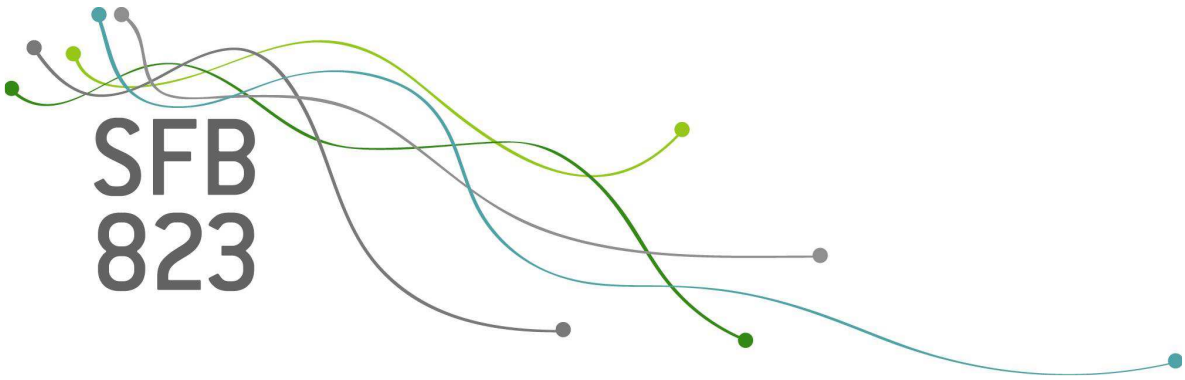
SFB 823

Bayesian analysis of reduced rank regression models using post-processing

Christian Aßmann, Jens Boysen-Hogrefe, Markus Pape

Nr. 13/2021

Discussion Paper



Bayesian analysis of reduced rank regression models using post-processing

Christian Aßmann*
Leibniz Institute for Educational Trajectories Bamberg, Germany
Chair of Survey Statistics and Data Analysis, Otto-Friedrich-Universität Bamberg, Germany

Jens Boysen-Hogrefe
Kiel Institute for the World Economy, Germany

Markus Pape
Department of Economics, Ruhr-Universität Bochum, Germany

June 23, 2021

Abstract

Bayesian estimation of reduced rank regression models requires careful consideration of the well known identification problem. We demonstrate that this identification problem can be handled efficiently by using prior distributions that restrict a part of the parameter space to the Stiefel manifold and post-processing the obtained Gibbs sampler output according to an appropriately specified loss function. This extends the possibilities for Bayesian inference in reduced rank regression models. Besides inference, we also discuss model selection in terms of posterior predictive assessment. We choose this approach because computing the marginal data likelihood under the identifying restrictions implies prohibitive computational burden. We illustrate the proposed approach with a simulation study and an empirical application.

JEL classification: C11; C31; C51; C52
Keywords: Bayesian Estimation; Reduced Rank Regression; Orthogonal Transformation; Model Selection; Stiefel Manifold; Posterior Predictive Assessment

Acknowledgments

We thank Roman Liesenfeld, Jörg Breitung, and participants of the statistics seminar of the University of Cologne as well as particpants of the ISBA 2018 poster session for helpful comments and suggestions provided on earlier versions of the paper. The first author acknowledges financial support by the Deutsche Forschungsgemeinschaft (DFG) within priority programme SPP 1646 under grants AS 368/3-1 and AS 368/3-2. The third author acknowledges support by the Collaborative Research Center 'Statistical modeling of nonlinear dynamic processes' (SFB 823, project A1) of the Deutsche Forschungsgemeinschaft (DFG).

*Corresponding author: C. Aßmann, https://orcid.org/0000-0002-4827-3150, Tel.: +49 951 863 3423, *E-mail address:* christian.assmann@uni-bamberg.de

1 Introduction

Bayesian analysis of reduced rank regression models is a well established tool in economics, psychology, or neuroscience, see Aguilar and West (2000), Chib et al. (2006),Woolley et al. (2010), Edwards (2010), or Sadtler et al. (2014). Reduced rank models can take various forms dealt with in the Bayesian context as discussed by Geweke (1996) or Başẗurk et al. (2017). Recent work covers several variants of reduced rank models, e.g. Man and Culpepper (2020), Chan et al. (2018), and Aßmann et al. (2016) deal with factor models, while Koop et al. (2010) discuss vector error correction models, and Zellner et al. (2014) show the link to models with instrumental variables.

Typically, reduced rank regression models need additional identifying restrictions to come up with interpretable empirical results. In certain setups this has troublesome consequences as the choice of ex-ante identifying restrictions can influence model evidence, see Chan et al. (2018), and the posterior distribution can exhibit multimodality, see Gelman and Rubin (1992), Lopes and West (2004), and Ročkov́a and George (2016). With regard to factor models as one prominent reduced rank regression model, multimodality can occur if identification is reached by constraining the loadings matrix to a positive lower triangular (PLT) matrix a priori as proposed by Geweke and Zhou (1996). As the constraints are imposed on particular elements of the loadings matrix, inference results may depend on the ordering of the variables. This is likewise observed by Carvalho et al. (2008). Altogether, the use of ex-ante identification via constraining the parameter space may influence inference results with respect to the model parameters and functions of these parameters. Hence, Chan et al. (2018) advise to refrain from this kind of identification. In this line, Aßmann et al. (2016) and Erosheva and Curtis (2017) propose ex-post approaches to achieve identified inference on factors and loadings.[1]

Likewise, in vector error correction models, where rank reduction is linked to the cointegration space, this cointegration space is only identified up to an arbitrary linear combination of the cointegration vectors. Several authors, e.g. Villani (2005), Kleibergen and van Dijk (1994), and Kleibergen and Paap (2002), suggest to enforce linear identifying restrictions a priori. Resulting Bayesian estimation is typically straightforward. However, enforcing linear restrictions a priori can induce estimation results depending on the ordering of the variables. To address this issue, several papers, e.g. Strachan (2003), Strachan and van Dijk (2003), and Strachan and Inder (2004), follow an alternative identification strategy related to the classical setup in Johansen (1988, 1991) and provide order-invariant Bayesian estimation approaches. However, Villani (2006) argues that point estimates based on the method in Strachan (2003) may provide counterintuitive interpretations and proposes an alternative ex-post point estimator for the cointegration space. However, this ex-post point estimator has the drawback that inference is only possible for the cointegration space as a whole and not for a specific cointegration vector. Thus, identifiying assumptions that allow for structural interpretations cannot

[1] Ex-post identification has also attracted wider use in the literature on finite mixture models, compare Celeux (1998), Celeux et al. (2000), and Stephens (2000). Ex-post identification can be motivated in terms of a decision-theoretic approach, see e.g. Stephens (2000), where a loss function is used to assess the difference between the parameter and the corresponding estimator.

be included. Further, Villani (2005) proposes a sampler where identification is not imposed by means of the prior distribution, however the reduced rank has to be taken into account. Finally, Koop et al. (2010) propose an approach that is order invariant and efficient. In this approach cointegrating vectors stem from the Stiefel manifold and Koop et al. (2010) also suggest a sampler that is easy to implement. Their approach solves the identification problem up to an orthogonal transformation.

In the light of the literature summarized above, the contribution of our paper is twofold. First, we develop an ex-post identification scheme for the sampler of Koop et al. (2010). By modifying the post-processing approach in Aßmann et al. (2016), we propose an identification scheme that allows to employ different additional identifying assumptions to reach inference for single cointegration vectors and not just the whole cointegration space. The suggested ex-post identification then allows for inference on the cointegrating vectors and extends the point estimator of Villani (2006). Second, we migrate the approach of Koop et al. (2010) to the class of static factor models. Modelling the factors on the Stiefel manifold ensures that the covariance of the factors is the identity matrix. Thus, it allows for the classical variance decomposition in factor analysis. We thereby extend the possibilities to perform ex-post identification in the context of reduced rank regressions involving identifying restrictions to the Stiefel manifold. Further, to avoid the prohibitively large computational burden involved in the calculation of marginal likelihoods for model selection, we propose the use of a posterior predictive assessment approach for model selection and discuss its performance by means of a simulation study for static factor models. The simulation study confirms the possibility to identify the reduced rank dimensionality correctly via posterior predictive assessment. We demonstrate the applicability of the suggested Bayesian approach in terms of an illustrative empirical example using exchange rate data of 22 currencies against the Euro. This empirical illustration highlights the capability of the suggested approach to obtain interpretable estimation results.

The paper proceeds as follows. Section 2 states the considered reduced rank regression models and discusses the involved identification issues. Section 3 provides the suggested ex-post approach towards identification in reduced rank regression models. Section 4 evaluates the suggested approach via simulation and numerical experiments. Section 5 provides details regarding the selection of rank order in terms of posterior predictive assessment. Section 6 provides an empirical illustration. Section 7 concludes.

2 Model setup, identification, and estimation

Following Geweke (1996), the reduced rank regression model setup including the vector error correction and the factor model can be stated as

$$Y = \Psi X + \Xi W + E, \tag{1}$$

where $Y = (y_1, \ldots, y_T)$ is the $P \times T$ matrix of dependent variables, $E = (e_1, \ldots, e_T)$ is the corresponding matrix of error terms with $\text{vec}(E)$ having multivariate normal distribution with mean zero and covariance $I_T \otimes \Sigma$, where Σ denotes a $P \times P$ covariance matrix.[2] Further, W is a $Q \times T$ matrix of explanatory variables with corresponding parameter matrix Ξ of size $P \times Q$. ΨX incorporates the reduced rank structure of the model, where Ψ denotes a $P \times J$ matrix and X a corresponding $J \times T$ matrix. The vector error correction model, see e.g. Villani (2006), arises when Y corresponds to first differences of observed variables, i.e. $Y = (\Delta y_1, \ldots, \Delta y_T)$ with $y_0 = \ldots = y_{2-K} = 0$ implying $\Delta y_1 = y_1$ and $\Delta y_0 = \ldots = \Delta y_{1-K} = 0$. Correspondingly, we have $X = (y_0, \ldots, y_{T-1})$ and W summarizing K lagged differences and exogenous variables, i.e.

$$W = \begin{pmatrix} \Delta y_0 & \cdots & \Delta y_{T-1} \\ \vdots & & \vdots \\ \Delta y_{1-K} & \cdots & \Delta y_{T-K} \\ Z_1 & \cdots & Z_T \end{pmatrix} \quad \text{and} \quad \Xi = (\Phi_1, \ldots, \Phi_K, \varphi),$$

where Z_t, $t = 1 \ldots, T$ denotes the vectors of exogenous variables each of dimension $M \times 1$. Hence, we have $J = P$ and $Q = KP + M$. A static factor model arises for $X = I_T$, $W = (Z_1, \ldots, Z_T)$ and $\Xi = \varphi$, $J = T$, and $Q = M$.[3]

The reduced rank structure is captured via decomposing $\Psi = \alpha\beta'$ with α denoting a parameter matrix of size $P \times R$ and β a parameter matrix of dimension $J \times R$, with $R \ll \min\{J, P\}$. In case of the vector error correction model α governs the adjustment back to equilibrium and $\beta' X$ denotes the T vectors of stationary departures from the R long run equilibria, while in case of the static factor model α denotes the matrix of factor loadings and β the matrix of factors. Then with

$$\Theta = (\text{vec}(\alpha)', \text{vec}(\beta)', \text{vec}(\Xi)', \text{vech}(\Sigma)')', \tag{2}$$

the resulting likelihood for both models is given as

$$\mathcal{L}(Y|\Theta, X, W) = \frac{|\Sigma|^{-\frac{T}{2}}}{(2\pi)^{\frac{TP}{2}}} \exp\left\{-\frac{1}{2}\text{tr}\left[\left(Y - \alpha\beta' X - \Xi W\right)' \Sigma^{-1} \left(Y - \alpha\beta' X - \Xi W\right)\right]\right\}. \tag{3}$$

For Σ and Ξ, we choose the commonly used conjugate priors as independent inverse Wishart and multivariate normal distributions with probability densities given as

$$\pi(\Sigma) \propto |\Omega_\Sigma|^{\frac{\nu_\Sigma}{2}} |\Sigma|^{-\frac{\nu_\Sigma + P + 1}{2}} \exp\left\{-\frac{1}{2}\text{tr}[\Omega_\Sigma \Sigma^{-1}]\right\}$$

[2] Note that in the following $\odot$ and $\otimes$ denote Hadamard and Kronecker tensor products as defined in Lütkepohl (1996) respectively, and ι and I denote a row vector of ones and an identity matrix of indicated size.

[3] In restricting the parameter space to the Stiefel manifold the model deviates from the static factor models e.g. applied in Aßmann et al. (2016) or Chan et al. (2018). The restriction ensures that the factors are uncorrelated and have unit variance. Without this restriction, the posterior distribution of the factors may exhibit correlation and the scaling depends on the prior variance of the loadings.

and

$$\pi(\Xi) \propto \exp\left\{-\frac{1}{2}\text{tr}[(\Xi-\mu_\Xi)\Omega_\Xi^{-1}(\Xi-\mu_\Xi)']\right\}.$$

The prior for α and β has to address the identification problem arising in factor and cointegration analysis, as for an invertible matrix D of dimension $R \times R$, we have

$$(\alpha D)(\beta D^{-1\prime})' = \alpha D D^{-1}\beta'$$

thus the likelihood is invariant under this transformation. One part of this identification problem is typically addressed via restricting the scaling of α or β, where restrictions on the scaling of β are prominent within the literature, see Villani (2005). We follow Koop et al. (2010) and Villani (2006) and restrict the scaling of β by assuming that $\beta'\beta = I_R$, i.e., β is semiorthogonal. The corresponding prior distribution is hence defined on the Stiefel manifold. In principle, the prior is constructed in relation to the free elements in α and β, where the number of free elements can be calculated as follows. With the semiorthogonality restriction imposed on β, i.e. $\beta'\beta = I_R$, it must hold that $\beta_{r,J} = \pm\sqrt{1-\sum_{j=1}^{J-1}\beta_{r,j}^2}$, where the sign is determined by the equality $\Psi = \alpha\beta'$. This reduces the number of free elements in β by R. Moreover, α and β can be replaced by $\alpha^* = \alpha D$ and $\beta^* = \beta D$. Then D can be represented as $D = E_D \cdot \text{diag}(\text{sgn}(l_{1,1}, \ldots, \text{sgn}(l_{R,R}))$, where E_D and L result from the QR-decomposition $\beta' = E_D L'$, see e.g. Golub and van Loan (2013), and hence, D is an orthogonal matrix. This reduces the number of free elements in β by another $R(R-1)/2$. Thus, the number of free elements in α is RP, and the number of free elements in β is $RJ - R(R+1)/2$. Koop et al. (2010) suggests to use a prior setup for α and β given as

$$\pi(\alpha,\beta|\Sigma) \propto |\Sigma|^{-\frac{R}{2}} \exp\left\{-\frac{1}{2}\text{tr}\left[\tau^{-1}\beta' C_\tau^{-1}\beta\alpha'\Sigma^{-1}\alpha\right]\right\}\mathcal{I}(\beta'\beta = I_R),$$

where $\mathcal{I}(\cdot)$ denotes the indicator matrix. This joint prior distribution relates to a marginal matrix angular central Gaussian distribution conditional on matrix C_τ for β and a conditional multivariate normal prior for α conditional on β with expected value zero and covariance matrix $\tau(\beta' C_\tau^{-1}\beta)^{-1} \otimes \Sigma$.[4] Although this prior distribution for α and β identifies the scaling of β, the so far implied posterior distribution

$$p(\Theta|Y,X,W) \propto \mathcal{L}(Y|\Theta,X,W)\pi(\Xi)\pi(\alpha,\beta|\Sigma)\pi(\Sigma),$$

remains invariant when restricting D to be an orthogonal matrix. To formalize, define for any

[4]Following Koop et al. (2010), C_τ is constructed as $C_\tau = CC' + \tau\tilde{C}\tilde{C}'$, where τ denotes a scaling hyperparameter and $\tilde{C}$ denotes the null space matrix of C, and $C = \hat{C}(\hat{C}'\hat{C})^{-.5}$ with elements of $\hat{C}$, say e.g. c, are each independently drawn from uniform density $\frac{1}{2}\mathcal{I}(-1 < c < 1)$. This ensures that C_τ is positive definite as required and provides the flexibility to possibly consider different a priori orientations and scalings. A discussion of the matrix angular central Gaussian distribution is provided in Chikuse (1990, 2003).

orthogonal $R \times R$ matrix D the transformation

$$\begin{aligned} H(D)\Theta &= (\text{vec}(\alpha D)', \text{vec}(\beta D)', \text{vec}(\Xi)', \text{vech}(\Sigma)')' \\ &= (\text{vec}(\tilde{\alpha})', \text{vec}(\tilde{\beta})', \text{vec}(\Xi)', \text{vech}(\Sigma)')' = \tilde{\Theta}, \end{aligned} \tag{4}$$

with

$$H(D) = \begin{pmatrix} (D' \otimes I_{P+J}) & 0 \\ 0 & I_{PQ+P(P+1)/2} \end{pmatrix}. \tag{5}$$

Taking into account that the transformation described in Equation (4) has no impact on the range of parameters and $d\tilde{\Theta} = |\det(H(D)^{-1})| d\Theta$ with $|\det((H(D)^{-1})| = 1$, also the marginal likelihood

$$\mathcal{M}(Y|X,W) = \int p(Y,\Theta|X,W)d\Theta = \int p(Y,\tilde{\Theta}|X,W)d\tilde{\Theta}, \tag{6}$$

is invariant and thus is the posterior distribution $p(\Theta|Y,X,W) = p(H(D)\Theta|Y,X,W)$.

The considered model setup allows for sampling from the posterior distribution via use of the Gibbs sampler, which we call the *unconstrained Gibbs sampler* in the following, because it only imposes scale restrictions on α and β. To facilitate efficient closed form sampling Koop et al. (2010) discuss a reparametrization using the transformation $A = \alpha(\alpha'\alpha)^{-\frac{1}{2}}$ and $B = \beta(\alpha'\alpha)^{\frac{1}{2}}$ with corresponding inverse functions given as $\alpha = A(B'B)^{\frac{1}{2}}$ and $\beta = B(B'B)^{-\frac{1}{2}}$ implying $\alpha\beta' = AB'$. This transformation yields

$$\pi(A,B|\Sigma) \propto |\Sigma|^{-\frac{R}{2}} \exp\left\{-\frac{1}{2}\text{tr}\left[\tau^{-1}A'\Sigma^{-1}AB'C_\tau^{-1}B\right]\right\} \mathcal{I}(A'A = I_R).$$

The considered reparametrization corresponds then to $\Theta_{AB} = (\text{vec}(A)', \text{vec}(B)', \text{vec}(\Xi)', \text{vech}(\Sigma)')'$ with corresponding posterior distribution

$$p(\Theta_{AB}|Y,X,W) \propto \mathcal{L}(Y|\Theta_{AB},X,W)\pi(\Xi)\pi(A,B|\Sigma)\pi(\Sigma). \tag{7}$$

Note that the posterior is also invariant when considering the reparametrization in terms of A and B and under any permutation of the P variables in Y, the corresponding rows of each α, β, Ξ, and the corresponding rows and columns of Σ, see also Appendix A. Following Koop et al. (2010) and considering the reparameterization in terms of A and B allowing for efficient sampling yields the following set of full conditional distributions. For the covariance matrix Σ we have

$$f(\Sigma|\cdot) \propto |\Sigma|^{-\frac{\varphi_\Sigma+P+R+1}{2}} \exp\left\{-\frac{1}{2}\text{tr}[\Psi_\Sigma\Sigma^{-1}]\right\},$$

with $\varphi_\Sigma = \nu_\Sigma + T - 1$ and $\Psi_\Sigma = \Omega_\Sigma + (Y - \alpha\beta'X - \Xi W)(Y - \alpha\beta'X - \Xi W)' + \frac{1}{\tau}\alpha\beta'C_\tau^{-1}\beta\alpha'$. For

the parameters Ξ, the full conditional distribution is given as

$$f(\Xi|\cdot) \propto \exp\left\{-\frac{1}{2}(\text{vec}(\Xi) - \varphi_\Xi)'\Psi_\Xi^{-1}(\text{vec}(\Xi) - \varphi_\Xi)\right\},$$

where $\Psi_\Xi = ((WW' \otimes \Sigma^{-1}) + \Omega_\Xi^{-1})^{-1}$ and $\varphi_\Xi = \Psi_\Xi(\text{vec}(\Sigma^{-1}(Y - \alpha\beta' X)W') + \Omega_\Xi^{-1}\mu_\Xi)$. For the full conditional of α we have

$$f(\alpha|\cdot) \propto \exp\left\{-\frac{1}{2}(\text{vec}(\alpha) - \varphi_\alpha)'\Psi_\alpha^{-1}(\text{vec}(\alpha) - \varphi_\alpha)\right\},$$

with $\Psi_\alpha = ((\beta' XX'\beta \otimes \Sigma^{-1}) + \frac{1}{\tau}(\beta' C_\tau^{-1}\beta \otimes \Sigma^{-1}))^{-1}$ and $\varphi_\alpha = \Psi_\alpha(\text{vec}(\Sigma^{-1}(Y - \Xi W)X'\beta))$. This draw of α is then transformed in $A = \alpha(\alpha'\alpha)^{-\frac{1}{2}}$. Given A, we have the full conditional of B as

$$f(B|\cdot) \propto \exp\left\{-\frac{1}{2}(\text{vec}(B) - \varphi_B)'\Psi_B^{-1}(\text{vec}(B) - \varphi_B)\right\},$$

with $\Psi_B = ((A'\Sigma^{-1}A \otimes X'X) + (A'\Sigma^{-1}A \otimes \frac{1}{\tau}C_\tau^{-1}))^{-1}$ and $\varphi_B = \Psi_B(\text{vec}(X'(Y - \Xi W)\Sigma^{-1}A))$. The draw of B is then transformed to obtain $\beta = B(B'B)^{-\frac{1}{2}}$. The prior hyperparameters used in estimation and simulation are documented in Table (1).

Given this sampling algorithm, a posterior sample can be established. However, this sample and the involved draws are subject to the identification invariance as described above. This makes estimation and inference feasible merely for quantities not subject to the identification problem, such as $\alpha\beta'$.[5] To obtain estimates for all other quantities, we propose the following post-processing procedure.

3 Solving the identification problem via post-processing

To address the identification problem arising in Bayesian analysis of reduced rank regressions, we adapt the ex-post identification approach by Aßmann et al. (2016) that introduces post-processing for factor models. In the following, we outline the post-processing approach we propose for reduced rank models with the parameter space of β restricted to the Stiefel manifold. The presentation is closely related to the one in Aßmann et al. (2016), since this is a modification of the approach for factor models. The modification is needed to guarantee that the result of the post-processing lies on the Stiefel manifold.

A loss function $L(\Theta^*, \Theta)$ is defined as a mapping of the estimators Θ^* from the set of possible estimators Ξ and each of the parameter values Θ within the parameter space on the real line, i.e. L :

[5] Villani (2006) proposes a point estimator for the cointegration vector β in the context of vector error correction models, which is the posterior mean cointegration space (PMCS) estimator. This estimator consists of the first R eigenvectors of $\widehat{\beta'\beta}$. This specific setup, however, does not allow for inference on single parameters. We use the PMCS estimator in our simulation studies in section 4 as a reference point.

$\Xi \times \Theta \to [0, \infty)$. The optimal estimator in terms of minimal expected loss is then defined as

$$\tilde{\Theta}^* = \arg\min_{\Theta^*} \int_{\Theta} L(\Theta^*, \Theta) p(\Theta|Y, X, W) d\Theta.$$

To solve the identification problem, we propose to extend the loss function approach in order to discriminate between invariant losses of estimators invoked under the transformation described in Equation (4) depending on an orthogonal matrix D. The extended loss function then takes the form

$$L(\Theta^*, \Theta) = \min_{D} \{L_D(\Theta^*, H(D)\Theta)\}, \quad \text{s.t. } D'D = I_R,$$

with $L_D(\Theta^*, H(D)\Theta)$ denoting for given Θ^* the loss invoked for any transformation of Θ as described in Equation (4).[6] As this minimization is done for each Θ, the parameter space of Θ is restricted ex-post via the corresponding first order conditions. As recommended by Larsson and Villani (2001), we use

$$L_D(\Theta^*, H(D)\Theta) \quad = \quad (H(D)\Theta - \Theta^*)'(H(D)\Theta - \Theta^*)$$

for the considered reduced rank model specifications.[7] Since the integral involved in the expected posterior loss is approximated via Monte Carlo (MC) methods the corresponding minimization problem takes the form

$$\{\{\tilde{D}^{(s)}\}_{s=1}^{S}, \tilde{\Theta}^*\} = \arg \min_{\{D^{(s)}\}_{s=1}^{S}, \Theta^*} \sum_{s=1}^{S} L_D(\Theta^*, H(D^{(s)})\Theta^{(s)}), \tag{8}$$

subject to $\beta^{*\prime}\beta^* = I_R$ and $D^{(s)\prime}D^{(s)} = I_R$, $s = 1, \ldots, S$, where $\Theta^{(s)}$, $s = 1, \ldots, S$ denotes a sample from the unconstrained posterior distribution. Note that all samples taking the form $\{H(D^{(s)})\Theta^{(s)}\}_{s=1}^{S}$ for arbitrary sequences of orthogonal matrices $\{D^{(s)}\}_{s=1}^{S}$ have the same posterior probability. For all elements of Θ^* not referring to β, the estimator implied by the defined loss function takes the form of

$$\overline{H(D)\Theta} = \frac{1}{S} \sum_{s=1}^{S} H(D^{(s)})\Theta^{(s)}. \tag{9}$$

For β, the elements of $\overline{H(D)\Theta}$ do not fulfill the restriction $\beta^{*\prime}\beta^* = I_R$. This restriction turns the minimization of the posterior expected loss with regard to β into an orthogonal Procrustes problem.

[6] Note that this extension corresponds to the choice of an optimal permutation that is proposed in the aforementioned relabeling literature, see e.g. Jasra et al. (2005), Section 5.1.

[7] Note that under general regularity conditions, see Cheng et al. (1999), the suggested loss function is first-order equivalent to the Kullback-Leibler distance, see Clarke et al. (1990) for a discussion of corresponding properties. Further, the suggested loss function corresponds to the square of the Frobenius norm.

The defined loss function implies

$$\arg\min_{\beta^*} \operatorname{tr}\left[-\beta^* \sum_{s=1}^{S} \beta^{(s)} D^{(s)}\right] \quad \text{s.t.} \quad \beta^{*\prime}\beta^* = I_R. \tag{10}$$

The structure of an orthogonal Procrustes problem arises also with regard to determining $D^{(s)}$, as the corresponding minimization problem arising from Equation (8) takes the form

$$\arg\min_{D^{(s)}} \operatorname{tr}[(\bar{\Lambda}^{(s)} D - \bar{\Lambda}^*)'(\bar{\Lambda}^{(s)} D - \bar{\Lambda}^*)], \quad \text{s.t.} \quad D^{(s)\prime} D^{(s)} = I_R, \tag{11}$$

with $\bar{\Lambda}^*$ denoting the estimator of the stacked matrix $\bar{\Lambda} = (\alpha', \beta')'$ and $\bar{\Lambda}^{(s)}$ denoting a draw of $\bar{\Lambda}$ from the unconstrained sampler. The post-processing approach as discussed here transforms the output from the unconstrained sampler given a fixed point, i.e. the estimator. The following paragraph outlines how a solution for the vector error correction model can be obtained via a sequential algorithm. The algorithm needs an initialization with regard to Θ^*, where we choose the last draw of the unconstrained sampler for convenience.[8]

Step 1 For given Θ^* the minimization problem implied by Equation (11) resembles the orthogonal Procrustes problem discussed by Kristof (1964) and Schönemann (1966), see also Golub and van Loan (2013). The solution involves the following calculations.

1.1 Define $\Upsilon_{D^{(s)}} = \bar{\Lambda}^{(s)\prime}\bar{\Lambda}^*$.

1.2 Do the singular value decomposition $\Upsilon_{D^{(s)}} = U_{D^{(s)}} M_{D^{(s)}} V'_{D^{(s)}}$, where $U_{D^{(s)}}$ and $V_{D^{(s)}}$ denote the matrix of eigenvectors of $\Upsilon_{D^{(s)}}\Upsilon'_{D^{(s)}}$ and $\Upsilon'_{D^{(s)}}\Upsilon_{D^{(s)}}$, respectively, and $M_{D^{(s)}}$ denotes a diagonal matrix of singular values, which are the square roots of the eigenvalues of $\Upsilon_{D^{(s)}}\Upsilon'_{D^{(s)}}$ and $\Upsilon'_{D^{(s)}}\Upsilon_{D^{(s)}}$. Note that the nonzero eigenvalues of $\Upsilon_{D^{(s)}}\Upsilon'_{D^{(s)}}$ and $\Upsilon'_{D^{(s)}}\Upsilon_{D^{(s)}}$ are identical.

1.3 Obtain the orthogonal transformation matrix $D^{(s)} = U_{D^{(s)}} V'_{D^{(s)}}$.

For further details on the derivation of this solution, see Schönemann (1966).

Step 2 Choose α^*, Φ^*, and Σ^* as implied by $\overline{H(D)\Theta}$. With regard to β^*, the minimization problem also takes the form of an orthogonal Procrustes problem, where the solution then involves the following calculations.

2.1 Define $\mathcal{S}_\beta = \sum_{s=1}^{S} \beta^{(s)} D^{(s)}$.

2.2 Do the singular value decomposition $\mathcal{S}_\beta = U_\beta M_\beta V'_\beta$, where U_β denotes the matrix of the eigenvectors of $\mathcal{S}_\beta\mathcal{S}'_\beta$ corresponding to the R largest eigenvectors, and V_β denotes the matrix of eigenvectors arising from $\mathcal{S}'_\beta\mathcal{S}_\beta$. Further, M_β denotes a diagonal matrix of

[8] Note that the first step is the same as in Aßmann et al. (2016). Step 2 entails the modification to cope with the restriction on the Stiefel manifold.

singular values, which are the square roots of the eigenvalues of $\mathcal{S}_\beta \mathcal{S}_\beta'$ and $\mathcal{S}_\beta' \mathcal{S}_\beta$. Note that the largest R eigenvalues of $\mathcal{S}_\beta \mathcal{S}_\beta'$ and $\mathcal{S}_\beta' \mathcal{S}_\beta$ are identical.

2.3 Obtain the semiorthogonal matrix $\beta^* = U_\beta V_\beta'$.

With regard to convergence of the post-processing algorithm, we have found that for arbitrary initial choices of Θ^* taken from the unconstrained sampler output, less than ten iterations usually suffice to achieve convergence to a fixed point Θ^* providing the Bayes estimator. Convergence is assumed if the sum of squared deviations between two successive Θ^* does not exceed a predefined threshold value, where we use 10^{-9}. The iterative procedure of the algorithm suggests to use the transformed output of the unconstrained sample, i.e. $H(D^{(s)})\Theta^{(s)}$, as input for the next iteration, thus, reducing required computer memory capacities. In this line, Vito et al. (2018) have found that also a single iteration may result in estimates sufficiently close to the final estimates. The post processed posterior sample provides then the basis to calculate posterior summary statistics including uncertainty measures allowing for inference. Note that all estimation and simulation routines have been implemented in MATLAB® and are available from the authors upon request.

4 Simulation and numerical experiments

To illustrate the properties of the suggested identification approach, we perform a simulation experiment for a vector error correction model. We use the posterior mean cointegration space (PMCS) estimator proposed by Villani (2006) to provide an orientation for the directed inference. We show that the point estimator is extremely close to the PMCS, while our approach also provides posterior distributions for each parameter.

The results obtained from the simulation experiment are based on the following setup. We simulate $T = 500$ observations following a vector error correction model with $R = 2$ cointegrating vectors for $P = 4$ variables. Moreover, we set $K = 3$. The values used within the data generating process (DGP) with regard to $\alpha\beta'$ and thus α and β are given in the first column of Table (2) and Table (3) respectively. To obtain a sample from the posterior distribution of α and β, we run the Koop et al. (2010) sampler with $S = 20000$ after a burn-in phase of 5000 iterations, incorporating the scale restriction $\beta\beta' = I_R$.

Figure 1 illustrates the identification problem by showing the circular shape of the posterior distribution when plotting pairwise parameter trajectories arising from the unconstrained Koop et al. (2010) sampler. The β matrices are semiorthogonal, so their columns have unit length. Note, however, that their rows vary in length. There is therefore some variability in the length of the plotted draws in terms of row vectors. A plot of the draws in terms of column vectors would make the semi-orthogonality property obvious, however, it would have to be four-dimensional. We therefore show the distribution of the lengths of the row and column vectors of β, respectively, in Figure 2. Note that these quantities are quantities that are invariant under the transformation described in

Equation (4), and that the distribution of the lengths of the column vectors of β is degenerate due to the incorporated scale restriction $\beta'\beta = I_R$.

Since the quantity $\alpha\beta'$ is invariant under the aforementioned transformation, we can use the output from the unconstrained Koop et al. (2010) sampler directly without post-processing for inference on this quantity. The second column of Table (2) shows the point estimates for $\alpha\beta'$ based on the output of the Koop et al. (2010) sampler, which is quite accurate. The third column shows the product of the point estimate for α and the transpose of the point estimate of β. We observe a substantial deviation between this result and the estimate for the invariant quantity $\alpha\beta'$ if the output from the sampler without post-processing is used. This reflects the equally imprecise estimates for α and β, which are not separately reported.

The effects of post-processing the output of the unconstrained sampler can also be seen by looking at the shape of the posterior distribution of β. The lower panels of Figure 1 indicate that the posterior distributions of the rows are no longer circular. They also allow for proper inference on the elements of α and β. If the output is post-processed, the estimate for the invariant quantity $\alpha\beta'$ stays the same, as can be seen in the fourth column of Table (2). The product of the point estimates for α and β, however, shown in the last column of Table (2), is now almost identical to the invariant estimate. This, in turn, indicates that the estimates for α and β must be much more precise than if no post-processing is applied.

Note that the estimates for α and β, and indeed the entire samples from the respective posterior distributions, can be transformed by a single orthogonal matrix to satisfy identifying assumptions. The point estimates and highest posterior density intervals (HPDIs) in Table (3) show the results of two different treatments of the same post-processed output in that respect. In the first five columns of Table (3), we choose the rotation that minimizes the Frobenius norm of the distance between the estimates and the parameters α and β that were used to simulate the data set. In the last five columns of Table (3), on the other hand, we choose the rotation that minimizes the Frobenius norm of the distance between the estimates and the PMCS estimate for β obtained from the approach by Villani (2006). Conversely, the PMCS estimate reported in the second column of Table (3) has been subject to the inverse rotation, in order to make it comparable to the remaining results.

The resulting point estimates for α and β are given in the third and eighth column of Table (3). In the latter column, the point estimate for β from the proposed post-processing approach is identical to the PMCS estimate. This happens because the singular value decomposition that is used to obtain the fix point β^* in the post-processing approach yields singular vectors that are identical up to an orthogonal transformation to the eigenvectors obtained from $\widehat{\beta'\beta}$ in the PMCS approach of Villani (2006). It is therefore possible to find an orthogonal transformation in the two dimensional real space that maps our estimate for β exactly onto the PMCS estimate. In the second column of Table (3), on the other hand, the rotated PMCS estimate differs from the rotated estimate from the post-processing approach, because the optimal rotation for the PMCS estimate only takes β into account, whereas the post-processing approach takes both α and β into account. To sum up, the

point estimates for β obtained from the post-processing approach are - depending on the chosen reference point - very similar or identical to the PMCS estimate. The post-processing approach, however, also provides point estimates for α and provides samples from the posterior distributions of both α and β, thus allowing for proper inference, as illustrated by the reported 84% HPDIs for all elements of α and β in Table (3).

5 Model selection via posterior predictive assessment

In a Bayesian context, model selection and specification is conceptually straightforward in terms of the marginal model likelihood $\mathcal{M}(Y|X,W)$ stated in Equation 6, see Chib (1995) and Kass and Raftery (1995). However, for the considered reduced rank regression model framework the use of the marginal likelihood for specifying the reduced rank dimension relating to the number of latent factors or cointegrating vectors on the Stiefel manifold, involves computational difficulties that prevent use the marginal likelihood for model selection purposes in the context of reduced rank regression models. Typically, the computation of the marginal likelihood is based on the full conditional distributions including the corresponding normalizing constants, see Chib (1995) and Chib and Jeliazkov (2001). As the functional form of the involved full conditional distribution for β is given as a Bingham-von Mises-Fisher distribution, as discussed by Chikuse (2003), Gupta and Nagar (2000) and Hoff (2009), the integrating constant required for computation of the marginal model likelihood involves Hayakawa polynomials, see Mathai et al. (1995) and Crowther (1975), or the hypergeometric function with matrix argument, see Herz (1955) and Koev and Edelman (2006). However, the analytical calculation is non trivial and the saddlepoint approximation suggested by Kume et al. (2013) generalizing the work of Butler and Wood (2003) and Kume (2005), does not provide sufficient numerical precision for typical dimensions relevant in application contexts. The same holds for alternative numerical approaches as power posterior sampling, see Friel and Pettitt (2008), as a version of thermodynamic integration closely related to annealed importance sampling, see Neal (2001), bridge sampling, see Meng and Wong (1996), and path sampling, see Gelman and Meng (1998).

Taking the aforementioned difficulties into account, we propose to make use of posterior predictive assessment, see Gelman et al. (1996), to perform model selection with regard to the dimensionality of the reduced rank structure. For this purpose, a defined fraction of Y, e.g. within the range from 1% to 10%, are discarded from the data, hence we can partition the data in discarded (Y^{DIS}) and remaining (Y^{REM}) observations. The partition implies $y_t^{\text{DIS}} = L_t^{\text{DIS}} y_t$ and $y_t^{\text{REM}} = L_t^{\text{REM}} y_t$ for all $t = 1\ldots,T$, where L_t^{DIS} and L_t^{REM}, $t = 1,\ldots,T$ denote appropriately defined elimination matrices. The discarded observations are then augmented to the parameter vector and subject to sampling within the Gibbs sampling algorithm. Posterior predictive assessment is then based on extending the Gibbs sampling scheme with the full conditional distributions of the discarded observations Y^{DIS}. For the factor model setup, this set of full conditional distributions is directly arising from the likelihood function given in Equation (3). Since the likelihood in the static factor model setup (F) can be

factored as

$$\mathcal{L}_{\mathrm{F}}(Y|\Theta, X, W) = \prod_{t=1}^{T} f_{\mathrm{F}}(y_t|\Theta, Z_t),$$

the corresponding posterior predictive distribution is given as

$$f_{\mathrm{F}}(Y^{\mathrm{DIS}}|Y^{\mathrm{REM}}, \Theta, X, W) = \prod_{t=1}^{T} f(y_t^{\mathrm{DIS}}|y_t^{\mathrm{REM}}, \Theta, Z_t)$$

where $f(y_t^{\mathrm{DIS}}|y_t^{\mathrm{REM}}, \Theta, Z_t)$ is given as multivariate normal as implied by multivariate normal distribution theory.

For the vector error correction setup (VECM), sampling from the set of full conditional distributions of the discarded observation values is more elaborate. First, the vector error correction model is reformulated as a vector autoregressive model in levels y_t, i.e.,

$$y_t = (I + \alpha\beta' + \Phi_1)y_{t-1} + \sum_{k=1}^{K}(\Phi_{k+1} - \Phi_k)y_{t-k} + e_t,$$

where $\Phi_{K+1} = 0$. The corresponding state space representation has $y_t^{\mathrm{DIS}} = L_t^{\mathrm{DIS}} y_t$ as the measurement equation, whereas the corresponding transition equation is given by

$$\tilde{Y}_t = \Gamma\tilde{Y}_{t-1} + \tilde{Z}_t\phi + \tilde{E}_t,$$

where $\tilde{Y}_t = (y_t, y_{t-1}, \ldots, y_{t-K})'$,

$$\Gamma = \begin{pmatrix} I_P + \alpha\beta' + \Phi_1 & \Phi_2 - \Phi_1 & \Phi_3 - \Phi_2 & \ldots & -\Phi_K \\ I_P & 0 & 0 & \ldots & 0 \\ 0 & I_P & 0 & & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \ldots & 0 & I_P & 0 \end{pmatrix},$$

$\tilde{Y}_{t-1} = (y_{t-1}, y_{t-2}, \ldots, y_{t-K-1})'$, $\tilde{Z}_t = (Z_t, 0, \ldots 0)'$, and $\tilde{E}_t = (e_t, 0, \ldots, 0)'$. A sample of all discarded values Y^{DIS} in the vector error correction context can then be obtained by iteratively sampling from the set of full conditional distributions of y_t for all y_t included in Y^{DIS}. The full conditional distribution of y_t corresponds to the smoothed distribution arising from forward (predicting) and backward (smoothing) recursion of the Kalman filter. The corresponding sample provides the basis for posterior predictive model assessment. Note that the involved predictive distribution (PR) is directly provided by the transition equation, whereas the full conditional distribution (SM) implied

via the backward smoothing recursion. Given the model setup, we have

$$f_{\text{VECM}}(y_t|y_1,\ldots,y_{t-1},y_{t+1},y_T,Z_1,\ldots,Z_T,\Theta) \propto \prod_{k=0}^{K} f(y_{t+k}|y_{t+k-1},\ldots,y_1,Z_1,\ldots,Z_T,\Theta),$$

where $f(y_{t+k}|y_{t+k-1},\ldots,y_1,Z_1,\ldots,Z_T,\Theta)$ for $k=0,\ldots,K$ corresponds to the predictive distribution as implied by the transition equation corresponding to a normal distribution with expected value and covariance matrix given as

$$\mu_{y_t}^{\text{PR}} = (I_P + \alpha\beta' + \Phi_1)y_{t-1} + \sum_{k=1}^{K}(\Phi_{k+1} - \Phi_k)y_{t-k} \quad \text{and} \quad \Omega_{y_t}^{\text{PR}} = \Sigma.$$

Hence, the full conditional distribution corresponds to a normal distribution as implied by

$$f_{\text{VECM}}(y_t|y_1,\ldots,y_{t-1},y_{t+1},y_T,Z_1,\ldots,Z_T,\Theta) \propto \prod_{k=0}^{K} \exp\left\{-\frac{1}{2}(y_{t+k} - \mu_{y_{t+k}}^{\text{P}})'\Sigma^{-1}(y_{t+k} - \mu_{y_{t+k}}^{\text{P}})\right\}.$$

The corresponding full conditional expectation and covariance are given as

$$\Omega_{y_t}^{\text{SM}} = [\Sigma^{-1} + (I_P + \alpha\beta' + \Phi_1)'\Sigma^{-1}(I_P + \alpha\beta' + \Phi_1) + \sum_{k=2}^{K}(\Phi_k - \Phi_{k-1})'\Sigma^{-1}(\Phi_k - \Phi_{k-1})]^{-1}$$

and $\mu_{y_t}^{\text{SM}} = \Omega_{y_t}^{\text{SM}}\kappa_{y_t}^{\text{SM}}$, where

$$\begin{aligned}
\kappa_{y_t}^{\text{SM}} \quad = \quad & \Sigma^{-1}\left((I_P + \alpha\beta' + \Phi_1 + \sum_{k=1}^{K}(\Phi_{k+1} - \Phi_k)y_{t-k-1})\right) + \\
& \left((I_p + \alpha\beta' + \Phi_1)'\Sigma^{-1}(y_{t+1} - \sum_{k=1}^{K}(\Phi_{k+1} - \Phi_k)y_{t-k})\right) + \\
& \sum_{k=1}^{K}\left((\Phi_{k+1} - \Phi_k)'\Sigma^{-1}\left(y_{t+k+1} - (I_P + \alpha\beta' + \Phi_1)y_{t+k} - \sum_{\substack{k'=1 \\ k'\neq k}}^{K}(\Phi_{k'+1} - \Phi_{k'})y_{t-k'+1}\right)\right).
\end{aligned}$$

Using the observed values as initializations of the discarded values, sampling of the set of discarded values $\{y_t^{\text{DIS}}\}_{t=1}^{T}$ is then possible via iteratively sampling from

$$f_{\text{VECM}}(y_t^{\text{DIS}}|y_1^{\text{COM}},\ldots,y_{t-1}^{\text{COM}},y_t^{\text{REM}},y_{t+1}^{\text{COM}},\ldots,y_T^{\text{COM}},Z_1,\ldots,Z_T,\Theta),$$

as implied by multivariate normal theory and with y_t^{COM} denoting if applicable the completed vector y_t, where the discarded values are replaced by their sampled counterparts.

Given a sample of the discarded values drawn from the posterior predictive distributions, model fit is measured as $SSE = \sum_{s=1}^{S}\text{vec}(Y - Y_{\text{COM}}^{(s)})'\text{vec}(Y - Y_{\text{COM}}^{(s)})$, where $Y_{\text{COM}}^{(s)}$ denotes the matrix of

completed observations with discarded values replaced by the draws from the posterior predictive distribution at each iteration $s = 1, \ldots, S$.[9] Model selection using posterior predictive assessment does not require a postprocessed sample, as the quantities involved in the posterior predictive distribution, i.e. the full conditional distribution of the discarded values, are all invariant quantities.

To highlight the precision of the posterior predictive assessment approach, the number of cross sections P, the number of observations in time T, the signal to noise ratio, and the way the information from the incomplete data sets is used are varied within a simulation study. The fraction of data missing is set to equal 1%, but the correspondingly implied partition is different for each incomplete data set.[10] The SSE is then calculated for all of these data sets, conditional on the same specific choice of the number of factors R. The choices for the parameters are $P = \{10, 20, 40, 80\}$, $T = \{100, 200\}$, and $R = \{2, 3\}$, and the signal-to-noise ratio is varied between 10 and 1. The simulation study hence covers the arising 32 scenarios. For each scenario, $G = 50$ data sets are simulated. From each data set, $J = 100$ incomplete versions are generated, removing 1% of the data at random. For each incomplete data set per scenario, the model is estimated for a set of candidate values given as $R^C = \{1, 2, 3, 4, 5\}$, thus providing five Gibbs sequences of length $S = 5,000$ after discarding burn-in sequences of length $2,000$. Now for each simulated data set, there are $J = 100$ vectors of size 5×1 containing the SSE values for the set of candidate values R^C. Hence $SSE_{g,j}(\tilde{R})$ with $g = 1, \ldots, G$ and $j = 1, \ldots, J$ denotes the sum of squared errors for the j^{th} incomplete version of the g^{th} simulated data set when the number of factors in the estimation is set to $\tilde{R}$. Next, in a bootstrap step, $L = 10,000$ bootstrap samples of size $Q = \{25, 100\}$ each are drawn with replacement along the j dimension for each of the $G = 50$ data sets per scenario. The index set $\mathcal{B}_{l,g}$ contains the indices of the bootstrapped elements, which range from 1 to J, including possible duplicates. We then calculate $C_{l,g}(\tilde{R}) = \frac{1}{Q} \sum_{q \in \mathcal{B}_{l,g}} SSE_{g,q}(\tilde{R})$ for all $\tilde{R} \in R^C$, and estimate the number of factors as $\hat{R}_{l,g} = \arg\min_{\tilde{R} \in R^C} C_{l,g}(\tilde{R})$. With $l \in \{1, \ldots, L\}$ and $g \in \{1, \ldots, G\}$, this gives us $500,000$ estimates for the number of factors per scenario.

Table (4) reports the corresponding shares for $\hat{R}$ from R^C for each scenario. Overall, the obtained results indicate that the chance to underestimate R is virtually zero for all scenarios, except those with $P = 10$ and $R = 3$, where a signal-to-noise ratio of 1 results in frequent underestimations. The underestimation is thereby more pronounced for the scenarios with $T = 100$. In the following, the scenarios with with $P = \{20, 40, 80\}$ are summarized. In these 30 scenarios, the number of factors is sometimes overestimated, but it must be noted that in all of these scenarios, the correct model is identified in more than 90% of the cases. On average, models are correctly identified in about 97% of all cases for the signal-to-noise ratio of 1 and in about 96% of all cases for the signal-to-noise ratio of

[9] In order to reduce the computational burden in terms of the required memory capacity, calculation of the SSE is based on the set of discarded values only, as not discarded values do not contribute to SSE.

[10] We have also inspected higher rates of discarded values, i.e., 2%, 5%, and 10%. While such higher rates of discarded rates generally only slightly increase the accuracy of model detection, the involved computational burden arising from required storage capacity becomes even larger. The reported combination of 1% discarded values and 100 incomplete versions of a data set can hence be recommended to handle the implicit trade-off between model detection accuracy and computational burden.

10. If Q is reduced to 25, the correct model is identified in more than 88% of the cases. On average, models are correctly identified in about 94% of all cases for the signal-to-noise ratio of 1 and in about 92% of all cases for the signal-to-noise ratio of 10.[11]

6 Empirical illustration

In this section, we illustrate the suggested ex-post approach using a data set from financial economics. This empirical illustration closely follows Frühwirth-Schnatter and Lopes (2018). The data set consists of monthly log returns of 22 exchange rates against the Euro from February 1999 to September 2018, see Figure 3.[12] The data are demeaned and standardized. In the first step, the posterior predictive assessment as described above is used to determine the appropriate number of factors. Of course, there is only one available data set here, however, the method is applied in the same fashion via generating 100 incomplete data sets from the available one, and then using the bootstrap procedure to produce 100 samples to determine $\hat{R}$. Figure 4 shows that the SSE is minimized for $\hat{R} = 2$, so that the model is estimated with two factors.[13]

After estimation, both the factors and factor loadings are orthogonally transformed. The orthogonal transformation performed here turns the first factor into a *U.S. dollar factor*, maximizing the loading on the first factor for the exchange rate between U.S. dollar and the Euro. The estimates of rotated factor loadings and corresponding 84% HPDIs are shown in Table 5, whereas the estimated factors and corresponding 84% HPDIs are displayed in Figure 5. Indeed, the rotated first factor is virtually identical to the exchange rate between U.S. dollar and the Euro, with a factor loading of 0.9999, or a correlation coefficient of 0.9998. The *U.S. dollar factor* also clearly shows the (flexible) peg between the U.S. dollar and the Hongkong dollar, which has a factor loading of 0.9997, and strong loadings with a number of south east Asian currencies, such as the Indonesian rupee, the Malaysian ringgit, the Philippine peso, the Singapore dollar, and the Thai baht. Less pronounced loadings are found for the Japanese yen, the Canadian dollar and the Korean won. The factor is virtually orthogonal to the Czech koruna, the Mexican peso, the Norwegian krone, the Swedish krona and the Romanian leu and affects the Polish zloty slightly negatively. The second factor cannot be linked to any particular exchange rate, but shows largest loadings for the Australian dollar and the

[11] We have also tried bootstrap samples of size $Q = 10$ and $Q = 1$, respectively, the latter corresponding to model choice based on a single incomplete data set. While for $Q = 10$, models are still correctly identified in about 85% of all cases on average, for $Q = 1$, this share drops to less than 60%.

[12] Data has been extracted from the European Central Banks Statistical Data Warehouse on September 20, 2018. The considered currency are Australian dollar (AUS), Canadian dollar (CAD), Swiss franc (CHF), Czech koruna (CZK), Danish krone (DKK), UK pound sterling (GBP), Hong Kong dollar (HKD), Indonesian rupiah (IDR), Japanese yen (JPY), South Korean won (KRW), Mexican peso (MXN), Malaysian ringgit (MYR), Norwegian krone (NOK), New Zealand dollar (NZD), Philippine peso (PHP), Polish zloty (PLN), Romanian leu (RON), Russian rouble (RUB), Swedish krona (SEK), Singapore dollar (SGD), Thai baht (THB), US dollar (USD).

[13] Figure 4 also shows the quantiles of the distribution for the average SSE, indicating that the distributions display quite strong separation for the different choices of R. In fact, the 88.35% quantile of the distribution for $R = 1$ equals the 11.65% quantile of the distribution for $R = 2$, and the 94.33% quantile of the distribution for $R = 2$ equals the 5.67% quantile of the distribution for $R = 3$.

Korean won. From the perspective of investors from the euro area this gives rise for the opportunity to diversify exchange rate risks. Overall, the estimation uncertainty for the *U.S. dollar factor* is substantially lower than for the second factor.

7 Conclusion

This paper discusses the handling of identifying restrictions on the Stiefel manifold in the context of reduced rank regressions. We illustrate the rotational invariance of the likelihood as one part of the identification problem. To provide Bayesian estimation, the sampler of Koop et al. (2010) is used to provide a sample from the posterior distribution. This paper proposes a post-processing algorithm for the posterior sample that allows for identification and directed inference, and thereby extends the possibilities to conduct valid inference, when the cointegration vectors or factors are restricted on the Stiefel manifold. The post-processing algorithm is an extended version of the ex-post algorithm proposed by Aßmann et al. (2016) for static and dynamic factor models. We illustrate how the post-processing works for vector error correction models via a simulation study and show an application of the sampling procedure suggested by Koop et al. (2010) for factor models. Further, we propose to use posterior predictive assessment to obtain model evidence and to compare models. We do so, because obtaining the marginal likelihood is computational extremely demanding when the Stiefel manifold is involved. Finally, our approach to the analysis of reduced rank models is illustrated in an empirical example. Future research may focus on alternative possibilities to provide model comparison and assessment in a Bayesian framework.

References

Abadir, K. and Magnus, J. (2005). *Matrix Algebra*. Econometric Exercises. Cambridge University Press.

Aguilar, O. and West, M. (2000). Bayesian Dynamic Factor Models and Portfolio Allocation. *Journal of Business & Economic Statistics*, 18(3):338–357.

Aßmann, C., Boysen-Hogrefe, J., and Pape, M. (2016). Bayesian analysis of static and dynamic factor models: An ex-post approach towards the rotation problem. *Journal of Econometrics*, 192(1):190–206.

Baştürk, N., Hoogerheide, L., and van Dijk, Herman K. (2017). Bayesian analysis of boundary and near-boundary evidence in econometric models with reduced rank. *Bayesian Analysis*, 12(3):879–917.

Butler, R. W. and Wood, A. T. (2003). Laplace approximation for bessel functions of matrix argument. *Journal of Computational and Applied Mathematics*, 155(2):359–382.

Carvalho, C. M., Chang, J., Lucas, J. E., Nevins, J. R., Wang, Q., and West, M. (2008). High-dimensional sparse factor modeling: Applications in gene expression genomics. *Journal of the American Statistical Association*, 103(484):1438–1456.

Celeux, G. (1998). Bayesian inference for mixtures: The label-switching problem. In Payne, R. and Green, P. J., editors, *COMPSTAT 98—Proc. in Computational Statistics*, pages 227–233.

Celeux, G., Hurn, M., and Robert, C. (2000). Computational and inferential difficulties with mixture posterior distributions. *Journal of the American Statistical Association*, 95(451):957–970.

Chan, J., Leon-Gonzalez, R., and Strachan, R. W. (2018). Invariant inference and efficient computation in the static factor model. *Journal of the American Statistical Association*, 113(522):819–828.

Cheng, M.-Y., Hall, P., and Turlach, B. A. (1999). High-derivative parametric enhancements of nonparametric curve estimators. *Biometrika*, 86(2):417–428.

Chib, S. (1995). Marginal likelihood from the gibbs output. *Journal of the American Statistical Association*, 90(432):1313–1321.

Chib, S. and Jeliazkov, I. (2001). Marginal likelihood from the metropolis–hastings output. *Journal of the American Statistical Association*, 96(453):270–281.

Chib, S., Nardari, F., and Shephard, N. (2006). Analysis of High Dimensional Multivariate Stochastic Volatility Models. *Journal of Econometrics*, 134(2):341–371.

Chikuse, Y. (1990). The matrix angular central gaussian distribution. *Journal of Multivariate Analysis*, 33(2):265–274.

Chikuse, Y. (2003). *Statistics on special manifolds*, volume 174 of *Lecture notes in statistics*. Springer, New York.

Clarke, B., Andrew, and Barron, R. (1990). Information-theoretic Asymptotics of Bayes Methods. *IEEE Transactions on Information Theory*, 36:453–471.

Crowther, N. (1975). The exact non-central distribution of a quadratic form in normal vectors. *South African Statistical Journal*, 9(1):27–36.

Edwards, M. C. (2010). A Markov Chain Monte Carlo Approach to Confirmatory Item Factor Analysis. *Psychometrika*, 75(3):474–497.

Erosheva, E. A. and Curtis, S. M. (2017). Dealing with Reflection Invariance in Bayesian Factor Analysis. *Psychometrika*, 82(2):295–307.

Friel, N. and Pettitt, A. N. (2008). Marginal likelihood estimation via power posteriors. *Journal of the Royal Statistical Society: Series B*, 70(3):589–607.

Frühwirth-Schnatter, S. and Lopes, H. (2018). Sparse Bayesian Factor Analysis when the Number of Factors is Unknown. *ArXiv e-prints*.

Gelman, A. and Meng, X.-L. (1998). Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical Science*, 13(2):163–185.

Gelman, A., Meng, X.-L., and Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, 6(4):733–760.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472.

Geweke, J. (1996). Bayesian reduced rank regression in econometrics. *Journal of Econometrics*, 75(1):121–146.

Geweke, J. and Zhou, G. (1996). Measuring the Pricing Error of the Arbitrage Pricing Theory. *Review of Financial Studies*, 9(2):557–587.

Golub, G. H. and van Loan, C. F. (2013). *Matrix Computations*. The Johns Hopkins University Press, 4th edition.

Gupta, A. K. and Nagar, D. K. (2000). *Matrix variate distributions*, volume 104 of *Chapman & Hall/CRC monographs and surveys in pure and applied mathematics*. Chapman & Hall, Boca Raton, FL.

Herz, C. S. (1955). Bessel functions of matrix argument. *Annals of Mathematics*, 61(3):474–523.

Hoff, P. D. (2009). Simulation of the matrix bingham–von mises–fisher distribution, with applications to multivariate and relational data. *Journal of Computational and Graphical Statistics*, 18(2):438–456.

Jasra, A., Holmes, C., and Stephens, D. (2005). Markov chain monte carlo methods and the label switching problem in bayesian mixture modeling. *Statistical Science*, 20(1):50–67.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12(2-3):231–254.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica*, 59(6):1551–1580.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795.

Kleibergen, F. and Paap, R. (2002). Priors, posteriors and bayes factors for a bayesian analysis of cointegration. *Journal of Econometrics*, 111(2):223–249.

Kleibergen, F. and van Dijk, H. (1994). On the shape of the likelihood/posterior in cointegration models. *Econometric Theory*, 10(3-4):514–551.

Koev, P. and Edelman, A. (2006). The efficient evaluation of the hypergeometric function of a matrix argument. *Mathematics of Computation*, 75(254):833–847.

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

Kristof, W. (1964). Die beste orthogonale Transformation zur gegenseitigen Überführung zweier Factorenmatrizen. *Diagnostica*, 10:87–90.

Kume, A. (2005). Saddlepoint approximations for the bingham and fisher-bingham normalising constants. *Biometrika*, 92(2):465–476.

Kume, A., Preston, S. P., and Wood, A. T. A. (2013). Saddlepoint approximations for the normalizing constant of fisher-bingham distributions on products of spheres and stiefel manifolds. *Biometrika*, 100(4):971–984.

Larsson, R. and Villani, M. (2001). A distance measure between cointegration spaces. *Economics Letters*, 70(1):21–27.

Lopes, H. F. and West, M. (2004). Bayesian Model Assessment in Factor Analysis. *Statistica Sinica*, 14:41–67.

Lütkepohl, H. (1996). *Handbook of matrices*. Wiley, Chichester and New York.

Man, A. X. and Culpepper, S. A. (2020). A mode-jumping algorithm for bayesian factor analysis. *Journal of the American Statistical Association*, 14:1–14.

Mathai, A. M., Provost, S. B., and Hayakawa, T. (1995). *Bilinear forms and zonal polynomials*, volume 102 of *Lecture notes in statistics*. Springer-Verlag, New York.

Meng, X.-L. and Wong, W. H. (1996). Simulating ratios of normalizing constants via a simple identity: A theoretical exploration. *Statistica Sinica*, 6:831–860.

Neal, R. M. (2001). Annealed importance sampling. *Statistics and Computing*, 11(2):125–139.

Ročková, V. and George, E. I. (2016). Fast bayesian factor analysis via automatic rotations to sparsity. *Journal of the American Statistical Association*, 111(516):1608–1622.

Sadtler, P. T., Quick, K. M., Golub, M. D., Chase, S. M., Ryu, S. I., Tyler-Kabara, E. C., Yu, B. M., and Batista, A. P. (2014). Neural constraints on learning. *Nature*, 512(7515):423–426.

Schönemann, P. H. (1966). A generalized solution to the orthogonal Procrustes Problem. *Psychometrika*, 31(1):1–10.

Stephens, M. (2000). Dealing with Label Switching in Mixture Models. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 62(4):795–809.

Strachan, R. W. (2003). Valid bayesian estimation of the cointegrating error correction model. *Journal of Business & Economic Statistics*, 21(1):185–195.

Strachan, R. W. and Inder, B. (2004). Bayesian analysis of the error correction model. *Journal of Econometrics*, 123(2):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.

Villani, M. (2005). Bayesian reference analysis of cointegration. *Econometric Theory*, 21(02).

Villani, M. (2006). Bayesian point estimation of the cointegration space. *Journal of Econometrics*, 134(2):645 – 664.

Vito, R. D., Bellio, R., Trippa, L., and Parmigiani, G. (2018). Bayesian multi-study factor analysis for high-throughput biological data.

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., and Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science (New York, N.Y.)*, 330(6004):686–688.

Zellner, A., Ando, T., Baştürk, N., Hoogerheide, L., and van Dijk, H. K. (2014). Bayesian analysis of instrumental variable models: Acceptance-rejection within direct monte carlo. *Econometric Reviews*, 33(1-4):3–35.

Tables

Table 1: Prior hyperparameter setting

parameter	distribution	hyperparameter
Σ	inverse Wishart	$\nu_\Sigma = 3,\ \Omega_\Sigma = 1/1000 I_P$
Ξ	multivariate normal	$\mu_\Xi = 0,\ \Omega_\Xi = 100 I_{PQ}$
β	matrix angular central Gaussian	$C_\tau = CC' + \tau\tilde{C}\tilde{C}'$, where $\tau = 1$ and $\tilde{C}$ denotes the null space matrix of C, and $C = \hat{C}(\hat{C}'\hat{C})^{-.5}$ with elements of $\hat{C}$, say e.g. c, are each independently drawn from uniform density $\frac{1}{2}\mathcal{I}(-1 < c < 1)$
α	multivariate normal	$\mu_\alpha = 0,\ \Omega_\alpha = \tau(\beta' C_\tau^{-1}\beta)^{-1} \otimes \Sigma$
A	matrix angular central Gaussian	$C_A = \tau(B' C_\tau^{-1} B)^{-1} \otimes \Sigma$
B	multivariate normal	$\mu_B = 0,\ \Omega_B = \tau(A'\Sigma^{-1}A)^{-1} \otimes C_\tau$

Table 2: Parameter values and estimates for $\alpha\beta'$

DGP	$\beta\beta' = I_R$		post-processed $\beta\beta' = I_R$	
$\text{vec}(\alpha\beta')$	$\text{vec}(\widehat{\alpha\beta'})$	$\text{vec}(\hat{\alpha}\hat{\beta}')$	$\text{vec}(\widehat{\alpha\beta'})$	$\text{vec}(\hat{\alpha}\hat{\beta}')$
-0.0177	-0.0155	-0.0097	-0.0155	-0.0155
0.0826	0.0795	0.0537	0.0795	0.0796
0.0559	0.0477	0.0327	0.0477	0.0478
-0.0375	-0.0409	-0.0294	-0.0409	-0.0411
-0.3033	-0.3302	-0.2251	-0.3302	-0.3302
0.2151	0.2354	0.1597	0.2354	0.2354
0.0249	0.0170	0.0114	0.0170	0.0169
0.0800	0.0688	0.0460	0.0688	0.0689
-0.4022	-0.4364	-0.2985	-0.4364	-0.4364
0.1913	0.2248	0.1530	0.2248	0.2249
-0.0399	-0.0367	-0.0253	-0.0367	-0.0367
0.1626	0.1467	0.1005	0.1467	0.1467
-0.0992	-0.1044	-0.0712	-0.1044	-0.1044
0.0710	0.0773	0.0525	0.0773	0.0773
0.0086	0.0072	0.0050	0.0072	0.0074
0.0258	0.0200	0.0132	0.0200	0.0199

Notes: DGP refers to parameter values used in the data generating process.

Table 3: Parameter values and estimates for α and β.

mapped onto α and β from DGP					mapped onto PMCS estimate for β				
DGP	PMCS	post-processed $\beta\beta' = I_R$			DGP	PMCS	post-processed $\beta\beta' = I_R$		
vec(α)		vec($\hat{\alpha}$)	84% HPDI		vec(α)		vec($\hat{\alpha}$)	84% HPDI	
-0.1981		-0.2140	-0.2387	-0.1889	-0.0230		-0.0279	-0.0408	-0.0146
0.1991		0.2097	0.1835	0.2369	-0.0800		-0.0723	-0.0866	-0.0588
0.0618		0.0502	0.0241	0.0752	-0.0729		-0.0617	-0.0747	-0.0472
0.0170		0.0077	-0.0196	0.0345	0.0640		0.0653	0.0509	0.0795
0.4740		0.5147	0.4658	0.5641	0.5132		0.5567	0.5052	0.6127
-0.2347		-0.2726	-0.3240	-0.2207	-0.2972		-0.3363	-0.3949	-0.2817
0.0399		0.0382	-0.0123	0.0909	0.0097		0.0129	-0.0439	0.0677
-0.1861		-0.1682	-0.2225	-0.1161	-0.1756		-0.1552	-0.2122	-0.0964
vec(β)	vec($\hat{\beta}$)	vec($\hat{\beta}$)	84% HPDI		vec(β)	vec($\hat{\beta}$)	vec($\hat{\beta}$)	84% HPDI	
0.7308	0.7422	0.7413	0.7183	0.7632	-0.7753	-0.7889	-0.7889	-0.8120	-0.7624
0.6427	0.6263	0.6276	0.5978	0.6606	-0.4229	-0.4032	-0.4032	-0.4398	-0.3701
-0.0776	-0.0685	-0.0656	-0.0856	-0.0473	0.4460	0.4352	0.4352	0.4163	0.4574
0.2166	0.2283	0.2286	0.2054	0.2502	-0.1453	-0.1602	-0.1602	-0.1832	-0.1335
0.2682	0.2756	0.2781	0.2662	0.2887	-0.0692	-0.0673	-0.0673	-0.0714	-0.0633
-0.3712	-0.3828	-0.3807	-0.3967	-0.3630	-0.6099	-0.6134	-0.6134	-0.6210	-0.6060
-0.8810	-0.8750	-0.8753	-0.8853	-0.8666	-0.7637	-0.7622	-0.7622	-0.7675	-0.7571
-0.1188	-0.1085	-0.1078	-0.1194	-0.0969	-0.1999	-0.1955	-0.1955	-0.1995	-0.1914

Notes: DGP denotes data generating process and PMCS denotes the posterior mean cointegration space estimators as suggested by Villani (2006).

Table 4: Results of the posterior predictive simulation study.

	$\hat{R}$ proportion for signal-to-noise ratio 1					$\hat{R}$ proportion for signal-to-noise ratio 10				
{Q,T,R,P}	1	2	3	4	5	1	2	3	4	5
$\{100, 100, 2, 10\}$	0.0000	0.9993	0.0007	0.0000	0.0000	0.0000	0.9953	0.0047	0.0000	0.0000
$\{100, 100, 3, 20\}$	0.0000	0.9862	0.0137	0.0000	0.0000	0.0000	0.9247	0.0737	0.0012	0.0004
$\{100, 100, 2, 40\}$	0.0000	0.9238	0.0675	0.0087	0.0000	0.0000	0.9463	0.0524	0.0013	0.0000
$\{100, 100, 3, 80\}$	0.0000	0.9949	0.0051	0.0000	0.0000	0.0000	0.9435	0.0562	0.0002	0.0001
$\{100, 100, 2, 10\}$	0.1540	0.2774	0.5656	0.0030	0.0000	0.0000	0.0000	0.9773	0.0227	0.0000
$\{100, 100, 3, 20\}$	0.0000	0.0000	0.9949	0.0048	0.0003	0.0000	0.0000	0.9195	0.0803	0.0002
$\{100, 100, 2, 40\}$	0.0000	0.0000	0.9437	0.0554	0.0009	0.0000	0.0000	0.9330	0.0424	0.0245
$\{100, 100, 3, 80\}$	0.0000	0.0000	0.9944	0.0056	0.0000	0.0000	0.0000	0.9125	0.0852	0.0023
$\{100, 200, 2, 10\}$	0.0000	0.9994	0.0006	0.0000	0.0000	0.0000	0.9999	0.0001	0.0000	0.0000
$\{100, 200, 3, 20\}$	0.0000	0.9751	0.0249	0.0000	0.0000	0.0000	0.9818	0.0180	0.0003	0.0000
$\{100, 200, 2, 40\}$	0.0000	0.9464	0.0492	0.0043	0.0000	0.0000	0.9582	0.0306	0.0111	0.0001
$\{100, 200, 3, 80\}$	0.0000	0.9561	0.0439	0.0000	0.0000	0.0000	0.9776	0.0224	0.0001	0.0000
$\{100, 200, 2, 10\}$	0.0000	0.5488	0.4512	0.0000	0.0000	0.0000	0.0000	0.9966	0.0034	0.0000
$\{100, 200, 3, 20\}$	0.0000	0.0000	0.9569	0.0431	0.0000	0.0000	0.0000	0.9691	0.0309	0.0000
$\{100, 200, 2, 40\}$	0.0000	0.0000	0.9587	0.0378	0.0035	0.0000	0.0000	0.9453	0.0537	0.0010
$\{100, 200, 3, 80\}$	0.0000	0.0000	0.9821	0.0179	0.0000	0.0000	0.0000	0.9402	0.0596	0.0002
$\{25, 100, 2, 10\}$	0.0000	0.9876	0.0124	0.0000	0.0000	0.0000	0.9717	0.0280	0.0004	0.0000
$\{25, 100, 3, 20\}$	0.0000	0.9465	0.0495	0.0028	0.0012	0.0000	0.8688	0.1212	0.0067	0.0033
$\{25, 100, 2, 40\}$	0.0000	0.8804	0.0918	0.0265	0.0013	0.0000	0.8921	0.0906	0.0159	0.0014
$\{25, 100, 3, 80\}$	0.0000	0.9753	0.0244	0.0003	0.0000	0.0000	0.9101	0.0841	0.0040	0.0019
$\{25, 100, 2, 10\}$	0.1689	0.2769	0.5397	0.0146	0.0000	0.0000	0.0000	0.9520	0.0472	0.0008
$\{25, 100, 3, 20\}$	0.0000	0.0000	0.9631	0.0289	0.0080	0.0000	0.0000	0.8731	0.1178	0.0091
$\{25, 100, 2, 40\}$	0.0000	0.0000	0.9043	0.0838	0.0120	0.0000	0.0000	0.8673	0.0981	0.0346
$\{25, 100, 3, 80\}$	0.0000	0.0000	0.9802	0.0192	0.0006	0.0000	0.0000	0.8802	0.1109	0.0089
$\{25, 200, 2, 10\}$	0.0000	0.9887	0.0112	0.0000	0.0000	0.0000	0.9933	0.0066	0.0000	0.0000
$\{25, 200, 3, 20\}$	0.0000	0.9329	0.0649	0.0013	0.0009	0.0000	0.9292	0.0643	0.0062	0.0002
$\{25, 200, 2, 40\}$	0.0000	0.9038	0.0791	0.0154	0.0018	0.0000	0.9085	0.0719	0.0168	0.0028
$\{25, 200, 3, 80\}$	0.0000	0.9427	0.0569	0.0004	0.0000	0.0000	0.9413	0.0547	0.0035	0.0005
$\{25, 200, 2, 10\}$	0.0000	0.5508	0.4492	0.0000	0.0000	0.0000	0.0000	0.9856	0.0144	0.0000
$\{25, 200, 3, 20\}$	0.0000	0.0000	0.9083	0.0875	0.0042	0.0000	0.0000	0.9397	0.0594	0.0009
$\{25, 200, 2, 40\}$	0.0000	0.0000	0.9194	0.0641	0.0165	0.0000	0.0000	0.8915	0.0949	0.0136
$\{25, 200, 3, 80\}$	0.0000	0.0000	0.9695	0.0287	0.0018	0.0000	0.0000	0.8985	0.0917	0.0098

Notes: R denotes the number of factors used for data generation, $\hat{R}$ denotes the estimated number of factors. For each cell in the table, a distinct parameter set was simulated, for each of which $G = 50$ distinct data sets were simulated. Out of each of these data sets, $J = 100$ different incomplete data sets were created, with 1% of the data missing from each incomplete data set, Q denotes the number of bootstrap samples, T denotes the number of observations for each of the P variables.

Table 5: Estimated factor loadings for the exchange rate data, after rotation (84% HPDIs in parentheses).

	first factor	second factor
AUD/EUR	0.3281 (0.2683, 0.3909)	0.7093 (0.6324, 0.7829)
CAD/EUR	0.5085 (0.4371, 0.5770)	0.4188 (0.3366, 0.5051)
CHF/EUR	0.4734 (0.3959, 0.5542)	0.0336 (-0.0635, 0.1266)
CZK/EUR	-0.1728 (-0.2553, -0.0855)	0.3305 (0.2315, 0.4326)
DKK/EUR	0.2086 (0.1217, 0.2995)	0.0392 (-0.0663, 0.1425)
GBP/EUR	0.3009 (0.2161, 0.3814)	0.2999 (0.1969, 0.3908)
HKD/EUR	0.9997 (0.9963, 1.0030)	0.0039 (-0.0020, 0.0100)
IDR/EUR	0.7922 (0.7468, 0.8353)	0.4019 (0.3507, 0.4545)
JPY/EUR	0.6690 (0.5977, 0.7330)	-0.0077 (-0.0881, 0.0692)
KRW/EUR	0.4931 (0.4385, 0.5544)	0.6536 (0.5847, 0.7234)
MXN/EUR	0.1344 (0.0560, 0.2110)	0.5165 (0.4272, 0.6162)
MYR/EUR	0.7908 (0.7375, 0.8382)	0.2727 (0.2126, 0.3295)
NOK/EUR	0.1088 (0.0301, 0.1834)	0.5500 (0.4582, 0.6462)
NZD/EUR	0.3605 (0.2899, 0.4340)	0.5254 (0.4425, 0.6115)
PHP/EUR	0.8227 (0.7713, 0.8717)	0.1545 (0.0947, 0.2114)
PLN/EUR	-0.3490 (-0.4231, -0.2783)	0.5294 (0.4429, 0.6185)
RON/EUR	-0.0603 (-0.1485, 0.0324)	-0.0819 (-0.1897, 0.0209)
RUB/EUR	0.2900 (0.2087, 0.3682)	0.4036 (0.3096, 0.4991)
SEK/EUR	-0.0173 (-0.0954, 0.0715)	0.4256 (0.3324, 0.5309)
SGD/EUR	0.8412 (0.8119, 0.8675)	0.4968 (0.4642, 0.5295)
THB/EUR	0.7985 (0.7567, 0.8412)	0.4194 (0.3719, 0.4719)
USD/EUR	0.9999 (0.9974, 1.0027)	0.0000 (-0.0056, 0.0055)

Figures

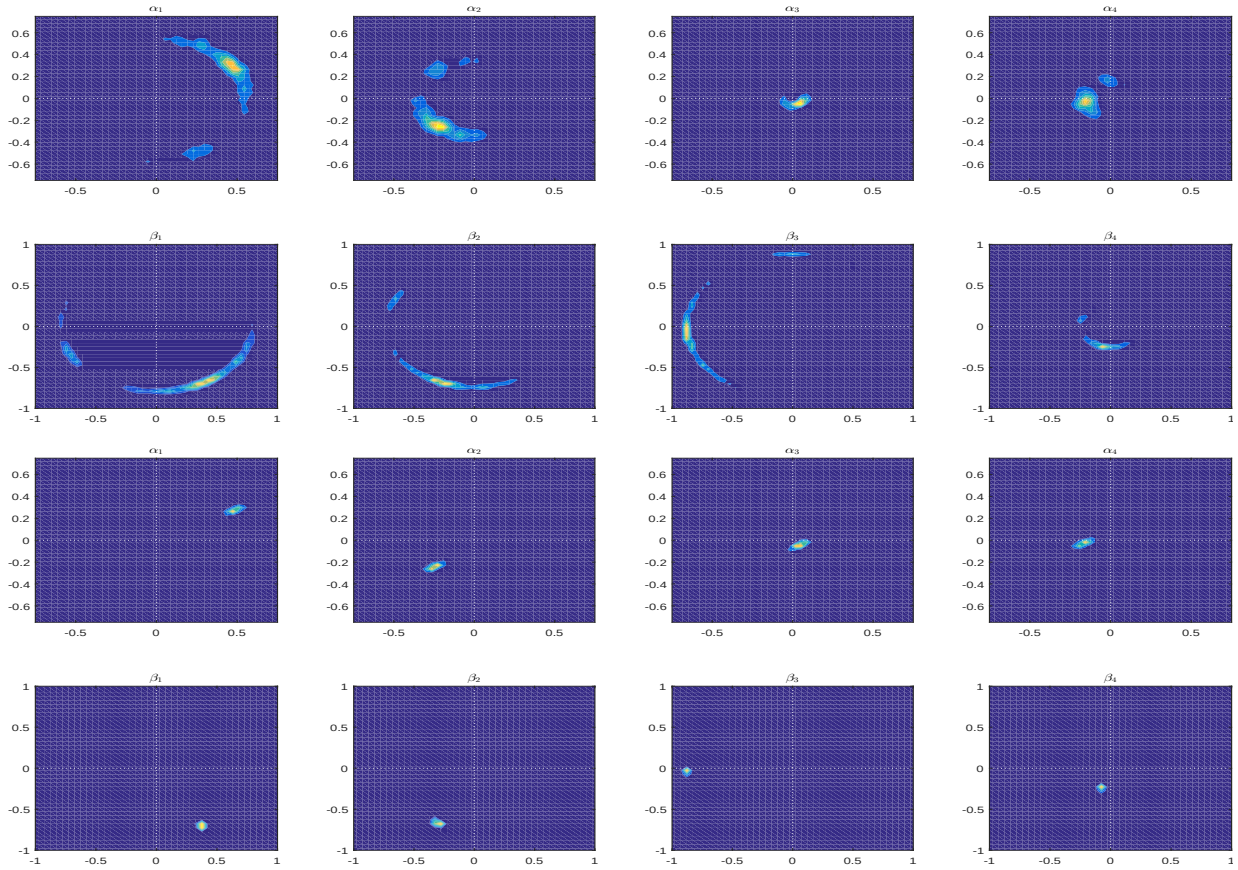


Figure 1: Distribution of row vectors of α and of β without (first and second row) and with post-processing (third and fourth row).

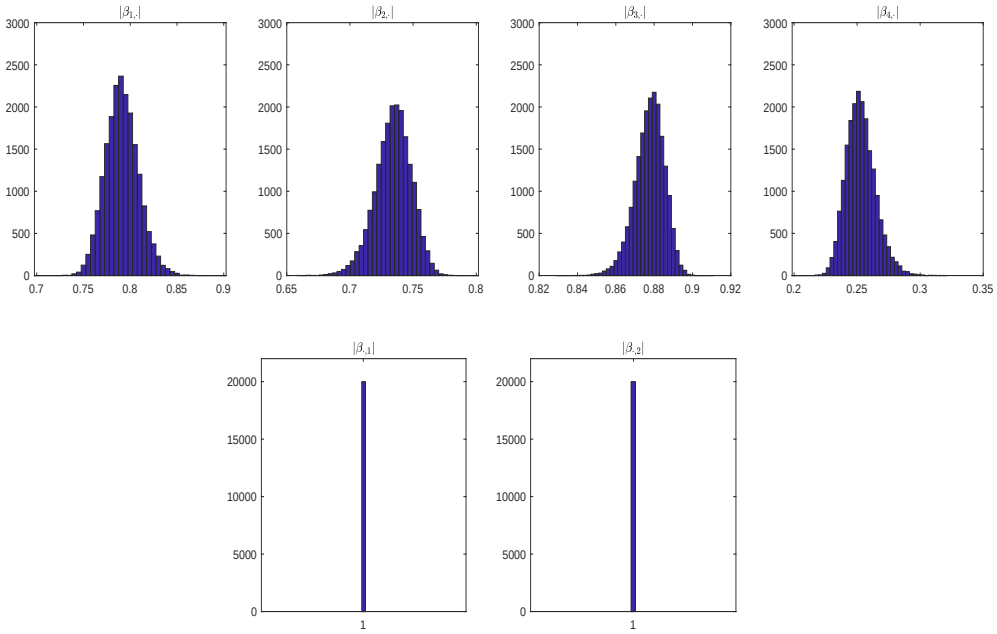


Figure 2: Distribution of row vector lengths (top) and column vector lengths (bottom) of β.

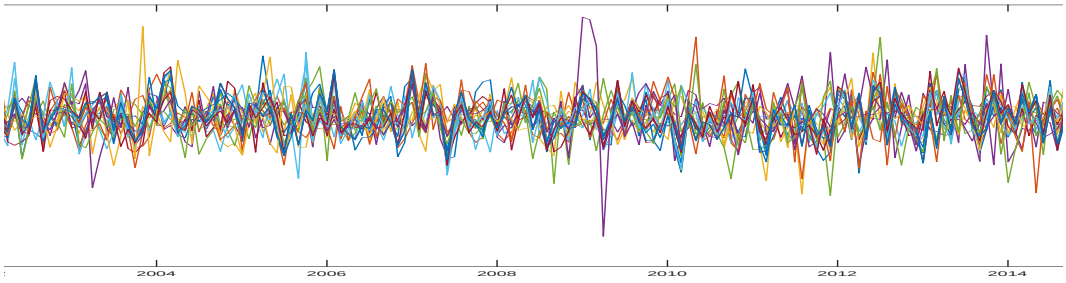


Figure 3: Demeaned and standardized monthly log returns based on the first trading day in a month for 22 currencies against the Euro from February 1999 until September 2018.

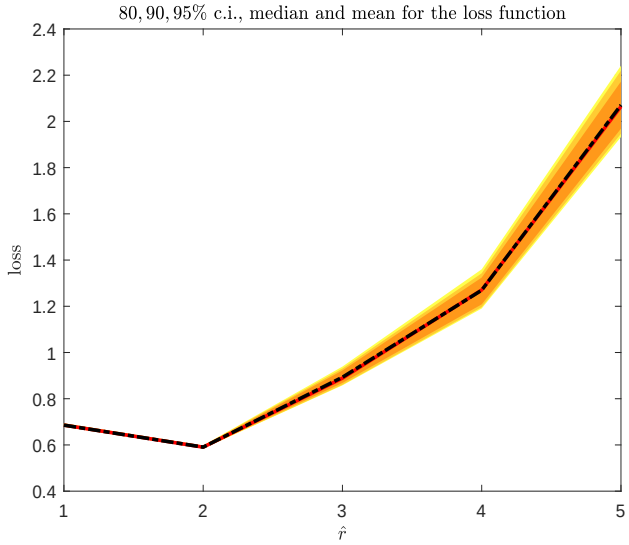


Figure 4: Loss for models with different numbers of factors applied to the exchange rate data using the posterior predictive assessment approach with choosing $Q = 100$.

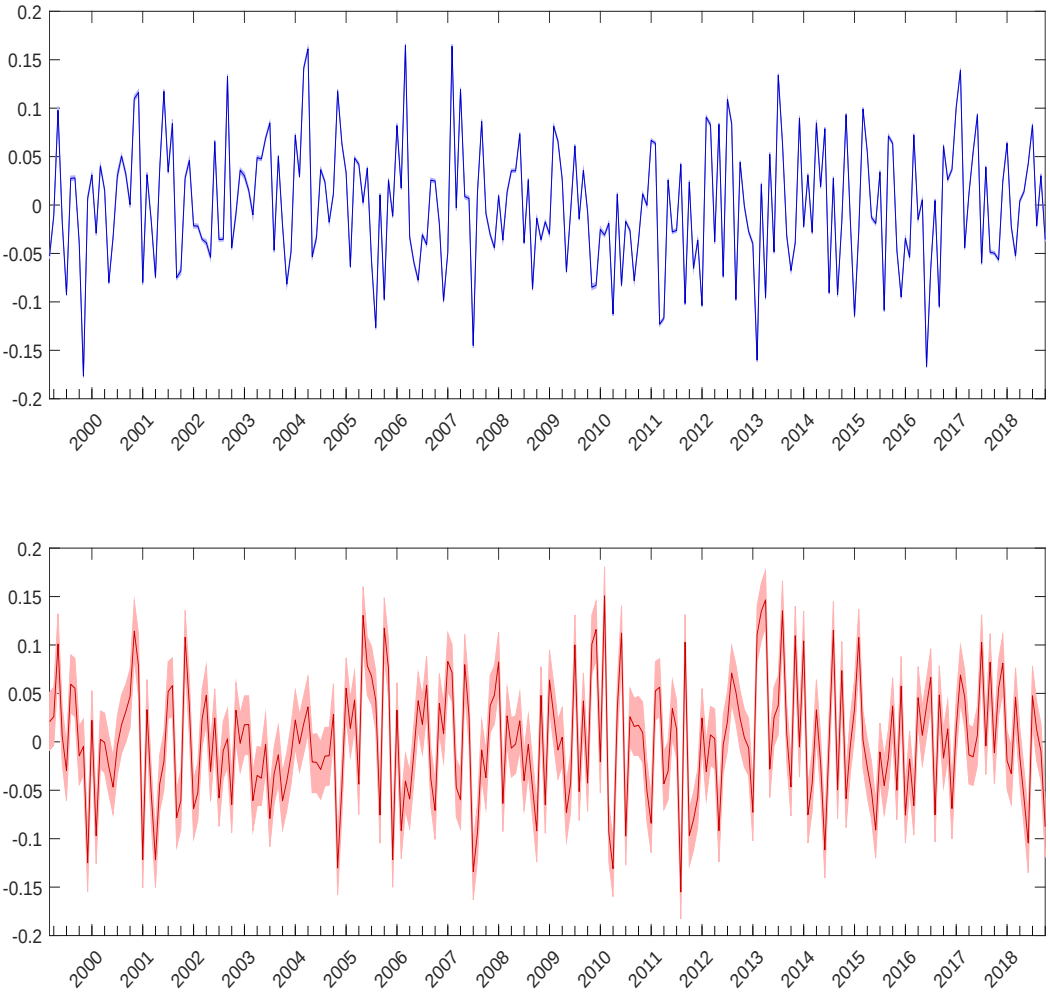


Figure 5: Estimated factors for the exchange rate data, after rotation. Blue denotes the *U.S. dollar factor*, and red denotes the second factor. Shaded areas denote 84% HPD intervals.

A Invariance of the posterior when considering the reparametrization in terms of A and B

The posterior distribution is also invariant under the reparametrization of α and β considered to facilitate efficient sampling, i.e.

$$A = \alpha(\alpha'\alpha)^{-\frac{1}{2}}, \quad B = \beta(\alpha'\alpha)^{\frac{1}{2}},$$

as vec(αD) and vec(βD) imply vec(AD) and vec(BD), i.e.,

$$(\alpha D)((\alpha D)'(\alpha D))^{-\frac{1}{2}} = \alpha(\alpha'\alpha)^{-\frac{1}{2}}D = AD, \tag{12}$$

and

$$(\beta D)((\alpha D)'(\alpha D))^{\frac{1}{2}} = \beta(\alpha'\alpha)^{\frac{1}{2}}D = BD. \tag{13}$$

This follows from the singular value decomposition of a real symmetric matrix as defined in Lütkepohl (1996), since

$$(\alpha'\alpha) \quad = \mathcal{S}_1\Lambda\mathcal{S}_1' \quad \text{and} \quad ((\alpha D)'(\alpha D)) \quad = \mathcal{S}_2\Lambda\mathcal{S}_2',$$

where $\mathcal{S}_1$ and $\mathcal{S}_2$ are orthogonal matrices and Λ is the diagonal matrix of eigenvalues the matrix $\alpha'\alpha$ and $D'\alpha'\alpha D$ as multiplying α with an orthogonal matrix D does not render the eigenvalues. Then

$$\mathcal{S}_2\Lambda\mathcal{S}_2' = D'\alpha'\alpha D = D'\mathcal{S}_1\Lambda\mathcal{S}_1'D,$$

implying $\mathcal{S}_2 = D'\mathcal{S}_1$. Then from the definition of the square root matrix given e.g. in Abadir and Magnus (2005), we have

$$(D'\alpha'\alpha D)^{-\frac{1}{2}} = \mathcal{S}_2\Lambda^{-\frac{1}{2}}\mathcal{S}_2' = D'\mathcal{S}_1\Lambda^{-\frac{1}{2}}\mathcal{S}_1'D = D'(\alpha'\alpha)^{-\frac{1}{2}}D.$$

Therefore, we obtain

$$\alpha D((\alpha D)'(\alpha D))^{-\frac{1}{2}} \quad = \quad \alpha DD'(\alpha'\alpha)^{-\frac{1}{2}}D = \alpha(\alpha'\alpha)^{-\frac{1}{2}}D = AD. \tag{14}$$

The result from Equation (3) can be derived analogously but using the square root of the elements in Λ in contrast to the inverse square root. Inserting AD and BD into the corresponding posterior distribution provided in Equation (7) reveals invariance under the orthogonal transformation described in Equation (4).