

## Article

# When It Counts—Econometric Identification of the Basic Factor Model Based on GLT Structures

Sylvia Frühwirth-Schnatter <sup>1</sup>, Darjus Hosszejni <sup>1,\*</sup> and Hedibert Freitas Lopes <sup>2,3</sup>

<sup>1</sup> Department of Finance, Accounting, and Statistics, WU Vienna University of Economics and Business, 1020 Vienna, Austria; sfruehwi@wu.ac.at

<sup>2</sup> School of Mathematical and Statistical Sciences, Arizona State University, Tempe, AZ 85281, USA

<sup>3</sup> Insper Institute of Education and Research, São Paulo 04546-042, Brazil

\* Correspondence: darjus.hosszejni@wu.ac.at

**Abstract:** Despite the popularity of factor models with simple loading matrices, little attention has been given to formally address the identifiability of these models beyond standard rotation-based identification such as the positive lower triangular (PLT) constraint. To fill this gap, we review the advantages of variance identification in simple factor analysis and introduce the generalized lower triangular (GLT) structures. We show that the GLT assumption is an improvement over PLT without compromise: GLT is also unique but, unlike PLT, a non-restrictive assumption. Furthermore, we provide a simple counting rule for variance identification under GLT structures, and we demonstrate that within this model class, the unknown number of common factors can be recovered in an exploratory factor analysis. Our methodology is illustrated for simulated data in the context of post-processing posterior draws in sparse Bayesian factor analysis.

**Keywords:** identifiability; sparsity; rank deficiency; rotational invariance; variance identification

**JEL Classification:** C11; C38; C63



**Citation:** Frühwirth-Schnatter, Sylvia; Darjus Hosszejni, and Hedibert Freitas Lopes. 2023. When It Counts—Econometric Identification of the Basic Factor Model Based on GLT Structures. *Econometrics* 11: 26. <https://doi.org/10.3390/econometrics11040026>

Academic Editor: Martin Wagner

Received: 24 February 2023

Revised: 29 October 2023

Accepted: 1 November 2023

Published: 20 November 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Ever since the pioneering work of [Thurstone \(1935, 1947\)](#), factor analysis has been a popular method to model the covariance matrix  $\Omega$  of correlated, multivariate observations  $\mathbf{y}_t$  of dimension  $m$  (see, e.g., [Anderson \(2003\)](#) for a comprehensive review). Assuming  $r$  uncorrelated factors in a basic factor model, for instance, yields the representation  $\Omega = \Lambda\Lambda^\top + \Sigma_0$ , with a  $m \times r$  factor loading matrix  $\Lambda$  and a diagonal matrix  $\Sigma_0$ . This considerable reduction in the number of parameters compared to the  $m(m+1)/2$  parameters of an unconstrained covariance matrix is the main motivation for applying factor models to covariance estimation, especially if  $m$  is large (see, among many others, [Fan et al. \(2008\)](#)). In addition, shrinkage estimation has been shown to lead to very efficient covariance estimation (see, for example, [Kastner \(2019\)](#) in Bayesian factor analysis and [Ledoit and Wolf \(2020\)](#) in a non-Bayesian context).

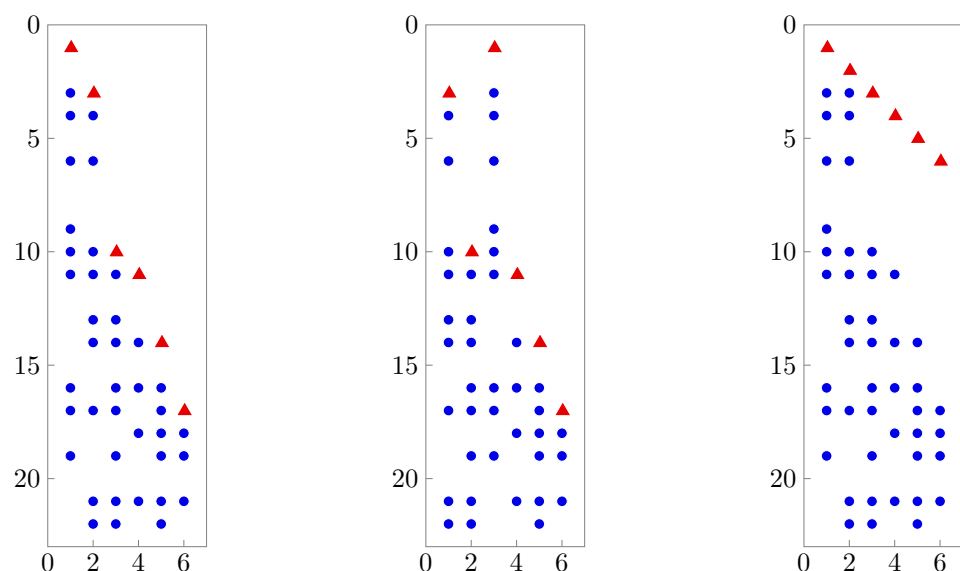
In numerous applications, factor analysis reaches beyond covariance modeling (see, among many others, [Forni et al. \(2009\)](#) in the context of structural factor models). From the very beginning, the goal of factor analysis has been to extract the underlying loading matrix  $\Lambda$  to understand the driving forces behind the observed correlation between the measurements (see, e.g., [Owen and Wang \(2016\)](#) for a recent review). However, also in this setting, the only source of information is the observed covariance of the data, making the decomposition of the covariance matrix  $\Omega$  into the cross-covariance matrix  $\Lambda\Lambda^\top$  and the variance  $\Sigma_0$  of the idiosyncratic errors more challenging than estimating only  $\Omega$  itself.

A huge amount of literature, dating back to [Koopmans and Reiersøl \(1950\)](#) and [Reiersøl \(1950\)](#), has addressed this problem of identification which can be resolved only by imposing additional structure on the factor model. [Anderson and Rubin \(1956\)](#) consider two kinds of

conditions for identification. The first problem, also known as solving rotational invariance, aims at identifying  $\Lambda$ , assuming that  $\Lambda\Lambda^\top$  is determined uniquely. This problem has received considerable attention in econometrics, statistics, and machine learning. The most popular condition for solving rotational invariance is to consider positive lower triangular (PLT) loading matrices (see, e.g., Geweke and Zhou (1996); Lopes and West (2004); West (2003)), although other strategies have been put forward (see, e.g., Neudecker (1990), Bai and Ng (2013), Aßmann et al. (2016), Chan et al. (2018), and Williams (2020)). In related strands, Anderson et al. (2016) examine the generic identifiability of dense vector autoregressive systems with mixed frequencies, and their theory is also applicable to static factor models.

In the second problem, Anderson and Rubin (1956) consider conditions for variance identification, i.e., unique identification of  $\Lambda\Lambda^\top$  and  $\Sigma_0$  assuming that the covariance matrix  $\Omega = \Lambda\Lambda^\top + \Sigma_0$  arises from a basic factor model. Examples in Bartholomew (1987), for instance, show that two different models could imply the same covariance matrix. Sufficient conditions for ensuring variance identification have received much less attention than the first problem. Conditions include the row-deletion property (Anderson and Rubin 1956) and a simple counting rule for the number of non-zero loading in each column of the factor loading matrix in the context of dedicated factor models (Conti et al. 2014); see also Bekker (1989) for related work.

In this work, we discuss conditions for identification based on ordered and unordered generalized lower triangular (GLT) structures which relax the PLT condition, see Figure 1 for illustration. This concept was first introduced in the unpublished work by Frühwirth-Schnatter and Lopes (2018) as part of an MCMC sampler for sparse Bayesian factor analysis where the number of factors is unknown. In the present paper, GLT structures are given a full and comprehensive mathematical treatment. It will be proven that GLT structures simultaneously address rotational invariance and variance identification in factor models. Variance identification relies on a counting rule for the number of non-zero elements in the loading matrix  $\Lambda$ , which is a sufficient condition that extends the previous work by Sato (1992) and Conti et al. (2014).



**Figure 1.** (Left): simple ordered GLT matrix with six factors and pivot rows  $(l_1, \dots, l_6) = (1, 3, 10, 11, 14, 17)$ . (Center): one of the  $2^6 \cdot 6!$  corresponding simple unordered GLT matrices with pivot rows  $(l_1, \dots, l_6) = (3, 10, 1, 11, 14, 17)$ . (Right): a corresponding simple PLT matrix, i.e., enforced non-zeros on the main diagonal, with pivot rows  $(l_1, \dots, l_6) = (1, 2, 3, 4, 5, 6)$ . Pivot rows are marked by triangles, unconstrained loadings are marked by circles, and zero loadings are left blank.

In addition, we show that unordered GLT structures are useful in exploratory factor analysis where the factor dimension  $r$  is unknown. Identification of the number of factors in applied factor analysis is a notoriously difficult problem for many latent factor models, with considerable ambiguity about which method works best, be it BIC-type criteria for approximate factor models (Bai and Ng 2002), marginal likelihoods for basic factor models (Lopes and West 2004), techniques from Bayesian non-parametrics involving infinite-dimensional factor models (Bhattacharya and Dunson 2011; Legramanti et al. 2020; Ročková and George 2017) or more heuristic procedures for dynamic factor models (Kaufmann and Schumacher 2019). Imposing an unordered GLT structure in exploratory factor analysis allows us to identify the true factor dimension by spotting spurious columns in a possibly overfitting model and to identify the true loading matrix  $\Lambda$  and the matrix  $\Sigma_0$ .

The theoretical results of the present paper are exploited in related work. Relying on ordered and unordered GLT structures, Frühwirth-Schnatter et al. (2023) develop an efficient reversible jump MCMC sampler for sparse Bayesian factor analysis under very general shrinkage priors when the number of factors is unknown and use the counting rules introduced in this paper for postprocessing the posterior draws.

The current paper is structured as follows. Section 2 reviews the role of identification in a basic factor model using illustrative examples. Section 3 introduces ordered and unordered GLT structures, proves identification for simple GLT structures, and shows that any full column-rank unconstrained loading matrix has a unique representation as a GLT matrix. Section 4 addresses variance identification under GLT structures. Section 5 discusses exploratory factor analysis under unordered GLT structures and addresses additional identification problems that arise in a basic factor model when the number of factors is unknown. Section 7 presents illustrative applications to simulated and empirical data. Section 8 concludes.

## 2. The Basic Factor Model

### 2.1. Model Definition

Let  $(\mathbf{y}_1, \dots, \mathbf{y}_T)$  be a sequence of observations, where  $\mathbf{y}_t = (y_{1t}, \dots, y_{mt})^\top$  for  $t = 1, \dots, T$  is a vector of  $m$  measurements assumed to arise from a multivariate normal distribution,  $\mathbf{y}_t \sim \mathcal{N}_m(\mathbf{0}, \Omega)$ , with zero mean and covariance matrix  $\Omega$ . In factor analysis, the correlation among the measurements in  $\mathbf{y}_t$  is assumed to be driven by a latent  $r$ -variate random variable  $\mathbf{f}_t = (f_{1t}, \dots, f_{rt})^\top$ , the so-called common factors, through the following observation equation:

$$\mathbf{y}_t = \Lambda \mathbf{f}_t + \epsilon_t, \quad (1)$$

where the  $m \times r$  matrix  $\Lambda$  containing the factor loadings  $\Lambda_{ij}$  is of full column rank,  $\text{rk}(\Lambda) = r$ , equal to the factor dimension  $r$ .

In the present paper, we focus on the so-called basic factor model where the observations  $(\mathbf{y}_1, \dots, \mathbf{y}_T)$  are assumed to be iid. Furthermore, the vector  $\epsilon_t = (\epsilon_{1t}, \dots, \epsilon_{mt})^\top$  accounts for the independent, idiosyncratic variation in each measurement and is distributed as  $\epsilon_t \sim \mathcal{N}_m(\mathbf{0}, \Sigma_0)$ , with  $\Sigma_0 = \text{Diag}(\sigma_1^2, \dots, \sigma_m^2)$  being a positive definite diagonal matrix. The common factors are orthogonal, meaning that  $\mathbf{f}_t \sim \mathcal{N}_r(\mathbf{0}, \mathbf{I}_r)$ , and independent of  $\epsilon_t$ . In this case, the observation Equation (1) implies the following covariance matrix  $\Omega$ , when we integrate with respect to the latent common factors  $\mathbf{f}_t$ :

$$\Omega = \Lambda \Lambda^\top + \Sigma_0. \quad (2)$$

Finally, the model assumes independence of  $\mathbf{f}_t$  and  $\mathbf{f}_s$ ,  $\mathbf{f}_t$  and  $\epsilon_s$ , and  $\epsilon_t$  and  $\epsilon_s$  for all  $s \neq t$ . All cross-sectional dependence among the  $m$  measurements in  $\mathbf{y}_t$  is explained through the latent common factors and the off-diagonal elements of  $\Lambda \Lambda^\top$  define the marginal covariance between any two measurements  $y_{i_1,t}$  and  $y_{i_2,t}$ :

$$\text{Cov}(y_{i_1,t}, y_{i_2,t}) = \Lambda_{i_1,\bullet} \Lambda_{i_2,\bullet}^\top, \quad (3)$$

where  $\Lambda_{i,\bullet}$  is the  $i$ th row of  $\Lambda$ . Consequently, we will refer to  $\Lambda\Lambda^\top$  as the cross-covariance matrix. Since the number of factors,  $r$ , is often considerably smaller than the number of measurements,  $m$ , (2) can be seen as a parsimonious representation of the dependence between the measurements, with considerably fewer parameters in the factor loading matrix  $\Lambda$  than the  $m(m-1)/2$  off-diagonal elements in an unconstrained covariance matrix  $\Omega$ .

## 2.2. Loading Matrices with a Simple Structure

The factor loading matrices given in Figure 1 are examples of simple structures in the sense specified by [Thurstone \(1947\)](#), where each variable loads only on specific factors and factors affect only a subset of variables. In traditional factor analysis, after estimating an identified matrix of loadings, factors are rotated so as to enhance a simple structure, the most popular procedure being Varimax (see, e.g., [Magnus and Neudecker \(2019, sct. 17.14\)](#)). In sparse Bayesian factor analysis, which will be discussed in more detail in Section 6.1, priors are chosen that encourage *automatic rotation* to a simple structure (see, e.g., [Ročková and George \(2017\)](#)).

Subsequently, we will make use of a representation of a factor loading matrix  $\Lambda$  with a simple structure called the *sparsity matrix*. The sparsity matrix  $\delta$  is a binary indicator matrix of 0 s and 1 s of the same dimension as  $\Lambda$  and contains the information which elements of a factor loading matrix with a simple structure are equal to 0 and which elements are unconstrained, i.e., if  $\delta_{ij} = 0$ , then  $\Lambda_{ij} = 0$ , while  $\Lambda_{ij} \in \mathbb{R}$  if  $\delta_{ij} = 1$ .

The sparsity matrix  $\delta$  contains a lot of information about the structure of  $\Lambda$  (see the left-hand matrix in Figure 1 for illustration). In total, six factors are needed to fully explain the variation in these 22 measurements. The fifth row contains only zeros, which tells us that measurement  $y_{5t}$  is uncorrelated with the remaining measurements, since  $\text{Cov}(y_{it}, y_{5t}) = 0$  for all  $i \neq 5$ . The measurements 1, 3, 10, 11, 14, and 17 each constitute a new factor. Hence, the variation in the first nine measurements can be explained by at most two factors which also load on some (but not all) of the remaining measurements. The rows which constitute a new factor will play an instrumental role for identification in the present paper. More specifically, the sparsity matrix on the left hand side tells us that the underlying loading matrix  $\Lambda$  has an ordered GLT structure, while the matrix in the middle is one of many unordered GLT structures that can be derived from  $\Lambda$ . The sparsity matrix on the right hand side indicates that the loading matrix  $\Lambda$  has the commonly applied PLT structure where the first six measurements lead the six factors. Ordered and unordered GLT structures will be discussed in full details in Section 3.

## 2.3. A Brief Review of Identification When the Number of Factors Is Known

Since the factors  $\mathbf{f}_t$  are unobserved, the only information available to estimate  $\Lambda$  and  $\Sigma_0$  is the covariance matrix  $\Omega$ , which creates well-known identification issues for the basic factor model. Consider, for example, the following factor loading matrix  $\Lambda$  and all loading matrices  $\beta = \Lambda\mathbf{P}_{ab}$  defined as a rotation of  $\Lambda$ :

$$\Lambda = \begin{pmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ \lambda_{41} & \lambda_{42} \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{pmatrix}, \quad \mathbf{P}_{ab} = \begin{pmatrix} \cos \alpha & (-1)^b \sin \alpha \\ -\sin \alpha & (-1)^b \cos \alpha \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_{11} & \beta_{21} \\ \beta_{21} & \beta_{22} \\ \beta_{31} & \beta_{32} \\ \beta_{41} & \beta_{42} \\ \beta_{51} & \beta_{52} \\ \beta_{61} & \beta_{62} \end{pmatrix}. \quad (4)$$

Then, for any  $\alpha \in [0, 2\pi)$  and  $b \in \{0, 1\}$ , the factor loading matrix  $\beta$  yields the same cross-covariance matrix for  $\mathbf{y}_t$  as  $\Lambda$ :

$$\beta\beta^\top = \Lambda\mathbf{P}_{ab}\mathbf{P}_{ab}^\top\Lambda^\top = \Lambda\Lambda^\top. \quad (5)$$

The rotational invariance apparent in (5) holds for any basic factor model (1), as is easily verified. Take any  $r \times r$  rotation matrix  $\mathbf{P}$  (i.e.,  $\mathbf{P}\mathbf{P}^\top = \mathbf{I}_r$ ) and define the basic factor model

$$\mathbf{f}_t^* \sim \mathcal{N}_r(\mathbf{0}, \mathbf{I}_r), \quad \mathbf{y}_t = \boldsymbol{\beta} \mathbf{f}_t^* + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}_m(\mathbf{0}, \boldsymbol{\Sigma}_0), \quad (6)$$

where  $\boldsymbol{\beta} = \boldsymbol{\Lambda}\mathbf{P}$  and  $\mathbf{f}_t^* = \mathbf{P}^\top \mathbf{f}_t$ . Obviously, both models imply the same covariance  $\boldsymbol{\Omega}$ , given by (2). Hence, without imposing further conditions,  $\boldsymbol{\Lambda}$  is in general not identified from the cross-covariance matrix  $\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top$ . The usual way of dealing with this rotational invariance is to constrain  $\boldsymbol{\Lambda}$  in such a way that the only possible rotation is the identity  $\mathbf{P} = \mathbf{I}_r$ . For orthogonal factors, at least  $r(r-1)/2$  restrictions on the elements of  $\boldsymbol{\Lambda}$  are needed to eliminate rotational indeterminacy (Anderson and Rubin 1956).

As Anderson and Rubin (1956) point out, additional insight can be gained by comparing the number of parameters in  $\boldsymbol{\Omega}$ , namely  $m(m+1)/2$ , with the number of parameters in the pair  $(\boldsymbol{\Lambda}, \boldsymbol{\Sigma}_0)$ , namely  $m(r+1)$ . If  $r(r-1)/2$  restrictions are imposed to eliminate rotational indeterminacy, this yields the necessary condition  $(m-r)^2 > m+r$  (see also Anderson (2003, sct. 14.2.2)). However, this condition is by no means sufficient for identification, as will be illustrated in Example (9) below, where  $m = 5$  and  $r = 2$  satisfies this condition, but identifiability fails.

A rigorous approach toward the identification of factor models was first offered by Anderson and Rubin (1956). Assume that a pair of parameters  $(\boldsymbol{\Lambda}, \boldsymbol{\Sigma}_0)$ , where  $\boldsymbol{\Lambda}$  is an  $m \times r$  matrix and  $\boldsymbol{\Sigma}_0$  is a positive definite diagonal matrix, satisfying (2) is given. Let  $(\boldsymbol{\beta}, \boldsymbol{\Sigma})$ , where  $\boldsymbol{\beta}$  is an  $m \times r$  matrix and  $\boldsymbol{\Sigma}$  is a positive definite diagonal matrix, be another pair of parameters satisfying (2). Then,

$$\boldsymbol{\Omega} = \boldsymbol{\beta}\boldsymbol{\beta}^\top + \boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top + \boldsymbol{\Sigma}_0,$$

and both pairs imply the same Gaussian distribution  $\mathbf{y}_t \sim \mathcal{N}_m(\mathbf{0}, \boldsymbol{\Omega})$  for every possible realization  $\mathbf{y}_t$ . For an *identified* model, it would follow that the two pairs of parameters are identical, i.e.,  $\boldsymbol{\beta} = \boldsymbol{\Lambda}$  and  $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$ . However, as discussed above, identification can be achieved for a basic factor model only by imposing conditions on  $\boldsymbol{\Lambda}$  and  $\boldsymbol{\Sigma}_0$ .

Anderson and Rubin (1956) consider two kinds of conditions for identification. First, they consider conditions assuming that  $\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top$  and  $\boldsymbol{\Sigma}_0$  are determined uniquely. According to their Lemma 5.1, any alternative loading matrix  $\boldsymbol{\beta}$  which satisfies  $\boldsymbol{\beta}\boldsymbol{\beta}^\top = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top$  (while  $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$ ) and, consequently, implies the same covariance matrix  $\boldsymbol{\Omega}$  as  $\boldsymbol{\Lambda}$ , is an orthogonal rotation  $\mathbf{P}$  of  $\boldsymbol{\Lambda}$ , i.e.,  $\boldsymbol{\beta} = \boldsymbol{\Lambda}\mathbf{P}$ . Conditions are then imposed on the structure of  $\boldsymbol{\Lambda}$  to solve this rotational invariance problem (see Section 2.4 for details). Second, they consider conditions that ensure that  $\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top$  and  $\boldsymbol{\Sigma}_0$  are, indeed, determined uniquely. In their Theorem 5.1, they formulate a row-deletion property as a sufficient condition to resolve the variance identification problem and ensure unique identification of the variance decomposition in (2) (see Section 2.5 for details). The literature on factor analysis often reduces the identification of factor models to the first problem; however, as we will argue in the present paper, variance identification is equally important, in particular for loading matrices with a simple structure.

#### 2.4. Conditions Resolving Rotational Invariance

Let us assume that variance identification holds and  $\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top$  and  $\boldsymbol{\Sigma}_0$  are determined uniquely from  $\boldsymbol{\Omega}$ . By far, the most popular constraints to deal with rotational invariance are positive lower triangular (PLT) loading matrices, where the upper triangular part is constrained to be zero and the main diagonal elements  $\Lambda_{11}, \dots, \Lambda_{rr}$  of  $\boldsymbol{\Lambda}$  are strictly positive.<sup>1</sup> The corresponding sparsity matrix  $\delta$  also exhibits a PLT structure (see the right-hand side of Figure 1 for illustration).

Despite its popularity, the PLT structure is restrictive, as outlined already by Jöreskog (1969). Let  $\boldsymbol{\beta}\boldsymbol{\beta}^\top$  be an arbitrary cross-covariance matrix with factor loading matrix  $\boldsymbol{\beta}$ . A PLT representation for  $\boldsymbol{\beta}\boldsymbol{\beta}^\top$  is possible if the top  $r$  rows of  $\boldsymbol{\beta}$  are linearly independent,



or, equivalently, if a rotation matrix  $\mathbf{P}$  exists such that  $\beta$  can be rotated into a PLT matrix  $\Lambda = \beta\mathbf{P}$ . However, as example (4) illustrates, this is not necessarily the case. Obviously,  $\Lambda$  is not a PLT matrix, since  $\Lambda_{22} = 0$ . Any of the possible rotations  $\beta = \Lambda\mathbf{P}_{ab}$  have *non-zero* elements above the main diagonal and are not PLT matrices either. This example demonstrates that the PLT representation is restrictive. To circumvent this problem in Example (4), one could reorder the measurements in an appropriate manner. However, in applied factor analysis, such an appropriate ordering is typically not known in advance and the choice of the first  $r$  measurements is an important modeling decision under PLT constraints (see, e.g., Lopes and West (2004) and Carvalho et al. (2008)).

An alternative identification condition, also dating back to Anderson and Rubin (1956), is assuming diagonality of  $\Lambda^\top \Sigma_0^{-1} \Lambda$  and requires the corresponding diagonal elements  $d_1, \dots, d_r$  to be different. In practice, however, weak identifiability might occur if some of the elements  $d_1, \dots, d_r$  are similar, even if not equal. Furthermore, in an attempt to formalize the criteria for simplicity of Thurstone (1947), Reiersøl (1950) put forward a condition which requires a minimum of  $r$  zero factor loadings in each column. However, this condition rules out the existence of a “market” factor that loads on all measurements as for the financial returns in our empirical case study in Section 7.2.

In Section 3, we discuss a new identification strategy to resolve rotational invariance in factor models based on the concept of generalized lower triangular (GLT) structures. Loosely speaking, GLT structures generalize PLT structures by freeing the position of the first non-zero factor loading in each column (see the examples in Figure 1 and the loading matrix  $\Lambda$  in (4)). Consequently, the use of GLT structures for rotational identification does not imply that the first  $r$  rows of the loading matrix are linearly independent, as the use of PLT structures would, but only that there exist  $r$  linearly independent rows in  $\Lambda$ . This weakened condition hints at a wider applicability of GLT structures compared to PLT structures. In particular, we show in Section 3.2 that a unique GLT structure  $\Lambda$  can be identified from any cross-covariance matrix  $\beta\beta^\top$ , provided that variance identification holds and, consequently,  $\beta\beta^\top$  itself is identified. Even if  $\beta\beta^\top$  is obtained from a loading matrix  $\beta$  that does not take the form of a GLT structure, such as the matrix  $\beta$  in (4), we show that an orthogonal matrix  $\mathbf{G}$  exists which represents  $\beta = \Lambda\mathbf{G}^\top$  as a rotation of an ordered GLT structure  $\Lambda$  and defines the unique rotation  $\Lambda = \beta\mathbf{G}$  of  $\beta$  into GLT. Hence, the GLT representation is unrestrictive in the sense of Anderson and Rubin (1956) and Jöreskog (1969), and is a generic way to resolve rotational invariance for any factor loading matrix.

### 2.5. Conditions for Variance Identification

Conditions that resolve rotational invariance typically take variance identification, i.e., identification of  $\Lambda\Lambda^\top$ , for granted (see, e.g., Geweke and Zhou (1996)). Variance identification refers to the problem that the idiosyncratic variances  $\sigma_1^2, \dots, \sigma_m^2$  in  $\Sigma_0$  are identified only from the diagonal elements of  $\Omega$ , as all elements in the cross-variance matrix  $\Lambda\Lambda^\top$  are independent of the  $\sigma_i^2$ s (see again (3)). To achieve variance identification of  $\sigma_i^2$  from  $\Omega_{ii} = \Lambda_{i\bullet}\Lambda_{i\bullet}^\top + \sigma_i^2$ , all factor loadings have to be identified solely from the off-diagonal elements of  $\Omega$ . Variance identification, however, is easily violated, as the following examples illustrate.

Let us consider a factor model with the following loading matrix (also known as a dedicated factor model; see, e.g., Conti et al. 2014):

$$\Lambda = \begin{pmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{pmatrix}. \quad (7)$$

The corresponding covariance matrix  $\Omega$  is given by:

$$\Omega = \begin{pmatrix} \lambda_{11}^2 + \sigma_1^2 & \lambda_{11}\lambda_{21} & \lambda_{11}\lambda_{31} & & & \\ \lambda_{11}\lambda_{21} & \lambda_{21}^2 + \sigma_2^2 & \lambda_{21}\lambda_{31} & & & \\ \lambda_{11}\lambda_{31} & \lambda_{21}\lambda_{31} & \lambda_{31}^2 + \sigma_3^2 & & & \\ & & & \lambda_{42}^2 + \sigma_4^2 & \lambda_{42}\lambda_{52} & \lambda_{42}\lambda_{62} \\ & & & \lambda_{42}\lambda_{52} & \lambda_{52}^2 + \sigma_5^2 & \lambda_{52}\lambda_{62} \\ & & & \lambda_{42}\lambda_{62} & \lambda_{52}\lambda_{62} & \lambda_{62}^2 + \sigma_6^2 \end{pmatrix}. \quad (8)$$

Let us assume that the sparsity matrix  $\delta$  of  $\Lambda$  is given (i.e., we know which elements in (7) are zero), but the specific values of the unconstrained loadings  $(\lambda_{11}, \dots, \lambda_{62})$  are unknown. An interesting question is the following. Knowing  $\Omega$  and  $\delta$ , can the unconstrained loadings  $\lambda_{11}, \dots, \lambda_{62}$  and the idiosyncratic variances  $\sigma_1^2, \dots, \sigma_m^2$  be identified uniquely? Given  $\Omega$ , the three nonzero covariances  $\text{Cov}(y_{1t}, y_{2t}) = \lambda_{11}\lambda_{21}$ ,  $\text{Cov}(y_{1t}, y_{3t}) = \lambda_{11}\lambda_{31}$  and  $\text{Cov}(y_{2t}, y_{3t}) = \lambda_{21}\lambda_{31}$  are available to identify the three factor loadings  $(\lambda_{11}, \lambda_{21}, \lambda_{31})$ . Similarly, the three nonzero covariances  $\text{Cov}(y_{4t}, y_{5t}) = \lambda_{42}\lambda_{52}$ ,  $\text{Cov}(y_{4t}, y_{6t}) = \lambda_{42}\lambda_{62}$  and  $\text{Cov}(y_{5t}, y_{6t}) = \lambda_{52}\lambda_{62}$  are available to identify the factor loadings  $(\lambda_{42}, \lambda_{52}, \lambda_{62})$ . Hence, variance identification holds. However, when we remove the last measurement from the loading factor matrix defined in (7), we obtain

$$\Lambda = \begin{pmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \end{pmatrix}, \quad (9)$$

and the corresponding covariance matrix  $\Omega$  reads:

$$\Omega = \begin{pmatrix} \lambda_{11}^2 + \sigma_1^2 & \lambda_{11}\lambda_{21} & \lambda_{11}\lambda_{31} & & & \\ \lambda_{11}\lambda_{21} & \lambda_{21}^2 + \sigma_2^2 & \lambda_{21}\lambda_{31} & & & \\ \lambda_{11}\lambda_{31} & \lambda_{21}\lambda_{31} & \lambda_{31}^2 + \sigma_3^2 & & & \\ & & & \lambda_{42}^2 + \sigma_4^2 & \lambda_{42}\lambda_{52} & \\ & & & \lambda_{42}\lambda_{52} & \lambda_{52}^2 + \sigma_5^2 & \end{pmatrix}.$$

While the three factor loadings  $(\lambda_{11}, \lambda_{21}, \lambda_{31})$  are still identified from the off-diagonal elements of  $\Omega$  as before, variance identification of  $\sigma_4^2$  and  $\sigma_5^2$  fails. Since  $\text{Cov}(y_{4t}, y_{5t}) = \lambda_{42}\lambda_{52}$  is the only non-zero element that depends on the loadings  $\lambda_{42}$  and  $\lambda_{52}$ , infinitely many different parameters  $(\lambda_{42}, \lambda_{52}, \sigma_4^2, \sigma_5^2)$  imply the same covariance matrix  $\Omega$ .

From these considerations, it is evident that a minimum of three non-zero loadings is necessary in each column to achieve variance identification, a condition which dates back to [Anderson and Rubin \(1956\)](#). At the same time, this condition is not sufficient. For illustration, consider the rotation  $\beta = \Lambda P_{\alpha b}$  of the loading matrix  $\Lambda$  in (9) with  $\alpha \neq \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$ , where all elements of the corresponding sparsity matrix are equal to 1. Although each column of  $\beta$  has six non-zero elements, variance identification obviously does not hold.

For more general loading matrices, variance identification is not as easily verified as for these examples and we rely in the present paper on the *row-deletion property* introduced by [Anderson and Rubin \(1956\)](#).

**Definition 1** (Row-deletion property **AR** ([Anderson and Rubin 1956](#))). An  $m \times r$  factor loading matrix  $\Lambda$  satisfies the row-deletion property if the following condition is satisfied: whenever an arbitrary row is deleted from  $\Lambda$ , two disjoint submatrices of rank  $r$  remain.

[Anderson and Rubin \(1956, Theorem 5.1\)](#) prove that the row-deletion property is a sufficient condition for the identification of  $\Lambda\Lambda^\top$  and  $\Sigma_0$  from the marginal covariance matrix  $\Omega$  given in (2). For any (not necessarily GLT) factor loading matrix  $\Lambda$ , the row-deletion property **AR** can be trivially tested by a step-by-step analysis, where every single row of  $\Lambda$

is sequentially deleted and the two distinct submatrices are determined from examining the remaining matrix (Hayashi and Marcoulides 2006). However, this procedure is inefficient in higher dimensions and simpler conditions for verifying variance identification under the row-deletion property **AR** are warranted.

The literature provides necessary conditions for **AR** that are based on counting the number of non-zero factor loadings in  $\Lambda$ , such as: for every nonsingular  $r$ -dimensional square matrix  $\mathbf{G}$ , the matrix  $\beta = \Lambda\mathbf{G}$  contains *at least three* nonzero factor loadings in each column, and *at least five* nonzero factor loadings in each pair of columns (Anderson and Rubin 1956). Sato (1992, Theorem 3.3) extends these necessary conditions in the following way: every subset of  $1 \leq q \leq r$  columns of  $\beta = \Lambda\mathbf{G}$  contains *at least*  $2q + 1$  nonzero factor loadings for every nonsingular matrix  $\mathbf{G}$ . We call this the 3579 counting rule for obvious reasons.

For illustration, apply the 3579 counting rule to all rotations  $\beta = \Lambda\mathbf{P}_{\alpha b}$  of the loading matrix  $\Lambda$  in Example (7) for which we already verified variance identification. We find that the counting rules are satisfied for all rotations  $\beta$ , and one might wonder if the 3579 counting rule can also lead to a *sufficient* criterion for variance identification under **AR**.

Sufficiency of counting rules was proven by Conti et al. (2014) in the context of a dedicated factor model, where the factor loading matrix  $\Lambda$  has a perfect simple structure, i.e., each measurement  $y_{it}$  loads on at most one factor, as in (7) and (9); however, the sparsity matrix  $\delta$ , i.e., the exact position of the non-zero elements, is unknown. Conti et al. (2014) consider a dedicated factor model with correlated factors,  $\mathbf{f}_t \sim \mathcal{N}_r(\mathbf{0}, \mathbf{R})$ , and prove that the following conditions are both necessary and sufficient for uniqueness of the variance decomposition: the correlation matrix  $\mathbf{R}$  is of full rank,  $\text{rk}(\mathbf{R}) = r$ , and each column of  $\Lambda$  contains at least three nonzero loadings.

In the present paper, we aim for simple structures with potentially more than one non-zero loading in each row and generalize this work to basic factor models with orthogonal factors. We impose a GLT condition on the loading matrices  $\Lambda$  and provide sufficient conditions for variance identification in Section 4. These conditions are formulated as counting rules for the sparsity matrix  $\delta$  of  $\Lambda$ . More specifically, if the 3579 counting rule of Sato (1992, Theorem 3.3) holds for  $\delta$ , then this is a sufficient condition for the row-deletion property **AR** and, consequently, for variance identification, except for a set of measure 0.

Assuming a GLT structure also avoids the need to check the counting rule for all rotations of a given loading matrix. For illustration, we return to example (9) and use the counting rule of Sato (1992) to verify variance identification for  $\beta$ . If no condition to resolve rotational invariance is imposed on  $\beta$ , then the counting rule of Sato (1992) has to be checked not only for  $\beta$ , but for all rotations  $\beta\mathbf{G}$  or, equivalently, for all possible rotations  $\Lambda\mathbf{P}_{\alpha b}$  of  $\Lambda$ . Nearly all rotations have five non-zero rows and do not violate the counting rules, except for the eight rotations where  $(\alpha, b) \in \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\} \times \{0, 1\}$  leads to trivial rotations of  $\Lambda$ :

$$\begin{pmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \end{pmatrix} \begin{pmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & -\lambda_{42} \\ 0 & -\lambda_{52} \end{pmatrix} \begin{pmatrix} -\lambda_{11} & 0 \\ -\lambda_{21} & 0 \\ -\lambda_{31} & 0 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \end{pmatrix} \begin{pmatrix} -\lambda_{11} & 0 \\ -\lambda_{21} & 0 \\ -\lambda_{31} & 0 \\ 0 & -\lambda_{42} \\ 0 & -\lambda_{52} \end{pmatrix} \\ \begin{pmatrix} 0 & \lambda_{11} \\ 0 & \lambda_{21} \\ 0 & \lambda_{31} \\ \lambda_{42} & 0 \\ \lambda_{52} & 0 \end{pmatrix} \begin{pmatrix} 0 & -\lambda_{11} \\ 0 & -\lambda_{21} \\ 0 & -\lambda_{31} \\ \lambda_{42} & 0 \\ \lambda_{52} & 0 \end{pmatrix} \begin{pmatrix} 0 & \lambda_{11} \\ 0 & \lambda_{21} \\ 0 & \lambda_{31} \\ -\lambda_{42} & 0 \\ -\lambda_{52} & 0 \end{pmatrix} \begin{pmatrix} 0 & -\lambda_{11} \\ 0 & -\lambda_{21} \\ 0 & -\lambda_{31} \\ -\lambda_{42} & 0 \\ -\lambda_{52} & 0 \end{pmatrix}. \quad (10)$$

On the other hand, if we impose an unordered GLT structure on  $\beta$ , then the set of all possible rotations  $\beta\mathbf{G}$  reduces to the eight permutations in (10) and lack of variance identification can be verified by applying the counting rule of Sato (1992) to a single one of them.



### 3. Solving Rotational Invariance through GLT Structures

#### 3.1. Ordered and Unordered GLT Structures

In this work, we introduce a new identification strategy to resolve rotational invariance based on the concept of generalized lower triangular (GLT) structures. Throughout this section, we assume that variance identification holds and  $\Lambda\Lambda^\top$  and  $\Sigma_0$  are uniquely determined from  $\Omega$ . First, we introduce the notion of *pivot rows* of a factor loading matrix  $\Lambda$ .<sup>2</sup>

**Definition 2 (Pivot rows).** Consider an  $m \times r$  factor loading matrix  $\Lambda$  with  $r$  non-zero columns. For each column  $j = 1, \dots, r$  of  $\Lambda$ , the pivot row  $l_j$  is defined as the row index of the first non-zero factor loading in column  $j$ , i.e.,  $\Lambda_{ij} = 0, \forall i < l_j$  and  $\Lambda_{l_j,j} \neq 0$ . The factor loading  $\Lambda_{l_j,j}$  is called the leading factor loading of column  $j$ .

Conditions on the pivot rows define different structures on the factor loading matrix. If the pivot rows lie on the main diagonal, i.e.,  $(l_1, \dots, l_r) = (1, \dots, r)$ , then the factor loading matrix  $\Lambda$  exhibits a PLT structure and the leading factor loadings are equal to  $\Lambda_{jj}$  for all columns  $j = 1, \dots, r$ . GLT structures also require the pivot rows  $(l_1, \dots, l_r)$  to be pairwise distinct, but impose less stringent conditions on their position than a PLT structure. We will distinguish between two types of GLT structures, namely ordered and unordered GLT structures. In the first case, the pivots  $l_1 < \dots < l_r$  are ordered by size (see Definition 3), whereas they may take arbitrary positions  $(l_1, \dots, l_r)$  for unordered GLT structures, as long as they are pairwise distinct (a more formal definition is given below). Examples of all three structures are displayed in Figure 1 for a model with  $r = 6$  factors. Obviously, GLT structures contain the PLT structure as the special case where  $l_j = j$  for  $j = 1, \dots, r$ .

**Definition 3 (Ordered GLT structures).** An  $m \times r$  factor loading matrix  $\Lambda$  with full column rank  $r$  has an ordered GLT structure if the pivot rows  $l_1, \dots, l_r$  of  $\Lambda$  are ordered, i.e.,  $l_1 < \dots < l_r$ , and the leading factor loadings are positive, i.e.,  $\Lambda_{l_j,j} > 0$  for  $j = 1, \dots, r$ .

Since the “diagonal” loadings  $\Lambda_{jj}$  are allowed to be zero for an ordered GLT structure, measurements different from the first  $r$  ones may lead the factors. For each factor  $j$ , the leading variable is the response variable  $y_{l_j,t}$  corresponding to the pivot row  $l_j$ . From a mathematical viewpoint, one could argue that the  $m$  measurements  $y_{1t}, \dots, y_{mt}$  can be rearranged such that a PLT structure holds and the diagonal of the first  $r$  rows of  $\Lambda$  has non-zero elements. However, in practice, it is not obvious which measurements are able to constitute a new factor; in particular, if we expect the loading matrix  $\Lambda$  to exhibit a simple structure, but the measurements have no natural grouping. Even if such a grouping exists, rearranging the measurements might be challenging, if more than one factor is needed to explain the group-specific covariance.

As opposed to PLT, under GLT structures, we may learn from the data whether the  $i$ th row of the loading matrix is linearly independent of the previous rows 1 to  $i - 1$ . Only in this case, the measurement  $y_{it}$  constitutes an additional factor and row  $i$  defines the next pivot. In the empirical case study in Section 7.2, for instance, the measurements are grouped by industry and the pivot rows learned from the data show that this ordering is in conflict with the PLT assumption (see also Figure 2).

Imposing an ordered GLT structure resolves rotational invariance if the pivot rows are known. For any two ordered GLT matrices  $\beta$  and  $\Lambda$  with identical pivot rows  $l_1, \dots, l_r$ , the identity  $\beta = \Lambda P$  evidently holds if  $P = I_r$ . GLT structures where the pivot rows are unknown will be discussed in Section 3.3.

In some treatment, it is customary to impose conditions that resolve rotational invariance up to column and sign switching (see, e.g., Conti et al. (2014)). This trivial form of rotational invariance does not impose any additional mathematical challenges and is often convenient from a computational viewpoint, in particular for Bayesian inference (see also Frühwirth-Schnatter et al. (2023)). More formally, so-called *signed permutations* are introduced to permute the columns of the factor loading matrix  $\Lambda$  and to reverse the sign

of all factor loadings in specific columns. Such a signed permutation  $\beta$  of a loading matrix  $\Lambda$  is defined as

$$\beta = \Lambda \mathbf{P}_{\pm} \mathbf{P}_{\rho}, \quad (11)$$

where the permutation matrix  $\mathbf{P}_{\rho}$  corresponds to one of the  $r!$  permutations of the  $r$  columns of  $\Lambda$  and introduces column switching. The reflection matrix  $\mathbf{P}_{\pm} = \text{Diag}(\pm 1, \dots, \pm 1)$  corresponds to one of the  $2^r$  possibilities to either keep or reverse the sign of each of the  $r$  columns of  $\Lambda$  and introduces sign switching. This generates a whole equivalence class of loading matrices given by all  $2^r r!$  signed permutations  $\beta = \Lambda \mathbf{P}_{\pm} \mathbf{P}_{\rho}$  of  $\Lambda$ .

It is easy to verify how identification up to signed permutations can be achieved for GLT structures. For  $r = 2$ , for instance, all eight signed permutations of the ordered GLT structure  $\Lambda$  defined in (9) are depicted in (10). Applying all signed permutations to an arbitrary ordered GLT structure leads to the specification of so-called *unordered* GLT structures as loading matrices where the pivot rows  $l_1, \dots, l_r$  simply occupy  $r$  different rows (see Definition 4).

**Definition 4 (Unordered GLT structures).** An  $m \times r$  factor loading matrix  $\beta$  with full column rank  $r$  has an unordered GLT structure if the pivot rows  $l_1, \dots, l_r$  of  $\beta$  are pairwise distinct.

In Definition 4, no order constraint is imposed on the pivot rows and no sign constraint is imposed on the leading factor loadings. This very general structure allows us to design highly efficient sampling schemes for Bayesian factor analysis under GLT structures (see Frühwirth-Schnatter et al. (2023)).

For unordered GLT structures with known pivots, rotational invariance is resolved only up to signed permutations; however, full identification can be easily obtained. Any unordered GLT structure  $\beta$  has (unordered) pivot rows  $l_1, \dots, l_r$  occupying different rows. The corresponding ordered GLT structure  $\Lambda$  is recovered from  $\beta$  by sorting the columns of  $\beta$  such that the pivot rows of  $\Lambda$  are equal to the order statistics  $l_{(1)} < \dots < l_{(r)}$  of the pivot rows  $l_1, \dots, l_r$  of  $\beta$  (see again Figure 1). This procedure resolves rotational invariance, since the pivot rows  $l_1, \dots, l_r$  in the unordered GLT structure are distinct. Furthermore, imposing the condition  $\Lambda_{l_i, j} > 0$  in each column  $j$  resolves sign switching: if  $\Lambda_{l_i, j} < 0$ , then the sign of all factor loadings  $\Lambda_{ij}$  in column  $j$  is reversed.

### 3.2. Rotation into GLT

Ordered GLT structures generalize the PLT constraint, but one might wonder how restrictive this condition still is. Theorem 1 proves that it is unrestrictive in the sense of Anderson and Rubin (1956) and Jöreskog (1969). We show that for any basic factor model with an unconstrained loading matrix  $\beta$  there exists an equivalent representation involving an ordered GLT structure  $\Lambda$ , which is related to  $\beta$  by an orthogonal transformation, provided that uniqueness of the variance decomposition holds. See Appendix A for a proof.

**Theorem 1 (Rotation into GLT).** Let  $\beta$  be an arbitrary loading matrix with full column rank  $r$  and let  $\beta_{l, \bullet}$  denote the  $l$ th row of  $\beta$ . Then, the following statements hold:

(a) There exists an equivalent unique representation of  $\beta$  involving an ordered GLT structure  $\Lambda$ ,

$$\beta = \Lambda \mathbf{G}^{\top}, \quad (12)$$

where  $\mathbf{G}$  is a rotation matrix.  $\Lambda$  is called the GLT representation of  $\beta$ ;

(b) To compute  $\mathbf{G}$  from  $\beta$ , first find the smallest row index  $l_1$  such that the  $l_1$ th row of  $\beta$  is not fully zero. Next, in an iterative manner, given indices  $(l_1, \dots, l_{i-1})$  for  $2 \leq i \leq r$ , find the smallest row index  $l_i$  such that  $\beta_{l_1, \bullet}, \dots, \beta_{l_{i-1}, \bullet}$  and  $\beta_{l_i, \bullet}$  together form a linearly independent set of vectors. After the last iteration, the rows  $\beta_{l_1, \bullet}, \dots, \beta_{l_r, \bullet}$  form an  $r \times r$  invertible matrix  $\tilde{\beta}$ . Then,  $\mathbf{G}$  is the ‘Q’ part of the QR-decomposition of  $\tilde{\beta}^{\top}$ .<sup>3</sup>

Does a similar result hold for PLT structures? The answer is definitely no, as has already been established in Section 2; in (4), for example. As mentioned above, ordered GLT structures encompass PLT structures as a special case. Hence, as a consequence of Theorem 1, if a PLT representation  $\Lambda$  exists for a loading matrix  $\beta = \Lambda P$ , then the ordered GLT representation in (12) *automatically* reduces to the PLT structure  $\Lambda$  with the “rotation into GLT” being  $G = P^\top$ . On the other hand, if the ordered GLT representation  $\Lambda$  differs from a PLT structure, then no equivalent PLT representation exists. Hence, forcing a PLT structure in the representation (1) may introduce a bias in estimating the marginal covariance matrix  $\Omega$ .

In practice, the pivot rows  $l_1, \dots, l_r$  of a GLT structure are unknown and need to be identified from the marginal covariance matrix  $\Omega$  for a given number of factors  $r$ . Given variance identification, i.e., assuming that the cross-covariance matrix  $\Lambda\Lambda^\top$  is identified, an interesting question regarding the identification of a GLT factor model is whether  $\Lambda$  is uniquely identified from  $\Lambda\Lambda^\top$  if the pivot rows are *unknown*. Non-trivial rotations  $\tilde{\Lambda} = \Lambda P$  of a loading matrix  $\Lambda$  might exist such that  $\tilde{\Lambda}\tilde{\Lambda}^\top = \Lambda\Lambda^\top$ , while the pivot rows  $\tilde{l}_1, \dots, \tilde{l}_r$  of  $\tilde{\Lambda}$  are different from the pivot rows  $l_1, \dots, l_r$  of  $\Lambda$ .

For ordered GLT structures with unknown pivots, we obtain as an immediate (and somewhat trivial) consequence of Theorem 1 that  $\tilde{\Lambda}$  is the unique ordered GLT representation of  $\Lambda$ , and therefore  $P = I_r$ . Indeed, when we compute the rotation matrix  $G$  in  $\Lambda = \tilde{\Lambda}G^\top$  as described in Part (b), we find that the matrix  $\tilde{\beta}$  is equal to the pivot rows of  $\Lambda$ . Therefore,  $\tilde{\beta}^\top$  is an upper triangular matrix of full rank and the ‘Q’ part of its QR-decomposition is equal to the identity matrix. Since  $\Lambda = \tilde{\Lambda}P^\top$  and rotation into GLT is unique, we obtain that  $P^\top = G^\top = I_r$ . This insight is formalized in Corollary 1.

**Corollary 1.** *An ordered GLT structure is uniquely identified, provided that uniqueness of the variance decomposition holds, i.e., if  $\Lambda$  and  $\tilde{\Lambda}$  are GLT matrices, respectively, with pivot rows  $l_1 < \dots < l_r$  and  $\tilde{l}_1 < \dots < \tilde{l}_r$  that satisfy  $\tilde{\Lambda}\tilde{\Lambda}^\top = \Lambda\Lambda^\top$ , then  $\tilde{\Lambda} = \Lambda$  and, consequently,  $(\tilde{l}_1, \dots, \tilde{l}_r) = (l_1, \dots, l_r)$ .*

For unordered GLT structures with unknown pivots, the factor loading matrix  $\Lambda$  is identified from  $\Lambda\Lambda^\top$  up to signed permutations and  $\tilde{\Lambda}$  and  $\Lambda$  have the same pivot rows  $l_1, \dots, l_r$ , provided that  $\Lambda\Lambda^\top$  is identified. This can be easily shown by extending Corollary 1 to unordered GLT structures, as any signed permutation  $\tilde{\Lambda} = \Lambda P_\pm$  of  $\Lambda$  is uniquely identified from  $\tilde{\Lambda}\tilde{\Lambda}^\top = \Lambda\Lambda^\top$ . To summarize, Theorem 1 together with Corollary 1 establish the existence and uniqueness of GLT structures for any basic factor model, provided that uniqueness of the variance decomposition holds.

### 3.3. Simple GLT Structures

In Definitions 3 and 4, “structural” zeros are introduced for a GLT structure for all factor loading above the pivot row  $l_j$ , while the factor loading  $\Lambda_{l_j,j}$  in the pivot row is non-zero by definition. We call  $\Lambda$  a *dense* GLT structure if all loadings below the pivot rows are unconstrained and can take any value in  $\mathbb{R}$ .

A simple GLT structure results if factor loadings at unspecified places below the pivot rows are zero and only the remaining loadings are unconstrained. As discussed in Section 2.2, any simple structure can be characterized by the sparsity matrix  $\delta$ , defined as a binary indicator matrix of 0/1s of the same size as  $\Lambda$ , where  $\delta = \mathbb{I}(\Lambda \neq 0)$  and the indicator function is applied element-wise. Evidently, the sparsity matrix of a GLT structure  $\Lambda$  exhibits the same pivots as  $\Lambda$ , regardless of whether the structure is ordered or unordered (see again Figure 1 for an illustration).

In sparse Bayesian factor analysis, single factor loadings take zero-values with positive probability and the corresponding sparsity matrix  $\delta$  is a random binary matrix that has to be identified from the data (see Section 6.1 for more details). Identification in sparse Bayesian factor analysis has to provide conditions under which the entire 0/1 pattern in  $\delta$  can be identified from a given covariance matrix  $\Omega$ , if  $\delta$  is unknown. Whether this is

possible hinges on variance identification, i.e., whether the decomposition of  $\Omega$  into  $\Lambda\Lambda^\top$  and  $\Sigma_0$  is unique. How variance identification can be verified for simple GLT structures is investigated in detail in Section 4. Let us assume at this point that variance identification holds, i.e., the cross-covariance matrix  $\Lambda\Lambda^\top$  is identified. Then, an important step toward the identification of a factor model with a simple structure is to verify whether the 0/1 pattern of  $\Lambda$ , characterized by the sparsity matrix  $\delta$ , is uniquely identified from  $\Lambda\Lambda^\top$ . If  $\Lambda$  exhibits an ordered GLT structure, then it follows immediately from Corollary 1 that the indicator matrix  $\delta$  is, indeed, uniquely identified from  $\Lambda\Lambda^\top$ , since  $\Lambda$  is identified and  $\delta_{ij} = 0$  if  $\Lambda_{ij} = 0$  for all  $i, j$ .

We would like to emphasize that in sparse Bayesian factor analysis with unconstrained loading matrices  $\Lambda$ , this is not necessarily the case. The sparsity matrix  $\delta$  is, in general, not uniquely identified from  $\Lambda\Lambda^\top$ , because rotations may change the zero pattern in  $\beta = \Lambda\mathbf{P}$ , while  $\beta\beta^\top = \Lambda\Lambda^\top$ . For illustration, let us return to Example (4) and assume that the sparsity matrix  $\delta$  is unknown. While  $\delta$  is uniquely identified from  $\Lambda\Lambda^\top$  under GLT, two distinct solutions  $\delta$  exist if the loading matrix is left unconstrained and any rotation  $\beta = \Lambda\mathbf{P}_{ab}$  of  $\Lambda$  is an admissible solution. For all rotations where  $(a, b) \in \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\} \times \{0, 1\}$ ,  $\beta$  corresponds to one of the eight signed permutations of  $\Lambda$  (similar to the rotations in (10)) and  $\delta$  is equal to the sparsity matrix of  $\Lambda$  up to column switching. For all other rotations, all elements of  $\beta$  are different from zero and  $\delta$  is simply a matrix of ones.

#### 4. Variance Identification for Simple GLT Structures

As mentioned in the previous sections, conditions imposed on the structure of a factor loading matrix  $\Lambda$  will resolve rotational invariance only if the uniqueness of the variance decomposition holds and the cross-covariance matrix  $\Lambda\Lambda^\top$  is identified. However, such conditions do not necessarily guarantee the uniqueness of the variance decomposition.

One exception is the popular factor analysis model where  $\Lambda$  takes the form of a *dense* PLT matrix, where all factor loadings below the main diagonal are unconstrained and may take any value in  $\mathbb{R}$ . For this model, condition AR and hence variance identification holds, except for a set of measure 0, provided that the condition

$$m \geq 2r + 1 \iff r \leq \frac{m-1}{2} \quad (13)$$

on the number of factors is satisfied for the given number  $m$  of measurements. For *simple* structures variance identification is easily violated; consider, e.g., a *simple* PLT loading matrix where only a single factor loading below the diagonal is nonzero in some column.

In Section 4.1, we derive sufficient conditions for variance identification of simple GLT structures based on the 3579 counting rule of Sato (1992, Theorem 3.3). In Section 4.2, we discuss how to verify variance identification for simple GLT structures in practice.

##### 4.1. Counting Rules for Variance Identification

First, of all, we need not constrain the factor loading matrix to take the form of an ordered GLT structure, since variance identification is invariant to signed permutations. If we can verify the variance identification for a single signed permutation  $\beta = \Lambda\mathbf{P}_\pm$  of a loading matrix  $\Lambda$ , as defined in (11), then variance identification of  $\Lambda$  holds, since  $\beta$  and  $\Lambda$  imply the same cross-covariance matrix  $\Lambda\Lambda^\top$ . Hence, we focus in this section on the variance identification of unordered GLT structures. We will show how to verify from the 0/1 pattern of the sparsity matrix  $\delta$  of an unordered, possibly simple GLT structure  $\beta$ , whether the row-deletion property AR holds for  $\beta$  and all its signed permutations. Our condition is a structural counting rule expressed solely in terms of the sparsity matrix  $\delta$  underlying  $\beta$ , and does not involve the values of the unconstrained factor loadings in  $\beta$ , which can take any value in  $\mathbb{R}$ .

Next, we recall the so-called *extended row-deletion property* in Definition 5, introduced by Tumura and Sato (1980) which applies to arbitrary loading matrices  $\beta$ .

**Definition 5** (Extended row-deletion property  $RD(r, s)$ ). An  $m \times r$  factor loading matrix  $\beta$  satisfies the row-deletion property  $RD(r, s)$  if the following condition is satisfied: whenever  $s \in \mathbb{N}_0$  rows are deleted from  $\beta$ , then two disjoint submatrices of rank  $r$  remain.

The row-deletion property of Anderson and Rubin (1956) results as a special case where  $s = 1$ . As will be shown in Section 5, the extended row-deletion properties  $RD(r, s)$  for  $s > 1$  are useful in exploratory factor analysis, when the factor dimension  $r$  is unknown. In Definition 6, we introduce a counting rule for binary matrices.

**Definition 6** (Counting rule  $CR(r, s)$ ). Let  $\delta$  be an  $m \times r$  binary matrix. For each  $q = 1, \dots, r$ , consider all submatrices  $\delta_{q,\ell}$ ,  $\ell = 1, \dots, \binom{r}{q}$ , built from  $q$  columns of  $\delta$ .  $\delta$  is said to satisfy the  $CR(r, s)$  counting rule for  $s \in \mathbb{N}_0$  if the matrix  $\delta_{q,\ell}$  has at least  $2\ell + s$  nonzero rows for all  $(q, \ell)$ .

Note that the counting rule  $CR(r, s)$ , like the extended row-deletion property  $RD(r, s)$ , is invariant to signed permutations. Lemma A1 in Appendix A summarizes further useful properties of  $CR(r, s)$ .

For a given binary matrix  $\delta$  of dimension  $m \times r$ , let  $\Theta_\delta$  be the space generated by the non-zero elements of all unordered simple GLT structures  $\beta$  with the same sparsity matrix  $\delta$  and all their  $2^r r! - 1$  signed permutations  $\beta \mathbf{P}_\pm \mathbf{P}_\rho$ . We prove in Theorem 2 that for simple GLT structures, the counting rule  $CR(r, s)$  and the extended row-deletion property  $RD(r, s)$  are equivalent conditions for all simple loading matrices in  $\Theta_\delta$  with the same sparsity matrix  $\delta$ , except for a set of measure 0.

**Theorem 2.** Let  $\delta$  be a binary  $m \times r$  matrix with an unordered GLT structure. Then, the following holds:

- (a) If  $\delta$  violates the counting rule  $CR(r, s)$ , then the extended row-deletion property  $RD(r, s)$  is violated for all simple structures  $\beta \in \Theta_\delta$  generated by  $\delta$ ;
- (b) If  $\delta$  satisfies the counting rule  $CR(r, s)$ , then the extended row-deletion property  $RD(r, s)$  holds for all simple structures  $\beta \in \Theta_\delta$  except for a set of measure 0.

See Appendix A for a proof. The special case  $s = 1$  is relevant for verifying the row-deletion property **AR**. It proves that, for unordered simple GLT structures, the 3579 counting rule of Sato (1992) is not only a necessary, but also a sufficient condition for **AR** to hold. In addition, this means that the counting rule needs to be verified only for the sparsity matrix  $\delta$  of a single signed permutation  $\beta = \Lambda \mathbf{P}_\pm \mathbf{P}_\rho$  rather than for every non-singular matrix  $\mathbf{G}$ . This result is summarized in Corollary 2.

**Corollary 2** (Variance identification rule for simple GLT structures). For any unordered simple GLT structure  $\beta$  of size  $m \times r$ , the following holds:

- (a) If a binary matrix  $\delta$  of size  $m \times r$  satisfies the 3579 counting rule, i.e., every column of  $\delta$  has at least three non-zero elements, every pair of columns at least five, and, more generally, every possible combination of  $q = 3, \dots, r$  columns has at least  $2q + 1$  non-zero elements, then variance identification is given for all simple unordered GLT structures  $\beta \in \Theta_\delta$  except for a set of measure 0; i.e., for any other factor decomposition of the marginal covariance matrix  $\Omega = \beta\beta^\top + \Sigma = \tilde{\beta}\tilde{\beta}^\top + \tilde{\Sigma}$ , where  $\tilde{\beta}$  is an unordered GLT matrix, it follows that  $\tilde{\Sigma} = \Sigma$ , i.e.,  $\tilde{\beta}\tilde{\beta}^\top = \beta\beta^\top$ , and  $\tilde{\beta} = \beta \mathbf{P}_\pm \mathbf{P}_\rho$ .
- (b) If a binary matrix  $\delta$  of size  $m \times r$  violates the 3579 counting rule, then for all  $\beta \in \Theta_\delta$ , the row-deletion property **AR** does not hold.
- (c) For  $r = 1$ ,  $r = 2$ , and  $r = 3$ , condition  $CR(r, 1)$  is both sufficient and necessary for variance identification.

A few comments are in order. If  $\delta$  satisfies  $CR(r, 1)$ , then **AR** holds for all  $\beta \in \Theta_\delta$  and a sufficient condition for variance identification is satisfied. As shown by Anderson and Rubin (1956), **AR** is a necessary condition for variance identification only for  $r = 1$  and



$r = 2$ . Tumura and Sato (1980, Theorem 3) show the same for  $r = 3$ , provided that  $m \geq 7$ . It follows that  $\text{CR}(r, 1)$  is a *necessary and sufficient* condition for variance identification for all models summarized in (c). In all other cases, variance identification may hold for loading matrices  $\beta \in \Theta_\delta$ , even if  $\delta$  violates  $\text{CR}(r, 1)$ .

The definition of unordered GLT structures given in Section 3 imposes no condition on the position of the pivot rows  $l_1, \dots, l_r$  beyond the assumption that they are distinct. This may lead to GLT structures that can never satisfy the 3579 rule, even if all elements below the pivot rows are non-zero. Consider, for instance, a GLT matrix with the pivot row in column  $r$  being equal to  $l_r = m - 1$ . The loading matrix has at most two nonzero elements in column  $r$  and violates the necessary condition for variance identification. This example shows that there is an upper bound for the pivot elements beyond which the 3579 rule can never hold. This insight is formalized in Definition 7.

**Definition 7.** An unordered GLT structure  $\beta$  fulfills condition **GLT-AR**, if the pivots  $l_1, \dots, l_r$  satisfy the following condition, where  $z_j$  is the rank of  $l_j$  in the ordered sequence  $l_{(1)} < \dots < l_{(r)}$ :

$$l_j \leq m - 2(r - z_j + 1). \quad (14)$$

For an ordered GLT structure  $\Lambda$  with pivots  $l_1 < \dots < l_r$ , condition **GLT-AR** reduces to:

$$l_j \leq m - 2(r - j + 1). \quad (15)$$

For the special case of a PLT structure where  $l_j = j$ , condition (15) reduces to the upper bound for the number of factors given in (13).

For dense GLT structures, condition **GLT-AR** is a sufficient condition for **AR**. However, for simple GLT structures with zeros below the pivot rows, **GLT-AR** is only a necessary condition for **AR**, as discussed above for an example, and the 3579 rule has to be verified explicitly. Very conveniently for verifying variance identification in factor analysis based on GLT structures, Theorem 2 and Corollary 2 operate solely on the sparsity matrix  $\delta$  summarizing the simple structure in  $\beta$ .

#### 4.2. Variance Identification in Practice

To verify  $\text{CR}(r, s)$  in practice, all submatrices of  $q$  columns have to be extracted from the sparsity matrix  $\delta$  to verify if at least  $2q + 1$  rows of this submatrix are non-zero. For  $q = 1, 2, r - 1, r$ , this condition is easily verified from simple functionals of  $\delta$ ; see Corollary 3, which follows immediately from Theorem 2 (see Appendix A for details).

**Corollary 3 (Simple counting rules for  $\text{CR}(r, s)$ ).** Let  $\delta$  be a  $m \times r$  unordered GLT sparsity matrix. The following conditions on  $\delta$  are necessary for  $\text{CR}(r, s)$  to hold:

$$\mathbf{1}_{r \times m} \cdot \delta + \delta^\top (\mathbf{1}_{m \times r} - \delta) \geq 4 + s - 2\mathbf{I}_r, \quad (16)$$

$$\mathbf{1}_{1 \times m} \cdot \mathbb{I}(\delta^* > 0) \geq 2r + s, \quad \delta^* = \delta \cdot \mathbf{1}_{r \times 1}, \quad (17)$$

$$\mathbf{1}_{1 \times m} \cdot \mathbb{I}(\delta^* > 0) \geq 2(r - 1) + s, \quad \delta^* = (\mathbf{1}_{m \times m} - \mathbf{I}_m) \cdot \delta, \quad (18)$$

where the indicator function  $\mathbb{I}(\delta^* > 0)$  is applied element-wise and  $\mathbf{1}_{n \times k}$  denotes an  $n \times k$  matrix of ones. For  $r \leq 4$ , these conditions are also sufficient for  $\text{CR}(r, s)$  to hold for  $\delta$ .

Using Corollary 3 for  $s = 1$ , one can efficiently verify if the 3579 counting rule and hence the row-deletion property **AR** holds for simple unordered GLT factor models with up to four ( $r \leq 4$ ) factors. For models with more than four factors ( $r > 4$ ), a more elaborated strategy is needed. After checking the conditions of Corollary 3,  $\text{CR}(r, s)$  could be verified for a given binary matrix  $\delta$  by iterating over all remaining  $r! / (q!(r - q)!)$  subsets of  $q = 3, \dots, r - 2$  columns of  $\delta$ . While this is a finite task, such a naïve approach may need to visit  $2^r - 1$  matrices in order to make a decision and the combinatorial explosion quickly

becomes an issue in practice as  $r$  increases. Recent work by [Hosszejni and Frühwirth-Schnatter \(2022\)](#) establishes the applicability of this framework for large models.<sup>4</sup>

## 5. Identification in Exploratory Factor Analysis

In this section, we discuss how the concept of GLT structures is helpful for addressing identification problems in exploratory factor analysis.

### 5.1. Exploratory Factor Analysis

Consider data  $\{\mathbf{y}_1, \dots, \mathbf{y}_T\}$  from a zero-mean multivariate Gaussian distribution, where an investigator wants to perform factor analysis since she expects that the covariances of  $m$  measurements are driven by common factors. In practice, the number of factors is often unknown and it may be uncertain whether all measurements are actually correlated. It is then common to employ exploratory factor analysis (EFA) by fitting the following basic factor model with an *assumed* maximum number of factors,  $H$ , to all measurements in  $\mathbf{y}_t$ :

$$\mathbf{y}_t = \beta_H \mathbf{f}_t + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}_m(\mathbf{0}, \Sigma_H), \quad (19)$$

where  $\beta_H$  is an  $m \times H$  loading matrix, not necessarily of full column rank,  $\mathbf{f}_t \sim \mathcal{N}_H(\mathbf{0}, \mathbf{I}_H)$ , and  $\Sigma_H$  is a diagonal matrix with strictly positive entries. The EFA model (19) is potentially overfitting in two ways. First, if some measurements in  $\mathbf{y}_t$  are uncorrelated with the remaining measurements, then  $\beta_H$  allows for too many non-zero rows. However, as will be discussed in Section 5.3, such irrelevant measurements are easily identified. Second, the assumed number of factors  $H$  is possibly larger than the true number of factors  $r$ , i.e.,  $\beta_H$  has too many columns. The goal is then to extract the true number of factors from the non-zero columns of  $\beta_H$ , collected in a  $m \times k$  submatrix  $\beta_k$  of rank  $k$ . Before we discuss this challenging problem in more detail in Section 6.3, additional identification issues for overfitting factor models have to be addressed.

We assume that the data are generated by a basic factor model with error covariance matrix  $\Sigma_0$  and a loading matrix  $\Lambda$  of factor dimension equal to  $r$ . Therefore,  $\mathbf{y}_t \sim \mathcal{N}_m(\mathbf{0}, \Omega)$ , where the covariance matrix  $\Omega$  has a representation as in (2):

$$\Omega = \Lambda \Lambda^\top + \Sigma_0. \quad (20)$$

Furthermore, we assume that variance identification holds for (20) and the true cross-correlation matrix  $\Lambda \Lambda^\top$  as well as  $\Sigma_0$  are uniquely determined from  $\Omega$ .

Two questions arise in this context and can be answered based on [Reiersøl \(1950\)](#). First, under which conditions the covariance matrix is implied by the loading matrix  $\beta_k$  extracted from the EFA model (19) equivalent to the true covariance matrix of the data, i.e., when does the following hold (note that  $\Sigma_k = \Sigma_H$ ):

$$\beta_k \beta_k^\top + \Sigma_k = \Omega? \quad (21)$$

Second, if such an equivalent representation exists, does variance identification still hold, i.e., can  $\beta_k \beta_k^\top$  and  $\Sigma_k$  be uniquely determined from (21)?

It follows from [Reiersøl \(1950\)](#) that no equivalent representation exists, if  $k \leq H < r$ . [Reiersøl \(1950\)](#) show that the true number of factors  $r$  is equal to the smallest value  $k$  that satisfies (21). Hence, if the assumed number of factors  $H$  in the EFA model is equal to the true number of factors  $r$  and  $\beta_H$  has full column rank, then such a representation obviously exists. Since variance identification of (20) holds, we obtain that  $\beta_H \beta_H^\top = \beta_r \beta_r^\top = \Lambda \Lambda^\top$  and  $\Sigma_H = \Sigma_r = \Sigma_0$ . Consequently,  $\beta_H = \beta_r = \Lambda \mathbf{P}$  is a rotation of  $\Lambda$ .

Equivalent representations also exist for  $H \geq k > r$ , in which case the EFA model is *overfitting*. It follows from [Reiersøl \(1950, Theorem 3.3\)](#) that any structure  $(\Lambda, \Sigma_0)$  in a basic

factor model of factor dimension  $r$  creates infinitely many solutions  $(\beta_k, \Sigma_k)$  of dimension  $k = r + 1, \dots, H$  which imply the same covariance matrix  $\Omega$  as  $(\Lambda, \Sigma_0)$ , i.e.,

$$\Omega = \beta_k \beta_k^\top + \Sigma_k = \Lambda \Lambda^\top + \Sigma_0, \quad (22)$$

where the rank of the loading matrix  $\beta_k$  is equal to  $k > r$  and  $\Sigma_k$  is a positive definite matrix different from  $\Sigma_0$  (see also Geweke and Singleton (1980)). Since infinitely many solutions can be created that differ in  $\Sigma_k$ , the decomposition (22) is no longer variance identified.

For illustration, we return to Example (4), where  $r = 2$ , and construct infinitely many solutions for  $k = 3$ . The first two columns of  $\beta_3$  are equal to  $\Lambda$ , the third column is a so-called *spurious* factor with a single non-zero loading, and  $\Sigma_3$  is defined as follows:

$$\beta_3 = \begin{pmatrix} \lambda_{11} & 0 & 0 \\ \lambda_{21} & 0 & \beta_{23} \\ \lambda_{31} & 0 & 0 \\ \lambda_{41} & \lambda_{42} & 0 \\ 0 & \lambda_{52} & 0 \\ 0 & \lambda_{62} & 0 \end{pmatrix}, \quad \Sigma_3 = \text{Diag}(\sigma_1^2, \sigma_2^2 - \beta_{23}^2, \sigma_3^2, \sigma_4^2, \sigma_5^2, \sigma_6^2). \quad (23)$$

We can place the spurious factor loading  $\beta_{i3}$  in any row  $i \in \{1, \dots, m\}$  and it can take any value satisfying  $0 < \beta_{i3}^2 < \sigma_i^2$ . It is easy to verify that any such pair  $(\beta_3, \Sigma_3)$  indeed implies the same covariance matrix  $\Omega$  as the pair  $(\Lambda, \Sigma_0)$ . Therefore, while variance identification holds for  $r = 2$ , it fails for  $k = 3$ , because infinitely many solutions with different error covariance matrices  $\Sigma_3$  are available, depending on the choice of  $i$  and the spurious factor loading  $\beta_{i3}$ . On the other hand, the spurious column is easily spotted for any such solution  $\beta_3$  and the true loading matrix  $\Lambda$  is clearly identified from the two remaining columns. One may even recover  $\Sigma_0$ , by adding  $\beta_{i3}^2$  to the  $i$ th diagonal element of  $\Sigma_k$ .

So far in this section, we imposed no conditions that resolve rotational invariance, either on the true loading matrix  $\Lambda$  or on the overfitting loading matrix  $\beta_3$ . In this case, additional solutions are obtained by rotating the loading matrices  $\beta_3$  in (23), e.g.,

$$\tilde{\beta}_3 = \begin{pmatrix} 0 & 0 & \lambda_{11} \\ \beta_{23} & 0 & \lambda_{21} \\ 0 & 0 & \lambda_{31} \\ 0 & -\lambda_{42} & \lambda_{41} \\ 0 & -\lambda_{52} & 0 \\ 0 & -\lambda_{62} & 0 \end{pmatrix}, \quad \tilde{\tilde{\beta}}_3 = \begin{pmatrix} -\lambda_{11} \sin \alpha & 0 & \lambda_{11} \cos \alpha \\ \beta_{23} \cos \alpha - \lambda_{21} \sin \alpha & 0 & \lambda_{21} \cos \alpha \\ -\lambda_{31} \sin \alpha & 0 & \lambda_{31} \cos \alpha \\ -\lambda_{41} \sin \alpha & \lambda_{42} & \lambda_{41} \cos \alpha \\ 0 & \lambda_{52} & 0 \\ 0 & \lambda_{62} & 0 \end{pmatrix}, \quad (24)$$

both combined with the same  $\Sigma_3$  as in (23). The first solution  $\tilde{\beta}_3$  is a signed permutation of  $\beta_3$ , while the second solution  $\tilde{\tilde{\beta}}_3$  combines a signed permutation of  $\beta_3$  with a rotation of the spurious and  $\Lambda$ 's first column involving  $\mathbf{P}_{\alpha b}$ . In  $\tilde{\beta}_3$ , the spurious column is still easily spotted and the true loading matrix  $\Lambda$  is identified from the two remaining columns of  $\beta_3$  up to column and sign switching.<sup>5</sup> However, for  $\tilde{\tilde{\beta}}_3$  the presence of a spurious column is no longer obvious. While the second column of  $\Lambda$  still pops up, the first column is disguised.

To summarize, further identifiability problems arise for  $r < k \leq H$  beyond the ones discussed in the previous sections for a basic factor model where  $r$  is known. We discuss these problems in a more formal manner in Section 5.2 and investigate the class of overfitting GLT structures where an unordered GLT condition is imposed on the non-zero columns  $\beta_k$  of the loading matrix  $\beta_H$  in the EFA model (19). We apply results by Tumura and Sato (1980) to this class and prove that under this condition, (a) spurious factors in  $\beta_k$  are as easily spotted as in  $\tilde{\beta}_3$ , and (b) the non-spurious columns are an unordered GLT representation of the true loading matrix  $\Lambda$ . Our strategy relies on the concept of extended variance identification and the extended row-deletion property introduced by Tumura and Sato (1980), where more than one row is deleted from the loading matrix. The extended

counting rule  $RD(r, s)$  with  $s > 1$  introduced in Definition 5 in Section 4 will be useful in this context.

### 5.2. “Revealing the Truth” in an Overfitting EFA Model

As illustrated in Section 5.1, fundamental identifiability problems arise if an overfitting EFA model with  $r < k \leq H$  is fitted to data arising from a basic factor model of factor dimension  $r$ . The question arises if we could, nevertheless, recover the true loading matrix  $\Lambda$  from the non-zero columns  $\beta_k$  of  $\beta_H$ . We will show how this can be achieved mathematically by combining the important work by Tumura and Sato (1980) with the framework of GLT structures.

We have demonstrated in Section 5.1, using examples (23) and (24) for a model where  $k = r + 1$ , how to construct infinitely many solutions  $(\beta_k, \Sigma_k)$  with the same covariance matrix  $\Omega = \beta_k \beta_k^\top + \Sigma_k$  as the true factor model. First, set the first  $r$  columns of  $\beta_k$  to  $\Lambda$  and append a spurious column to its right with the single non-zero loading  $\beta_{l_k, k}$  lying in any row  $l_k \in \{1, \dots, m\}$  taking any value that satisfies  $0 < \beta_{l_k, k}^2 < \sigma_{l_k}^2$ . Then, reduce the idiosyncratic variance in row  $l_k$  to  $\sigma_{l_k}^2 - \beta_{l_k, k}^2$ , and finally apply an arbitrary rotation  $P$ :

$$\beta_k = \left( \Lambda \mid \begin{array}{c} 0 \\ \beta_{l_k, k} \\ 0 \end{array} \right) P, \quad \Sigma_k = \text{Diag}(\sigma_1^2, \dots, \sigma_{l_k}^2 - \beta_{l_k, k}^2, \dots, \sigma_m^2). \quad (25)$$

The following questions arise: under which conditions is (25) an exhaustive representation of all possible solutions  $\beta_k$  of rank  $k$  where the *degree of overfitting* defined as  $s = k - r$  is equal to one? How can all possible solutions  $\beta_k$  be represented if  $s > 1$ ?

Such identifiability problems in overfitting EFA models have been analyzed in depth by Tumura and Sato (1980). They provide a general representation of the factor loading matrix  $\beta_k$  in an overfitting representation (22) with  $k > r$ . In addition, they show that a stronger condition than  $RD(r, 1)$  is needed for  $\Lambda$  in the underlying variance decomposition (21) to ensure that only spurious factors, and no additional common factors, are present in (21).

**Theorem 3.** (Tumura and Sato 1980, Theorem 1) Suppose that  $\Omega$  has a decomposition as in (20) with  $r$  factors, and that for some  $S \in \mathbb{N}$  with  $2r + S + 1 \leq m$  the extended row-deletion property  $RD(r, 1 + S)$  holds for  $\Lambda$ . If  $\Omega$  has another decomposition such that  $\Omega = \beta_k \beta_k^\top + \Sigma_k$  where  $\beta_k$  is an  $m \times (r + s)$ -matrix of rank  $k = r + s$  with  $1 \leq s \leq S$ , then there exists an orthogonal matrix  $T_k$  of rank  $k$  such that

$$\beta_k T_k = \left( \Lambda \quad M_s \right), \quad \Sigma_k = \Sigma_0 - M_s M_s^\top, \quad (26)$$

where the off-diagonal elements of  $M_s M_s^\top$  are zero.

The  $m \times s$ -matrix  $M_s$  is a so-called *spurious factor loading matrix* that does not contribute to explaining the covariance in  $y_t$ , since

$$\beta_k \beta_k^\top + \Sigma_k = \beta_k T_k T_k^\top \beta_k^\top + \Sigma_k = \Lambda \Lambda^\top + M_s M_s^\top + (\Sigma_0 - M_s M_s^\top) = \Lambda \Lambda^\top + \Sigma_0 = \Omega.$$

While this theorem is an important result, Tumura and Sato (1980) did not impose conditions that resolve rotational invariance, either on the true loading matrix  $\Lambda$  or on the overfitting loading matrix  $\beta_k$ . However, without such conditions, the factor loading matrix  $\beta_k$  in an overfitting EFA model does not immediately “reveal the truth”, as the separation of  $\beta_k$  into the true factor loading matrix  $\Lambda$  and the spurious factor loading matrix  $M_s$  is possible only up to an arbitrary rotation  $T_k$  of  $\beta_k$ .

However, the “truth” in an overfitting EFA model can be recovered if Tumura and Sato (1980, Theorem 1) is applied within the class of unordered GLT structures introduced in this paper. Under the condition that  $\Lambda$  is a GLT structure which satisfies the extended

row-deletion property  $RD(r, 1 + S)$ , we prove in Theorem 4 the following result. If the factor loading matrix  $\beta_k$  in an overfitting EFA model satisfies the unordered GLT condition, then  $\beta_k$  has a representation, where the rotation in (26) reduces to a signed permutation  $T_k = P_{\pm} P_{\rho}$ . Furthermore, the spurious factor loading matrix  $M_s$  takes the form of a *spurious unordered GLT structure*, introduced in Definition 8.

**Definition 8** (Spurious unordered GLT structures). An  $m \times s$  unordered GLT structure  $M_s$  with pivots rows  $\{n_1, \dots, n_s\}$  is a *spurious unordered GLT structure* if all columns exhibit a single nonzero loading in the corresponding pivot row.

**Theorem 4.** Let  $\Lambda$  be an  $m \times r$  factor loading matrix with an unordered GLT structure with pivot rows  $l_1, \dots, l_r$  and assume that  $\Lambda$  obeys the extended row-deletion property  $RD(r, 1 + S)$  for some  $S \in \mathbb{N}$ . Assume that the  $m \times k$  matrix  $\beta_k$  in the EFA variance decomposition  $\Omega = \beta_k \beta_k^T + \Sigma_k$  is of rank  $\text{rk}(\beta_k) = k = r + s$ , where  $1 \leq s \leq S$ . If an unordered GLT condition is imposed on  $\beta_k$ , then (26) reduces to

$$\beta_k P_{\pm} P_{\rho} = \begin{pmatrix} \Lambda & M_s \end{pmatrix}, \quad \Sigma_k = \Sigma_0 - M_s (M_s)^T,$$

where  $M_s$  is a *spurious unordered GLT structure* with pivot rows  $n_1, \dots, n_s$  which are distinct from the  $r$  pivot rows in  $\Lambda$ . Hence,  $r$  columns of  $\beta_k$  are a signed permutation of the true loading matrix  $\Lambda$ , while the remaining  $s$  columns of  $\beta_k$  are a *spurious unordered GLT structure* with pivots rows  $n_1, \dots, n_s$ .

See Appendix A for a proof. Theorem 4 is employed in Section 6.3 to recover the number of factors from a Bayesian inference of the EFA model (19) using sparsity priors.

### 5.3. Identifying Irrelevant Variables

In applied factor analysis, the assumption that each measurement  $y_{it}$  is correlated with at least one other measurement is too restrictive, because irrelevant measurements might be present that are uncorrelated with all the other measurements. As argued by Boivin and Ng (2006) and Kaufmann and Schumacher (2017) for various latent factor models, it is useful to identify such variables. Within the framework of sparse Bayesian factor analysis, irrelevant variables are identified in Kaufmann and Schumacher (2017) by exploring the sparsity matrix  $\delta$  of a factor loading matrix  $\Lambda$  with respect to zero rows. Since  $\text{Cov}(y_{it}, y_{lt}) = 0$  for all  $l \neq i$  if the entire  $i$ th row of  $\Lambda$  is zero (see also (3)), the presence of  $m_0$  irrelevant measurements causes the corresponding  $m_0$  rows of  $\Lambda$  and  $\delta$  to be zero.

Let us investigate the identification of the zero rows in  $\Lambda$  and the corresponding sparsity matrix  $\delta$  for the case that the assumed and the true number of factors in the EFA model (19) are identical, i.e.,  $H = r$ . Provided that variance identification of (20) in the underlying model holds, we obtain that  $\beta_r = \Lambda P$  is a rotation of  $\Lambda$ . Therefore, the position of the zero rows both in  $\Lambda$  and  $\beta_r$  are identical and all irrelevant variables can be identified from  $\beta_r$  or the corresponding sparsity matrix  $\delta_r$ , regardless of the conditions imposed to resolve rotational invariance.

To ensure variance identification under condition AR, the loading matrix  $\Lambda$  has to satisfy the row-deletion property  $RD(r, 1)$ . If  $\Lambda$  contains  $m_0$  zero row, then a necessary condition for  $RD(r, 1)$  is that  $2r + 1 \leq m - m_0$ . This leads to a tighter bound for  $r$  than (13), namely

$$m - m_0 \geq 2r + 1 \iff r \leq \frac{m - m_0 - 1}{2}. \quad (27)$$

Hence, there is a trade-off between  $m_0$  and  $r$ : the more irrelevant measurements are included among the  $m$  measurements, the smaller the maximum number of factors  $r$  can be.



## 6. Identifying the Number of Factors in Sparse Bayesian Factor Analysis

### 6.1. Sparse Bayesian Factor Analysis

Sparse Bayesian factor analysis operates in the exploratory factor analysis (EFA) model (19), which allows up to  $H$  factors, but the true factor dimension  $r$  is unknown (see, among many others, Ročková and George (2017), Frühwirth-Schnatter and Lopes (2018), and Ohn and Kim (2022)). Often, spike-and-slab priors are employed, where the elements  $\beta_{ij}$  of the loading matrix  $\beta_H$  in the EFA model a priori are allowed to be exactly zero with positive probability. This is achieved through a suitable prior on the corresponding  $m \times H$  sparsity matrix  $\delta_H$ . In each column  $j$  of  $\delta_H$ , the binary indicators  $\delta_{ij}$  are active a priori with a column-specific probability  $\tau_j$ , i.e.,  $\Pr(\delta_{ij} = 1 | \tau_j) = \tau_j$  for  $i = 1, \dots, m$ , where the probabilities  $\tau_1, \dots, \tau_H$  arise from an exchangeable shrinkage process (ESP) prior:

$$\tau_j | H \sim \mathcal{B}\left(\gamma \frac{\alpha}{H}, \gamma\right), \quad j = 1, \dots, H. \quad (28)$$

If  $\gamma = 1$ , then (28) is a so-called one-parameter-beta (1PB) prior, otherwise (28) is a so-called two-parameter-beta (2PB) prior. The 1PB prior converges to the Indian buffet process prior (Teh et al. 2007) for  $H \rightarrow \infty$ .

This specification leads to a Dirac-spike-and-slab prior for the loadings  $\beta_{ij}$  in  $\beta_H$ ,

$$\beta_{ij} | \kappa, \sigma_i^2, \tau_j \sim (1 - \tau_j) \Delta_0 + \tau_j \mathcal{N}(0, \kappa \sigma_i^2), \quad (29)$$

where a Gaussian slab distribution is assumed and the scale of the prior depends on the idiosyncratic variance  $\sigma_i^2$  and a random global shrinkage parameter  $\kappa$ . The priors  $\sigma_i^2 \sim \mathcal{G}^{-1}(c^\sigma, b^\sigma)$  and  $\kappa \sim \mathcal{G}^{-1}(c^\kappa, b^\kappa)$  are assumed for  $\sigma_1^2, \dots, \sigma_m^2$  and  $\kappa$ . Other slab distributions are possible (see, e.g., Zhao et al. (2016) and Frühwirth-Schnatter et al. (2023)).

As shown by Frühwirth-Schnatter (2023), the ESP prior (28) has a representation as a cumulative shrinkage process (CUSP) prior (Legramanti et al. 2020). For the 1PB prior, e.g., the decreasing order statistics  $\tau_{(1)} > \dots > \tau_{(H)}$  of the slab probabilities,  $\tau_1, \dots, \tau_H$  can be expressed by the following multiplicative (stick-breaking) representation in terms of independent beta random variables, i.e., for  $j = 1, \dots, H$ :

$$\tau_{(j)} = \prod_{\ell=1}^j v_\ell, \quad v_\ell \sim \mathcal{B}\left(\alpha \frac{H - \ell + 1}{H}, 1\right), \quad \ell = 1, \dots, H. \quad (30)$$

With the largest slab probability following  $\tau_{(1)} \sim \mathcal{B}(\alpha, 1)$ , the subsequent slab probabilities  $\tau_{(j)} = \tau_{(j-1)} v_j$  are rapidly converging to zero as  $j$  increases. Hence, the columns of the loading matrix  $\beta_H$  (if permuted according to the order statistics  $\tau_{(1)} > \dots > \tau_{(H)}$ ) are increasingly pulled toward 0 and the ESP prior induces column sparsity, with the number of non-zero columns  $k$  in  $\beta_H$  being considerably smaller than  $H$  a priori.

The hyperparameters  $\alpha$  and  $\gamma$  of the ESP prior are instrumental in controlling prior column sparsity and retrieving the number of factors from an EFA model (see Section 6.3) and are learned from the data under the priors  $\alpha \sim \mathcal{G}(a^\alpha, b^\alpha)$  and  $\gamma \sim \mathcal{G}(a^\gamma, b^\gamma)$ .

### 6.2. MCMC Estimation

For a given choice of hyperparameters  $(c^\sigma, b^\sigma, c^\kappa, b^\kappa, a^\alpha, b^\alpha, a^\gamma, b^\gamma)$ , Markov chain Monte Carlo (MCMC) methods are applied to sample from the posterior distribution  $p(\beta_H, \Sigma_H, \delta_H | \mathbf{y})$ , given  $T$  multivariate observations  $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_T)$  (see, e.g., Kaufmann and Schumacher (2019) among many others). In Frühwirth-Schnatter et al. (2023), such a sampler is developed for GLT factor models. To move between factor models of different factor dimensions, Frühwirth-Schnatter et al. (2023) exploit Theorem 4 proven in the present paper to add and delete spurious columns through a reversible jump MCMC (RJMCMC) sampler. We refer to Frühwirth-Schnatter et al. (2023) for full details of this algorithm.

### 6.3. Identifying the Number of Factors

Identification of the number of factors is a notoriously difficult problem which is closely related to the intrinsic identifiability problems of overfitting EFA models discussed in Section 5.2. Any EFA model, where  $H$  is overfitting the true number of factors, may generate decompositions of  $\Omega$  which contain spurious columns. This ambiguity makes estimating the number of factors in applied factor analysis challenging.

A common procedure is to apply an incremental procedure, by increasing the maximum number of factors  $H$  step by step, and to use model selection criteria such as information criteria (Aßmann et al. 2016; Bai and Ng 2002) or Bayes factors (Lee and Song 2002; Lopes and West 2004) to choose the number of factors. Kaufmann and Schumacher (2019) estimate a sparse dynamic factor model with an increasing number  $H$  of potential factors and use so-called “extracted factor representation” during post-processing MCMC draws to select the number of factors. In sparse Bayesian analysis, the number of factors is usually estimated in one sweep jointly with all unknown parameters using shrinkage priors, such as the ESP prior (28). For instance, based on the prior multiplicative gamma process (Bhattacharya and Dunson 2011), Carvalho et al. (2008) infer  $r$  as the number columns that remain after removing columns in the loading matrix with a few nonzero elements in a fairly heuristic manner.

In the present paper, we suggest a one-sweep approach within the framework of sparse Bayesian inference of the overfitting EFA model (19). We identify the number of factors by post-processing the posterior draws of the matrix  $\beta_k$  containing the  $k$  nonzero columns of the loading matrix  $\beta_H$ , where we rely on the insights gained in Section 5.2 regarding the role of variance identification and spurious columns.<sup>6</sup>

Analyzing the problem from the viewpoint of variance identification is helpful in understanding some of the fundamental difficulties. If the number of factors is unknown, then we need to find a decomposition of  $\Omega$  as in (21), where  $\beta_k \beta_k^\top$  is identified. If variance identification holds, then (20) and (21) are equivalent and unique. Therefore,  $k = r$ , and we can identify the true loading matrix  $\Lambda = \beta_r \mathbf{P}$  from  $\beta_r$  up to a rotation  $\mathbf{P}$ . On the other hand, for any decomposition (21) which is not variance identified, we can deduce that  $k$  is bigger than the true number of factors.

In an overfitting EFA model, many posterior draws with  $k$  nonzero columns will have a representation as in Theorem 3 or Theorem 4, and contain a submatrix  $\mathbf{M}_s$  with  $s$  spurious columns. Hence, these draws violate even the most simple conditions for variance identification. As a consequence, the number of non-zero columns  $k$  overestimates  $r$  since  $k = r + s$  or, equivalently,  $r = k - s$ . These insights show that verifying variance identification is essential for recovering the true number of factors, and this has implications for applied factor analysis. Most importantly, methods of inferring the number of factors from the *rank* or the *number of non-zero columns* of the posterior draws  $\beta_H$  in an overfitting factor model are prone to overestimate the number of factors.

In addition, we operate under the GLT condition and rely on the mathematically justified representation of  $\beta_k$  in an overfitting EFA model provided by Theorem 4. Under an unordered GLT condition, spurious columns among the non-zero columns  $\beta_k$  are easily spotted, as the  $m \times s$  spurious factor loading matrix  $\mathbf{M}_s$  has an extremely simple structure with a single non-zero loading in each column.<sup>7</sup> Furthermore, the remaining columns  $\beta_r$  are an unordered GLT representation of the true factor loading matrix  $\Lambda$ , if variance identification holds and  $\beta_r \beta_r^\top$  is identified.

These considerations imply the following strategy for postprocessing the posterior draws obtained by the RJMCMC procedure of Frühwirth-Schnatter et al. (2023). For each posterior draw of the non-zero columns  $\beta_k$  of  $\beta_H$ , the active columns  $\beta_r$  (i.e., all columns with at least two non-zero elements) and the corresponding sparsity matrix  $\delta_r$  are determined. If  $\delta_r$  satisfies the counting rule  $\text{CR}(r, 1)$ , then  $\beta_r$  is a signed permutation of  $\Lambda$  by virtue of Theorem 4. The corresponding error covariance matrix is equal to  $\Sigma_r = \Sigma_k + \mathbf{M}_s (\mathbf{M}_s)^\top$ , where  $\mathbf{M}_s$  contains the spurious columns of  $\beta_k$ . These variance identified draws are kept for further inference and the number of columns of  $\beta_r$  is considered a

posterior draw of the unknown factor dimension  $r$ . The posterior distribution  $p(r|\mathbf{y})$  can be estimated from these draws and the posterior mode provides a point estimator of the number of factors  $r$ .

The proposed strategy of recovering the number of factors through a sparse Bayesian analysis of an overfitting EFA model is illustrated in Section 7 for simulated as well as real data (see also Frühwirth-Schnatter et al. (2023) for further applications).

## 7. Illustrative Applications

### 7.1. An Illustrative Simulation Study

For illustration, we perform a simulation study with  $m = 30$  and  $T = 150$ , and consider three different scenarios for the factor loading matrix  $\Lambda$ . In all three scenarios,  $r_{\text{true}} = 5$  factors are assumed, but the sparsity matrix  $\delta$  is quite different. The first setting is a *dedicated* factor model, where the first six variables load on factor 1, the next six variables load on factor 2, and so forth, and the final six variables load on factor 5. The second scenario is a *block* factor model, where the first 15 observations load only on factors 1 and 2, while the remaining 15 observations only load on factors 3, 4, and 5, and the covariance matrix has a block-diagonal structure. All factor loadings within a block are non-zero. The third scenario is a *dense* factor loading matrix without any zero loadings. For all three scenarios, non-zero factor loadings are drawn as  $\lambda_{ij} = (-1)^{b_{ij}}(1 + 0.1\mathcal{N}(0, 1))$ , where the exponent  $b_{ij}$  is a binary variable with  $\Pr(b_{ij} = 1) = 0.2$ . With the exception of the first scenario, no GLT condition is imposed on the simulated loading matrix. In all three scenarios,  $\Sigma_0 = \mathbf{I}$ . A total of 21 data sets are sampled under these three scenarios from the basic factor model (1).

The overfitting exploratory factor model (19) with the maximum number of factors  $H = 14$  being equal to the upper bound defined in (13) is fitted to each simulated data set. Regarding rotational invariance, we compare a model where the non-zero columns of the factor loading matrix  $\beta_k$  extracted from  $\beta_H$  obey an unordered GLT structure with a model where  $\beta_H$  is left unconstrained.

Inference is based on the Bayesian approach described in Section 6.1. We consider both a 1PB and a 2PB shrinkage prior on the sparsity matrix  $\delta_H$ , and select the following priors:  $\sigma_i^2 \sim \mathcal{G}^{-1}(2.5, 1.5)$ ,  $i = 1, \dots, 30$ ,  $\kappa \sim \mathcal{G}^{-1}(5, 25)$ ,  $\alpha \sim \mathcal{G}(6, 2)$  and  $\gamma \sim \mathcal{G}(6, 6)$ . Using the RJMCMC algorithm of Frühwirth-Schnatter et al. (2023), 3000 posterior draws are generated after a burn-in of 2000 draws.<sup>8</sup> As discussed in Section 6.3, the active columns  $\beta_r$  are retrieved for each MCMC draw of the loading matrix  $\beta_H$ . Under the GLT condition,  $\beta_r$  is checked for variance identification using the counting rule  $\text{CR}(r, 1)$ . For sparse Bayesian factor analysis with unstructured loading matrices, the draws of  $\beta_r$  are not screened for variance identification and inference is based on all draws.

For each of the 21 simulated data sets, we evaluate all 12 combinations of data scenarios, structural constraints (GLT versus unconstrained) and priors on the sparsity matrix (1PB versus 2PB) through Monte Carlo estimates of the following statistics: to assess the performance in estimating the true number  $r_{\text{true}}$  of factors, we consider the mode  $\hat{r}$  of the posterior distribution  $p(r|\mathbf{y})$  as a point estimator of  $r$ . In addition, we consider the magnitude of the posterior ordinate  $p(\hat{r} = r_{\text{true}}|\mathbf{y})$  as a measure of how strongly the posterior distribution concentrates around the true value. The closer this value is to 1, the smaller is the uncertainty in estimating  $r$ . To assess the accuracy in estimating the true covariance matrix  $\Omega_0 = \Lambda\Lambda^\top + \Sigma_0$  of the data via the posterior draws of  $\Omega_r = \beta_r\beta_r^\top + \Sigma_r$ , we consider the mean squared error (MSE) defined by

$$\text{MSE}_\Omega = \sum_i \sum_{\ell \leq i} \mathbb{E}((\Omega_{r,i\ell} - \Omega_{0,i\ell})^2 | \mathbf{y}) / (m(m+1)/2),$$

which accounts both for variance and bias. Table 1 reports, for all 12 combinations, the median, the 5% and the 95% quantile of these statistics across all simulated data sets. For GLT structures, the fraction of variance identified draws is also reported and is, in general, pretty high.

**Table 1.** Sparse Bayesian factor analysis under GLT and unconstrained structures (EFA) under a 1PB prior ( $\alpha \sim \mathcal{G}(6, 2)$ ) and a 2PB prior ( $\alpha \sim \mathcal{G}(6, 2), \gamma \sim \mathcal{G}(6, 6)$ ). The true number of factors is equal to 5. GLT and EFA-V use only the variance identified draws ( $M_V$  is the percentage of variance identified draws), EFA uses all posterior draws.

Scenario		Prior	$M_V$ Med (QR)	$\hat{r}$ Med (QR)	$p(\hat{r} = r_{\text{true}} \mathbf{y})$ Med (QR)	$\text{MSE}_\Omega$ Med (QR)
Dedic	GLT	1PB	97.0 (91.5, 98.3)	5 (5, 5)	0.90 (0.94, 0.99)	0.018 (0.014, 0.030)
		2PB	97.6 (87.7, 98.9)	5 (5, 5)	0.99 (0.83, 1.00)	0.019 (0.016, 0.027)
	EFA	1PB	-	5 (5, 6)	0.66 (0.09, 0.79)	0.020 (0.015, 0.026)
		2PB	-	5 (5, 6)	0.69 (0.36, 0.80)	0.019 (0.014, 0.024)
	EFA-V	1PB	80.3 (49.8, 87.0)	5 (5, 6)	0.81 (0.17, 0.91)	0.020 (0.015, 0.026)
		2PB	82.6 (63.4, 87.9)	5 (5, 6)	0.84 (0.53, 0.92)	0.019 (0.014, 0.024)
Block	GLT	1PB	96.5 (39.4, 98.9)	5 (5, 5)	0.99 (0.28, 0.99)	0.12 (0.08, 0.18)
		2PB	98.7 (61.9, 99.4)	5 (5, 5)	0.99 (0.54, 1.00)	0.10 (0.08, 0.14)
	EFA	1PB	-	5 (4, 5)	0.78 (0.22, 0.88)	0.14 (0.11, 0.20)
		2PB	-	5 (4, 5)	0.79 (0.08, 0.89)	0.12 (0.08, 0.24)
	EFA-V	1PB	87.0 (55.0, 91.5)	5 (4, 5)	0.89 (0.09, 0.96)	0.14 (0.11, 0.20)
		2PB	85.9 (28.3, 90.4)	5 (4, 5)	0.92 (0.03, 0.97)	0.12 (0.08, 0.24)
Dense	GLT	1PB	95.7 (84.6, 98.6)	5 (5, 5)	0.98 (0.92, 0.99)	0.67 (0.44, 1.12)
		2PB	99.4 (90.8, 99.8)	5 (5, 5)	0.99 (0.93, 1.00)	0.68 (0.51, 1.18)
	EFA	1PB	-	5 (5, 6)	0.76 (0.43, 0.85)	0.54 (0.39, 0.76)
		2PB	-	5 (5, 5)	0.80 (0.66, 0.91)	0.59 (0.43, 0.90)
	EFA-V	1PB	84.4 (76.0, 90.2)	5 (5, 6)	0.89 (0.57, 0.95)	0.54 (0.39, 0.76)
		2PB	89.7 (80.4, 93.9)	5 (5, 5)	0.93 (0.77, 0.98)	0.59 (0.43, 0.90)

Med is the median and QR are the 5% and the 95% quantile of the various statistics over the 21 simulated data sets.

Several conclusions can be drawn from Table 1. First, of all, sparse Bayesian factor analysis under the GLT constraint successfully recovers the true number of factors in all three scenarios. For most of the simulated data sets, the posterior ordinate  $p(\hat{r} = r_{\text{true}}|\mathbf{y})$  is larger than 0.9. Sparse Bayesian factor analysis with unstructured loading matrices is also quite successful in recovering  $r_{\text{true}}$ , but with less confidence. Both over- and underfitting is observed, and  $p(\hat{r} = r_{\text{true}}|\mathbf{y})$  is much smaller than under a GLT structure. For both structures, the 2PB prior yields higher posterior ordinates than the 1PB prior.

Recently, [Hosszejni and Frühwirth-Schnatter \(2022\)](#) proved that the counting rule  $\text{CR}(r, 1)$  can also be applied to verify variance identification of  $\beta_r$  for unconstrained loading matrices. As is evident from Table 1, the fraction of variance identified draws is, however, much smaller than under GLT structures. Nevertheless, inference with respect to the number of factors can be improved also for an unconstrained EFA model by rejecting all draws of  $\beta_r$  that do not obey the counting rule  $\text{CR}(r, 1)$ .

## 7.2. A Real Data Example

For further illustration, we consider monthly log returns from the New York Stock Exchange (NYSE) for two industry sectors. For each month  $t = 1, \dots, T$  from February 1999 till August 2019 (i.e.,  $T = 247$ ), we consider  $m = 28$  firms. The first 10 measurements in  $\mathbf{y}_t$  correspond to all firms in the energy sector, the remaining 18 measurements to all firms in the health care sector.<sup>9</sup> We compare sparse Bayesian factor analysis with maximum likelihood estimation and to this aim demean and standardize the data.

Bayesian inference is based on the approach described in Section 6.1. The overfitting exploratory factor model (19) is fitted to this data set under an unordered GLT structure with the maximum number of factors  $H = 13$  being equal to the upper bound defined in (13). MCMC estimation is run for 30,000 iterations after a burn-in of 20,000 using the

RJMCMC algorithm of Frühwirth-Schnatter et al. (2023). Prior choices are exactly as in Section 7.1 and we report results for the 2PB prior.

As discussed in Section 6.3, the active columns  $\beta_r$  are retrieved for each MCMC draw of the loading matrix  $\beta_H$  and checked for variance identification using the counting rule  $CR(r, 1)$ . The fraction of variance identified MCMC draws for the 2PB prior is equal to 63%. The number of columns for each variance identified draw  $\beta_r$  changes due to the dimension-changing nature of the RJMCMC algorithm and can be considered a posterior draw of the number of active factors  $r$  by virtue of Theorem 4. As shown in Table 2, the posterior distribution  $p(r|y)$  derived in this manner is rather disperse and models from four up to six factors receive considerable posterior probability.

**Table 2.** Posterior distribution  $p(r|y)$  of the number of factors based on all variance identified posterior draws.

$r$	0–3	4	5	6	7	8	9–13
$p(r y)$	0	0.28	0.45	0.25	0.02	0.001	0

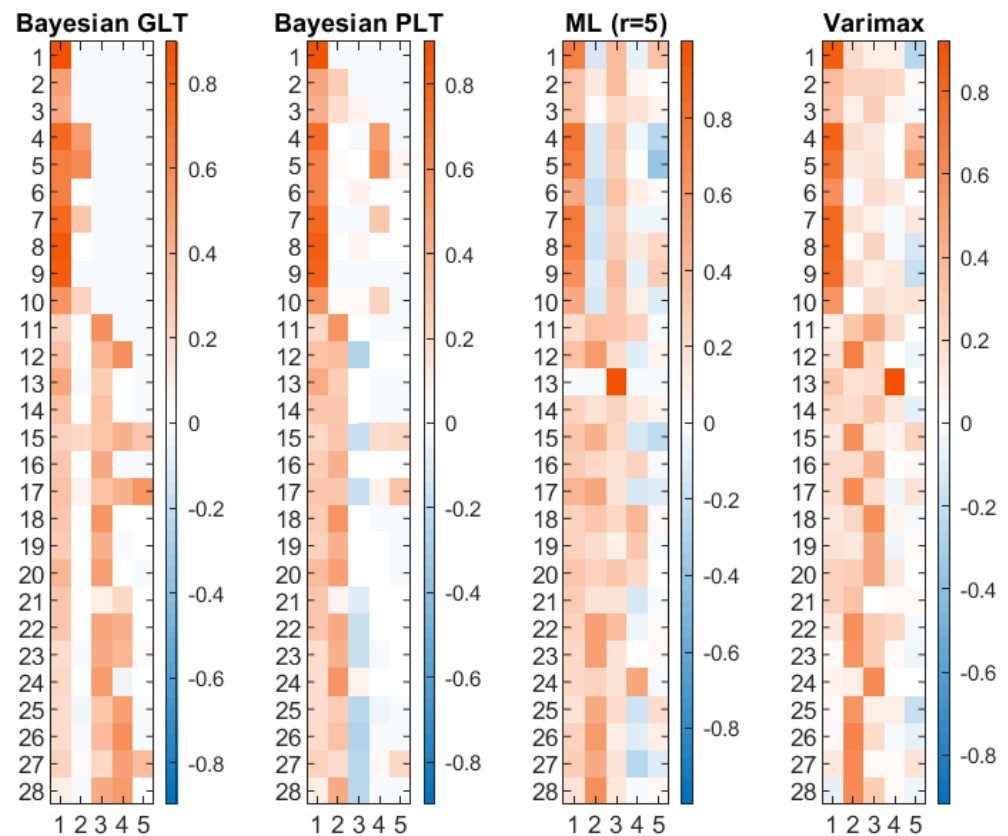
We proceed with the posterior mode estimator of the number of factors and compare various estimators of the factor loading matrix  $\Lambda$  for a basic factor model (1) where  $r = 5$  is assumed. First, we estimate  $\hat{\Lambda}_{ML}$  using maximum likelihood methods and apply the Varimax procedure to rotate  $\hat{\Lambda}_{ML}$  into a simple structure  $\hat{\Lambda}_{VM}$ .<sup>10</sup>

Second, we derive two Bayesian estimators where we impose, respectively, the PLT and the GLT condition. Under the GLT condition we learn the pivot rows  $\mathbf{l} = (l_1, \dots, l_5)$  from the data, while  $\mathbf{l} = (1, \dots, 5)$  under the PLT condition. To identify simple PLT and GLT structures, we perform variable selection beyond the pivot rows and assume the uniform prior  $\tau_j \sim \mathcal{B}(1, 1)$  (instead of 1PB or 2PB shrinkage priors), since the number of factors is known. The prior on the factor loadings and the idiosyncratic variances is the same as before. 10,000 posterior draws are generated by adjusting the algorithm of Frühwirth-Schnatter et al. (2023) to these models. The draws are screened for variance identification using the counting rule  $CR(5, 1)$ . The average of the variance identified draws immediately yields the Bayesian estimator  $\hat{\Lambda}_{PLT}$  under the PLT condition.

Under the GLT condition, the sampler operates within an unordered GLT structure with unknown pivot rows. All variance identified posterior draws are identified up to a signed permutation. Full identification is achieved by resolving column and sign switching for each posterior draw. Column switching is resolved by ordering the pivot rows such that  $l_1 < \dots < l_5$  and an ordered GLT structure is imposed on  $\Lambda$  by reordering the columns accordingly. Among these ordered GLT posterior draws of  $\Lambda$ , the combination of pivot rows  $\hat{\mathbf{l}} = (1, 4, 11, 12, 15)$  is visited most often. All ordered GLT draws where the pivot rows coincide with  $\hat{\mathbf{l}}$  are averaged to derive the Bayesian estimator  $\hat{\Lambda}_{GLT}$  of the GLT factor loading matrix  $\Lambda$ . Beforehand, sign switching in the posterior draws is resolved by imposing the constraint  $\Lambda_{11} > 0$ ,  $\Lambda_{42} > 0$ ,  $\Lambda_{11,3} > 0$ ,  $\Lambda_{12,4} > 0$  and  $\Lambda_{15,5} > 0$  on  $\Lambda$ .

All four estimators are depicted in Figure 2. Note that the measurements are grouped by industry. The pivot rows learned under GLT from the data clearly show that this ordering is in conflict with the PLT assumption. The varimax estimator  $\hat{\Lambda}_{VM}$ , as well as both Bayesian estimators  $\hat{\Lambda}_{PLT}$  and  $\hat{\Lambda}_{GLT}$ , exhibit a simple structure, and sparsity is more pronounced for the two Bayesian estimators due to variable selection on the factor loadings.





**Figure 2.** Comparing estimators of the loading matrix  $\Lambda$  for the NYSE data (1–10, firms in the energy sector; 11–28, firms in the health care sector) for a basic factor model with  $r = 5$  factors. Heatmaps from left to right: Bayesian estimator  $\hat{\Lambda}_{\text{GLT}}$  of the GLT representation of the loading matrix; Bayesian estimator  $\hat{\Lambda}_{\text{PLT}}$  obtained by imposing a PLT condition; ML estimator  $\hat{\Lambda}_{\text{ML}}$  and the corresponding Varimax rotation  $\hat{\Lambda}_{\text{VM}}$ . Red, white, and blue colors denote positive, zero, and negative values, respectively.

In our opinion, the Bayesian GLT estimator  $\hat{\Lambda}_{\text{GLT}}$  allows a clearer interpretation of the four factors than the other estimators. In particular, forcing a PLT assumption leads to a less clear interpretation of the factors. The first factor is a market factor that loads on all 28 firms. The second factor captures additional correlations among the 10 firms in the energy sector, as well as cross-sectional correlations with specific firms in the health care sector. Three additional factors are present that are specific to firms in the health care sector, with factor three loading on nearly all firms in this sector.

## 8. Concluding Remarks

We have given a full and comprehensive mathematical treatment to generalized lower triangular (GLT) structures, an identification strategy that relaxes the popular positive lower triangular (PLT) assumption for factor loadings matrices. We have proven that GLT retains PLT's good properties: uniqueness and rotational invariance. At the same time, and unlike PLT, a GLT structure exists for any factor loadings matrix; i.e., it is not a restrictive assumption. Furthermore, we have shown that verifying variance identification under GLT structures is simple and is based purely on the zero-nonzero pattern of the factor loadings matrix. Additionally, we have embedded the GLT model class into exploratory factor analysis with unknown factor dimension and discussed how easily spurious factors and irrelevant variables are recognized in that setup. At the end, we demonstrated our framework in a simulation study and for real data.

**Author Contributions:** Conceptualization, S.F.-S., D.H. and H.F.L.; methodology, S.F.-S., D.H. and H.F.L.; software, S.F.-S. and D.H.; formal analysis, S.F.-S. and D.H.; investigation, S.F.-S. and D.H.; data curation, S.F.-S. and D.H.; writing—original draft preparation, S.F.-S. and D.H.; writing—review and editing, S.F.-S., D.H. and H.F.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** Hedibert Freitas Lopes receives support from FAPESP Grant 2018/04654-9.

**Data Availability Statement:** Restrictions apply to the availability of these data. Data were obtained from Bloomberg and are available from the authors if Bloomberg gives permission.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Appendix A

**Proof of Theorem 1.** First, note that  $\mathbf{G}$ , the result of the procedure described in part (b), exists: there exists a set of  $r$  linearly independent rows in  $\beta$  due to its rank  $r$ , and therefore the described procedure always succeeds at finding row indices  $(l_1, \dots, l_r)$ . We show that  $\Lambda = \beta \mathbf{G}$  is an ordered GLT matrix with pivot rows  $(l_1, \dots, l_r)$  and positive leading elements. This will prove both part (a), which describes that a rotation such as  $\mathbf{G}$  exists, and part (b), which describes how said  $\mathbf{G}$  is constructed.

Let the QR-decomposition of  $\tilde{\beta}^\top$  be  $\tilde{\beta}^\top = \mathbf{G}\tilde{\Lambda}^\top$ , where  $\tilde{\Lambda}$  is the  $r \times r$  lower triangular submatrix of  $\Lambda$  that consists of, clearly, the  $l_1$ th,  $\dots$ ,  $l_r$ th rows of  $\Lambda$ . We choose  $\mathbf{G}$  such that  $\tilde{\Lambda}^\top$  has positive diagonal elements; we can do so according to Golub and Van Loan (2013), and this way  $\mathbf{G}$  is also unique. We proceed column-by-column in  $\Lambda$ , starting with the first column. Note that  $\Lambda_{l_1,1} = \tilde{\Lambda}_{1,1}$  is positive, therefore  $l_1$  is a candidate for a pivot row. Next, assume that  $l_1$  is not a pivot row because  $\Lambda_{i,1}$  has a nonzero entry  $\Lambda_{i,1}$  for  $i < l_1$ . This implies that  $\beta_{i,\bullet}$  contains a nonzero entry, which contradicts how  $l_1$  is found by the procedure. Therefore,  $l_1$  is the pivot row for the first column.

Now, we examine the second column of  $\Lambda$ . Note that  $\Lambda_{l_2,2} = \tilde{\Lambda}_{2,2}$  is positive, so  $l_2$  is a candidate for a pivot row. Also note that  $\Lambda_{l_1,2}$  is zero because it is the second element of  $\tilde{\Lambda}_{1,\bullet}$ . Next, assume that  $l_2$  is not a pivot row because  $\Lambda_{i,2}$  has a nonzero entry  $\Lambda_{i,2}$  for  $i < l_2$ . This implies that  $\Lambda_{i,\bullet}$  is linearly independent from  $\Lambda_{l_1,\bullet}$ , which implies that  $\beta_{i,\bullet}$  is linearly independent from  $\beta_{l_1,\bullet}$ , which contradicts how  $l_2$  is found by the procedure. Therefore,  $l_2 > l_1$  is the pivot row for the second column.

We proceed in this manner, applying the observations for the second column to the third and later columns, until we find that  $l_1 < l_2 < \dots < l_r$  are the pivot rows of  $\Lambda$  and  $\Lambda$  is ordered GLT. This concludes the proof.  $\square$

**Lemma A1.** The counting rule  $CR(r, s)$  has the following properties:

- (a)  $CR(r, s)$  holds for  $\delta$  if  $CR(q, s)$  holds for every submatrix of  $q \in \{1, \dots, r\}$  columns of  $\delta$ ;
- (b) If  $CR(r, s)$  holds for  $\delta$  and arbitrary  $\tilde{s} \leq s$  rows are deleted from  $\delta$ , then the remaining matrix satisfies  $CR(r, s - \tilde{s})$ ;
- (c) If  $CR(r, s)$  holds for  $\delta$  and some or all zero rows are removed from  $\delta$ , then  $CR(r, s)$  also holds for the remaining matrix.

The proof is straightforward.

**Proof of Theorem 2.** Any matrix  $\beta \in \Theta_\delta$  has the same non-zero rows as  $\delta$ . Hence, if  $CR(r, s)$  does not hold for  $\delta$ , then it also does not hold for any  $\beta \in \Theta_\delta$ . According to Theorem 3.4. by Sato (1992) with their rotation  $\mathbf{G}$  being the identity, this implies that  $RD(r, s)$  is violated for all  $\beta \in \Theta_\delta$ . This proves part (a).

We prove part (b) by induction. If  $CR(1, s)$  holds for a  $m \times 1$  sparsity vector  $\delta$ , then at least  $m_1 \geq 2 + s$  elements of  $\delta$  are different from 0. It trivially follows that all  $\beta \in \Theta_\delta$  have same number of non-zero elements. After deleting  $s$  elements, two subvectors with at least one non-zero element can be formed and  $RD(1, s)$  is satisfied for  $\beta$ . For any  $r \geq 2$ , assume

that part (b) of Theorem 2 holds for  $r - 1$  and that the counting rule  $\text{CR}(r, s)$  holds for an  $m \times r$  ordered GLT sparsity matrix  $\delta$ . A suitable permutation of the rows of  $\delta$  yields:

$$\Pi_r \delta = \begin{pmatrix} \delta^c & \mathbf{0} \\ \delta^b & \delta^A \end{pmatrix},$$

where  $\delta^A$  is a GLT sparsity matrix with  $r - 1$  columns,  $\delta^b$  and  $\delta^c$  are column vectors, and  $\delta^c$  contains  $d_1 \geq 1$  non-zero elements and no zero elements. According to Lemma A1  $\delta^A$  satisfies  $\text{CR}(r - 1, s)$  and the first column  $((\delta^c)^\top, (\delta^b)^\top)^\top$  satisfies  $\text{CR}(1, s)$ . Consequently,  $\delta^b$  contains at least  $2 + s - d_1$  non-zero elements. Let  $\Lambda \in \Theta_\delta$  be an ordered GLT matrix. If the same  $s$  rows are deleted from  $\Pi_r \delta$  and  $\Pi_r \Lambda$ , we obtain the following matrices:

$$\tilde{\delta} = \begin{pmatrix} \tilde{\delta}^c & \mathbf{0} \\ \tilde{\delta}^b & \tilde{\delta}^A \end{pmatrix}, \quad \tilde{\Lambda} = \begin{pmatrix} \tilde{\Lambda}^c & \mathbf{0} \\ \tilde{\Lambda}^b & \tilde{\Lambda}^A \end{pmatrix},$$

where  $0 \leq s_1 \leq \min(d_1, s)$  non-zero elements are deleted from the vector  $\delta^c$  and  $d_1 - s_1$  non-zero elements remain in the vectors  $\tilde{\delta}^c$  and  $\tilde{\Lambda}^c$ , while the vectors  $\tilde{\delta}^b$  and  $\tilde{\Lambda}^b$  contain  $d_2 \geq \max(0, 2 + s_1 - d_1)$  non-zero elements. Since we removed  $s - s_1$  rows from  $\delta^A$ , according to Lemma A1(b), the sparsity matrix  $\tilde{\delta}^A$  satisfies  $\text{CR}(r - 1, s_1)$  and, hence,  $\tilde{\Lambda}^A$  obeys  $\text{RD}(r - 1, s_1)$  except for a set of measure 0. We proceed with those matrices  $\tilde{\Lambda}^A$  where  $\text{RD}(r - 1, s_1)$  holds. If further  $s_1$  rows are deleted from  $\tilde{\Lambda}^A$ , then a matrix results which contains two sub matrices  $\mathbf{A}_1$  and  $\mathbf{A}_2$  of rank  $r - 1$ . Let the  $s_1 \times (r - 1)$  matrix  $\mathbf{B}$  contain the rows that were deleted from of  $\tilde{\Lambda}^A$ . If the same rows are deleted from  $\tilde{\Lambda}^b$ , then the vector  $\mathbf{b}$  containing the deleted elements has at least  $\max(0, 2 - (d_1 - s_1))$  non-zero elements. Next, we consider three cases. First, if  $d_1 - s_1 \geq 2$ , then we use two of the  $d_1 - s_1$  non-zero elements of  $\tilde{\delta}^c$  to define the following submatrices of  $\Lambda$ :

$$\begin{pmatrix} \tilde{\delta}_{i_1}^c & \mathbf{0} \\ \times & \mathbf{A}_1 \end{pmatrix}, \quad \begin{pmatrix} \tilde{\delta}_{i_2}^c & \mathbf{0} \\ \times & \mathbf{A}_2 \end{pmatrix}. \quad (\text{A1})$$

Both matrices obviously have rank  $r$ . Second, if  $d_1 - s_1 = 1$ , then we use the only non-zero element of  $\tilde{\delta}^c$  and one of the non-zero elements of  $\mathbf{b}$ , denoted by  $\mathbf{b}_{i_2}$ , and the corresponding row  $\mathbf{B}_{i_2, \bullet}$  of  $\mathbf{B}$  to define the following submatrices of  $\Lambda$ :

$$\begin{pmatrix} \tilde{\delta}_{i_1}^c & \mathbf{0} \\ \times & \mathbf{A}_1 \end{pmatrix}, \quad \begin{pmatrix} \mathbf{b}_{i_2} & \mathbf{B}_{i_2, \bullet} \\ \times & \mathbf{A}_2 \end{pmatrix}. \quad (\text{A2})$$

The first matrix obviously has rank  $r$ . The rank of the second matrix is at least equal to  $\text{rk}(\mathbf{A}_2) = r - 1$ . The row vector  $\mathbf{B}_{i_2, \bullet}$  contains  $0 \leq d_3 \leq r - 1$  non-zero elements, which take arbitrary values in  $\mathbb{R}$ . Hence, the set of matrices where the row  $(\mathbf{b}_{i_2} \mathbf{B}_{i_2, \bullet})$  is linearly dependent on the other rows and rank deficiency occurs has measure zero. Finally, if  $d_1 - s_1 = 0$ , then we use two of the at least two non-zero elements in  $\mathbf{b}$ , denoted by  $\mathbf{b}_{i_1}$  and  $\mathbf{b}_{i_2}$ , and the corresponding rows  $\mathbf{B}_{i_1, \bullet}$  and  $\mathbf{B}_{i_2, \bullet}$  of  $\mathbf{B}$  to define the following submatrices of  $\Lambda$ :

$$\begin{pmatrix} \mathbf{b}_{i_1} & \mathbf{B}_{i_1, \bullet} \\ \times & \mathbf{A}_1 \end{pmatrix}, \quad \begin{pmatrix} \mathbf{b}_{i_2} & \mathbf{B}_{i_2, \bullet} \\ \times & \mathbf{A}_2 \end{pmatrix}. \quad (\text{A3})$$

Using the same argument as above, both matrices are of rank  $r$ , except for a set of measure 0. This proves that  $\text{RD}(r, s)$  holds for all GLT matrices  $\Lambda \in \Theta_\delta$ , except for a set of measure 0. The counting rule  $\text{CR}(r, s)$  is invariant to signed permutations of  $\delta$ . Therefore, if  $\text{CR}(r, s)$  implies  $\text{RD}(r, s)$  for an ordered GLT matrix  $\Lambda \in \Theta_\delta$ , then this holds for all signed permutations  $\beta = \Lambda \mathbf{P} \pm \mathbf{P}_\rho$  of  $\Lambda$ . This completes the proof of part (b), since the set where  $\text{RD}(r, s)$  does not hold is a finite union of sets of measure 0.  $\square$

**Proof of Corollary 3.** The conditions in Corollary 3 follow immediately from Theorem 2. The  $(j, l)$ th elements of the matrix on the left hand of (16) is given by  $d_j + \sum_{i=1}^m \delta_{ij}(1 - \delta_{il})$ , where  $d_j = \sum_{i=1}^m \delta_{ij}$  is the total number of non-zero indicators in column  $j$ . The diagonal elements ( $j = l$ ) are equal to  $d_j$  (since  $\delta_{ij}(1 - \delta_{ij}) = 0$ ) and check if each column contains at least  $2 + s$  non-zero indicators. The off-diagonal elements ( $j \neq l$ ) count the number of nonzero rows in columns  $j$  and  $l$ . Hence, the matrix on the right hand of (16) has diagonal elements equal to  $2 + s$  and off-diagonal elements equal to  $4 + s$ . The column vector  $\delta^*$  in (17) is equal to the number of non-zero indicators in each row. Hence, (17) verifies if the total number of nonzero rows of  $\delta$  is at least equal to  $2r + s$ . Finally, (18) verifies if each submatrix of  $r - 1$  columns has at least  $2r - 1$  nonzero rows. The  $j$ th column of the matrix  $\delta^*$  in (18) is the number of non-zero indicators in each row of the submatrix  $\delta_{-j}$  excluding the  $j$ th column. The matrix  $\mathbb{I}(\delta^* > 0)$  indicates nonzero rows in  $\delta_{-j}$  and the  $j$ th element of the row vector  $\mathbf{1}_{1 \times m} \cdot \mathbb{I}(\delta^* > 0)$  counts the number of nonzero rows in  $\delta_{-j}$ .  $\square$

**Proof of Theorem 4.** We start by proving further properties of the spurious factor matrix  $\mathbf{M}_s$  in representation (26) beyond the characterization given in Tumura and Sato (1980, Theorem 1). More specifically, we show that the spurious cross-covariance matrix  $\mathbf{M}_s \mathbf{M}_s^\top = \mathbf{D}_s$  is equal to a diagonal matrix of rank  $s$ , with  $s$  nonzero entries  $d_{n_1}, \dots, d_{n_s}$  in rows  $n_1, \dots, n_s$ . From  $\text{rk}(\beta_k \mathbf{T}_k) = \min(\text{rk}(\beta_k), \text{rk}(\mathbf{T}_k)) = r + s$ , we obtain that  $\mathbf{M}_s$  has full column rank  $\text{rk}(\mathbf{M}_s) = s$ . Therefore,  $\text{rk}(\mathbf{D}_s) = \text{rk}(\mathbf{M}_s) = s$  and only  $s$  diagonal elements  $d_{n_1}, \dots, d_{n_s}$  of  $\mathbf{D}_s$  in rows  $n_1, \dots, n_s$  are different from 0. It follows that  $\mathbf{M}_s$  has exactly the same  $s$  nonzero rows  $n_1, \dots, n_s$  as  $\mathbf{D}_s$ : using for each row  $\mathbf{M}_{i\cdot}$  of  $\mathbf{M}_s$  that  $\mathbf{M}_{i\cdot} \mathbf{M}_{i\cdot}^\top = \|\mathbf{M}_{i\cdot}\|_2^2 = d_i$ , it follows for any  $i \notin \{n_1, \dots, n_s\}$  that  $\|\mathbf{M}_{i\cdot}\|_2^2 = 0$  and, therefore,  $\mathbf{M}_{i\cdot} = \mathbf{0}$ , whereas the remaining rows with  $i \in \{n_1, \dots, n_s\}$  are nonzero, since  $\|\mathbf{M}_{i\cdot}\|_2^2 > 0$ . The submatrix  $\mathbf{M}_0$  of nonzero rows in  $\mathbf{M}_s$  satisfies  $\mathbf{M}_0 \mathbf{M}_0^\top = \mathbf{D}_0^2$  with  $\mathbf{D}_0^2 = \text{Diag}(d_{n_1}, \dots, d_{n_s})$  being a diagonal matrix of rank  $s$ . It follows that  $\mathbf{D}_0^{-1} \mathbf{M}_s \mathbf{D}_0^{-1} \mathbf{M}_s^\top = \mathbf{I}$ , hence  $\mathbf{D}_0^{-1} \mathbf{M}_s = \mathbf{Q}$  for any arbitrary rotation matrix  $\mathbf{Q}$  of rank  $s$ . Therefore,  $\mathbf{M}_0 = \mathbf{D}_0 \mathbf{Q}$ .

These results allow following representation of the  $\beta_k \mathbf{T}_k$  in (26). Let  $\beta_k^*, \Sigma_k^*, \mathbf{M}_s^*, \Lambda^*$ , and  $\Sigma_0^*$ , be the matrices that result from deleting the rows  $n_1, \dots, n_s$  (and for  $\Sigma_k$  and  $\Sigma_0$  also the columns) from the matrices  $\beta_k, \Sigma_k, \mathbf{M}_s, \Lambda$ , and  $\Sigma_0$ . Since  $\mathbf{M}_s^* = \mathbf{O}$ , we obtain:

$$\beta_k^* \mathbf{T}_k = \begin{pmatrix} \Lambda^* & \mathbf{O} \end{pmatrix}, \quad \Sigma_k^* = \Sigma_0^*. \quad (\text{A4})$$

Condition  $\text{RD}(r, 1 + S)$  for  $\Lambda$  implies that  $\Lambda^*$  satisfies condition  $\text{RD}(r, 1)$ , and the variance decomposition  $\Omega^* = \Lambda^* (\Lambda^*)^\top + \Sigma_0^*$  is unique. Hence,  $\beta_k^* (\beta_k^*)^\top = \Lambda^* (\Lambda^*)^\top$  and  $\beta_k^*$  has reduced rank  $\text{rk}(\beta_k^*) = \text{rk}(\Lambda^*) = r$ . Regarding the  $s$  rows  $\beta_k^d$  and  $\Lambda^d$  that were deleted from, respectively,  $\beta_k$  and  $\Lambda$ , we obtain

$$\beta_k^d \mathbf{T}_k = \begin{pmatrix} \Lambda^d & \mathbf{M}_0 \end{pmatrix}, \quad \mathbf{M}_0 = \mathbf{D}_0 \mathbf{Q}, \quad (\text{A5})$$

where  $\mathbf{D}_0$  is a diagonal matrix and  $\mathbf{Q}$  is an arbitrary rotation matrix, both of rank  $s$ . These results are valid regardless of the conditions imposed on  $\beta_k$  to resolve rotational invariance (if any). If an unordered GLT condition is imposed, then it can be shown that  $\mathbf{M}_0$  is a spurious unordered GLT matrix and  $\mathbf{T}_k$  reduces to a signed permutation. Without loss of generality, we assume that the true loading matrix  $\Lambda$  also takes the form of an unordered GLT structure.

First, we show that under an unordered GLT condition  $\beta_k$  takes a similar form as  $\beta_k \mathbf{T}_k$  does in (A4) and (A5), up to a signed permutation.  $\beta_k$  has  $\text{rk}(\beta_k) = r + s$  distinct pivot rows. Let  $n_1, \dots, n_s$  be the non-zero rows in the spurious loading matrix  $\mathbf{M}_s$ . When these rows are deleted from  $\beta_k$ , the resulting matrix  $\beta_k^*$  has reduced rank  $r$  which implies that  $s$  among the  $r + s$  pivot rows of  $\beta_k$  are identical to  $n_1, \dots, n_s$ . Furthermore,  $\beta_k^*$  contains a  $m \times r$  submatrix  $\beta_r^*$  that obeys an unordered GLT condition with the remaining  $r$  rows serving as pivots. Since both  $\beta_k^*$  and  $\beta_r^*$  have rank  $r$ , the remaining  $m \times s$  submatrix of  $\beta_k^*$

has rank zero, and is equal to a nullmatrix except for a set of measure zero. Hence, a signed permutation matrix  $\mathbf{P}$  of size  $r + s$  can be used to reorder the columns of  $\beta_k^*$ :

$$\beta_k^* \mathbf{P} = \begin{pmatrix} \beta_r^* & \mathbf{O} \end{pmatrix}. \quad (\text{A6})$$

If the same signed permutation is applied to  $\beta_k^d$ , then we obtain:

$$\beta_k^d \mathbf{P} = \begin{pmatrix} \beta_r^d & \beta_s \end{pmatrix}, \quad (\text{A7})$$

where  $\beta_r^d$  is an  $s \times r$  matrix. The  $s \times s$  submatrix  $\beta_s$  has an unordered GLT structure with pivot rows  $n_1, \dots, n_s$  and is equal to a lower triangular matrix  $\tilde{\mathbf{L}}$  of rank  $s$  up to column switching, i.e.,  $\beta_s = \tilde{\mathbf{L}}\tilde{\mathbf{P}}_\rho$ .

Next, let us investigate the spurious submatrix  $\mathbf{M}_0$  in (A5). Since under an unordered GLT condition, any rotation  $\beta_k \mathbf{T}_k$  generated from  $\beta_k$  through (26) also exhibits an unordered GLT structure, it follows from (A4) and (A5) that  $\mathbf{M}_0$  is an unordered GLT matrix with pivot rows  $n_1, \dots, n_s$ . Therefore,  $\mathbf{M}_0 = \mathbf{L}\mathbf{P}_\rho$  is equal to a lower triangular matrix  $\mathbf{L}$  up to column switching. From  $\mathbf{L}\mathbf{L}^\top = \mathbf{M}_0\mathbf{M}_0^\top = \mathbf{D}_0^2$ , it follows that  $\mathbf{L} = \mathbf{D}_0\mathbf{P}_\pm$  is equal to the diagonal matrix  $\mathbf{D}_0$  up to sign switching and, therefore:

$$\mathbf{M}_0 = \mathbf{D}_0\mathbf{P}_M, \quad (\text{A8})$$

where  $\mathbf{P}_M$  is a signed permutation of size  $s$ . This proves the first claim that under the unordered GLT framework, that the spurious factor matrix of any rotation  $\beta_k \mathbf{T}_k$  generated through (26) can be represented as a spurious unordered GLT matrix  $\mathbf{M}_s$ .

To complete the proof, we show that the rotation matrix  $\mathbf{T}^*$  defined by  $\mathbf{T}^* := \mathbf{P}^\top \mathbf{T}_k$  is a signed permutation.  $\mathbf{T}^*$  is split in the following way:

$$\mathbf{T}^* = \begin{pmatrix} \mathbf{T}_1^* & \mathbf{T}_3^* \\ (\mathbf{T}_3^*)^\top & \mathbf{T}_2^* \end{pmatrix}, \quad (\text{A9})$$

with square matrices  $\mathbf{T}_1^*$  and  $\mathbf{T}_2^*$  of size  $r$  and  $s$ . We obtain from (A4) and (A6):

$$\beta_k^* \mathbf{T}_k = \beta_k^* \mathbf{P} \mathbf{P}^\top \mathbf{T}_k = \begin{pmatrix} \beta_r^* \mathbf{T}_1^* & \beta_r^* \mathbf{T}_3^* \end{pmatrix} = \begin{pmatrix} \Lambda^* & \mathbf{O} \end{pmatrix}.$$

Since  $\beta_r^*$  has full column rank, we obtain from  $\beta_r^* \mathbf{T}_3^* = \mathbf{O}$  by left multiplication with  $((\beta_r^*)^\top \beta_r^*)^{-1}(\beta_r^*)^\top$  that  $\mathbf{T}_3^* = \mathbf{O}$ . Furthermore,  $\beta_r^* \mathbf{T}_1^* = \Lambda^*$ . Application of Corollary 1 to the unordered GLT matrix  $\beta_r^*$ , which satisfies  $\beta_r^* (\beta_r^*)^\top = \Lambda^* (\Lambda^*)^\top$ , yields  $\beta_r^* = \Lambda^* \mathbf{P}_r$ , where  $\mathbf{P}_r$  is a signed permutation of size  $r$ . From  $\beta_r^* \mathbf{T}_1^* = \Lambda^* = \beta_r^* \mathbf{P}_r^\top$ , we obtain that  $\mathbf{T}_1^* = \mathbf{P}_r^\top$  is a signed permutation. Finally, we obtain from (A5), (A7), and (A8):

$$\beta_k^d \mathbf{T}_k = \beta_k^d \mathbf{P} \mathbf{T}^* = \begin{pmatrix} \beta_r^d & \beta_s \end{pmatrix} \mathbf{T}^* = \begin{pmatrix} \beta_r^d \mathbf{P}_r^\top & \tilde{\mathbf{L}}\tilde{\mathbf{P}}_\rho \mathbf{T}_2^* \end{pmatrix} = \begin{pmatrix} \Lambda^d & \mathbf{D}_0 \mathbf{P}_M \end{pmatrix}.$$

It follows that  $\tilde{\mathbf{L}}\tilde{\mathbf{L}}^\top = \mathbf{D}_0^2$ , and therefore the lower triangular matrix  $\tilde{\mathbf{L}}$  is equal to  $\mathbf{D}_0$  up to sign switching. Consequently,  $\tilde{\mathbf{L}}\tilde{\mathbf{P}}_\rho \mathbf{T}_2^* = \mathbf{D}_0 \mathbf{P}_s \mathbf{T}_2^* = \mathbf{D}_0 \mathbf{P}_M$ , where  $\mathbf{P}_s$  is a signed permutation of size  $s$ . It follows immediately that  $\mathbf{T}_2^* = \mathbf{P}_s^\top \mathbf{P}_M$  is also a signed permutation matrix of size  $s$ . Therefore, both the rotation matrix  $\mathbf{T}^*$  defined in (A9), as well as  $\mathbf{T}_k = \mathbf{T}^* \mathbf{P}$ , are signed permutations of size  $r + s$ .  $\square$

## Notes

- <sup>1</sup> The sign condition on  $\Lambda_{ij}$  is needed to avoid sign switching, since  $2^r - 1$  loading matrices  $\beta$  can be constructed, which differ from  $\Lambda$  by a sign switch in a subset of columns, but yield the same cross-covariance matrix  $\beta\beta^\top = \Lambda\Lambda^\top$  (see also Section 3). Without sign conditions,  $\Lambda$  would not be identified. However, any other sign condition, such as  $\Lambda_{ii} < 0, i = 1, \dots, r$ , could be applied.
- <sup>2</sup> Our use of the term “pivot” is inspired by the concept of pivot columns in a row reduced echelon form (RREF), which is the result of Gauss–Jordan elimination. In particular, if  $\Lambda^\top$  is in RREF, then pivot rows of  $\Lambda$  are the pivot columns of  $\Lambda^\top$ . For more details, see, e.g., Anton and Rorres (2013).



- 3 The pivot rows  $l_1, \dots, l_r$  thus coincide with the pivot columns of the row reduced echelon form (RREF) of  $\beta^\top$ .
- 4 Their algorithm for the efficient verification of  $\text{CR}(r, 1)$  is implemented in R and MATLAB. The computer code is publicly available at <https://github.com/hdarjus/sparvaride> (accessed on 31 October 2023) and, respectively, <https://github.com/hdarjus/sparvaride-matlab> (accessed on 31 October 2023).
- 5 Note that the non-spurious columns of  $\tilde{\beta}_3$  form an unordered GLT structure, while  $\Lambda$  is GLT.
- 6 Evidently, zero columns (if any) in a posterior draw of  $\beta_H$  can be ignored, since  $\beta_H \beta_H^\top = \beta_k \beta_k^\top$ .
- 7 It should be noted that Dirac-spike-and-slab priors such as (29) are useful in this regard, since they are able to identify the  $(m - 1)s$  exact zeros in the columns corresponding to spurious factors. Under continuous shrinkage priors \*(see, e.g., Bhattacharya and Dunson (2011); Ročková and George (2017)), how to identify and remove spurious factors is not straightforward.
- 8 This algorithm is designed for inference in EFA models under the GLT condition, but can be easily extended to models with unconstrained loading matrices  $\beta_H$ .
- 9 See Frühwirth-Schnatter et al. (2023) for a case study involving additional industry sectors.
- 10 These computation were carried out using the function `factoran` in MATLAB.

## References

- Anderson, Brian David Outram, Manfred Deistler, Elisabeth Felsenstein, Bernd Funovits, Lukas Koelbl, and Mohsen Zamani. 2016. Multivariate AR systems and mixed frequency data: G-identifiability and estimation. *Econometric Theory* 32: 793–826. [CrossRef]
- Anderson, Theodore Wilbur. 2003. *An Introduction to Multivariate Statistical Analysis*, 3rd ed. Chichester: Wiley.
- Anderson, Theodore Wilbur, and Herman Rubin. 1956. Statistical inference in factor analysis. Paper presented at Third Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, USA, December 26–31, Volume V, pp. 111–50.
- Anton, Howard, and Chris Rorres. 2013. *Elementary Linear Algebra*, 11th ed. Hoboken: Wiley Global Education.
- Aßmann, Christian, Jens Boysen-Hogrefe, and Markus Pape. 2016. Bayesian analysis of static and dynamic factor models: An ex-post approach toward the rotation problem. *Journal of Econometrics* 192: 190–206. [CrossRef]
- Bai, Jushan, and Serena Ng. 2002. Determining the number of factors in approximate factor models. *Econometrica* 70: 191–221. [CrossRef]
- Bai, Jushan, and Serena Ng. 2013. Principal components estimation and identification of static factors. *Journal of Econometrics* 176: 18–29. [CrossRef]
- Bartholomew, David John. 1987. *Latent Variable Models and Factor Analysis*. London: Charles Griffin.
- Bekker, Paul A. 1989. Identification in restricted factor models and the evaluation of rank conditions. *Journal of Econometrics* 41: 5–16. [CrossRef]
- Bhattacharya, Anirban, and David Brian Dunson. 2011. Sparse Bayesian infinite factor models. *Biometrika* 98: 291–306. [CrossRef]
- Boivin, Jean, and Serena Ng. 2006. Are more data always better for factor analysis? *Journal of Econometrics* 132: 169–94. [CrossRef]
- Carvalho, Carlos M., Jeffrey Chang, Joseph E. Lucas, Joseph R. Nevins, Quanli Wang, and Mike West. 2008. High-dimensional sparse factor modeling: Applications in gene expression genomics. *Journal of the American Statistical Association* 103: 1438–56. [CrossRef]
- Chan, Joshua, Roberto Leon-Gonzalez, and Rodney W. Strachan. 2018. Invariant inference and efficient computation in the static factor model. *Journal of the American Statistical Association* 113: 819–28. [CrossRef]
- Conti, Gabriella, Sylvia Frühwirth-Schnatter, James Joseph Heckman, and Rémi Piatek. 2014. Bayesian exploratory factor analysis. *Journal of Econometrics* 183: 31–57. [CrossRef]
- Fan, Jianqing, Yingying Fan, and Jinchi Lv. 2008. High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics* 147: 186–97. [CrossRef]
- Forni, Mario, Domenico Giannone, Marco Lippi, and Lucrezia Reichlin. 2009. Opening the black box: Structural factor models with large cross sections. *Econometric Theory* 25: 1319–47. [CrossRef]
- Frühwirth-Schnatter, Sylvia. 2023. Generalized cumulative shrinkage process priors with applications to sparse Bayesian factor analysis. *Philosophical Transactions of the Royal Society A*, forthcoming. [CrossRef] [PubMed]
- Frühwirth-Schnatter, Sylvia, and Hedibert F. Lopes. 2018. Sparse Bayesian factor analysis when the number of factors is unknown. *arXiv* arXiv:1804.04231.
- Frühwirth-Schnatter, Sylvia, Darjus Hosszejni, and Hedibert Freitas Lopes. 2023. Sparse Bayesian factor analysis when the number of factors is unknown. *arXiv* arXiv:2301.06459.
- Geweke, John Frederick, and Guofu Zhou. 1996. Measuring the pricing error of the arbitrage pricing theory. *Review of Financial Studies* 9: 557–87. [CrossRef]
- Geweke, John Frederick, and Kenneth James Singleton. 1980. Interpreting the likelihood ratio statistic in factor models when sample size is small. *Journal of the American Statistical Association* 75: 133–37. [CrossRef]
- Golub, Gene H., and Charles F. Van Loan. 2013. *Matrix Computations*, 4th ed. Baltimore: Johns Hopkins University Press.
- Hayashi, Kentaro, and George A. Marcoulides. 2006. Examining identification issues in factor analysis. *Structural Equation Modeling* 13: 631–45. [CrossRef]
- Hosszejni, Darjus, and Sylvia Frühwirth-Schnatter. 2022. Cover it up! Bipartite graphs uncover identifiability in sparse factor analysis. *arXiv* arXiv:2211.00671.
- Jöreskog, Karl Gustav. 1969. A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika* 34: 183–202.

- Kastner, Gregor. 2019. Sparse Bayesian time-varying covariance estimation in many dimensions. *Journal of Econometrics* 210: 98–115. [\[CrossRef\]](#)
- Kaufmann, Sylvia, and Christian Schumacher. 2017. Identifying relevant and irrelevant variables in sparse factor models. *Journal of Applied Econometrics* 32: 1123–44. [\[CrossRef\]](#)
- Kaufmann, Sylvia, and Christian Schumacher. 2019. Bayesian estimation of sparse dynamic factor models with order-independent and ex-post identification. *Journal of Econometrics* 210: 116–34. [\[CrossRef\]](#)
- Koopmans, Tjalling Charles, and Olav Reiersøl. 1950. The identification of structural characteristics. *The Annals of Mathematical Statistics* 21: 165–81. [\[CrossRef\]](#)
- Ledoit, Olivier, and Michael Wolf. 2020. The power of (non-)linear shrinking: A review and guide to covariance matrix estimation. *Journal of Financial Econometrics* 20: 187–218. [\[CrossRef\]](#)
- Lee, Sik-Yum, and Xin-Yuan Song. 2002. Bayesian selection on the number of factors in a factor analysis model. *Behaviormetrika* 29: 23–39. [\[CrossRef\]](#)
- Legramanti, Sirio, Daniele Durante, and David B. Dunson. 2020. Bayesian cumulative shrinkage for infinite factorizations. *Biometrika* 107: 745–52. [\[CrossRef\]](#)
- Lopes, Hedibert Freitas, and Mike West. 2004. Bayesian model assessment in factor analysis. *Statistica Sinica* 14: 41–67.
- Magnus, Jan R., and Heinz Neudecker. 2019. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Hoboken: John Wiley & Sons.
- Neudecker, Heinz. 1990. On the identification of restricted factor loading matrices: An alternative condition. *Journal of Mathematical Psychology* 34: 237–41. [\[CrossRef\]](#)
- Ohn, Ilsang, and Yongdai Kim. 2022. Posterior consistency of factor dimensionality in high-dimensional sparse factor models. *Bayesian Analysis* 17: 491–514. [\[CrossRef\]](#)
- Owen, Art B., and Jingshu Wang. 2016. Bi-cross-validation for factor analysis. *Statistical Science* 31: 119–39. [\[CrossRef\]](#)
- Reiersøl, Olav. 1950. On the identifiability of parameters in Thurstone's multiple factor analysis. *Psychometrika* 15: 121–49. [\[CrossRef\]](#) [\[PubMed\]](#)
- Ročková, Veronika, and Edward I. George. 2017. Fast Bayesian factor analysis via automatic rotation to sparsity. *Journal of the American Statistical Association* 111: 1608–22. [\[CrossRef\]](#)
- Sato, Manabu. 1992. A study of an identification problem and substitute use of principal component analysis in factor analysis. *Hiroshima Mathematical Journal* 22: 479–524. [\[CrossRef\]](#)
- Teh, Yee Whye, Dilan Görür, and Zoubin Ghahramani. 2007. Stick-breaking construction for the Indian buffet process. Paper presented at Eleventh International Conference on Artificial Intelligence and Statistics, Proceedings of Machine Learning Research (PMLR), San Juan, Puerto Rico, March 21–24. Edited by Marina Meila and Xiaotong Shen, vol. 2, pp. 556–63.
- Thurstone, Louis Leon. 1935. *The Vectors of Mind*. Chicago: University of Chicago.
- Thurstone, Louis Leon. 1947. *Multiple Factor Analysis*. Chicago: University of Chicago.
- Tumura, Yosiro, and Manabu Sato. 1980. On the identification in factor analysis. *TRU Mathematics* 16: 121–31.
- West, Mike. 2003. Bayesian factor regression models in the “large p, small n” paradigm. In *Bayesian Statistics 7*. Edited by José Miguel Bernardo, María Jesús Bayarri, James Orvis Berger, Alexander Philip Dawid, David Heckerman, Adrian Frederick Melhuish Smith and Mike West. Oxford: Oxford University Press, pp. 733–42.
- Williams, Benjamin. 2020. Identification of the linear factor model. *Econometric Reviews* 39: 92–109. [\[CrossRef\]](#)
- Zhao, Shiwen, Chuan Gao, Sayan Mukherjee, and Barbara E. Engelhardt. 2016. Bayesian group factor analysis with structured sparsity. *Journal of Machine Learning Research* 17: 1–47.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.