additional lag of Δx_t are appended to the bottom of Z_2 when k increases by one unit.

We are now in a position to state the following result:

Corollary 1: *Suppose that $p(\Gamma|\Omega)$ is given by equation (7) and $r = p$. The marginal likelihood for lag order k is then:*

$$p\big(\mathfrak{D}|k\big) = k_1^*\Gamma_p\big(T^* + q - d\big)|\Sigma_\Gamma|^{-p/2}|Z_2 M_D Z_2' + \Sigma_\Gamma^{-1}|^{-p/2}\lambda_\alpha^{-p^2}|C_1|^{-p/2}|S|^{-(T^*+q-d)/2}, \tag{32}$$

for $k = 1, \ldots, k^*$. For $k = 1$, the two determinants involving Σ_Γ are replaced with unity. The multiplicative constant is:

$$k_1^* = \frac{|A|^{q/2}}{\pi^{(T^*-d)p/2}|DD'|^{p/2}\Gamma_p(q)}.$$

Proof: This follows directly from Proposition 3. □

Notice that k_1^*, $\Gamma_p(T^* + q - d)$, and $\lambda_\alpha^{-p^2}$ appear in all marginal likelihoods. These three constants may therefore be dispensed with. Moreover, the marginal likelihood for any k is given as an analytic expression and can therefore be calculated without first having made any draws from the posterior distributions. Hence, the posterior probability of lag order k (conditional on $r = p$) can quickly be determined in practise using (32), Bayes rule, and a prior on k. In particular, the posterior lag order probability is:

$$p(k|\mathfrak{D}) = \frac{p(\mathfrak{D}|k)p(k)}{\sum_{i=1}^{k^*} p(\mathfrak{D}|i)p(i)}, \qquad k = 1, \ldots, k^*, \tag{33}$$

where, e.g., a uniform prior on k may be used, i.e., $p(k) = 1/k^*$ for all $k \in \{1, \ldots, k^*\}$.

It should be emphasized that this approach to lag order determination uses the same prior distribution in both the lag order and cointegration rank steps and may therefore be preferable to relying on, e.g., fractional marginal likelihoods, as in Villani (2001b), given that cointegration rank analysis will follow lag order determination.

While the above approach to selecting k has the advantage of being computationally convenient, it neither allows for a determination of the joint posterior probability of a particular pair (r, k), nor for a determination of the marginal posterior probabilities of r and k. By properly using the results on the computation of the posterior rank probabilities (conditional on k) and the above setup for the posterior lag order probabilities, it is possible to calculate $p(r, k|\mathfrak{D})$ through the marginal likelihood expressions in Section 4.2 for the informative $\Gamma|\Omega$ prior. It should be kept in mind that the multiplicative constant is given by k_1^* and not by k_1 as the determinants involving Σ_Γ depend on the choice of k. Furthermore, a uniform prior on (r, k) implies that $p(r, k) = 1/(k^*(p + 1))$, with $p(r|k) = p(r) = 1/(p + 1)$, and $p(k|r) = p(k) = 1/k^*$. The calculation of marginal posterior probabilities of r and k is straightforward once we have calculated the marginal likelihood for each possible pair (r, k).

4.5. *Over-Identified β when Φ and Γ are Nuisance Parameters*

Let me now consider Gibbs sampling from the posterior distribution of (α, ψ, Ω) when β may involve over-identifying restrictions and Φ and Γ are considered nuisance parameters. Most model parameters are typically located in these two matrices and the Gibbs updating steps for them usually dominate the total computing time. This can also have an impact on the time to convergence for the Gibbs sampler.

The next Proposition therefore gives the conditional posteriors necessary to perform a (marginal) Gibbs sampler to generate draws from $p(\alpha, \psi, \Omega|\mathfrak{D}, r)$.

Proposition 5: *Suppose β satisfies equation* (9) *when $r \in \{1, \ldots, p-1\}$ and $s \leq (p-r)r$, then*

$$\Omega|\alpha, \beta, \mathfrak{D}, r \sim IW_p\big(WNW' + A + (1/\lambda_\alpha^2)\alpha\beta'\beta\alpha', T + q + r - p(n-1) - d\big), \tag{34}$$

where $W = Z_0 - \alpha\beta' Z_1$, while N and n are defined in Proposition 2.

$$\mathrm{vec}(\alpha)|\beta, \Omega, \mathfrak{D}, r \sim N_{pr}\big(\tilde{\mu}_\alpha, \tilde{\Sigma}_\alpha\big), \tag{35}$$

where $\tilde{\mu}_\alpha = \mathrm{vec}(\hat{\alpha})$, $\tilde{\Sigma}_\alpha = [(\beta' C_1 \beta)^{-1} \otimes \Omega]$, while $\hat{\alpha}$ and C_1 are defined in Proposition 2.

$$\psi|\alpha, \Omega, \mathfrak{D}, r \sim N_s(\tilde{\mu}_\psi, \tilde{\Sigma}_\psi), \tag{36}$$

where $\tilde{\mu}_\psi = \tilde{\Sigma}_\psi H' \tilde{\Sigma}_\beta^{-1}(\tilde{\mu}_\beta - h)$, and $\tilde{\Sigma}_\psi = (H' \tilde{\Sigma}_\beta^{-1} H)^{-1}$. Furthermore, $\tilde{\mu}_\beta = \tilde{\Sigma}_\beta \mathrm{vec}(Z_1 N Z_0' \Omega^{-1}\alpha)$ whereas $\tilde{\Sigma}_\beta^{-1} = [\alpha'\Omega^{-1}\alpha \otimes C_1]$.

Proof: See Appendix D. □

Notice that integration of the full conditional posterior distribution of Ω with respect to Φ and Γ results in a loss of degrees of freedom equal to $d + p(k-1)$. This is one cost for the lower computation time when Gibbs sampling is based on Proposition 5 rather than on Proposition 1. Another cost is that the covariance matrices of the conditional distributions of α and ψ in Proposition 5 are greater than or equal to the covariances matrices of the conditionals distributions of these parameters in Proposition 1.

The posterior mode estimator of (α, β, Ω) under the exactly identifying restrictions in (4) can be computed using similar techniques as in Proposition 4. In this case we would replace $(T+p+q+r+m+1)$ in the definitions of the S_{ij} matrices with $(T+p+q+r+m+1-d-p(k-1))$. This means that the posterior mode estimators of (α, β) are identical for the full posterior density $p(\alpha, \psi, \Phi, \Gamma, \Omega|\mathfrak{D}, r)$ and the marginal posterior density $p(\alpha, \psi, \Omega|\mathfrak{D}, r)$, while the posterior mode estimators of Ω for the density $p(\alpha, \psi, \Omega|\mathfrak{D}, r)$ is proportional to the posterior mode estimator of Ω for $p(\alpha, \psi, \Phi, \Gamma, \Omega|\mathfrak{D}, r)$. Since the constant of proportionality converges to unity as $T \to \infty$, these two posterior mode estimators of Ω are asymptotically equivalent. Moreover, under over-identifying restrictions on β, we can apply the switching algorithm discussed in Section 4.3 for the full posterior density. Again, the posterior mode estimator of (α, ψ) is identical to the posterior estimator of these parameters for the full posterior, while the posterior mode estimators of Ω are proportional and asymptotically equivalent.

4.6. *Marginal Posterior Distributions*

The marginal posterior distributions of Φ, Γ, and Ω are available by first marginalizing their densities conditional on α and β, and, second, by using posterior draws of α and β. The distribution of Ω conditional on α and β is given above in Proposition 5 and belongs to the inverted Wishart family, while the distributions of Γ and Φ conditional on α and β are matrix t. In fact, from the proof of Proposition 2 it can be deduced that

$$\Gamma|\alpha, \beta, \mathfrak{D}, r \sim t_{p \times p(k-1)}\left[\hat{\Gamma}, \left(A + (1/\lambda_\alpha^2)\alpha\beta'\beta\alpha' + WNW'\right)^{-1}, \left(Z_2 M_D Z_2' + \Sigma_\Gamma^{-1}\right)^{-1}, T+q+r-pn-d\right], \quad (37)$$

where $\hat{\Gamma} = WM_D Z_2'[Z_2 M_D Z_2' + \Sigma_\Gamma^{-1}]^{-1}$, $W = Z_0 - \alpha\beta' Z_1$, while N, M_D, and n are defined in Proposition 2. Similarly, for Φ it can be shown that

$$\Phi|\alpha, \beta, \mathfrak{D}, r \sim t_{p \times d}\left[\hat{\Phi}, \left(A + (1/\lambda_\alpha^2)\alpha\beta'\beta\alpha' + WNW'\right)^{-1}, \left(DM_{Z_2}D'\right)^{-1}, T+q+r-pn-d\right], \quad (38)$$

where $\hat{\Phi} = WM_{Z_2}D'(DM_{Z_2}D')^{-1}$ and $M_{Z_2} = I_T - Z_2'(Z_2 Z_2' + \Sigma_\Gamma^{-1})^{-1}Z_2$. It should be noted that these partially marginalized conditional distributions do not depend on the choice $\beta = \beta_c$. Alternatively, we can compute partially marginalized posteriors of Φ, Γ, and Ω by first marginalizing their densities conditional on β, and then using posterior draws of β. These conditional posteriors are still matrix t for Φ and Γ conditional on β, respectively, and inverted Wishart for Ω conditional on β.

The marginal distribution of the individual α parameters can be calculated as follows. Let α_{ij} denote the element in row i and column j of α. From the relationship between the matrix t and the distribution of an individual element (cf. Section 3.1, or Box and Tiao, 1973, Section 8.4.3), and from Proposition 2 we know that

$$\alpha_{ij}|\beta, \mathfrak{D}, r \sim t_{1\times 1}\left[\hat{\alpha}_{ij}, 1/e_i^{(p)\prime}\left(A + Z_0 N[Z_0 - \hat{\alpha}\beta' Z_1]'\right)e_i^{(p)}, e_j^{(r)\prime}\left(\beta' C_1 \beta\right)^{-1} e_j^{(r)}, T+q-(pn+d)\right],$$

where $e_i^{(s)}$ is the i:th column of I_s.

The value of the marginal posterior of α_{ij} can now be approximated at a point $\tilde{\alpha}_{ij}$ by evaluating the marginal conditional density of α_{ij} at this point for a draw of β from its posterior and averaging over all such β draws. That is, we compute

$$\hat{p}\left(\tilde{\alpha}_{ij}|\mathfrak{D},r\right) = \frac{1}{G}\sum_{g=1}^{G} p\left(\tilde{\alpha}_{ij}|\beta^{(g)},\mathfrak{D},r\right).$$

From the ergodic theorem (cf. Tierney, 1994) $\hat{p}(\tilde{\alpha}_{ij}|\mathfrak{D},r) \xrightarrow{\text{a.s.}} p(\tilde{\alpha}_{ij}|\mathfrak{D},r)$ as $G \to \infty$. This calculation can be performed over a grid of $\tilde{\alpha}_{ij}$ values providing us with a sequence of pairs of $(\tilde{\alpha}_{ij}, \hat{p}(\tilde{\alpha}_{ij}|\mathfrak{D},r))$ that can be used to approximate the marginal posterior of α_{ij}, where a finer grid improves the approximation. We can compute integer moments as well as other properties of the marginal posterior from this sequence.

Moments of the marginal posterior of α_{ij} can also be estimated from the marginal conditional distribution. The marginal posterior mean, for example, can be estimated by

$$E\left[\alpha_{ij}|\mathfrak{D},r\right] = \frac{1}{G}\sum_{g=1}^{G} E\left[\alpha_{ij}|\beta^{(g)},\mathfrak{D},r\right],$$

where the expectation on the right hand side is given by $\hat{\alpha}_{ij}$ and the sample average of this term is an estimate of its expectation with respect to $\beta|\mathfrak{D},r$. This estimator is unbiased, consistent, and converges in distribution to a Gaussian at the rate $\sqrt{G}$; see Tierney (1994) for details.

Regarding precision, note that the numerical standard error of the estimated marginal posterior mean is not equal to the square root of $\text{Var}(\hat{\alpha}_{ij})/G$ since the Gibbs output $\beta^{(1)},\ldots,\beta^{(G)}$ is dependent. Letting γ_s denote the sth order autocovariance of $\hat{\alpha}_{ij}$, then the numerical standard error of the estimated posterior mean can be calculated as the square root of

$$\frac{1}{G}\left(\gamma_0 + 2\sum_{s=1}^{\bar{G}}\gamma_s\frac{\bar{G}+1-s}{\bar{G}+1}\right),$$

where $\bar{G} \leq G$ is an integer such that the autocorrelation function tappers off; cf. Newey and West (1987). Similarly, and as suggested by Chib (1995), we can calculate the numerical standard error of $\hat{p}(\tilde{\alpha}|\mathfrak{D},r)$, needed for the marginal likelihood of cointegration rank r in equation (20), using the above sample variance estimator but with

$$\gamma_s = \frac{1}{G}\sum_{g=s+1}^{G}\left(p\left(\tilde{\alpha}|\beta^{(g)},\mathfrak{D},r\right) - \hat{p}\left(\tilde{\alpha}|\mathfrak{D},r\right)\right)\left(p\left(\tilde{\alpha}|\beta^{(g-s)},\mathfrak{D},r\right) - \hat{p}\left(\tilde{\alpha}|\mathfrak{D},r\right)\right).$$

An estimator of the marginal posterior variance of α_{ij} can be determined from the well known identity $\text{Var}[\alpha_{ij}|\mathfrak{D},r] = E[\text{Var}(\alpha_{ij}|\beta,\mathfrak{D},r)] + \text{Var}[E(\alpha_{ij}|\beta,\mathfrak{D},r)]$. That is,

$$\text{Var}\left[\alpha_{ij}|\mathfrak{D},r\right] = \frac{1}{G}\sum_{g=1}^{G}\text{Var}\left[\alpha_{ij}|\beta^{(g)},\mathfrak{D},r\right] + \frac{1}{G}\sum_{g=1}^{G}\left(E\left[\alpha_{ij}|\beta^{(g)},\mathfrak{D},r\right] - E\left[\alpha_{ij}|\mathfrak{D},r\right]\right)^2,$$

where the second term on the right hand side is an estimate of the variance (with respect to $\beta|\mathfrak{D},r$) of $\hat{\alpha}_{ij}$, while the first term is an estimate of the expected value (with respect to $\beta|\mathfrak{D},r$) of the conditional variance of α_{ij}. The latter variance is given by

$$\text{Var}\left[\alpha_{ij}|\beta^{(g)},\mathfrak{D},r\right] = \frac{e_j^{(r)\prime}\left(\beta^{(g)\prime}C_1\beta^{(g)}\right)^{-1}e_j^{(r)}e_i^{(p)\prime}\left(A + Z_0N[Z_0 - \hat{\alpha}\beta^{(g)\prime}Z_1]'\right)e_i^{(p)}}{T+q-(pn+d)-1}.$$

The marginal posteriors of the other parameters can be numerically determined in a similar way. If $\beta = \beta_c$, then the Student t distribution can be used for the marginal conditional posterior of the individual Ψ parameters. In this case a posterior sample of α is required. On the other hand, if β is subject to over-identifying restrictions, then the normal distribution can be employed as the marginal conditional posterior of the individual ψ parameters, while a posterior sample of α and Ω is needed for the approximation of its marginal posterior. It should be kept in mind that the marginal posterior of Ψ does not have any finite integer moments, while ψ is likely to have finite integer moments up to the order of the number of over-identifying restrictions; see, e.g., Bauwens and Lubrano (1996).

For Γ and Φ we can likewise make use of equations (37)–(38), respectively, and a sample of posterior draws of α and β when we wish to calculate the marginal posteriors of their individual elements. As mentioned above, the posteriors of these parameters conditional on β are also matrix t and, hence, can alternatively be used given a posterior sample of β.

5. The Demand for Euro Area M3

A number of studies have recently attempted to estimate the long-run demand for M3 in the euro area; see, e.g., Brand and Cassola (2000), Calza, Gerdesmeier, and Levy (2001), Coenen and Vega (2001), Kontolemis (2002), BDW (2003), Carstensen (2003, 2004), and references therein. A common finding in these studies is that the estimated income elasticity of long-run money demand tends to exceed unity and is sometimes closer to 1.5 than to 1. In addition, the interest rate semi-elasticities are often imprecisely estimated in the sense that the error bands are very wide.

All these studies apply classical methods for estimating the parameters of interest and investigating the uncertainty of the point estimates. With the exception of BDW, who also make use of bootstrapping, inference is solely conducted on the basis of asymptotic theory. This means that, e.g., the income elasticity is biased in small samples and the size and the direction of the bias is unknown.[5]

Since the trace test for determining the number of long-run relations in the Gaussian maximum likelihood setting often has low power in small samples and thus leads to under-estimating the number of such relations, below I will consider two Bayesian approaches to this issue. Furthermore, I compare the results on income elasticities and interest rate semi-elasticities for long-run money demand from the Bayesian analysis to results obtained using maximum likelihood. Before I consider these issues, the BDW model is briefly presented along with hypotheses about the cointegration relations.

5.1. *The Data and Hypotheses about β*

The benchmark money demand system for the euro area in BDW consists of 6 variables: real M3, m_t, inflation measured by annualized quarterly changes of the GDP deflator, Δp_t, real GDP, y_t, the short-term (3-month) market interest rate, $i_{s,t}$, the long-term (10-year) market interest rate, i_{lt}, and the own rate of return of M3, $i_{o,t}$. The interest rates are all measured in annual percentage rates, while the remaining variables are measured in natural logarithms of seasonally adjusted data multiplied by 100. Hence, I use a different scale for the variables than BDW do. Finally, the money stock, GDP, and the GDP deflator have been aggregated using the irrevocably fixed exchange rates, while M3 weights have been used for the aggregation of the interest rate. For the cointegrated VAR model the authors set the deterministic vector $D_t = 1$ without imposing any restrictions on Φ. Hence, as long as there is one unit root the endogenous variables are allowed to be dominated by a linear deterministic trend.

---

[5] Alternative factors that may explain not just a small sample bias of the income elasticity estimate as well as of the interest rate (semi-)elasticities, but a measurement related bias are discussed in depth by Reynard (2004, 2005).

The sample considered by BDW is 1980:Q2–2001:Q4. Since additional and especially interesting data are now available I will examine data until 2004:Q4. It thus includes the period of exceptional liquidity preference between 2001 and mid-2003. The portfolio shifts into liquidity has resulted in the annual growth rate of nominal M3 to rise far above the reference value of 4.5 percent with a peak above 8 percent. As a consequence, the ECB has attempted to construct portfolio shift corrected measures of M3; see European Central Bank (2004) for details. In this paper I use the M3 series based on the index of notional stocks which thus incorporates the effects of portfolio shifts into money.

With $x_t = [m_t\ \Delta p_t\ y_t\ i_{s,t}\ i_{l,t}\ i_{o,t}]'$, BDW finds evidence of two cointegration relations in their study and considered a range of potential cointegration vectors. In this paper I will focus on one of their β models for $r = 2$. Namely,

$$\begin{aligned} m_t - \beta_y y_t - \beta_o\big(i_{s,t} - i_{o,t}\big) &\sim I(0), \\ i_{o,t} - \beta_\pi \Delta p_t - \beta_s i_{s,t} - \beta_l i_{l,t} &\sim I(0). \end{aligned} \tag{39}$$

The first relation is interpreted as a long-run money demand relation, while the second may be regarded as a pricing relation for the own rate of return on M3. The parameters of interest are β_y, the income elasticity, and β_o, the opportunity cost semi-elasticity. Given such an interpretation for the long-run parameters we expect $\beta_y > 0$ and $\beta_o < 0$.

Given the set of variables in the model we may also consider other candidates for stationary relations. Such candidates include stationary interest rates and inflation, as well as stationary interest rate spreads and stationary real interest rates.

In Figure 1 the long-run money demand relation has been plotted based on the maximum likelihood estimates for data until 2001:Q4 and a cointegration rank of 2. In this case the point estimates of $\beta_y = 1.37$ and $\beta_o = -0.44$ and the sample mean has been extracted from the series. It is striking that from some point in late 2001 or early 2002, the deviation from steady state is positive and trending upwards. Hence, it seems unlikely that a model which neglects this behavior can be considered well specified. Rather, this graph suggests that there is a missing variable and such a variable may need to reflect the increased preference for liquidity towards the end of the sample.

In this paper I will therefore change the definition of the deterministic term for the BDW model. Specifically, I will include a broken linear trend from $t_0 = 2001\text{:Q4}$ until 2004:Q4. The main reason for picking such a break date is the increased geopolitical and economic uncertainty that followed in the aftermath of September, 11, 2001. Hence, I will use $D_t = [1\ \ d_t]'$, where

$$d_t = \begin{cases} t - t_0 + 1 & \text{if } t_0 \leq t \leq T, \\ 0 & \text{otherwise.} \end{cases}$$

From a time series model point of view, the d_t variable introduces the possibility of a quadratic trend influencing the endogenous variables after t_0; see, e.g., Johansen (1996, pp. 80–83). To avoid this we can restrict the parameters on d_t, denoted by Φ_d, such that they satisfy the restrictions $\Phi_d = \alpha\mu_d$. This yields exactly $p - r$ restrictions on Φ_d. In this paper I will not impose such restrictions since it is inconsistent with an informative prior for α and a non-informative prior for Φ. The general issue of including such restrictions for the parameters on deterministic variables in a Bayesian cointegrated VAR is examined by

Villani (2005c) for an informative prior of Φ. This is an important issue since Φ is directly related to the steady state of the system.

5.2. *Hyperparameters of the Prior Distribution*

Before we consider the determination of a proper lag order and cointegration rank benchmark values of the hyperparameters of the prior need to be established. Since the interest rates, inflation, and the first differences of real money and income are measured in percentage rates, the elements of Ω can be expected to take on values between -1 and 1. In this paper I let the matrix $A = \lambda_A I_p$, where $\lambda_A = 1/5$ while $q = p + 2$. This choice for q minimizes the effect of A on the posterior subject to a finite expected value of Ω. An alternative is to let A be equal to the maximum likelihood estimate of Ω for the largest cointegration rank and the largest lag order, denoted by $\hat{\Omega}^{(p,k^*)}$:

$$A = \hat{\Omega}^{(6,6)} = \begin{bmatrix} 0.084 & & & & & \\ -0.097 & 0.326 & & & & \\ 0.007 & -0.013 & 0.095 & & & \\ -0.019 & 0.006 & 0.013 & 0.087 & & \\ 0.006 & -0.002 & 0.002 & 0.036 & 0.046 & \\ -0.002 & 0.000 & 0.001 & 0.022 & 0.009 & 0.008 \end{bmatrix}.$$

Such a choice for A is not a proper Bayesian solution since the prior here depends on the data. Still, from $\hat{\Omega}^{(6,6)}$ it can be seen that my choice for A approximately has the same scale as the maximum likelihood estimate of Ω.

The hyperparameter $\lambda_\alpha = 0.7$ and its value depends on the chosen scale for the endogenous variables. As $\lambda_\alpha \to 0$ it is very likely that $p(r|\mathfrak{D}) \to p(r)$ and, hence, that data is completely uninformative above the rank. Furthermore, as λ_α becomes very large we expect $p(r = 0|\mathfrak{D}) \to 1$ since the joint prior becomes non-informative about r and Bartlett's paradox then applies.

Under the informative prior on $\Gamma|\Omega$, the hyperparameter $\lambda_b = 1.5$ while $\lambda_l = 1$. As λ_b becomes very large, the prior approaches the non-informative prior on $\Gamma|\Omega$. Finally, the maximum lag order $k^* = 6$, while the priors on the cointegration rank and lag order are uniform with $p(r) = p(r|k) = 1/7$ and $p(k) = p(k|r) = 1/6$.

5.3. *Cointegration Rank and Lag Order Analysis*

The results on lag order determination based on Corollary 1 are presented in Table 1.[6] Almost all the posterior probability mass is given to $k = 2$, the lag order selected by BDW. This choice is also preferred in models where $\lambda_l \in [0.5, 1.5]$, while $\lambda_l = 2$ gives roughly equal probabilities for $k \geq 2$. Increasing the value of λ_b has the expected effect of increasing the posterior probability for, in particular, $k = 1$. For instance, $\lambda_b = 2$ results in a 99 percent posterior probability that $k = 1$. Smaller values for λ_b similarly result in higher posterior probabilities for longer lag orders, where, e.g., $\lambda_b = 0.7$ leads to a posterior probability of $k = 2$ equal to 56 percent while the rest of the probability mass is found for $k \geq 3$.

[6] All computation in this paper have been carried out with Structural VAR. This software is available for downloaded from my website at: `http://www.texlips.net/svar/`.

Furthermore, the λ_α parameter has only minor effects on the posterior lag order probabilities for the full rank models. For very small values the probability that $k = 3$ dominates, while values above 1.5 give 83 percent probability on $k = 2$ and the remaining 17 percent on $k = 1$. The results are also relatively robust with respect to the Ω prior hyperparameters. Once λ_A is above $3/5$ the largest posterior probability is assigned to $k = 1$, while values of q in excess of 40 leads to $k = 6$ having the highest posterior probability.

For comparison, we may also examine if standard information criteria are consistent with $k = 2$. The SBC (Schwarz, 1978) and the FML (Fractional Marginal Likelihood; see Villani, 2001b) criteria suggest $k = 1$, while HQ (Quinn, 1980) and AIC (Akaike, 1974) prefer $k = 2$ when a maximum of 6 lags is considered. Moreover, based on the FML, which is asymptotically similar to the SBC, $p(k = 1|\mathfrak{D}) = 0.78$, $p(k = 2|\mathfrak{D}) = 0.19$, $p(k = 3|\mathfrak{D}) = 0.02$, and almost 0 for higher k.

Let us therefore turn to the analysis of the cointegration rank conditional on $k = 2$. An approximation formula for the marginal likelihood $p(\mathfrak{D}|r)$ was suggested by Corander and Villani (2004). As in Villani (2001b) for the lag order, the fractional Bayes approach proposed by O'Hagan (1995, 1997) is used. This yields a simple expression that only requires the maximum likelihood estimates of Ω for the various cointegration ranks, denoted by $\hat{\Omega}^{(r)}$. The fractional marginal likelihood for rank r can then be approximated by:

$$p_b\big(\mathfrak{D}|r\big) \propto \frac{\Gamma_p(T - f_r)}{\Gamma_p(f_m - f_r)} \left[\frac{|\hat{\Omega}^{(r)}|}{|\hat{\Omega}^{(0)}|}\right]^{-(T-f_m)/2}, \tag{40}$$

where $f_r = d + p(k-1) + r + (p-r)r/p$ (the number of free parameter in each equation), $f_m = d + p(k+1)$, and $b = f_m/T$ is the minimal fraction of the data used to "train" the improper prior $p(\alpha, \beta, \Phi, \Gamma, \Omega|r) \propto |\Omega|^{-(p+1)/2}$ into a proper distribution for the largest model ($r = p$). Notice that the term within brackets is the likelihood ratio for cointegration rank 0 versus r. As shown by Corander and Villani (2004, Theorem 1), the log of the approximate fractional marginal likelihood in (40) is an $O_p(1)$ approximation of the log marginal likelihood given our choice for f_r.

In Table 2 the classical LR trace tests (Johansen, 1996) are given, along with bootstrapped p-values, as well as logs of the fractional marginal likelihoods and the rank probabilities based on these. If we consider a nominal size of 5 (10) percent, then the trace tests suggest a cointegration rank of 2 (3), while the fractional marginal likelihoods give roughly 95 percent probability to rank 4, and 5 percent to rank 2.

To calculate the posterior rank probabilities from the marginal likelihoods in Section 4.2 we need to select a Gibbs sampler for obtaining posterior draws of β. In this paper I have opted for the sampler in Proposition 5. Implementation of the sampler requires initial values for the parameters. Since any parameter values from their support are valid, I use the posterior mode estimator in Proposition 4. After 2,500 burn-in draws, an additional 5,000 draws were simulated from the posterior distribution of the model parameters. Convergence of the sampler has been checked through visual inspection of the relative constancy of the recursive posterior median point estimates of the parameters. More elaborate approaches to study convergence of the Gibbs sampler (and other MCMC methods) are discussed by, e.g., Bauwens et al. (1999) and Geweke (1999).

The posterior cointegration rank probabilities under the Bayesian approach conditional on $k = 2$ as well as the log marginal likelihood values and their numerical standard errors are given in the last six columns of Table 2. Under the non-informative $\Gamma|\Omega$ prior, the preferred rank is 1, with roughly 20 percent probability given to $r = 2$ and 15 percent to $r = 0$. For the informative $\Gamma|\Omega$ prior we find more support for $r > 1$ with approximately the same posterior probabilities for $r = 1$ and $r = 2$. Hence, decreasing the value for the λ_b parameter tends to increase the posterior probabilities for models with a higher rank. Once $\lambda_b = 0.5$, the posterior probabilities of $r = 2$ and $r = 3$ are roughly equal and dominate the others.

Regarding robustness with respect to the other hyperparameters, the posterior probabilities of $r = 0$ and $r = 1$ tend to rise when λ_α increases, while the probabilities of higher cointegration ranks tend to increase when this hyperparameter decreases. However, as λ_α becomes sufficiently low, the low cointegration rank models again seem to dominate. These results are obtained for both models with an informative and a non-informative $\Gamma|\Omega$ prior. Similarly, higher (lower) values of λ_A leads to posterior probabilities that favor lower (higher) cointegration rank models, while higher values of q increases the support for models with a higher cointegration rank.

In Section 4.4 I discussed a procedure for computing joint and marginal posterior probabilities of the cointegration rank and the lag order. This procedure requires that the informative $\Gamma|\Omega$ prior is used. The results from this analysis is summarized in Table 3. Concerning joint posterior probabilities, the highest probability is obtained for the pair $(r = 2, k = 2)$, with the pair $(r = 1, k = 2)$ obtaining a probability just below this maximum. For the marginal lag order probabilities we find that $k = 2$ is the preferred choice with roughly 94 percent of the probability, while $k = 3$ obtains about 6 percent. The marginal cointegration rank probabilities give roughly the same probability to $r = 1$ and $r = 2$. Given that $k = 2$ dominates so strongly in the lag order dimension, it is not surprising that the marginal rank probabilities are very similar to those found when we conditioned on $k = 2$ in Table 2.

To summarize our results thus far, the lag order appears to be quite well determined and in what follows I will condition the analysis on $k = 2$. The different Bayesian methods I have considered give very different suggestions for the choice of a proper cointegration rank. The fractional pseudo-Bayesian approach strongly suggests that $r = 4$ is the best choice, while the two more elaborate Bayesian approaches, with either a non-informative or an informative $\Gamma|\Omega$ prior, point towards a lower rank. The precise choice for these approaches is, however, less evident. If we, e.g., prefer having a prior with a large λ_b value, then $r = 1$ obtains most support, while moderate to small λ_b values give greater support to $r = 2$ (or larger). Given that the BDW model specifies that $r = 2$ we will condition on this value below. It should be emphasized that the choice of cointegration rank for the current data is not clear-cut. In particular, the choices of certain prior hyperparameters tend to greatly influence the posterior results. Hence, the data itself is not as informative about the cointegration rank as one would hope for or as, for instance, the classical analysis may suggest.

5.4. *Long-Run Money Demand*

Table 4 provides point estimates and 95 percent error bands for the income elasticity and the opportunity cost semi-elasticity. The two long-run relations in (39) imply 3 over-identifying restrictions on β. When evaluating these with a *LR*-test we obtain a value of 5.23, corresponding to a *p*-value of 16 (41) percent when compared with the asymptotic (bootstrap) distribution.

From a Bayesian perspective we can again examine the posterior probability of the model for β. For the two competing β models used for the *LR*-test, i.e., the restricted and the unrestricted β models, we can compute their posterior probabilities through an approximation of the marginal likelihood suggested by Draper (1995). Let $M_{r,i}$ denote the model with cointegration rank r and β satisfying (9) for a given pair (h_i, H_i) of linear restrictions matrices. The following approximation of the marginal likelihood for $M_{r,i}$ can then be considered:

$$\ln p\big(\mathfrak{D}|M_{r,i}\big) \approx \frac{n_{r,i}}{2}\ln\big(2\pi\big) + \ln p\big(\mathfrak{D}|\hat{\theta}_{r,i}\big) - \frac{n_{r,i}}{2}\ln T. \tag{41}$$

The number of free parameters for model $M_{r,i}$ is given by $n_{r,i}$, while $p(\mathfrak{D}|\hat{\theta}_{r,i})$ is the maximum of the likelihood function under $M_{r,i}$. Given equal prior probabilities $(1/2)$ of the models, the posterior probability of the restricted β model is 82 percent. Hence, both the classical and the Bayesian assessments about these models favor the restricted β model.

The marginal distributions of the income elasticity and the opportunity cost semi-elasticity are plotted in Figure 2 along with the bounds of the 95 percent confidence bands and the posterior mode estimates. From these graphs it can be seen that the choice of prior has only minor influence on the marginal posteriors. While the mode shifts to the right under the informative priors, this shift is small. Moreover, the densities are unimodal and are approximately symmetric.

From Table 4 we find that the maximum likelihood estimate of the income elasticity is close to the point estimates from the marginal posterior distributions of both the non-informative and the informative $\Gamma|\Omega$ prior. All these estimates are in line with the point estimate provided by BDW for the sample 1980:Q2–2001:Q4. Furthermore, the 95 percent classical error bands are very similar for this parameter and have lower and upper bounds that come close to the 95 percent confidence bands of the marginal posterior distributions.

Regarding the opportunity cost semi-elasticity the maximum likelihood estimate is bigger (smaller) than the point estimates from the marginal posterior distributions under the non-informative (informative) $\Gamma|\Omega$ prior. The 95 percent classical error bands differ greatly, with the asymptotic unconditional band being the largest. Nevertheless, the empirical evidence from both the classical and the Bayesian assessment confirms BDW's finding that there is very little information in the data about the interest rate semi-elasticity of long-run money demand.

6. Concluding Comments

Building on recent work by Villani (2005b), I have presented an approach to Bayesian inference in cointegrated VAR models with Gaussian errors. Relative to other Bayesian cointegration approaches, Villani's setup is particularly interesting since it uses a natural uniform prior on the cointegration space. The current study extends Villani's framework in two important dimensions. First, the prior distribution of the parameters on lagged endogenous variables can be proper, thereby making it possible to determine the lag order of the VAR as well as the cointegration rank jointly. In addition, marginal posterior probabilities of these two parameters can be calculated.

Second, to compute the posterior rank and lag order probabilities we require the marginal likelihoods of the various models we wish to compare. Analytical expressions of these data density value measures are available for the full and zero cointegration rank models conditional on any finite lag order, but not for the reduced rank cases where numerical methods are needed. Villani (2005b) considered three simulation based methods to handle this issue and in his Monte Carlo study the so called marginal likelihood identity (MLI) approach consistently outperformed the other two algorithms. The MLI, due to Chib (1995), requires a point estimator of the parameters and this estimator should preferably have high posterior density. A natural candidate is therefore the posterior mode and an analytical expression of this point estimator is provided for the just identified case, thus greatly facilitating the numerical evaluation of the marginal likelihoods.

The Bayesian procedures based on a proper (informative) and an improper (non-informative) prior on the short-run dynamics are applied to the Bruggeman, Donati, and Warne (2003) money demand system for the euro area. The preferred specification in BDW has two cointegration relations and two lags in a system comprising real money, real income, inflation, a short and long term rate as well as the own rate of return of M3, and the data set used by BDW is extended from 2001:Q4 to 2004:Q4. In addition to applying the two Bayesian procedures, a fractional pseudo-Bayesian approach, suggested by Corander and Villani (2004), and the classical maximum likelihood approach, due to Johansen (1988, 1991), extended with bootstrapping, are used for comparisons. Given a fixed lag order, the classical trace tests for the cointegration rank suggests two long-run relations, while the pseudo-Bayesian prefers four. In contrast, the two Bayesian approaches I have presented tend to prefer a lower cointegration rank (given the selected prior) than the pseudo-Bayesian procedure. When the improper prior is used for the short-run dynamics, the preferred rank is one, while the informative prior gives roughly equal weight to the models with rank one and with rank two.

The cointegration rank and the lag order have also been studied based on their marginal posterior probabilities given a maximum lag order of six. This requires an informative prior on the short-run dynamics and the results strongly support the use of two lags and the finding that the models with one and two cointegration relations have approximately the same posterior probabilities and are preferred over the other models. It should be emphasized, however, that the empirical results on cointegration rank are sensitive to the selected values of certain parameters of the prior distribution.

The parameters of long-run money demand have been studied within the cointegrated VAR model with rank two. For the income elasticity the Bayesian approaches yield similar point estimates as the maximum likelihood estimator. Furthermore, the 2.5 and 97.5 percent quantiles of its marginal posterior distribution are roughly equal to the lower and upper limits of the 95 percent error bands from the classical analysis. The point estimates of the opportunity cost semi-elasticity, however, are more sensitive to the choice of prior and the maximum likelihood estimate lies between the estimates from the marginal posteriors of the two Bayesian approaches. Still, the Bayesian analysis confirms the finding in BDW that the data has little information about this parameter.

TABLE 1: Lag order determination based on full rank models with an informative prior on $\Gamma|\Omega$ with a maximum lag order equal to 6.

k	$\ln p(\mathfrak{D}\|k, r = 6)$	$p(k\|\mathfrak{D}, r = 6)$
1	−330.03	0.02
2	−326.12	0.98
3	−337.37	0.00
4	−344.44	0.00
5	−344.26	0.00
6	−348.04	0.00

NOTE: The log marginal likelihood values are calculated using Corollary 1 without removing multiplicative constants and with $\lambda_\alpha = 0.7$, $\lambda_b = 1.5$, $\lambda_l = 1$, $A = (1/5)I_6$, and $q = 8$. The sample is given by 1981:Q4–2004:Q4 for all k.

TABLE 2: Cointegration rank determination using classical trace tests with bootstrapped p-values, the fractional marginal likelihood, and the marginal likelihood conditional on $k = 2$ lags.

	Classical		Fractional		Non-informative $\Gamma\|\Omega$			Informative $\Gamma\|\Omega$		
r	LR_{tr}	p-value	$\ln p_b(\mathfrak{D}\|r)$	$p(r\|\mathfrak{D})$	$\ln p(\mathfrak{D}\|r)$	st. err.	$p(r\|\mathfrak{D})$	$\ln p(\mathfrak{D}\|r)$	st. err.	$p(r\|\mathfrak{D})$
0	156.13	0.00	694.94	0.00	−327.25	—	0.15	−344.55	—	0.01
1	100.19	0.00	702.45	0.00	−325.79	0.03	0.63	−340.44	0.02	0.37
2	57.74	0.08	711.30	0.05	−326.99	0.08	0.19	−340.30	0.13	0.42
3	30.14	0.22	704.38	0.00	−328.84	0.10	0.03	−341.18	0.17	0.17
4	5.13	0.93	714.30	0.95	−332.01	0.04	0.00	−342.93	0.04	0.03
5	0.01	0.94	705.90	0.00	−337.81	0.04	0.00	−348.33	0.04	0.00
6	—	—	705.90	0.00	−344.77	—	0.00	−355.04	—	0.00

NOTE: Numerical standard errors are computed for the log marginal likelihood values for cointegration rank 1–5 using the Newey and West (1987) correction as discussed in Section 4.6 and the delta method. The sample is 1980:Q4–2004:Q4

TABLE 3: Joint and marginal posterior probabilities of the cointegration rank and the lag order.

	k						
r	1	2	3	4	5	6	$p(r\|\mathfrak{D})$
0	0.00	0.02	0.05	0.00	0.00	0.00	0.07
1	0.00	0.37	0.01	0.00	0.00	0.00	0.38
2	0.00	0.41	0.00	0.00	0.00	0.00	0.41
3	0.00	0.13	0.00	0.00	0.00	0.00	0.13
4	0.00	0.01	0.00	0.00	0.00	0.00	0.01
5	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$p(k\|\mathfrak{D})$	0.00	0.94	0.06	0.00	0.00	0.00	1.00

NOTE: See Table 1 for details.

TABLE 4: Point estimates of the income elasticity and the interest rate semi-elasticity of long-run money demand along with 95 percent bands for classical and Bayesian methods.

Method	Statistic	β_y	β_o
Classical	ML estimate	1.38	−0.26
	Asymptotic conditional error band	[1.34, 1.42]	[−0.63, 0.11]
	Asymptotic unconditional error band	[1.31, 1.44]	[−1.08, 0.54]
	Bootstrap percentile-t error band	[1.32, 1.45]	[−0.91, 0.45]
Bayesian	Posterior mode	1.37	−0.29
non-informative	Posterior median	1.37	−0.31
$p(\Gamma\|\Omega)$	Posterior mean	1.37	−0.31
	95 percent confidence bands	[1.30, 1.44]	[−1.13, 0.56]
Bayesian	Posterior mode	1.38	−0.17
informative	Posterior median	1.38	−0.17
$p(\Gamma\|\Omega)$	Posterior mean	1.39	−0.18
	95 percent confidence bands	[1.32, 1.45]	[−1.05, 0.67]

NOTE: Bayesian point estimators are taken from the marginal posterior distribution of β_y and β_o, respectively. The asymptotic conditional error band is constructed from an estimate of the conditional standard error. The asymptotic unconditional error band is calculated by selecting the parameter of interest from a range of values. Conditional on each such value, the other free parameters are re-estimated by maximum likelihood and the log-likelihood value is recorded. A 95 percent bound for the log-likelihood value is constructed by subtracting $\chi^2_{0.95}(1)/2$ from the largest log-likelihood value. The 95 percent unconditional error band for the parameter of interest are then constructed based on the parameter values that have log-likelihood values equal to this bound. For the bootstrap percentile-t error band I have used 999 bootstrap replications.

FIGURE 1: Long-run money demand relation 1980:Q4–2004:Q4 based on maximum likelihood estimates for data until 2001:Q4.

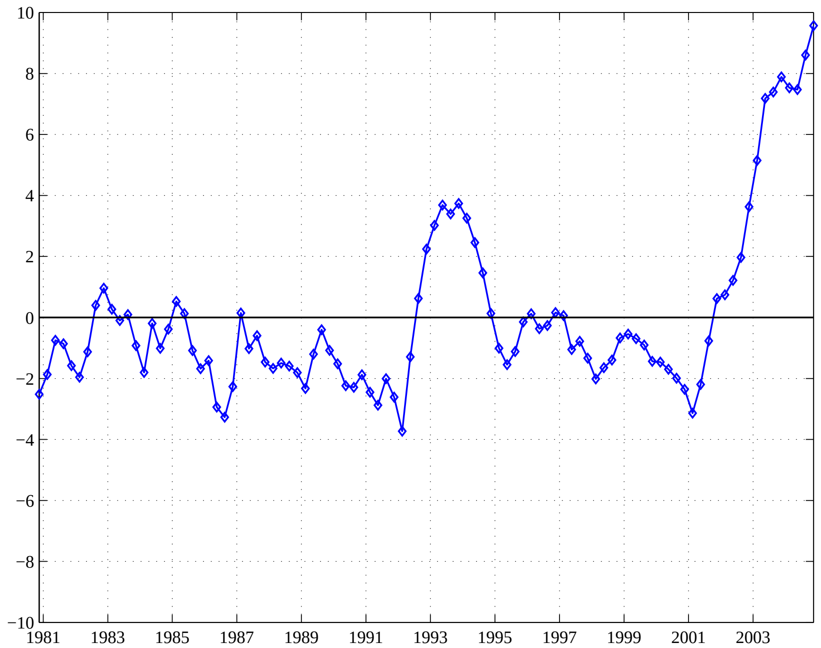
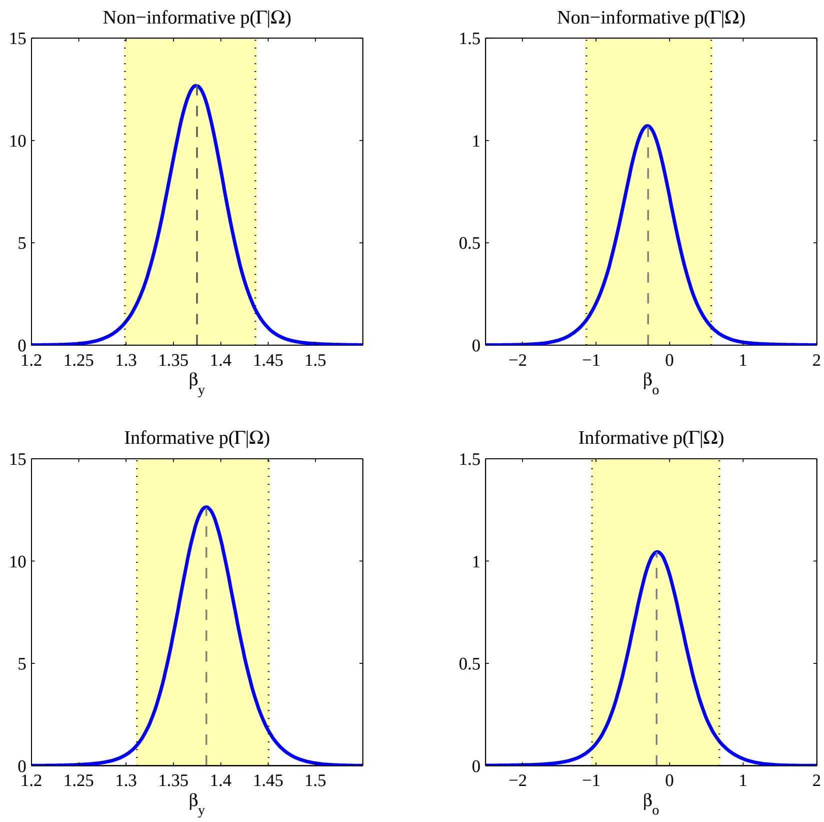


FIGURE 2: Marginal posterior densities of the income elasticity (β_y) and the opportunity cost semi-elasticity (β_o) under the non-informative and the informative priors with posterior mode and 95 percent confidence bands.



(i) Premultiplying $\beta = \beta_c$ by K we get $K\beta = [I_r \; \Psi']'$, i.e., the linear normalization of β used by Villani (2005b). The results follow directly from the proof of Theorem 4.6 in Villani (2005b) and the use of Definition 1 above, where the proof of that Theorem is located in Villani (2001a, Theorem 4.5).

(ii) An analytic expression for $p(\alpha, \beta, \mathfrak{D}|r)$ can be derived from $p(\alpha, \beta, \Phi, \Gamma, \Omega, \mathfrak{D}|r)$ by integrating the latter with respect to Ω, Φ, and Γ. Multiplying the data density by the prior density we have that

$$\begin{aligned} p\big(\alpha, \beta, \Phi, \Gamma, \Omega, \mathfrak{D}|r\big) &= \big(2\pi\big)^{-Tp/2}\big(2\pi\big)^{-mp/2} c_r |\Sigma_\Gamma|^{-p/2} |\Omega|^{-(T+p+q+r+m+1)/2} \\ &\quad \times \exp\left(-\frac{1}{2}\operatorname{tr}\big[\Omega^{-1}\big(\varepsilon\varepsilon' + A + \lambda_\alpha^{-2}\alpha\beta'\beta\alpha' + \Gamma\Sigma_\Gamma^{-1}\Gamma'\big)\big]\right), \end{aligned} \tag{A.1}$$

where $m = p(k-1)$. Integrating (A.1) with respect to Ω is directly achieved by using Definition 2 for the inverted Wishart distribution. After some manipulations this provides us with:

$$\begin{aligned} p\big(\alpha, \beta, \Phi, \Gamma, \mathfrak{D}|r\big) &= \nu_r \pi^{-Tp/2} \pi^{-mp/2} \pi^{(2p-r)r/2} \lambda_\alpha^{-pr} |A|^{q/2} |\Sigma_\Gamma|^{-p/2} \frac{\Gamma_p(T+q+r+m)\Gamma_r(p)}{\Gamma_p(q)\Gamma_r(r)} \\ &\quad \times \Big|\varepsilon\varepsilon' + A + \lambda_\alpha^{-2}\alpha\beta'\beta\alpha' + \Gamma\Sigma_\Gamma^{-1}\Gamma'\Big|^{-(T+q+r+m)/2}. \end{aligned} \tag{A.2}$$

Next, for $X = Z_0 - \Gamma Z_2 - \alpha\beta' Z_1$, we have that

$$\varepsilon\varepsilon' = XM_D X' + \big(\Phi' - \hat{\Phi}'\big)' DD' \big(\Phi' - \hat{\Phi}'\big),$$

where $\hat{\Phi} = XD'(DD')^{-1}$ and M_D is defined in Proposition 2. Integration of (A.2) with respect to Φ' can be performed by using Definition 1 for the matrix t distribution. After some algebra this gives us:

$$\begin{aligned} p\big(\alpha, \beta, \Gamma, \mathfrak{D}|r\big) &= \nu_r \pi^{-(T-d)p/2} \pi^{-mp/2} \pi^{-(2p-r)r/2} \lambda_\alpha^{-pr} |A|^{q/2} |\Sigma_\Gamma|^{-p/2} \frac{\Gamma_p(T+q+r+m-d)\Gamma_r(p)}{\Gamma_p(q)\Gamma_r(r)} \\ &\quad \times |DD'|^{-p/2} \Big|A + \lambda_\alpha^{-2}\alpha\beta'\beta\alpha' + \Gamma\Sigma_\Gamma^{-1}\Gamma' + XM_D X'\Big|^{-(T+q+r+m-d)/2}. \end{aligned} \tag{A.3}$$

Now, for $W = Z_0 - \alpha\beta' Z_1$ we have that

$$\Gamma\Sigma_\Gamma^{-1}\Gamma' + XM_D X' = WNW' + \big(\Gamma' - \hat{\Gamma}'\big)' \big[Z_2 M_D Z_2' + \Sigma_\Gamma^{-1}\big] \big(\Gamma' - \hat{\Gamma}'\big),$$

where $\hat{\Gamma} = WM_D Z_2'[Z_2 M_D Z_2' + \Sigma_\Gamma^{-1}]^{-1}$ and N is defined in Proposition 2. Integrating (A.3) with respect to Γ' is again achieved through the definition of the matrix t distribution. This eventually gives us:

$$\begin{aligned} p\big(\alpha, \beta, \mathfrak{D}|r\big) &= \nu_r \pi^{-(T-d)p/2} \pi^{-(2p-r)r/2} \lambda_\alpha^{-pr} |A|^{q/2} |\Sigma_\Gamma|^{-p/2} \frac{\Gamma_p(T+q+r-d)\Gamma_r(p)}{\Gamma_p(q)\Gamma_r(r)} \\ &\quad \times |DD'|^{-p/2} |Z_2 M_D Z_2' + \Sigma_\Gamma^{-1}|^{-p/2} \Big|A + \lambda_\alpha^{-2}\alpha\beta'\beta\alpha' + WNW'\Big|^{-(T+q+r-d)/2}. \end{aligned} \tag{A.4}$$

Finally, by construction $p(\alpha|\beta, \mathfrak{D}, r) \propto p(\alpha, \beta, \mathfrak{D}|r)$ and $p(\Psi|\alpha, \mathfrak{D}, r) \propto p(\alpha, \beta, \mathfrak{D}|r)$. Using the same procedures as for part (i), the marginal conditional distributions of α and Ψ are obtained. □

Appendix B: Proof of Proposition 3

The proof for $p(\Gamma|\Omega) = 1$ is presented in the appendix of Villani (2001a, Theorem 5.2). When $p(\Gamma|\Omega)$ is given by equation (7), the case $r = 0$ is determined directly from (A.4) by letting $\Gamma_r(p) = \Gamma_r(r) = \nu_r = 1$ and $\alpha = \beta = 0$.

For the case $r = p$, we start from (A.4) with $\Pi = \alpha\beta'$ and $\nu_p = 1$. Hence,

$$\begin{aligned} p(\Pi, \mathfrak{D}|r = p) &= \pi^{-(T-d)p/2} \pi^{-p^2/2} \lambda_\alpha^{-p^2} |A|^{q/2} |\Sigma_\Gamma|^{-p/2} \frac{\Gamma_p(T+q+p-d)}{\Gamma_p(q)} \\ &\quad \times |DD'|^{-p/2} |Z_2 M_D Z_2' + \Sigma_\Gamma^{-1}|^{-p/2} \Big|A + \lambda_\alpha^{-2}\Pi\Pi' + WNW'\Big|^{-(T+q+p-d)/2}, \end{aligned} \tag{B.1}$$

where $W = Z_0 - \Pi Z_1$. It is straightforward to show that

$$A + \lambda_\alpha^{-2}\Pi\Pi' + WNW' = S + \big(\Pi' - \hat{\Pi}'\big)' C_1 \big(\Pi' - \hat{\Pi}'\big),$$

where S, C_1, and $\hat{\Pi}$ are defined in Proposition 2. Integration of (B.1) with respect to Π' can again be achieved by utilizing the properties of the matrix t density. This gives the claimed result. □

The full posterior density under the $\Gamma|\Omega$ prior in (7) is proportional to (A.1). The log posterior density can therefore, apart from a constant, be expressed as:

$$\begin{aligned}\ln p\big(\alpha,\beta,\Phi,\Gamma,\Omega|\mathfrak{D},r\big) &= -\frac{T+p+q+r+m+1}{2}\ln\big(|\Omega|\big)\\ &\quad -\frac{1}{2}\operatorname{tr}\big[\Omega^{-1}\big(\varepsilon\varepsilon'+A+\lambda_\alpha^{-2}\alpha\beta'\beta\alpha'+\Gamma\Sigma_\Gamma^{-1}\Gamma'\big)\big].\end{aligned}\tag{C.1}$$

The posterior mode is obtained by maximizing (C.1) with respect to $\alpha,\beta,\Phi,\Gamma,\Omega$. The partial derivatives of (C.1) with respect to $\operatorname{vec}(\Phi)$ and $\operatorname{vec}(\Gamma)$ are given by (see, e.g., Magnus and Neudecker, 1988, for results on matrix differentials and partial derivatives):

$$\begin{aligned}\frac{\partial\ln p\big(\alpha,\beta,\Phi,\Gamma,\Omega|\mathfrak{D},r\big)}{\partial\operatorname{vec}(\Phi)'} &= \operatorname{vec}\big(\varepsilon D'\big)'\big[I_d\otimes\Omega^{-1}\big],\\ \frac{\partial\ln p\big(\alpha,\beta,\Phi,\Gamma,\Omega|\mathfrak{D},r\big)}{\partial\operatorname{vec}(\Gamma)'} &= \operatorname{vec}\big(\varepsilon Z_2'-\Gamma\Sigma_\Gamma^{-1}\big)'\big[I_m\otimes\Omega^{-1}\big].\end{aligned}$$

Setting these two equations to zero and solving for Φ and Γ as functions of (α,β) we get equations (29)–(30). Substituting these functions back into the expression for ε we find that:

$$Z_0-\Phi(\alpha,\beta)D-\Gamma(\alpha,\beta)Z_2-\alpha\beta'Z_1 = Z_0N-\alpha\beta'Z_1N,$$

where the $T\times T$ matrix N is defined in Proposition 2. Moreover, substituting $\Gamma(\alpha,\beta)$ for Γ in $\Gamma\Sigma_\Gamma^{-1}\Gamma'$ we get $(Z_0-\alpha\beta'Z_1)Q(Z_0-\alpha\beta'Z_1)'$, where

$$Q = M_DZ_2'\big(Z_2M_DZ_2'+\Sigma_\Gamma^{-1}\big)^{-1}\Sigma_\Gamma^{-1}\big(Z_2M_DZ_2'+\Sigma_\Gamma^{-1}\big)^{-1}Z_2M_D.$$

After some algebra it follows that $NN'+Q=N$. Hence, the concentrated log posterior density can, apart from a constant, be expressed as:

$$\begin{aligned}\ln p\big(\alpha,\beta,\Omega|\mathfrak{D},r\big) &= -\frac{T+p+q+r+m+1}{2}\ln\big(|\Omega|\big)\\ &\quad -\frac{1}{2}\operatorname{tr}\big[\Omega^{-1}\big((Z_0-\alpha\beta'Z_1)N(Z_0-\alpha\beta'Z_1)'+A+\lambda_\alpha^{-2}\alpha\beta'\beta\alpha'\big)\big].\end{aligned}\tag{C.2}$$

The partial derivatives with respect to α and Ω are now:

$$\begin{aligned}\frac{\partial\ln p\big(\alpha,\beta,\Omega|\mathfrak{D},r\big)}{\partial\operatorname{vec}(\alpha)'} &= \operatorname{vec}\big((Z_0-\alpha\beta'Z_1)NZ_1'\beta-\lambda_\alpha^{-2}\alpha\beta'\beta\big)'\big[I_r\otimes\Omega^{-1}\big],\\ \frac{\partial\ln p\big(\alpha,\beta,\Omega|\mathfrak{D},r\big)}{\partial\operatorname{vec}(\Omega)'} &= \frac{1}{2}\operatorname{vec}\big(A+\lambda_\alpha^{-2}\alpha\beta'\beta\alpha'+(Z_0-\alpha\beta'Z_1)N(Z_0-\alpha\beta'Z_1)'\\ &\quad -(T+p+q+r+m+1)\Omega\big)'\big[\Omega^{-1}\otimes\Omega^{-1}\big].\end{aligned}$$

Setting these two equations to zero and solving for α and Ω as functions of β we get

$$\alpha(\beta) = S_{01}\beta\big(\beta'S_{11}\beta\big)^{-1},\tag{C.3}$$

while $\Omega(\beta)$ is given by (28). Substituting these expression back into (C.2) we find that the concentrated log posterior density, apart from a constant, is:

$$\ln p\big(\beta|\mathfrak{D},r\big) = -\frac{T+p+q+r+m+1}{2}\ln\Big|S_{00}-S_{01}\beta\big(\beta'S_{11}\beta\big)^{-1}\beta'S_{10}\Big|.\tag{C.4}$$

The choice of β which maximizes this function, denoted by $\bar\beta_{\mathrm{PM}}$, is given by Johansen (1996, Theorem 6.1). The posterior mode cointegration space estimator in (24) implicitly uses the $r(r+1)/2$ identifying restrictions $\bar\beta_{\mathrm{PM}}'S_{11}\bar\beta_{\mathrm{PM}}=I_r$ and the further $r(r-1)/2$ restrictions $\bar\beta_{\mathrm{PM}}'S_{10}S_{00}^{-1}S_{01}\bar\beta_{\mathrm{PM}}=\Lambda=\operatorname{diag}(\lambda_1,\ldots,\lambda_r)$. We therefore find that

$$\begin{aligned}\ln p\big(\bar\beta_{\mathrm{PM}}|\mathfrak{D},r\big) &= -\frac{T+p+q+r+m+1}{2}\ln|S_{00}||I_p-S_{00}^{-1}S_{01}\bar\beta_{\mathrm{PM}}\bar\beta_{\mathrm{PM}}'S_{10}|\\ &= -\frac{T+p+q+r+m+1}{2}\ln|S_{00}||I_r-\bar\beta_{\mathrm{PM}}'S_{10}S_{00}^{-1}S_{01}\bar\beta_{\mathrm{PM}}|\\ &= -\frac{T+p+q+r+m+1}{2}\ln|S_{00}||I_r-\Lambda|.\end{aligned}$$

The posterior mode estimators of α_c and Ψ follow directly from their relations to β.

The posterior density $p(\alpha, \Psi|\mathfrak{D}, r)$ is proportional to equation (A.4). Hence, apart from a constant the log posterior density can be expressed as:

$$\ln p\big(\alpha, \beta|\mathfrak{D}, r\big) = -\frac{T+q+r-p(n-1)-d}{2} \ln \Big|(Z_0 - \alpha\beta' Z_1)N(Z_0 - \alpha\beta' Z_1)' + \lambda_\alpha^{-2}\alpha\beta'\beta\alpha' + A\Big|, \quad \text{(C.5)}$$

where the integer n is defined in Proposition 2. The partial derivative with respect to α is here given by

$$\begin{aligned}\frac{\partial \ln p\big(\alpha, \beta|\mathfrak{D}, r\big)}{\partial \mathrm{vec}(\alpha)'} &= [T+q+r-p(n-1)-d]\mathrm{vec}\big((Z_0 - \alpha\beta' Z_1)NZ_1'\beta - \lambda_\alpha^{-2}\alpha\beta'\beta\big)' \\ &\quad \times \Big[I_r \otimes \big((Z_0 - \alpha\beta' Z_1)N(Z_0 - \alpha\beta' Z_1)' + \lambda_\alpha^{-2}\alpha\beta'\beta\alpha' + A\big)^{-1}\Big].\end{aligned}$$

Setting this equation to zero and solving for α as a function of β we obtain (C.3). Substituting this back into (C.5) we get, apart from a constant,

$$\ln p\big(\beta|\mathfrak{D}, r\big) = -\frac{T+q+r-p(n-1)-d}{2} \ln \Big|S_{00} - S_{01}\beta\big(\beta' S_{11}\beta\big)^{-1}\beta' S_{10}\Big|. \quad \text{(C.6)}$$

This function is maximized for the same choice of β as the function in (C.4). Hence, the posterior mode estimators of α_c and Ψ are identical to those for the full posterior density. □

Appendix D: Proof of Proposition 5

Consider the case when $p(\Gamma|\Omega)$ is given by equation (7). From equation (A.1) we know that

$$\begin{aligned}p\big(\alpha, \beta, \Phi, \Gamma, \Omega|\mathfrak{D}, r\big) &\propto |\Omega|^{-(T+p+q+r+m+1)/2} \\ &\quad \times \exp\left(-\frac{1}{2}\mathrm{tr}\big[\Omega^{-1}\big(\varepsilon\varepsilon' + A + \lambda_\alpha^{-2}\alpha\beta'\beta\alpha' + \Gamma\Sigma_\Gamma^{-1}\Gamma'\big)\big]\right),\end{aligned} \quad \text{(D.1)}$$

where $m = p(k-1)$. For $X = Z_0 - \Gamma Z_2 - \alpha\beta' Z_1$ we have that

$$\varepsilon\varepsilon' = XM_D X' + \big(\Phi - \hat{\Phi}\big)DD'\big(\Phi - \hat{\Phi}\big)',$$

where $\hat{\Phi} = XD'(DD')^{-1}$ and M_D is defined in Proposition 2. Integrating the right hand side of (D.1) with respect to Φ can be achieved by utilizing the definition of the density for the matricvariate normal distribution and the relationship between the trace and the vec operator. This yields

$$\begin{aligned}p\big(\alpha, \beta, \Gamma, \Omega|\mathfrak{D}, r\big) &\propto |\Omega|^{-(T+p+q+r+m+1-d)/2} \\ &\quad \times \exp\left(-\frac{1}{2}\mathrm{tr}\big[\Omega^{-1}\big(XM_D X' + A + \lambda_\alpha^{-2}\alpha\beta'\beta\alpha' + \Gamma\Sigma_\Gamma^{-1}\Gamma'\big)\big]\right).\end{aligned} \quad \text{(D.2)}$$

Next we note that for $W = Z_0 - \alpha\beta' Z_1$

$$\Gamma\Sigma_\Gamma^{-1}\Gamma' + XM_D X' = WNW' + \big(\Gamma - \hat{\Gamma}\big)\big[Z_2 M_D Z_2' + \Sigma_\Gamma^{-1}\big]\big(\Gamma - \hat{\Gamma}\big)',$$

where $\hat{\Gamma} = WM_D Z_2'[Z_2 M_D Z_2' + \Sigma_\Gamma^{-1}]^{-1}$ and N is defined in Proposition 2. Integrating the right hand side of (D.2) with respect to Γ is again made simple by using the definition of the matricvariate normal density. This provides us with

$$\begin{aligned}p\big(\alpha, \beta, \Omega|\mathfrak{D}, r\big) &\propto |\Omega|^{-(T+p+q+r+m+1-d-p(k-1))/2} \\ &\quad \times \exp\left(-\frac{1}{2}\mathrm{tr}\big[\Omega^{-1}\big(WNW' + A + \lambda_\alpha^{-2}\alpha\beta'\beta\alpha'\big)\big]\right).\end{aligned} \quad \text{(D.3)}$$

Since $p(\Omega|\alpha, \beta, \mathfrak{D}, r) \propto p(\alpha, \beta, \Omega|\mathfrak{D}, r)$ and we recognise the kernel of the inverted Wishart density, the result (34) follows directly. Similarly, $p(\alpha|\beta, \Omega, \mathfrak{D}, r)$ and $p(\psi|\alpha, \Omega, \mathfrak{D}, r)$ are proportional to the second term on the right hand side of (D.3). By similar matrix manipulations on quadratic forms to those above and recognizing the kernel of the matricvariate normal density, the results (35) and (36) can be shown to hold. When $p(\Gamma|\Omega) = 1$ we can simply set $m = \Sigma_\Gamma^{-1} = 0$ and perform similar derivations to those above for $Z = [D' \;\; Z_2']'$. □

References

Ahn, S. K., and Reinsel, G. C. (1990). "Estimation for Partially Non-Stationary Multivariate Autoregressive Processes." *Journal of the American Statistical Association*, *85*, 813–823.

Akaike, H. (1974). "A New Look at the Statistical Model Identification." *IEEE Transactions on Automatic Control*, *AC-19*, 716–723.

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

Bauwens, L., and Giot, A. (1998). "A Gibbs Sampling Approach to Cointegration." *Computational Statistics*, *13*, 339–368.

Bauwens, L., and Lubrano, M. (1996). "Identification Restrictions and Posterior Densities in Cointegrated VAR Systems." In T. B. Fomby (Ed.), *Advances in Econometrics, Volume 11, Part B* (pp. 3–28). London: JAI Press.

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

Bauwens, L., and Richard, J.-F. (1985). "A 1-1 Poly-*t* Random Variable Generator with Application to Monte Carlo Integration." *Journal of Econometrics*, *29*, 19–46.

Bauwens, L., and van Dijk, H. K. (1990). "Bayesian Limited Information Analysis Revisited." In J. J. Gabszewicz, J.-F. Richard, and L. Wolsey (Eds.), *Economic decision-making: Games, econometrics and optimisation* (pp. 385–424). Amsterdam: North-Holland.

Bernardo, J. M., and Ramón, J. M. (1998). "An Introduction to Bayesian Reference Analysis: Inference on the Ratio of Multinomial Parameters." *The Statistician*, *47*, 101–135.

Box, G. E. P., and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis.* Reading, Massachusetts: Addison-Wesley.

Brand, C., and Cassola, N. (2000). *A Money Demand System for the Euro Area.* (ECB Working Paper No. 39)

Bruggeman, A., Donati, P., and Warne, A. (2003). *Is the Demand for Euro Area M3 Stable?* (ECB Working Paper No. 255.)

Calza, A., Gerdesmeier, D., and Levy, J. (2001). *Euro Area Money Demand: Measuring the Opportunity Costs Appropriately.* (IMF Working Paper No. 01/179)

Carstensen, K. (2003). *Is European Money Demand Still Stable?* (Kiel Working Paper No. 1179, Kiel Institute for World Economics, Kiel, Germany)

Carstensen, K. (2004). *Stock Market Downswing and the Stability of EMU Money Demand.* (Manuscript, Kiel Institute for World Economics, Kiel, Germany)

Casella, G., and George, E. I. (1992). "Explaining the Gibbs Sampler." *The American Statistician*, *46*, 167–174.

Chib, S. (1995). "Marginal Likelihood from the Gibbs Output." *Journal of the American Statistical Association*, *90*, 1313–1321.

Coenen, G., and Vega, J. L. (2001). "The Demand for M3 in the Euro Area." *Journal of Applied Econometrics*, *16*, 727–748.

Corander, J., and Villani, M. (2004). "Bayesian Assessment of Dimensionality in Multivariate Reduced Rank Regression." *Statistica Neerlandica*, *58*, 255–270.

Draper, D. (1995). "Assessment and Propagation of Model Uncertainty." *Journal of the Royal Statistical Society B*, *57*, 45–97.

Drèze, J. H. (1977). "Bayesian Regression Analyses Using Poly-*t* Densities." *Journal of Econometrics*, *6*, 329–354.

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

European Central Bank. (2004). "Monetary Analysis in Real Time." *Monthly Bulletin*. (October)

Geweke, J. (1996). "Bayesian Reduced Rank Regression in Econometrics." *Journal of Econometrics*, *75*, 121–146.

Geweke, J. (1999). "Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communication." *Econometric Reviews*, *18*, 1–73.

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

James, A. T. (1954). "Normal Multivariate Analysis and the Orthogonal Group." *Annals of Mathematical Statistics*, *25*, 40–74.

Johansen, S. (1988). "Statistical Analysis of Cointegration Vectors." *Journal of Economic Dynamics and Control*, *12*, 231–254.

Johansen, S. (1991). "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models." *Econometrica*, *59*, 1551–1580.

Johansen, S. (1996). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models* (2nd ed.). Oxford: Oxford University Press.

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

Johansen, S., and Juselius, K. (1992). "Testing Structural Hypoitheses in a Multivariate Cointegration Analysis of the PPP and UIP for the UK." *Journal of Econometrics*, *53*, 211–244.

Kadiyala, K. R., and Karlsson, S. (1997). "Numerical Methods for Estimation and Inference in Bayesian VAR-Models." *Journal of Applied Econometrics*, *12*, 99–132.

Kleibergen, F., and Paap, R. (2002). "Priors, Posteriors and Bayes Factors for a Bayesian Analysis of Cointegration." *Journal of Econometrics*, *111*, 223–249.

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

Kontolemis, Z. G. (2002). *Money Demand in the Euro Area: Where Do We Stand (Today)?* (IMF Working Paper No. 02/185)

Koop, G., Leon-Gonzalez, R., and Strachan, R. (2006). *Bayesian Inference in a Cointegrating Panel Data Model.* (Department of Economics, University of Leicester, Working Paper No. 06/2)

Koop, G., Strachan, R. W., van Dijk, H. K., and Villani, M. (2006). "Bayesian Approaches to Cointegration." In T. C. Mills and K. Patterson (Eds.), *Palgrave Handbook on Econometrics: Volume 1 Econometric Theory* (pp. 871–898). Basingstoke: Palgrave MacMillan.

Lindley, D. V. (1957). "A Statistical Paradox." *Biometrika*, *44*, 187–192.

Litterman, R. B. (1986). "Forecasting with Bayesian Vector Autoregressions — Five Years of Experience." *Journal of Business & Economic Statistics*, *4*, 25–38.

Magnus, J., and Neudecker, H. (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics.* Chichester: John Wiley.

Newey, W. K., and West, K. D. (1987). "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica*, *55*, 703–708.

O'Hagan, A. (1995). "Fractional Bayes Factors for Model Comparisons." *Journal of the Royal Statistical Society*, *B57*, 99–138.

O'Hagan, A. (1997). "Properties of Intrinsic and Fractional Bayes Factors." *Test*, *6*, 101–118.

Park, J. Y. (1992). "Canonical Cointegrating Regressions." *Econometrica*, *60*, 119–143.

Pesaran, M. H., and Shin, Y. (2002). "Long-Run Structural Modelling." *Econometric Reviews*, *21*, 49–87.

Phillips, P. C. B. (1989). "Spherical Matrix Distributions and Cauchy Quotients." *Statistics & Probability Letters*, *8*, 51–53.

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

Phillips, P. C. B., and Hansen, B. E. (1990). "Statistical Inference in Instrumental Variables Regression with I(1) Processes." *Review of Economic Studies*, *57*, 99–125.

Phillips, P. C. B., and Loretan, M. (1991). "Estimating Long-Run Economic Equilibria." *Review of Economic Studies*, *58*, 407–436.

Quinn, B. G. (1980). "Order Determination for a Multivariate Autoregression." *Journal of the Royal Statistical Society*, *B42*, 182–185.

Reinsel, G. C., and Ahn, S. K. (1992). "Vector Autoregressive Models with Unit Roots and Reduced Rank Structure: Estimation, Likelihood Ratio Test, and Forecasting." *Journal of Time Series Analysis*, *13*, 353–375.

Reynard, S. (2004). "Financial Market Participation and the Apparent Instability of Money Demand." *Journal of Monetary Economics*, *51*, 1297–1317.

Reynard, S. (2005). *Money and the Great Disinflation.* (Manuscript, Swiss National Bank, Zürich, Switzerland)

Robertson, J. C., and Tallman, E. W. (1999). "Vector Autoregressions: Forecasting and Reality." *Federal Reserve Bank of Atlanta Economic Review*, *84*, 4–18.

Saikkonen, P. (1991). "Asymptotically Efficient Estimation of Cointegrating Vectors." *Econometric Theory*, *7*, 1–21.

Schwarz, G. (1978). "Estimating the Dimension of a Model." *Annals of Statistics*, *6*, 461–464.

Sims, C. A., and Zha, T. (1998). "Bayesian Methods for Dynamic Multivariate Models." *International Economic Review*, *39*, 949–968.

Stock, J. H., and Watson, M. W. (1993). "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems." *Econometrica*, *61*, 783–820.

Strachan, R. W. (2003). "Valid Bayesian Estimation of the Cointegrating Error Correction Model." *Journal of Business & Economic Statistics*, *21*, 185–195.

Strachan, R. W., and Inder, B. (2004). "Bayesian Analysis of the Error Correction Model." *Journal of Econometrics*, *123*, 307–325.

Strachan, R. W., and van Dijk, H. K. (2003). "Bayesian Model Selection with an Uninformative Prior." *Oxford Bulletin of Economics and Statistics*, *65*, 863–876.

Strachan, R. W., and van Dijk, H. K. (2004). *Valuing Structure, Model Uncertainty and Model Averaging in Vector Autoregressive Processes.* (Econometric Institute Report EI 2004–23, Erasmus University Rotterdam)

Strachan, R. W., and van Dijk, H. K. (2005). *Improper Priors with Well Defined Bayes Factors.* (Department of Economics, University of Leicester, Working Paper No. 05/4)

Tierney, L. (1994). "Markov Chains for Exploring Posterior Distributions." *Annals of Statistics*, *22*, 1701–1762. (With discussion.)

Villani, M. (2000). *Aspects of Bayesian Cointegration.* PhD Thesis, Stockholm University, Sweden.

Villani, M. (2001a). *Bayesian Reference Analysis of Cointegration.* (Research Report No. 2001:1, Department of Statistics, Stockholm University)

Villani, M. (2001b). "Fractional Bayesian Lag Length Inference in Multivariate Autoregressive Processes." *Journal of Time Series Analysis*, *22*, 67–86.

Villani, M. (2005a). *Bayesian Inference of General Linear Restrictions on the Cointegration Space.* (Sveriges Riksbank Working Paper Series No. 189)

Villani, M. (2005b). "Bayesian Reference Analysis of Cointegration." *Econometric Theory*, *21*, 326–357.

Villani, M. (2005c). *Inference in Vector Autoregressive Models with an Informative Prior on the Steady State.* (Sveriges Riksbank Working Paper Series No. 181)

Villani, M., and Warne, A. (2003). *Monetary Policy Analysis in a Small Open Economy Using Bayesian Cointegrated Structural VARs.* (ECB Working Paper No. 296)

Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics.* New York: John Wiley.

European Central Bank Working Paper Series

For a complete list of Working Papers published by the ECB, please visit the ECB's website (http://www.ecb.int)

665 "The euro as invoicing currency in international trade" by A. Kamps, August 2006.

666 "Quantifying the impact of structural reforms" by E. Ernst, G. Gong, W. Semmler and L. Bukeviciute, August 2006.

667 "The behaviour of the real exchange rate: evidence from regression quantiles" by K. Nikolaou, August 2006.

668 "Declining valuations and equilibrium bidding in central bank refinancing operations" by C. Ewerhart, N. Cassola and N. Valla, August 2006.

669 "Regular adjustment: theory and evidence" by J. D. Konieczny and F. Rumler, August 2006.

670 "The importance of being mature: the effect of demographic maturation on global per-capita GDP" by R. Gómez and P. Hernández de Cos, August 2006.

671 "Business cycle synchronisation in East Asia" by F. Moneta and R. Rüffer, August 2006.

672 "Understanding inflation persistence: a comparison of different models" by H. Dixon and E. Kara, September 2006.

673 "Optimal monetary policy in the generalized Taylor economy" by E. Kara, September 2006.

674 "A quasi maximum likelihood approach for large approximate dynamic factor models" by C. Doz, D. Giannone and L. Reichlin, September 2006.

675 "Expansionary fiscal consolidations in Europe: new evidence" by A. Afonso, September 2006.

676 "The distribution of contract durations across firms: a unified framework for understanding and comparing dynamic wage and price setting models" by H. Dixon, September 2006.

677 "What drives EU banks' stock returns? Bank-level evidence using the dynamic dividend-discount model" by O. Castrén, T. Fitzpatrick and M. Sydow, September 2006.

678 "The geography of international portfolio flows, international CAPM and the role of monetary policy frameworks" by R. A. De Santis, September 2006.

679 "Monetary policy in the media" by H. Berger, M. Ehrmann and M. Fratzscher, September 2006.

680 "Comparing alternative predictors based on large-panel factor models" by A. D'Agostino and D. Giannone, October 2006.

681 "Regional inflation dynamics within and across euro area countries and a comparison with the US" by G. W. Beck, K. Hubrich and M. Marcellino, October 2006.

682 "Is reversion to PPP in euro exchange rates non-linear?" by B. Schnatz, October 2006.

683 "Financial integration of new EU Member States" by L. Cappiello, B. Gérard, A. Kadareja and S. Manganelli, October 2006.

684 "Inflation dynamics and regime shifts" by J. Lendvai, October 2006.

685 "Home bias in global bond and equity markets: the role of real exchange rate volatility" by M. Fidora, M. Fratzscher and C. Thimann, October 2006

686 "Stale information, shocks and volatility" by R. Gropp and A. Kadareja, October 2006.

687 "Credit growth in Central and Eastern Europe: new (over)shooting stars?" by B. Égert, P. Backé and T. Zumer, October 2006.

688 "Determinants of workers' remittances: evidence from the European Neighbouring Region" by I. Schiopu and N. Siegfried, October 2006.

689 "The effect of financial development on the investment-cash flow relationship: cross-country evidence from Europe" by B. Becker and J. Sivadasan, October 2006.

690 "Optimal simple monetary policy rules and non-atomistic wage setters in a New-Keynesian framework" by S. Gnocchi, October 2006.

691 "The yield curve as a predictor and emerging economies" by A. Mehl, November 2006.

692 "Bayesian inference in cointegrated VAR models: with applications to the demand for euro area M3" by A. Warne, November 2006.

ISSN 1561081-0
9 771561 081005