European University Institute

DEPARTMENT OF ECONOMICS

EUI Working Papers

ECO 2011/14
DEPARTMENT OF ECONOMICS

ADAPTING THE LITTERMAN PRIOR FOR COINTEGRATED VARs

Michal Markun

EUROPEAN UNIVERSITY INSTITUTE, FLORENCE
DEPARTMENT OF ECONOMICS

Adapting the Litterman Prior for Cointegrated VARs

MICHAL MARKUN

EUI Working Paper **ECO** 2011/14

ISSN 1725-6704

Printed in Italy
European University Institute
Badia Fiesolana
I – 50014 San Domenico di Fiesole (FI)
Italy
www.eui.eu
cadmus.eui.eu

Adapting the Litterman prior for cointegrated VARs

Michał Markun*

Department of Economics, European University Institute,

Via della Piazzuola 43, I-50133 Firenze, Italy.

e-mail: michal.markun@eui.eu

Abstract

The paper presents a novel prior for Bayesian VAR models, characterized by explicit modelling of cointegration that avoids certain unattractive restrictive properties of the priors used previously. The potential of the prior for easy elicitation from the well-established Litterman beliefs is demonstrated. An efficient procedure for sampling from posterior distribution is provided. The favourable outcome of the forecast comparison exercise gives further support for the use of the methods proposed.

1 Introduction

The vector autoregressive model (VAR), persuasively advocated by Sims[1980] some 30 years ago, has been widely adopted, and together with its various descendants such as mixed-frequency VARs, VARs for large data-sets, VARs with time-varying parameters or structural VARs, extensively used by the economics profession ever since. It provides a flexible description of dynamic interactions in a multivariate system, yet using prior information very sparsely – only the variables that enter the system and the lag length need to be specified. However, the sufficient representation of the data-generating

*I would like to thank Helmut Lütkepohl, Rodney Strachan, Domenico Giannone, Karl Schlag, Marco Lombardi and Piotr Jelonek for their comments, and the participants of the EUI Time Series Econometrics Working Group, DIW Macroeconometric Workshop 2010 and EIEF Second Carlo Giannini Ph.D. Workshop in Econometrics where this work has been presented. The article is part of the author's Ph.D. thesis.

process may necessitate many variables or high lag orders, which causes the intertemporal relations to become more difficult to capture. This so-called *curse of dimensionality* is apparent in the poor forecast performance of such models, for example. Thus, the researcher faces a difficult trade-off and may ultimately be led to restrict the model. The classical zero restrictions, however, are extreme and completely silence the data in some dimensions. It may be therefore preferable to adopt a Bayesian solution, which is more flexible, by imposing a non-degenerate probability distributions for the parameters (the so-called *prior distribution*). The Bayesian framework possesses an additional advantage in that it allows one to make exact small-sample probabilistic statements about parameters, or to incorporate uncertainty deriving from both unknown future disturbances hitting the system and unknown parameter values into forecasts.

The problem with this approach, often prohibitive, is the choice of prior. Thus, an important moment in the development of Bayesian VARs was the contribution of Litterman[1979], where one of the earliest priors was proposed, also called *the Minnesota prior* (see also Doan, Litterman and Sims[1986] and Litterman[1986] for useful exposition). Owing to its appeal as an easy and accurate expression of prior beliefs about the nature of macroeconomic variables and accuracy in predicting them, it gained considerable attention and became a benchmark in literature. However, its deficiencies came to light soon. It is characterized by restrictive treatment of the residual covariance matrix – it is assumed to be diagonal and fixed. Moreover, it is essentially an univariate modelling framework with no cross-equation parameter restrictions. As is well-known, at least one such type of restriction, namely cointegration, plays an important role in econometric research. Therefore, there have been subsequent developments aimed at extending this basic prior. Firstly by Kadiyala and Karlsson[1994, 1997] to allow cross-equation dependencies. Then by Sims, Waggoner and Zha in various papers written by the authors over many years in various combinations (see Sims and Zha[1998], Waggoner and Zha[2003] and the most recent Sims, Waggoner and Zha[2008]). Finally by various authors that proposed priors for cointegrated systems, reviewed in Koop, Strachan, van Dijk and Villani[2006]).

The priors for VAR models do not model cointegration restriction explicitly: the original Litterman prior rules it out, the priors by Kadiyala and Karlsson[1997] merely 'allow for cointegration', and the prior by Sims, Waggoner and Zha at most 'favour cointegration'. Cointegration in VAR systems, or rank reduction of the long-run impact matrix in the corresponding VECM reparametrization, is a functional restriction on its parameters and therefore special methods are needed to implement Bayesian analysis – see Koop *et al.*[2006] for an overview and a summary. One problem is the resulting nonlinearity of the model, causing problems for posterior inference. Another problem analysed in econometric literature is the choice of prior on the matrix of cointegrating vectors – as it is known such a matrix is not in general identified and two methods of identification have been proposed: linear and nonordinal. The basis for the present study paper in the first line of research is Villani[2005], extended and applied in Warne[2006] and Villani and Warne[2003], who was the first to propose a satisfactory noninformative prior for cointegrated systems and to present a Gibbs sampler to simulate the posterior distribution of the parameters. In the second line of research it is Koop, León-González and Strachan[2010], extending and related to *inter alia* Strachan[2003], Strachan and Inder[2004], Strachan and van Dijk[2003], who present a simple and efficient algorithm to work with nonordinal normalization.

The present study draws from these two strands of research and proposes a new prior, an elicitation method and a posterior sampler. The new prior is motivated by the restrictiveness of the previously-used priors. The prior by Villani[2005] displays a certain well-known symmetry property unattractive both from theoretical and practical viewpoint: a Kronecker-type covariance matrix of dynamic adjustment coefficients. It is widely acknowledged in literature that it makes it too restrictive a framework for modelling, and in the context of the VAR model this led to favouring the normal–diffuse over the normal–Wishart or diffuse one (Kadiyala and Karlsson[1997]). Furthermore, in the Villani[2005] prior I will show another restriction on the covariance matrix of the long-run impact matrix, so far unnoticed in literature, and stress its incompatibility with the Litterman beliefs. I will show how it can be relaxed to allow easy elicitation from the Litterman beliefs. After all, a VECM model is just a reparametrization plus a restriction on the VAR model — why, then, while constructing a prior should we forget about the Litterman prior for VAR and construct some prior for VECM anew? With the new prior, researchers no longer have to renounce their Litterman beliefs, which are well-entrenched in the literature, but instead they can incorporate them easily in an efficient framework for inference about cointegrated systems. To bridge the gap between the Litterman/normal–diffuse priors and the Villani[2005] prior, the general idea is to couple a suitable prior for the long-run matrix with the normal–diffuse prior on short-run dynamics and stochastic structure parameters. As it proves useful to work with nonordinal normalization, Koop *et al.*[2010] paper is of key importance for the inference, as it provides an efficient Gibbs algorithm for drawing samples from posterior distribution, which makes the framework suitable for routine use.

The paper is structured as follows. In the next section I will introduce the VECM model. In section 3 I first present relevant existing prior and posterior distributions for VAR and VECM models. The main part of this paper, subsection 3.3, introduces the new prior, discusses its properties and presents the sampling algorithm. The empirical example gives further support for the use of the new framework by demonstrating forecast accuracy improvements over a range of benchmarks. The final section concludes the paper. There are two appendices with formal proofs and a third presenting additional results for the empirical example.

2 The model

Our aim is to describe the dynamic behaviour of K endogenous variables over a period of T points in time: y_{kt} for $k = 1, \dots, K$ and $t = 1, \dots, T$. The standard VAR model reads

$$y_t = d_t D + \sum_{p=1}^{P} y_{t-p} A_p + e_t \quad \text{and} \quad e_t' \overset{iid}{\sim} N(0, \Psi), \tag{1}$$

for $t = 1, \dots, T$, where y_t is a $(1 \times K)$ vector of endogenous variables, d_t is a $(1 \times K_d)$ vector of deterministic variables, e_t is a $(1 \times K)$ vector of residual errors, P is the lag order – it is assumed known and we also assume that P initial observations (prior to time $t = 1$) are available, the covariance matrix Ψ is $(K \times K)$ symmetric positive-definite (SPD). In modelling cointegrated systems, however, the following VECM reparametrization of (1) is used

$$\Delta y_t = d_t D + y_{t-1}\Pi + \sum_{p=1}^{P-1} \Delta y_{t-p}\Gamma_p + e_t \quad \text{and} \quad e_t' \overset{iid}{\sim} N(0, \Psi). \tag{2}$$

As is well-known, cointegration is equivalent to a rank reduction of matrix Π, expressed therefore as $\Pi = \beta\alpha'$ where α and β are $(K \times r)$ matrices of full column rank $(r \le K)$.

Recall the relationship between the parameters $D, A_1, \dots, A_P$ of the VAR model and $D, \Pi, \Gamma_1, \dots, \Gamma_{P-1}$ of the VECM model. The formulas below will be used for eliciting the hyperparameters for the proposed prior of the VECM model (2) from Litterman beliefs expressed in terms of the VAR model. Note first that the parameter D is unchanged, and in case of rank reduction $\Pi = \beta\alpha'$. Then to go from (1) to (2)

$$\begin{aligned} \Pi &= A_1 + \ldots + A_P - I_K \\ \Gamma_p &= -(A_{p+1} + \ldots + A_P), \qquad p = 1, \dots, P-1. \end{aligned} \tag{3}$$

3 Prior–Posterior pairs

In this section priors for the VAR model will be briefly reviewed. Next, details on the Villani[2005] and Koop *et al.*[2010] priors will be given. In particular, I will discuss in detail the two restrictive features that make them inconsistent with the Litterman beliefs. This provides a motivation for the next subsection where the new prior will finally be proposed, the elicitation procedure will be presented and the Gibbs sampler for posterior inference in the new framework will be provided.

3.1 Litterman and normal–diffuse priors

The construction starts from equation-wise representation of the VAR model (1) as follows

$$y_k = X\gamma_k + e_k \quad \text{and} \quad e_k \sim N(0, \psi_{kk} I_T), \tag{4}$$

where now $y_k = [y_{k1}, \ldots, y_{kT}]'$ is a vector of observations on the kth variable throughout the whole observation period, X is built of x_ts by stacking, where $x_t = [d_t, y_{t-1}, \ldots, y_{t-P}]$, γ_k is the kth column of $\Gamma = [D', A_1', \ldots, A_P']'$ and contains parameters on explanatory variables in the model for the kth variable.

The belief underlying the construction of the Litterman prior is that the variables in the system are random walks without deterministic terms; moreover, it assumes independence between error terms of different equations, residual standard error known, as well as no interrelationships between equations' parameters. The beliefs about parameter values are supposed to be less vague the higher lag parameter matrix they refer to. For a given lag, they are more precise about the mean if they refer to variables other than the variable whose regression equation is being considered. The beliefs described above are operationalized by assuming dogmatic zero restrictions on the covariance matrix $\Psi = \{\psi_{ij}\}$, namely that $\psi_{ij} = 0$ for $i \neq j$ while for $i = j = k$ it assumes $\psi_{kk} = s_k^2$ – residual standard error of a P-lag univariate AR for variable k including an intercept, and by assuming normal distributions on parameters. Thus,

$$\gamma_k \sim N(\tilde{\mu}_k, \tilde{\Sigma}_k) \text{ for } k = 1, \ldots, K, \quad \text{and } \gamma_{k_1} \perp \gamma_{k_2} \text{ if } k_1 \neq k_2. \tag{5}$$

The mean and the covariance matrices are given by

$$\tilde{\mu}_k = \begin{cases} 1 & \text{coef. lag 1 var. } k \\ 0 & \text{otherwise} \end{cases} \qquad \tilde{\Sigma}_k = \begin{cases} \pi_3 \sigma_k^2 & \text{det. terms} \\ \frac{\pi_2 \sigma_k^2}{p^{\pi_4} \sigma_l^2} & \text{coef. lag } p \text{ var. } l \\ \frac{\pi_1}{p^{\pi_4}} & \text{coef. lag } p \text{ var. } k \end{cases}$$

Note that $\tilde{\Sigma}_k$s are diagonal as it is rather difficult to have any prior beliefs about the possible correlation between parameters.

The prior is parametrized by four hyperparameters: π_1, π_2, π_3 and π_4, their interpretation is obvious from $\tilde{\Sigma}_k$, which facilitates their elicitation: π_1 determines the tightness of beliefs on own-variable's lags, π_2 – on other variables, π_3 determines the tightness of beliefs on deterministic terms, while π_4 the rate of beliefs tightening with increasing lag. The possibility of differing variances across the variables in the system is accommodated by including in the matrix $\tilde{\Sigma}_k$ the scale factors σ_k^2 set at s_k^2.

The application of the Bayes' theorem to model (4) with the Litterman prior (5) yields the following posterior distribution

$$(\gamma_k|y_k, X) \sim N(\bar{\mu}_k, \bar{\Sigma}_k) \text{ for } k = 1, \ldots, K, \quad \text{and } \gamma_{k_1} \perp \gamma_{k_2} \text{ if } k_1 \neq k_2,$$

where the posterior moments are $\bar{\Sigma}_k = (X'X/\psi_{kk} + \tilde{\Sigma}_k^{-1})^{-1}$ and $\bar{\mu}_k = \bar{\Sigma}_k(X'y_k/\psi_{kk} + \tilde{\Sigma}_k^{-1}\tilde{\mu}_k)$.

To motivate the need to extend the Litterman prior, let us recall its main property: it is essentially of a univariate nature that does not account for possible dependencies between the equations (arising both from correlation between error terms and from correlation between parameters of different equations). The main message of Kadiyala and Karlsson[1997] is that using a prior distribution that allows for the above-mentioned dependencies gives rise to better forecasts. Four such priors are suggested and their advantages and drawbacks are discussed both from theoretical and numerical points of view. The prior that attracted particular attention in the literature is the normal–diffuse prior (more on this issue, in the context of the VECM model, can be found in the next subsection).

For the full-information analysis the following formulation of model (1) is used

$$y = (I_K \otimes X)\gamma + e \quad \text{and} \quad e \sim N(0, \Psi \otimes I_T), \tag{6}$$

where $y = [y_1', \ldots, y_K']'$ and $\gamma = \text{vec}(\Gamma) = [\gamma_1', \ldots, \gamma_K']'$. The normal–diffuse prior allows for a non-diagonal random residual covariance matrix Ψ. It combines the Litterman beliefs on parameters of dynamic structure γ and, assuming independence between the two groups of parameters, a diffuse prior on stochastic structure Ψ. In short,

$$\gamma \sim N(\tilde{\mu}, \tilde{\Sigma}), \quad \Psi \propto |\Psi|^{-(K+1)/2} \qquad \text{and} \qquad \gamma \perp \Psi, \tag{7}$$

where Litterman construction and hyperparameters are used further to specify prior moments: stack $\tilde{\mu}_k$s into $\tilde{\mu}$ and build $\tilde{\Sigma}$ as a block-diagonal matrix with blocks $\tilde{\Sigma}_k$s. The off-block-diagonal zeros in $\tilde{\Sigma}$ express a lack of knowledge rather than a genuine belief in the parameters' independence; still, note that unlike in the case of the pure Litterman prior, the posterior does not impose zero correlation between γ_{k_1} and γ_{k_2} for $k_1 \neq k_2$ – the data, thus, is allowed to 'speak' on the matter of correlation between parameters of different equations. The prior on Ψ is improper; off-diagonal elements are not restricted to zero.

The prior is inconvenient in the sense that it does not allow us to obtain analytically posterior distributions. Posterior inference is implemented by way of a Gibbs sampler with the following full conditional posterior distributions

$$(\gamma|\Psi, y, X) \sim N(\bar{\mu}, \bar{\Sigma}) \qquad \text{and} \qquad (\Psi|\gamma, y, X) \sim iW(\bar{\Psi}, T),$$

where iW denotes the inverse Wishart distribution (see Appendix B for definitions). The parameters are given by $\bar{\Sigma} = ((\Psi^{-1} \otimes X'X) + \tilde{\Sigma}^{-1})^{-1}$, $\bar{\mu} = \bar{\Sigma}((\Psi^{-1} \otimes X'X)\hat{\gamma} + \tilde{\Sigma}^{-1}\tilde{\mu})$, and $\bar{\Psi} = (\Gamma - \hat{\Gamma})'X'X(\Gamma - \hat{\Gamma}) + \hat{E}'\hat{E}$. $\hat{E} = (Y - X\hat{\Gamma})$ where $\hat{\Gamma}$ is the OLS estimator of Γ, $\hat{\Gamma} = (X'X)^{-1}X'Y$.

3.2 Villani[2005] and Koop *et al.*[2010] priors.

The approach to Bayesian analysis of VECM models that I will adopt and subsequently modify is based on Villani[2005]. I will start by presenting its main characteristics relevant to the present paper, inviting the reader to consult the original paper for details. The starting point is the VECM model (2). To overcome the problem of nonidentifiability of β in the factorization $\Pi = \beta\alpha'$, linear normalization $\beta = [I_r, B']'$ is used.

The prior is the following (here Γ_V is obtained by stacking $\Gamma_1, \ldots, \Gamma_{P-1}$, note the difference with respect to other subsections)

$$p(\alpha, \beta, D, \Gamma_V, \Psi, r) = p(\alpha, \beta, D, \Gamma_V, \Psi|r)p(r), \tag{8}$$

where

$$\begin{aligned} p(\alpha, \beta, D, \Psi|r) &\propto |\Psi|^{-(K+q+r+1)/2} \exp\left(-\frac{1}{2}\text{tr}[\Psi^{-1}(A + (1/\lambda_\alpha^2)\alpha\beta'\beta\alpha')]\right), \\ p(\Gamma_V|\Psi) &\propto |\Psi \otimes \Sigma_{\Gamma_V}|^{-1/2} \exp\left(-\frac{1}{2}\text{tr}[\Psi^{-1}\Gamma_V'\Sigma_{\Gamma_V}^{-1}\Gamma_V]\right). \end{aligned}$$

Note, it is assumed that $p(\Gamma_V|\alpha, \beta, D, \Psi, r) = p(\Gamma_V|\Psi)$, therefore conditionally on Ψ parameters of the long-run impact matrix are independent of the parameters of short-run dynamics. The matrix Σ_{Γ_V} is given a block-diagonal form imitating Litterman construction, with blocks $\Sigma_{\Gamma_V,p} = \frac{\lambda_b^2}{p^{2\lambda_l}} I_K$, for $p = 1, \ldots, P-1$. The distribution of $\Gamma_V|\Psi$ is therefore the $N_{K(P-1)\times K}(0, \Psi, \Sigma_{\Gamma_V})$ (more on this and other matricvariate distributions in Appendix B).

Villani[2005] was the first to provide a solution to the debate on an uninformative prior for cointegration space, as it can be shown that $\mathrm{sp}(\beta)$ is uniformly distributed on the Grassmann manifold (see Appendix A). The paper also allows straightforward inference as all full conditional posterior density functions necessary to implement a Gibbs sampler are provided. The hyperparameters to be provided by the user are the following: q, A, λ_α, λ_b and λ_l. In the cited paper, a clear interpretation of the hyperparameters is provided by presenting various marginal and conditional distributions, allowing the user to better understand the construction of the prior. I shall concentrate on two of them.

Marginal prior distribution of Γ_V is the matricvariate t distribution $t_{K(P-1)\times K}(0, \Sigma_{\Gamma_V}^{-1}, A, q-K)$, so its mean is zero and covariance matrix $\mathrm{var}(\mathrm{vec}(\Gamma_V)) = E(\Psi) \otimes \Sigma_{\Gamma_V}$ where $E(\Psi) = A/(q-K-1)$ (if exists). Note the restrictive Kronecker structure of the covariance. This is the first restrictive feature of the prior that I will relax in the next subsection. A similar restrictive structure was the reason to question the appropriateness of the normal–Wishart and diffuse priors by Kadiyala and Karlsson[1997]. Their solution consisted of assuming the normal–diffuse prior, and in the next subsection I will similarly adapt the normal–diffuse prior to the VECM model.

As to the second restrictive feature, it has so far not been discussed in literature. Note that what is missing and potentially interesting in the cited paper, is the distribution of the long-run impact matrix Π as a whole, or at least its summaries such as first two moments. It can be shown (it is a special case of a more general situation presented in the next subsection, see also Appendix A) that the first two moments of Π under this prior are

$$E(\Pi) = 0 \qquad \text{and} \qquad V(\mathrm{vec}(\Pi)|r) = \lambda_\alpha^2 E(\Psi) \otimes \frac{r}{K} I_K.$$

The second moment, in my view, has a very restrictive form. In particular, it is inconsistent with beliefs on Π as expressed by the Litterman prior (even if its symmetric variant, with $\pi_1 = \pi_2$, is taken as a simplification). As shown below, in the elicitation procedure for the new prior, in such a case the prior moments of $\Pi = A_1 + \ldots + A_P - I_K$ are given by

$$E(\Pi) = 0 \qquad \text{and} \qquad V(\mathrm{vec}(\Pi)) = \tilde{D} \otimes \tilde{\Omega},$$

with matrices $\tilde{D}$ and $\tilde{\Omega}$, being functions of the data and Litterman's hyperparameters $\pi_1, \ldots, \pi_4$, to be defined. $\tilde{\Omega}$ is not even diagonal, let alone proportional to identity. Therefore, in the next subsection, a prior along the lines of Villani[2005] will be built, possessing the required moments as above.

I shall review the necessary facts from the paper by Koop *et al.*[2010]. The authors focus on eliciting identified cointegration *space*, spanned by the cointegrating vectors, rather than vectors themselves. Another contribution is in deriving corresponding posterior simulation methods, which had been complicated for such priors. From the outset the preferred normalization is that $\beta'\beta = I_r$ (where r is the number of cointegrating vectors), whose great advantage is that it does not restrict the cointegration space in a way that linear normalization does. The prior for β, and thus $\mathrm{sp}(\beta)$, is the so called *matrix angular central Gaussian* (see the next subsection and Appendix A for precise definitions) with density

$$p(\beta|r) \propto |\beta' P_\tau^{-1} \beta|^{-K/2},$$

for $\beta'\beta = I_r$, where it is through P_τ that a prior information on the location of cointegration space may be introduced (and it is shown how to do it). Then

$$\alpha|\beta \sim N_{K,r}(0, \nu(\beta' P_{1/\tau} \beta)^{-1}, G),$$

where G may or may not equal Ψ – depending on whether priors exploit the conjugacy or require additional flexibility (as it is in our case below). Such a setup will be also useful for the purpose of the present paper. However, the elicitation procedures will differ.

The authors motivate their MCMC algorithm by computational difficulties encountered in previous work, arising from the identification of β. Although $\alpha|\beta$ can easily be generated, drawing from the conditional posterior of β is nonstandard. The problem can be overcome in a smart way, using the idea of Liu[1994], as follows. Introduce notation: $\kappa = (\alpha'\alpha)^{1/2}$, $B = \beta\kappa$ and $A = \alpha\kappa^{-1}$ so that

$$\beta\alpha' = (\beta\kappa)(\alpha\kappa^{-1})' = BA'.$$

Now, however, $A'A = I_r$ and B is unrestricted. As the posterior sampler that Koop *et al.*[2010] propose switches between these two parameterizations, it is useful to express the prior similarly. Their Proposition 1 (p. 230) states that the prior on (α, β) expressed as above is equivalent to the following prior on (A, B)

$$p(A) \propto |A'G^{-1}A|^{-K/2} \qquad \text{and} \qquad B \sim N_{K,r}(0, (A'G^{-1}A)^{-1}, \nu P_\tau).$$

Then a Gibbs sampler based on these two parameterizations is proposed. I will come back to it when I present the full algorithm for the model comprising all the dynamic, deterministic and stochastic parameters.

3.3 The new proposal

This is the central subsection of the paper. We are now equipped with necessary tools to introduce the new prior, discuss its properties, advantages and limitations, and to present the sampling algorithm.

The prior

The prior that I propose is the following

$$p(\alpha, \beta, \gamma, \Psi|r) = p(\alpha, \beta|r)p(\gamma)p(\Psi), \tag{9}$$

where

$$\begin{aligned} p(\alpha, \beta|r) &\propto \exp(-\frac{1}{2}\mathrm{tr}(\tilde{D}^{-1}\alpha\beta'\tilde{S}^{-1}\beta\alpha')), \\ p(\gamma) &\propto \exp(-\frac{1}{2}(\gamma - \tilde{v})'\tilde{V}^{-1}(\gamma - \tilde{v})), \\ p(\Psi) &\propto |\Psi|^{-\frac{K+1}{2}}. \end{aligned}$$

$\beta'\beta = I_r$ is assumed as a normalization and $\tilde{S} = \frac{K}{r}\tilde{\Omega}$. This time $\Gamma = [D', \Gamma'_1, \ldots, \Gamma'_{P-1}]$, *i.e.* the deterministic terms is included in this parameter (this is a small difference with respect to Villani[2005], which makes the new prior more flexible and more in line with the normal–diffuse of Kadiyala and Karlsson[1997]). The prior is parameterized by three matrices $\tilde{D}$, $\tilde{\Omega}$ and $\tilde{V}$.

The prior has a structure similar to the normal–diffuse prior, in that the parameters of the dynamic and stochastic structure are independent *a priori*. Like in the normal–diffuse prior, the prior distribution of Ψ is diffuse. As to short-run and deterministic terms parameter matrices, note that I relax the Kronecker-type restriction on the covariance matrix of Villani[2005] and use normal distribution with an unrestricted covariance matrix instead. As in Villani[2005], I keep the distribution of long-run matrix independent of other parameters. The prior has some limitations that make it only an approximation of the ideal situation: firstly, as the distribution of Π would not be independent of that of Γ_ps if the distributions of these parameters were obtained from the Litterman prior according to eqs. (3), this only refers to the prior distribution and does not impose any restriction on the posterior, thus I believe this point is not too restrictive; also note, that the covariance matrix of Π will have a Kronecker structure *a priori*, which is not in line with the prior variance implied by the Litterman prior (with $\pi_1 \neq \pi_2$) *via* $\Pi = A_1 + \ldots + A_P - I_K$, but from the reading of Villani[2005] it seems unavoidable.

I shall present the key properties of the prior. In the derivations below I use well-known properties of conditional expectation, matricvariate distributions and matrix algebra operations reviewed in Appendix B. I will focus on the most important factor $p(\alpha, \beta|r)$, and address the issue of moments of Π raised during the discussion of the Villani[2005] paper. It has the following conditional structure

$$\begin{aligned}
p(\alpha|\beta, r) &= f_{N_{K\times r}}(\alpha|0, \tilde{D}, (\beta'\tilde{S}^{-1}\beta)^{-1}), \\
p(\beta|r) &\propto \int p(\alpha, \beta|r)d\alpha \\
&= \int \exp(-\frac{1}{2}\text{vec}(\alpha)'(\beta'\tilde{S}^{-1}\beta \otimes I_K)\text{vec}(\alpha))d\text{vec}(\alpha) \\
&= |(\beta'\tilde{S}^{-1}\beta \otimes I_K)^{-1}|^{\frac{1}{2}} \\
&= |\beta'\tilde{S}^{-1}\beta|^{-\frac{K}{2}}.
\end{aligned}$$

The normalization used is that $\beta'\beta = I_r$. sp(β) is not distributed uniformly over Grassman manifold any longer, but possesses a so-called *matrix angular central Gaussian* distribution with parameter $\tilde{S}$, or $\beta \sim MACG(\tilde{S})$ (see Appendix A). For the mean of the whole long-run impact matrix I obtain

$$E(\Pi) = E(\beta\alpha') = E_\beta(E(\beta\alpha'|\beta)) = E_\beta(\beta E(\alpha'|\beta)) = E_\beta(\beta 0) = 0,$$

whereas for the variance I get

$$\begin{aligned}
V(\text{vec}(\Pi)|r) &= V(\text{vec}(\beta\alpha')) \\
&= E_\beta(E(\text{vec}(\beta\alpha')\text{vec}(\beta\alpha')'|\beta)) \\
&= E_\beta((I_K \otimes \beta)E(\text{vec}(\alpha')\text{vec}(\alpha')'|\beta)(I_K \otimes \beta')) \\
&= E_\beta((I_K \otimes \beta)(\tilde{D} \otimes (\beta'\tilde{S}^{-1}\beta)^{-1})(I_K \otimes \beta')) \\
&= \tilde{D} \otimes E_\beta(\beta(\beta'\tilde{S}^{-1}\beta)^{-1}\beta').
\end{aligned}$$

Now Lemma 1 below, proved in Appendix A, shows that $E_\beta(\beta(\beta'\tilde{S}^{-1}\beta)^{-1}\beta') = \tilde{\Omega}$.

Lemma 1. *Suppose β is distributed as $MACG(\tilde{S})$ with SPD matrix $\tilde{S} \in \mathbb{R}^{K\times K}$, then the expectation $E_\beta(\beta(\beta'\tilde{S}^{-1}\beta)^{-1}\beta') = \frac{r}{K}\tilde{S}$.*

In short, the proposed prior possesses the following moments

$$E(\Pi) = 0 \qquad \text{and} \qquad V(\text{vec}(\Pi)|r) = \tilde{D} \otimes \tilde{\Omega},$$

as required.

The prior is parameterized by three matrices $\tilde{D}$, $\tilde{\Omega}$ and $\tilde{V}$. The structure is rich enough, so that these matrices in turn will be matched with appropriate matrices derived from Litterman prior *via* VAR to VECM transformation (3), as shown in the paragraph on elicitation below. The second-order structures of parameters under both priors will thus agree.

The prior is parameterized by the same set of hyperparameters $\pi_1,\ldots,\pi_4$ as the priors considered in subsection 3.1. No additional hyperparameters are needed, as it is in case of Villani[2005] or Koop *et al.*[2010]. Moreover, the hyperparameters retain approximately their original interpretation. Hence, the structure of the prior should be transparent enough to the end user.

The elicitation

We will first elicit the hyperparameters $\tilde{D}$ and $\tilde{S}$ for the prior distribution of the factors of the long-run impact matrix, and then $\tilde{V}$ for the remaining parameters of the dynamic structure. As has been explained, the structure of the proposed prior, the prior for parameters of the VECM model (2), allows us to derive its hyperparameters from the well known prior distribution on the parameters of the VAR model (1), the Litterman prior discussed in subsection 3.1.

As for $\tilde{D}$ and $\tilde{S}$, it has been explained that for this part I will assume the symmetric variant of the Litterman prior. Assume $\pi_1 = \pi_2$ and then

$$\tilde{\tilde{\Omega}} = \left[\begin{array}{c|c|ccccc|c}
\pi_3 I_{K_d} & & & & & & & \\ \hline
& \ddots & & & & & & \\ \hline
& & \frac{\pi_1}{p^{\pi_4}\sigma_1^2} & & & & & \\
& & & \ddots & & & & \\
& & & & \frac{\pi_1}{p^{\pi_4}\sigma_k^2} & & & \\
& & & & & \ddots & & \\
& & & & & & \frac{\pi_1}{p^{\pi_4}\sigma_K^2} & \\ \hline
& & & & & & & \ddots
\end{array}\right], \quad \tilde{D} = \begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_K^2 \end{bmatrix}.$$

The $\{\sigma_k^2\}$s are unknown and estimated from the data as usual in the Litterman-type constructions. Denote $\Gamma^{var} = [D', A_1', \ldots, A_P']$, the parameters of the original VAR model (1). Under the (symmetric) Litterman prior $\Gamma^{var} \sim N_{(Kd+KP)\times K}(M, \tilde{\tilde{\Omega}}, \tilde{D})$, with properly specified mean M. Now, with

$$\Phi = [0_{K\times Kd}, I_K, \ldots, I_K]$$

(there are P copies of I_K) we have $\Pi = \Phi\Gamma^{var} - I_K$ (see eqs. (3)) and therefore $\Pi \sim N_{K\times K}(0, \tilde{\Omega}, \tilde{D})$, where $\tilde{\Omega} = \Phi\tilde{\tilde{\Omega}}\Phi'$ and $\tilde{D}$ are both understood to be hyperparameters of the prior distribution. Using this construction the moments of Π are given by

$$E(\Pi) = 0 \qquad \text{and} \qquad V(\text{vec}(\Pi)) = \tilde{D} \otimes \tilde{\Omega},$$

as I showed the proposed prior has the same structure, and thus, the moments can be matched.

For $\tilde{V}$, note that $\Gamma = \Phi\Gamma^{var}$ (see eqs. (3)), or $\gamma = (I_K \otimes \Phi)\gamma^{var}$ in vectorized form, where this time

$$\Phi = \begin{bmatrix} I_{Kd} & 0 & 0 & \dots & 0 \\ 0 & 0 & -I_K & \dots & -I_K \\ & & \ddots & & \\ 0 & 0 & 0 & \dots & -I_K \end{bmatrix}.$$

Therefore, $\tilde{v} = (I_K \otimes \Phi)\tilde{\mu}$, and $\tilde{V} = (I_K \otimes \Phi)\tilde{\Sigma}(I_K \otimes \Phi')$ where the prior moments of γ^{var}, $\tilde{\mu}$ and $\tilde{\Sigma}$, were defined in the subsection on the normal–diffuse prior below eqs. (7).

The posterior inference

The posterior inference is implemented with a Gibbs sampler, as is common in literature on Bayesian methods and fortunately possible in our framework. For the setup of the sampler, the paper by Koop *et al.*[2010] is instrumental. Their idea is as follows. Note, in the above prior, as well as later in the posterior, α is unrestricted and β is semi-orthogonal. The latter restriction complicates drawing samples from the posterior (obtaining draws from $\alpha|\beta$ then is easy). The proposed prior is also not conjugate for our VECM model (in such a case, drawing samples would be easy too). On the other hand, by defining $A = \alpha(\alpha'\alpha)^{-1/2}$ and $B = \beta(\alpha'\alpha)^{-1/2}$, now it is A that is semi-orthogonal and B unrestricted, moreover $\beta\alpha' = BA'$. Prior distribution on (α, β) can be equivalently expressed as a prior on (A, B). Therefore, the authors explain, a draw from (α, β) (prior or posterior, conditionally on r and other parameters) can be obtained by starting with a given (α^0, β^0), drawing $\alpha^1 \sim p(\alpha|\beta)$, transforming $A = \alpha^1(\alpha^{1\prime}\alpha^1)^{-1/2}$, drawing $B \sim p(B|A)$ and finally transforming $\beta^1 = B(B'B)^{-1/2}$. The reader is invited to read the original paper and Liu[1994] for details. The following theorem implements their idea and presents the full conditional posterior density functions necessary to set up the algorithm for drawing samples from the joint posterior distribution of the parameters. The theorem is proved in Appendix B by combining derivations from Kadiyala and Karlsson[1997] and Koop *et al.*[2010].

Theorem 1. *Given model (2), which can be rewritten as*

$$Z_d = Z_l\beta\alpha' + X\Gamma + E \qquad \text{and} \qquad E \sim N_{T\times K}(0, I_T, \Psi),$$

where Z_d is $(T \times K)$ matrix obtained by stacking Δy_ts, Z_l $(T \times K)$ by stacking y_{t-1}s, X $(T \times (K_d + KP))$ by stacking $x_t = [d_t, \Delta y_{t-1}, \dots, \Delta y_{t-P}]$s. Given *the prior (9), the Gibbs sampler is implemented by drawing sequentially from the following*

- $(\alpha|\beta, \Gamma, \Psi, data)$: $\text{vec}(\alpha) \sim N(\bar{a}, \bar{D})$ *where*

$$\begin{aligned} \bar{D} &= (\beta' Z_l' Z_l \beta \otimes \Psi^{-1} + \beta' \tilde{S}^{-1} \beta \otimes \tilde{D}^{-1})^{-1}, \\ \bar{a} &= \bar{D}\text{vec}(\Psi^{-1}(Z_d - X\Gamma)' Z_l \beta); \end{aligned}$$

- $(\beta|\alpha, \Gamma, \Psi, data)$: $\beta = B(B'B)^{-1/2}$ *where* $\text{vec}(B) \sim N(\bar{b}, \bar{S})$ *and*

$$\begin{aligned} \bar{S} &= (A'\Psi^{-1}A \otimes Z_l' Z_l + A'\tilde{D}^{-1}A \otimes \tilde{S}^{-1})^{-1}, \\ \bar{b} &= \bar{S}\text{vec}(Z_l'(Z_d - X\Gamma)\Psi^{-1}A), \end{aligned}$$

  where $A = \alpha(\alpha'\alpha)^{-1/2}$*;*

- $(\Gamma|\alpha, \beta, \Psi, data)$: $\gamma \sim N(\bar{v}, \bar{V})$ *where*

$$\begin{aligned} \bar{V} &= (\Psi^{-1} \otimes X'X + \tilde{V}^{-1})^{-1}, \\ \bar{v} &= \bar{V}((\Psi^{-1} \otimes X'X)\hat{\gamma} + \tilde{V}^{-1}\tilde{v}), \end{aligned}$$

  for $\hat{\gamma} = \text{vec}(\hat{\Gamma})$*,* $\hat{\Gamma} = (X'X)^{-1}X'(Z_d - Z_l\beta\alpha')$*;*

- $(\Psi|\alpha, \beta, \Gamma, data)$: $\Psi \sim iW(\bar{\Psi}, T)$ *where*

$$\bar{\Psi} = (\Gamma - \hat{\Gamma})'X'X(\Gamma - \hat{\Gamma}) + \hat{E}'\hat{E},$$

  where $\hat{E} = Z_d - Z_l\beta\alpha - X\hat{\Gamma}$ *and* $\hat{\Gamma} = (X'X)^{-1}X'(Z_d - Z_l\beta\alpha')$*.*

As a starting point for the algorithm, the ML estimate will be used[1], thus reducing the necessaity for a convergence burn-in period.

4 Empirical Example

In this section I will show an example of how the methods developed can be used. As the proposed prior is supposed to approximate the Litterman/normal–diffuse priors within a class of cointegrated priors, it is of course desirable to compare these four priors between them. To this end I shall re-examine a widely-known dataset, originally used by Kadiyala and Karlsson[1997] in their paper on comparing forecasting accuracy of various

[1]Recall that Johansen's procedure uses linear normalization, but moving from linear to nonordinal normalizations has been explained in subsection 3.2 and can easily be implemented to obtain starting points.

priors. The exercise is supposed to be just a rough test on the proposed prior, not a proper forecasting method horse-race whose aim would be to argue in favour of one particular method. Still, anticipating the results, the forecast comparison shows that the proposed prior provides a better overall (in a sense to be defined) improvement over benchmarks than the Litterman or the normal–diffuse priors[2].

Data and model

The data[3] consists of quarterly time series that start in the first quarter of 1948 and end in the fourth quarter of 1986. The variables are OBS, RGNPG, INFLA, UNEMP, LM1, INVEST, CPRATE and CBI; *i.e.* observation number, real growth rate of GNP (annualized quarterly changes in real GNP), inflation rate (annualized quarterly changes in the implicit GNP deflator), unemployment (percentage of civilian labour force), natural logarithm of average of monthly M1 data, gross private domestic investment (nominal), 4–6 month commercial paper rate (averages of daily rates), changes in business inventory (nominal at annual rates). RGNPG is obtained from seasonally adjusted nominal GNP and the seasonally adjusted GNP deflator. M1, UNEMP, INVEST and CBI are seasonally adjusted. The data is shown on Fig. 1. As we can see, the seven series display a wide range of possible behaviours, including trending (INVEST) and even possible I(2) type nonstationarity (LM1).

[2]Note, that the resulting convenient factorization under the Litterman prior (of the posterior distribution of $\gamma = \mathrm{vec}(\Gamma)$ into independent distributions of the γ_ks) is a consequence of assumed diagonality of Ψ. At the time of the construction of the prior, this was an important simplification as it implied a sequence of inversions of smaller matrices rather than one inversion of a matrix of bigger dimensions (compare the dimensions of $\bar{\Sigma}_k$s above and $\bar{\Sigma}$ for the normal–diffuse prior at the end of this subsection). However, for the normal–diffuse prior I have made large inversions nevertheless; moreover, it would be rather unjustified to compare models that assume Ψ diagonal with diagonal elements fixed, with models that relax these two restrictions at the same time. Thus, in the empirical section I will work with a model where Ψ is fixed at an estimated value (without diagonality imposed), and call it Litterman. As the benchmark I will also use the original (or pure) Litterman prior too. This can be also thought of as a variant of the normal–diffuse prior, where Ψ is fixed at an estimated value. This gives a more detailed view of the performance of the methods considered.

[3]`http://www.econ.queensu.ca/jae/1997-v12.2/kadiyala-karlsson/`.

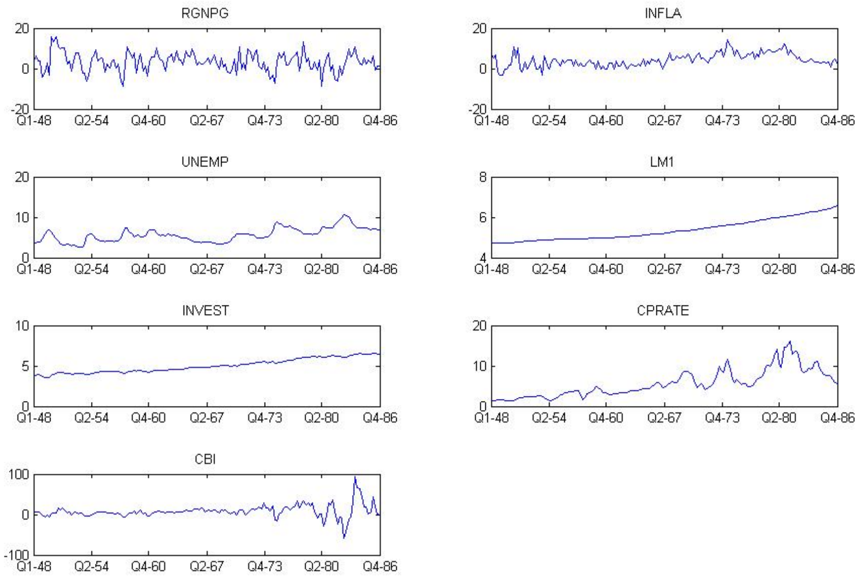

Figure 1: The data (source: Kadiyala and Karlsson[1997]).

As in Kadiyala and Karlsson[1997], I will use a standard VAR setup (1) with an intercept term as the only deterministic term and a lag length of $P = 6$. As to the cointegration rank, its determination is not a part of the experiment, but I simply assume a cointegration rank $r = 5$ – this choice has some support in a basic cointegration analysis carried out in JMulTi using the Johansen procedure (available on request).

Forecasting and measures of forecast accuracy

I will obtain forecasts at different horizons using a chain rule of forecasting, as the means of posterior predictive densities. Thus I will implement the same procedure as Kadiyala and Karlsson[1997], where the reader is referred to for details.

For forecasts of individual variables it is standard procedure to use the square root of the mean squared forecast error (RMSFE) as a measure of forecast accuracy. To measure forecasting power at the aggregate level I will use the log-determinant of the cross-products of forecast errors at horizon h, *i.e.* $\ln|E_h| = \ln|(Y_e - Y_e^f(h))'(Y_e - Y_e^f(h))|$, where the Y_e denotes observations on variables in the evaluation period and $Y_e^f(h)$ the corresponding forecasts made h-quarters back; it has a long tradition in econometrics dating back at least to Doan *et al.*[1986] (see the original paper for justification).

Experimental setup

The sample consists of 183 quarters. As in Kadiyala and Karlsson[1997], the period of 27 quarters 1980q2 to 1986q4 forms evaluation period. For each quarter, using each of the priors, three forecasts $Y_e^f(h)$ for $h = 1, 4, 8$ are obtained, and $\ln|E_h|$ is computed. Fixed, rolling[4] and recursive forecasting schemes are implemented. However, for the sake of brevity only the results for the recursive scheme, as the most often used, are discussed. The full set of results is available in the appendix. Note also that in principle four benchmarks are potentially worth considering: a no-change forecast, a univariate AR, VAR, and, last but not least, the pure Litterman prior (*i.e.* the prior with a diagonal fixed residual covariance matrix) which other priors are supposed to supplant. The results presented take as a benchmark the no-change forecast. The result for other benchmarks are also presented in the appendix (excluding the figures to save space).

As the second order structure of all of them is determined similarly and the hyperparameters retain approximately the original interpretation, I use the same set of hyperparameters as Kadiyala and Karlsson[1997] (however, I decided to change π_3 to a small value of 0.0001, and thus go back to the original Litterman setup from which Kadiyala and Karlsson[1997] deviated in their paper). This I believe is sufficient to ensure that the comparison is fair to all the methods considered. The hyperparameters are thus: $\pi_1 = 0.04$, $\pi_2 = 0.0036$, $\pi_3 = 0.0001$ and $\pi_4 = 1$.

Finally, the burn-in period for the Litterman and the normal–diffuse priors was set to $Nbin = 500$ and the Markov chains used for forecasting were of length $N = 200$, due to the nonlinearity in the case of the new prior for the VECM model, I used $Nbin = 1000$ and $N = 1000$. Graphical inspection of the draws did not reveal any convergence problems.

Results

Overall, the presented results are favourable to the proposed prior. Note first the organization of the results for the experiment in Table 1. Thus, I am considering priors for VAR and the new proposal for VECM. Both can assume the residual covariance matrix fixed at the estimated value or not. Thus, given a chosen scheme of forecasting and a benchmark, we obtain for each forecast horizon a (2×2) table of relative $\ln|E_h|$s. Here I concentrate on

[4]In the case of a rolling window, for each forecast horizon the maximal sample size is used, *i.e.* the sample size is the same as available for the first forecasted period, when the window is rolled and the model re-estimated

the most popular recursive scheme of forecasting and a no-change benchmark. The full set of results is available in Appendix C.

Table 1: Organization of the results on the relative $\ln|E_h|$ for a given forecasting scheme, benchmark and forecast horizon h

Forecast horizon	Ψ-fixed	Ψ-diffuse
VAR priors	Litterman prior	normal–diffuse prior
VECM priors	the new prior	the new prior

The relative gains or losses in forecast accuracy are presented in Table 2 below. Note the qualitative uniformity of the conclusions across forecast horizons: for each horizon the methods assuming fixed and known residual covariance matrix Ψ fare better than the methods assuming prior $p(\Psi) \propto |\Psi|^{-\frac{K+1}{2}}$. Furthermore, methods modelling cointegration explicitly fare better than the methods that do not model cointegration (except for random Ψ for two years forecast). Overall, the method that gives best results is the proposed in subsection 3.3 prior with a fixed residual covariance matrix Ψ.

Table 2: The relative $\ln|E_h|$ for all priors, with the no-change forecast as a benchmark, under the recursive scheme of forecasting.

h=1	Ψ-fixed	Ψ-diffuse
noCI	0.8498	0.9078
CI-priors	0.8380	0.8937
h=4	Ψ-fixed	Ψ-diffuse
noCI	0.9056	0.9580
CI-priors	0.9027	0.9547
h=8	Ψ-fixed	Ψ-diffuse
noCI	0.9329	0.9779
CI-priors	0.9273	0.9795

On the other hand, the performance of the methods under consideration for each variable separately is investigated on Figure 2. For one-quarter-ahead forecasts, the proposed priors are rather indistinguishable from the point of view of RMSFE for individual variables. Only for two-years-ahead forecasts, the differences appear. However, a uniform ranking of priors' forecast performance is impossible as their relative accuracy differs from one variable to another. The reader has surely noticed a striking regularity: for each variable all the methods give very similar forecast errors to each other. They are all higher, or all lower than the benchmark. In the working paper version I demonstrated that such a pattern is also obtained from a wider comparison of VAR priors, in a unified framework and balanced environment of optimizing choice of hyperparameters. I conjectured that this fact may have gone unnoticed due to the omission of the optimizing step in the prior elicitation procedure in some previous papers. This regularity should not be therefore considered a failure of the new prior in comparison to VAR priors, I believe, but rather an issue pertinent to the model itself.

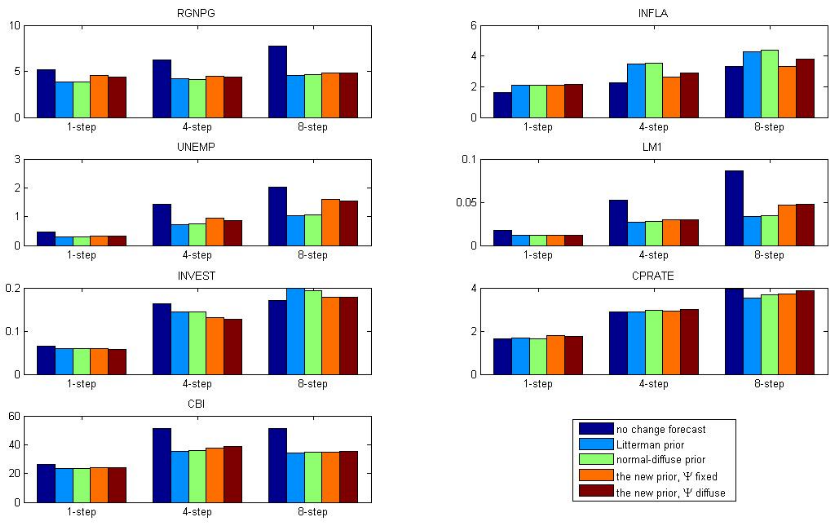

Figure 2: Individual RMSFE for all priors, with the no-change forecast as a benchmark, under the recursive scheme of forecasting.

Summary

To sum up, the new VECM prior provides a modest improvement over the Litterman/normal–diffuse priors. In fact, the proposed prior turns out to give largest gains with respect to any benchmark. It is also noteworthy that the priors that assume a fixed, albeit unrestricted, residual covariance matrix, fare uniformly better than those with Ψ random, being a parameter to be simulated, which can be interesting for applied forecasters.

5 Conclusions

The theoretical contribution of this paper is the construction of a prior for cointegrated systems expressing approximate Litterman beliefs, that is less restrictive than priors used to date. The proposed prior gives applied researchers yet another modelling framework that may better express their beliefs and suit their needs in a fairly flexible way. Eliciting priors is difficult in general – with this new prior I give researchers a simple alternative for cointegrated systems that is based on widespread and well-established beliefs about the properties of macroeconomic variables that can be elicited automatically. The efficient Gibbs sampler allows the framework to be used on a routine basis. In a forecast comparison the proposed prior displayed improvements over popular benchmarks.

A Appendix. Proof of Lemma 1

The aim of this appendix is to prove the following

Lemma 1. *Let $\tilde{S} \in \mathbb{R}^{K\times K}$ be SPD matrix and let $\beta \in \mathbb{R}^{K\times r}$ be such that $\beta'\beta = I_r$. Assume β is distributed according to* matrix angular central Gaussian *distribution with parameter $\tilde{S}$, denoted $\beta \sim MACG(\tilde{S})$, i.e. β has density $p(\beta) \propto |\beta'\tilde{S}^{-1}\beta|^{-\frac{K}{2}}$. Then $E_\beta(\beta(\beta'\tilde{S}^{-1}\beta)^{-1}\beta') = \frac{r}{K}\tilde{S}$.*

The proof will be immediate and will be presented at the end of this section, but first a number of notions and results pertaining to the field of *directional statistics* must be introduced. This section provides only the minimum necessary for the proof. The interested reader is invited to consult the book by Chikuse[2003], on which this whole section is based and to which I refer throughout without explicit mentioning, for a more in-depth treatment.

The *Stiefel manifold* $V_{r,K}$ is the space whose points are *r-frames* in $\mathbb{R}^K$, where a r-frame in $\mathbb{R}^K$ ($r \leq K$) is a set of r orthonormal vectors in $\mathbb{R}^K$. The Stiefel manifold $V_{r,K}$ is represented by the set of $(K \times r)$ matrices X such that $X'X = I_r$, thus $V_{r,K} = \{X \in \mathbb{R}^{K\times r} \colon\ X'X = I_r\}$. For $r = K$, $V_{r,K}$ is the *orthogonal group* $O(r)$ of $(r \times r)$ orthonormal matrices. A point in $V_{r,K}$ may also be called an *orientation* extending the notion of a direction for $r = 1$. The *Grassmann manifold* $G_{r,K-r}$ is the space whose points are *r-planes* $\mathcal{V}$, that is, r-dimensional hyperplanes in $\mathbb{R}^K$ containing the origin.

The following equivalent definition of the Grassmann manifold is of particular interest and use in its statistical analysis. To each r-plane $\mathcal{V}$ in $G_{r,K-r}$ corresponds a unique $(K \times K)$ orthogonal projection matrix P idempotent of rank r onto $\mathcal{V}$. If the r columns of a $(K \times K)$ matrix Y in $V_{r,K}$ span $\mathcal{V}$, we have $YY' = P$. Letting $P_{r,K-r}$ denote the set of all $(K \times K)$ orthogonal projection matrices idempotent of rank r, the matrix space $P_{r,K-r}$ is a manifold equivalent to the Grassmann manifold $G_{r,K-r}$, and the statistical analysis on the equivalent manifold $P_{r,K-r}$ can be conducted as on the Grassmann manifold.

On the above defined manifolds uniform distributions may be constructed as normalized unique invariant measures. These invariant measures are represented through their differential forms, to be specific, exterior differential forms of maximum degree, by taking exterior product of suitable linear differential forms of degree one.[5] As the manifolds are compact, the total

[5]This is actually a constructive proof in a special case of Stiefel or Grassmann manifold of a more general theorem, where the existence of a unique invariant measure on a topological space under a transitive compact topological group of transformations of this space into itself under a certain assumption of continuity is proved.

masses of these measures may be evaluated, thus yielding the normalizing factor. Let us denote the unique invariant measures on the manifolds $V_{r,K}$ and $G_{r,K-r}$ by (dX) and (dP), and the normalized measures by $[dX]$ and $[dP]$, respectively. We shall not concern ourselves with the mathematical intricacies of the construction, as it is not a purpose of this section and the distribution of interest is a nonuniform distribution $MACG$ that will be introduced later, whose density function, however, is expressed *with respect to* the normalized measure $[dX]$.

The uniform distribution on $P_{r,K-r}$ has mathematical expectation

$$E(P) = \frac{r}{K} I_r.$$

Furthermore, random matrices distributed uniformly on Stiefel and Grassmann manifold are related, in a sense that, $P = XX'$ is uniform on $P_{r,K-r}$ if and only if X is uniform on $V_{r,K}$. The proof of Lemma 1 will consist of reducing our problem to these two well-known facts.

Let Z be an $(K \times r)$ $(r \leq K)$ matrix of rank r. Then we can define the unique *polar decomposition* of Z as

$$Z = H_Z T_Z^{1/2}, \text{ with } H_Z = Z(Z'Z)^{-1/2} \text{ and } T_Z = Z'Z,$$

so that the *orientation* $H_Z \in V_{r,K}$. $A^{-1/2} = (A^{1/2})^{-1}$ where $A^{1/2}$ denotes the unique square root of a positive semidefinite matrix A.

A point in the Stiefel manifold can be therefore obtained as the orientation of a $(K \times r)$ matrix Z. Moreover, some families of nonuniform distributions on the Stiefel manifold can be generated that way, as the distributions of the orientation of a random matrix. Their distributions are then derived from the distribution of the random matrix whose orientation is under consideration (the same refers to the distribution of T_Z). In particular, if Z has the matricvariate normal distribution $Z \sim N_{K\times r}(0; \tilde{S}, I_r)$, we have the density function of H_Z (with respect to the uniform distribution on the Stiefel manifold)

$$f(H_Z) = |\tilde{S}|^{-r/2} |H_Z' \tilde{S}^{-1} H_Z|^{-K/2}.$$

Thus, $H_Z \sim MACG(\tilde{S})$. Note as a special case, that when $\tilde{S} = I_K$, H_Z is distributed uniformly on $V_{r,K}$.

The last piece of information that we need to present the proof of Lemma 1, is the relation between the distribution of the orientation of a given random matrix Z and the distribution of the orientation of a linear transformation of Z. Assume Z is a $(K \times r)$ random matrix, and $Y = BZ$ where B is a $(K \times K)$ nonsingular matrix. Denote the respective polar decompositions by $Z = H_Z T_Z^{1/2}$ and $Y = H_Y T_Y^{1/2}$. If f_{H_Z} is the density function of H_Z, then the density function of H_Y is given by

$$f_{H_Y}(H_Y) = |B|^{-r}|W'W|^{-K/2} f_{H_Z}(H_W),$$

where $W = B^{-1}H_Y$ and $H_W = W(W'W)^{-1/2}$. In particular, if $H_Z \sim MACG(\Sigma)$ then $Y = BZ \sim MACG(B\Sigma B')$, if $H_Z \sim MACG(I_K)$ (*i.e.* H_Z has uniform distribution over $V_{r,K}$) then $Y = BZ \sim MACG(BB')$, and finally if $H_Z \sim MACG(\Sigma)$ then for B such that $BB' = \Sigma^{-1}$ the transformation $Y = BZ$ has uniform distribution over $V_{r,K}$.

Proof. The precise meaning of the expectation in question is

$$\mathcal{I} = E_\beta(\beta(\beta'\tilde{S}^{-1}\beta)^{-1}\beta') = \int_{V_{r,K}} \beta(\beta'\tilde{S}^{-1}\beta)^{-1}\beta'|\tilde{S}|^{-r/2}|\beta'\tilde{S}^{-1}\beta|^{-K/2}[d\beta].$$

As we know, we can think of the $MACG(\tilde{S})$ distribution of the orientation β, as arising from distribution of the polar decomposition of $(K \times r)$ matrix Z distributed as $N_{K\times r}(0; \tilde{S}, I_r)$, $\beta = H_Z$ where $Z \sim N_{K\times r}(0; \tilde{S}, I_r)$. Make a transformation $Y = \tilde{S}^{-1/2}Z$. Then simple matrix algebra yields $H_Z(H_Z'\tilde{S}^{-1}H_Z)^{-1}H_Z' = \tilde{S}^{1/2}H_Y H_Y'\tilde{S}^{1/2}$. On the other hand H_Y has uniform distribution on $V_{r,K}$, $f_{H_Y}(H_Y) = 1$, which is equivalent to $H_Y H_Y' = P$ having uniform distribution on $P_{r,K-r}$. Thus,

$$\begin{aligned}
\mathcal{I} &= \int_{V_{r,K}} H_Z(H_Z'\tilde{S}^{-1}H_Z)^{-1}H_Z'|\tilde{S}|^{-r/2}|H_Z'\tilde{S}^{-1}H_Z|^{-K/2}[dH_Z] \\
&= \int_{V_{r,K}} \tilde{S}^{1/2}H_Y H_Y'\tilde{S}^{1/2}[dH_Y] = \tilde{S}^{1/2}\Big(\int_{P_{r,K-r}} P[dP]\Big)\tilde{S}^{1/2} \\
&= \tilde{S}^{1/2}\Big(\frac{r}{K}I_K\Big)\tilde{S}^{1/2} = \frac{r}{K}\tilde{S}.
\end{aligned}$$

□

B Appendix. Proof of Theorem 1

I shall first present definitions of matricvariate distributions and vector operations used in this paper and in the proof below (Bauwens, Lubrano and Richard[1999], Gupta and Nagar[2000], Lütkepohl[2006] are good references).

Let $X \in \mathbb{R}^{p\times q}$, $M \in \mathbb{R}^{p\times q}$, $P \in \mathbb{R}^{p\times p}$ and $Q \in \mathbb{R}^{q\times q}$ be SPD matrices. X has a *matricvariate normal* distribution with parameters M, Q and P, denoted $X \sim N_{p\times q}(M, Q, P)$, if its pdf is given by

$$p(X) \propto |Q \otimes P|^{-1/2} \exp\big(-\frac{1}{2}\mathrm{tr}[Q^{-1}(X - M)'P^{-1}(X - M)]\big).$$

Moreover, if $X \sim N_{p\times q}(M, P, Q)$ then $E(X) = M$ and $V(\text{vec}(X)) = Q \otimes P$; if $X \sim N_{p\times q}(M, P, Q)$ then $X' \sim N_{q\times p}(M', Q, P)$.

Let $\Sigma \in \mathbb{R}^{q\times q}$ and $A \in \mathbb{R}^{q\times q}$ be SPD matrices and $\nu \geqslant q$. Σ has an *inverse-Wishart* distribution with parameters A and ν, denoted $\Sigma \sim iW_q(A, \nu)$ if its pdf is given by

$$p(\Sigma) \propto |\Sigma|^{-\frac{\nu+q+1}{2}} \exp\big(-\frac{1}{2}\text{tr}(\Sigma^{-1}A)\big).$$

Let $X \in \mathbb{R}^{p\times q}$, $M \in \mathbb{R}^{p\times q}$, $P \in \mathbb{R}^{p\times p}$ and $Q \in \mathbb{R}^{q\times q}$ be SPD matrices, and $\nu \geqslant 0$. X has a *matricvariate Student's t* distribution with parameters M, P, Q and degrees of freedom ν, denoted $X \sim t_{p\times q}(M, P, Q, \nu)$ if its pdf is given by

$$p(X) \propto |I_q + Q^{-1}(X - M)'P(X - M)|^{-\frac{\nu+p+q}{2}}.$$

Let A be $(p \times q)$ matrix, then there exists a matrix K_{pq}, called commutation matrix, such that $\text{vec}(A') = K_{pq}\text{vec}(A)$ (or equivalently $\text{vec}(A) = K_{pq}\text{vec}(A')$ and $K_{pq} = K_{qp}$).

Assume that A, B and C are matrices. In the proof below I will use the following matrix operations:
Op.1) specifically, let A be $(m \times n)$ matrix, B be $(p \times q)$ matrix, then $K_{pm}(A \otimes B)K_{nq} = (B \otimes A)$;
Op.2) if A and B are square matrices of dimensions m and n, respectively, then $|A \otimes B| = |A|^n|B|^m$;
Op.3) $(A \otimes B)' = (A' \otimes B')$, $(A \otimes B)^{-1} = (A^{-1} \otimes B^{-1})$ if the inverses of A and B exist, $(A \otimes B)(C \otimes D) = (AC \otimes BD)$ if the dimensions agree;
Op.4) $\text{tr}(ABC) = \text{vec}(A')'(C' \otimes I)\text{vec}(B)$ if the dimensions agree;
Op.5) $\text{vec}(ABC) = (C' \otimes A)\text{vec}(B)$ if the dimensions agree.

For the purpose of proving Theorem 1, recall the notation used. The model is expressed in matrix form as

$$Z_d = Z_l\beta\alpha' + X\Gamma + E \qquad \text{and} \qquad E \sim N_{T\times K}(0, I_T, \Psi),$$

Z_d and Z_l are $(T\times K)$, and X $(T\times(K_d+KP))$ matrix. The prior conditionally on cointegration rank is given by

$$p(\alpha, \beta, \gamma, \Psi|r) = p(\alpha, \beta|r)p(\gamma)p(\Psi),$$

where

$$\begin{aligned}
p(\alpha, \beta|r) &\propto \exp(-\frac{1}{2}\text{tr}(\tilde{D}^{-1}\alpha\beta'\tilde{S}^{-1}\beta\alpha')),\\
p(\gamma) &\propto \exp(-\frac{1}{2}(\gamma - \tilde{v})'\tilde{V}^{-1}(\gamma - \tilde{v})),\\
p(\Psi) &\propto |\Psi|^{-\frac{K+1}{2}},
\end{aligned}$$

where $\gamma = \text{vec}(\Gamma)$ and $\beta'\beta = I_r$ is used as normalization. The proof is based on combining the derivations from Kadiyala and Karlsson[1997] and Koop *et al.*[2010].

To derive the first two full conditional density functions we assume Γ and Ψ are fixed and known. Denote $Y = Z_d - X\Gamma$, then rewrite the model as $Y = Z_l\beta\alpha' + E$ where $E \sim N_{T\times K}(0, I_T, \Psi)$. Then

• full conditional density function $(\alpha|\beta, \Gamma, \Psi, data)$

$$
\begin{aligned}
&p(\alpha|\beta, \Gamma, \Psi, data)\\
&\propto \exp\Big(-\frac{1}{2}\text{tr}(\Psi^{-1}(Y - Z_l\beta\alpha')'(Y - Z_l\beta\alpha'))\Big)\exp\Big(-\frac{1}{2}\text{tr}(\tilde{D}^{-1}\alpha\beta'\tilde{S}^{-1}\beta\alpha')\Big)\\
&\propto \exp\Big(-\frac{1}{2}(\text{vec}((Y - Z_l\beta\alpha')'))'(I\otimes\Psi^{-1})\text{vec}((Y - Z_l\beta\alpha')')\\
&\quad + \text{vec}(\alpha)'(\beta'\tilde{S}^{-1}\beta\otimes\tilde{D}^{-1})\text{vec}(\alpha))\Big)\\
&\propto \exp\Big(-\frac{1}{2}(-2\text{vec}(\alpha)'(\beta'Z_l'\otimes I)(I\otimes\Psi^{-1})\text{vec}(Y')\\
&\quad + \text{vec}(\alpha)'(\beta'Z_l'\otimes I)(I\otimes\Psi^{-1})(Z_l\beta\otimes I)\text{vec}(\alpha) + \text{vec}(\alpha)'(\beta'\tilde{S}^{-1}\beta\otimes\tilde{D}^{-1})\text{vec}(\alpha))\Big)\\
&\propto \exp\Big(-\frac{1}{2}(2\text{vec}(\alpha)'\underbrace{[\beta'Z_l'Z_l\beta\otimes\Psi^{-1} + \beta'\tilde{S}^{-1}\beta\otimes\tilde{D}^{-1}]}_{\bar{D}^{-1}}\text{vec}(\alpha)\\
&\quad - 2\text{vec}(\alpha)'\bar{D}^{-1}\underbrace{\bar{D}(\beta'Z_l'\otimes\Psi^{-1})\text{vec}(Y')}_{\bar{\alpha}} + \bar{\alpha}\bar{D}^{-1}\bar{\alpha})\Big)\\
&\propto \exp\Big(-\frac{1}{2}(\text{vec}(\alpha) - \bar{\alpha})'\bar{D}^{-1}(\text{vec}(\alpha) - \bar{\alpha})\Big).
\end{aligned}
$$

The second $\propto$ follows from Op.4) and Op.5), the third $\propto$ from $\text{vec}((Y - Z_l\beta\alpha')') = \text{vec}(Y') - (Z_l\beta\otimes I)\text{vec}(\alpha)$ by Op.5), the fourth $\propto$ from Op.3) and I also use $\bar{\alpha} = \bar{D}\text{vec}(\Psi^{-1}Y'Z_l\beta)$ by Op.5).

Hence, $\text{vec}(\alpha) \sim N(\bar{a}, \bar{D})$ where $\bar{D} = (\beta'Z_l'Z_l\beta\otimes\Psi^{-1} + \beta'\tilde{S}^{-1}\beta\otimes\tilde{D}^{-1})^{-1}$ and $\bar{a} = \bar{D}\text{vec}(\Psi^{-1}(Z_d - X\Gamma)'Z_l\beta)$.

• full conditional density function $(\beta|\alpha, \Gamma, \Psi, data)$
We shall use the equivalence of distribution of (α, β) under the prior $p(\alpha, \beta|r)$ with the distribution of (α, β) where $\alpha = A(B'B)^{1/2}$ and $\beta = B(B'B)^{-1/2}$,

under $p(A, B|r) \propto \exp(-\frac{1}{2}\mathrm{tr}(\tilde{S}^{-1}BA'\tilde{D}^{-1}AB'))$ where $A'A = I_r$.

$$
\begin{aligned}
&p(B|A, \Gamma, \Psi, data)\\
&\propto \exp\Big(-\frac{1}{2}\mathrm{tr}(\Psi^{-1}(Y - Z_lBA')'(Y - Z_lBA'))\Big)\exp\Big(-\frac{1}{2}\mathrm{tr}(\tilde{S}^{-1}BA'\tilde{D}^{-1}AB')\Big)\\
&\propto \exp\Big(-\frac{1}{2}(\mathrm{vec}((Y - Z_lBA')')'(I\otimes\Psi^{-1})\mathrm{vec}((Y - Z_lBA')')\\
&\quad + \mathrm{vec}(B)'(A'\tilde{D}^{-1}A\otimes\tilde{S}^{-1})\mathrm{vec}(B))\Big)\\
&\propto \exp\Big(-\frac{1}{2}(-2\mathrm{vec}(B)'K'_{Kr}(Z'_l\otimes A')(I\otimes\Psi{-1})K_{TK}\mathrm{vec}(Y)\\
&\quad + \mathrm{vec}(B)'K'_{Kr}(Z'_l\otimes A)(I\otimes\Psi^{-1})(Z_l\otimes A)K_{Kr}\mathrm{vec}(B)\\
&\quad + \mathrm{vec}(B)'(A'\tilde{D}^{-1}A\otimes\tilde{S}^{-1})\mathrm{vec}(B))\Big)\\
&\propto \exp\Big(-\frac{1}{2}(\mathrm{vec}(B)'\underbrace{[A'\Psi^{-1}A\otimes Z'_lZ_l + A'\tilde{D}^{-1}A\otimes\tilde{S}^{-1}]}_{\bar{S}^{-1}}\mathrm{vec}(B)\\
&\quad - 2\mathrm{vec}(B)'\bar{S}^{-1}\underbrace{\bar{S}(A'\Psi^{-1}\otimes Z'_l)\mathrm{vec}(Y)}_{\bar{b}} + \bar{b}'\bar{S}^{-1}\bar{b})\Big)\\
&\propto \exp\Big(-\frac{1}{2}(\mathrm{vec}(B) - \bar{b})'\bar{S}^{-1}(\mathrm{vec}(B) - \bar{b})\Big).
\end{aligned}
$$

The third $\propto$ follows from $\mathrm{vec}((Y - Z_lBA')') = \mathrm{vec}(Y') - (Z_l\otimes A)\mathrm{vec}(B') = \mathrm{vec}(Y') - (Z_l\otimes A)K_{Kr}\mathrm{vec}(B)$ by the definition of the commutation matrix. The fourth $\propto$ follows from $K'_{Kr}(Z'_l\otimes A')(I\otimes\Psi{-1})K_{TK} = (A'\Psi^{-1}\otimes Z'_l)$ by Op.1) and Op.3), and $K'_{Kr}(Z'_l\otimes A)(I\otimes\Psi^{-1})(Z_l\otimes A)K_{Kr} = (A'\Psi^{-1}A\otimes Z'_lZ_l)$ similarly; also here $\bar{b} = \bar{S}(A'\Psi^{-1}\otimes Z'_l)\mathrm{vec}(Y) = \bar{S}\mathrm{vec}(Z'_lY\Psi^{-1}A)$ by Op.5).

Therefore, $\beta = B(B'B)^{-1/2}$ where $\mathrm{vec}(B) \sim N(\bar{b}, \bar{S})$ for $\bar{S} = (A'\Psi^{-1}A\otimes Z'_lZ_l + A'\tilde{D}^{-1}A\otimes\tilde{S}^{-1})^{-1}$ and $\bar{b} = \bar{S}\mathrm{vec}(Z'_l(Z_d - X\Gamma)\Psi^{-1}A)$ $(A = \alpha(\alpha'\alpha)^{-1/2})$.

To derive the last two full conditional density functions we assume α and β are fixed and known. Denote $Y = Z_d - Z_l\beta\alpha'$, then rewrite the model as $Y = X\Gamma + E$ where $E \sim N_{T\times K}(0, I_T, \Psi)$, finally vectorize it to obtain $y = (I_K\otimes X)\gamma + e$ where $e \sim N(0, \Psi\otimes I_T)$ (and $y = \mathrm{vec}(Y)$, $\gamma = \mathrm{vec}(\Gamma)$ and

$e = \mathrm{vec}(E)$). Then the likelihood is

$$
\begin{aligned}
&L(\gamma, \Psi|\alpha, \beta, data)\\
&\propto |\Psi \otimes I_T|^{-1/2} \exp\left(-\frac{1}{2}(y-(I_K \otimes X)\gamma)'(\Psi \otimes I_T)^{-1}(y-(I_K \otimes X)\gamma)\right)\\
&\propto |\Psi|^{-T/2} \exp\big(-\frac{1}{2}((y-(I\otimes X)\hat{\gamma})'(\Psi^{-1}\otimes I)(y-(I\otimes X)\hat{\gamma})\\
&\quad -2(y-(I\otimes X)\hat{\gamma})'(\Psi^{-1}\otimes I)(I\otimes X)(\gamma-\hat{\gamma})\\
&\quad +(\gamma-\hat{\gamma})'(I\otimes X')(\Psi^{-1}\otimes I)(I\otimes X)(\gamma-\hat{\gamma}))\big)\\
&\propto |\Psi|^{-T/2} \exp\left(-\frac{1}{2}(\gamma-\hat{\gamma})'(\Psi^{-1}\otimes X'X)(\gamma-\hat{\gamma}) - \frac{1}{2}\mathrm{tr}(\Psi^{-1}\hat{E}'\hat{E}) + \mathrm{tr}(\hat{E}X(\Gamma-\hat{\Gamma}\Psi^{-1})\right)\\
&\propto p_N(\gamma|\hat{\gamma}, \Psi\otimes(X'X)^{-1}) \cdot p_{iW}(\Psi|\hat{E}'\hat{E}, T-\bar{K}-K-1),
\end{aligned}
$$

where $\bar{K} = K_d + KP$, $\hat{E} = (Y - X\hat{\Gamma})$ where $\hat{\Gamma}$ is the OLS estimator of Γ, $\hat{\Gamma} = (X'X)^{-1}X'Y$ and $\hat{\gamma} = \mathrm{vec}(\hat{\Gamma})$. The second $\propto$ follows from Op.2). The third $\propto$ follows because $y-(I\otimes X)\hat{\gamma} = \mathrm{vec}(\hat{E})$ and $(y-(I\otimes X)\hat{\gamma})'(\Psi^{-1}\otimes I)(y-(I\otimes X)\hat{\gamma}) = \mathrm{vec}(\hat{E})(\Psi^{-1}\otimes I)\mathrm{vec}(\hat{E}) = \mathrm{tr}(\Psi^{-1}\hat{E}'\hat{E})$ by Op.4), and similarly $(y-(I\otimes X)\hat{\gamma})'(\Psi^{-1}\otimes I)(I\otimes X)(\gamma-\hat{\gamma}) = \mathrm{tr}(\hat{E}'X(\Gamma-\hat{\Gamma})\Psi^{-1})$ by Op.3) and Op.4); Op.3) was also used to simplify matrix products of Kronecker products. The last $\propto$ follows because $\hat{E}X = 0$ by classic argument of least-squares. Then, using similar arguments as above, we obtain

- full conditional density function $(\Gamma|\alpha, \beta, \Psi, data)$

By conditioning $p(\gamma, \Psi|\alpha, \beta, data)$ on Ψ

$$
\begin{aligned}
&p(\gamma|\Psi, \alpha, \beta, data)\\
&\propto p_N(\gamma|\hat{\gamma}, \Psi\otimes(X'X)^{-1})p_N(\gamma|\tilde{v}, \tilde{V})\\
&\propto \exp\left(-\frac{1}{2}(\gamma'\underbrace{((\Psi^{-1}\otimes X'X)+\tilde{V}^{-1})}_{\bar{V}^{-1}}\gamma - 2\gamma'\bar{V}^{-1}\underbrace{\bar{V}((\Psi^{-1}\otimes X'X)\hat{\gamma}+\tilde{\Sigma}^{-1}\tilde{v})}_{\bar{v}}\right)\\
&\propto \exp\left(-\frac{1}{2}(\gamma-\bar{v})'\bar{V}^{-1}(\gamma-\bar{v})\right).
\end{aligned}
$$

Hence, $\gamma \sim N(\bar{v}, \bar{V})$ where $\bar{V} = (\Psi^{-1}\otimes X'X + \tilde{V}^{-1})^{-1}$ and $\bar{v} = \bar{V}((\Psi^{-1}\otimes X'X)\hat{\gamma} + \tilde{V}^{-1}\tilde{v})$ for $\hat{\gamma} = \mathrm{vec}(\hat{\Gamma})$.

- full conditional density function $(\Psi|\alpha, \beta, \Gamma, data)$

By conditioning $p(\gamma, \Psi|\alpha, \beta, data)$ on γ

$$
\begin{aligned}
&p(\Psi|\gamma, \alpha, \beta, data)\\
&\propto p_N(\gamma|\hat{\gamma}, \Psi \otimes (X'X)^{-1}) p_{iW}(\Psi|\hat{E}'\hat{E}, T-\bar{K}-K-1)|\Psi|^{-\frac{K+1}{2}}\\
&\propto |\Psi^{-1} \otimes X'X|^{1/2} \exp\left(-\frac{1}{2}(\gamma-\hat{\gamma})'(\Psi^{-1} \otimes X'X)(\gamma-\hat{\gamma})\right)\\
&\qquad \cdot |\Psi|^{-\frac{(T-\bar{K}-K-1)+K+1}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}(\Psi^{-1}\hat{E}'\hat{E})\right)|\Psi|^{-\frac{K+1}{2}}\\
&\propto |\Psi|^{-\frac{T+K+1}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}(\Psi^{-1}(\underbrace{(\Gamma-\hat{\Gamma})'X'X(\Gamma-\hat{\Gamma})+\hat{E}'\hat{E}}_{\bar{\Psi}}))\right).
\end{aligned}
$$

Hence, $\Psi \sim iW(\bar{\Psi}, T)$ where $\bar{\Psi} = (\Gamma-\hat{\Gamma})'X'X(\Gamma-\hat{\Gamma}) + \hat{E}'\hat{E}$, where $\hat{E} = Z_d - Z_l\beta\alpha - X\hat{\Gamma}$ and $\hat{\Gamma} = (X'X)^{-1}X'(Z_d - Z_l\beta\alpha')$.

C Appendix. Additional results for the empirical example

Table 3: Relative $\ln |E_h|$, fixed scheme of forecasting

(a) benchmark: no change forecast

h=1	Ψ-fixed	Ψ-diffuse
noCI	0.9273	1.0341
CI-priors	0.9078	1.0092
h=4	Ψ-fixed	Ψ-diffuse
noCI	0.9556	1.0175
CI-priors	0.9474	1.0035
h=8	Ψ-fixed	Ψ-diffuse
noCI	0.9497	0.9864
CI-priors	0.9404	0.9749

(b) benchmark: univariate AR

h=1	Ψ-fixed	Ψ-diffuse
noCI	0.9692	1.0808
CI-priors	0.9488	1.0548
h=4	Ψ-fixed	Ψ-diffuse
noCI	0.9898	1.0539
CI-priors	0.9814	1.0394
h=8	Ψ-fixed	Ψ-diffuse
noCI	1.0141	1.0533
CI-priors	1.0041	1.0410

(c) benchmark: VAR

h=1	Ψ-fixed	Ψ-diffuse
noCI	0.6454	0.7197
CI-priors	0.6318	0.7024
h=4	Ψ-fixed	Ψ-diffuse
noCI	0.7843	0.8351
CI-priors	0.7776	0.8236
h=8	Ψ-fixed	Ψ-diffuse
noCI	0.8780	0.9119
CI-priors	0.8693	0.9012

(d) benchmark: pure L-prior

h=1	Ψ-fixed	Ψ-diffuse
noCI	1.0259	1.1441
CI-priors	1.0044	1.1166
h=4	Ψ-fixed	Ψ-diffuse
noCI	1.0091	1.0745
CI-priors	1.0005	1.0597
h=8	Ψ-fixed	Ψ-diffuse
noCI	1.0184	1.0578
CI-priors	1.0084	1.0453

Table 4: Relative $\ln|E_h|$, rolling scheme of forecasting

(a) benchmark: no change forecast

h=1	Ψ-fixed	Ψ-diffuse
noCI	0.8389	0.9109
CI-priors	0.8232	0.8879
h=4	Ψ-fixed	Ψ-diffuse
noCI	0.8947	0.9664
CI-priors	0.8935	0.9519
h=8	Ψ-fixed	Ψ-diffuse
noCI	0.9202	0.9763
CI-priors	0.9174	0.9581

(b) benchmark: univariate AR

h=1	Ψ-fixed	Ψ-diffuse
noCI	0.8691	0.9437
CI-priors	0.8528	0.9199
h=4	Ψ-fixed	Ψ-diffuse
noCI	0.8813	0.9519
CI-priors	0.8801	0.9376
h=8	Ψ-fixed	Ψ-diffuse
noCI	0.8912	0.9456
CI-priors	0.8885	0.9279

(c) benchmark: VAR

h=1	Ψ-fixed	Ψ-diffuse
noCI	0.6575	0.7139
CI-priors	0.6452	0.6959
h=4	Ψ-fixed	Ψ-diffuse
noCI	0.7689	0.8305
CI-priors	0.7679	0.8180
h=8	Ψ-fixed	Ψ-diffuse
noCI	0.8004	0.8492
CI-priors	0.7979	0.8333

(d) benchmark: pure L-prior

h=1	Ψ-fixed	Ψ-diffuse
noCI	1.0025	1.0886
CI-priors	0.9837	1.0611
h=4	Ψ-fixed	Ψ-diffuse
noCI	0.9901	1.0695
CI-priors	0.9888	1.0534
h=8	Ψ-fixed	Ψ-diffuse
noCI	0.9870	1.0472
CI-priors	0.9839	1.0276

Table 5: Relative $\ln|E_h|$, recursive scheme of forecasting

(a) benchmark: no change forecast

h=1	Ψ-fixed	Ψ-diffuse
noCI	0.8498	0.9078
CI-priors	0.8380	0.8937
h=4	Ψ-fixed	Ψ-diffuse
noCI	0.9056	0.9580
CI-priors	0.9027	0.9547
h=8	Ψ-fixed	Ψ-diffuse
noCI	0.9329	0.9779
CI-priors	0.9273	0.9795

(b) benchmark: univariate AR

h=1	Ψ-fixed	Ψ-diffuse
noCI	0.8839	0.9442
CI-priors	0.8717	0.9295
h=4	Ψ-fixed	Ψ-diffuse
noCI	0.8989	0.9509
CI-priors	0.8961	0.9477
h=8	Ψ-fixed	Ψ-diffuse
noCI	0.9118	0.9557
CI-priors	0.9063	0.9573

(c) benchmark: VAR

h=1	Ψ-fixed	Ψ-diffuse
noCI	0.6739	0.7199
CI-priors	0.6645	0.7087
h=4	Ψ-fixed	Ψ-diffuse
noCI	0.7721	0.8168
CI-priors	0.7697	0.8140
h=8	Ψ-fixed	Ψ-diffuse
noCI	0.8336	0.8738
CI-priors	0.8286	0.8752

(d) benchmark: pure L-prior

h=1	Ψ-fixed	Ψ-diffuse
noCI	0.9975	1.0655
CI-priors	0.9836	1.0489
h=4	Ψ-fixed	Ψ-diffuse
noCI	0.9920	1.0494
CI-priors	0.9889	1.0458
h=8	Ψ-fixed	Ψ-diffuse
noCI	0.9926	1.0405
CI-priors	0.9866	1.0421

References

Luc Bauwens, Michel Lubrano, and Jean-Francois Richard. *Bayesian Inference in Dynamic Econometric Models*. Oxford University Press, 1999.

Yasuko Chikuse. *Statistics on special manifolds*. Springer Verlag, 2003.

Thomas Doan, Robert B. Litterman, and Christopher A. Sims. Forecasting and conditional projection using realistic prior distribution. Technical report, 1986.

A.K. Gupta and D.K. Nagar. *Matrix variate distributions*. Chapman & Hall/CRC monographs and surveys in pure and applied mathematics. Chapman & Hall, 2000. ISBN 9781584880462.

K. Rao Kadiyala and Sune Karlsson. Numerical aspects of bayesian VAR-modeling. Working Paper Series in Economics and Finance 12, Stockholm School of Economics, March 1994.

K. Rao Kadiyala and Sune Karlsson. Numerical methods for estimation and inference in Bayesian VAR-models. *Journal of Applied Econometrics*, 12 (2):99–132, March-Apr 1997.

Gary Koop, Roberto León-González, and Rodney Strachan. Efficient posterior simulation for cointegrated models with priors on the cointegration space. *Econometric Reviews*, 29(2):224–242, 2010.

Gary Koop, Rodney Strachan, Herman van Dijk, and Mattias Villani. *Palgrave Handbook on Econometric: Volume 1 Econometric Theory*, chapter Bayesian Approaches to Cointegration. Basingstoke: MacMillan, 2006.

Robert B. Litterman. Techniques of forecasting using vector autoregressions. Technical report, 1979.

Robert B Litterman. Forecasting with bayesian vector autoregressions-five years of experience. *Journal of Business & Economic Statistics*, 4(1):25–38, January 1986.

Jun S. Liu. The collapsed gibbs sampler in bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association*, 89(427):958–966, 1994. ISSN 01621459.

Helmut Lütkepohl. *New introduction to multiple time series analysis*. Springer, 2006. ISBN 9783540262398.

Christopher A Sims. Macroeconomics and reality. *Econometrica*, 48(1):1–48, January 1980.

Christopher A. Sims, Daniel F. Waggoner, and Tao Zha. Methods for inference in large multiple-equation markov-switching models. *Journal of Econometrics*, 146(2):255–274, October 2008.

Christopher A. Sims and Tao Zha. Bayesian methods for dynamic multivariate models. *International Economic Review*, 39(4):949–68, November 1998.

Rodney W Strachan. Valid bayesian estimation of the cointegrating error correction model. *Journal of Business & Economic Statistics*, 21(1):185–95, January 2003.

Rodney W. Strachan and Herman K. Dijk. Bayesian model selection with an uninformative prior. *Oxford Bulletin of Economics and Statistics*, 65(s1):863–876, December 2003.

Rodney W. Strachan and Brett Inder. Bayesian analysis of the error correction model. *Journal of Econometrics*, 123(2):307–325, December 2004.

Mattias Villani. Bayesian reference analysis of cointegration. *Econometric Theory*, 21(02):326–357, April 2005.

Mattias Villani and Anders Warne. Monetary policy analysis in a small open economy using Bayesian cointegrated structural VARs? Working Paper Series 296, European Central Bank, December 2003.

Daniel F. Waggoner and Tao Zha. A Gibbs sampler for structural vector autoregressions. *Journal of Economic Dynamics and Control*, 28(2):349–366, November 2003.

Anders Warne. Bayesian inference in cointegrated VAR models with application to the demand for Euro area M3. Working Paper Series 692, European Central Bank, November 2006.

European University Institute
European University Press