

## Secondary Publication



Aßmann, Christian; Boysen-Hogrefe, Jens; Pape, Markus

### Post-processing for Bayesian analysis of reduced rank regression models with orthonormality restrictions

Date of secondary publication: 19.12.2024

Version of Record (Published Version), Article

Persistent identifier: urn:nbn:de:bvb:473-irb-1055138

#### Primary publication

Aßmann, Christian; Boysen-Hogrefe, Jens; Pape, Markus (2024): Post-processing for Bayesian analysis of reduced rank regression models with orthonormality restrictions, in: Advances in statistical analysis : AStA, Berlin ; Heidelberg: Springer, Vol. 108, Nr. 3, pp. 577–609, doi: 10.1007/s10182-023-00489-5.

#### Legal Notice

This work is protected by copyright and/or the indication of a licence. You are free to use this work in any way permitted by the copyright and/or the licence that applies to your usage. For other uses, you must obtain permission from the rights-holders.

This document is made available under a Creative Commons license.



The license information is available online:

<https://creativecommons.org/licenses/by/4.0/legalcode>



# Post-processing for Bayesian analysis of reduced rank regression models with orthonormality restrictions

Christian Aßmann<sup>1,2</sup> · Jens Boysen-Hogrefe<sup>3</sup> · Markus Pape<sup>4</sup>

Received: 11 August 2022 / Accepted: 10 October 2023 / Published online: 20 December 2023  
© The Author(s) 2023

## Abstract

Orthonormality constraints are common in reduced rank models. They imply that matrix-variate parameters are given as orthonormal column vectors. However, these orthonormality restrictions do not provide identification for all parameters. For this setup, we show how the remaining identification issue can be handled in a Bayesian analysis via post-processing the sampling output according to an appropriately specified loss function. This extends the possibilities for Bayesian inference in reduced rank regression models with a part of the parameter space restricted to the Stiefel manifold. Besides inference, we also discuss model selection in terms of posterior predictive assessment. We illustrate the proposed approach with a simulation study and an empirical application.

**Keywords** Bayesian estimation · Post-processing · Reduced rank regression · Orthogonal transformation · Model selection · Stiefel manifold · Posterior predictive assessment

**JEL Classification** C11 · C31 · C51 · C52

## 1 Introduction

Bayesian analysis of reduced rank regression models is a well-established tool in economics, psychology, and neuroscience, see Aguilar and West (2000), Chib et al. (2006), Woolley et al. (2010), Edwards (2010), Sadtler et al. (2014), Geweke (1996),

---

✉ Christian Aßmann  
christian.assmann@uni-bamberg.de

<sup>1</sup> Leibniz Institute for Educational Trajectories, Bamberg, Germany

<sup>2</sup> Chair of Survey Statistics and Data Analysis, Otto-Friedrich-Universität Bamberg, Bamberg, Germany

<sup>3</sup> Kiel Institute for the World Economy, Kiel, Germany

<sup>4</sup> Department of Economics, Ruhr-Universität Bochum, Bochum, Germany

and Baştürk et al. (2017). Factor models constitute a special case of reduced rank regression models, see Man and Culpepper (2020), Chan et al. (2018), and Aßmann et al. (2016) for corresponding Bayesian approaches. The same holds for vector error correction models, which are discussed in Koop et al. (2010) from a Bayesian perspective. Eventually, Zellner et al. (2014) relate reduced rank regression models to models with instrumental variables.

Typically, reduced rank regression models require identifying restrictions to come up with interpretable estimation 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). Moreover, the posterior distribution can exhibit multimodality, see Gelman and Rubin (1992), Lopes and West (2004), and Ročková and George (2016). With regard to factor models, which are a prominent class of reduced rank regression models, multimodality can occur if identification is reached by constraining the loading matrix to a positive lower triangular (PLT) matrix a priori as proposed by Geweke and Zhou (1996). More generally, if the constraints are imposed on particular elements of the loading matrix, inference results may depend on the ordering of the variables. This is likewise observed by Carvalho et al. (2008). Altogether, ex-ante identification, which is achieved by constraining the parameter space, may influence inference results for the quantities of interest, i.e., 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 Eroshova and Curtis (2017) suggest ex-post approaches to achieve directed inference on factors and loadings.<sup>1</sup> The post-processing algorithm proposed in Aßmann et al. (2016) addresses static and dynamic factor models whose scaling restrictions are formulated in terms of restricted moments of the prior distribution. This approach is hence not suited to handle factor models with orthonormality restrictions. If the restrictions are placed such that they reflect the classical model assumptions of Thurstone (1935), regarding zero-mean unit-scale uncorrelated factors, draws from the posterior distribution produce correlated factors with non-zero mean and non-unit scaling. This also carries over to the posterior factor estimates. As the major contribution of this paper, we provide a modified post-processing algorithm that yields perfectly orthonormal factor estimates. In fact, the modified approach allows for Bayesian inference on factors and loadings corresponding to inference on factors via principal component analysis in a frequentist setup. Furthermore, the approach can be applied not only to static factor models, but also to other reduced rank regression models with orthonormality restrictions. The new post-processing approach proposed in this paper is based on the sampler by Koop et al. (2010) for vector error correction models.

---

<sup>1</sup> 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.

Due to cointegration, vector error correction models also belong to the class of reduced rank models. Cointegration spaces are only identified up to an arbitrary linear combination of the corresponding 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. Enforcing linear restrictions a priori, however, can likewise induce estimation results to depend 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. Also, Villani (2006) argues that point estimates based on the method in Strachan (2003) may provide counter-intuitive interpretations and proposes an alternative ex-post point estimator for the cointegration space. Further, resulting Bayesian estimation is typically straightforward in terms of Markov chain Monte Carlo (MCMC) techniques. Villani (2005) proposes an MCMC sampler where identification is imposed by means of linear identifying restrictions on the cointegration space implying, via a Cauchy density on the unrestricted elements, a uniform distribution of the entire cointegration space. The resulting ex-post point estimator has the drawback that inference is only feasible for the cointegration space as a whole and not for a specific cointegration vector. Thus, identifying assumptions that permit structural interpretations cannot be incorporated. Finally, Koop et al. (2010) propose an approach that is order-invariant and computationally efficient. In this approach, cointegrating vectors stem from the Stiefel manifold, and Koop et al. (2010) also suggest an MCMC sampling approach that is easy to implement. Chan et al. (2018) adapt this sampling approach to factor models, albeit with a set of identifying restrictions different from the one discussed in the following. Note that in Koop et al. (2010) the identification problem is solved up to an orthogonal transformation. The sampler by Koop et al. (2010) is the starting point for the novel ex-post identification approach in the light of orthonormality restrictions proposed in this paper.

To implement the ex-post approach, we discuss two appropriate loss functions formulated in terms of squared Frobenius norms, either involving orthogonal transformations or orthogonal complements to assess estimation loss when the parameter space is partly restricted to the Stiefel manifold. We assess similarities and differences of the loss functions and derive corresponding post-processing algorithms that fit the model setup based on the MCMC sampling approach of Koop et al. (2010). The post-processing scheme allows to employ different additional identifying assumptions to reach inference for directed factors and loadings, like e.g., Varimax or Quartimax. In the context of vector error correction models, inference is possible for single cointegration vectors and not just the entire cointegration space. The suggested ex-post identification also provides the point estimator of Villani (2006) in terms of a decision-theoretic approach under an appropriately defined loss function. In a further contribution of this paper, we assess the possibility to use posterior predictive assessment for model selection in reduced rank models. For factor models with orthonormality restrictions, Chan et al. (2018) point at possible computational difficulties when using the Savage-Dickey density ratio to determine the number of factors. The same also applies to the calculation of marginal model likelihoods for

model selection. Thus, we propose to use 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.

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 toward identification in reduced rank regression models and discusses two alternative loss functions. 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

$$Y = \underset{(P \times T)}{\Pi} \underset{(P \times J)(J \times T)}{X} + \underset{(P \times Q)(Q \times T)}{\Xi} W + \underset{(P \times T)}{E}, \tag{1}$$

where  $Y = (y_1, \dots, y_T)$  is the  $P \times T$  matrix of dependent variables.  $E = (e_1, \dots, e_T)$  is the corresponding matrix of error terms with  $\text{vec}(E)$  following a multivariate normal distribution with mean zero and covariance  $I_T \otimes \Sigma$ , where  $\Sigma$  denotes a  $P \times P$  covariance matrix.<sup>2</sup> Further,  $W$  is a  $Q \times T$  matrix of explanatory variables with corresponding parameter matrix  $\Xi$  of size  $P \times Q$ .  $\Pi X$  incorporates the reduced rank structure of the model, where  $\Pi$  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.,  $Y = (\Delta y_1, \dots, \Delta y_T)$  with appropriately chosen initial conditions for  $y_0, y_{-1}, \dots, y_{2-K}$  and corresponding values for  $\Delta y_1, \Delta y_0, \dots, \Delta y_{1-K}$ . Accordingly, we have  $X = (y_0, \dots, y_{T-1})$  and  $W$  summarizing  $K$  lagged differences and exogenous variables, i.e.,

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

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

The reduced rank structure is captured via decomposing  $\Pi = \alpha\beta'$  with  $\alpha$  denoting a parameter matrix of dimension  $P \times R$  and  $\beta$  a parameter matrix of dimension

<sup>2</sup> Note that in the following,  $\odot$  and  $\otimes$  denote Hadamard and Kronecker tensor products as defined in Lütkepohl (1996), respectively, and  $I$  denotes an identity matrix of indicated size.

$J \times R$ , with  $R \ll \min\{J, P\}$ . In the case of the vector error correction model,  $\alpha$  governs the adjustment back to equilibrium and  $\beta'X$  denotes the  $T$  vectors of stationary departures from the  $R$  long run equilibria, while in the case of the static factor model,  $\alpha$  denotes the matrix of factor loadings and  $\beta$  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

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

For  $\Sigma$  and  $\Xi$ , we choose the commonly used conjugate priors as independent inverse Wishart and multivariate normal distributions with probability densities given as

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

and

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

The prior for  $\alpha$  and  $\beta$  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

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

Thus, the likelihood is invariant under this transformation. One part of the implied identification problem is typically addressed by restricting the scaling of  $\alpha$  or  $\beta$ , where restrictions on the scaling of  $\beta$  are prominent within the literature, see Villani (2005). We follow Strachan and Inder (2004), Villani (2006), and Koop et al. (2010) and restrict the scaling of  $\beta$  by assuming that  $\beta' \beta = I_R$ , i.e.,  $\beta$  is a semiorthogonal matrix and therefore has orthonormal column vectors. The corresponding prior distribution is hence defined on the Stiefel manifold.<sup>3</sup> Moreover,  $\alpha$  and  $\beta$  can be

<sup>3</sup> In restricting the parameter space to the Stiefel manifold via orthonormality restrictions, the model deviates from the static factor models applied, e.g., in Almann et al. (2016), where the scaling restrictions are formulated in terms of moment restrictions given as  $E[\beta' \beta] = I_R$  or  $E[\text{vec}(\beta') \text{vec}(\beta)'] = I_{RJ}$ . The orthonormality restrictions ensure that the estimated factors are uncorrelated and have unit variance. Without the orthonormality restrictions, the posterior distribution may exhibit correlation between the factors, and the factor scaling depends on the prior variance of the loadings and may deviate from one. The prior is hence constructed in relation to the free elements in  $\alpha$  and  $\beta$ , where the number of free elements can be calculated as follows. With the orthonormality restriction imposed on  $\beta$ , i.e.,  $\beta' \beta = I_R$ , it must hold that  $\beta_{j,r} = \pm \sqrt{1 - \sum_{i=1}^{j-1} \beta_{i,r}^2}$ , where the sign is determined by the equality  $\Pi = \alpha \beta'$ . This reduces the number of free elements in  $\beta$  by  $R$ .

replaced by  $\alpha^* = \alpha D$  and  $\beta^* = \beta D$ , where  $D$  is an orthogonal matrix.<sup>4</sup> Then  $D$  can be represented as  $D = E_D \cdot \text{diag}(\text{sgn}(l_{1,1}), \text{sgn}(l_{2,2}), \dots, \text{sgn}(l_{R,R}))$ , where  $E_D$  and  $L$  result from the QR decomposition  $\beta' = E_D L'$  with  $\text{sgn}(\cdot)$  providing the sign of the argument, see e.g., Golub and van Loan (2013). This reduces the number of free elements in  $\beta$  by another  $R(R - 1)/2$ . Thus, the number of free elements in  $\alpha$  is  $RP$ , and the number of free elements in  $\beta$  is  $RJ - R(R + 1)/2$ . Koop et al. (2010) suggest to use a prior setup for  $\alpha$  and  $\beta$  given as

$$\pi(\alpha, \beta | \Sigma) \propto |\Sigma|^{-\frac{R}{2}} \exp \left\{ -\frac{1}{2} \text{tr} [v^{-1} \beta' C_\tau^{-1} \beta \alpha' \Sigma^{-1} \alpha] \right\} \mathcal{I}(\beta' \beta = I_R), \tag{4}$$

where  $\mathcal{I}(\cdot)$  denotes the indicator matrix. This joint prior distribution corresponds to a marginal matrix angular central Gaussian distribution with parameter  $C_\tau$  for  $\beta$  and a conditional multivariate normal prior for  $\alpha$  conditional on  $\beta$  with expected value zero and covariance matrix  $v(\beta' C_\tau^{-1} \beta)^{-1} \otimes \Sigma$  with  $v$  denoting a shrinkage parameter, see Appendix 1 for further details.<sup>5</sup> Although this prior distribution for  $\alpha$  and  $\beta$  identifies the scaling of  $\beta$ , 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  $D$  is an orthogonal matrix. To formalize, define for any orthogonal  $R \times R$  matrix  $D$  the transformation

$$\begin{aligned} H(D)\Theta &= (\text{vec}(\alpha D)')', \text{vec}(\beta D)', \text{vec}(\Xi)', \text{vech}(\Sigma)')' \\ &= (\text{vec}(\tilde{\alpha})', \text{vec}(\tilde{\beta})', \text{vec}(\Xi)', \text{vech}(\Sigma)')' = \tilde{\Theta}, \end{aligned} \tag{5}$$

with

<sup>4</sup> As suggested by a reviewer, it is instructive to study a simplified setup, namely

$$\beta = \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix}.$$

The invariance of  $\beta$  with respect to orthogonal transformations implies that  $\beta' \beta = \beta^{*'} \beta^*$ , where  $\beta^* = \beta D$ , and  $D$  is a conformably dimensioned orthogonal matrix. In this simplified setup, we therefore have  $D \in \{-1, +1\}$ , and for any choice of  $D$  and  $\theta$ ,  $\beta$  is semiorthogonal, as  $(\pm \sin \theta)^2 + (\pm \cos \theta)^2 = 1$  always holds, and for  $R = 1$ , we need not care about mutually orthogonal columns in  $\beta$ . Further assume that  $\alpha \in \mathbb{R}$ , and hence,  $\Pi = \alpha \beta'$  is a bivariate vector. Now assume that  $\Pi$  has identical elements, which implies that  $\theta = \frac{\pi}{4}$  or  $\theta = \frac{3\pi}{4}$ , where the latter is simply the former with  $D = -1$ . In this simplified setup, it would be possible to restrict the prior and allow for  $\theta < \pi$  only, but instead, we choose an orthogonally invariant prior for  $\alpha$  and  $\beta$  as this is tractable for higher dimension of  $\beta$  as well.

<sup>5</sup> Following Koop et al. (2010),  $C_\tau$  is constructed as  $C_\tau = CC' + \tau \tilde{C} \tilde{C}'$ , where  $\tau$  denotes a scaling hyperparameter and  $\tilde{C}$  denotes the null space matrix of  $C$ , i.e., its orthogonal complement, and  $C = \hat{C}(\hat{C}' \hat{C})^{-\frac{1}{2}}$  where the elements of  $\hat{C}$ , denoted as  $c$ , are each independently drawn from the uniform density  $\frac{1}{2} \mathcal{I}(-1 < c < 1)$ . The choice of  $\tau = 1$  ensures a uniform prior on the Stiefel manifold, where  $0 < \tau < 1$  induces shrinkage of the cointegration space toward  $CC'$ . This setup ensures that  $C_\tau$  is positive definite as required and provides the flexibility to possibly consider different a priori orientations and scalings, see Appendix 1 for further details. A discussion of the matrix angular central Gaussian distribution is provided in Chikuse (1990, 2003).

$$H(D) = \begin{pmatrix} (D' \otimes I_{P+J}) & 0 \\ 0 & I_{PQ+P(P+1)/2} \end{pmatrix}. \tag{6}$$

Taking into account that the transformation described in Eq. (5) has no impact on the range of parameters and  $d\tilde{\Theta} = |\det(H(D)^{-1})|d\Theta$  with  $|\det(H(D)^{-1})| = 1$ , the marginal likelihood

$$\mathcal{M}(Y|X, W) = \int p(Y, \Theta|X, W)d\Theta = \int p(Y, \tilde{\Theta}|X, W)d\tilde{\Theta}, \tag{7}$$

is also invariant and thus is the posterior distribution  $p(\Theta|Y, X, W) = p(H(D)\Theta|Y, X, W)$ . The invariance arises hence for a model setup conditional on  $\beta$ , in contrast to the marginalized model setup used in Aßmann et al. (2016) to assess the invariance.

The considered model setup allows for sampling from the posterior distribution using an MCMC sampler which we call the *rotation-invariant* sampler in the following, as it ensures that  $\beta$  has zero-mean uncorrelated unit-length column vectors, but does not impose an orientation. Accordingly, we also refer to the obtained samples and the posterior distribution they originate from as rotation-invariant. 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  $\alpha\beta' = AB'$ .<sup>6</sup> As stated in Koop et al. (2010), this transformation yields

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

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

$$p(\Theta_{AB}|Y, X, W) \propto \mathcal{L}(Y|\Theta_{AB}, X, W)\pi(\Xi)\pi(A, B|\Sigma)\pi(\Sigma). \tag{8}$$

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  $\alpha$ ,  $\beta$ ,  $\Xi$ , and the corresponding rows and columns of  $\Sigma$ , see also Appendix 1. Following Koop et al. (2010) and considering the reparametrization in terms of  $A$  and  $B$ , allowing for efficient sampling, yields the following set of full conditional distributions. For the covariance matrix  $\Sigma$  we have

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

<sup>6</sup> Note that the absolute value of the determinant of the Jacobian matrix of this transformation is unity, since  $\frac{\partial \text{vec}(\beta)}{\partial \text{vec}(A')} = 0$ ,  $\frac{\partial \text{vec}(\alpha)}{\partial \text{vec}(A')} = (B'B)^{\frac{1}{2}} \otimes I_P$ , and  $\frac{\partial \text{vec}(\beta)}{\partial \text{vec}(B')} = (B'B)^{-\frac{1}{2}} \otimes I_P$ , where the normalized matrix  $(B'B)^{-\frac{1}{2}}$  is unaffected by changes in  $B$ , see also proof of Proposition 1 in Koop et al. (2010) and proof of Theorem 4.2 in Villani (2005).

**Table 1** Prior hyperparameter setting

Parameter	Distribution	Hyperparameter
$\Sigma$	Inverse Wishart	$\mu_\Sigma = 3, \Omega_\Sigma = 1/1000I_p$
$\Xi$	Multivariate normal	$\mu_\Xi = 0, \Omega_\Xi = 100I_{pQ}$
$\beta$	Matrix angular central Gaussian	$C_\tau = CC' + \tau\tilde{C}\tilde{C}'$ , with $\tau = 1$ and $\tilde{C}$ denoting the null space matrix of $C$ , and $C = \hat{C}(\hat{C}'\hat{C})^{-.5}$ where elements of $\hat{C}$ , say e.g., $c$ , are each independently drawn from uniform density $\frac{1}{2}\mathcal{I}(-1 < c < 1)$ , see Appendix 1 for further details
$\alpha$	Multivariate normal	$\mu_\alpha = 0, \Omega_\alpha = v(\beta'C_\tau^{-1}\beta)^{-1} \otimes \Sigma, v = 1$
$A$	Matrix angular central Gaussian	$C_A = v(B'C_\tau^{-1}B)^{-1} \otimes \Sigma, v = 1$
$B$	Multivariate normal	$\mu_B = 0, \Omega_B = v(A'\Sigma^{-1}A)^{-1} \otimes C_\tau, v = 1$

with  $\varphi_\Sigma = \mu_\Sigma + T - 1$  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  $\Xi$ , the full conditional distribution is given as

$$f(\Xi|Y, X, W, \alpha, \beta, \Sigma) \propto \exp \left\{ -\frac{1}{2}(\text{vec}(\Xi) - \varphi_\Xi)' \Psi_\Xi^{-1} (\text{vec}(\Xi) - \varphi_\Xi) \right\},$$

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  $\alpha$  we have

$$f(\alpha|Y, X, W, \beta, \Xi, \Sigma) \propto \exp \left\{ -\frac{1}{2}(\text{vec}(\alpha) - \varphi_\alpha)' \Psi_\alpha^{-1} (\text{vec}(\alpha) - \varphi_\alpha) \right\},$$

with  $\Psi_\alpha = ((\beta'XX'\beta \otimes \Sigma^{-1}) + \frac{1}{v}(\beta'C_\tau^{-1}\beta \otimes \Sigma^{-1}))^{-1}$  and  $\varphi_\alpha = \Psi_\alpha(\text{vec}(\Sigma^{-1}(Y - \Xi W)X'\beta))$ .

This draw of  $\alpha$  is then transformed into  $A = \alpha(\alpha'\alpha)^{-\frac{1}{2}}$ . Given  $A$ , we have the full conditional of  $B$  as

$$f(B|Y, X, W, A, \Xi, \Sigma) \propto \exp \left\{ -\frac{1}{2}(\text{vec}(B) - \varphi_B)' \Psi_B^{-1} (\text{vec}(B) - \varphi_B) \right\},$$

with  $\Psi_B = ((A'\Sigma^{-1}A \otimes X'X) + (A'\Sigma^{-1}A \otimes \frac{1}{v}C_\tau^{-1}))^{-1}$  and  $\varphi_B = \Psi_B(\text{vec}(X'(Y - \Xi W)\Sigma^{-1}A))$ .

The draw of  $B$  is then transformed into  $\beta = B(B'B)^{-\frac{1}{2}}$ . The prior hyperparameters used in estimation and simulation are documented in Table 1.

With this sampling algorithm, a posterior sample can be obtained. 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  $\alpha\beta'$ . To obtain estimates and posterior distributions for all quantities, we propose the following post-processing procedure.

### 3 Solving the identification problem via post-processing

In the following, we outline the post-processing approach for reduced rank models with the parameter space of  $\beta$  restricted to the Stiefel manifold. The presentation follows Aßmann et al. (2016), as it is a modification of the approach for factor models involving prior moment restrictions. The modification guarantees that the estimator based on post-processing lies on the Stiefel manifold and hence satisfies the imposed orthonormality restrictions.

A loss function  $L(\Theta^*, \Theta)$  defines a mapping of the estimators  $\Theta^*$  from the set of possible estimators and each of the parameter values  $\Theta$  within the parameter space onto the real line. The optimal estimator in terms of minimal expected loss is then defined as

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

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 Eq. (5) 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  $\Theta^*$  the loss invoked for any transformation of  $\Theta$  as described in Eq. (5).<sup>7</sup> As this minimization is done for each  $\Theta$ , the parameter space of  $\Theta$  is restricted ex-post via the corresponding first order conditions. Since the integral involved in the expected posterior loss is approximated via Monte Carlo (MC) methods, the corresponding minimization problem takes the form

$$\{ \{\hat{D}^{(s)}\}_{s=1}^S, \hat{\Theta}^* \} = \arg \min_{\{D^{(s)}\}_{s=1}^S, \Theta^*} \sum_{s=1}^S L_D(\Theta^*, H(D^{(s)}), \Theta^{(s)}), \tag{9}$$

subject to  $\beta^{*'}\beta^* = I_R$  and  $D^{(s)'}D^{(s)} = I_R$  for all  $s = 1, \dots, S$ , where  $\Theta^{(s)}$ ,  $s = 1, \dots, S$  denotes a sample from the rotation-invariant 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.

Loss functions are typically formulated in terms of squared Frobenius norms given as

$$L_D(\Theta^*, H(D^{(s)}), \Theta^{(s)}) = \text{tr} \left[ \mathcal{K}(\Theta^*, H(D^{(s)}), \Theta^{(s)}) \mathcal{K}(\Theta^*, H(D^{(s)}), \Theta^{(s)})' \right], \tag{10}$$

<sup>7</sup> 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), Subsection 5.1.

where  $\mathcal{K}(\cdot)$  denotes a function of the involved quantities  $\Theta^*$ ,  $\Theta^{(s)}$ , and  $H(D^{(s)})$  for  $s = 1 \dots, S$ .<sup>8</sup> A typical choice is the Euclidean norm, which results in our first loss function

$$\mathcal{K}_{\text{EOT}}(\Theta^*, H(D^{(s)}), \Theta^{(s)}) = \begin{pmatrix} \text{vec}(\alpha^{(s)}D^{(s)} - \alpha^*) \\ \text{vec}(\beta^{(s)}D^{(s)} - \beta^*) \\ \text{vec}(\Xi^{(s)} - \Xi^*) \\ \text{vech}(\Sigma^{(s)} - \Sigma^*) \end{pmatrix}'$$

which involves an orthogonal transformation affecting the directed quantities  $\alpha$  and  $\beta$  and is hence labelled as EOT. Larsson and Villani (2001) discuss the properties of the Euclidean norm without involving an orthogonal transformation for assessing losses on the Stiefel manifold and suggest the use of a squared Frobenius norm involving orthogonal complements (FOC) of the quantities defined on the Stiefel manifold. This concept is incorporated in our second loss function, which is

$$\mathcal{K}_{\text{FOC}}(\Theta^*, H(D^{(s)}), \Theta^{(s)}) = \begin{pmatrix} \text{vec}(\alpha^{(s)}D^{(s)} - \alpha^*) \\ \text{vec}(\beta_{\perp}^{*'}\beta^{(s)}D^{(s)}) \\ \text{vec}(\Xi^{(s)} - \Xi^*) \\ \text{vech}(\Sigma^{(s)} - \Sigma^*) \end{pmatrix}'$$

with  $\perp$  denoting the orthogonal complement, i.e., the null space, of the indicated matrix. We will consider both possibilities to assess the loss, illustrate and discuss the strong similarities of the two loss functions in assessing and minimizing expected loss, and point at the implied differences with regard to handling of the orthogonal invariance. For all elements of  $\Theta^*$  not referring to  $\beta$ , the estimator implied by the defined loss functions takes the form of an arithmetic mean given as

$$\overline{H(D)\Theta} = \frac{1}{S} \sum_{s=1}^S H(D^{(s)})\Theta^{(s)}. \tag{11}$$

For  $\beta$ , however, the elements of  $\overline{H(D)\Theta}$  do not fulfill the restriction  $\beta^{*'}\beta^* = I_R$ .

This restriction in combination with the considered EOT loss function turns the minimization of the posterior expected loss with regard to  $\beta$  into an orthogonal Procrustes problem. The defined EOT loss functions implies

$$\hat{\beta}_{\text{EOT}}^* = \arg \min_{\beta^*} \text{tr} \left[ -\beta^{*'} \sum_{s=1}^S \beta^{(s)}D^{(s)} \right], \quad \text{s.t.} \quad \beta^{*'}\beta^* = I_R. \tag{12}$$

<sup>8</sup> 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 and Andrew, and Barron, R. (1990) for a discussion of corresponding properties.

Note that the structure of the minimization problem for determining  $D^{(s)}$  as arising from Eq. (10) under the EOT loss takes the form

$$\arg \min_{D^{(s)}} \text{tr}[(\bar{\Lambda}^{(s)} D^{(s)} - \bar{\Lambda}^*)'(\bar{\Lambda}^{(s)} D^{(s)} - \bar{\Lambda}^*)], \quad \text{s.t. } D^{(s)'} D^{(s)} = I_R, \quad (13)$$

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 rotation-invariant sampler. The post-processing approach as discussed here transforms the output from the rotation-invariant sampler given a fixed point, i.e., the estimator.

In case the loss is assessed in terms of the FOC loss function, the estimator is based on orthogonal complements. Hence, in this loss function, the sequence of the orthogonal transformation matrices  $D^{(s)}, s = 1, \dots, S$  is not linked to the corresponding sequence  $\beta^{(s)}, s = 1, \dots, S$ . Considering  $D' = D^{-1}$  and

$$\begin{aligned} \text{vec}(\beta_{\perp}^{*'} \beta D)' \text{vec}(\beta_{\perp}^{*'} \beta D) &= \text{tr}[(\beta D)' \beta_{\perp}^{*'} \beta_{\perp}^{*'} (\beta D)] = \text{tr}[D' \beta' \beta_{\perp}^{*'} \beta_{\perp}^{*'} \beta D] \\ &= \text{tr}[\beta' \beta_{\perp}^{*'} \beta_{\perp}^{*'} \beta] = \text{tr}[\beta^{*'} \beta_{\perp} \beta_{\perp}' \beta^*] = \text{tr}[\beta^{*'} (I_R - \beta \beta') \beta^*], \end{aligned} \quad (14)$$

see also Larsson and Villani (2001), the minimization problem for  $\beta^*$  in case of the FOC loss function takes the form

$$\hat{\beta}_{\text{FOC}}^* = \arg \min_{\beta^*} \text{tr} \left[ \beta^{*'} \left( \sum_{s=1}^S \beta_{\perp}^{(s)} \beta_{\perp}^{(s)'} \right) \beta^* \right], \quad \text{s.t. } \beta^{*'} \beta^* = I_R. \quad (15)$$

Since

$$\text{tr} \left[ \beta^{*'} \left( \sum_{s=1}^S \beta_{\perp}^{(s)} \beta_{\perp}^{(s)'} \right) \beta^* \right] = \text{tr} \left[ \beta^{*'} \left( \sum_{s=1}^S (I_R - \beta^{(s)} \beta^{(s)'}) \right) \beta^* \right] = RS - \text{tr} \left[ \beta^{*'} \left( \sum_{s=1}^S \beta^{(s)} \beta^{(s)'} \right) \beta^* \right]$$

given the relations stated in Eq. (14), the minimization problem stated in (15) is equivalent to

$$\hat{\beta}_{\text{FOC}}^* = \arg \max_{\beta^*} \text{tr} \left[ \beta^{*'} \left( \sum_{s=1}^S \beta^{(s)} \beta^{(s)'} \right) \beta^* \right], \quad \text{s.t. } \beta^{*'} \beta^* = I_R. \quad (16)$$

The structure of the minimization problem for determining  $D^{(s)}$  as arising from Eq. (10) takes the form

$$\arg \min_{D^{(s)}} \text{tr}[(\alpha^{(s)} D^{(s)} - \alpha^*)'(\alpha^{(s)} D^{(s)} - \alpha^*)'], \quad \text{s.t. } D^{(s)'} D^{(s)} = I_R. \quad (17)$$

Given this, the similarities and differences between the two loss functions shall be discussed. The implied estimators for the undirected quantities  $\Sigma$  and  $\Xi$  are the same. The loss functions, however, have a slightly different approach toward assessing distances on the Stiefel manifold. As pointed out by Larsson and Villani (2001), the Euclidean distance not involving an orthogonal transformation between two  $P$ -dimensional vectors is strictly increasing with respect to the angle between

them. For instance, the Euclidean and FOC loss functions for  $R = 1$  are given as  $\propto 1 - \cos(\eta)$  and  $\propto \sin^2(\eta)$ , respectively, where  $\eta \in (0, \pi)$  denotes the angle between the two considered vectors on the Stiefel manifold. While the Euclidean loss function reaches its maximum at  $\pi$ , the FOC loss function increases between 0 and  $\pi/2$  and decreases between  $\pi/2$  and  $\pi$ . Thus, the situation with maximal loss for the Euclidean loss function is in fact the situation where both vectors span the same space.

This setup changes if the Euclidean loss function takes an orthogonal transformation into account, resulting in the EOT loss function. The sequence  $D^{(s)}$ ,  $s = 1, \dots, S$  captures the information about the orientation in the  $\alpha^{(s)}$  and  $\beta^{(s)}$ , respectively, thus solving the aforementioned problem in the Euclidean loss function. Consider as a simple example the case  $R = 1$ , where the orthogonal transformation takes the form  $\pm 1$ . The corresponding EOT loss function is given as  $\propto \min\{1 - \cos(\eta), 1 - \cos(\eta - \pi)\}$ , which also increases between 0 and  $\pi/2$  and decreases between  $\pi/2$  and  $\pi$ , just like the FOC loss function. Note, however, that the FOC loss function deliberately works with quantities in which unwanted orthogonal transformations in the  $\beta^{(s)}$  cancel out, see Eq. (15).<sup>9</sup> Hence, for arbitrary orthogonal transformations in the  $\alpha^{(s)}$  and  $\beta^{(s)}$ , which occur in the sampling process, there is no effect on the estimator for  $\beta^*$  if the FOC loss function is used, but the estimator for  $\beta^*$  is affected by such transformations if the EOT loss function is used. However, in such a situation, the orthogonal transformations are fully captured in the sequence  $D^{(s)}$ ,  $s = 1, \dots, S$ . If the FOC loss function is used, an according sequence is also obtained, albeit only based on the  $\alpha^{(s)}$ , see Eq. (17). Thus the estimator for  $\alpha^*$  changes under both loss functions for orthogonal transformations in the  $\alpha^{(s)}$  and  $\beta^{(s)}$ . To summarize, the FOC loss function is able to provide an estimator for  $\beta^*$  that is entirely unaffected by orthogonal transformations in the draws, whereas the EOT loss function is able to provide estimators for  $\alpha^*$  and  $\beta^*$  that are aligned with each other such that the  $D^{(s)}$ ,  $s = 1, \dots, S$  affect both of them in virtually the same manner.<sup>10</sup>

Both in simulation and empirical studies, the obtained estimates for  $\alpha^*$  and  $\beta^*$  undergo a final orthogonal mapping, either onto the parameters used in the simulation, or to satisfy an identification constraint, for instance, a lower triangular form with positive diagonal elements, or according to some criterion, such as Varimax. If the EOT loss function is used, the mapping involves the same orthogonal matrix for  $\alpha^*$  and  $\beta^*$ . For the FOC loss function, on the other hand, the information in the  $D^{(s)}$ ,  $s = 1, \dots, S$  affects the initial estimation of  $\alpha^*$ , but not that of  $\beta^*$ , which is an undirected estimator. In simulation studies, where the parameter values for  $\alpha$  and  $\beta$  are at hand, this simply means that the orthogonal mappings for  $\alpha^*$  and  $\beta^*$  are distinct. In applications, on the other hand, this is not possible. As the initially obtained estimates for  $\alpha^*$  and  $\beta^*$  are not aligned with each other in the first place, an orthogonal mapping that turns  $\alpha^*$  into a positive lower triangular matrix, see Sect. 2, or results

<sup>9</sup> Note that for an orthogonal matrix  $D$ , it holds for  $\tilde{\beta} = \beta D$  that  $\beta\beta' = \tilde{\beta}\tilde{\beta}'$  and  $\beta_{\perp}\beta'_{\perp} = \tilde{\beta}_{\perp}\tilde{\beta}'_{\perp}$ .

<sup>10</sup> As  $\alpha$  and  $\beta$  are sampled conditional on each other within the collapsed sampler, the rotational variation present in a single draw  $\alpha^{(s)}$  is highly correlated with the rotational variation present in the corresponding draw  $\beta^{(s)}$  for all  $s = 1, \dots, S$ , but nevertheless is not exactly the same.

in the Varimax solution for  $\alpha^*$ , would yield an arbitrary estimator if applied to  $\beta^*$ . If  $\beta^*$  is of interest beyond the mere cointegration space that it spans, it is therefore advisable to align the FOC estimators  $\alpha_{FOC}^*$  and  $\beta_{FOC}^*$  with each other prior to this step. Two approaches to achieve this are described below. As will be shown in the simulation study and the empirical illustration, not only the differences between the estimators for  $\alpha^*$  are negligible, but also those between the estimators for  $\beta^*$ . In this sense, this paper points out that the Euclidean norm involving an orthogonal transformation is well suited for estimators defined on the Stiefel manifold. Further, the equivalence of the optimization problems stated in Eqs. (15) and (16) implies that the estimator resulting from the FOC loss function is the posterior mean cointegration space (PMCS) estimator discussed in Villani (2006).

The following paragraphs outline how a solution for the reduced rank regression models can be obtained via sequential algorithms in case of the EOT and FOC loss functions. The post-processing approach as discussed here transforms the output from the rotation-invariant sampler given a fixed point, i.e., the estimator. The algorithms need an initialization with regard to  $\Theta^*$ , where we choose the last draw of the rotation-invariant sampler for convenience.

**EOT** For the EOT loss function, the algorithm takes the following steps.

**EOT 1** For given  $\Theta^*$  the minimization problem implied by Eq. (13) 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.

**EOT 1.1** Define  $Y_{D^{(s)}} = \bar{\Lambda}^{(s)'} \bar{\Lambda}^*$ .

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

**EOT 1.3** Obtain the orthogonal transformation matrix as  $D^{(s)} = U_{D^{(s)}} V_{D^{(s)}}'$ .  
For further details on the derivation of this solution, see Schönemann (1966).

**EOT 2** Choose  $\alpha^*$ ,  $\Xi^*$ , and  $\Sigma^*$  as implied by Eq. (11). With regard to  $\beta^*$ , the minimization problem given in Eq. (12) also takes the form of an orthogonal Procrustes problem, where the solution then involves the following calculations.

**EOT 2.1** Define  $S_\beta = \sum_{s=1}^S \beta^{(s)} D^{(s)}$ .

**EOT 2.2** Do the singular value decomposition  $S_\beta = U_\beta M_\beta V_\beta'$ , where  $U_\beta$  denotes the matrix of the eigenvectors of  $S_\beta S_\beta'$ , and  $V_\beta$  denotes the matrix

of eigenvectors of  $S'_\beta S_\beta$ . Further,  $M_\beta$  denotes a diagonal matrix of singular values, which are the square roots of the eigenvalues of  $S'_\beta S_\beta$ , which are also the nonzero eigenvalues of  $S_\beta S'_\beta$ .

**EOT 2.3** Obtain the semiorthogonal matrix  $\beta_{\text{EOT}}^* = U_\beta \mathcal{J}_{\text{EOT}} V'_\beta$ , where the matrix  $\mathcal{J}_{\text{EOT}} = [I_R \ 0_{R \times (P-R)}]'$  selects the  $R$  largest eigenvectors of  $S_\beta S'_\beta$  based on a corresponding implicit sorting. The information from the  $D^{(s)}$ ,  $s = 1, \dots, S$  matrices contained in  $S_\beta$  accordingly rotates the eigenvectors such that  $\beta_{\text{EOT}}^*$  is aligned with  $\alpha^*$ .

**FOC** For the FOC loss function, the sequential algorithm involves the following steps.

**FOC 1** For given  $\Theta^*$  the minimization problem implied by Eq. (17) resembles again an orthogonal Procrustes problem but involves only  $\alpha^*$  and  $\alpha^{(s)}$ ,  $s = 1, \dots, S$ . The solution involves the following calculations.

**FOC 1.1** Define  $\tilde{Y}_{D^{(s)}} = \alpha^{(s)'} \alpha^*$ .

**FOC 1.2** Do the singular value decomposition  $\tilde{Y}_{D^{(s)}} = \tilde{U}_{D^{(s)}} \tilde{M}_{D^{(s)}} \tilde{V}'_{D^{(s)}}$ , where  $\tilde{U}_{D^{(s)}}$  and  $\tilde{V}_{D^{(s)}}$  denote the matrix of eigenvectors of  $\tilde{Y}_{D^{(s)}} \tilde{Y}'_{D^{(s)}}$  and  $\tilde{Y}'_{D^{(s)}} \tilde{Y}_{D^{(s)}}$ , respectively, and  $\tilde{M}_{D^{(s)}}$  denotes a diagonal matrix of singular values, which are the square roots of the eigenvalues of  $\tilde{Y}_{D^{(s)}} \tilde{Y}'_{D^{(s)}}$  and  $\tilde{Y}'_{D^{(s)}} \tilde{Y}_{D^{(s)}}$ . Note that the nonzero eigenvalues of  $\tilde{Y}_{D^{(s)}} \tilde{Y}'_{D^{(s)}}$  and  $\tilde{Y}'_{D^{(s)}} \tilde{Y}_{D^{(s)}}$  are identical.

**FOC 1.3** Obtain the orthogonal transformation matrix as  $D^{(s)} = \tilde{U}_{D^{(s)}} \tilde{V}'_{D^{(s)}}$ .

**FOC 2** Choose  $\alpha^*$ ,  $\Xi^*$ , and  $\Sigma^*$  as implied by Eq. (11). With regard to  $\beta^*$ , the minimization problem under FOC loss given in Eq. (15) is solved by the following calculations, see also Lütkepohl (1996).

**FOC 2.1** Calculate the orthogonal complement  $\beta_\perp^{(s)}$  for each  $s = 1, \dots, S$  as the null space of the involved matrix, i.e.,  $\beta_\perp^{(s)} = \text{null}(\beta^{(s)'})$  for all  $s = 1, \dots, S$ .

**FOC 2.2** Define  $\tilde{S}_\beta = \sum_{s=1}^S \beta_\perp^{(s)} \beta_\perp^{(s)'}$ . As discussed above, step FOC 2.2 could alternatively be based on the optimization problem stated in Eq. (16). The resulting estimator consists of the eigenvectors corresponding to the  $R$  largest eigenvalues of the matrix  $\sum_{s=1}^S \beta^{(s)} \beta^{(s)'}$  and is thus the PMCS estimator of Villani (2006).

**FOC 2.3** Do the spectral decomposition of  $\tilde{S}_\beta = \tilde{V}_\beta \tilde{D}_\beta \tilde{V}'_\beta$ , where  $\tilde{V}_\beta$  denotes the matrix of the eigenvectors and  $\tilde{D}_\beta$  denotes the diagonal matrix of eigenvalues.

FOC 2.4 Obtain the semiorthogonal matrix  $\beta_{\text{FOC}}^* = \tilde{V}_\beta \mathcal{J}_{\text{FOC}}$ , where the matrix  $\mathcal{J}_{\text{FOC}} = [0_{R \times (P-R)} \ I_R]'$  selects the  $R$  smallest eigenvectors of  $\tilde{\mathcal{S}}_\beta$  based on a corresponding implicit sorting.

Note that  $\tilde{\mathcal{S}}_\beta$  in FOC 2.2 is unaffected by arbitrary orthogonal transformations of the  $\beta^{(s)}$ ,  $s = 1, \dots, S$ . Hence,  $\beta_{\text{FOC}}^*$  is an undirected estimator, whereas  $\alpha^*$  is a directed one. It is possible, however, to obtain an orthogonal mapping that aligns the estimators for  $\alpha^*$  and  $\beta_{\text{FOC}}^*$  with each other with regard to  $\Pi^* = \frac{1}{S} \sum_{s=1}^S \alpha^{(s)} \beta^{(s)'}.$  The required matrix  $D_A$  can be obtained via minimizing the squared Frobenius norm, i.e.,

$$D_A = \arg \min_D \text{tr}[(\alpha^*(\beta_{\text{FOC}}^* D)' - \Pi^*)'(\alpha^*(\beta_{\text{FOC}}^* D)' - \Pi^*)],$$

taking the form of an orthogonal Procrustes problem as well. This approach, however, only provides a directed point estimator for  $\tilde{\beta}_{\text{FOC}}^* = \beta_{\text{FOC}}^* D_A.$  To allow for inference, the orthogonal transformation matrices  $D^{(s)}$ ,  $s = 1, \dots, S$  from FOC 1.3 can be used to transform the corresponding  $\beta^{(s)}$ ,  $s = 1, \dots, S$  matrices. This aligns each  $\beta^{(s)}$  with the corresponding  $\alpha^{(s)}$ , and hence turns  $\beta_{\text{FOC}}^*$  into a directed estimator, as does the matrix  $D_A$  if the orthogonal transformation matrices are not applied. The cointegration space remains the same under this transformation, so information on the orientation and on the distribution of  $\beta$  is added, while none of the previously contained information is lost.

With regard to convergence of the post-processing algorithms, we have found that for arbitrary initial choices of  $\Theta^*$  taken from the rotation-invariant sampler output, less than ten iterations usually suffice to achieve convergence to a fixed point  $\hat{\Theta}^*$  providing the Bayes estimator. Convergence is assumed if the sum of squared deviations between two successive  $\hat{\Theta}^*$  does not exceed a predefined threshold value, where we use  $10^{-9}$ . In case of the EOT loss function, the iterative procedure of the algorithm suggests to use the transformed output of the rotation-invariant sample, i.e.,  $H(D^{(s)})\Theta^{(s)}$ , as input for the next iteration, thus reducing required computer memory capacities. In this line, De Vito et al. (2021) have found that also a single iteration may result in estimates sufficiently close to the final estimates. The post-processed posterior sample then provides 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. Within the supplementary material all necessary files to reestimate the empirical illustration are provided in form of a zip-archive.

### 4 Simulation and numerical experiments

To illustrate the properties of the suggested post-processing approach, we perform a simulation experiment for a vector error correction model. We show that the point estimator under EOT loss and the FOC estimator are extremely close, thus showing

the adequacy of the EOT and the FOC loss functions. 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  $K = 3$ . We assume throughout the simulation study  $y_0 = \dots = y_{2-K} = 0$  and correspondingly  $\Delta y_1 = y_1$  and  $\Delta y_0 = \dots, \Delta y_{1-K} = 0$  as initial conditions for the data generating process (DGP), where parameter values used within the DGP with regard to  $\alpha\beta'$  and thus  $\alpha$  and  $\beta$  are given in the first column of Table 2 and Table 3, respectively. To obtain a sample from the posterior distribution of  $\alpha$  and  $\beta$ , we run the rotation-invariant sampler of Koop et al. (2010) with  $S = 20,000$  after a burn-in phase of 5,000 iterations. Note that the rotation-invariant sampler ensures that the orthonormality restriction  $\beta'\beta = I_R$  holds for each draw.

Figure 1 illustrates the identification problem by showing the circular shape of the posterior distribution when plotting pairwise parameter trajectories arising from the rotation-invariant sampler of Koop et al. (2010). The  $\beta$  matrices are semiorthogonal, so their columns have unit length. An illustration analogous to Fig. 1 could show this in four dimensions only, hence we add Fig. 2, which depicts the distribution of the lengths of the row vectors (upper part) and the column vectors (lower part). Note that the vector lengths are invariant under the transformation described in Eq. (5), and that the distribution of the lengths of the column vectors of  $\beta$  is degenerate due to the scale restriction implied by  $\beta'\beta = I_R$ .

Since the quantity  $\alpha\beta'$  is invariant under the aforementioned transformation, we can use the output from the rotation-invariant Koop et al. (2010) sampler for inference on this quantity with no post-processing required. The second column of Table 2 shows the point estimates for  $\alpha\beta'$  based on the output of the rotation-invariant Koop et al. (2010) sampler, which is quite accurate. Note that in the EOT approach, the estimates for  $\alpha$  and  $\beta$ , and indeed the entire samples from the respective posterior distributions, can be transformed by a single orthogonal matrix to satisfy identifying assumptions. In the FOC approach, this is only possible for  $\alpha$ , as  $\beta$  is an undirected estimator. We can augment the information contained in the  $D^{(s)}$  matrices, however, to obtain directed point estimates for  $\beta$  and the according distributions, and then proceed in the same way as in the EOT approach. The third column shows the product of the point estimate for  $\alpha$  and the transpose of the point estimate of  $\beta$ . 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  $\alpha$  and  $\beta$ , which are not separately reported, and point out that both, sampling of  $\alpha$  and  $\beta$  within the rotation-invariant sampler, are sources of the orthogonal invariance. 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  $\alpha$  and  $\beta$  under the two considered loss functions FOC and EOT as shown in the last two columns of Table 2, are now almost identical to the invariant estimate. This, in turn, indicates that the estimates after post-processing for  $\alpha$  and  $\beta$  must be much more precise than if no post-processing is applied. Further, the results obtained for the two alternative loss functions are almost identical. This underlines the conceptual similarities as discussed above. The effects of post-processing the output of the rotation-invariant sampler can also be seen by looking at the shape of the posterior distribution of  $\beta$ .

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

DGP	$\beta\beta' = I_R$		post-processed $\beta\beta' = I_R$		
			FOC	EOT	
$\text{vec}(\alpha\beta')$	$\text{vec}(\hat{\alpha}\hat{\beta}')$	$\text{vec}(\hat{\alpha}\hat{\beta}')$	$\text{vec}(\hat{\alpha}\hat{\beta}')$	$\text{vec}(\hat{\alpha}\hat{\beta}')$	$\text{vec}(\hat{\alpha}\hat{\beta}')$
-0.0177	-0.0256	-0.0039	-0.0256	-0.0254	-0.0255
0.0826	0.0938	0.0148	0.0938	0.0943	0.0941
0.0559	0.0647	0.0103	0.0647	0.0652	0.0650
-0.0375	-0.0527	-0.0091	-0.0527	-0.0531	-0.0531
-0.3033	-0.3310	-0.0533	-0.3310	-0.3310	-0.3310
0.2151	0.2560	0.0408	0.2560	0.2561	0.2560
0.0249	0.0534	0.0083	0.0534	0.0535	0.0533
0.0800	0.0146	0.0021	0.0146	0.0147	0.0147
-0.4022	-0.4209	-0.0674	-0.4209	-0.4211	-0.4210
0.1913	0.2299	0.0364	0.2299	0.2299	0.2300
-0.0399	-0.0104	-0.0019	-0.0104	-0.0105	-0.0104
0.1626	0.0885	0.0144	0.0885	0.0884	0.0884
-0.0992	-0.1096	-0.0177	-0.1096	-0.1096	-0.1096
0.0710	0.0853	0.0136	0.0853	0.0853	0.0853
0.0086	0.0182	0.0028	0.0182	0.0181	0.0181
0.0258	0.0044	0.0007	0.0044	0.0045	0.0045

DGP refers to parameter values used in the data generating process. The second-last and last columns, respectively, show the estimates obtained under the suggested EOT loss function and the FOC loss function following Larsson and Villani (2001)

The lower panels of Fig. 1 indicate that the posterior distributions of the rows are no longer circular. They also allow for proper inference on the elements of  $\alpha$  and  $\beta$ .

The point estimates and highest posterior density intervals (HPDIs) in Table 3 show the results of two different transformations applied to the same output after post-processing in that respect. HPDIs were calculated using the HPDI estimation algorithm according to Chen and Shao (1999), following Chen et al. (2000). In the upper part of Table 3, we choose the transformation that minimizes the Frobenius norm of the distance between the estimates and the parameters  $\alpha$  and  $\beta$  used to simulate the data.

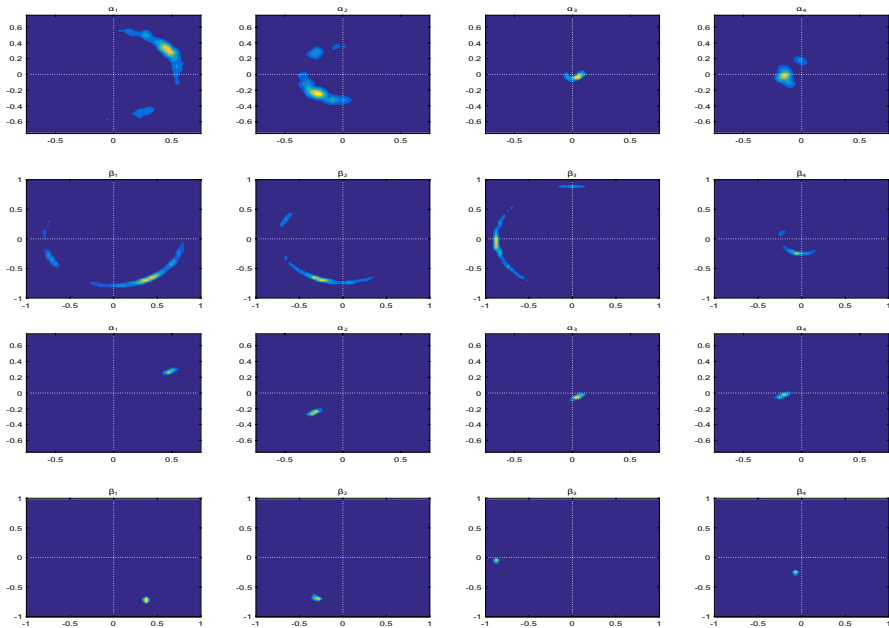
In the lower part of Table 3, on the other hand, we choose the rotation that minimizes the Frobenius norm of the distance between the EOT estimate and the FOC estimate for  $\beta$ .<sup>11</sup> The resulting point estimates for  $\alpha$  and  $\beta$  show that estimates obtained under the two alternative loss functions can both be well transformed to match the parameters of the DGP. In fact, the point estimates for the FOC and the EOT approach are very similar. The Frobenius norm of the distance for  $\alpha$  is 0.1306 in both cases. The Frobenius norm of the distance for  $\beta$  is 0.0172 for the FOC

<sup>11</sup> Note that the estimator obtained under FOC loss is equivalent to the PMCS estimator suggested by Villani (2006).

**Table 3** Parameter values and estimates for  $\alpha$  and  $\beta$ 

Mapped onto DGP						
DGP	FOC			EOT		
$\text{vec}(\alpha)$	$\text{vec}(\hat{\alpha})$	95% HPDI		$\text{vec}(\hat{\alpha})$	95% HPDI	
-0.1981	-0.2236	-0.2596	-0.1854	-0.2237	-0.2662	-0.1813
0.1991	0.2349	0.1929	0.2787	0.2346	0.1941	0.2746
0.0618	0.0867	0.0462	0.1267	0.0865	0.0473	0.1280
0.0170	-0.0350	-0.0754	0.0038	-0.0349	-0.0756	0.0042
0.4740	0.4995	0.4269	0.5723	0.4994	0.4278	0.5713
-0.2347	-0.2818	-0.3491	-0.2165	-0.2820	-0.3517	-0.2114
0.0399	0.0050	-0.0655	0.0805	0.0049	-0.0670	0.0815
-0.1861	-0.0983	-0.1709	-0.0328	-0.0982	-0.1693	-0.0312
$\text{vec}(\beta)$	$\text{vec}(\hat{\beta})$	95% HPDI		$\text{vec}(\hat{\beta})$	95% HPDI	
0.7308	0.7365	0.6840	0.7851	0.7358	0.6877	0.7863
0.6427	0.6370	0.5910	0.6820	0.6379	0.5955	0.6798
-0.0776	-0.0732	-0.1310	-0.0106	-0.0709	-0.1093	-0.0295
0.2166	0.2156	0.1807	0.2492	0.2159	0.1811	0.2490
0.2682	0.2767	0.2363	0.3169	0.2786	0.2515	0.3064
-0.3712	-0.3787	-0.4217	-0.3309	-0.3770	-0.4047	-0.3513
-0.8810	-0.8746	-0.8971	-0.8499	-0.8747	-0.8957	-0.8508
-0.1188	-0.1233	-0.1484	-0.0983	-0.1228	-0.1434	-0.1024
Mapped onto FOC						
DGP	FOC			EOT		
$\text{vec}(\alpha)$	$\text{vec}(\hat{\alpha})$	95% HPDI		$\text{vec}(\hat{\alpha})$	95% HPDI	
-0.0516	-0.0413	-0.0594	-0.0225	-0.0412	-0.0634	-0.0173
-0.0633	-0.0723	-0.0924	-0.0522	-0.0720	-0.0953	-0.0493
-0.0733	-0.0786	-0.1014	-0.0564	-0.0784	-0.1007	-0.0556
0.0737	0.0776	0.0548	0.1025	0.0775	0.0536	0.1014
0.5111	0.5457	0.4658	0.6253	0.5457	0.4653	0.6255
-0.3012	-0.3597	-0.4356	-0.2823	-0.3597	-0.4386	-0.2828
0.0056	-0.0369	-0.1171	0.0446	-0.0369	-0.1179	0.0457
-0.1717	-0.0697	-0.1445	0.0061	-0.0697	-0.1471	0.0046
$\text{vec}(\beta)$	$\text{vec}(\hat{\beta})$	95% HPDI		$\text{vec}(\hat{\beta})$	95% HPDI	
-0.7703	-0.7796	-0.8345	-0.7226	-0.7796	-0.8344	-0.7233
-0.3883	-0.3814	-0.4377	-0.3202	-0.3813	-0.4281	-0.3314
0.4879	0.4791	0.4186	0.5324	0.4790	0.4332	0.5214
-0.1339	-0.1314	-0.1701	-0.0893	-0.1314	-0.1688	-0.0918
-0.1123	-0.1055	-0.1376	-0.0729	-0.1055	-0.1137	-0.0970
-0.6325	-0.6354	-0.6619	-0.6083	-0.6354	-0.6473	-0.6225
-0.7377	-0.7353	-0.7660	-0.7041	-0.7353	-0.7481	-0.7235
-0.2077	-0.2108	-0.2237	-0.1973	-0.2108	-0.2192	-0.2018

DGP denotes data generating process, FOC the estimates obtained for the Frobenius norm involving orthogonal complements and EOT the estimates for the Euclidean norm involving an orthogonal transformation

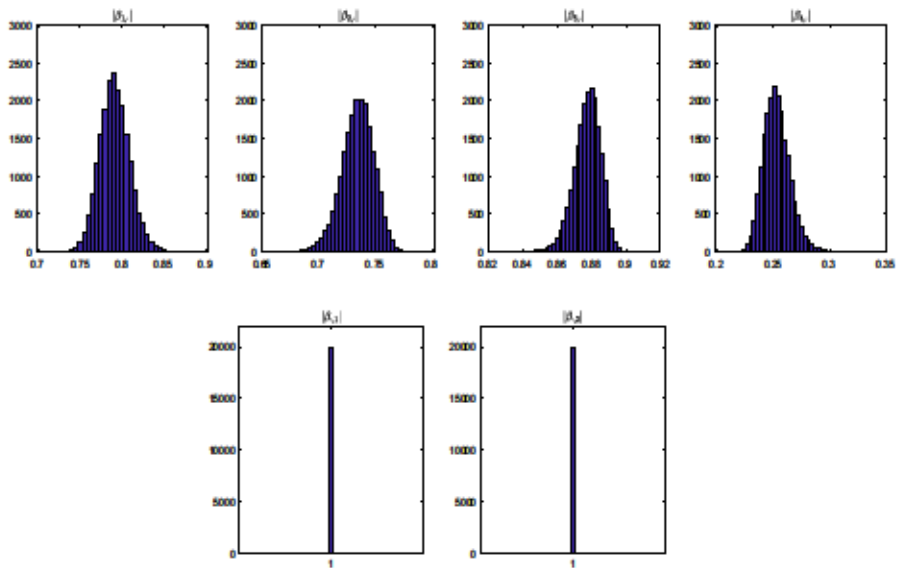


**Fig. 1** Distribution of row vectors of  $\alpha$  and of  $\beta$  without post-processing (first and second row) and with post-processing (third and fourth row)

approach and 0.0170 for the EOT approach. More substantial differences can be found for the HPDIs, which are overall slightly narrower for  $\alpha$ , but wider for  $\beta$  in the FOC approach, compared to the EOT approach. This is due to the fact that the FOC approach relies only on the information in  $\alpha$  to determine the transformation matrices, which results in reduced variation in the draws of  $\alpha$  and increased variation for the draws of  $\beta$ . Similar results are obtained when mapping the estimates onto the FOC estimate. The resulting Frobenius norms are necessarily identical here. Again, the HPDIs for  $\alpha$  are slightly narrower in the FOC approach, and those for  $\beta$  are wider.

### 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 Eq. (7), see Chib (1995) and Kass and Raftery (1995). In reduced rank regression models, model selection amounts to specifying the reduced rank dimension, i.e., the number of latent factors or the number of cointegrating vectors on the Stiefel manifold. Evaluating the marginal model likelihood for this type of model, however, involves substantial computational difficulties. Typically, the computation of the marginal model likelihood is based on the full conditional distributions including



**Fig. 2** Distribution of row vector lengths (top) and column vector lengths (bottom) of  $\beta$

the corresponding normalizing constants, see Chib (1995) and Chib and Jeliazkov (2001). As the functional form of the involved full conditional distribution for  $\beta$  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 nontrivial 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). Chan et al. (2018) report computational difficulties for the Savage-Dickey density ratio as well.

Given the aforementioned difficulties with common model selection approaches, 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. Note that this approach involves only invariant quantities and thus does not depend on an assumed loss function. For this purpose, a defined fraction of the data  $Y$ , e.g., within the range from 1 to 10%, are discarded. Hence, we can partition the data into discarded ( $Y^{\text{DIS}}$ ) and remaining ( $Y^{\text{REM}}$ ) observations, where  $O = PT = O_{\text{DIS}} + O_{\text{REM}}$  where  $O_{\text{DIS}}$  and  $O_{\text{REM}}$  denote the number of discarded and remaining observations, respectively. 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  $t = 1, \dots, T$ , where  $L_t^{\text{DIS}}$  and  $L_t^{\text{REM}}$ ,  $t = 1, \dots, T$  denote appropriately defined elimination matrices. The discarded observations are then augmented to the parameter vector and subject to sampling within the Gibbs sampling algorithm. Kaufmann and Kugler (2010) provide a similar approach for handling outlying and missing values. Posterior predictive assessment is then based on extending the Gibbs sampling scheme with the full conditional distributions of the discarded observations  $Y^{\text{DIS}}$ . For the factor model setup, this set of full conditional distributions is directly arising from the likelihood function given in Eq. (3). Since the likelihood in the static factor model setup (F) can be factorized as

$$\mathcal{L}_F(Y|\Theta, X, W) = \prod_{t=1}^T f_F(y_t|\Theta, Z_t),$$

the corresponding posterior predictive distribution is given as

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

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

For the vector error correction model (VECM) setup, sampling from the set of full conditional distributions of the discarded observation values is more elaborate. First, the VECM 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  $y_t^{\text{DIS}} = L_t^{\text{DIS}}y_t$  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  $\tilde{Y}_t = (y_t, y_{t-1}, \dots, y_{t-K})'$ ,

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

$\tilde{Y}_{t-1} = (y_{t-1}, y_{t-2}, \dots, y_{t-K-1})'$ ,  $\tilde{Z}_t = (Z_t, 0, \dots, 0)'$ , and  $\tilde{E}_t = (e_t, 0, \dots, 0)'$ . A sample of all discarded values  $Y^{\text{DIS}}$  in the VECM context can then be obtained by iteratively

sampling from the set of full conditional distributions of  $y_t$  for all  $y_t$  included in  $Y^{\text{DIS}}$ . 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) is implied via the backward smoothing recursion. Given the model setup, we have

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

where  $f(y_{t+k}|y_{t+k-1}, \dots, y_1, Z_1, \dots, Z_T, \Theta)$  for  $k = 0, \dots, K$  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

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

The corresponding full conditional expectation ( $\mu_{y_t}^{\text{SM}}$ ) and covariance ( $\Omega_{y_t}^{\text{SM}}$ ) are given as  $\mu_{y_t}^{\text{SM}} = \Omega_{y_t}^{\text{SM}} \kappa_{y_t}^{\text{SM}}$ , where

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

and

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

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

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

as implied by multivariate normal theory and with  $y_t^{\text{COM}}$  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  $SSE = \sum_{s=1}^S \frac{1}{O_{\text{DIS}}} \text{vec}(Y - Y_{\text{COM}}^{(s)})' \text{vec}(Y - Y_{\text{COM}}^{(s)})$ , 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  $s = 1, \dots, S$ .<sup>12</sup> Model selection using posterior predictive assessment does not require a post-processed 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, we vary 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 in a simulation study involving a static factor model setup. We set the fraction of discarded values to 1%, but the correspondingly implied partition is different for each incomplete data set.<sup>13</sup> 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  $J = 100$  five-dimensional vectors, containing the  $SSE$  values for the set of candidate values  $R^C$ . Hence  $SSE_{g,j}(\tilde{R})$  with  $g = 1, \dots, G$  and  $j = 1, \dots, J$  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, we obtain samples of size  $Q = 25$  and  $Q = 100$ , respectively, by drawing with replacement from the  $J = 100$  vectors of  $SSE$  values obtained for the set of candidate values  $R^C$ . Proceeding accordingly, we create  $L = 10,000$  such bootstrap samples for each of the  $G = 50$  simulated data sets per scenario. Each bootstrap sample can be referred to by the index set  $\mathcal{B}_{l,g}$ , which contains the indices

<sup>12</sup> 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 non-discarded values do not contribute to  $SSE$ . The  $SSE$  also involves a trade-off between computational burden and variance, where larger values of  $O_{\text{DIS}}$  increase computational burden, but also reduce the variance of the  $SSE$ .

<sup>13</sup> Note that this choice implies reduced computational burden but possibly larger uncertainty in the calculation of the  $SSE$  compared to higher fractions of discarded data. We have also inspected higher rates of discarded data, i.e., 2%, 5%, and 10%. While such higher rates of discarded values generally only slightly increase the accuracy of model detection, the involved computational burden arising from required storage capacity increases substantially. 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.

of the bootstrapped elements in bootstrap sample  $l$  for the simulated data set  $g$ . These indices range from 1 to  $J$ , and the index set may contain duplicate entries. We then calculate the average  $SSE$  for each bootstrap sample for every candidate value in  $R^C$ , i.e.,  $C_{l,g}(\tilde{R}) = \frac{1}{Q} \sum_{q \in \mathcal{B}_{l,g}} SSE_{g,q}(\tilde{R})$  for all  $\tilde{R} \in R^C$ . Eventually, we estimate for each  $l$  and each  $g$  the number of factors as the  $\tilde{R}$  from  $R^C$  that yields the lowest average  $SSE$ , i.e.,  $\hat{R}_{l,g} = \arg \min_{\tilde{R} \in R^C} C_{l,g}(\tilde{R})$ . With  $l \in \{1, \dots, L\}$  and  $g \in \{1, \dots, G\}$ , this gives us 500,000 estimates for the number of factors per scenario and per value chosen for  $Q$ , denoted as  $\hat{R}$ .

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 more pronounced for the scenarios with  $T = 100$ . In the following, the scenarios with  $P = \{20, 40, 80\}$  are summarized. In these 24 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 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.<sup>14</sup>

## 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 Fig. 3.<sup>15</sup> The data are demeaned and standardized. In the first step, the posterior predictive assessment is used to determine the appropriate number of factors. The method described in Sect. 5 is applied via generating  $J = 100$  incomplete data sets from the available one, and then using the bootstrap procedure to produce  $L = 10,000$  samples of size  $Q = 100$  to determine  $\hat{R}$ . Discarding 1% of the data,  $\hat{R} = 2$  is chosen in 96.9% of all cases, and  $\hat{R} = 3$  is chosen in 3.1% of all cases. If we discard 5% of the data,  $\hat{R} = 2$

<sup>14</sup> 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%.

<sup>15</sup> Data have been extracted from the European Central Bank's Statistical Data Warehouse on September 20, 2018. The considered currencies 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 Złoty (PLN), Romanian Leu (RON), Russian Rouble (RUB), Swedish Krona (SEK), Singapore Dollar (SGD), Thai Baht (THB), US Dollar (USD).

is chosen in 100% of all cases. We therefore estimate the model with two factors. To allow for directed estimates and inference for  $\beta$  in the FOC approach, we again augment the information contained in the sequence  $D^{(s)}$ ,  $s = 1, \dots, S$ .

After estimation, both the factors and factor loadings are orthogonally transformed to obtain an economically interpretable solution. The orthogonal transformation performed here turns the first factor into a *US Dollar factor*, maximizing the loading on the first factor for the exchange rate between the US Dollar and the Euro. The loading on the second factor for this pair of exchange rates is zero accordingly. The resulting loading matrix is hence a version of the positive lower triangular form of  $\alpha$  mentioned in Sect. 3. The required orthogonal transformation matrix  $D_{\text{PLT}}$  is obtained by reordering the rows of  $\alpha^*$  such that the USD/EUR exchange rate forms the first row, and all remaining rows of  $\alpha^*$  are shifted downwards. This yields the row-permuted matrix

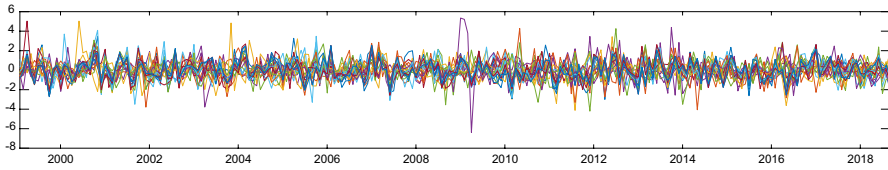
$$\alpha_P^* = \begin{pmatrix} 0_{1 \times (R-1)} & 1 \\ I_{R-1} & 0_{(R-1) \times 1} \end{pmatrix} \alpha^*.$$

Next, the QR decomposition  $\alpha_P^{*'} = E_D L'$  is used to obtain  $D_{\text{PLT}} = E_D \cdot \text{diag}(\text{sgn}(l_{1,1}), \dots, \text{sgn}(l_{R,R}))$ , as described in Sect. 2. The reported estimates then correspond to  $\alpha^* D_{\text{PLT}}$ . The estimates of rotated factor loadings and corresponding 95% HPDIs are shown in Table 5, whereas the estimated factors and corresponding 95% HPDIs are displayed in Fig. 4. Again, estimates resulting from both FOC and EOT loss functions are reported. The upper parts of Table 5 and Fig. 4 corresponds to FOC estimates, whereas the lower parts correspond to EOT estimates. Further, we like to stress that HPDIs can only be interpreted for a each loading on its own. Indeed, the rotated first factor is virtually perfectly correlated with the exchange rate between the US Dollar and the Euro, with a factor loading of 1.0026 as the FOC estimate and 1.0027 as the EOT estimate. The *US Dollar factor* also clearly shows the (flexible) peg between the US Dollar and the Hongkong Dollar, which has a factor loading of 1.0026 (FOC) and 1.0027 (EOT), respectively, and strong loadings with a number of south east Asian currencies, such as the Indonesian Rupiah, 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 Złoty slightly negatively. The second factor cannot be linked to any particular exchange rate, but shows the largest loadings for the Australian Dollar and the 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 *US Dollar factor* is substantially lower than that for the second factor. With regard to differences between the FOC and EOT loss function, we find the point estimates to be virtually identical, whereas the empirical illustration shows slightly broader HPDIs for the EOT loss function approach. This reflects that for the EOT loss function orthogonal invariance is contributed by  $\alpha$  and  $\beta$ .

**Table 4** Results of the posterior predictive simulation study

{Q,T,R,P}	$\hat{R}$ proportion for signal-to-noise ratio 1					$\hat{R}$ proportion for signal-to-noise ratio 10				
	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
{100,100,2,20}	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
{100,100,2,80}	0.0000	0.9949	0.0051	0.0000	0.0000	0.0000	0.9435	0.0562	0.0002	0.0001
{100,100,3,10}	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
{100,100,3,40}	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
{100,200,2,20}	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
{100,200,2,80}	0.0000	0.9561	0.0439	0.0000	0.0000	0.0000	0.9776	0.0224	0.0001	0.0000
{100,200,3,10}	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
{100,200,3,40}	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
{25,100,2,20}	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
{25,100,2,80}	0.0000	0.9753	0.0244	0.0003	0.0000	0.0000	0.9101	0.0841	0.0040	0.0019
{25,100,3,10}	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
{25,100,3,40}	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
{25,200,2,20}	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
{25,200,2,80}	0.0000	0.9427	0.0569	0.0004	0.0000	0.0000	0.9413	0.0547	0.0035	0.0005
{25,200,3,10}	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
{25,200,3,40}	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

$R$  denotes the number of factors used for data generation,  $T$  denotes the number of observations for each of the  $P$  variables,  $\hat{R}$  denotes the estimated number of factors. For each scenario (combination of  $T$ ,  $R$ ,  $P$  and signal-to-noise ratio), a distinct parameter set was simulated, for each of which  $G = 50$  distinct data sets were generated. Out of each of these data sets,  $J = 100$  different incomplete data sets were created, with 1% of the data discarded from each incomplete data set. The  $J$  incomplete data sets were used to create bootstrap samples of size  $Q$ . The number of bootstrap samples per scenario was  $L = 10,000$  throughout



**Fig. 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

### 7 Conclusion

This paper discusses the handling of orthonormality restrictions constraining parts of the parameter space to the Stiefel manifold in the context of reduced rank regressions via a novel post-processing algorithm. The output of the rotation-invariant sampler of Koop et al. (2010) is the starting point for the ex-post algorithm. We consider appropriate formulations of loss functions and propose corresponding post-processing algorithms for the posterior sample that allow for identification and directed inference. Thereby, the possibilities to conduct valid inference for cointegration vectors or factors restricted to the Stiefel manifold are extended. We discuss the differences and similarities implied by defining as a loss function for the parameters defined on the Stiefel manifold either a Euclidean distance function involving an orthogonal transformation or the Frobenius norm involving orthogonal complements for handling the orthogonal invariance present in the output of the rotation-invariant sampler. We illustrate how the post-processing works for vector error correction models in 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 computationally extremely demanding when the Stiefel manifold is involved. Overall, the results suggest that the two alternative loss functions lead to virtually equivalent results. 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.

### Appendix A: Prior distribution on $\alpha$ and $\beta$

The prior distribution for  $\alpha$  and  $\beta$  provided in Eq. (4) corresponds to a marginal matrix angular central Gaussian distribution with parameter  $C_\tau$  for  $\beta$ , see Chikuse (1990, 2003), and a conditional multivariate normal prior for  $\alpha$  conditional on  $\beta$  with expected value zero and covariance matrix  $\tau(\beta' C_\tau^{-1} \beta)^{-1} \otimes \Sigma$ , hence  $\pi(\alpha, \beta | \Sigma) \propto \pi(\alpha | \beta, \Sigma) \pi(\beta)$ , where

$$\pi(\beta) \propto |C_\tau|^{-\frac{R}{2}} |\beta' C_\tau^{-1} \beta|^{-\frac{p}{2}} I(\beta' \beta = I_R)$$

and

$$\pi(\alpha|\beta, \Sigma) \propto |\nu(\beta' C_\tau^{-1} \beta)^{-1} \otimes \Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \text{vec}(\alpha)' (\nu(\beta' C_\tau^{-1} \beta)^{-1} \otimes \Sigma)^{-1} \text{vec}(\alpha) \right\}.$$

This implies the form for  $\pi(\alpha, \beta|\Sigma)$  given in Eq. (4) since we have that

$$|\nu(\beta' C_\tau^{-1} \beta)^{-1} \otimes \Sigma|^{-\frac{1}{2}} = \nu^{-\frac{RP}{2}} |(\beta' C_\tau^{-1} \beta)|^{\frac{P}{2}} |\Sigma|^{-\frac{R}{2}}$$

and

$$\begin{aligned} -\frac{1}{2} \text{vec}(\alpha)' (\nu(\beta' C_\tau^{-1} \beta)^{-1} \otimes \Sigma)^{-1} \text{vec}(\alpha) &= -\frac{1}{2} \text{vec}(\alpha)' (\nu^{-1} (\beta' C_\tau^{-1} \beta) \otimes \Sigma^{-1}) \text{vec}(\alpha) \\ &= -\frac{1}{2} \text{tr} [\nu^{-1} \beta' C_\tau^{-1} \beta \alpha' \Sigma^{-1} \alpha]. \end{aligned}$$

Note that with regard to  $\pi(\beta)$  the choice  $\tau = 1$  implies a noninformative, i.e., uniform prior on the Stiefel manifold, see Chikuse (1990). Opting additionally for  $\nu = 1$  implies for  $\pi(\alpha|\beta, \Sigma)$  that scaling is proportional to  $(\beta' C_\tau^{-1} \beta)^{-1} \otimes \Sigma$  with no further shrinkage considered.

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

The posterior distribution is also invariant under the reparametrization of  $\alpha$  and  $\beta$  considered to facilitate efficient sampling, i.e.,

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

as  $\text{vec}(\alpha D)$  and  $\text{vec}(\beta D)$  imply  $\text{vec}(AD)$  and  $\text{vec}(BD)$ , i.e.,

$$(\alpha D)((\alpha D)'(\alpha D))^{-\frac{1}{2}} = \alpha(\alpha' \alpha)^{-\frac{1}{2}} D = AD, \tag{18}$$

and

$$(\beta D)((\beta D)'(\beta D))^{\frac{1}{2}} = \beta(\beta' \beta)^{\frac{1}{2}} D = BD. \tag{19}$$

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

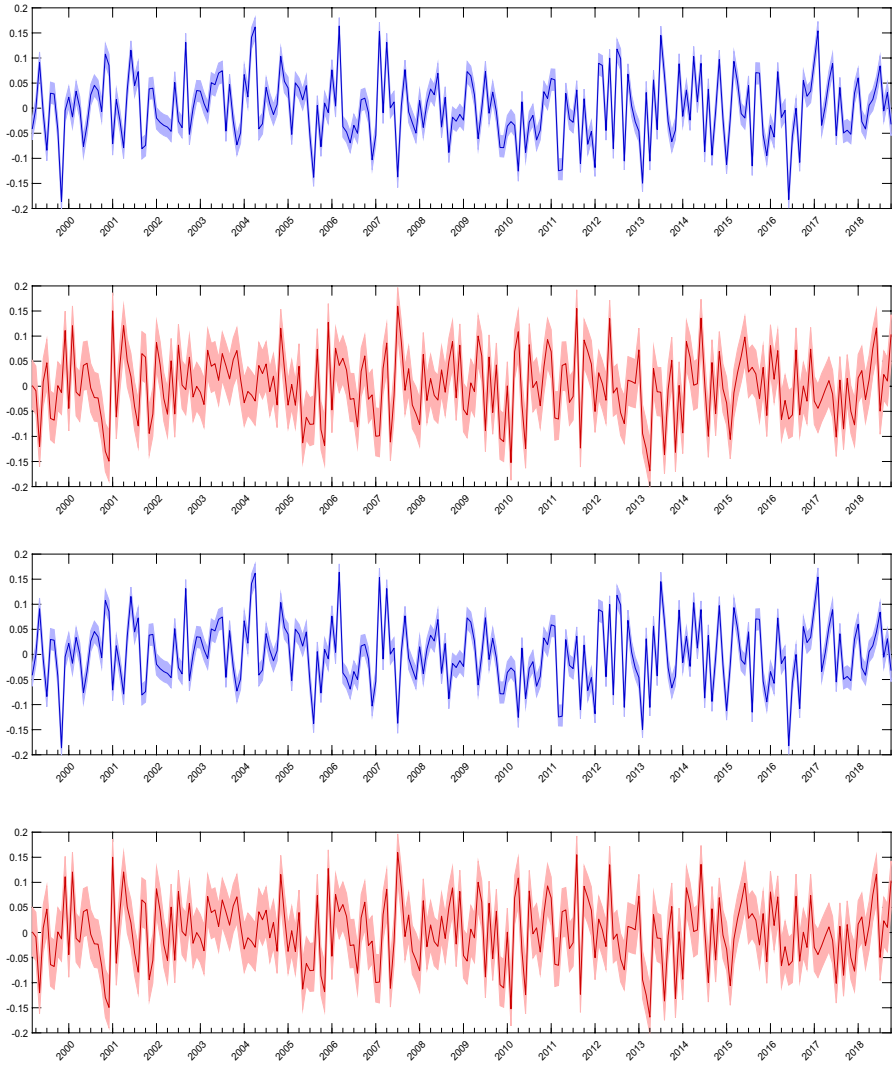
$$(\alpha' \alpha) = S_1 \Lambda S_1' \quad \text{and} \quad ((\alpha D)'(\alpha D)) = S_2 \Lambda S_2',$$

where  $S_1$  and  $S_2$  are orthogonal matrices and  $\Lambda$  is the diagonal matrix of eigenvalues of both the matrix  $\alpha' \alpha$  and the matrix  $D' \alpha' \alpha D$ , as multiplying  $\alpha$  by an orthogonal matrix  $D$  does not change the eigenvalues. Then

$$S_2 \Lambda S_2' = D' \alpha' \alpha D = D' S_1 \Lambda S_1' D,$$

**Table 5** Estimated factor loadings for the exchange rate data, after rotation (95% HPDIs in parentheses)

	FOC loss function		EOT loss function	
	first factor	second factor	first factor	second factor
	AUD/EUR	0.3105 (0.2218,0.3963)	0.7573 (0.6475,0.8613)	0.3105 (0.2231,0.3997)
CAD/EUR	0.4984 (0.3974,0.5986)	0.4665 (0.3480,0.5874)	0.4983 (0.3981,0.6015)	0.4665 (0.3470,0.5885)
CHF/EUR	0.5015 (0.3849,0.6155)	-0.0030 (-0.1344,0.1352)	0.5013 (0.3848,0.6151)	-0.0030 (-0.1427,0.1300)
CZK/EUR	-0.2222 (-0.3414,-0.1061)	0.4380 (0.2983,0.5855)	-0.2222 (-0.3416,-0.1051)	0.4380 (0.2909,0.5788)
DKK/EUR	0.2107 (0.0823,0.3367)	0.0465 (-0.1103,0.1948)	0.2106 (0.0822,0.3368)	0.0465 (-0.1080,0.1981)
GBP/EUR	0.2813 (0.1653,0.4025)	0.3611 (0.2181,0.5000)	0.2813 (0.1621,0.4005)	0.3611 (0.2176,0.5009)
HKD/EUR	1.0026 (0.9656,1.0403)	0.0033 (-0.0430,0.0510)	1.0027 (0.9647,1.0395)	0.0033 (-0.0367,0.0454)
IDR/EUR	0.8161 (0.7479,0.8841)	0.3681 (0.2889,0.4473)	0.8161 (0.7493,0.8853)	0.3681 (0.2869,0.4456)
JPY/EUR	0.7128 (0.6171,0.8099)	-0.0858 (-0.1978,0.0253)	0.7127 (0.6193,0.8119)	-0.0858 (-0.1973,0.0301)
KRW/EUR	0.5155 (0.4268,0.6059)	0.5954 (0.4844,0.7003)	0.5155 (0.4302,0.6111)	0.5953 (0.4836,0.7011)
MXN/EUR	0.1246 (0.0116,0.2367)	0.5506 (0.4171,0.6908)	0.1246 (0.0125,0.2407)	0.5505 (0.4171,0.6907)
MYR/EUR	0.8128 (0.7368,0.8887)	0.2450 (0.1572,0.3333)	0.8127 (0.7370,0.8885)	0.2450 (0.1567,0.3345)
NOK/EUR	0.0679 (-0.0448,0.1688)	0.6518 (0.5191,0.7837)	0.0679 (-0.0463,0.1705)	0.6518 (0.5163,0.7806)
NZD/EUR	0.3691 (0.2656,0.4759)	0.5163 (0.3919,0.6449)	0.3690 (0.2643,0.4763)	0.5163 (0.3877,0.6429)
PHP/EUR	0.8566 (0.7821,0.9290)	0.1000 (0.0138,0.1843)	0.8566 (0.7822,0.9291)	0.1000 (0.0167,0.1883)
PLN/EUR	-0.3952 (-0.4949,-0.2975)	0.6244 (0.5024,0.7420)	-0.3951 (-0.4943,-0.2939)	0.6243 (0.5010,0.7421)
RON/EUR	-0.0813 (-0.2097,0.0500)	-0.0332 (-0.1868,0.1219)	-0.0813 (-0.2099,0.0501)	-0.0331 (-0.1899,0.1189)
RUB/EUR	0.2691 (0.1555,0.3824)	0.4639 (0.3285,0.6022)	0.2691 (0.1572,0.3858)	0.4639 (0.3319,0.6057)
SEK/EUR	-0.0654 (-0.1799,0.0512)	0.5302 (0.3864,0.6744)	-0.0654 (-0.1862,0.0474)	0.5301 (0.3861,0.6745)
SGD/EUR	0.8645 (0.8141,0.9132)	0.4431 (0.3847,0.5017)	0.8645 (0.8156,0.9140)	0.4431 (0.3886,0.5004)
THB/EUR	0.8269 (0.7643,0.8967)	0.3663 (0.2903,0.4434)	0.8269 (0.7593,0.8920)	0.3663 (0.2904,0.4430)
USD/EUR	1.0026 (0.9664,1.0394)	0.0000 (-0.0459,0.0474)	1.0027 (0.9665,1.0394)	0.0000 (-0.0414,0.0410)



**Fig. 4** Estimated factors for the exchange rate data, after rotation. Blue denotes the *US dollar factor*, and red denotes the *second factor*. Shaded areas denote 95% HPDIs. Top two graphs show estimates with the FOC loss function, bottom two graphs show estimates with the EOT loss function

implying  $S_2 = D' S_1$ . Then from the definition of the square root matrix given, e.g., in Abadir and Magnus (2005), we have

$$(D' \alpha' \alpha D)^{-\frac{1}{2}} = S_2 \Lambda^{-\frac{1}{2}} S_2' = D' S_1 \Lambda^{-\frac{1}{2}} S_1' D = D' (\alpha' \alpha)^{-\frac{1}{2}} D.$$

Therefore, we obtain

$$\alpha D((\alpha D)'(\alpha D))^{-\frac{1}{2}} = \alpha D D'(\alpha' \alpha)^{-\frac{1}{2}} D = \alpha(\alpha' \alpha)^{-\frac{1}{2}} D = AD. \quad (20)$$

The result from Eq. (3) can be derived analogously, using the square root of the elements in  $\Lambda$  instead of the inverse square root. Inserting  $AD$  and  $BD$  into the corresponding posterior distribution provided in Eq. (8) reveals invariance under the orthogonal transformation described in Eq. (5).

**Acknowledgements** We thank Roman Liesenfeld, Jörg Breitung, and participants of the statistics seminar of the University of Cologne as well as participants 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).

**Funding** Open Access funding enabled and organized by Projekt DEAL.

## Declarations

**Conflict of interest** The authors declare to have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abadir, K., Magnus, J.: *Matrix Algebra*. Cambridge University Press, *Econometric Exercises* (2005)
- Aguilar, O., West, M.: Bayesian dynamic factor models and portfolio allocation. *J. Bus. Econom. Stat.* **18**(3), 338–357 (2000)
- Aßmann, C., Boysen-Hogrefe, J., Pape, M.: Bayesian analysis of static and dynamic factor models: an ex-post approach towards the rotation problem. *J. Econom.* **192**(1), 190–206 (2016)
- Baştürk, N., Hoogerheide, L., van Dijk, Herman K.: Bayesian analysis of boundary and near-boundary evidence in econometric models with reduced rank. *Bayesian Anal.* **12**(3), 879–917 (2017)
- Butler, R.W., Wood, A.T.: Laplace approximation for Bessel functions of matrix argument. *J. Comput. Appl. Math.* **155**(2), 359–382 (2003)
- Carvalho, C.M., Chang, J., Lucas, J.E., Nevins, J.R., Wang, Q., West, M.: High-dimensional sparse factor modeling: applications in gene expression genomics. *J. Am. Stat. Assoc.* **103**(484), 1438–1456 (2008)
- Celex, G.: 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 (1998)
- Celex, G., Hurn, M., Robert, C.: Computational and inferential difficulties with mixture posterior distributions. *J. Am. Stat. Assoc.* **95**(451), 957–970 (2000)
- Chan, J., Leon-Gonzalez, R., Strachan, R.W.: Invariant inference and efficient computation in the static factor model. *J. Am. Stat. Assoc.* **113**(522), 819–828 (2018)
- Chen, M.-H., Shao, Q.-M.: Monte Carlo estimation of Bayesian credible and HPD intervals. *J. Comput. Graph. Stat.* **8**(1), 69–92 (1999)

- Chen, M.-H., Shao, Q.-M., and Ibrahim, J. G.: Computing Bayesian Credible and HPD Intervals, pages 213–235. Springer New York, New York, NY (2000)
- Cheng, M.-Y., Hall, P., Turlach, B.A.: High-derivative parametric enhancements of nonparametric curve estimators. *Biometrika* **86**(2), 417–428 (1999)
- Chib, S.: Marginal likelihood from the Gibbs output. *J. Am. Stat. Assoc.* **90**(432), 1313–1321 (1995)
- Chib, S., Jeliazkov, I.: Marginal likelihood from the metropolis-hastings output. *J. Am. Stat. Assoc.* **96**(453), 270–281 (2001)
- Chib, S., Nardari, F., Shephard, N.: Analysis of high dimensional multivariate stochastic volatility models. *J. Econom.* **134**(2), 341–371 (2006)
- Chikuse, Y.: The matrix angular central Gaussian distribution. *J. Multivar. Anal.* **33**(2), 265–274 (1990)
- Chikuse, Y.: Statistics on Special Manifolds. Lecture notes in statistics, vol. 174. Springer, New York (2003)
- Clarke, B., Barron, A.R.: Information-theoretic asymptotics of Bayes methods. *IEEE Trans. Inf. Theory* **36**, 453–471 (1990)
- Crowther, N.: The exact non-central distribution of a quadratic form in normal vectors. *South Afr. Stat. J.* **9**(1), 27–36 (1975)
- De Vito, R., Bellio, R., Trippa, L., Parmigiani, G.: Bayesian multi-study factor analysis for high-throughput biological data. *Ann. Appl. Stat.* **15**(4), 1723–1741 (2021)
- Edwards, M.C.: A Markov chain Monte Carlo approach to confirmatory item factor analysis. *Psychometrika* **75**(3), 474–497 (2010)
- Erosheva, E.A., Curtis, S.M.: Dealing with reflection invariance in Bayesian factor analysis. *Psychometrika* **82**(2), 295–307 (2017)
- Friel, N., Pettitt, A.N.: Marginal likelihood estimation via power posteriors. *J. R. Stat. Soc. Ser. B* **70**(3), 589–607 (2008)
- Frühwirth-Schnatter, S. and Lopes, H. (2018). Sparse Bayesian Factor Analysis when the Number of Factors is Unknown. ArXiv e-prints
- Gelman, A., Meng, X.-L.: Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. *Stat. Sci.* **13**(2), 163–185 (1998)
- Gelman, A., Meng, X.-L., Stern, H.: Posterior predictive assessment of model fitness via realized discrepancies. *Stat. Sin.* **6**(4), 733–760 (1996)
- Gelman, A., Rubin, D.B.: Inference from iterative simulation using multiple sequences. *Stat. Sci.* **7**(4), 457–472 (1992)
- Geweke, J.: Bayesian reduced rank regression in econometrics. *J. Econom.* **75**(1), 121–146 (1996)
- Geweke, J., Zhou, G.: Measuring the pricing error of the arbitrage pricing theory. *Rev. Financ. Stud.* **9**(2), 557–587 (1996)
- Golub, G.H., van Loan, C.F.: *Matrix Computations*, 4th edn. The Johns Hopkins University Press (2013)
- 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.: Bessel functions of matrix argument. *Ann. Math.* **61**(3), 474–523 (1955)
- Hoff, P.D.: Simulation of the matrix Bingham-von Mises-Fisher distribution, with applications to multivariate and relational data. *J. Comput. Graph. Stat.* **18**(2), 438–456 (2009)
- Jasra, A., Holmes, C., Stephens, D.: Markov Chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Stat. Sci.* **20**(1), 50–67 (2005)
- Johansen, S.: Statistical analysis of cointegration vectors. *J. Econ. Dyn. Control* **12**(2–3), 231–254 (1988)
- Johansen, S.: Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica* **59**(6), 1551–1580 (1991)
- Kass, R.E., Raftery, A.E.: Bayes factors. *J. Am. Stat. Assoc.* **90**(430), 773–795 (1995)
- Kaufmann, S., Kugler, P.: A monetary real-time conditional forecast of euro area inflation. *J. Forecast.* **29**(4), 388–405 (2010)
- Kleibergen, F., Paap, R.: Priors, Posteriors and Bayes Factors for a Bayesian Analysis of Cointegration. *J. Econom.* **111**(2), 223–249 (2002)
- Kleibergen, F., van Dijk, H.: On the shape of the likelihood/posterior in cointegration models. *Econom. Theory* **10**(3–4), 514–551 (1994)
- Koev, P., Edelman, A.: The efficient evaluation of the hypergeometric function of a matrix argument. *Math. Comput.* **75**(254), 833–847 (2006)
- Koop, G., León-González, R., Strachan, R.W.: Efficient posterior simulation for cointegrated models with priors on the cointegration space. *Econom. Rev.* **29**(2), 224–242 (2010)

- Kristof, W.: Die beste orthogonale Transformation zur gegenseitigen Überführung zweier Factorenmatrizen. *Diagnostica* **10**, 87–90 (1964)
- Kume, A.: Saddlepoint approximations for the Bingham and Fisher-Bingham normalising constants. *Biometrika* **92**(2), 465–476 (2005)
- Kume, A., Preston, S.P., Wood, A.T.A.: Saddlepoint approximations for the normalizing constant of Fisher-Bingham distributions on products of spheres and Stiefel manifolds. *Biometrika* **100**(4), 971–984 (2013)
- Larsson, R., Villani, M.: A distance measure between cointegration spaces. *Econ. Lett.* **70**(1), 21–27 (2001)
- Lopes, H.F., West, M.: Bayesian model assessment in factor analysis. *Stat. Sin.* **14**, 41–67 (2004)
- Lütkepohl, H.: *Handbook of Matrices*. Wiley, Chichester and New York (1996)
- Man, A.X., Culpepper, S.A.: A mode-jumping algorithm for Bayesian factor analysis. *J. Am. Stat. Assoc.* **114**, 1–14 (2020)
- Mathai, A.M., Provost, S.B., Hayakawa, T.: *Bilinear Forms and Zonal Polynomials*. Lecture notes in statistics, vol. 102. Springer-Verlag, New York (1995)
- Meng, X.-L., Wong, W.H.: Simulating ratios of normalizing constants via a simple identity: a theoretical exploration. *Stat. Sin.* **6**, 831–860 (1996)
- Neal, R.M.: Annealed importance sampling. *Stat. Comput.* **11**(2), 125–139 (2001)
- Ročková, V., George, E.I.: Fast Bayesian factor analysis via automatic rotations to sparsity. *J. Am. Stat. Assoc.* **111**(516), 1608–1622 (2016)
- Sadtlir, P.T., Quick, K.M., Golub, M.D., Chase, S.M., Ryu, S.I., Tyler-Kabara, E.C., Yu, B.M., Batista, A.P.: Neural constraints on learning. *Nature* **512**(7515), 423–426 (2014)
- Schönemann, P.H.: A generalized solution to the orthogonal procrustes problem. *Psychometrika* **31**(1), 1–10 (1966)
- Stephens, M.: Dealing with label switching in mixture models. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **62**(4), 795–809 (2000)
- Strachan, R.W.: Valid Bayesian estimation of the cointegrating error correction model. *J. Bus. Econ. Stat.* **21**(1), 185–195 (2003)
- Strachan, R.W., Inder, B.: Bayesian analysis of the error correction model. *J. Econ.* **123**(2), 307–325 (2004)
- Strachan, R.W., van Dijk, H.K.: Bayesian model selection with an uninformative prior. *Oxford Bull. Econ. Stat.* **65**, 863–876 (2003)
- Thurstone, L.L.: *The Vectors of Mind*. The University of Chicago Press, Chicago (1935)
- Villani, M.: Bayesian reference analysis of cointegration. *Econom. Theory* **21**(02), 326–357 (2005)
- Villani, M.: Bayesian point estimation of the cointegration space. *J. Econom.* **134**(2), 645–664 (2006)
- 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., van Dijk, H.K.: Bayesian analysis of instrumental variable models: acceptance-rejection within direct Monte Carlo. *Econom. Rev.* **33**(1–4), 3–35 (2014)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.