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

Working Paper

The directional identification problem in Bayesian factor analysis: An ex-post approach

Economics Working Paper, No. 2012-11

Provided in Cooperation with:
Christian-Albrechts-University of Kiel, Department of Economics

Suggested Citation: Aßmann, Christian; Boysen-Hogrefe, Jens; Pape, Markus (2012) : The directional identification problem in Bayesian factor analysis: An ex-post approach, Economics Working Paper, No. 2012-11, Kiel University, Department of Economics, Kiel

This Version is available at:
http://hdl.handle.net/10419/65679

C | A | U
Christian-Albrechts-Universität zu Kiel
Department of Economics
Economics Working Paper
No 2012-11

the directional identification problem in bayesian factor analysis: an ex-post approach

by Christian Aßmann, Jens Boysen-Hogrefe, and Markus Pape

issn 2193-2476

The Directional Identification Problem in Bayesian Factor Analysis: An Ex-Post Approach

Christian Aßmann
Chair of Statistics and Econometrics and National Educational Panel Study
Otto-Friedrich-Universität Bamberg, Germany

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

Markus Pape*
Institute of Statistics and Econometrics
Christian-Albrechts-Universität Kiel, Germany

October 9, 2012

Abstract

Due to their well-known indeterminacies, factor models require identifying assumptions to guarantee unique parameter estimates. For Bayesian estimation, these identifying assumptions are usually implemented by imposing constraints on certain model parameters. This strategy, however, may result in posterior distributions with shapes that depend on the ordering of cross-sections in the data set. We propose an alternative approach, which relies on a sampler without the usual identifying constraints. Identification is reached ex-post based on a Procrustes transformation. Resulting posterior estimates are ordering invariant and show favorable properties with respect to convergence and statistical as well as numerical accuracy.

JEL classification: C11; C31; C38; C51; C52

Keywords: Bayesian Estimation; Factor Models; Multimodality; Ordering Problem; Orthogonal Transformation

*Corresponding author: M. Pape, Tel.: +49-431-8803399, *E-mail address:* markus.pape@stat-econ.uni-kiel.de

1 Introduction

Bayesian Factor analysis is a well-established tool in econometrics with a host of applications in economics and finance (e.g. Geweke and Zhou, 1996; Otrok and Whiteman, 1998; Kose et al., 2003; Bernanke et al., 2005). Latent factors influence observable data through factor loadings, where identifying assumptions are required to obtain unique estimates for the factors and loadings. Geweke and Zhou (1996) suggested an identification scheme that has been used in many applications, sometimes in the slightly modified form of Aguilar and West (2000). It is also common to use overidentifying restrictions, where certain factor loadings are set to zero, implying the assumption that the corresponding cross-sectional units are not affected by the corresponding factors (e.g. by Kose et al., 2003). Especially for a large number of cross-sections, this has led to the development of sparse factor models (West, 2003). In a different approach, Frühwirth-Schnatter and Lopes (2009) develop a sampler that selects the cross-sections used for identification during the sampling process.

For the exactly identified factor model, the ordering of the cross-sections should have no effect on the inference results. Lopes and West (2004), however, find that model selection criteria used to choose the number of factors are influenced by the way the cross-sections are ordered. Carvalho (2006) and Carvalho et al. (2008) find that parameter estimates differ, depending on the ordering of the data defining which are the first K observations serving as the *founders* of the model. They develop a hierarchical approach to find the most appropriate subset of cross-sections to be used for identification.

We will address the causes of these difficulties, analyzing the effect of the set of identifying constraints suggested by Geweke and Zhou (1996). These constraints are implemented by using truncated prior distributions. Hence, the restrictions guarantee a unique maximum of the likelihood underlying the posterior distribution. They do not, however, guarantee the non-existence of local modes. This has already been observed by Rubin and Thayer (1982, 1983) in the context of maximum likelihood estimation of explanatory and confirmatory factor models. The choice of constraints influences the shape of the likelihood and thus the posterior distribution, effecting the behavior of the Gibbs sampler. Consequently, ex-ante identification influences the inference on the parameters of the factor model and functions of these with respect to directed parameters, like factor loadings.

Instead of enforcing constraints on the parameter space ex-ante, leading to the aforementioned undesirable properties, we propose to identify the parameters ex-post based on orthogonal transformations of the unconstrained Gibbs sampler output. Ex-post identification approaches are well known in the econometric literature in the context of finite mixture models. Similar to factor models,

finite mixture models are typically not identified. Labels of the mixture components can be changed around. Thus, given symmetric priors, the posterior distribution has multiple modes. Identification can be achieved by imposing an ordering of the labels with respect to at least one of the parameters that are subject to label switching. However, if this identifying assumption is introduced by prior distributions, the choice of the constraint may have a significant impact on posterior estimates. Thus, while the identifying assumption does not change the maximum posterior value, it has consequences for the shape of the posterior. In addition, it is often observed in finite mixture models with ex ante identifying assumptions that posterior distributions show local modes. The occurrence of such local modes has severe consequences for the mixing behavior of the Gibbs sampler. In general, convergence cannot be guaranteed. To cure this problem, Frühwirth-Schnatter (2001, 2006) propose to skip identifying assumptions and instead enforce relabeling within the sampler. In the context of finite mixtures, identification is achieved by applying relabeling algorithms, see Celeux (1998) and Stephens (2000).

A factor model without directional identification can be interpreted as a continuous mixture. The mixing takes place via orthogonal transformations, hence we call the sampler *orthogonally mixing* and the corresponding output *orthogonally mixed*. We will use this output and adapt the remedy that is proposed for finite mixture models, i.e. introducing an ex-post identification strategy by likewise applying orthogonal transformations. Our approach uses the orthogonal Procrustes transformation by Schönemann (1966) to recover meaningful directed parameter estimates. The orthogonal Procrustes transformation can deal with label and sign switching, both of which result in a finite number of mixture components, but also with rotations, which result in an infinite number of mixture components. We hence extend the post-screening literature with respect to the considered model classes, developing an approach for factor models, and with respect to the type of mixing, developing an approach that goes beyond finite mixtures. Our approach does not require any constraints for the loadings matrix and is thus purely exploratory, with no need of theoretical reasoning before estimating the model. Inference results obtained under different orderings of the data are therefore equivalent under an orthogonal transformation.

To illustrate the properties of ex-post identification, we provide a simulation study within the framework of a static factor model comparing ex-post identification inference results with those from the constrained sampling approach by Geweke and Zhou (1996), which provides identification through a positive lower triangular loadings matrix. We check both samplers for their convergence properties, as well as statistical and numerical accuracy. Convergence is generally obtained faster for the ex-post identification scheme. Statistical accuracy is similar to that of the ex-ante identification

scheme for parameters invariant under orthogonal transformations and if the ex-ante scheme does not produce pathological posterior distributions. Such pathological cases do not occur under ex-post identification. The numerical accuracy of estimates is much higher for the ex-post identification scheme.

Eventually, we analyze a data set of ten equity indices, addressing the ordering problem. While the ordering of the data series has an effect on the parameter estimates for the ex-ante identification scheme, the parameter estimates from the ex-post identification scheme do not depend on the ordering of the series.

The paper proceeds as follows. Section 2 reviews the identification problem for factor models and its relation to orthogonal transformations and demonstrates why the typically used identification schemes do not succeed in preventing such orthogonal transformations. Section 3 introduces the concept of ex-post identification. Section 4 provides a simulation study that shows the improvements for the ex-post identification scheme. Section 5 provides an empirical application, and section 6 concludes.

2 Model setup and directional identification

Assume a factor model of the following form

$$Y = F\Lambda' + U, \tag{1}$$

where $Y = (y_1, \ldots, y_T)'$ is an $T \times N$ matrix of observable demeaned data, $F = (f_1, \ldots, f_T)'$ denotes a $T \times K$ matrix of K latent factors and $\Lambda = (\lambda_1, \ldots, \lambda_N)'$ represents the $N \times K$ matrix of loadings. $U = (u_1, \ldots, u_T)'$ is a $T \times N$ matrix of errors, where $u_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Sigma_u)$. The corresponding likelihood is given as

$$f(Y|F, \Lambda, \Sigma_u) = \prod_{t=1}^{T} |\Sigma_u|^{-.5} (2\pi)^{-.5N} \exp\{-.5(y_t - f_t\Lambda')'\Sigma_u^{-1}(y_t - f_t\Lambda')\}. \tag{2}$$

Introducing priors for Λ and Σ_u as

$$\pi(\Lambda) = \prod_{i=1}^{N} (2\pi)^{-.5K} \exp\{-.5\lambda_i' I_K^{-1} \lambda_i\}, \tag{3}$$

$$\pi(\Sigma_u) = 2^{-\frac{mN}{2}} \frac{1}{\Gamma_N(\frac{m}{2})} |S_0|^{\frac{m}{2}} |\Sigma_u|^{-\frac{m+N+1}{2}} \exp\left\{-\frac{1}{2}\text{tr}(S_0\Sigma_u^{-1})\right\}, \tag{4}$$

or, if $\Sigma_u = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2)$,

$$\pi(\Sigma_u) \quad = \quad \prod_{i=1}^{N} \frac{\alpha_{0i}^{\beta_{0i}}}{\Gamma(\alpha_{0i})} \sigma_i^{-2(\alpha_{0i}+1)} \exp\left\{-\frac{\beta_{0i}}{\sigma_{u_i}^2}\right\}, \tag{5}$$

and assuming that f_t is a priori normally distributed with mean zero and covariance I_K, i.e.

$$\pi(F) \quad = \quad \prod_{t=1}^{T} (2\pi)^{-.5K} \exp\{-.5 f_t I_K^{-1} f_t'\}, \tag{6}$$

allows to derive the marginalized likelihood as

$$f(Y|\Lambda, \Sigma_u) \quad = \quad \prod_{t=1}^{T} |(\Sigma_u + \Lambda\Lambda')|^{-.5} (2\pi)^{-.5N} \exp\{-.5 y_t' (\Sigma_u + \Lambda\Lambda')^{-1} y_t\}. \tag{7}$$

As the likelihood as well as the priors are invariant under orthogonal transformations, the posterior distribution under investigation requires identifying assumptions to allow for a meaningful Bayesian analysis. Orthogonal transformations include rotations, reflections and permutations of the K-dimensional factors f_t and loadings λ_i and are represented by matrices D having the property

$$DD' = D'D = I_K. \tag{8}$$

Without further identifying assumptions, iterative sampling from the full conditional distributions derived from the above setup, see Appendix A for details, would result in parameter trajectories subject to element-wise orthogonal transformations, thus allowing for no meaningful inference with respect to the directed parameters and variables, i.e. the factors and factor loadings.

To ensure identification, Geweke and Zhou (1996) propose to constrain the parameter space of the loadings to a positive lower triangular (PLT) matrix, i.e.

$$\Lambda = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \ldots & & & \lambda_{1N} \\ 0 & \lambda_{22} & \lambda_{23} & \ldots & & \lambda_{2N} \\ \vdots & & \ddots & \ddots & & \vdots \\ 0 & \ldots & 0 & \lambda_{KK} & \ldots & \lambda_{KN} \end{pmatrix}', \quad \text{with} \quad \lambda_{ii} > 0, i = 1, \ldots, K. \tag{9}$$

Imposing this PLT constraint on the loadings matrix guarantees that no point in the parameter space that satisfies the constraint can undergo an orthogonal transformation other than the identity

and still satisfy the constraint. The zero restrictions rule out rotation and label switching, and the positivity constraints rule out sign switching. Since the PLT constraint is conditional on the ordering of the data, there are $N!/(N-K)!$ different possibilities at hand to achieve identification. We denote an identification approach relying on a data ordering O_1 as $PLT|O_1$ identification scheme. Considering that for each loadings matrix Λ there exists a unique orthogonal matrix D_Λ such that ΛD_Λ satisfies constraints of the PLT form, see e.g. Muirhead (1982), the following transformation allows for mapping of the parameter space under one PLT identification scheme $PLT|O_1$ into that under a different one $PLT|O_2$, i.e.

$$PLT|O_1 \rightarrow PLT|O_2 = \{D_{\Lambda_{O_1}} : \Lambda_{O_1} D_{\Lambda_{O_1}} \in PLT|O_2\} \quad \text{for all} \quad \Lambda_{O_1} \in PLT|O_1 \tag{10}$$

The transformation in equation (10) involves an infinite number of orthogonal matrices D_Λ.[1] To obtain the same posterior distribution up to an orientation, however, the transformation should be done by a single orthogonal matrix for all points. Further, the position of the constraints of $PLT|O_2$ towards the posterior distribution obtained under $PLT|O_1$ determines which set of orthogonal matrices transforms how much probability mass.[2]

As pointed out by Lopes and West (2004) and Carvalho et al. (2008), inference results depend on the data ordering and thus on the selected PLT scheme. Consider reordering a sample initially following ordering O_1 in such a way that it satisfies ordering O_2 afterwards. Using the aforementioned transformation, it is now possible to map each point obtained from the Gibbs sampler under $PLT|O_1$ into a point that could have been obtained under $PLT|O_2$, i.e.

$$\left\{\tilde{\Lambda}_{O_2}^{(s)}\right\}_{s=1}^{S} \quad = \quad \left\{\tilde{\Lambda}_{O_1}^{(s)} D_{\Lambda_{O_1}}^{(s)}\right\}_{s=1}^{S}. \tag{11}$$

Note, however, that this mapping rule has a mixing effect. The resulting mixtures in such cases display features like multimodality and non-ellipticity complicating inference as noted e.g. by Ardia and Hoogerheide (2010). Thus identification via introduction of PLT constraints may have undesirable consequences for the shape of the posterior distribution, which causes numerical problems for the Gibbs sampler and complicates the interpretation and inference derived from the Gibbs output.

Appendix B discusses a numerical example for the behavior of the unconstrained Gibbs sampler

[1] All points satisfying both sets of constraints are transformed by the identity matrix, whereas all points with identical row vectors for the *founders* under the second set of constraints are transformed by the same orthogonal matrix.

[2] Moreover, for a comprehensive analysis of the effect of such transformations on the probability space, each of the possible $N!/(N-K)!$ orderings should be considered.

and the Gibbs sampler under PLT constraints. We simulate a data set for a model with $K = 2$ factors and $N = 10$ cross-sections, of which the first five are arranged in three different orderings, while the remaining five stay identical. The data are analyzed by means of the unconstrained Gibbs sampler and the PLT constrained Gibbs sampler. The first row of figure 1 shows the unconstrained output as Gibbs sequences and bivariate contour plots for cross section eight under the three orderings, the second row shows the according PLT constrained output. In both cases, 20,000 draws have been discarded as burn-in, another 20,000 have been kept. The unconstrained Gibbs sampler output is unfit for inference under all three orderings. The PLT constrained sampler, on the other hand, looks suitable for inference under the third ordering, but displays non-ellipticity under the second, and non-ellipticity and multimodality under the first ordering.

3 An ex-post identification scheme

In the following we propose an ex-post identification scheme for Bayesian analysis of factor models. Ex-post identification approaches are well known to the Bayesian literature, especially in the context of finite mixture models (Frühwirth-Schnatter, 2001, 2006). In this context they are used when label switching (Redner and Walker, 1984) occurs in the output of an unconstrained sampler. Richardson and Green (1997) advise to use different identifiability constraints when postprocessing the MCMC output. For finite mixtures Stephens (2000) and Frühwirth-Schnatter (2001) propose the use of relabeling algorithms that screen the output of the unconstrained sampler and sort the labels to minimize some divergence measures, e.g. Kullback-Leibler distances. The main idea behind the relabeling approach in finite mixtures is that the output of the unconstrained sampler in fact stems from a mixture distribution. The mixing is discrete and occurs via permutations of the labels. The relabeling algorithm reverses the transformation and thus the mixing.

The unconstrained Gibbs sampler for the factor model generates an *orthogonally mixed sample*, i.e. under the absence of parameter constraints, the obtained draws lack orientation and thus do not allow for meaningful inference via the calculation of arithmetic means. We suggest to apply an analogy to the finite mixture literature discussed above in order to solve the directional identification problem. However, since orthogonal mixing does not only involve relabeling but also reflections and rotations, existing relabeling algorithms do not apply here. Instead we propose the use of the orthogonal Procrustes (OP) transformation as devised by Schönemann (1966) to recover unique posterior estimates from the orthogonally mixed sample. The ex-post identification is achieved via post-screening the unconstrained Gibbs sampling output. To achieve identification, the following

criterion is considered

$$Q = \sum_{i=1}^{N} (\tilde{\lambda}_i(D) - \lambda_i^*)'(\tilde{\lambda}_i(D) - \lambda_i^*) = \text{tr}\left[(\tilde{\Lambda}(D) - \Lambda^*)(\tilde{\Lambda}(D) - \Lambda^*)'\right], \tag{12}$$

where $\Lambda^* = (\lambda_1^*, \ldots, \lambda_N^*)'$ denotes a chosen fixed point and $\tilde{\Lambda}(D) = (\tilde{\lambda}_1(D), \ldots, \tilde{\lambda}_N(D))'$ denotes a draw from the unconstrained distribution conditional on an unknown orthogonal transformation D. The relabeling is then concerned with determining a set of orthogonal matrices and a fixed point, i.e.

$$\{\{D^{(r)}\}_{r=1}^R, \Lambda^*\} = \arg\min \sum_{r=1}^{R} \text{tr}\left[(\tilde{\Lambda}^{(r)}(D^{(r)}) - \Lambda^*)(\tilde{\Lambda}^{(r)}(D^{(r)}) - \Lambda^*)'\right]. \tag{13}$$

The minimization is achieved iteratively via a two-step optimization, based on an initial choice of Λ^*. Note that the result is identified up to an orthogonal transformation, since the right side of equation (13) yields the same result if $\{D^{(r)}\}_{r=1}^R$ and Λ^* are all orthogonally transformed by the same matrix. Hence the initial choice of Λ^* effects the orientation of the eventual Λ^*. Apart from this effect on the orientation of the result, we find robustness of the algorithm with respect to the initial choice of Λ^*, as long as it stems from the orthogonally mixing posterior distribution. For convenience, we choose the last draw from the unconstrained sampler.

Step 1 Operationalizing $\tilde{\Lambda}^{(r)}(D^{(r)})$ as $\tilde{\Lambda}^{(r)}D^{(r)}$ yields the following minimization problem for $D^{(r)}$, i.e.

$$D^{(r)} = \arg\min \text{tr}\left[(\tilde{\Lambda}^{(r)}D^{(r)} - \Lambda^*)(\tilde{\Lambda}^{(r)}D^{(r)} - \Lambda^*)'\right] \quad \text{subject to} \quad D^{(r)\prime}D^{(r)} = I. \tag{14}$$

The solution of this orthogonal Procrustes (OP) problem is provided by Schönemann (1966). It involves the following calculations

1.1 Define $S_r = \tilde{\Lambda}^{(r)\prime}\Lambda^*$.

1.2 Do the eigenvalue decomposition $S_r'S_r = V_{1,r}M_rV_{1,r}'$.

1.3 Do the eigenvalue decomposition $S_rS_r' = V_{2,r}M_rV_{2,r}'$.

1.4 Find the unique reflection matrix R_r for which $((R_rV_{2,r})'S_rV_{1,r})_{ii} \geq 0$ for all $i = 1, \ldots, K$.

1.5 Obtain the orthogonal transformation matrix $D^{(r)} = (R_rV_{2,r})V_{1,r}'$.

1.6 Transform the observed matrix and obtain $\tilde{\Lambda}^{(r)} = \tilde{\Lambda}^{(r)}D^{(r)}$.

For further details on the derivation of this solution, see Appendix C. Note that if the dispersion

between the cross sections is rather large, the solutions may be improved via consideration of weights. Thus Step 1.1 above is altered into

1.1a Define $S_r = \tilde{\Lambda}^{(r)\prime} W \Lambda^*$,

where the weights matrix has to be diagonal. We initialize the weights as the inverses of the estimated lengths of the loadings vectors, i.e.

$$w_{ii} = R \left(\sum_{r=1}^{R} \sqrt{\tilde{\lambda}_i^{(r)\prime} \tilde{\lambda}_i^{(r)}} \right)^{-1}, \quad i = 1, \ldots, N. \tag{15}$$

Consecutively, we use as weights a function of the number of factors and the determinants of the estimated covariance matrices, which are a measure invariant to orthogonal transformations, i.e.

$$w_{ii} = \det \left(\frac{1}{R} \sum_{r=1}^{R} (\tilde{\lambda}_i^{(r)} - \lambda_i^*)(\tilde{\lambda}_i^{(r)} - \lambda_i^*)' \right)^{-\frac{1}{K}}, \quad i = 1, \ldots, N. \tag{16}$$

This procedure will be called *Weighted Orthogonal Procrustes* (WOP) transformation in the following.

Step 2 Choose Λ^* as

$$\lambda_i^* = \frac{1}{R} \sum_{r=1}^{R} \tilde{\lambda}_i^{(r)} D^{(r)}, \quad i = 1, \ldots, N. \tag{17}$$

Step 1 thus minimizes the (weighted) distance between the transformed observations and the current Λ^*. Hence the best choice for Λ^* would be the mean of the oriented sample we aim to restore. Of course, this mean is unobservable. Still, the mean of an approximation to the oriented sample, obtained in step 1, is a better choice for Λ^* than the initial arbitrary draw. Therefore we update Λ^* to this value. With a better approximation to the mean, conversely, we are able to obtain a better approximation to the oriented sample. For arbitrary initial choices of Λ^* taken from the orthogonally mixing sample, five iterations usually suffice to achieve convergence to a fixed point Λ^*, which is also the posterior mean of Λ. The sequence of factor matrices is accordingly transformed. To obtain a well-interpretable result, the user is free to apply an orthogonal transformation to the resulting ex-post identified posterior afterwards.

Note that for this procedure to work, the $\{\tilde{\Lambda}^{(r)} \tilde{\Lambda}^{(r)\prime}\}_{r=1}^{R}$ must have converged to a stationary

distribution and the sampler must be well-mixing for $\{\tilde{\Lambda}^{(r)}\tilde{\Lambda}^{(r)\prime}\}_{r=1}^{R}$ to ensure that the moments can be recovered properly. It does not matter, however, how the orthogonal transformation matrices are distributed, as long as they are all orthogonal. Mixing can be sped up by occasionally adding random orthogonal transformation steps. Appendix D gives a simulation experiment for the removal of orthogonal mixing via the WOP approach. We apply the WOP on the output of the unconstrained sampler shown in the first row of figure 1 and obtain the output shown in the third row. Note that the orientation of the posterior distributions under WOP is arbitrary and hence a common orientation has been identified, treating the posterior means obtained under all three orderings as an orthogonally mixed sample.

4 Simulation study

In this section, we perform a simulation study to assess the properties of our proposed WOP ex-post identification approach. As a benchmark we also apply the PLT ex-ante identification approach. We will analyze 27 different model setups, which all have some features in common. First, for every cross section, at least 20% of the variation is explained by the factors. Allowing for data series whose variation cannot be explained by any of the factors potentially complicates the estimation of the factors, see Boivin and Ng (2006). Second, the explanatory power of the *founders* of Carvalho et al. (2008), i.e. the loadings that are used to identify the model in the PLT approach, is exactly the average of all variables loading on the same factor. This implies that the identification constraints are not extraordinarily well, but also not extraordinarily badly chosen. Third, the parameters always satisfy the conditions of the PLT constraints, i.e. the loadings matrices used in the simulation are all positive lower triangular. We choose this property to allow for direct comparison of the results under PLT with the simulated parameters without further transformations. The WOP identification scheme, on the other hand, by definition will not provide a result that satisfies these constraints. To reach comparable results, the resulting final distribution under the WOP scheme will be transformed with respect to the simulated parameters. The transformation matrix is obtained by mapping the mean of the posterior distribution onto the simulated parameters via OP.[3] Afterwards all sampled points of the posterior distribution under the WOP scheme are transformed by the resulting transformation matrix. To get a fair comparison with the PLT scheme we do the same with the PLT samples, too, and refer to them as transformed PLT. Note that in exploratory

[3]Note that Doz et al. (2011) choose a similar approach to compare their factor estimates with the simulated factors, however, they use an OLS regression to determine the required transformation matrix, resulting in an approximately orthogonal matrix, while OP provides an exactly orthogonal matrix.

factor analysis, it is common to perform an orthogonal transformation of the estimates. Thus, the orthogonal transformation that maps the posterior distributions onto the simulated parameters is needed to reach sensible results if we like to analyze distributions of directional parameters like loadings or factors.

We consider data sets with short ($T = 30$), medium ($T = 60$), and long ($T = 150$) factors and with $N = 10$, $N = 40$, and $N = 100$ cross-sections. Each of these setups is estimated for models with $K = 2$, $K = 3$ and $K = 4$ factors. The parameters are summarized in $\Theta = (\Lambda, \Sigma_u, F)$ with Σ_u assumed diagonal. The factor loadings are simulated according to the aforementioned conditions. We investigate the corresponding 27 different scenarios. In all cases, the number of factors is assumed to be known. We use inverse Gamma priors for the variances with $\alpha_{0i} = \beta_{0i} = 1$ for all $i = 1, \ldots, N$.

The length of the burn-in may vary, depending on the model size and the sampler we use. We set the initial burn-in to 1,000 draws. During this period, the PLT approach is allowed to jointly flip around the signs of each pair of factors and loadings vectors if the full conditional density for one or multiple loadings vectors indicates significantly negative loadings where the constraints require positive ones. This will help the constrained sampler in the case of inconveniently chosen starting values. Beyond this initial burn-in phase of 1,000 draws, we will determine the required length of the burn-in, monitoring convergence via the statistics provided in Geweke (1991) for the invariant parameters. This procedure ensures that we actually obtain a sample from the posterior distribution.

For the post-screening via the WOP scheme, we perform five iterations to reach convergence. We find that PLT generally requires substantially more iterations to converge than the unconstrained sampler. Especially in setups with more than two factors, we often require a burn-in sequence three to five times longer than for WOP.

Due to scaling problems occurring in large N small T settings, the RMSE does not reflect the performance of the samplers accurately with respect to the factors and loadings. We thus consider a scaling-invariant measure, the correlations between the (OP transformed) mean of the posterior distributions and the simulated parameters. The correlation results are given in tables 1 and 2. Correlations between posterior means and true values under the WOP scheme are at least slightly higher than under the PLT scheme, if transformed PLT results are considered for factors and loadings. The improvement by the WOP scheme is more pronounced in models with more than two factors. The WOP scheme provides at least as good results as the PLT scheme for all model setups.

The RMSEs for the covariance parameters, which are invariant to orthogonal transformations and not affected by scaling issues, are given in table 3 and are almost identical for PLT and WOP. Since

directional identification is not required to estimate the covariance parameters, this comparison is in fact one between PLT and the unconstrained sampler.

Table 4 gives a measure of the correlation between the factors and loadings. We evaluate the estimates for the (undirected) systematic part of the data $F\Lambda'$, which can be obtained without directional identification, and the directed estimates of the factors F and loadings Λ. Out of these estimates, we calculate a divergence measure given as

$$\sqrt{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\frac{1}{R}\sum_{r=1}^{R}\left(\tilde{f}_t^{(r)}\tilde{\lambda}_i^{(r)}\right)-\left(\frac{1}{R}\sum_{r=1}^{R}\tilde{f}_t^{(r)}\right)\left(\frac{1}{R}\sum_{r=1}^{R}\tilde{\lambda}_i^{(r)}\right)\right)^2}. \tag{18}$$

We find that divergence between the undirected and directed estimates for the factors and loadings is substantially larger for PLT than for WOP, indicating that the posterior distributions of the factors and loadings are non-elliptical and possibly multimodal. Eventually, we look at the 85% coverage intervals, see e.g. Hoff (2009), of the systematic part of the data $F\Lambda'$ and the covariances and find that for both measures, the simulated values are in fact covered in 85% of the cases, indicating that identification has little impact on the variance decomposition.

In the next step, we assess the numerical properties of the posterior sampler by means of Monte Carlo errors. For parameters invariant to orthogonal transformations, differences between both approaches are not very pronounced (see tables 6 and 7). In some setups, the unconstrained sampler provides slightly better results. However, for parameters that depend on directional identification the numerical standard errors of the WOP approach are much smaller (see tables 8 and 9).

Hence, the ex-ante identification has almost no impact on the inference of parameters invariant to orthogonal transformations. Modest improvements can be obtained by skipping ex-ante identification and applying the unconstrained sampler. If inference on parameters that require directional identification is concerned, the WOP approach provides better results. Correlations between true and estimated factors and loadings are generally moderately, for some setups substantially higher. The numerical accuracy of the estimates can be improved by WOP compared to PLT.

One source for the poorer performance of the PLT approach lies in the shapes of the posterior distributions that are implied by the ex-ante constraints. As discussed before, PLT may generate non-elliptical posterior distributions or even posterior distributions with multiple modes. Such distributions are much harder to handle with the Gibbs sampler. Convergence of the Gibbs sampler can be very poor under such circumstances (Woodard, 2011). Such convergence problems can be eliminated by the use of the unconstrained sampler and the ex-post identification scheme WOP.

5 Empirical example

We analyze a data set of growth rates on 10 equity indices over a period from 1973Q3 until 2011Q3, obtained from DataStream®. The industries considered are oil and gas (OILGS), basic materials (BMATR), industrial goods (INDUS), consumer goods (CNSMG), health care (HLTHC), consumer services (CNSMS), telecommunications (TELCM), utilities (UTILS), financial services (FINAN) and technology (TECNO). Based on a preliminary analysis based on the information criteria of Bai and Ng (2002) we estimate factor models with two factors.

Recall that the PLT approach allows for $N!/(N-K)!$ different choices for the *founders* of the model. For a large number of cross-sections, possibly in combination with many factors, this quickly becomes infeasible, as has been noted by Carvalho (2006) and Carvalho et al. (2008). In our setup with ten series and two factors, there are 90 different orderings that we can compare to analyze the ordering effect on the constrained sampler. We start by putting the first series in the first position, alternating between the second to tenth series for the second position. Next, we put the second series in the first position and alternate between the remaining ones, and so forth. If the ordering of the data had no effect on the estimates, it should be possible to orthogonally transform the estimated factors and loadings vectors per cross-section obtained under different orderings into each other.

To get rid of the orthogonal variation in the estimates across the orderings, we treat both sets of the 90 posterior means of the loadings as orthogonally mixed samples. Afterwards, they are both transformed by OP such that their means satisfy the PLT constraints. The top two plots of figure 2 show the accordingly transformed factors for PLT for all 90 orderings of the data. Both factors look quite similar, with a moderate degree of dispersion. The bottom two plots, on the other hand, display a much smaller degree of variation across the 90 orderings of the data. The absence of identification constraints helps to obtain results that do not depend on the ordering of the series. Likewise, the variation in the estimated loadings over the different orderings, which is shown in table 10, is much larger for the PLT estimates than for the WOP estimates, despite the means being similar. Eventually, we will look at the idiosyncratic covariance parameters. The estimates are shown in table 11. There is no significant difference between any parameters for PLT and WOP, so the estimated idiosyncratic part of the model does not differ under both estimation schemes.

The empirical example underlines that both PLT and the unconstrained sampler provide similar results as long as directional invariant parameters are concerned. However, PLT has the disadvantage that results vary with the ordering of the data, while WOP provides results that do not depend on the ordering of the data.

6 Conclusion

In Bayesian estimation of factor models, constraints on the parameter space ensuring identifiability of factors and loadings can be troublesome. The PLT identification scheme results in posterior densities which may be non-elliptical or may even have multiple local modes. In any case, the shape of the posterior density depends on the ordering of the data. This characteristic of the PLT constrained posterior densities, reflected in the output of the constrained Gibbs sampler, make the use of an ex-post identification scheme attractive.

We suggest to refrain from ex-ante identification and instead use the output of the unconstrained sampler, which, we argue, stems from an orthogonally mixing distribution. It is possible to remove the mixing from the unconstrained Gibbs output, using the orthogonal Procrustes transformation devised by Schönemann (1966), which we extend by a weighting component. We find that the resulting Bayes estimates yield numerically superior and, in terms of statistical accuracy, competitive results, compared to the PLT approach. In those cases where using the constrained sampler results in non-well-behaved posterior densities, whose means may be no viable estimators, the WOP approach does not experience any problems. Results obtained under different WOP schemes can be orthogonally transformed into each other and into economically interpretable results. Note that the approach may also be suited for identification of dynamic factor models. However, a detailed discussion of this issue is beyond the scope of this paper.

Estimating a two-factor model for a data set of ten equity indices, we confirm that the PLT identification scheme is sensitive to ordering. Our ex-post identification approach based on the weighted orthogonal Procrustes, on the other hand, yields very similar results under all possible different orderings. Thus it seems recommendable to use the ex-post identification procedure to avoid problems with the posterior distributions and thus the estimates derived therefrom, obtaining identical parameter estimates, independent of the ordering of the data series.

Acknowledgements

For very helpful comments and thoughtful suggestions they provided on earlier versions of the paper, we thank Roman Liesenfeld, Uwe Jensen, Sylvia Frühwirth-Schnatter, and participants of the Statistische Woche (September 2012).

References

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

Ardia, D. and Hoogerheide, L. F. (2010). Bayesian Estimation of the GARCH(1,1) Model with Student-t Innovations. Tinbergen Institute Discussion Papers 10-045/4, Tinbergen Institute.

Bai, J. and Ng, S. (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica*, 70:191–221.

Bernanke, B., Boivin, J., and Eliasz, P. S. (2005). Measuring the Effects of Monetary Policy: A Factor-augmented Vector Autoregressive (FAVAR) Approach. *The Quarterly Journal of Economics*, 120(1):387–422.

Boivin, J. and Ng, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics*, 132(1):169–194.

Carvalho, C. M. (2006). Structure and sparsity in high-dimensional multivariate analysis. PhD Thesis, Department of Statistical Science, Duke University.

Carvalho, C. M., Chang, J., Lucas, J. E., Nevins, J. R., Wang, Q., and West, M. (2008). High-Dimensional Sparse Factor Modeling: Applications in Gene Expression Genomics. *Journal of the American Statistical Association*, 103(4):1438–1456.

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

Doz, C., Giannone, D., and Reichlin, L. (2011). A two-step estimator for large approximate dynamic factor models based on Kalman filtering. *Journal of Econometrics*, 164(1):188–205.

Frühwirth-Schnatter, S. (2001). Fully Bayesian Analysis of Switching Gaussian State Space Models. *Annals of the Institute of Statistical Mathematics*, 53(1):31–49.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer, New York.

Frühwirth-Schnatter, S. and Lopes, H. F. (2009). Parsimonious Bayesian Factor Analysis when the Number of Factors is Unknown. Technical report, University of Chicago Booth School of Business.

Geweke, J. (1991). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. Staff Report 148, Federal Reserve Bank of Minneapolis.

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

Hoff, P. D. (2009). *A First Course in Bayesian Statistical Methods*. Springer, New York.

Kose, M. A., Otrok, C., and Whiteman, C. H. (2003). International Business Cycles: World, Region, and Country-Specific Factors. *American Economic Review*, 93(4):1216–1239.

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

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

Otrok, C. and Whiteman, C. H. (1998). Bayesian Leading Indicators: Measuring and Predicting Economic Conditions in Iowa. *International Economic Review*, 39(4):997–1014.

Redner, R. A. and Walker, H. F. (1984). Mixture Densities, Maximum Likelihood and the EM Algorithm. *SIAM Review*, 26(2):195–239.

Richardson, S. and Green, P. J. (1997). On Bayesian Analysis of Mixtures With an Unknown Number of Components. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 59(4):731–792.

Rubin, D. and Thayer, D. (1982). EM algorithms for ML factor analysis. *Psychometrika*, 47(1):69–76.

Rubin, D. and Thayer, D. (1983). More on EM for ML factor analysis. *Psychometrika*, 48(2):253–257.

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

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

Woodard, D. B. (2011). Detecting Poor Convergence of Posterior Samplers due to Multimodality. mimeo.

Tables

			PLT				transformed PLT				WOP			
N	T	K	f_1	f_2	f_3	f_4	f_1	f_2	f_3	f_4	f_1	f_2	f_3	f_4
10	30	2	0.7919	−0.0852			0.9275	0.8597			0.9282	0.8600		
10	30	3	0.7063	0.5650	0.2690		0.8706	0.8438	0.6988		0.8834	0.8675	0.8246***	
10	30	4	0.7703	0.6956	0.2713	0.4045	0.8466	0.8066	0.6582	0.5815	0.8627	0.8122	0.7566***	0.6937***
10	60	2	0.9196	0.8552			0.9355	0.8820			0.9357	0.8825		
10	60	3	0.7844	0.7330	0.1747		0.9013	0.8883	0.8622		0.9015	0.8891	0.8598	
10	60	4	0.8488	0.7608	0.2141	0.4064	0.8732	0.8335	0.7120	0.6615	0.8789	0.8411	0.7625*	0.7712***
10	150	2	0.9397	0.8859			0.9443	0.8917			0.9443	0.8917		
10	150	3	0.8904	0.8669	0.6972		0.9108	0.8991	0.8813		0.9104	0.8985	0.8796	
10	150	4	0.8593	0.8197	0.3732	0.4661	0.8829	0.8656	0.7910	0.7891	0.8864	0.8645	0.7975	0.8125
40	30	2	0.7112	0.1386			0.9565	0.9600			0.9568	0.9610		
40	30	3	0.7250	0.3386	0.5826		0.9309	0.9232	0.9253		0.9585*	0.9513**	0.9627*	
40	30	4	0.6780	0.6614	0.6288	0.4179	0.8860	0.8944	0.9013	0.8622	0.9505***	0.9389***	0.9489***	0.9363***
40	60	2	0.8455	0.2508			0.9660	0.9619			0.9658	0.9625		
40	60	3	0.6891	0.5885	0.5689		0.9440	0.9478	0.9532		0.9582	0.9602	0.9644*	
40	60	4	0.7709	0.7865	0.7109	0.3867	0.9395	0.9375	0.9549	0.9316	0.9499*	0.9435	0.9644**	0.9472**
40	150	2	0.9635	0.9555			0.9695	0.9639			0.9695	0.9639		
40	150	3	0.8696	0.8511	0.8826		0.9601	0.9622	0.9662		0.9603	0.9625	0.9670	
40	150	4	0.9113	0.9121	0.9145	0.7460	0.9531	0.9488	0.9652	0.9537	0.9531	0.9487	0.9651	0.9537
100	30	2	0.8301	0.4782			0.9877	0.9854			0.9877	0.9854		
100	30	3	0.7392	0.5132	0.3420		0.9831	0.9760	0.9761		0.9863*	0.9793	0.9806	
100	30	4	0.8302	0.5740	0.3693	0.4085	0.9617	0.9511	0.9563	0.9418	0.9828***	0.9803*	0.9781***	0.9781**
100	60	2	0.6130	0.4441			0.9885	0.9871			0.9886	0.9871		
100	60	3	0.8102	0.6115	0.0839		0.9812	0.9801	0.9821		0.9853*	0.9834**	0.9842	
100	60	4	0.7743	0.4787	0.2797	0.4816	0.9754	0.9666	0.9694	0.9633	0.9840***	0.9820***	0.9822***	0.9801***
100	150	2	0.9751	0.9195			0.9887	0.9869			0.9887	0.9870		
100	150	3	0.9663	0.8699	0.8056		0.9863	0.9845	0.9846		0.9867	0.9850	0.9845	
100	150	4	0.9413	0.8800	0.5744	0.5490	0.9869	0.9850	0.9837	0.9796	0.9871	0.9852	0.9834	0.9801

Table 1: Average correlation between estimated and true factors. *,** and *** indicate significantly higher correlation for WOP at the 10%, 5% and 1% level for a mean-difference t-test.

			PLT				transformed PLT				WOP			
N	T	K	λ_1	λ_2	λ_3	λ_4	λ_1	λ_2	λ_3	λ_4	λ_1	λ_2	λ_3	λ_4
10	30	2	0.8412	−0.0585			0.9505	0.9125			0.9544	0.9213		
10	30	3	0.7753	0.5763	0.1846		0.9201	0.8550	0.7449		0.9409*	0.9121**	0.9104***	
10	30	4	0.8219	0.7093	0.2725	0.5834	0.9133	0.8446	0.7294	0.7932	0.9285	0.8770*	0.8604***	0.8800***
10	60	2	0.9607	0.9307			0.9743	0.9580			0.9744	0.9582		
10	60	3	0.8374	0.7627	0.1366		0.9685	0.9646	0.9439		0.9722	0.9679	0.9481	
10	60	4	0.9281	0.8373	0.2242	0.5693	0.9615	0.9305	0.8245	0.8605	0.9699	0.9425	0.9215***	0.9426***
10	150	2	0.9892	0.9755			0.9936	0.9832			0.9936	0.9834		
10	150	3	0.9710	0.9467	0.7665		0.9895	0.9852	0.9833		0.9896	0.9853	0.9827	
10	150	4	0.9613	0.9189	0.3988	0.6240	0.9851	0.9741	0.9513	0.9675	0.9877*	0.9776	0.9674*	0.9806*
40	30	2	0.7034	0.1485			0.9359	0.9381			0.9426	0.9444		
40	30	3	0.7122	0.3248	0.5670		0.9029	0.9021	0.9208		0.9405**	0.9397***	0.9579**	
40	30	4	0.6703	0.6492	0.6126	0.4307	0.8564	0.8581	0.8974	0.8499	0.9367***	0.9216***	0.9477***	0.9208***
40	60	2	0.8380	0.2469			0.9671	0.9639			0.9706	0.9707*		
40	60	3	0.6716	0.5821	0.5718		0.9413	0.9568	0.9651		0.9676	0.9693	0.9740*	
40	60	4	0.7565	0.7683	0.7407	0.4002	0.9469	0.9520	0.9651	0.9500	0.9638***	0.9642***	0.9740***	0.9629**
40	150	2	0.9810	0.9803			0.9887	0.9874			0.9887	0.9874		
40	150	3	0.8806	0.8587	0.9158		0.9846	0.9868	0.9889		0.9864	0.9875	0.9892	
40	150	4	0.9305	0.9231	0.9564	0.7830	0.9857	0.9841	0.9891	0.9848	0.9859	0.9844	0.9893	0.9849
100	30	2	0.7923	0.4507			0.9570	0.9482			0.9578	0.9487		
100	30	3	0.7190	0.5074	0.3162		0.9475	0.9403	0.9365		0.9525	0.9424	0.9415	
100	30	4	0.7828	0.5289	0.3409	0.4060	0.9229	0.9176	0.9057	0.8984	0.9467**	0.9415**	0.9330***	0.9317**
100	60	2	0.6166	0.4494			0.9762	0.9754			0.9778	0.9758		
100	60	3	0.8018	0.5895	0.0484		0.9714	0.9660	0.9662		0.9746	0.9688	0.9707	
100	60	4	0.7520	0.4683	0.2404	0.4691	0.9617	0.9458	0.9450	0.9480	0.9726***	0.9679***	0.9702***	0.9645***
100	150	2	0.9775	0.9255			0.9911	0.9894			0.9912	0.9894		
100	150	3	0.9659	0.8714	0.8012		0.9884	0.9878	0.9872		0.9884	0.9879	0.9872	
100	150	4	0.9370	0.8570	0.5478	0.5258	0.9883	0.9856	0.9853	0.9847	0.9890	0.9873**	0.9869**	0.9852

Table 2: Average correlation between estimated and true loadings (column) vectors. *,** and *** indicate significantly higher correlation for WOP at the 10%, 5% and 1% level for a mean-difference t-test.

			PLT					WOP				
N	T	K	min	lq	med	uq	max	min	lq	med	uq	max
10	30	2	0.0717	0.1044	0.1762	0.2094	0.2525	0.0716	0.1049	0.1759	0.1972	0.2488
10	30	3	0.1023	0.1459	0.1528	0.1749	0.2513	0.1024	0.1483	0.1530	0.1745	0.2547
10	30	4	0.0954	0.1399	0.1699	0.1994	0.2797	0.1001	0.1300	0.1617	0.2058	0.2753
10	60	2	0.0603	0.0892	0.1178	0.1438	0.1719	0.0607	0.0891	0.1162	0.1415	0.1737
10	60	3	0.0809	0.0929	0.1142	0.1344	0.1716	0.0824	0.0921	0.1146	0.1322	0.1928
10	60	4	0.0752	0.1145	0.1228	0.1645	0.2725	0.0760	0.1139	0.1217	0.1620	0.2699
10	150	2	0.0285	0.0458	0.0728	0.0893	0.1111	0.0286	0.0458	0.0727	0.0874	0.1115
10	150	3	0.0517	0.0596	0.0704	0.0795	0.1085	0.0528	0.0600	0.0710	0.0801	0.1257
10	150	4	0.0618	0.0783	0.0839	0.1088	0.1592	0.0628	0.0765	0.0891	0.1073	0.1535
40	30	2	0.0824	0.1423	0.1754	0.2071	0.2580	0.0832	0.1422	0.1750	0.2065	0.2575
40	30	3	0.0823	0.1175	0.1472	0.1837	0.2628	0.0821	0.1172	0.1481	0.1842	0.2636
40	30	4	0.0784	0.1107	0.1402	0.1738	0.2537	0.0780	0.1107	0.1392	0.1738	0.2397
40	60	2	0.0600	0.1050	0.1244	0.1361	0.1794	0.0604	0.1051	0.1231	0.1335	0.1786
40	60	3	0.0397	0.0736	0.1042	0.1204	0.1785	0.0399	0.0736	0.1024	0.1207	0.1772
40	60	4	0.0439	0.0763	0.0932	0.1172	0.1925	0.0441	0.0762	0.0933	0.1172	0.1642
40	150	2	0.0360	0.0635	0.0806	0.0901	0.1052	0.0361	0.0637	0.0805	0.0901	0.1054
40	150	3	0.0294	0.0456	0.0614	0.0731	0.1126	0.0292	0.0457	0.0606	0.0731	0.1008
40	150	4	0.0307	0.0431	0.0529	0.0654	0.1029	0.0306	0.0430	0.0531	0.0655	0.1035
100	30	2	0.0534	0.1232	0.1588	0.1889	0.2913	0.0567	0.1237	0.1595	0.1873	0.2925
100	30	3	0.0544	0.1088	0.1421	0.1773	0.2675	0.0547	0.1087	0.1420	0.1755	0.2684
100	30	4	0.0627	0.0994	0.1281	0.1569	0.2627	0.0629	0.0999	0.1284	0.1557	0.2629
100	60	2	0.0333	0.0878	0.1137	0.1351	0.2307	0.0357	0.0875	0.1131	0.1347	0.1891
100	60	3	0.0367	0.0794	0.0973	0.1179	0.2082	0.0371	0.0795	0.0970	0.1176	0.2085
100	60	4	0.0374	0.0654	0.0871	0.1076	0.1773	0.0373	0.0653	0.0870	0.1077	0.1781
100	150	2	0.0251	0.0516	0.0659	0.0829	0.1033	0.0255	0.0518	0.0659	0.0828	0.1033
100	150	3	0.0206	0.0483	0.0620	0.0765	0.1116	0.0206	0.0484	0.0621	0.0765	0.1113
100	150	4	0.0236	0.0391	0.0530	0.0657	0.1051	0.0237	0.0392	0.0529	0.0661	0.1053

Table 3: RMSEs for the error covariance parameters. Instead of reporting all N parameters per model, we only give the minimum (min), lower quartile (lq), median (med), upper quartile (uq) and maximum (max).

N	T	K	PLT		WOP	
10	30	2	1.7904	(2.0800)	0.3521	(0.0340)
10	30	3	6.7457	(1.7820)	0.6784	(0.1900)
10	30	4	6.8205	(1.7907)	1.0251	(0.1718)
10	60	2	0.8220	(0.7367)	0.2516	(0.0256)
10	60	3	3.9047	(3.1545)	0.4518	(0.1650)
10	60	4	6.8975	(2.0395)	0.8802	(0.2243)
10	150	2	0.3924	(0.1389)	0.1520	(0.0097)
10	150	3	1.8427	(0.8217)	0.2575	(0.0219)
10	150	4	7.1776	(3.0651)	0.6017	(0.2180)
40	30	2	3.9896	(5.6306)	0.4565	(0.0103)
40	30	3	11.5812	(4.7491)	0.7220	(0.0210)
40	30	4	16.1750	(3.7290)	1.0221	(0.0182)
40	60	2	3.0296	(5.7009)	0.4425	(0.0129)
40	60	3	9.9685	(6.8669)	0.6890	(0.0184)
40	60	4	11.9271	(7.5055)	0.9603	(0.0226)
40	150	2	0.5065	(0.2057)	0.2705	(0.0105)
40	150	3	4.8346	(5.3977)	0.4010	(0.0146)
40	150	4	4.5695	(1.5449)	0.5291	(0.0137)
100	30	2	4.9078	(5.1580)	0.3041	(0.0163)
100	30	3	11.2491	(6.6144)	0.4593	(0.0201)
100	30	4	19.4439	(5.7920)	0.6157	(0.0278)
100	60	2	7.0157	(11.7991)	0.4394	(0.0171)
100	60	3	10.1146	(8.8365)	0.6923	(0.0217)
100	60	4	21.4234	(10.3978)	0.9679	(0.0235)
100	150	2	1.2406	(0.5314)	0.4673	(0.0192)
100	150	3	4.9478	(5.0763)	0.7336	(0.0183)
100	150	4	12.2751	(5.5279)	1.0342	(0.0177)

Table 4: Divergence between the systematic part and the factors and loadings. Standard deviations in parentheses.

			PLT				WOP			
N	T	K	Σ_U		$F\Lambda'$		Σ_U		$F\Lambda'$	
10	30	2	0.8667	(0.1028)	0.8196	(0.0448)	0.8733	(0.0944)	0.8371	(0.0438)
10	30	3	0.8433	(0.1135)	0.8268	(0.0378)	0.8267	(0.1143)	0.8487	(0.0341)
10	30	4	0.8767	(0.1251)	0.8228	(0.0355)	0.8867	(0.1074)	0.8453	(0.0330)
10	60	2	0.8600	(0.1192)	0.8378	(0.0341)	0.8600	(0.1133)	0.8465	(0.0336)
10	60	3	0.8300	(0.1119)	0.8458	(0.0311)	0.8233	(0.1006)	0.8554	(0.0277)
10	60	4	0.8533	(0.1196)	0.8348	(0.0274)	0.8533	(0.1196)	0.8466	(0.0245)
10	150	2	0.8500	(0.1253)	0.8431	(0.0263)	0.8533	(0.1279)	0.8477	(0.0252)
10	150	3	0.8767	(0.1165)	0.8503	(0.0175)	0.8633	(0.1159)	0.8557	(0.0177)
10	150	4	0.8100	(0.1213)	0.8351	(0.0223)	0.8333	(0.1241)	0.8441	(0.0230)
40	30	2	0.8625	(0.0424)	0.8295	(0.0295)	0.8608	(0.0419)	0.8369	(0.0293)
40	30	3	0.8417	(0.0585)	0.8466	(0.0284)	0.8383	(0.0556)	0.8539	(0.0278)
40	30	4	0.8633	(0.0579)	0.8474	(0.0210)	0.8625	(0.0544)	0.8586	(0.0191)
40	60	2	0.8617	(0.0512)	0.8406	(0.0227)	0.8667	(0.0527)	0.8483	(0.0233)
40	60	3	0.8533	(0.0439)	0.8384	(0.0187)	0.8567	(0.0435)	0.8466	(0.0177)
40	60	4	0.8458	(0.0542)	0.8409	(0.0174)	0.8433	(0.0537)	0.8486	(0.0162)
40	150	2	0.8750	(0.0549)	0.8500	(0.0152)	0.8667	(0.0558)	0.8520	(0.0148)
40	150	3	0.8533	(0.0495)	0.8382	(0.0133)	0.8525	(0.0480)	0.8430	(0.0126)
40	150	4	0.8700	(0.0407)	0.8477	(0.0129)	0.8692	(0.0424)	0.8516	(0.0128)
100	30	2	0.8637	(0.0346)	0.8503	(0.0240)	0.8643	(0.0346)	0.8534	(0.0246)
100	30	3	0.8727	(0.0333)	0.8515	(0.0174)	0.8740	(0.0323)	0.8553	(0.0176)
100	30	4	0.8680	(0.0287)	0.8515	(0.0133)	0.8697	(0.0294)	0.8553	(0.0132)
100	60	2	0.8523	(0.0316)	0.8483	(0.0179)	0.8540	(0.0306)	0.8529	(0.0176)
100	60	3	0.8550	(0.0264)	0.8486	(0.0163)	0.8570	(0.0272)	0.8523	(0.0157)
100	60	4	0.8563	(0.0255)	0.8488	(0.0180)	0.8577	(0.0245)	0.8537	(0.0183)
100	150	2	0.8623	(0.0360)	0.8469	(0.0150)	0.8640	(0.0335)	0.8483	(0.0149)
100	150	3	0.8493	(0.0362)	0.8477	(0.0119)	0.8517	(0.0353)	0.8493	(0.0119)
100	150	4	0.8543	(0.0380)	0.8510	(0.0133)	0.8523	(0.0381)	0.8533	(0.0129)

Table 5: Share of parameters contained in the 85% coverage interval. Standard deviations in parentheses.

			PLT					WOP				
N	T	K	min	lq	med	uq	max	min	lq	med	uq	max
10	30	2	0.0019	0.0023	0.0036	0.0101	0.0369	0.0013	0.0019	0.0021	0.0039	0.0056
10	30	3	0.0030	0.0034	0.0042	0.0072	0.0155	0.0030	0.0032	0.0036	0.0043	0.0051
10	30	4	0.0025	0.0034	0.0054	0.0074	0.0158	0.0025	0.0034	0.0048	0.0058	0.0100
10	60	2	0.0010	0.0011	0.0016	0.0021	0.0024	0.0010	0.0011	0.0014	0.0019	0.0028
10	60	3	0.0016	0.0021	0.0033	0.0048	0.0223	0.0014	0.0018	0.0023	0.0031	0.0050
10	60	4	0.0014	0.0027	0.0049	0.0082	0.0235	0.0014	0.0025	0.0036	0.0048	0.0098
10	150	2	0.0007	0.0008	0.0010	0.0015	0.0041	0.0007	0.0008	0.0011	0.0014	0.0041
10	150	3	0.0013	0.0013	0.0016	0.0028	0.0061	0.0010	0.0010	0.0018	0.0030	0.0050
10	150	4	0.0013	0.0022	0.0045	0.0102	0.0142	0.0014	0.0022	0.0027	0.0073	0.0105
40	30	2	0.0010	0.0020	0.0026	0.0035	0.0712	0.0009	0.0016	0.0022	0.0024	0.0034
40	30	3	0.0009	0.0017	0.0020	0.0030	0.0536	0.0009	0.0015	0.0019	0.0024	0.0035
40	30	4	0.0009	0.0017	0.0021	0.0029	0.0198	0.0009	0.0016	0.0020	0.0025	0.0046
40	60	2	0.0010	0.0014	0.0017	0.0023	0.0426	0.0008	0.0012	0.0014	0.0018	0.0020
40	60	3	0.0007	0.0011	0.0015	0.0018	0.0383	0.0007	0.0010	0.0013	0.0015	0.0022
40	60	4	0.0009	0.0012	0.0014	0.0018	0.0130	0.0006	0.0011	0.0013	0.0016	0.0021
40	150	2	0.0006	0.0007	0.0008	0.0010	0.0013	0.0005	0.0007	0.0009	0.0010	0.0011
40	150	3	0.0004	0.0006	0.0007	0.0009	0.0017	0.0004	0.0006	0.0007	0.0009	0.0012
40	150	4	0.0005	0.0006	0.0008	0.0009	0.0021	0.0004	0.0006	0.0007	0.0009	0.0012
100	30	2	0.0007	0.0015	0.0019	0.0024	0.0331	0.0007	0.0014	0.0018	0.0022	0.0040
100	30	3	0.0009	0.0016	0.0020	0.0025	0.0252	0.0007	0.0014	0.0017	0.0022	0.0039
100	30	4	0.0005	0.0013	0.0018	0.0021	0.0167	0.0006	0.0012	0.0016	0.0021	0.0035
100	60	2	0.0006	0.0010	0.0013	0.0016	0.1129	0.0004	0.0009	0.0011	0.0014	0.0021
100	60	3	0.0005	0.0009	0.0011	0.0015	0.0485	0.0004	0.0009	0.0011	0.0013	0.0022
100	60	4	0.0004	0.0008	0.0010	0.0013	0.0573	0.0003	0.0007	0.0009	0.0013	0.0019
100	150	2	0.0003	0.0006	0.0007	0.0009	0.0016	0.0003	0.0005	0.0007	0.0009	0.0012
100	150	3	0.0003	0.0005	0.0007	0.0008	0.0019	0.0003	0.0005	0.0006	0.0008	0.0013
100	150	4	0.0002	0.0005	0.0006	0.0007	0.0059	0.0002	0.0005	0.0006	0.0007	0.0011

Table 6: Numerical standard errors for the idiosyncratic variances. Instead of reporting all N parameters per model, we only give the minimum (min), lower quartile (lq), median (med), upper quartile (uq) and maximum (max).

			PLT					WOP				
N	T	K	min	lq	med	uq	max	min	lq	med	uq	max
10	30	2	0.0023	0.0069	0.0103	0.0184	0.1811	0.0017	0.0030	0.0037	0.0048	0.0115
10	30	3	0.0033	0.0070	0.0101	0.0161	0.1201	0.0021	0.0044	0.0052	0.0063	0.0112
10	30	4	0.0035	0.0062	0.0081	0.0121	0.0399	0.0024	0.0050	0.0060	0.0077	0.0184
10	60	2	0.0014	0.0027	0.0034	0.0044	0.0206	0.0013	0.0026	0.0031	0.0038	0.0066
10	60	3	0.0021	0.0044	0.0060	0.0098	0.0721	0.0021	0.0036	0.0042	0.0051	0.0138
10	60	4	0.0026	0.0056	0.0088	0.0158	0.0648	0.0024	0.0042	0.0051	0.0067	0.0230
10	150	2	0.0009	0.0024	0.0028	0.0038	0.0153	0.0011	0.0024	0.0029	0.0035	0.0128
10	150	3	0.0020	0.0035	0.0043	0.0058	0.0229	0.0016	0.0032	0.0039	0.0047	0.0143
10	150	4	0.0022	0.0048	0.0073	0.0116	0.0432	0.0023	0.0040	0.0050	0.0072	0.0368
40	30	2	0.0011	0.0036	0.0050	0.0075	0.2929	0.0006	0.0021	0.0026	0.0032	0.0067
40	30	3	0.0014	0.0032	0.0041	0.0057	0.2192	0.0013	0.0025	0.0029	0.0035	0.0074
40	30	4	0.0019	0.0036	0.0044	0.0056	0.0954	0.0015	0.0030	0.0036	0.0044	0.0094
40	60	2	0.0012	0.0027	0.0037	0.0061	0.3544	0.0009	0.0019	0.0023	0.0026	0.0050
40	60	3	0.0013	0.0031	0.0040	0.0057	0.2187	0.0011	0.0021	0.0024	0.0029	0.0054
40	60	4	0.0017	0.0031	0.0038	0.0049	0.0961	0.0011	0.0025	0.0029	0.0033	0.0054
40	150	2	0.0008	0.0016	0.0019	0.0022	0.0326	0.0007	0.0015	0.0018	0.0021	0.0037
40	150	3	0.0009	0.0019	0.0022	0.0026	0.0580	0.0009	0.0017	0.0020	0.0023	0.0040
40	150	4	0.0011	0.0023	0.0026	0.0031	0.0691	0.0012	0.0021	0.0023	0.0026	0.0046
100	30	2	0.0006	0.0021	0.0027	0.0034	0.2647	0.0005	0.0016	0.0021	0.0027	0.0052
100	30	3	0.0007	0.0027	0.0034	0.0042	0.1325	0.0007	0.0021	0.0025	0.0031	0.0058
100	30	4	0.0011	0.0025	0.0031	0.0039	0.1838	0.0009	0.0022	0.0026	0.0033	0.0064
100	60	2	0.0004	0.0017	0.0023	0.0032	0.5552	0.0003	0.0012	0.0015	0.0019	0.0048
100	60	3	0.0010	0.0022	0.0028	0.0037	0.3094	0.0007	0.0015	0.0018	0.0022	0.0049
100	60	4	0.0008	0.0021	0.0025	0.0032	0.3579	0.0007	0.0017	0.0020	0.0024	0.0052
100	150	2	0.0004	0.0011	0.0013	0.0015	0.0565	0.0004	0.0011	0.0013	0.0015	0.0032
100	150	3	0.0005	0.0013	0.0016	0.0019	0.0805	0.0004	0.0012	0.0014	0.0017	0.0031
100	150	4	0.0007	0.0015	0.0018	0.0021	0.1221	0.0005	0.0014	0.0016	0.0019	0.0035

Table 7: Numerical standard errors for the product of factors and loadings. Instead of reporting all TN parameters per model, we only give the minimum (min), lower quartile (lq), median (med), upper quartile (uq) and maximum (max).

			PLT					WOP				
N	T	K	min	lq	med	uq	max	min	lq	med	uq	max
10	30	2	0.0061	0.0413	0.0917	0.1631	0.4039	0.0016	0.0031	0.0038	0.0049	0.0124
10	30	3	0.0091	0.0402	0.0620	0.1031	0.3038	0.0021	0.0046	0.0055	0.0066	0.0121
10	30	4	0.0108	0.0250	0.0326	0.0436	0.0794	0.0033	0.0054	0.0067	0.0090	0.0230
10	60	2	0.0011	0.0028	0.0035	0.0044	0.0171	0.0012	0.0026	0.0031	0.0039	0.0067
10	60	3	0.0068	0.0443	0.0764	0.1323	0.3798	0.0021	0.0036	0.0043	0.0053	0.0138
10	60	4	0.0103	0.0398	0.0653	0.0959	0.2798	0.0024	0.0044	0.0054	0.0072	0.0245
10	150	2	0.0010	0.0025	0.0031	0.0040	0.0166	0.0011	0.0024	0.0029	0.0035	0.0130
10	150	3	0.0024	0.0048	0.0062	0.0086	0.0229	0.0016	0.0032	0.0039	0.0047	0.0147
10	150	4	0.0033	0.0352	0.0698	0.1120	0.4182	0.0019	0.0039	0.0051	0.0076	0.0401
40	30	2	0.0010	0.0045	0.0072	0.0120	0.3279	0.0008	0.0021	0.0027	0.0033	0.0073
40	30	3	0.0184	0.0745	0.1085	0.1469	0.4012	0.0014	0.0026	0.0032	0.0038	0.0081
40	30	4	0.0123	0.0413	0.0594	0.0836	0.2189	0.0014	0.0032	0.0039	0.0048	0.0099
40	60	2	0.0012	0.0028	0.0039	0.0064	0.3636	0.0008	0.0020	0.0023	0.0027	0.0050
40	60	3	0.0046	0.0361	0.0622	0.1041	0.4426	0.0010	0.0021	0.0025	0.0030	0.0059
40	60	4	0.0025	0.0635	0.1069	0.1687	0.5094	0.0008	0.0026	0.0031	0.0035	0.0057
40	150	2	0.0008	0.0017	0.0022	0.0028	0.0232	0.0007	0.0015	0.0018	0.0021	0.0037
40	150	3	0.0009	0.0037	0.0054	0.0079	0.0443	0.0009	0.0017	0.0020	0.0023	0.0041
40	150	4	0.0016	0.0038	0.0050	0.0066	0.0894	0.0011	0.0021	0.0023	0.0026	0.0046
100	30	2	0.0013	0.0269	0.0612	0.1115	0.3320	0.0005	0.0016	0.0021	0.0027	0.0053
100	30	3	0.0010	0.0241	0.0495	0.0932	0.4309	0.0004	0.0021	0.0025	0.0031	0.0060
100	30	4	0.0067	0.0568	0.0861	0.1411	0.5027	0.0008	0.0022	0.0027	0.0034	0.0071
100	60	2	0.0005	0.0430	0.0931	0.1727	0.6783	0.0003	0.0012	0.0016	0.0020	0.0049
100	60	3	0.0058	0.0560	0.0905	0.1392	0.5380	0.0007	0.0015	0.0019	0.0023	0.0051
100	60	4	0.0116	0.0633	0.0956	0.1441	0.5691	0.0006	0.0018	0.0021	0.0026	0.0054
100	150	2	0.0004	0.0022	0.0040	0.0066	0.0450	0.0004	0.0011	0.0013	0.0015	0.0031
100	150	3	0.0009	0.0114	0.0243	0.0454	0.1913	0.0004	0.0013	0.0015	0.0017	0.0032
100	150	4	0.0035	0.0342	0.0509	0.0776	0.2902	0.0005	0.0015	0.0017	0.0020	0.0035

Table 8: Numerical standard errors for the product of factors and loadings (mean of F times mean of Λ). Instead of reporting all TN parameters per model, we only give the minimum (min), lower quartile (lq), median (med), upper quartile (uq) and maximum (max).

			PLT					WOP				
N	T	K	min	lq	med	uq	max	min	lq	med	uq	max
10	30	2	0.0418	0.0687	0.1266	0.1670	0.2508	0.0023	0.0036	0.0053	0.0094	0.0110
10	30	3	0.0254	0.0413	0.0673	0.1284	0.1424	0.0035	0.0067	0.0099	0.0119	0.0151
10	30	4	0.0138	0.0186	0.0231	0.0295	0.0553	0.0060	0.0084	0.0094	0.0112	0.0161
10	60	2	0.0013	0.0024	0.0042	0.0054	0.0088	0.0014	0.0030	0.0034	0.0061	0.0097
10	60	3	0.0335	0.0888	0.1022	0.1304	0.1694	0.0029	0.0054	0.0058	0.0069	0.0089
10	60	4	0.0151	0.0264	0.0552	0.0753	0.1169	0.0044	0.0057	0.0070	0.0095	0.0112
10	150	2	0.0008	0.0018	0.0028	0.0042	0.0047	0.0006	0.0020	0.0029	0.0041	0.0050
10	150	3	0.0014	0.0036	0.0059	0.0075	0.0096	0.0012	0.0028	0.0034	0.0035	0.0072
10	150	4	0.0019	0.0334	0.0632	0.0897	0.1018	0.0017	0.0031	0.0041	0.0068	0.0120
40	30	2	0.0018	0.0098	0.0130	0.0208	0.1463	0.0017	0.0068	0.0090	0.0144	0.0262
40	30	3	0.0452	0.1444	0.2029	0.2894	0.4837	0.0036	0.0093	0.0152	0.0190	0.0318
40	30	4	0.0259	0.0780	0.1054	0.1285	0.1870	0.0062	0.0119	0.0177	0.0208	0.0362
40	60	2	0.0030	0.0053	0.0080	0.0102	0.0667	0.0019	0.0044	0.0064	0.0085	0.0135
40	60	3	0.0414	0.0744	0.1002	0.1291	0.2485	0.0029	0.0056	0.0073	0.0106	0.0186
40	60	4	0.0085	0.0923	0.1439	0.2369	0.3367	0.0016	0.0080	0.0108	0.0131	0.0227
40	150	2	0.0015	0.0023	0.0034	0.0051	0.0103	0.0012	0.0020	0.0027	0.0041	0.0081
40	150	3	0.0019	0.0047	0.0063	0.0088	0.0139	0.0014	0.0025	0.0039	0.0054	0.0088
40	150	4	0.0025	0.0043	0.0053	0.0069	0.0101	0.0023	0.0033	0.0043	0.0056	0.0076
100	30	2	0.0231	0.1106	0.1966	0.2828	0.6475	0.0021	0.0062	0.0097	0.0132	0.0247
100	30	3	0.0064	0.1216	0.2091	0.3205	0.7269	0.0021	0.0098	0.0124	0.0180	0.0284
100	30	4	0.0226	0.1918	0.3061	0.4732	0.8942	0.0047	0.0152	0.0202	0.0253	0.0401
100	60	2	0.0320	0.1918	0.2821	0.3603	0.6203	0.0016	0.0067	0.0091	0.0118	0.0231
100	60	3	0.0526	0.1915	0.2300	0.3319	0.6053	0.0034	0.0101	0.0148	0.0192	0.0370
100	60	4	0.0353	0.1547	0.2773	0.3667	0.5878	0.0051	0.0162	0.0213	0.0275	0.0423
100	150	2	0.0022	0.0055	0.0081	0.0109	0.0282	0.0021	0.0046	0.0068	0.0094	0.0156
100	150	3	0.0056	0.0196	0.0405	0.0655	0.1084	0.0021	0.0063	0.0084	0.0111	0.0153
100	150	4	0.0105	0.0411	0.0656	0.0959	0.1667	0.0024	0.0080	0.0100	0.0133	0.0198

Table 9: Numerical standard errors for the sum of the squared loadings. Instead of reporting all N parameters per model, we only give the minimum (min), lower quartile (lq), median (med), upper quartile (uq) and maximum (max).

	PLT		WOP	
	λ_1	λ_2	λ_1	λ_2
OILGS	0.7797	0.0000	0.7814	0.0000
	(0.0946)	(0.0886)	(0.0350)	(0.0315)
BMATR	0.6899	0.5076	0.6965	0.5151
	(0.0246)	(0.0201)	(0.0147)	(0.0133)
INDUS	0.7406	0.5735	0.7525	0.5777
	(0.0233)	(0.0196)	(0.0130)	(0.0117)
CNSMG	0.5473	0.6572	0.5532	0.6679
	(0.0197)	(0.0212)	(0.0056)	(0.0060)
HLTHC	0.4571	0.6603	0.4662	0.6665
	(0.0289)	(0.0293)	(0.0157)	(0.0146)
CNSMS	0.5506	0.8030	0.5580	0.8135
	(0.0326)	(0.0358)	(0.0068)	(0.0081)
TELCM	0.3868	0.5411	0.4025	0.5377
	(0.0251)	(0.0244)	(0.0190)	(0.0171)
UTILS	0.5777	0.3082	0.6044	0.2899
	(0.0469)	(0.0498)	(0.0242)	(0.0209)
FINAN	0.6155	0.6145	0.6332	0.6129
	(0.0157)	(0.0178)	(0.0133)	(0.0116)
TECNO	0.5219	0.5490	0.5291	0.5566
	(0.0232)	(0.0233)	(0.0156)	(0.0145)

Table 10: Estimated factor loadings parameters for equity indices. Standard deviations over the 90 different orderings in parentheses. Results have been rotated such that their mean takes the positive lower triangular shape.

	PLT	WOP
OILGS	0.3577	0.3929
	(0.0439)	(0.0382)
BMATR	0.2710	0.2739
	(0.0045)	(0.0036)
INDUS	0.1280	0.1284
	(0.0042)	(0.0034)
CNSMG	0.2747	0.2754
	(0.0012)	(0.0009)
HLTHC	0.3602	0.3573
	(0.0038)	(0.0033)
CNSMS	0.0561	0.0605
	(0.0044)	(0.0044)
TELCM	0.5643	0.5616
	(0.0048)	(0.0045)
UTILS	0.5603	0.5470
	(0.0294)	(0.0268)
FINAN	0.2497	0.2431
	(0.0080)	(0.0082)
TECNO	0.4293	0.4260
	(0.0043)	(0.0041)

Table 11: Estimated idiosyncratic variance parameters for equity indices. Standard deviations over the 90 different orderings in parentheses.

Figures

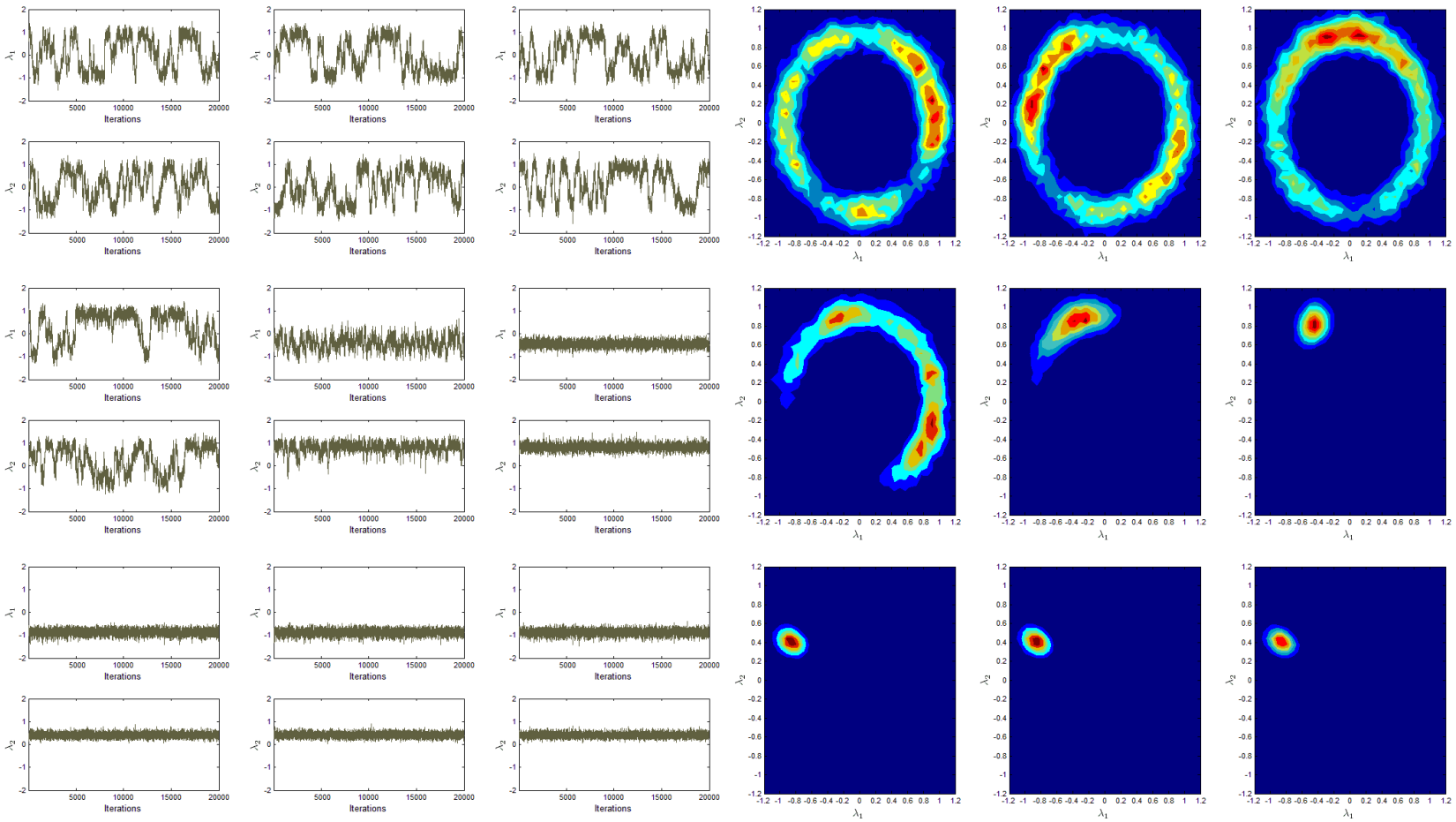

Figure 1: Gibbs sequences and contour plots for bivariate posterior distributions from the unconstrained sampler (top), the PLT constrained sampler (middle), and the unconstrained sampler after five WOP iterations and orientation (bottom). All plots show loadings on cross-section eight, for the three different orderings. Sample size is 20,000 draws, with the preceding 20,000 draws discarded as burn-in.

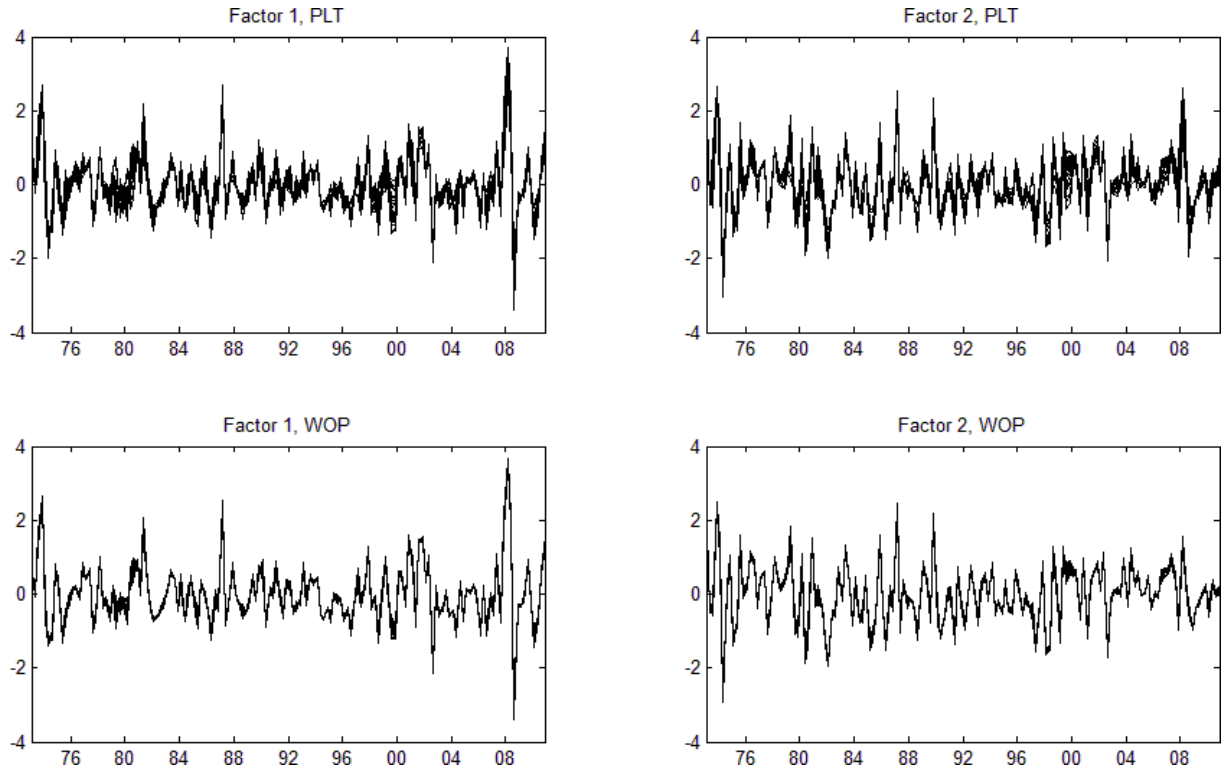

Figure 2: Estimated factors from equity indices. Factors in the first row are PLT results, factors in the second row are WOP results.

A The unconstrained Gibbs sampler

For the model described in Equation (1) and prior distributions given in Equations (3) to (6) the unconstrained sampling is achieved via iterative sampling from the following full conditional distributions, given as

$$f(F|\Lambda,\Sigma_u,Y) = \prod_{t=1}^{T}(2\pi)^{-\frac{k}{2}}|\Omega_{f_t}|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(f_t-\mu_{f_t})'\Omega_{f_t}^{-1}(f_t-\mu_{f_t})\right\}, \quad (19)$$

where $\Omega_{f_t} = (\Lambda'\Sigma_u^{-1}\Lambda + I)^{-1}$ and $\mu_{f_t} = \Omega_{f_t}(\Lambda'\Sigma_u^{-1}y_t)$. Throughout this paper, we will assume diagonality for Σ_u resulting in

$$f(\Sigma_u|F,\Lambda,Y) = \prod_{i=1}^{N}\frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)}\left(\frac{1}{\sigma_i^2}\right)^{\alpha_i-1}\exp\{-\frac{1}{\sigma_i^2}\beta_i\}, \quad (20)$$

where $\alpha_i = \frac{1}{2}T + \alpha_{0j}$ and $\beta_i = \frac{1}{2} + \sum_{t=1}^{T}(y_{it} - \lambda_i' f_t)^2 + \beta_{0i}$ for all $i = 1,\ldots,N$ and $\alpha_{0i} = \beta_{0i} = 1$ for all $i = 1\ldots,N$. The form of the full conditional distribution for Λ depends on the assumptions concerning Σ_u. Since $\Sigma_u = \mathrm{diag}(\sigma_1^2,\ldots,\sigma_n^2)$, then the full conditional distribution can be factorized along dimension N yielding

$$f(\Lambda|F,Y,\Sigma_u) = \prod_{i=1}^{N}(2\pi)^{-\frac{K}{2}}|\Omega_{\lambda_i}|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(\lambda_i-\mu_{\lambda_i})'\Omega_{\_i}^{-1}(\lambda_i-\mu_{\lambda_i})\right\}, \quad (21)$$

where $\Omega_{\lambda_i} = (\frac{1}{\sigma_i^2}f_t f_t' + I)^{-1}$ and $\mu_{\lambda_i} = \Omega_{\lambda_i}(\frac{1}{\sigma_i^2}\sum_{t=1}^{T}y_{it}f_t)$. In case Σ_u is not diagonal, factorization along the N dimension does not work.

B Numerical illustration for non-ellipticity and multimodality

For illustration purposes, we will consider a data set Y with $N = 10$ cross-sections and time dimension $T = 60$, which is driven by $K = 2$ static factors. This data set is simulated using as parameters

$$\Lambda = \begin{pmatrix} 0.100 & -0.200 & 0.500 & 0.600 & 0.100 & 0.174 & -0.153 & -0.470 & 0.186 & -0.577 \\ 0.000 & 0.200 & -0.100 & 0.400 & -0.900 & 0.429 & -0.392 & 0.652 & 0.282 & -0.541 \end{pmatrix}' \quad (22)$$

and

$$\Sigma_U \quad = \quad \text{diag}(0.990, 0.920, 0.740, 0.480, 0.180, 0.786, 0.823, 0.354, 0.886, 0.374). \tag{23}$$

We will consider three different orderings of the data. The first is the above ordering, denoted $Y|O_1$, the second orders the series as 2,3,1,4,5,6,7,8,9,10, denoted $Y|O_2$, and the third uses the ordering 5,2,1,3,4,6,7,8,9,10, denoted $Y|O_3$.

If normal priors are assumed for the loadings and the factors, the posterior distribution has a continuum of maxima that are all orthogonal transformations of each other. The first row of figure 1 shows the Gibbs output from an unconstrained sampler as it is denoted in Appendix A for the loadings of cross-section eight for $Y|O_1$, $Y|O_2$ and $Y|O_3$. Since orthogonal transformations between the Gibbs sweeps are not ruled out, the sampler moves along the circle in all three cases.

The following example will illustrate the reasons for non-ellipticity and multimodality, using the three orderings of the simulated data set. Consider the first ordering of the data $Y|O_1$, such that

$$Y|O_1 \quad = \quad F|O_1\Lambda'|O_1 + U|O_1 \qquad \text{with } U|O_1 = (u_1|O_1, \ldots, u_T|O_1)' \text{ and } u_t|O_1 \ \sim \ \mathcal{N}(0, \Sigma_u|O_1). \tag{24}$$

We can use a permutation matrix $P_{1,2}$ to rearrange the cross-sections and obtain the second ordering of the data $Y|O_2$, i.e.

$$Y|O_2 \quad = \quad Y|O_1P_{1,2} \qquad \text{with} \quad P_{1,2} = \begin{pmatrix} 0 & 0 & 1 & 0_{(1\times 7)} \\ 1 & 0 & 0 & 0_{(1\times 7)} \\ 0 & 1 & 0 & 0_{(1\times 7)} \\ 0_{(7\times 1)} & 0_{(7\times 1)} & 0_{(7\times 1)} & I_7 \end{pmatrix}. \tag{25}$$

Consequently, we obtain

$$\begin{aligned} Y|O_2 \quad &= \quad F|O_2\Lambda'|O_2 + U|O_2 \\ &= \quad F|O_1D_{P'_{1,2}\Lambda|O_1}D'_{P'_{1,2}\Lambda|O_1}\Lambda'|O_1P_{1,2} + U|O_1P_{1,2}, \end{aligned} \tag{26}$$

where $D_{P'_{1,2}\Lambda|O_1}$ is the orthogonal matrix that maps $P'_{1,2}\Lambda|O_1$ in such a way that it satisfies the PLT

constraints. This implies that

$$\Lambda|O_2 \;=\; P'_{1,2}\Lambda|O_1 D_{P'_{1,2}\Lambda|O_1} \tag{27}$$

and

$$\Sigma_u|O_2 \;=\; P'_{1,2}\Sigma_u|O_1 P_{1,2}. \tag{28}$$

Now consider one point from the posterior density of $\Lambda|O_1$. Take for example the loadings matrix derived from the principal components solution to the factor model for $Y|O_1$, which is then transformed to satisfy the PLT constraints, i.e.

$$\Lambda^*|O_1 \;=\; \begin{pmatrix} 0.411 & -0.333 & 0.319 & -0.249 & 0.827 & -0.435 & 0.537 & -0.807 & -0.179 & 0.327 \\ 0.000 & 0.419 & -0.720 & -0.716 & -0.027 & -0.252 & 0.364 & 0.378 & -0.360 & 0.656 \end{pmatrix}'. \tag{29}$$

This point can be transformed, using $P_{1,2}$ and $D_{P'_{1,2}\Lambda^*|O_1} = \begin{pmatrix} -0.6227 & -0.7824 \\ 0.7824 & -0.6227 \end{pmatrix}$, which is a rotation matrix with angle $\gamma = 0.714\pi$, to obtain

$$\begin{aligned} & P'_{1,2}\Lambda^*|O_1 D_{P'_{1,2}\Lambda^*|O_1} \\ = & \begin{pmatrix} 0.536 & -0.762 & 0.256 & -0.405 & -0.536 & 0.074 & -0.050 & 0.799 & -0.171 & 0.309 \\ 0.000 & 0.199 & -0.321 & 0.640 & 0.631 & 0.497 & -0.647 & 0.396 & 0.364 & -0.664 \end{pmatrix}', \end{aligned} \tag{30}$$

which is the same as $\Lambda^*|O_2$, i.e. the loadings matrix from the principal components solution for $Y|O_2$, transformed to satisfy the PLT constraints. Accordingly, the estimate for the covariance matrix for $Y|O_1$,

$$\Sigma^*_u|O_1 \;=\; \text{diag}(0.831,\ 0.713,\ 0.380,\ 0.426,\ 0.315,\ 0.747,\ 0.579,\ 0.205,\ 0.838,\ 0.463), \tag{31}$$

can be transformed to

$$P'_{1,2}\Sigma^*_u|O_1 P_{1,2} \;=\; \text{diag}(0.713,\ 0.380,\ 0.831,\ 0.426,\ 0.315,\ 0.748,\ 0.579,\ 0.205,\ 0.838,\ 0.463), \tag{32}$$

which is the same as $\Sigma_u^*|O_2$.

Next, consider another point from the posterior density of $\Lambda|O_1$ in the vicinity of $\Lambda^*|O_1$, say

$$\Lambda^\dagger|O_1 = \begin{pmatrix} 0.411 & -0.333 & 0.319 & -0.249 & 0.828 & -0.435 & 0.537 & -0.807 & -0.179 & 0.327 \\ 0.000 & \mathbf{0.319} & -0.720 & -0.716 & -0.027 & -0.252 & 0.364 & 0.378 & -0.360 & 0.656 \end{pmatrix}'. \tag{33}$$

We apply the same permutation and rotation matrices $P_{1,2}$ and $D_{P'_{1,2}\Lambda^*|O_1}$, and obtain

$$P'_{1,2}\Lambda^\dagger|O_1 D_{P'_{1,2}\Lambda^*|O_1} = \begin{pmatrix} 0.457 & -0.762 & 0.256 & -0.405 & -0.536 & 0.074 & -0.050 & 0.799 & -0.171 & 0.309 \\ \mathbf{0.062} & 0.199 & -0.321 & 0.640 & -0.631 & 0.497 & -0.647 & 0.396 & 0.364 & -0.664 \end{pmatrix}', \tag{34}$$

which does not satisfy the PLT constraints. If we do not chose the rotation matrix $D_{P'_{1,2}\Lambda^*|O_1}$, but the rotation matrix $D_{P'_{1,2}\Lambda^\dagger|O_1} = \begin{pmatrix} -0.7225 & -0.6914 \\ 0.6914 & -0.7225 \end{pmatrix}$, representing a rotation by an angle $\gamma = 0.757\pi$, we obtain

$$P'_{1,2}\Lambda^\dagger|O_1 D_{P'_{1,2}\Lambda^\dagger|O_1} = \begin{pmatrix} 0.461 & -0.728 & -0.297 & -0.315 & -0.616 & 0.140 & -0.137 & 0.845 & -0.120 & 0.217 \\ \mathbf{0.000} & 0.300 & -0.284 & 0.689 & -0.553 & 0.483 & -0.634 & 0.285 & 0.384 & -0.700 \end{pmatrix}'. \tag{35}$$

Thus the matrix $D_{P'_{1,2}\Lambda^*|O_1}$ cannot be applied to validly transform all points of the posterior density for $\Lambda|O_1$ into points of the posterior density for $\Lambda|O_2$. Conversely, there exist points in the posterior density of $\Lambda|O_2$ that can only be reached if a different rotation, namely $D_{P'_{1,2}\Lambda^\dagger|O_1}$, is applied.

	$\Lambda^*\|O_1, \Sigma_u^*\|O_1$	$\Lambda^*\|O_2, \Sigma_u^*\|O_2$	$\Lambda^*\|O_3, \Sigma_u^*\|O_3$
initial value	-755.9969	-755.9969	-755.9969
left rotation by $\gamma = -\frac{1}{5}\pi$	-757.5443	-761.1392	-761.4656
right rotation by $\gamma = \frac{1}{5}\pi$	-757.6931	-758.5168	-763.0706

Table 12: Log likelihood values for rotations of the PCA estimates.

Next, we will consider the principal components estimates for all three orderings $Y|O_1$, $Y|O_2$ and $Y|O_3$, which are afterwards transformed by $D_{\Lambda^*|O_1}$, $D_{\Lambda^*|O_2}$ and $D_{\Lambda^*|O_3}$, respectively, to satisfy the PLT constraint. Of course, all three solutions reach the same log likelihood value, which is -755.9969. Now consider a rotation of each loadings matrix to the left and to the right by some arbitrary angle, say $\frac{1}{5}\pi$. Obviously, the resulting value will no longer satisfy the PLT constraint, so we have to impose it separately. Table 12 shows the results of this experiment: For $Y|O_1$, imposing the constraints has little effect, so the log likelihood value stays nearly the same. For $Y|O_2$, if we rotate to the right, the likelihood again does not change much. For a rotation to the left, we observe a more pronounced change. Eventually, for $Y|O_3$, we see that both a rotation to the left and the right effects the log likelihood value to a greater extent. Figure 3 shows the same for the whole circle. Consider the solid line first: Whereas for $Y|O_1$, the log likelihood is nearly flat, it is flat on the left and steep on the right side for $Y|O_2$, whereas it is steep on both sides for $Y|O_3$. Note that the unit scale prior ensures that the contribution of the prior along the circle is identical, so the log likelihood must only be shifted to obtain the log posterior. It is easily seen that under the first ordering, the sampler can move along the circle much more easily than under the third ordering. For the second ordering, it can easily move to the right, but not to the left. Thus the ordering has an effect on the sampler's potential to move along the circle, or the range of rotations it will cover. For a wide range of rotations, represented by a flat log likelihood along the circle, the resulting posterior densities will lose their ellipticity property. This can be verified by comparing figure 3 with the sampler's output in the second row of figure 1.

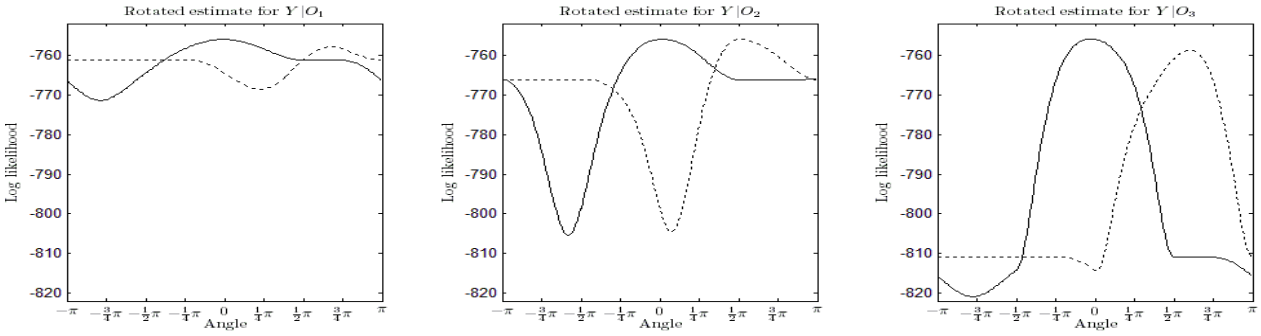

Figure 3: Log likelihood values of the principal components estimates, rotated along the circle, with constraints imposed.

Now note that rotations are not the only orthogonal transformations we must consider. There are also reflections and permutations, i.e. sign and label switchings of the factors and loadings. These may likewise occur while sampling. In the $\mathbb{R}^2$, all orthogonal transformations are expressible as a

product of one rotation and one permutation, so we are able to cover all the potential orthogonal transformations by additionally taking label switching into account. Obviously, the only label switching there is in a two-factor model is a switching of the first and second factor. Thus we exchange the factor labels and perform the same rotations as before, denoted by the dashed line in figure 3. We see that a second mode emerges. This mode is not located on the opposite side from the first one, nor is the distance between the first and second mode identical for all three orderings. We may thus consider this a case of genuine and non-systematic multimodality. Of course, the second mode is more likely reached by the sampler if it can cover larger parts of the circle. Thus we find that the sampler for $Y|O_1$ reaches a second mode, while the other two samplers do not, as the second row of figure 1 shows.

C The orthogonal Procrustes transformation

The orthogonal Procrustes problem is a minimization problem of the following form: Assume that we have two matrices X and Y, which both have dimension $N \times K$. We want to find an orthogonal matrix A that solves

$$XA = Y + E \quad \text{s.t.} \quad AA' = A'A = I_K \quad \text{with} \quad \text{tr}(E'E) = \min. \tag{36}$$

This problem has been solved by Schönemann (1966) in the following way: Write

$$g = g_1 + g_2 \tag{37}$$

with the minimization as

$$g_1 = \text{tr}(E'E) = \text{tr}(A'X'XA - 2A'X'Y + Y'Y) \tag{38}$$

and the constraint as

$$g_2 = \text{tr}(\Lambda(A'A - I)), \tag{39}$$

where Λ is a $K \times K$ matrix of Lagrange multipliers.

Taking the derivative with respect to A yields

$$\frac{\partial g}{\partial A} = (X'X + X'X)A - 2X'Y + A(\Lambda + \Lambda') \stackrel{!}{=} 0. \quad (40)$$

Rearranging the terms, we obtain

$$\frac{\Lambda + \Lambda'}{2} = A'X'Y - A'(X'X)A, \quad (41)$$

where the term on the left hand side and $A'(X'X)A$ are symmetric, so $A'X'Y$ must be symmetric as well, i.e.

$$A'X'Y = Y'XA, \quad (42)$$

or, equivalently,

$$X'Y = A'Y'XA. \quad (43)$$

If we write the square of the latter, we obtain

$$X'YY'X = A'Y'XAA'X'YA = A'YXX'YA, \quad (44)$$

since $AA' = I$. We can perform a spectral decomposition on both $X'YY'X$ and $Y'XX'Y$, resulting in

$$WDW' = AVDV'A', \quad (45)$$

where the matrix of eigenvalues D is the same for both decompositions, while the eigenvectors V and W are different. Now

$$W = AV, \quad (46)$$

and, consequently,

$$A = WV'. \quad (47)$$

We know have a necessary condition for A to be an orthogonal projection of X onto Y. Since we

need a minimum, we must consider

$$\begin{aligned} \mathrm{tr}(E'E) &= \mathrm{tr}(A'X'XA - 2A'X'Y + Y'Y) \\ &= \mathrm{tr}(X'X + Y'Y) - 2\mathrm{tr}(A'X'Y). \end{aligned} \tag{48}$$

Since this expression has to be minimized and the first term is fixed, we have to maximize the second term.

Plugging in the orthogonal projection solution for A from equation 47, we have

$$\begin{aligned} \mathrm{tr}(A'X'Y) &= \mathrm{tr}(VW'X'Y) \\ &= \mathrm{tr}(VW'WD^{0.5}V') \\ &= \mathrm{tr}(WW'D^{0.5}V'V) \\ &= \mathrm{tr}(D^{0.5}), \end{aligned} \tag{49}$$

where we use the singular value decomposition

$$X'Y = WD^{0.5}V' \tag{50}$$

and some properties of the trace function. Thus, solving for $D^{0.5}$ in equation (50), we obtain

$$D^{0.5} = W'X'YV = R'_W W'X'YVR_V, \tag{51}$$

where the column signs in V and W are undetermined, so the last equality holds for reflection matrices R_V and R_W. Since the solution in equation (47) is in fact

$$A = WV' = WR_W R_V V' = WR_{WV}V' \tag{52}$$

and the product of two reflection matrices is a reflection matrix itself, we merely have to find the R_{WV} that minimizes $D^{0.5}$ out of the K^2 potential solutions. This is obtained by complete enumeration.

D Simulation experiment for removing orthogonal mixing

An *orthogonally mixing* sampler generates a sequence of draws $\{\tilde{\Lambda}^{(s)}\}_{s=1}^{S}$ from a distribution that is subject to orthogonal transformations D^s after each draw s, generating an *orthogonally mixed* sample.

The orientation of the distribution in draw s relative to the orientation in draw 1 is thus $\tilde{D}^{(s)} = D^1 D^2 \dots D^{s-1}$. Note that $\tilde{D}^{(s)}$ is likewise an orthogonal matrix. Hence, the orthogonally mixed sample can also be interpreted as $\{\tilde{\Lambda}^{(s)}\tilde{D}^{(s)}\}_{s=1}^{S}$, where $\tilde{\Lambda}^{(s)}$ represents draw s from the directionally identified distribution, and $\tilde{D}^{(s)}$ represents the orthogonal transformation component. Both $\tilde{\Lambda}^{(s)}$ and $\tilde{D}^{(s)}$ stem from hidden Markov processes. We are not interested in the properties of $\tilde{D}^{(s)}$, but rather want to get rid of it. Nonetheless, sampling $\tilde{\Lambda}^{(s)}\tilde{D}^{(s)}$ should help to preserve the properties of $\tilde{\Lambda}^{(s)}$ better than sampling $\tilde{\Lambda}^{(s)}$ under identification constraints. Afterwards, we use the WOP algorithm to remove the effect of the $\tilde{D}^{(s)}$ from the sample and obtain $\{\tilde{\Lambda}^{(s)}D\}_{s=1}^{S}$, where D is a unique orthogonal matrix.

Using no directional identification constraints, the unconstrained Gibbs sampler for the factor model discussed in Appendix A is an orthogonally mixing sampler in this sense. Having previously obtained a draw $\tilde{\Lambda}^{(s)}D^s$, the moments of the full conditional distribution of F subject to the unobservable D^s from the previous iteration are

$$\begin{aligned}
\Omega_{f_t}|D^s &= (D^{s\prime}\Lambda'\Sigma_u^{-1}\Lambda D^s + I)^{-1} \\
&= (D^{s\prime}(\Lambda'\Sigma_u^{-1}\Lambda + I)D^s)^{-1} \\
&= D^{s\prime}(\Lambda'\Sigma_u^{-1}\Lambda + I)^{-1}D^s \\
&= D^{s\prime}\Omega_{f_t}D^s
\end{aligned} \tag{53}$$

and

$$\begin{aligned}
\mu_{f_t}|D^s &= (D^{s\prime}\Omega_{f_t}D^s)(D^{s\prime}\Lambda\Sigma_u y_t) \\
&= D^{s\prime}\Omega_{f_t}\Lambda\Sigma_u y_t \\
&= D^{s\prime}\mu_{f_t}.
\end{aligned} \tag{54}$$

Analogously, the moments of the full conditional distribution of Λ subject to the unobservable matrix D^s from the previous iteration are

$$\begin{aligned}
\Omega_{\lambda_i}|D^s &= (\sigma_i^{-2}D^{s\prime}F'FD^s + I)^{-1} \\
&= (D^{s\prime}(\sigma_i^{-2}F'F + I)D^s)^{-1} \\
&= D^{s\prime}(\sigma_i^{-2}F'F + I)^{-1}D^s \\
&= D^{s\prime}\Omega_{\lambda_i}D^s
\end{aligned} \tag{55}$$

and

$$\begin{aligned}\mu_{\lambda_i}|D^s &= D^{s\prime}\Omega_{\lambda_i}D^s(\sigma_i^{-2}D^{s\prime}F'y_i)\\ &= D^{s\prime}\Omega_{\lambda_i}(\sigma_i^{-2}F'y_i)\\ &= D^{s\prime}\mu_{\lambda_i}. \end{aligned} \tag{56}$$

The following sweep for Λ will then yield $\tilde{\Lambda}^{(s+1)}D^{s+1}$, where the unobservable $\tilde{\Lambda}^{(s+1)}$ is a draw from the full conditional distribution prior to its latest orthogonal transformation D^{s+1}. If we want all the $\tilde{\Lambda}^{(s)}$ to originate from a distribution with the same directional identification, we use the above reasoning and replace D^s by $\tilde{D}^{(s)}$ in the conditional moments.

The following experiment will demonstrate the capability of the weighted orthogonal Procrustes (WOP) procedure to recover parameters from an orthogonally mixed sample. As we generate the sample, we will follow the notion of $\tilde{\Lambda}^{(s)}$ and $\tilde{D}^{(s)}$ as separate unobservable processes, where the $\tilde{\Lambda}^{(s)}$ are drawn from a directionally identified distribution and the $\tilde{D}^{(s)}$ are arbitrary orthogonal matrices.

We simulate a sequence of matrices $\{\tilde{\Lambda}^{(s)}\}_{s=1}^S$, where $S = 10000$ and $\tilde{\Lambda}^{(s)} = (\tilde{\lambda}_1^{(s)\prime}, \ldots, \tilde{\lambda}_N^{(s)\prime})$ with $N = 100$ and $\tilde{\lambda}_i^{(s)} \sim \mathcal{N}(\mu_i, \Sigma_i)$ for all $i = 1, \ldots, N$ with the parameters μ_i drawn from a K-variate standard Normal distribution with $K = 6$ and Σ_i drawn from a K-variate scaled Wishart distribution $\frac{1}{\nu}\mathcal{W}(\Sigma, \nu)$ with $\Sigma = I_K$ and $\nu = 10$.

Next, we create subsamples containing the first k columns of each matrix, where $k \in \{2, 3, 4, 5, 6\}$, i.e. $\{\tilde{\Lambda}_{[1:k]}^{(s)}\}_{z=1}^Z$. For each of the resulting five samples, we then draw a sequence of k-dimensional rotation matrices $\{\tilde{Q}_k^{(s)}\}_{s=1}^S$, a sequence of k-dimensional permutation matrices $\{\tilde{P}_k^{(s)}\}_{s=1}^S$, and a sequence of k-dimensional reflection matrices $\{\tilde{R}_k^{(s)}\}_{s=1}^S$, afterwards multiplying them to obtain $\{\tilde{D}_k^{(s)} = \tilde{Q}_k^{(s)}\tilde{P}_k^{(s)}\tilde{R}_k^{(s)}\}_{s=1}^S$, which is again a sequence of orthogonal matrices. We then use the orthogonal matrices of the appropriate dimension to create the orthogonally mixed samples, i.e. $\{\tilde{\Lambda}_{[1:k]}^{\tilde{D}^{(s)}} = \tilde{\Lambda}_{[1:k]}^{(s)}\tilde{D}_k^{(s)}\}_{s=1}^S$.

Now we consider only the first n rows of each matrix for each of the five samples, where $n \in \{10, 20, 60, 100\}$. Note that the entries of the first five rows and two columns in the data before adding the orthogonal variation is identical throughout the resulting 20 samples. We can thus evaluate the effect of increasing the number of rows or the number of columns based on these reference parameters.

Consider the WOP procedure for $\{\tilde{\Lambda}_{[1:n,1:k]}^{\tilde{D}^{(s)}}\}_{s=1}^S$, with fixed point ${}_{(g)}\Lambda_{[1:n,1:k]}^*$ and weights matrix $W_{(g)}$ with $g \in \{0, 1, 2, \ldots, G-1\}$ denoting the number of completed iterations of the sampler, where for $g = 0$, we have fixed point ${}_{(0)}\Lambda_{[1:n,1:k]}^* = \tilde{\Lambda}_{[1:n,1:k]}^{\tilde{D}^{(S)}}$, i.e. the last observation, and

$W_{(0)} = \text{diag}({}_{(0)}w_{11}, \ldots, {}_{(0)}w_{nn})$ with ${}_{(0)}w_{ii} = S\left(\sum_{s=1}^{S}\sqrt{\tilde{\lambda}_i^{\tilde{D}^{(s)}\prime}\tilde{\lambda}_i^{\tilde{D}^{(s)}}}\right)^{-1}$, and for $g \geq 1$, we use as fixed point ${}_{(g)}\Lambda^*_{[1:n,1:k]} = \frac{1}{S}\sum_{s=1}^{S}{}_{(g)}\tilde{\Lambda}^{(s)}_{[1:n,1:k]}$ and $W_{(g)} = \text{diag}({}_{(g)}w_{11}, \ldots, {}_{(g)}w_{nn})$ with ${}_{(g)}w_{ii} = \det\left(\frac{1}{S}\sum_{s=1}^{S}({}_{(g)}\tilde{\lambda}_i^{(s)} - {}_{(g)}\lambda_i^*)({}_{(g)}\tilde{\lambda}_i^{(s)} - {}_{(g)}\lambda_i^*)'\right)^{-\frac{1}{k}}$, where $\{{}_{(g)}\tilde{\Lambda}^{(s)}_{[1:n,1:k]}\}_{s=1}^{S}$ denotes the output of the WOP procedure from iteration g. Hence the output after G iterations is

$$\{{}_{(G)}\tilde{\Lambda}^{(s)}_{[1:n,1:k]}\}_{s=1}^{S} \quad = \quad WOP(\{\tilde{\Lambda}^{\tilde{D}^{(s)}}_{[1:n,1:k]}\}_{s=1}^{S}, G). \tag{57}$$

We apply the WOP procedure to each of the orthogonally mixed samples, running it for $G = 5$ iterations. The restored sample has an orientation different from the original one. Thus we determine the transformation matrix for the (unweighted) orthogonal Procrustes that projects the mean of the restored sample onto the mean of the original sample before orthogonal mixing. This transformation is then applied to each matrix, aligning the restored data with the mean of the original data.

Eventually, we calculate the element-wise percentiles of the original data before orthogonal mixing and the restored data after removing the orthogonal mixing and compare them. If the original data has been properly restored, the percentiles should be very close to each other. Figure 4 shows the QQ plots for $k = 2$ and $k = 5$, with increasing number of rows used per matrix. Considering parameter $\lambda_{5,2}$ for $k = 2$, we find that the first WOP iteration does not restore the percentiles well if only information from 10 rows is available. If we add more rows, allowing more information to be used, the problem disappears. Likewise, additional WOP iterations help to match the restored with the original percentiles. Now consider parameter $\lambda_{1,1}$ for $k = 5$. If only ten rows are available - a case where the identification constraints for the factor model are not satisfied - additional WOP iterations do not ensure a good percentile matching. Adding more information by using additional rows of each matrix, however, helps to fix this problem.

We now measure the squared deviation between the 1st, 5th, 25th, 50th, 75th, 95th and 99th percentiles of the simulated and the restored data. Table 13 reports the average squared deviations for k ranging from 2 to 6. The left seven columns in each table show the results if the empirical mean from the unobservable data before mixing is used, the right seven columns show the results if only information from the orthogonally mixed sample is used. Consider first the left seven columns, where the mean before mixing is assumed to be known. We see that for increasing G and n, the average squared deviations for all the considered quantiles approach zero. Thus for sufficiently large

G and n,

$$WOP\left(\{\tilde{\Lambda}^{\tilde{D}^{(s)}}_{[1:n,1:k]}\}_{s=1}^{S}, G\middle|\ \frac{1}{S}\sum_{s=1}^{S}\tilde{\Lambda}^{(s)}_{[1:n,1:k]}\right) \approx \{\tilde{\Lambda}^{(s)}_{[1:n,1:k]}D\}_{s=1}^{S}, \tag{58}$$

where D is a unique orthogonal matrix very close to the identity matrix. Hence we can restore the information from the simulated data out of the orthogonally mixed sample if we only know the mean of the simulated data. The effect of additional iterations of the sampler is almost negligible, so we may choose G as low as 1 or 2.

Now we compare the results from the right seven columns per table with those from the left seven columns per table. We see that the average squared deviations from the sampler that uses information about the - unobservable - mean of the simulated data and the sampler that does not use this information approach each other as G increases. Thus, for sufficiently large G and n,

$$WOP(\{\tilde{\Lambda}^{\tilde{D}^{(s)}}_{[1:n,1:k]}\}_{s=1}^{S}, G) \approx \{\tilde{\Lambda}^{(s)}_{[1:n,1:k]}D\}_{s=1}^{S}, \tag{59}$$

i.e. we do not need the mean of the simulated data to recover other information about the data. Note, however, that the effect of additional iterations of the sampler is quite large here, so we may choose a larger value for G than if the mean of the simulated data is known, where each iteration improves improves inference about the mean, which serves as the fixed point in the subsequent iteration of the sampler. Moreover, D is not close to the identity matrix any more, but instead close to the orthogonal projection of ${}_{(0)}\Lambda^{*}_{[1:n,1:k]}$ onto the mean of the simulated data.

Overall, we observe that the average squared deviation for the central quantiles is much smaller than for the tails, which is easily explained by the fact that each normally distributed sample contains more information about the center of the distribution than about its tails. Moreover, for increasing values of k, we require larger values of n, because each additional dimension adds a source of uncertainty. The angular deviation between vectors in the $\mathbb{R}^2$ is univariate, while the angular deviation between vectors in the $\mathbb{R}^3$ is bivariate, etc.

Figure 4: QQ plots for original and restored data, showing results for the first 5 rows and 2 columns. From top to bottom n is 10, 20, 60 and 100. Left column uses matrices with $k = 2$ columns, right column uses matrices with $k = 5$ columns. Light and dark grey markers denote results after one and five WOP iterations, respectively.

k	n	G	q01	q05	q25	q50	q75	q95	q99	q01	q05	q25	q50	q75	q95	q99
2	10	1	0.0127	0.0063	0.0015	0.0009	0.0031	0.0098	0.0224	0.4215	0.1678	0.0603	0.0543	0.0861	0.3272	0.6349
2	10	5	0.0116	0.0057	0.0013	0.0007	0.0024	0.0082	0.0193	0.0116	0.0057	0.0013	0.0007	0.0024	0.0082	0.0193
2	20	1	0.0040	0.0020	0.0002	0.0003	0.0013	0.0028	0.0054	0.0589	0.0106	0.0030	0.0032	0.0055	0.0182	0.0737
2	20	5	0.0028	0.0016	0.0001	0.0002	0.0009	0.0019	0.0033	0.0028	0.0016	0.0001	0.0003	0.0009	0.0019	0.0033
2	60	1	0.0005	0.0005	0.0000	0.0001	0.0002	0.0004	0.0018	0.0011	0.0004	0.0001	0.0001	0.0004	0.0008	0.0038
2	60	5	0.0005	0.0004	0.0000	0.0001	0.0002	0.0003	0.0013	0.0005	0.0004	0.0000	0.0001	0.0002	0.0003	0.0013
2	100	1	0.0006	0.0002	0.0000	0.0000	0.0001	0.0001	0.0007	0.0007	0.0002	0.0000	0.0000	0.0002	0.0005	0.0021
2	100	5	0.0004	0.0002	0.0000	0.0000	0.0001	0.0001	0.0004	0.0004	0.0002	0.0000	0.0000	0.0001	0.0001	0.0004
3	10	1	0.0127	0.0063	0.0015	0.0009	0.0031	0.0098	0.0224	0.4215	0.1678	0.0603	0.0543	0.0861	0.3272	0.6349
3	10	5	0.0116	0.0057	0.0013	0.0007	0.0024	0.0082	0.0193	0.0116	0.0057	0.0013	0.0007	0.0024	0.0082	0.0193
3	20	1	0.0040	0.0020	0.0002	0.0003	0.0013	0.0028	0.0054	0.0589	0.0106	0.0030	0.0032	0.0055	0.0182	0.0737
3	20	5	0.0028	0.0016	0.0001	0.0002	0.0009	0.0019	0.0033	0.0028	0.0016	0.0001	0.0003	0.0009	0.0019	0.0033
3	60	1	0.0005	0.0005	0.0000	0.0001	0.0002	0.0004	0.0018	0.0011	0.0004	0.0001	0.0001	0.0004	0.0008	0.0038
3	60	5	0.0005	0.0004	0.0000	0.0001	0.0002	0.0003	0.0013	0.0005	0.0004	0.0000	0.0001	0.0002	0.0003	0.0013
3	100	1	0.0006	0.0002	0.0000	0.0000	0.0001	0.0001	0.0007	0.0007	0.0002	0.0000	0.0000	0.0002	0.0005	0.0021
3	100	5	0.0004	0.0002	0.0000	0.0000	0.0001	0.0001	0.0004	0.0004	0.0002	0.0000	0.0000	0.0001	0.0001	0.0004
4	10	1	0.1380	0.0750	0.0221	0.0109	0.0157	0.0489	0.0950	0.4836	0.2968	0.2224	0.2959	0.4969	1.0023	1.4729
4	10	5	0.2112	0.1234	0.0285	0.0110	0.0288	0.1227	0.2606	0.1092	0.0627	0.0156	0.0064	0.0121	0.0470	0.1047
4	20	1	0.0247	0.0107	0.0030	0.0019	0.0056	0.0125	0.0324	0.4811	0.2774	0.1838	0.2126	0.3193	0.6419	0.9942
4	20	5	0.0140	0.0096	0.0022	0.0012	0.0041	0.0122	0.0297	0.0140	0.0098	0.0022	0.0012	0.0042	0.0119	0.0294
4	60	1	0.0045	0.0019	0.0007	0.0004	0.0011	0.0021	0.0039	0.0172	0.0056	0.0020	0.0032	0.0074	0.0204	0.0712
4	60	5	0.0024	0.0018	0.0005	0.0003	0.0007	0.0021	0.0051	0.0024	0.0018	0.0005	0.0003	0.0007	0.0021	0.0051
4	100	1	0.0017	0.0006	0.0002	0.0001	0.0003	0.0009	0.0034	0.0053	0.0022	0.0005	0.0006	0.0010	0.0022	0.0082
4	100	5	0.0011	0.0006	0.0001	0.0001	0.0002	0.0008	0.0027	0.0011	0.0006	0.0001	0.0001	0.0002	0.0008	0.0027
5	10	1	0.2047	0.1092	0.0328	0.0153	0.0236	0.0698	0.1410	0.1958	0.0908	0.0574	0.0871	0.1405	0.2851	0.4681
5	10	5	0.2209	0.1162	0.0388	0.0199	0.0336	0.0992	0.1991	0.0885	0.0430	0.0129	0.0080	0.0139	0.0375	0.0640
5	20	1	0.0492	0.0234	0.0082	0.0045	0.0092	0.0206	0.0408	0.2097	0.1067	0.0931	0.1431	0.2415	0.4678	0.7365
5	20	5	0.0279	0.0144	0.0039	0.0022	0.0054	0.0139	0.0256	0.0289	0.0147	0.0041	0.0021	0.0050	0.0135	0.0246
5	60	1	0.0059	0.0022	0.0008	0.0005	0.0014	0.0028	0.0054	0.0183	0.0085	0.0026	0.0036	0.0095	0.0271	0.0848
5	60	5	0.0056	0.0027	0.0008	0.0004	0.0012	0.0032	0.0061	0.0056	0.0027	0.0008	0.0004	0.0012	0.0032	0.0061
5	100	1	0.0020	0.0008	0.0002	0.0001	0.0005	0.0011	0.0019	0.0046	0.0022	0.0007	0.0008	0.0015	0.0019	0.0079
5	100	5	0.0020	0.0008	0.0002	0.0001	0.0004	0.0010	0.0022	0.0020	0.0008	0.0002	0.0001	0.0004	0.0010	0.0022
6	10	1	0.6325	0.3461	0.0871	0.0182	0.0328	0.1715	0.3615	0.4471	0.2312	0.0665	0.0573	0.1242	0.3694	0.6909
6	10	5	1.5331	0.8137	0.1947	0.0416	0.0799	0.4303	0.9185	0.9713	0.5245	0.1151	0.0167	0.0468	0.2902	0.6082
6	20	1	0.1417	0.0750	0.0186	0.0053	0.0137	0.0530	0.1099	0.1353	0.1015	0.1362	0.2178	0.3755	0.7393	1.1287
6	20	5	0.1192	0.0663	0.0149	0.0047	0.0102	0.0377	0.0816	0.1007	0.0545	0.0126	0.0042	0.0093	0.0302	0.0630
6	60	1	0.0118	0.0059	0.0015	0.0005	0.0014	0.0057	0.0123	0.0485	0.0161	0.0065	0.0101	0.0189	0.0595	0.1497
6	60	5	0.0095	0.0045	0.0012	0.0004	0.0012	0.0044	0.0107	0.0095	0.0045	0.0012	0.0004	0.0012	0.0044	0.0107
6	100	1	0.0039	0.0020	0.0005	0.0002	0.0006	0.0018	0.0042	0.0106	0.0041	0.0011	0.0015	0.0029	0.0070	0.0159
6	100	5	0.0039	0.0017	0.0005	0.0002	0.0005	0.0018	0.0047	0.0039	0.0017	0.0005	0.0002	0.0005	0.0018	0.0047

Table 13: Average squared deviations between simulated and restored samples for $k = 2$ to $k = 6$, measured at the 1st (q01), 5th (q05), 25th (q25), 50th (q50), 75th (q75), 95th (q95) and 99th (q99) percentiles. The first seven columns report results for known means of the simulated data, the next seven columns report results for unknown means. n denotes the number of rows used, G denotes the number of iterations of the WOP procedure.