A geometric approach to factor model identification

Sylvia Kaufmann and Markus Pape

Working Paper 24.06

A geometric approach to factor model identification

Sylvia Kaufmann* and Markus Pape†

November 4, 2024

Abstract

We use the geometric representation of factor models to represent the factor loading structure by sets corresponding to unit-specific non-zero loadings. We formulate global and local identification conditions based on set conditions. We propose two algorithms to efficiently evaluate Sato (1992)'s counting rule. We demonstrate the efficiency and the performance of the algorithms with a simulation study. An application to exchange rate returns illustrates the approach.

Keywords: Factor model, Global and local identification, Set representation.

JEL: C10, C52, C55, E44, F31

*Study Center Gerzensee, Dorfstrasse 2, 3115 Gerzensee, Switzerland, and University of Basel
†Department of Economics, Ruhr-University Bochum, Universitaetsstrasse 150, 44801 Bochum, Germany

1 Introduction

Identification in factor modelling often relates to identifying the scale of factors and the orientation of the factor basis determining factor ordering and sign. The more fundamental issue of whether the decomposition into common and idiosyncratic variation is (uniquely) identified has eclipsed, as increasingly high-dimensional data sets have become available (N large), and, relying on their informativeness, few factors (K small) are usually expected to capture the bulk of data covariation. In exploratory analysis, standard non-parametric methods extract factors by principal or frequency components analysis (Stock and Watson, 2002; Forni et al., 2005). While in these approaches factors are determined and ordered according to the average share of covariation explained across series, the interpretation of factors gets increasingly blurred when they are extracted from high-dimensional sets of heterogeneous data. The situation is similar when factor extraction is based on singular value decomposition (Hoff, 2007; Chan et al., 2018). In the parametric framework, factor identification and ordering is usually induced by imposing restrictions on the factor loading matrix. Widely used approaches impose restrictions before estimation (Geweke and Zhou, 1996; Aguilar and West, 2000; Bernanke et al., 2005) while more recent ones apply efficient and order-invariant estimation, where factor identification and ordering is obtained by post-processing (Aßmann et al., 2016; Chan et al., 2018; Kaufmann and Schumacher, 2019).

Originating in psychometrics, a primary goal in confirmatory factor analysis is to obtain, besides factor identification, a factor interpretation. This is induced either by rotating factors to maximize the share of factor-specific variation in specific units, see e.g. Thurstone (1935), Cattell (1978) or Bollen (1989), or imposing a simple structure on the factor loading matrix to determine factor interpretation, i.e. imposing zero loadings on series not expected to determine or be loaded by specific factors, see e.g. Jöreskog (1969), Jennrich (1978), Jöreskog (1979), Millsap (2001) or Mulaik (2010). These approaches have lately been applied also in econometrics (Bai and Ng, 2013; Kose et al., 2003; Moench et al., 2013). Sparse (exploratory) factor analysis draws on the advantage of inducing a simple structure into the factor loading matrix to extract relevant factors or eventually obtain factor interpretation (West, 2003; Lucas et al., 2006; Beyeler and Kaufmann, 2021). Recent advances in econometrics propose estimation procedures also relying on sparse principal components (Zou et al., 2006; Despois and Doz, 2023) or regularized estimation (Freyaldenhoven, 2023).

Identification has been discussed since the very early stages of factor modelling (Ledermann, 1937; Anderson and Rubin, 1956). Global identification dealt with the question of unique identification of the variance decomposition, related to the number of factors identifiable from the covariance structure of the data (Shapiro, 1982). Imposing a simple structure in confirmatory factor analysis raised similar issues (Millsap, 2001; Sato, 1992). Nowadays, sparse exploratory factor analysis in high-dimensional data settings raises the issue of whether a sparse reduced rank representation is (uniquely) identifiable from the data covariance (Frühwirth-Schnatter et al., 2023; Kaufmann and Pape, 2023), given that in a factor model of potentially increasing dimension the factor loading matrix may become increasingly sparse.

This raises anew the issue of the identification of the variance decomposition (global identification) and factors (rotational identification), which motivates the present paper. We review

relevant global and local identification and uniqueness conditions that have been documented in the literature. Besides conditions derived in Anderson and Rubin (1956) and Sato (1992) related to global identification, we will review conditions for mode or rotational identification (Bekker, 1986; Neudecker, 1990; Bai and Wang, 2014). Our contribution relies on the geometric representation of factor models (Lawley and Maxwell, 1971) to derive a geometric approach to identification. Full-rank, rotational identification is based on set identification restrictions, where only (at least) K instead of $2^K - 1$ sets of subspaces lying in the K-dimensional factor space need to be populated for full-rank, rotational identification. Sato (1992) derived a counting rule which provides a necessary condition for global identification of the reduced-rank variance decomposition. Checking the rule needs an exponentially growing combinatorial number of evaluations $(2^K - 1)$, which becomes quickly computationally prohibitive for factor models of increasing dimensions. Based on our geometric approach to identification, we provide two algorithms, the $\mathcal{O}(K^2)$ and the set-based algorithm for checking efficiently Sato's counting rule with a high precision.

Related to our paper, Frühwirth-Schnatter et al. (2023) present a thorough analysis of identification based on the so-called generalized lower triangular (GLT) structure, where the leading non-zero loadings of factors (pivot elements) need to be in different rows (pivot rows). While imposing or rotating into a GLT structure may serve the evaluation of model identification, we will argue that imposing a GLT structure for estimation is prone to the same problem as the widely applied positive lower triangular (PLT) identification constraints. Imposing a GLT structure is not order-invariant, may induce an orientation of the factor basis destroying or biasing the sparse structure underlying the data. Our approach to identification suggests that requiring the loadings leading the pivot elements to equal zero is potentially misleading when the purpose is to estimate a sparse factor loading matrix. Rather, it suffices to determine pivot rows or *pivot series*, who when re-ordered first in the data set would lead to an unordered GLT structure in the factor loading matrix. The result of our set identification procedure can help in determining those pivot series.

In the next section, we introduce the factor model and briefly review published results and issues related to global and local identification, relevant for motivating our contribution. Section 3 relies on the geometric interpretation of factor models to introduce the geometric approach to identification. Based on these considerations, we propose two algorithms, the $\mathcal{O}(K^2)$ and the set-based algorithm, which allow an efficient evaluation of Sato's counting rule (Sato, 1992). Section 4 documents the efficiency and the performance of the algorithms by comparing them to the plain evaluation of Sato's rule. An application to exchange rate returns illustrates the approach. Section 5 concludes.

2 Representation and identification

We review results concerning global and local identification of factor models published in the literature, which are relevant for motivating and introducing the geometric approach to identification.

2.1 Representation

Collect observed data in vector $Y = (y_1', \ldots, y_T')'$, where each y_t, $t = 1, \ldots, T$, denotes an $N \times 1$ vector of variables or items y_{it}, $i = 1, \ldots, N$, and can be represented as

$$y_t = \Lambda f_t + \epsilon_t, \tag{1}$$
$$E\left(f_t f_t'\right) = I_K,\ E\left(\epsilon_t \epsilon_t'\right) = \Sigma_\epsilon,\ \Sigma_\epsilon \text{ diagonal}, \tag{2}$$

with $K << N$ and where f_t is a $K \times 1$ vector of latent factors, $\Lambda = \{\lambda_{ij} | i = 1, \ldots, N, j = 1, \ldots, K\}$ is the $N \times K$ factor loading matrix and ϵ_t is a $N \times 1$ vector of idiosyncratic components.[1] As common variation is captured by the factor component only, Σ_ϵ is diagonal and $E(f_t \epsilon_t') = 0$. Finally, identification issues we discuss in the following are independent of distributional assumptions; therefore, we do not specify any so far. We assume that first and second (unconditional) moments are, respectively, zero and constant, which means that observed data in (1) follows a covariance-stationary process.

Underlying factors are usually unobserved and have to be extracted from observed data variation, $\Sigma_y = E(y_t y_t')$:

$$\Sigma_y = \Lambda\Lambda' + \Sigma_\epsilon \tag{3}$$

Finding a solution to (3) does not only mean mathematically solving the system of $N(N+1)/2$ independent equations. A valid decomposition requires Σ_ϵ to be positive definite and $\Sigma_y - \Sigma_\epsilon$ positive semi-definite and of lower-rank K.

Questions that arise are (a) does a solution exist and is it unique, which concerns *global* identification; (b) is Σ_ϵ unique, which concerns *local* identification, and (c) for an identified solution, how to determine the orientation of the factor basis which includes factor order and sign, which concerns *rotational* or *mode* identification. We briefly review results concerning the first two issues before elaborating in more details on the last one.

2.2 Global and local identification

The most general concept of identification implies that for the decomposition into Λ and Σ_ϵ, there exists only a single solution. Several results have established the necessary and sufficient conditions for identifying the factor model, some of which are discussed below. However, a complete solution to the global identification problem has not yet been achieved, see e.g. Millsap (2011).

Definition 1: Model (3) is *globally identified (unique)*, if there is a (unique) Σ_ϵ with non-negative diagonal elements which solves (3) and yields a common covariance matrix $\Sigma_y - \Sigma_\epsilon$ that is positive-semidefinite and of reduced rank K (Anderson and Rubin, 1956, AR56).

[1] We assume without loss of generality an identity covariance matrix for factors, given that correlated factors $\tilde{f}_t$ can be de-correlated by using e.g. a Cholesky decomposition of the factor covariance: $E\left(\tilde{f}_t \tilde{f}_t'\right) = \Sigma_{\tilde{f}} = LL'$; $L^{-1}\Sigma_{\tilde{f}} L^{-1'} = I_K$. When post-multiplying $\tilde{\Lambda}$ with L, the factor model with correlated factors is observationally equivalent to system (1). $y_t = \tilde{\Lambda} L L^{-1} \tilde{f}_t + \epsilon_t = \Lambda f_t + \epsilon_t$.

Definition 2: Model (3) is *locally unique*, if in the neighbourhood of Σ_ϵ there is no other idiosyncratic variance matrix $\tilde{\Sigma}_\epsilon$ that solves (3) and yields a common covariance $\Sigma_y - \tilde{\Sigma}_\epsilon$ that is positive-definite and of reduced rank K (Shapiro, 1985, 1989).

A sufficient condition for global uniqueness is given in Theorem 5.1 of AR56 (adjusted here in notation)

> *A sufficient condition for identification of Σ_ϵ and Λ up to multiplication on the right by an orthogonal matrix is that if any row of Λ is deleted there remain two disjoint submatrices of rank K.*

Further theorems provide either necessary or necessary and sufficient conditions for global identification of particular models, see also the overview in Bollen (1989, ch. 7).

Definition 1 includes stronger conditions than the criterion given by Ledermann's bound (Ledermann, 1937), which equates the number of equations with the number of unknowns in (3) and yields $\varphi(N) = (2N + 1 - \sqrt{8N+1})/2$. The bound yields a hypothetical criterion about the number of factors needed to obtain a reduced-rank decomposition as in (3). It is not very useful, though, because an algebraic solution does not exclude cases where Σ_ϵ contains negative elements (Heywood case) or $\Sigma_y - \Sigma_\epsilon$ is not positive-semidefinite.[2]

As regards the Ledermann bound, Shapiro (1982) showed that it is almost surely a lower bound for the number of factors needed to obtain a reduced rank common covariance.[3] Shapiro (1985) further showed that any Σ_ϵ which solves (3) is almost surely non-unique if K is above the Ledermann bound and almost surely locally unique when K is at or below the Ledermann bound. Bekker and Berge (1997) then showed that Σ_ϵ is almost surely globally unique if K is strictly below the Ledermann bound.

---

[2] A noteworthy insight from counting is that at least three loadings must be non-zero to identify one factor, as a minimum of six covariances is needed to identify six unknown parameters:

$$\Sigma_y = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2 & \sigma_1\sigma_3 \\ \sigma_2\sigma_1 & \sigma_2^2 & \sigma_2\sigma_3 \\ \sigma_3\sigma_1 & \sigma_3\sigma_2 & \sigma_3^2 \end{bmatrix} = \begin{bmatrix} \lambda_1^2 + \sigma_{\epsilon 1}^2 & \lambda_1\lambda_2 & \lambda_1\lambda_3 \\ \lambda_2\lambda_1 & \lambda_2^2 + \sigma_{\epsilon 2}^2 & \lambda_2\lambda_3 \\ \lambda_3\lambda_1 & \lambda_3\lambda_2 & \lambda_3^2 + \sigma_{\epsilon 3}^2 \end{bmatrix}.$$

Theorem 4.2 of AR56 states additional conditions for a solution satisfying a factor representation (adjusted in notation)

> *A necessary and sufficient condition that Σ_y be a covariance matrix of a factor analysis model with one factor is that $N(N-1)/2 - N$ independent tetrad conditions are satisfied and*
>
> $$0 \leq \frac{\sigma_{ki}\sigma_{ij}}{\sigma_{kj}} \leq \sigma_i^2, \ \sigma_{kj} \neq 0$$
>
> *for one pair $(j \neq k)$ for each i.*

When $N = 3$, no tetrad conditions are left and the solution is exact: $\sigma_{\epsilon,k} = \sigma_k^2 - \lambda_k^2$, $k = 1, \ldots, K$. When $N = 4$, two additional tetrad conditions need to be fulfilled: $\sigma_{31}\sigma_{42} - \sigma_{41}\sigma_{32} = 0$, $\sigma_{21}\sigma_{34} - \sigma_{31}\sigma_{24} = 0$. For $K > 1$ conditions become more complicated (Wilson and Worcester, 1939, AR56). From these considerations, we conclude that whether a representation (3) exists depends ultimately on characteristics of the data covariance Σ_y (AR56, last paragraph of Section 5).

[3] That is, there are "almost no" sample covariance matrices that have factor structure with the number of factors lower than the Ledermann bound.

Note that AR56 do not discuss the Ledermann bound explicitly, but evaluate the number of degrees of freedom of the (static) factor model, which is

$$c = \frac{N(N+1)}{2} - N - \left(NK - \frac{K(K-1)}{2} \right) = \frac{(N-K)^2 - N - K}{2}. \tag{4}$$

They state that an algebraic solution is possible when $c \leq 0$ whereas when $c > 0$, there is generally no solution. In the case of $c \leq 0$, the solution may not be unique. Hence, it may be appropriate to increase c, e.g. by imposing additional restrictions on Λ. Both confirmatory factor analysis and sparse factor analysis increase c, albeit usually not in a fashion that results in a unique solution. An increase in the number of factors K, on the other hand, reduces c. In practical context, the case $c > 0$ seems much more relevant. As it is almost impossible to find a parsimonious solution for a factor representation of a given sample covariance matrix S_y, see Shapiro (1982), it is usually assumed that $S_y \neq \Sigma_y$, and $\hat{\Sigma}_y = \hat{\Lambda}\hat{\Lambda}' + \hat{\Sigma}_\epsilon$. Implicitly, $\hat{\Sigma}_y$ thus "satisfies some c conditions" (AR56), imposed through the structure of the factor model. Whether $\hat{\Sigma}_y$ is a plausible estimate, however, also depends on the type of model used. For instance, a sparse factor model with many factors is much more flexible than a dense model with fewer factors, even if the number of parameters is the same.

Sato (1992) frames global identification into the analysis of correlation, and represents regions of unique decompositions geometrically. He qualifies and extends results of AR56, in particular derives a necessary conditions for Theorem 5.1 of AR56 to hold (Sato, 1992, Theorem 3.3., adjusted here in notation):

> *A necessary condition for satisfying the condition of Theorem [5.1 in AR56] is that the submatrices which consist of each q columns of ΛG have at least $(2q+1)$ nonzero rows for every nonsingular G $(q = 1, 2, \ldots, K)$.*

This "counting rule" (Frühwirth-Schnatter et al., 2023) encompasses Theorem 5.6 in AR56, which states a necessary condition for global uniqueness to be that each column of ΛG (for every non-singular matrix G) should have at least three non-zero loadings. The results prove useful to assess global and local identification in $K << N$ settings, where the factor loading matrix may be sparse (West, 2003).

2.3 Rotational or mode identification

Definition 3: A locally unique model defines an *equivalence class*. The equivalence class contains all solutions for which $\Sigma_y - \Sigma_\epsilon = \Lambda\Lambda'$ is positive-semidefinite and has reduced rank K, see also Millsap (2001).

Definition 4: A solution in an equivalence class which identifies the orientation and scale of the factor basis (rotational identification), factor position and sign, is called a *mode-identified* solution.

Assuming a locally unique solution to (3) is given, it is clear that (3) is identified up to rotation, given that $\Lambda\Lambda'$ can always be replaced by $\Lambda QQ'\Lambda'$ using an orthogonal matrix Q,

or, if $\Sigma_{\tilde{f}} \neq I_K$, $\tilde{\Lambda}\Sigma_{\tilde{f}}\tilde{\Lambda}'$ can always be replaced by $\Lambda^*\Sigma_{f^*}\Lambda^{*\prime} = \tilde{\Lambda}QQ^{-1}\Sigma_{\tilde{f}}Q^{-1\prime}Q'\tilde{\Lambda}'$ using an arbitrary, non-singular matrix Q.

A first set of restrictions fixes the basis and the scale of the factor space. An obvious assumption is to set $\Sigma_f = I_K$, which defines an orthogonal basis of unit scale. This provides $K(K+1)/2$ restrictions. Further, assuming that every rotation $Q^{-1}\Sigma_f Q^{-1\prime} = I_K$ should not induce a change in the assumption $\Sigma_f = I$ reduces the set of rotation matrices Q to orthogonal matrices, of which $K(K-1)/2$ elements are free. Therefore, to fix the orientation of the basis, we need an additional $K(K-1)/2$ restrictions.

AR56 suggest three possible sets of restrictions to obtain rotational identification, the first two of which are typically used in exploratory factor analysis. The first set of constraints restricts the $K(K-1)/2$ upper diagonal elements of Λ to zero, i.e. $\lambda_{ik} = 0$ for $k > i$ (Geweke and Singleton, 1981; Geweke and Zhou, 1996). The position of the zeros additionally defines factor order. The first factor hence loads only on variable 1, while the second loads on variable 1 and 2 and so on. Factor sign is identified by requiring $\lambda_{kk} > 0$ for $k = 1, \ldots, K$, i.e. diagonal elements to be positive. The second set of constraints restricts Λ to satisfy $\Lambda'\Lambda = D$ with D diagonal (Stock and Watson, 2002). In this case, factor order is determined by requiring that the nonzero diagonal elements in D be distinct and ordered in increasing or decreasing order of magnitude. The last set of constraints requires that $\Lambda'\Sigma_\epsilon^{-1}\Lambda$ is a diagonal matrix, whose nonzero diagonal elements are distinct and ordered in increasing or decreasing order of magnitude.

Other identification constraints assume Σ_f diagonal, albeit without unit scaling. Then, in addition to requiring $\lambda_{ik} = 0$ for $k > i$, the elements $\lambda_{kk} = 1$ for $k = 1, \ldots, K$ are set to unity. This setup is used in e.g. West et al. (2001). Alternatively, Σ_f may be left altogether unconstrained, which yields an unscaled oblique factor basis. For rotational identification we then need K^2 restrictions. Constraining the leading $K \times K$ submatrix of Λ, $\Lambda_K = I_K$ represents a possibility to achieve rotational identification, as suggested in Jöreskog (1979) and applied in Bernanke et al. (2005). In this case, the first K variables are the factors, i.e. they define factor position and factor scale. The restrictions imposed on the leading K rows of Λ may alternatively also be spread across any subset of K rows in Λ. Imposing them on the leading K rows is just a matter of convenience. However, even though these constraints guarantee mode identification, they are not likelihood invariant with respect to variable ordering and may lead to serious estimation issues, see e.g. Millsap (2001), Chan et al. (2018). An illustration is given in Subsection 3. Recent advances therefore suggest to estimate just-identified or unrestricted sparse factor models and recover mode identification by processing the estimation output (Aßmann et al., 2016; Kaufmann and Schumacher, 2019; Despois and Doz, 2023).

Mode-identification in a dynamic factor model[4] needs essentially the same K^2 number of restrictions to identify the basis, scale and orientation of the factor space. As in the static case, identification of factor position and sign requires additional restrictions. The analogue to setting $\Sigma_f = I$ is to set the covariance matrix of the factor innovations to the identity matrix, $\Sigma_\eta = I$. This renders factors conditionally independent. Although dynamic factor processes

[4]Parametric extensions of (1) may include dynamic processes for f_t and ϵ_t, $f_t = \phi_1 f_{t-1} + \cdots + \phi_p f_{t-p} + \eta_t$ and $\epsilon_{it} = \psi_1 \epsilon_{i,t-1} + \cdots + \psi_q \epsilon_{i,t-q} + \upsilon_{it}$, respectively.

are assumed to be diagonal in some applications (Kose et al., 2003; Otrok and Whiteman, 1998), this does not have to be imposed generally in a $K << N$ (sparse) environment (Beyeler and Kaufmann, 2021). A minimum of $K(K-1)/2$ zero restrictions on Λ and a specific sparse structure are necessary for mode identification, see below.

Discussions in confirmatory factor analysis have shown that imposing additional constraints on the factor loading matrix may result in specifications that lack unique mode identification (Jennrich, 1978).[5] These issues translate to high-dimensional settings in confirmatory and exploratory factor analysis as well, where Λ may be sparse or subject to more zero restrictions than necessary for mode identification. In particular, some results show that multiple sparse structures may be the rule rather than the exception. Algina (1980) provided an explanation for the counterfactual example published in Jennrich (1978) and derived first rules concerning mode identification based on a rank condition. Bekker (1986) and Neudecker (1990) expressed the set of constraints placed on the factor loadings matrix as $R\text{vec}(\Lambda) = r$. Bekker (1986) showed that a necessary and sufficient condition for mode identification requires that[6]

$$\text{rk}\begin{pmatrix} D_K^+ \\ R(I_K \otimes \Lambda) \end{pmatrix} = K^2,$$

and Neudecker (1990) proved that

$$\text{rk}\begin{pmatrix} D_N^+(\Lambda \otimes I_N) \\ R \end{pmatrix} = NK$$

is equivalent to Bekker's condition under the assumption $\text{rk}\,(\Lambda) = K$. More recently, Bai and Wang (2014) extend these results to mode identification conditions for dynamic factor

[5] For example, Λ, although identified when re-ordering time series, can be transformed to ΛC (again identified when re-ordering time series), with a different loading structure below the identical first three rows of loadings (Householder transformation):

$$\Lambda = \begin{bmatrix} 1.41 & 0 & 1.12 & 0 \\ 1.42 & -0.79 & 0 & 0 \\ 0.67 & 0.89 & -0.10 & -1.09 \\ -1.21 & 0 & 0 & 0.03 \\ 0.72 & -1.07 & 0.32 & -1.21 \\ 1.63 & -0.81 & 0.31 & 0 \\ 0.81 & -0.34 & 0.23 & -0.41 \\ 1.03 & 0 & 0 & 0 \\ 0.73 & 0.33 & -0.16 & -0.77 \\ -0.30 & -0.75 & 0.63 & 0.37 \end{bmatrix} \rightarrow \Lambda C = \begin{bmatrix} 1.41 & 0 & 1.12 & 0 \\ 1.42 & -0.79 & 0 & 0 \\ 0.67 & 0.89 & -0.10 & -1.09 \\ -0.99 & 0.38 & -0.27 & 0.50 \\ 1.52 & 0.37 & -0.69 & 0.55 \\ 1.67 & -0.73 & 0.26 & 0.09 \\ 0.99 & 0 & 0 & 0 \\ 0.84 & -0.35 & 0.24 & -0.43 \\ 0.76 & 0.38 & -0.21 & -0.70 \\ 0 & -0.20 & 0.24 & 1.05 \end{bmatrix},$$

where $C = I_K - 2vv'$, $v = \text{null}(\Lambda^*)$ and Λ^* collecting the first rows of Λ determined by $\min(i_{max}, k_{max} - 1)$, where i_{max} and k_{max} refer to the largest row and column with a non-zero element on the upper diagonal of Λ (Here $i_{max} = 4$, $k_{max} - 1 = 3$), and

$$v = null\left(\begin{bmatrix} 1.41 & 0 & 1.12 & 0 \\ 1.42 & -0.79 & 0 & 0 \\ 0.67 & 0.89 & -0.10 & -1.09 \end{bmatrix}\right) \rightarrow C = \begin{bmatrix} 0.81 & -0.34 & 0.24 & -0.41 \\ -0.34 & 0.39 & 0.42 & -0.74 \\ 0.24 & 0.42 & 0.70 & 0.52 \\ -0.41 & -0.74 & 0.52 & 0.09 \end{bmatrix}.$$

[6] The matrix D_K^+ is the Moore-Penrose inverse of the duplication matrix D_K, which transforms the half-vectorization of a $K \times K$ symmetric matrix X into its vectorization as $D_K\text{vech}(X) = \text{vec}(X)$.

models, $\Phi(L)f_t = B(L)\eta_t$, where $\Phi(L)$, $B(L)$ represent polynomials in the lag operator. By defining

$$\Phi = [\Phi_1, \ldots, \Phi_p], \tilde{\Phi} = \begin{bmatrix} \Phi_1 \\ \vdots \\ \Phi_p \end{bmatrix}, B = [B_1, \ldots, B_q], \tilde{B} = \begin{bmatrix} B_1 \\ \vdots \\ B_q \end{bmatrix},$$

and expressing restrictions as $R_\Lambda \text{vec}(\Lambda) = r_\Lambda$, $R_\Phi \text{vec}(\Phi) = r_\Phi$, $R_B \text{vec}(B) = r_B$, the rank condition becomes:

$$\text{rk} \begin{pmatrix} D_K^+ \\ R_\Lambda \left(I_K \otimes \Lambda\right) \\ R_\Phi \left[I_K \otimes \tilde{\Phi} - \Phi' \otimes I_K\right] \\ R_B \left[I_K \otimes \tilde{B} - B' \otimes I_k\right] \end{pmatrix} = K^2.$$

Importantly, these rank conditions do not guard against singleton or spurious factor loadings (only one non-zero factor loading in a column), leading to an indeterminacy between factor and idiosyncratic variance.[7]. Hence, fulfilled rank conditions do not imply global identification. Global identification is assumed to hold in a first place.

3 A geometric approach to identification

3.1 Full-rank mode identification: Set conditions

Rank conditions can be derived based on a geometric interpretation of factor models, where Σ_f spans a possibly correlated factor basis[8] and each row λ_i in Λ represents weights attached to basis vectors and corresponds to cartesian coordinates in a K-dimensional space (Lawley and Maxwell, 1971). It is useful to introduce some geometric and topological concepts. Whereas the considered concepts can generally be defined for various fields, we are only interested in real numbers, see e.g. Boothby (2002) and James (1976). Denote as a K-frame a set of K independent column vectors in the $\mathbb{R}^N$ with $K < N$, or, as an $N \times K$ matrix with full column rank. The set of all K-frames in the $\mathbb{R}^N$ is then denoted as the (real) non-compact Stiefel manifold $V(K, N)$.[9] As the K independent column vectors in a K-frame span the K-dimensional (real) vector space $\mathbb{R}^K$, we consider its k-dimensional subspaces $k < K$. The set of all k-dimensional linear subspaces of $\mathbb{R}^K$ defines the (real) Grassmann manifold $Gr(k, K)$.

[7] Assume that factor K has a nonzero loading only for variable i. Then the variance for variable i can be decomposed as

$$\sum_{k=1}^{K-1} \lambda_{ik}^2 \sigma_{fk}^2 + \underbrace{\lambda_{iK}^2}_{\equiv 1} \sigma_{fK}^2 + \sigma_i^2 = \sum_{k=1}^{K-1} \lambda_{ik}^2 \sigma_{fk}^2 + (\sigma_{fK}^2 + c) + (\sigma_i^2 - c)$$

for $c \in (-\sigma_{fK}^2, \sigma_i^2]$. The case $c = -\sigma_{fK}^2$ is inconsistent with the unity constraint and would imply that factor K can be omitted from the analysis. Conversely, the case $c = \sigma_i^2$, implies that all remaining variable i's variation unexplained by the first $K-1$ factors is assigned to factor K.

[8] $\Sigma_f = I_K$ corresponds to an orthonormal factor basis.

[9] Note that we don't require the columns of the K-frame to be orthogonal, as sometimes defined.

For instance, $Gr(1,2)$ is the set of all lines through the origin in a plane. Finally, the set of all orthogonal $K \times K$ matrices is denoted as the (real) orthogonal group $O(K)$, corresponding to an orthogonal factor basis.

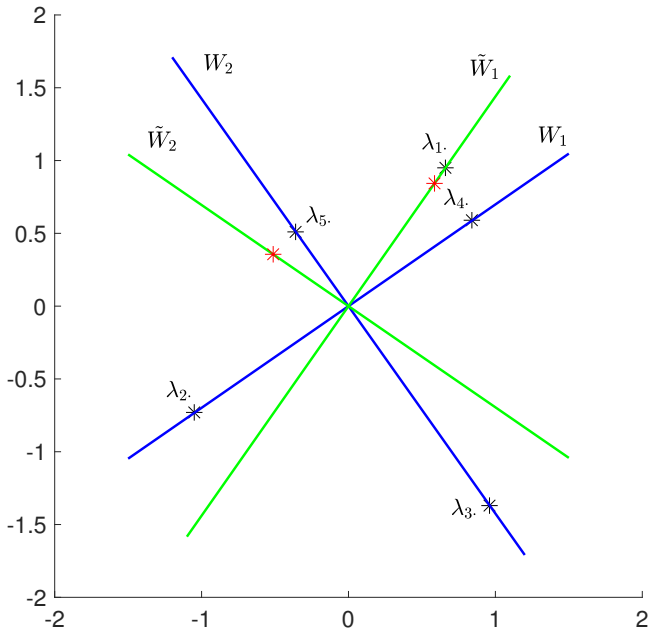

Figure 1: Coordinates representing five rows of two-dimensional factor loadings, of which each of two pairs are located in a 1-dimensional subspace (blue lines). The green lines correspond to a factor basis induced by a rotation into a GLT structure. The red stars indicate factor loadings when additional moderate shrinkage is induced into the GLT structure.

A necessary condition for the column space of Λ to exist is that Λ has full column rank K, such that a subset of rows of Λ, say Λ_a, exists that spans the $\mathbb{R}^K$. Equivalently, defining the vector space W that contains all row vectors of Λ (the K-frame defined above), i.e. $W = \{\lambda_{i\cdot} | i = 1, \ldots, N\}$, it must hold that $W \notin Gr(k, K)$ for $k < K$.

For example,[10] Figure 1 plots the following loading matrices as coordinates:

$$\Lambda = \begin{pmatrix} \Lambda_a \\ \hline \Lambda_b \end{pmatrix} = \left(\begin{array}{rr} 0.66 & 0.95 \\ -1.05 & -0.73 \\ \hline 0.96 & -1.37 \\ 0.84 & 0.59 \\ -0.36 & 0.51 \end{array}\right), \quad \tilde{\Lambda} = \left(\begin{array}{rr} 1.09 & 0.40 \\ -1.28 & 0.00 \\ \hline 0.00 & -1.68 \\ 1.02 & 0.00 \\ 0.00 & 0.63 \end{array}\right), \quad \Lambda^{GLT} = \left(\begin{array}{rr} 1.16 & 0.00 \\ -1.20 & 0.45 \\ \hline -0.58 & -1.57 \\ 0.96 & -0.35 \\ 0.21 & 0.59 \end{array}\right), \quad (5)$$

where coordinates for Λ are specified in terms of the x- and y-axis. Two pairs of row vectors in Λ, are each located in a 1-dimensional subspace, $\{W_1, W_2\} \in Gr(1,2)$ for $W_1 = \{\lambda_{2\cdot}, \lambda_{4\cdot}\}$ and $W_2 = \{\lambda_{3\cdot}, \lambda_{5\cdot}\}$. Both subspaces span an orthogonal factor basis $W_1 \perp W_2$, indicated with blue lines. The sparse loading matrix $\tilde{\Lambda}$ corresponds to the rotated factor basis. The example also

[10] We use the same numerical example as in Kaufmann and Pape (2023)

illustrates the importance of choosing units when setting pre-defined identification restrictions onto the factor loading matrix. Choosing either $\lambda_{2\cdot}$ and $\lambda_{4\cdot}$ or $\lambda_{3\cdot}$ and $\lambda_{5\cdot}$ as leading units Λ_a in Λ combined with identification restrictions such as lower diagonal or diagonal, would fail in identifying a second factor as each set of units is loaded by a single factor only. On the other hand, using the first two rows of Λ to rotate into a (generalized) lower triangular structure (Frühwirth-Schnatter et al., 2023) destroys the sparse structure of the loading matrix.[11] The green lines reflect the rotated factor basis corresponding to Λ^{GLT} in which the factor loading structure has been rotated into a lower triangular structure. These considerations motivate to base inference on order-invariant estimation and identify factors, including their order and sign (rotational identification), after estimation by processing the posterior output (Aßmann et al., 2016; Chan et al., 2018; Kaufmann and Schumacher, 2019).

Definition 1. *Subset enumeration (single index)*

Consider the set $\mathcal{I} = \{1, \ldots, K\}$ with $K \geq 1$. The power set $\mathcal{P}(\mathcal{I})$ contains all 2^K subsets of $\mathcal{I}$, which can be enumerated by $\mathcal{I}_w$, where $w = \sum_{i=1}^{K} 2^{i-1} 1_{\{i \in \mathcal{I}_w\}}$. $\diamond$

Definition 2. *Subset enumeration (double index)*

Consider the set $\mathcal{I} = \{1, \ldots, K\}$ with $K \geq 1$. The power set $\mathcal{P}(\mathcal{I})$ contains all 2^K subsets of $\mathcal{I}$, with $\binom{K}{k}$ subsets of dimension k for $0 \leq k \leq K$. To distinguish between subsets of different sizes, use index k to indicate the size of the subset and index l to enumerate the subsets of the same size. Hence, for every $\mathcal{I}_{k,l} \subseteq \{1, \ldots, K\}$ it holds that $|\mathcal{I}_{k,l}| = k$ and $\psi(\mathcal{I}_{k,l}) = \sum_{i=1}^{K} 2^{K-i} 1_{\{i \in \mathcal{I}_{k,l}\}}$, such that $\psi(\mathcal{I}_{k,l}) > \psi(\mathcal{I}_{k,h})$ for every $l < h$. $\diamond$

For instance, let $\mathcal{I} = \{1, 2, 3\}$. Then there exist three two-elemental subsets of $\mathcal{I}$, namely $\mathcal{I}_3 = \{1, 2\}$ with $2^{1-1} + 2^{2-1} = 1 + 2 = 3$, $\mathcal{I}_5 = \{1, 3\}$ with $2^{1-1} + 2^{3-1} = 1 + 4 = 5$ and $\mathcal{I}_6 = \{2, 3\}$ with $2^{2-1} + 2^{3-1} = 2 + 4 = 6$. It holds that $\mathcal{I}_3 = \mathcal{I}_{2,1}$, $\mathcal{I}_5 = \mathcal{I}_{2,2}$ and $\mathcal{I}_6 = \mathcal{I}_{2,3}$, since $|\mathcal{I}_3| = |\mathcal{I}_5| = |\mathcal{I}_6| = 2$, and $\psi(\mathcal{I}_{2,1}) = \psi(\mathcal{I}_3) = 2^{3-1} + 2^{3-2} = 4 + 2 = 6 > \psi(\mathcal{I}_{2,2}) = \psi(\mathcal{I}_5) = 2^{3-1} + 2^{3-3} = 4 + 1 = 5 > \psi(\mathcal{I}_{2,3}) = \psi(\mathcal{I}_6) = 2^{3-2} + 2^{3-3} = 2 + 1 = 3$.

Definition 3. *Subspace representation of an orthogonal factor model*

[11] Given the order non-invariance of the (generalized) lower triangular structure, inducing additional shrinkage for factor loadings below the leading non-zero loading of each factor when estimating the model might additionally bias the inference on the underlying factors and factor loading matrix. For example, inducing moderate shrinkage may pull $\lambda_{4\cdot}$ and $\lambda_{5\cdot}$ to lie on $\tilde{W}_1$ and $\tilde{W}_2$, respectively:

$$\Lambda^{GLT} = \begin{pmatrix} 1.16 & 0.00 \\ -1.20 & 0.45 \\ \hline -0.58 & -1.57 \\ 0.96 & -0.35 \\ 0.21 & 0.59 \end{pmatrix}, \Lambda^{GLT,shrink} = \begin{pmatrix} 1.16 & 0.00 \\ -1.20 & 0.45 \\ \hline -0.58 & -1.57 \\ 1.03 & 0.00 \\ 0.00 & 0.62 \end{pmatrix}, \tag{6}$$

see the coordinates in red in Figure 1. Obviously, this procedure blurs factor interpretation, biases the strong correlation across either pair of units 2 and 4 or 3 and 5, and overemphasizes the importance of Factors 1 and 2 for, respectively, Units 4 and 5.

For a general subspace representation of an orthogonal factor model with K factors, let

$$W_k \in Gr(1,K),\ W_{k_i} \perp W_{k_j},\ k_i \neq k_j,\ k_i, k_j \in \{1,\ldots,K\},$$

and

$$W_{\mathcal{I}_{k,l}} = \bigcup_{j \in \mathcal{I}_{k,l}} W_j \quad \text{and} \quad W_{\mathcal{I}_w} = \bigcup_{j \in \mathcal{I}_w} W_j.$$

Note that for $\mathcal{I}_{K,1} = \mathcal{I}_w$ with $w = 2^K - 1$, $W_{\mathcal{I}_{K,1}} = \bigcup_{j=1}^{K} W_j$ spans the $\mathbb{R}^K$. $\diamond$

Recall from Equation (4) that Λ has $NK - \frac{K(K-1)}{2}$ distinctly identified parameters. Assume a dense Λ, mode-identified by a lower diagonal structure. We obtain $\lambda_{i\cdot} \in W_{\mathcal{I}_{i,1}}$ for $i \in \{1,\ldots,K-1\}$, and $\lambda_i \in W_{\mathcal{I}_{K,1}}$ for all $i \geq K$. The equivalence class can be obtained by $P_N \Lambda P_K H$, where P_N and P_K are N-dimensional row- and K-dimensional column-permutation matrices, respectively, and $H \in O(K)$ is a rotation matrix. Sparse loading matrices, on the other hand, imply $\lambda_i \in W_{\mathcal{I}_w}$, where $0 \leq |\mathcal{I}_w| < K$ for (almost) every $i \in \{1,\ldots,N\}$.

A special case obtains if $\{\lambda_i\}_{i=1}^N$ can be partitioned into K subsets where the k^{th} subset $\{\lambda_{i_k\cdot}\}_{i_k=1}^{N_k} \in W_k$. This is the congeneric factor model of Jöreskog (1971) with orthogonal factors. If the factors are allowed to be correlated, the constraint $W_{k_i} \perp W_{k_j}$ can be replaced by $W_{k_i} \neq W_{k_j}$ for all $k_i \neq k_j$, $k_i, k_j \in \{1,\ldots,K\}$, which renders the model a lot more flexible.

For a generic (sparse) loadings structure, the required rank condition and rotational identification can be checked by means of the subspaces in Definition 3. We assign each row vector $\{\lambda_i\}_{i=1}^N$ to the set $\mathcal{I}_w$ which corresponds to its non-zero loadings, $\lambda_i \in W_{\mathcal{I}_w}$.[12] We denote by N_w the number of rows $\lambda_{i\cdot}$ assigned to $\mathcal{I}_w$, $\{\lambda_{i_w\cdot}\}_{i_w=1}^{N_w} \in W_{\mathcal{I}_w}$.

Full rank and rotation identification conditions for Λ can be defined in terms of set conditions. In particular, full rank identification does not require all index sets to be non-empty. Rather, it suffices that at least K subsets be non-empty and span the K-dimensional space. Rotational identification can be based on K elements, each one taken from a different set. We formulate two conditions.

Definition 4. Set identification: For a (sparse) factor loading matrix Λ, assign each row $\lambda_{i\cdot}$ to the set $\mathcal{I}_w$ corresponding exactly to the non-zero factor loadings, $\lambda_{i\cdot} \in W_{\mathcal{I}_w}$. Λ is set identified, if at least K sets are non-empty and the union over these sets spans the $\mathbb{R}^K$ space, $\bigcup_{w|N_w>0} \mathcal{I}_w = \{1,\ldots,K\}$.

Condition 1. *Full rank set identification: A set identified (sparse) factor loading matrix Λ has full column rank, $\Lambda \in V(K,N)$, if K rows of factor loadings, $\lambda_{i_k\cdot}$, $k = 1,\ldots,K$, each assigned to a different set, $\mathcal{I}_{w_{i_k}} \neq \mathcal{I}_{w_{i_l}}$, $k \neq l$, are set identified and form a full-rank matrix.*

[12] As we are interested in characterizing the sparse structure of λ, we assign each row to exactly one $\mathcal{I}_w$, the one that corresponds uniquely to the non-zero loadings. Note that generically, every $\lambda_{i\cdot} \in W_{\mathcal{I}_w}$ is also element of $W_{\mathcal{I}_x}$, for every $\mathcal{I}_x \supseteq \mathcal{I}_w$.

Condition 2. *Set-based mode-identification: A (sparse) factor loading matrix is mode-identified, if K rows of factor loadings, $\lambda_{i_k \cdot}$, $k = 1, \ldots, K$, each assigned to a different set, $\mathcal{I}_{w_{i_k}} \neq \mathcal{I}_{w_{i_l}}$, $k \neq l$, are full rank set identified and include at least $K(K-1)/2$ zero loading restrictions.*

Remark 1. Set identification is not satisfied if there are two or more columns in Λ with the same structure of non-zero loadings. In this case, applying a QR decomposition to these columns reveals whether they have full rank. If they do, the QR decomposition induces a GLT structure with additional zero constraints, while preserving the loading structure in the remaining columns.

Remark 2. Mode identification is also obtained upon re-ordering appropriately the K elements chosen to evaluate full rank set identification as leading units of the data set. Λ may then be rotated into a GLT structure, inducing at least $K(K-1)/2$ zero restrictions required for mode identification.

Remark 3. Generally, more than one mode-identified sparse solution may underlie empirical data. In these cases, the mode of interest will be determined by e.g. the most sparse representation or the factors and their interpretation of interest. In this sense, set-based mode identification encompasses multiple mode-identified solutions (Kaufmann and Pape, 2023).

The matrix Λ in the example given by Jennrich (1978), see also Footnote 5, can be used to illustrate Remarks 2 and 3. Both matrices, the original matrix Λ and the Householder-transformed matrix ΛC, can be re-ordered by suitable permutation matrices P_{N_1} and P_{N_2}, respectively, to obtain loading matrices mode-identified by GLT and PLT, respectively:

$$P_1\Lambda = \begin{pmatrix} 1.03 & 0 & 0 & 0 \\ 1.42 & -0.79 & 0 & 0 \\ 1.41 & 0 & 1.12 & 0 \\ 1.63 & -0.81 & 0.31 & 0 \\ 0.67 & 0.89 & -0.10 & -1.09 \\ -1.21 & 0 & 0 & 0.03 \\ 0.72 & -1.07 & 0.32 & -1.21 \\ 0.81 & -0.34 & 0.23 & -0.41 \\ 0.73 & 0.33 & -0.16 & -0.77 \\ -0.30 & -0.75 & 0.63 & 0.37 \end{pmatrix}, \; P_{N_2}\Lambda C = \begin{pmatrix} 0.99 & 0 & 0 & 0 \\ 1.42 & -0.79 & 0 & 0 \\ 1.41 & 0 & 1.12 & 0 \\ 0.67 & 0.89 & -0.10 & -1.09 \\ -0.99 & 0.38 & -0.27 & 0.50 \\ 1.52 & 0.37 & -0.69 & 0.55 \\ 1.67 & -0.73 & 0.26 & 0.09 \\ 0.84 & -0.35 & 0.24 & -0.43 \\ 0.76 & 0.38 & -0.21 & -0.70 \\ 0 & -0.20 & 0.24 & 1.05 \end{pmatrix}.$$

3.2 Global identification: Sato's $\mathcal{O}(K^2)$ algorithm

Full rank set identification of Λ does not guarantee global identification (see Subsection 2.2), and we may refer to Sato (1992)'s counting rule to evaluate whether this necessary condition is fulfilled. Based on the sets $|\mathcal{I}_w| > 0$, we may re-frame the counting rule as a set-based condition.

Condition 3. *Sato's set-based counting rule: A necessary condition for a mode-identified (sparse) factor loading matrix to satisfy the condition of Theorem 5.1 in AR56 is that for all non-empty sets $\mathcal{I}_w$, the following condition holds:*

$$\sum_{x|\mathcal{I}_x \supseteq \mathcal{I}_w} N_x + \sum_{x|\mathcal{I}_x \subset \mathcal{I}_w} N_x \geq 2|\mathcal{I}_w| + 1, \; \textit{for all } \mathcal{I}_w \textit{ with } |\mathcal{I}_w| > 0,$$

where N_x is the number of elements assigned to $\mathcal{I}_x$.

With an increasing number of factors, a full evaluation of Condition 3 becomes computationally expensive, as the number of sets to evaluate, $2^K - 1$, grows exponentially in K. In the following, we propose two algorithms to evaluate Condition 3, which both need a considerably lower number of evaluations. The outcome of the algorithms, although not sufficient, is necessary for Condition 3 to hold. Nevertheless, as demonstrated by simulation, both algorithms detect most matrices that violate the counting rule.

The first algorithm, Sato's $\mathcal{O}(K^2)$ algorithm, needs a number of evaluations that grows at most by K^2 rather than 2^K.

Algorithm 1. *Sato's $\mathcal{O}(K^2)$ algorithm:*

1. Set $k = K$, the number of columns in Λ.
2. Delete zero rows from Λ, and determine the number of rows of Λ, which is n.

   If $n < 2k + 1$, the counting rule is violated, and the algorithm returns an error (stops).
3. For all $j = 1, \ldots, k$,

   (a) count the number of nonzero elements in the j^{th} column of Λ, and denote it as $c_j = \sum_{w|I_w \supseteq \{j\}} N_w$. If any $c_j < 3$, the counting rule is violated, and the algorithm returns an error (stops).

   (b) same as (a), and determine $N_{\mathcal{I}_{1,j}}$.
4. (a) Determine $\mathcal{J} = \{j^* | c_{j^*} = \max_j(c_j)\}$.

   (b) Determine $\mathcal{J} = \{j^* | N_{\mathcal{I}_{1,j^*}} = \max_j(N_{\mathcal{I}_{1,j}})\}$.

   If $|\mathcal{J}| = 1$, choose $j^* \in \mathcal{J}$. Otherwise, if $|\mathcal{J}| > 1$, choose $j^* \in \mathcal{J}$ at random. If $|\mathcal{J}| = 0$, choose $j^* \in \{1, \ldots, k\}$ at random.

   Delete column j^* from Λ.
5. Set $k := k - 1$.

   If $k \geq 1$, proceed with step 2.

   If $k = 0$, Λ most probably satisfies the counting rule.

Remark 4. The algorithm includes two variants, a) and b) in Steps 3. and 4.. In variant a) the column with a maximum number of non-zero loadings is eliminated, while in variant b) the column with the maximum number of exclusive loadings is discarded.

The second algorithm relies in the first place on the evaluation of populated non-empty sets, $\mathcal{I}_w$ with $N_w > 0$, and included subsets, $\mathcal{I}_v \subseteq \mathcal{I}_w$. In particular for matrices of increasing factor dimension with a large degree of sparsity, the number of populated non-empty sets is potentially quite lower than $2^K - 1$. Focusing on these considerably accelerates the evaluation of the counting rule. We formulate the following condition:

Condition 4. *Sato's reduced set-based counting rule: A necessary condition for a mode-identified (sparse) factor loading matrix to satisfy the condition of Theorem 5.1 in AR56 is that for all populated non-empty sets $\mathcal{I}_w$ with $N_w > 0$, the following condition holds:*

$$\sum_{x|\mathcal{I}_x \supseteq \mathcal{I}_v} N_x + \sum_{x|\mathcal{I}_x \subset \mathcal{I}_v} N_x \geq 2|\mathcal{I}_v| + 1, \text{ for all } \mathcal{I}_v \subseteq \mathcal{I}_w \text{ with } |\mathcal{I}_w| > 0 \text{ and } N_w > 0$$

where N_x is the number of elements assigned to $\mathcal{I}_x$.

Remark 5. Condition 4 is equivalent to Condition 3 if the set $\{1, \ldots, K\}$ is populated.

The second algorithm uses Condition 4.

Algorithm 2. *Sato's set-based algorithm:*

1. Set $k = K$, the number of columns in Λ.
2. Delete zero rows from Λ, and determine the number of rows of Λ, which is n.

   If $n < 2k + 1$, the counting rule is violated, and the algorithm returns an error (stops).
3. Determine the populated non-empty sets, $\mathcal{I}_w$ with $N_w > 0$ and evaluate the reduced set-based Condition 4.

   If Condition 4 does not hold, the counting rule is violated, and the algorithm returns an error (stops).
4. Determine $\mathcal{J} = \{j | j \in \mathcal{I}_w = \mathcal{I}_{1,j}, N_w > 0\}$. If $|\mathcal{J}| = 0$, set $\mathcal{J} = \{j | j = 1, \ldots, K\}$.

   Determine $\mathcal{J}^* = \left\{ j^* | c_{j^*} = \max_j(c_j), c_j = \sum_{w|I_w \supseteq \{j\}} N_w, j \in \mathcal{J} \right\}$.

   If $|\mathcal{J}^*| = 1$, choose $j^* \in \mathcal{J}$, the column with the maximum number of non-zero loadings, and delete column j^* from Λ.

   If $|\mathcal{J}^*| > 1$, choose (Factor) j^+ with the least number of cross-loaded units, i.e. units loaded by other factors, $\mathcal{J}^+ = \left\{ j^+ | j^+ = \min_{j^*} \left(\sum_{w|\mathcal{I}_w \supset \{j^*\}} N_w \right), j^* \in \mathcal{J}^* \right\}$. If $|\mathcal{J}^+| > 1$ choose j^+ randomly. Delete column j^+ from Λ.
5. Set $k := k - 1$.

   If $k \geq 1$, proceed with step 2.

   If $k = 0$, Λ most probably satisfies the counting rule.

4 Efficiency and illustration

In this section, we evaluate the efficiency of the proposed algorithm by simulation and present an empirical application to illustrate the set-based identification procedure.

4.1 Efficiency

Table 1 documents the run-time efficiency of Sato's $\mathcal{O}(K^2)$ algorithm relative to the evaluation of all sets, as required in Condition 3. We generate 1,000 sparse matrices of different dimensions $3K \times K$ for $K \in \{3, \ldots, 16\}$. Table 1 shows that while for small K, $K \leq 5$, full evaluation is faster than the $\mathcal{O}(K^2)$ algorithm, runtime quickly increases exponentially with increasing dimension. Processing a large number of posterior draws of (sparse) factor loading matrices (e.g. Kaufmann and Pape (2023)) may thus become computationally very costly, if not almost infeasible, for very large models. Both proposed algorithms provide a procedure to circumvent the exponential increase in runtime.

Table 1: Runtimes (in seconds) for evaluating 1,000 matrices of dimension $3K \times K$. Evaluation of all sets (Condition 3), and based on Sato's $\mathcal{O}(K^2)$ algorithm, variant a).

K	Condition 3	$\mathcal{O}(K^2)$ algorithm
3	0.0694	0.1583
4	0.1013	0.1629
5	0.1937	0.2004
6	0.3805	0.2418
7	0.6980	0.2660
8	1.5336	0.3441
9	2.8039	0.3502
10	5.2279	0.3720
11	10.0200	0.3851
12	20.1041	0.4257
13	39.1514	0.4629
14	80.9280	0.5329
15	185.6150	0.6932
16	389.1633	0.7289

Table 2 provides an overview of the number of matrices falsely identified by the algorithms to satisfy the counting rule, out of 10,000 simulated matrices per scenario. For the scenarios we simulate matrices with $K \in \{3, \ldots, 12\}$ columns and $3K$ rows, and where a varying percentage of them satisfies the counting rule. Note that unlike in Condition 3, the algorithms check only a (small) subset of all $2^K - 1$ sets to consider. They may therefore identify false positives (the counting rule holds, while in fact it does not hold), but never false negatives (the counting rule does not hold, while in fact it does). Scenarios with a large share of matrices satisfying the counting rule leave less room for false positives than scenarios with a low share of matrices satisfying the counting rule. This explains why the number of errors is higher at the bottom of the table than at the top. In the 95% scenarios, the number of false positives can go up to 500, while in the 0% scenarios, all 10,000 matrices can produce false positives. Variant a) of the $\mathcal{O}(K^2)$ algorithm produces the highest number of errors. It is clearly outperformed by variant b) of the $\mathcal{O}(K^2)$ algorithm, which in turn is slightly outperformed by the set-based algorithm. Another interesting observation is that as K increases, the number of false positives also first increases, usually reaching a maximum for values of K between 6 and 8, before falling again

to much lower values for K beyond 10. The proposed algorithms therefore produce more precise results for scenarios where the efficiency gain in terms of computational time is also larger, see Table 1.

Table 2: Out of 10,000 simulated matrices of dimension $3K \times K$, the number of false-identified matrices satisfying the counting rule, as detected by the $\mathcal{O}(K^2)$ algorithm with deletion of most populated columns (A1a, variant a)), the $\mathcal{O}(K^2)$ algorithm with deletion of column representing most populated single-element set (A1b, variant b)), the set-based algorithm (A2).

K	3	4	5	6	7	8	9	10	11	12
95% satisfying the counting rule										
A1a	0	3	2	1	1	0	0	0	0	0
A1b	0	0	0	0	1	0	1	0	0	0
A2	0	1	0	0	0	0	0	0	0	0
90% satisfying the counting rule										
A1a	0	3	2	3	0	5	1	1	0	0
A1b	0	0	1	0	1	0	0	0	0	0
A2	0	0	0	0	0	2	0	0	0	0
75% satisfying the counting rule										
A1a	2	9	2	7	7	7	8	2	2	1
A1b	0	0	1	1	2	3	4	1	1	0
A2	0	1	1	4	2	2	3	2	1	0
50% satisfying the counting rule										
A1a	3	1	16	24	10	11	6	3	2	1
A1b	0	3	2	3	7	4	5	5	4	1
A2	0	0	3	6	7	4	5	1	2	1
25% satisfying the counting rule										
A1a	4	15	13	21	31	12	14	7	3	2
A1b	0	2	3	6	4	5	4	6	1	3
A2	0	2	2	6	10	3	6	3	2	1
10% satisfying the counting rule										
A1a	4	17	21	34	24	20	8	5	7	1
A1b	0	2	5	9	4	11	5	4	8	3
A2	0	2	3	13	10	3	0	3	5	0
0% satisfying the counting rule										
A1a	6	12	21	28	26	23	17	13	5	3
A1b	0	2	6	6	15	7	6	10	4	6
A2	0	2	6	12	8	7	5	7	4	2

We also conclude that an increasing share of draws not fulfilling the counting rule, leads to an increasing number of false positives identified by both algorithms, and eventually suggests that the model is mis-specified, for example overfitting the data.

4.2 Illustration

For the empirical illustration, we analyze monthly log returns in exchange rates for 22 currencies against the euro, covering the period from January 2000 to December 2007, see also Frühwirth-Schnatter et al. (2024). We first analyze a model with $K = 4$ factors. We use the unconstrained rotation sampler (Kaufmann and Pape, 2023) which uncovers two distinct modes for the loading matrix. Each mode is then estimated via confirmatory factor analysis (CFA), and the posterior means of factor loadings are presented in the left two panels of Figure 2. Table 3 lists the number of rows in Λ that are assigned to each of the subsets $\mathcal{I}_w \subseteq \{1, 2, 3, 4\}$. The factor loading matrices of both modes are set-identified, as the number of populated non-empty subsets is $7 > 4 = K$ in each mode, and these sets span the factor space $\bigcup_{w|N_w>0} \mathcal{I}_w = \{1, \ldots, 4\}$.

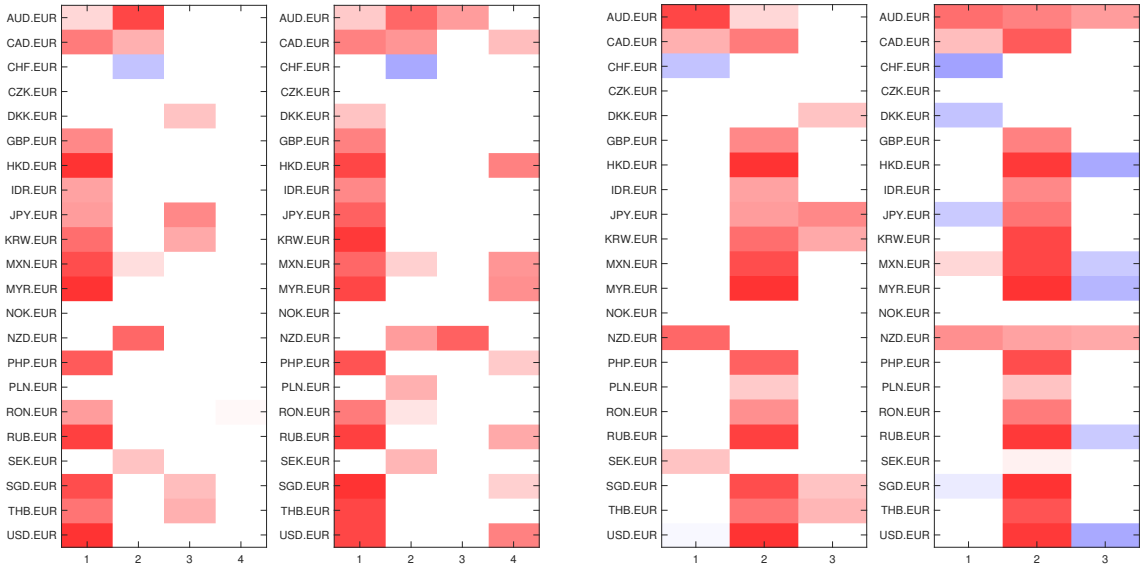

Figure 2: Loading pattern for exchange rates data, $K = 4$ (left two panels) and $K = 3$ (right two panels), two modes each.

When choosing $\lambda_{6,\cdot} \in W_{\mathcal{I}_{1,1}}$ (British pound), $\lambda_{14,\cdot} \in W_{\mathcal{I}_{1,2}}$ (New Zealand dollar), $\lambda_{5,\cdot} \in W_{\mathcal{I}_{1,3}}$ (Danish kroner) and $\lambda_{16,\cdot} \in W_{\mathcal{I}_{1,4}}$ (Polish zloty) in Mode 1 and $\lambda_{6,\cdot} \in W_{\mathcal{I}_{1,1}}$ (British pound), $\lambda_{19,\cdot} \in W_{\mathcal{I}_{1,2}}$ (Swedish kroner), $\lambda_{14,\cdot} \in W_{\mathcal{I}_{2,3}}$ (New Zealand dollar) and $\lambda_{22,\cdot} \in W_{\mathcal{I}_{2,4}}$ (US dollar) in Mode 2 we obtain submatrices of the following form:

$$\Lambda_{a,1} = \begin{pmatrix} * & 0 & 0 & 0 \\ 0 & * & 0 & 0 \\ 0 & 0 & * & 0 \\ 0 & 0 & 0 & * \end{pmatrix} \quad \text{and} \quad \Lambda_{a,2} = \begin{pmatrix} * & 0 & 0 & 0 \\ 0 & * & 0 & 0 \\ 0 & * & * & 0 \\ * & 0 & 0 & * \end{pmatrix}$$

These submatrices are of full-rank and contain at least $4(4-1)/2 = 6$ zero elements. They are both PLT identified. We conclude that both modes are set-based mode-identified.

Table 3: $K = 4$: Mode-specific set population. $N_w(1)$ and $N_w(2)$ report the number of rows in, respectively, Mode 1 and 2, assigned to each set. [a]

set	elements	$N_w(1)$	$N_w(2)$
$\mathcal{I}_0 = \mathcal{I}_{0,1}$	$= \{\}$	2	2
$\mathcal{I}_1 = \mathcal{I}_{1,1}$	$= \{1\}$	7	6
$\mathcal{I}_2 = \mathcal{I}_{1,2}$	$= \{2\}$	3	3
$\mathcal{I}_3 = \mathcal{I}_{2,1}$	$= \{1,2\}$	3	1
$\mathcal{I}_4 = \mathcal{I}_{1,3}$	$= \{3\}$	1	0
$\mathcal{I}_5 = \mathcal{I}_{2,2}$	$= \{1,3\}$	4	0
$\mathcal{I}_6 = \mathcal{I}_{2,3}$	$= \{2,3\}$	0	1
$\mathcal{I}_7 = \mathcal{I}_{3,1}$	$= \{1,2,3\}$	0	1
$\mathcal{I}_8 = \mathcal{I}_{1,4}$	$= \{4\}$	1	0
$\mathcal{I}_9 = \mathcal{I}_{2,4}$	$= \{1,4\}$	1	6
$\mathcal{I}_{10} = \mathcal{I}_{2,5}$	$= \{2,4\}$	0	0
$\mathcal{I}_{11} = \mathcal{I}_{3,2}$	$= \{1,2,4\}$	0	2
$\mathcal{I}_{12} = \mathcal{I}_{2,6}$	$= \{3,4\}$	0	0
$\mathcal{I}_{13} = \mathcal{I}_{3,3}$	$= \{1,3,4\}$	0	0
$\mathcal{I}_{14} = \mathcal{I}_{3,4}$	$= \{2,3,4\}$	0	0
$\mathcal{I}_{15} = \mathcal{I}_{4,1}$	$= \{1,2,3,4\}$	0	0

[a] Note that the loadings on the fourth factor are very small, with 0.024 for the Polish zloty, and 0.055 for the Romanian lei.

Applying Sato's $\mathcal{O}(K^2)$ and the set-based algorithm to the posterior modes plotted in the left two panels of Figure 2, immediately reveals that the counting rule is violated. In the first iteration, while $n = 20 > 9 = 2K + 1$, Step 3 of both algorithms determines $c_4 < 3$, which violates the counting rule.[13]

To improve model specification, we estimated a model with $K = 3$. The unconstrained rotation sampler again uncovers two distinct modes. The posterior means of mode-specific factor loadings, estimated via CFA, are reported in the right two panels of Figure 2. Table 4 lists the number of rows assigned to subsets $\mathcal{I}_w \subseteq \{1,2,3\}$ for each mode. Both modes are set-identified as the number of populated non-empty sets is $5 > 3$ and the sets span the factor space $K = 3$.

Choosing $\lambda_{14,\cdot} \in W_{\mathcal{I}_{1,1}}$ (New Zealand dollar), $\lambda_{6,\cdot} \in W_{\mathcal{I}_{1,2}}$ (British pound) and $\lambda_{5,\cdot} \in W_{\mathcal{I}_{1,3}}$ (Danish kroner) in Mode 1 and $\lambda_{5,\cdot} \in W_{\mathcal{I}_{1,1}}$ (Danish kroner), $\lambda_{6,\cdot} \in W_{\mathcal{I}_{1,2}}$ (British pound) and $\lambda_{22,\cdot} \in W_{\mathcal{I}_{2,3}}$ (US dollar) in Mode 2, we obtain mode-specific submatrices

$$\Lambda_{a,1} = \begin{pmatrix} * & 0 & 0 \\ 0 & * & 0 \\ 0 & 0 & * \end{pmatrix} \qquad \text{and} \qquad \Lambda_{a,2} = \begin{pmatrix} * & 0 & 0 \\ 0 & * & 0 \\ 0 & * & * \end{pmatrix},$$

[13] Note that the algorithms also identify most posterior draws to violate the counting rule.

Table 4: $K = 3$: Mode-specific set population. $N_w(1)$ and $N_w(2)$ denote the number of rows assigned to subsets in, respectively, Mode 1 and 2.

set	elements	$N_w(1)$	$N_w(2)$
$\mathcal{I}_0 = \mathcal{I}_{0,1}$	$= \{\}$	2	2
$\mathcal{I}_1 = \mathcal{I}_{1,1}$	$= \{1\}$	3	2
$\mathcal{I}_2 = \mathcal{I}_{1,2}$	$= \{2\}$	9	8
$\mathcal{I}_3 = \mathcal{I}_{2,1}$	$= \{1,2\}$	3	3
$\mathcal{I}_4 = \mathcal{I}_{1,3}$	$= \{3\}$	1	0
$\mathcal{I}_5 = \mathcal{I}_{2,2}$	$= \{1,3\}$	0	0
$\mathcal{I}_6 = \mathcal{I}_{2,3}$	$= \{2,3\}$	4	4
$\mathcal{I}_7 = \mathcal{I}_{3,1}$	$= \{1,2,3\}$	0	3

which both are of full rank and contain more than 3 zero restricted loadings. The posterior modes are set-based mode-identified.

Sato's $\mathcal{O}(K^2)$ algorithm now goes through three iterations, as conditions for the counting rule to hold are never violated. Figure 3 shows the steps of variant a) of the algorithm. In the first iteration, Factor 2 is eliminated in both modes, and in the next ones remaining Factor 1 is successively eliminated.Likewise, the set-based algorithm successively removes factors in the same order. Tables 5 and 6 summarize the sets considered in each iteration when evaluating, respectively, Modes 1 and 2.

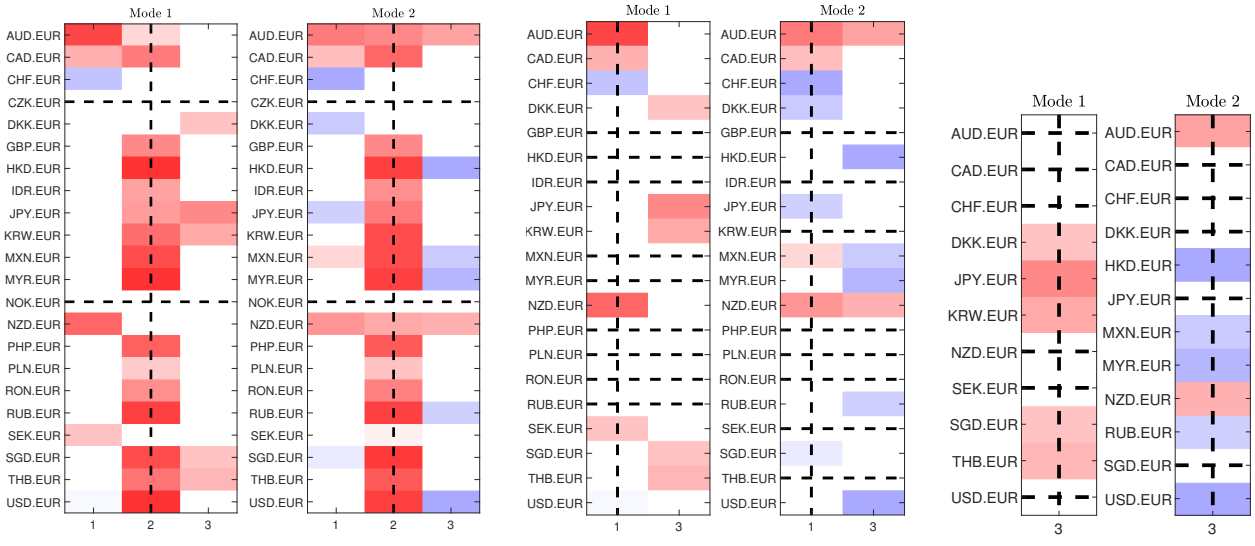

Figure 3: Loading pattern for exchange rates data, $K = 3$, two modes (left panel). Loading pattern remaining after elimination of Factor 2 (middle panel) and Factor 1 (right panel).

Table 5: $K = 3$, Mode 1: Sets evaluated in the set-based algorithm. The factor eliminated is indicated by a *.

Iteration	1					2		3
	$\mathcal{I}_w$					$\mathcal{I}_w$		$\mathcal{I}_w$
$\mathcal{I}_x$	$\{1\}$	$\{2\}^*$	$\{1,2\}$	$\{3\}$	$\{2,3\}$	$\{1\}^*$	$\{2\}$	$\{1\}^*$
$\{1\}$	3		3			6		5
$\{2\}$		9	9		9		5	
$\{1,2\}$	3	3	3		3			
$\{3\}$				1	1			
$\{2,3\}$		4	4	4	4			
$\sum_{x\|\mathcal{I}_x \supseteq \mathcal{I}_w} N_x + \sum_{x\|\mathcal{I}_x \subset \mathcal{I}_w} N_x$	6	16	19	5	17	6	5	5
$2\|\mathcal{I}_w\| + 1$	3	3	5	3	5	3	3	3
	✓	✓	✓	✓	✓	✓	✓	✓

Table 6: $K = 3$, Mode 2: Sets evaluated in the set-based algorithm. The factor eliminated is indicated by a *.

Iteration	1							2			3
	$\mathcal{I}_w$							$\mathcal{I}_w$			$\mathcal{I}_w$
$\mathcal{I}_x$	$\{1\}$	$\{2\}^*$	$\{1,2\}$	$\{2,3\}$	$\{1,2,3\}$	$\{3\}$	$\{1,3\}$	$\{1\}^*$	$\{2\}$	$\{1,2\}$	$\{1\}^*$
$\{1\}$	2		2		2		2	5		5	7
$\{2\}$		8	8	8	8				4	4	
$\{1,2\}$	3	3	3	3	3			3	3	3	
$\{2,3\}$		4	4	4	4	4	4				
$\{1,2,3\}$	3	3	3	3	3	3	3				
$\sum_{x\|\mathcal{I}_x \supseteq \mathcal{I}_w} N_x + \sum_{x\|\mathcal{I}_x \subset \mathcal{I}_w} N_x$	8	18	20	18	20	7	12	8	7	12	7
$2\|\mathcal{I}_w\| + 1$	3	3	5	5	7	3	5	3	3	5	3
	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

5 Conclusion

Following the geometric representation of factor models by Lawley and Maxwell (1971), we present a geometric approach to identification of factor models. We introduce the concept of set identification, and formulate conditions to check whether full-rank, mode identification holds. Set identification is not sufficient for global identification, however. Therefore, we make use of the counting rule proposed by Sato (1992), which we relate to the set identification concept. The number of sets to be checked can be greatly reduced by applying one of the two algorithms we propose for evaluating whether the counting rule holds.

A simulation study reveals that the proposed algorithms correctly identify matrices violating the counting rule in almost all cases and matrices satisfying the counting rule in all cases. Runtimes are greatly reduced, in particular for matrices of large dimensions, i.e. for models with many factors. In an application, we use an exchange rates data set, to which we apply the unconstrained rotation sampler from Kaufmann and Pape (2023) to identify two distinct modes, assuming $K = 4$. While set identification conditions hold, our algorithms indicate a violation of the counting rule, and hence, a misspecified model. We next estimate a model with $K = 3$ and find that set identification conditions as well as the counting rule hold for both modes. We conclude that this model is correctly specified.

References

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

Algina, J. (1980). A Note on Identification in the Oblique and Orthogonal Factor Models. *Psychometrika*, 45(3):393–396.

Anderson, T. and Rubin, H. (1956). *Statistical Inference in Factor Models*, volume 5 of *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, pages 111–150. University of California Press.

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

Bai, J. (2013). Panel data models and factor analysis. In Acemoglu, D., Arellano, M., and Dekel, E., editors, *Advances in Economics and Econometrics*, volume 3, pages 437–483. Cambridge University Press, Cambridge UK.

Bai, J. and Ng, S. (2013). Principal components estimation and identification of static factors. *Journal of Econometrics*, 176:18–29.

Bai, J. and Wang, P. (2014). Identification theory for high dimensional static and dynamic factor models. *Journal of Econometrics*, 178(2):794–804.

Bekker, P. A. (1986). A Note on Identification of Restricted Factor Loading Matrices. *Psychometrika*, 51(4):607–611.

Bekker, P. A. and Berge, J. T. (1997). Generic global identification in factor analysis. *Linear Algebra and its Applications*, 264:255–264.

Bernanke, B. S., Boivin, J., and Eliasz, P. (2005). Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach. *Quarterly Journal of Economics*, 120:387–422.

Beyeler, S. and Kaufmann, S. (2021). Reduced-form factor augmented VAR - Exploiting sparsity to include meaningful factors. *Journal of Applied Econometrics*, 36:989–1012.

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. John Wiley & Sons, New York.

Boothby, W. M. (2002). *An introduction to differentiable manifolds and Riemannian geometry*, volume 120 of *Pure and applied mathematics*. Academic Press, New York, 2nd edition.

Cattell, R. B. (1978). *The Techniques of Simple Structure Rotation*, pages 118–154. Springer US, Boston, MA.

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

Despois, T. and Doz, C. (2023). Identifying and interpreting the factors in factor models via sparsity: Different approaches. *Journal of Applied Econometrics*, 38:533–555.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2005). The Generalized Dynamic Factor Model: One-Sided Estimation and Forecasting. *Journal of the American Statistical Association*, 100:830–840.

Freyaldenhoven, S. (2023). Identification through sparsity in factor models: The l_1-rotation criterion. Working Paper 20-25, Federal Reserve Bank of Philadelphia.

Frühwirth-Schnatter, S., Hosszejni, D., and Freitas Lopes, H. (2023). When it counts – econometric identification of the basic factor model based on glt structures. *Econometrics*, 11:Article 26.

Frühwirth-Schnatter, S., Hosszejni, D., and Lopes, H. F. (2024). Sparse Bayesian factor analysis when the number of factors is unknown. *Bayesian Analysis*, 00:??–??

Geweke, J. and Zhou, G. (1996). Measuring the pricing error of the arbitrage pricing theory. *The Review of Financial Studies*, 9:557–587.

Geweke, J. F. and Singleton, K. J. (1981). Maximum likelihood "confirmatory" factor analysis of economic time series. *International Economic Review*, 22:37–54.

Hoff, P. D. (2007). Model averaging and dimension selection for the singular value decomposition. *Journal of the American Statistical Association*, 102:674–685.

James, I. M. (1976). *The Topology of Stiefel Manifolds*, volume 24 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, UK.

Jennrich, R. I. (1978). Rotational Equivalence of Factor Loading Matrices with Prespecified Values. *Psychometrika*, 43:421–426.

Jöreskog, K. (1969). A General Approach to Confirmatory Maximum Likelihood Factor Analysis. *Psychometrika*, 34(2):183–202.

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, 36(2):109–133.

Jöreskog, K. G. (1979). Author's Addendum to: A General Approach to Confirmatory Factor Analysis. In Magidson, J., editor, *Advances in Factor Analysis and Structural Equation Models*. Abt Books, Cambridge, MA.

Kaufmann, S. and Pape, M. (2023). Bayesian (non-)unique sparse factor modelling. Working Paper 23.04, Study Center Gerzensee.

Kaufmann, S. and Schumacher, C. (2019). Bayesian estimation of sparse dynamic factor models with order-independent and ex-post mode identification. *Journal of Econometrics*, 210:116–134.

Kose, M. A., Otrok, C., and Whiteman, C. H. (2003). International business cycles: World, region, and country-specific factors. *American Economic Review*, 93:1216–1239.

Lawley, D. and Maxwell, A. (1971). *Factor Analysis as a Statistical Method*. Butterworths, London, 2nd edition.

Ledermann, W. (1937). On the Rank of the Reduced Correlational Matrix in Multiple-Factor Analysis. *Psychometrika*, 2(2):85–93.

Lucas, J., Carvalho, C., Wang, Q., Bild, A., Nevins, J., and West, M. (2006). Sparse Statistical Modelling in Gene Expression Genomics. In Do, K. A., Mueller, P., and Vannucci, M., editors, *Bayesian Inference for Gene Expression and Proteomics*. Cambridge University Press, Cambridge UK.

Millsap, R. E. (2001). When Trivial Constraints Are Not Trivial: The Choice of Uniqueness Constraints in Confirmatory Factor Analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 8(1):1–17.

Millsap, R. E. (2011). *Statistical Approaches to Measurement Invariance*. Routledge, New York.

Moench, E., Ng, S., and Potter, S. (2013). Dynamic hierarchical factor models. *Review of Economics and Statistics*, 95:1811–1817.

Mulaik, S. A. (2010). *Foundations of Factor Analysis*. Chapman & Hall, Boca Raton, FL, 2nd edition.

Muthén, B. O. (1991). Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement*, 28(4):338–354.

Neudecker, H. (1990). On the identification of restricted factor loading matrices: An alternative condition. *Journal of Mathematical Psychology*, 34(2):237–241.

Otrok, C. and Whiteman, C. H. (1998). Bayesian leading indicators: Measuring and predicting economic conditions in Iowa. *International Economic Review*, 39:997–1014.

Sato, M. (1992). A study of an identification problem and substitute use of principal components analysis in factor analysis. *Hiroshima Mathematical Journal*, 22:479–524.

Shapiro, A. (1982). Rank reducibility of a symmetric matrix and sampling theory of minimum trace factor analysis. *Psychometrika*, 47:187–199.

Shapiro, A. (1985). Identifiability of factor analysis: some results and open problems. *Linear Algebra and its Applications*, 70:1–7.

Shapiro, A. (1989). Identifiability of factor analysis: Some results and open problems. *Linear Algebra and its Applications*, 125:149.

Stock, J. H. and Watson, M. W. (2002). Forecasting Using Principal Components from a Large Number of Predictors. *Journal of the American Statistical Association*, 97(460):1167–1179.

Thurstone, L. L. (1935). *The Vectors of Mind*. The University of Chicago Press, Chicago.

West, M. (2003). Bayesian Factor Regression Models in the "Large p, Small n" Paradigm. In *Bayesian Statistics 7*, pages 723–732. Oxford University Press.

West, M., Blanchette, C., Dressman, H., Huang, E., Ishida, S., Spang, R., Zuzan, H., Olson, J. A., Marks, J. R., and Nevins, J. R. (2001). Predicting the Clinical Status of Human Breast Cancer by Using Gene Expression Profiles. *Proc. Natl. Acad. Sci. U.S.A.*, 98(20):11462–11467.

Wilson, E. B. and Worcester, J. (1939). The resolution of tests into two general factors. *Proc. Nat. Acad. Sci.*, 25:20–25.

Zou, H., Hastie, T., and Tibshirani, R. (2006). Sparse principal component analysis. *Journal of Computational and Graphical Statistics*, 15:265–286.