Identification Through Sparsity in Factor Models: the ℓ_1-rotation criterion

Simon Freyaldenhoven

Federal Reserve Bank of Philadelphia*

November 21, 2022

Abstract

Linear factor models are generally not identified. We provide sufficient conditions for identification: under a sparsity assumption, we can estimate the individual loading vectors using a novel rotation criterion that minimizes the ℓ_1-norm of the loading matrix. This enables economic interpretation of the factors. The assumption of sparsity in the loading matrix is testable and we propose such a test. Existing rotation criteria are theoretically unjustified and perform worse in our simulations. We illustrate our method in two economic applications.

JEL codes: C38, C51, C55

KEYWORDS: identification, factor models, sparsity, local factors

*Email: simon.freyaldenhoven@phil.frb.org. I thank Jushan Bai, Richard Crump, Chris Hansen, Adam McCloskey, Jesse Shapiro, and seminar audiences at Brown University, the Bundesbank, the Federal Reserve Bank of New York, the Federal Reserve Bank of Philadelphia, the University of Connecticut, the University of Geneva, as well as participants of the 2019 Greater New York Area Econometrics Colloquium, the 2019 European Conference of the Econometrics Community, the 2020 World Congress of the Econometric Society, the 2022 RCEA Conference on Recent Developments in Economics, Econometrics and Finance and the 2022 Barcelona Summer Forum for helpful comments and suggestions. Joseph Huang, Nathan Schor, and Le Xu provided excellent research assistance.

1 Introduction

Factor models are subject to a rotational indeterminacy, meaning that the individual factors and loading vectors are only identified up to a rotation. Although this rotational indeterminacy prohibits any economic interpretation of the estimated factors, even seminal papers in economics (e.g., Stock and Watson 2002, Ludvigson and Ng 2009[1]) often include a discussion on the economic interpretations of individual factors, usually preceded by the caveat that such an interpretation is theoretically unjustified. For example, Stock and Watson [2002] remark,

> "Because the factors are identified only up to a $k \times k$ matrix, detailed discussion of the individual factors is unwarranted. Nevertheless, [...] Figure 1 therefore displays the R^2 of the regression of the 215 individual time series against each of the six empirical factors [...] Broadly speaking, the first factor loads primarily on output and employment; the second on interest rate spreads, unemployment rates and capacity utilization rates [...]."

Similarly, Ludvigson and Ng [2009] state,

> "Because the factors are identifiable only up to an $r \times r$ matrix, a detailed interpretation of the individual factors would be inappropriate. Moreover, we caution that any labeling of the factors is imperfect, because each is influenced to some degree by all the variables in our large dataset and the orthogonalization means that no one of them will correspond exactly to a precise economic concept like output or unemployment, which are naturally correlated. Nonetheless, it is useful to show that the factors capture relevant macroeconomic information. We do so here by briefly characterizing the factors as they relate to the underlying variables in our panel dataset. Figure 1 shows that the first factor loads heavily on measures of employment and production [...]."

We show that the assumption of sparsity in the loading matrix can solve this indeterminacy, allowing a researcher to estimate how the individual factors affect the observed variables. Sparsity in the loading matrix is natural in many economic applications. It is implied by the presence of local factors – factors that affect only a subset of the observables. Economic examples include industry-specific shocks in a firm-level dataset.

Formally, our first result is that the true loading matrix Λ^* achieves the minimum of the ℓ_0-norm across rotations of the loading matrix under a sparsity assumption. Intuitively this states that any rotation of a sparse loading vector will be less sparse. However, a rotation

[1]The two papers have a combined citation count of more than 4500 as of October 2022.

criterion based directly on the sparsity pattern (ℓ_0-norm) of the loading matrix will generally be infeasible. Our main result then establishes that the true loading matrix Λ^* also achieves a minimum of the ℓ_1-norm across rotations. Our proposed ℓ_1-rotation criterion thus enables a researcher to consistently estimate the individual loading vectors of any local factors. Our rotation criterion is easy to implement in practice, and simply requires a $\sqrt{n}$-consistent estimate of the loading space as a starting point. Despite the resemblance to regularized estimation methods with an ℓ_1-penalty, such as Sparse Principal Component Analysis, we emphasize that there is no "shrinkage" involved in our estimator. Instead, we use the ℓ_1-norm as a criterion to select the most sparse loading matrix Λ from among a set of rotations. Applying our criterion to both an international panel of daily stock returns and a panel of US macroeconomic indicators enables us to identify individual loading vectors in both cases and to better understand the economic structure of the data.

As a second contribution we introduce of a criterion that can be used to determine whether local factors are present in a given dataset. Our test effectively consists of counting the number of "small" loadings in the most sparse rotation of the loading matrix, and comparing it to the number of small loadings that could be expected if the true loading vector was nonzero everywhere. Using our testing criterion, we find strong evidence for the existence of local factors in both applications.

Despite the large literature on both factor models and sparsity, little work has been done on the intersection of the two. Arguably one reason is that the sparsity pattern in the loading matrix is generally not invariant to rotations of the loading vectors. Kristensen [2017] considers adding an ℓ_1-penalty to the estimation of the loading matrix via principal components to induce sparsity in the loading matrix. However, even if the true loadings are sparse, the principal components estimate an arbitrary rotation, which will generally not be sparse, even in population. Bai and Ng [2013] provide an overview of the different normalizations commonly used in economics. A choice of normalization is equivalent to a choice of rotation for the estimated factors. In practice, this choice often appears to be based on statistical convenience rather than economic arguments. We argue that sparsity in the loading matrix, which is both economically appealing and statistically testable, provides a more natural normalization in many settings.

A related literature considers hierarchical factor models with a known group structure (e.g., Boivin and Ng 2006, Moench et al. 2013, Choi et al. 2018). Unlike those papers, we neither require the group structure to be known a priori, nor require a hierarchical model in which each outcome belongs to only one group. Ando and Bai [2017], Uematsu and Yamagata [2022] and Freyaldenhoven [2022] also do not require knowledge of the group structure a priori, but the focus of the first two papers is on estimation of the factor space, and the focus of the second paper is on estimating the number of factors. Neither addresses identification of individual factors.

While work on the estimation of sparse principal components in the statistics literature (e.g., Jolliffe et al. [2003], Zou et al. [2006]) is naturally related to the estimation of sparse factors, principal components by definition do not suffer from the same rotational indeterminacy. For a Bayesian perspective, see Ročková and George [2016] and Kaufmann and Schumacher [2019], who use sparse priors to encourage sparsity in the loading matrix.

A large and popular literature already exists that considers rotation criteria aimed to simplify the loading matrix in factor models, going back to at least Carroll [1953] and Kaiser [1958] (also see Katz and Rohlf 1974, Rozeboom 1991, Jennrich 2006).[2] However, existing rotation criteria are generally missing formal consistency results. To the best of our knowledge, our ℓ_1-rotation is the first rotation criterion that comes with theoretical guarantees to recover the true loadings vectors under a sparsity assumption. We also find that our criterion performs better than existing heuristics across our simulations.

The paper proceeds as follows. After setting up our model and fixing notation in Section 2, we discuss a simple example and give an intuitive discussion of our results in Section 3. Section 4 includes our formal identification results. In Section 4.1, we show that the true loading matrix Λ^* is the unique minimum of the ℓ_0-norm across rotations under exact sparsity, and introduce a testing criterion to determine whether local factors are present in a given dataset. We further establish that Λ^* is also a minimum of the ℓ_1-norm across rotations. Section 4.2 extends our results to allow for $\sqrt{n}$-consistent initial estimates of the loading space and approximate sparsity in the true loading vectors. Section 5 provides Monte Carlo evidence that supports our asymptotic results in finite sample. In Section 6, we apply our results to a panel of individual stock returns as well as a panel of US macroeconomic indicators. In the latter application we also illustrate how local factors can create the "Illusion of sparsity" (cf. Giannone et al. [2021]) in forecast models.

[2] For example, the Varimax criterion (Kaiser 1958) is widely used across fields with more than 9000 citations as of August 2021 and is included in many major statistical software applications (e.g. R, Matlab and SAS).

2 Preliminaries

We use standard notation in the literature on factor models and assume X follows a factor structure:

$$\underset{(n\times 1)}{X_t} = \underset{(n\times r)}{\Lambda^*}\underset{(r\times 1)}{F_t} + \underset{(n\times 1)}{e_t} \ \forall t, \qquad \text{or more compactly,} \qquad \underset{(T\times n)}{X} = \underset{(T\times r)}{F}\underset{(r\times n)}{\Lambda^{*'}} + \underset{(T\times n)}{e}, \tag{1}$$

where $\Lambda^* = [\lambda^*_{1\bullet}; \lambda^*_{2\bullet}; ...; \lambda^*_{n\bullet}]' = [\lambda^*_{\bullet 1}\lambda^*_{\bullet 2}...\lambda^*_{\bullet r}]$ denotes the matrix of true factor loadings, and F denotes the unobserved factors. We use the running indices s, t for the T observations, i, j for the n variables, and k, l for the r factors throughout. To rule out pathological cases, we will assume throughout that $rank(\Lambda^*) = r$.

Let $tr(A)$ denote the trace of a matrix A. We use the Frobenius norm for matrices, such that $\|A\|^2 = tr(A'A) = \sum_{i,j} a_{ij}^2$. Similarly, unless otherwise noted, $\|A\|_1$ and $\|A\|_0$ will be entrywise (pseudo-)norms, such that, for instance, $\|A\|_0$ will count the non-zero entries of a matrix A. We use the term generalized permutation matrix for a matrix P^* that can be expressed as the product of an invertible diagonal matrix D and a permutation matrix P, with its dimension usually obvious from context. A set in a superscript of a vector x, always denoted by a script letter (e.g., $\mathcal{G}$), defines a vector $x^{\mathcal{G}}$ such that $x_i^{\mathcal{G}} = x_i$ whenever $i \in \mathcal{G}$ and $x_i^{\mathcal{G}} = 0$ otherwise. We write $a_n \asymp b_n$ for two sequences a_n, b_n if $a_n = O(b_n)$ and $b_n = O(a_n)$. We normalize the length of the true loading vectors throughout, and impose that $\sum_{i=1}^n \lambda_{il}^{*2} = n$ for $l = 1, \ldots, r$. Clearly, such a normalization of a loading vector $\lambda^*_{\bullet k}$ and its corresponding factor F_k is immaterial.

Equation (1) is observationally equivalent for different rotations of the loadings and factors. To see this, let H denote an arbitrary nonsingular matrix. We can redefine $\Lambda^0 = \Lambda^*(H')^{-1}$ and $F^0 = FH$. This rotation may well be oblique since H does not need to be unitary, and we make no assumption that either the factors or the loading vectors are orthogonal. In our view, there is no reason a priori to believe that the underlying factors, and in particular the loading vectors, are necessarily orthogonal.

Among others, Bai and Ng [2002] showed in their seminal paper that in factor models of large dimensions, we can consistently estimate the number of factors under some regularity conditions. We will therefore assume the true number of factors r to be known in the remainder of this paper.[3] Throughout the paper, we assume the data has been centered, such that

[3] See also Ahn and Horenstein [2013] and Onatski [2010] for alternative ways to determine the number of factors. Freyaldenhoven [2022] addresses the issue of estimating the number of factors under the presence of local factors, affecting only a subset of the observables.

$\mathbb{E}(X_i) = 0$. All proofs and auxiliary lemmata are relegated to the Online Appendix.

3 Intuition

We start with a stylized example and an intuitive discussion of our proposed criterion.

3.1 A Stylized Example

To fix ideas, consider the following simple factor model with two factors for a vector x_t of dimension $n = 207$:

$$x_t = \lambda^*_{\bullet 1} F_{1t} + \lambda^*_{\bullet 2} F_{2t} + e_t, \quad t = 1, \dots, T, \tag{2}$$

where $\lambda^*_{\bullet k}$ denotes the vector of loadings for factor k (denoted by F_{kt}), and e_t an idiosyncratic noise component. We discuss the data-generating process in more detail in Section 5. Suppose both factors are local with the structure of the loading matrix Λ^* given by

$$\Lambda^* = \begin{bmatrix} \lambda^*_{1:m_1,1} & 0 \\ 0 & \lambda^*_{(n+1)-m_2:n,2} \end{bmatrix}, \tag{3}$$

where $m_1 = m_2 = 120$. Thus, 120 outcomes are affected by the first factor, and 120 outcomes are affected by the second factor. Note that, with $n = 207$, some outcomes are affected by both factors. For the non-zero entries we set $\lambda^*_{ik} \overset{i.i.d.}{\sim} U(0.1, 2.9)$. Figure 1 visualizes the resulting loading matrix Λ^*.

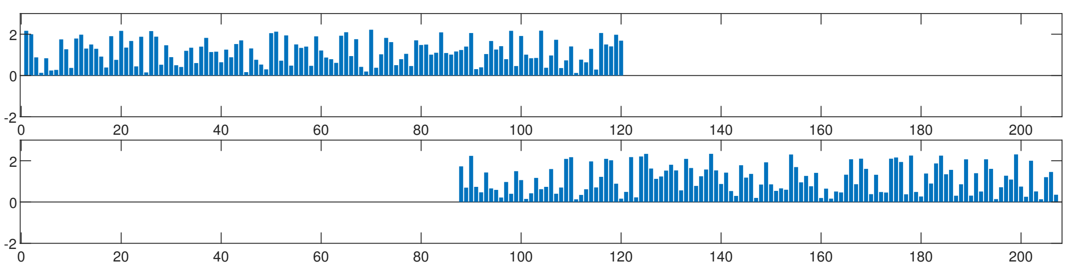


Figure 1: Illustration of true loading matrix Λ^* for stylized DGP. Top panel depicts $\lambda^*_{\bullet 1}$, bottom panel $\lambda^*_{\bullet 2}$. For each factor, the loadings associated with all 207 outcomes are depicted.

Under standard regularity conditions in the literature, it is well known that we can obtain estimates $\lambda^0_{\bullet 1}, \lambda^0_{\bullet 2}$, such that

$$\begin{aligned} \lambda^0_{i1} &= H_{11}\lambda^*_{i1} + H_{12}\lambda^*_{i2} + o_p(1) \\ \lambda^0_{i2} &= H_{21}\lambda^*_{i1} + H_{22}\lambda^*_{i2} + o_p(1), \end{aligned} \tag{4}$$

where H is an unknown non-singular rotation matrix (e.g. Bai 2003). Thus, the estimates $\lambda^0_{\bullet 1}$ and $\lambda^0_{\bullet 2}$ will in population be linear combinations of the true loading vectors $\lambda^*_{\bullet 1}$ and $\lambda^*_{\bullet 2}$. We make the following two observations (for now ignoring the $o_p(1)$ term in Equation (4)):

1. **Observation 1: Linear combinations of sparse loading vectors are generally dense**
   For an arbitrary linear combination of the true loading vectors $\lambda^0_{\bullet 1} = H_{11}\lambda^*_{\bullet 1} + H_{12}\lambda^*_{\bullet 2}$ with $H_{11}, H_{12} \neq 0$ we will generally have $\lambda^0_{i1} \neq 0$ for $i = 1, \ldots, n$. Thus, even though the true loading vector $\lambda^*_{\bullet 1}$ is sparse (cf. Figure 1), a generic estimate $\lambda^0_{\bullet 1}$ will generally have non-zero entries everywhere.

2. **Observation 2: There exists a linear combination of the estimated loading vectors that is sparse**
   Since $\lambda^0_{\bullet 1}$ and $\lambda^0_{\bullet 2}$ are linear combinations of $\lambda^*_{\bullet 1}$ and $\lambda^*_{\bullet 2}$, it follows that $\lambda^*_{\bullet 1}$ and $\lambda^*_{\bullet 2}$ are also linear combinations of $\lambda^0_{\bullet 1}$ and $\lambda^0_{\bullet 2}$. In other words, there must exist weights w_1 and w_2, such that $\lambda^*_{\bullet 1} = w_1\lambda^0_{\bullet 1} + w_2\lambda^0_{\bullet 2}$. It then also follows that, if $\lambda^*_{\bullet 1}$ is sparse, there must exist a linear combination of $\lambda^0_{\bullet 1}$ and $\lambda^0_{\bullet 2}$ that is sparse.

Together, these two observations form the key insight of the paper: The sparsity pattern in the loading matrix is not invariant to rotations and can be used to achieve identification. We next illustrate our approach to identification in this stylized DGP. By construction, the Principal Component estimator Λ^0 will estimate a rotation H of the true loadings and factors that satisfies $\lambda^{0\prime}_{\bullet 1}\lambda^0_{\bullet 2} = 0$ and $F^{0\prime}_{\bullet 1}F^0_{\bullet 2} = 0$.[4] Figure 2 depicts this estimate. In line with Observation 1, the rotation matrix H inherent to the Principal Component estimator results in an estimate of the loading matrix with no discernible sparsity pattern. Further, comparing Figures 1 and 2, we conclude that neither of the estimated loading vectors closely resembles $\lambda^*_{\bullet 1}$ or $\lambda^*_{\bullet 2}$.

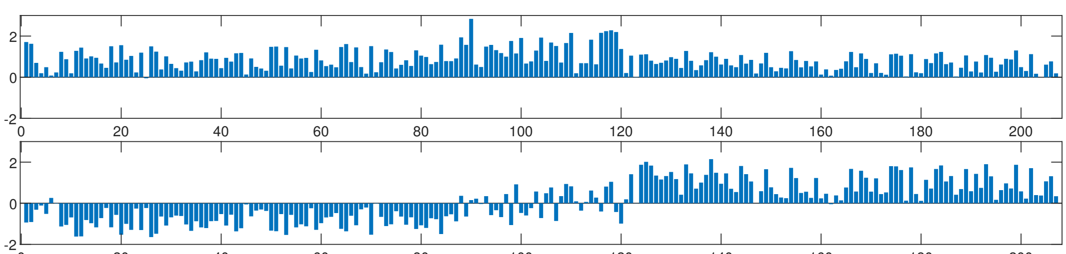


Figure 2: Illustration of Principal Component estimate Λ^0 for stylized DGP. Top panel depicts $\lambda^0_{\bullet 1}$, bottom panel $\lambda^0_{\bullet 2}$. For each factor, the loadings associated with all 207 outcomes are depicted.

[4] To compute the Principal Component estimator, we take the singular value decomposition $X = UDV'$. The leading r columns of V are used as $\lambda^0_{\bullet 1}, \ldots, \lambda^0_{\bullet r}$.

Following Observation 2, we are next interested in identifying a linear combination of $\lambda^0_{\bullet 1}$ and $\lambda^0_{\bullet 2}$ that is sparse. Because a rotation criterion that is directly based on the number of non-zero elements will generally be infeasible (we return to this later), our proposed estimator takes Λ^0 as a starting point and is equal to the rotation of Λ^0 that minimizes the ℓ_1-norm of the loading vectors. Figure 3 depicts the value of $\|\lambda_{\bullet k}\|_1$ across rotations in the space spanned by the Principal Component estimator Λ^0. Specifically, it depicts how $\|\lambda_{\bullet k}\|_1 = \|w_1\lambda^0_{\bullet l1} + w_2\lambda^0_{\bullet l2}\|_1$ changes as we vary the weights w_1, w_2, under the restriction that $w_1^2 + w_2^2 = 1$. A convenient way to enforce this restriction, and to depict the result graphically, is to let $[w_1, w_2] = [sin(\theta), cos(\theta)]$, and depict $\|\lambda_{\bullet k}\|_1$ as a function of the angle θ. This is depicted in Figure 3.

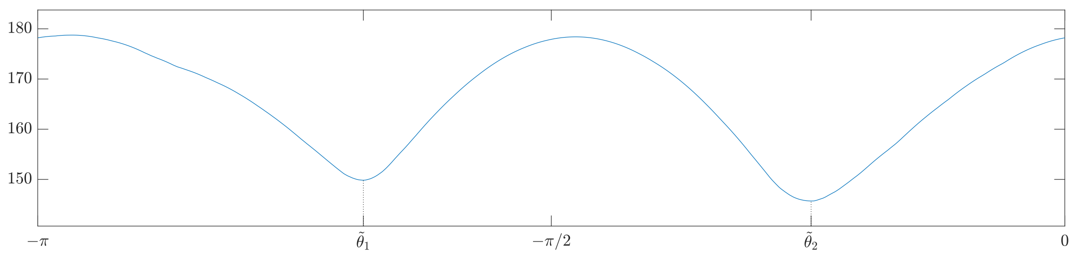


Figure 3: Objective function across rotations in the space spanned by the initial estimate Λ^0. Depicted is $\|\lambda_{\bullet k}\|_1 = \|sin(\theta)\lambda^0_{\bullet l1} + cos(\theta)\lambda^0_{\bullet l2}\|_1$ as a function of the angle θ.

We find two local minima at angles $\tilde{\theta}_1$ and $\tilde{\theta}_2$. The first minimum $\tilde{\theta}_1$ corresponds to weights of $[w_1\ w_2] = [-0.70\ 0.71]$, and consequently an estimated loading vector of $\tilde{\lambda}_{\bullet 1} = -0.70\lambda^0_{\bullet 1} + 0.71\lambda^0_{\bullet 2}$. The second minimum $\tilde{\theta}_2$ corresponds to weights of $[w_1, w_2] = [-0.84, -0.54]$, and consequently a second estimated loading vector of $\tilde{\lambda}_{\bullet 2} = -0.84\lambda^0_{\bullet 1} - 0.54\lambda^0_{\bullet 2}$. Combining $[\tilde{\lambda}_{\bullet 1}, \tilde{\lambda}_{\bullet 2}] = \tilde{\Lambda}$, we obtain our proposed estimator for Λ^*. $\tilde{\Lambda}$ is depicted in Figure 4. Comparing Figures 1 and 4, we conclude that $\tilde{\Lambda}$ is close to Λ^*, and that we are able to identify the individual columns of Λ^* using our proposed criterion.

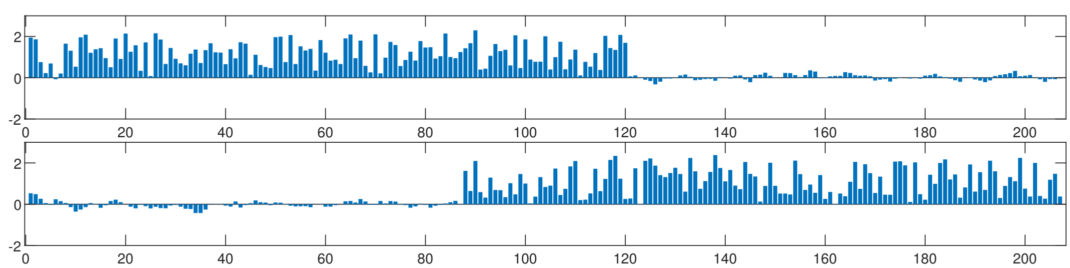


Figure 4: Illustration of $\tilde{\Lambda}$, the non-singular "rotated" matrix with the smallest ℓ_1-norm $\|\Lambda\|_1 = \sum_{i,k} |\lambda_{ik}|$ for stylized DGP. Top panel depicts $\tilde{\lambda}_{\bullet 1}$, bottom panel $\tilde{\lambda}_{\bullet 2}$. For each factor, the loadings associated with all 207 outcomes are depicted.

Remark 1. Even though $\tilde{\Lambda}$ is close to Λ^*, we note that $\tilde{\lambda}_{ik} \neq 0$ for all i, k. This is is expected because the preliminary estimate Λ^0 is subject to estimation error, and our method does not impose any regularization. Having identified the correct rotation of Λ^*, we conjecture that standard methods in regularized estimation, or even simple thresholding, can be used to further improve the estimate $\tilde{\Lambda}$ in practice. We leave this as an interesting avenue for future research.

Alternatively, we can approximate the ℓ_0-norm directly for each rotation by counting the number of "small" loadings. Figure 5 depicts the number of small loadings λ_{ik} across rotations in the space spanned by the initial estimate Λ^0, again as a function of the angle θ (we formally define "small" in Section 4.1.2). The angles $\tilde{\theta}_1$ and $\tilde{\theta}_2$ in Figure 5 are those found by minimizing the ℓ_1-norm of the loadings and are identical to the local minima depicted in Figure 3. While in this case, with just two factors, it is feasible to find a rotation that is close to Λ^* based on a visual inspection of the number of small loadings across rotations, the discontinuities and large number of local extrema of this function make this approach infeasible in higher dimensions (we expand on this in Online Appendix B).

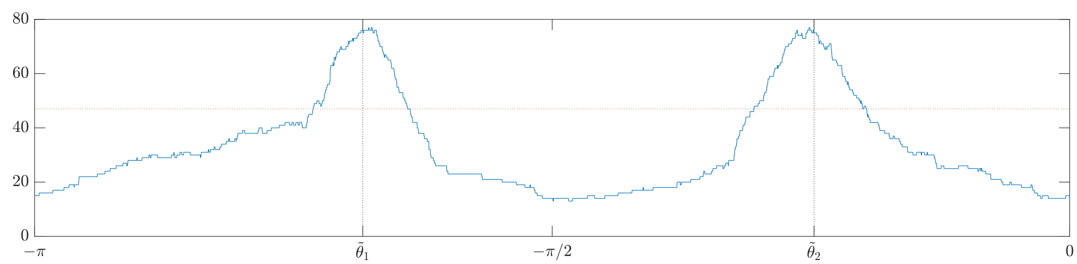


Figure 5: Depicted is an approximation of the ℓ_0-norm, $Q_0 = \sum_{i=1}^{n} \mathbf{1}_{|\lambda_{ik}|<1/log(n)}$, where $\lambda_{\bullet k} = sin(\theta)\lambda^0_{\bullet l1} + cos(\theta)\lambda^0_{\bullet l2}$, as a function of the angle θ. Horizontal dashed red line represents critical value for testing whether there are local factors in the data.

We also use Figure 5 to illustrate how one can use the estimate $\tilde{\Lambda}$ to test for the existence of local factors in a given dataset. In Section 4.1.2 we introduce a test that effectively consists of counting the number of small loadings in the most sparse estimated loading vector $\tilde{\lambda}_{\bullet k}$ and comparing it to the number of small loadings that would be expected if the true loading vector was non-zero everywhere with normally distributed loadings. In Figure 5 our test corresponds to checking whether, for either of the angles $\tilde{\theta}_1$ and $\tilde{\theta}_2$, the number of small coefficients, indicated by the blue line, is larger than a critical value, indicated by the horizontal red dashed line. Based on Figure 5, we conclude that we successfully detect the presence of local factors in this instance.[5]

[5]In Online Appendix A, we depict the equivalent of Figure 5 if both factors affect all outcomes. There, our

3.2 Connection to Existing Rotation Criteria

A number of widely used rotation criteria exist aimed at simplifying the loading matrix, going back to at least Carroll [1953] and Kaiser [1958]. These existing criteria usually use quartic functions of the loadings and maximize a variant of the following criterion function $Q(\cdot)$ over rotations of an initial estimate Λ^0:

$$Q(\Lambda^0 R) = Q(\Lambda) = \sum_{k=1}^{r}\sum_{l=1}^{k-1}(\sum_{i=1}^{n}\lambda_{ik}^2\lambda_{il}^2 - \frac{c}{n}\sum_{i=1}^{n}\lambda_{ik}^2\sum_{j=1}^{n}\lambda_{jl}^2). \tag{5}$$

If we consider only orthogonal rotations for now (which is equivalent to restricting R to be orthonormal), (5) simplifies to

$$Q(\Lambda^0 R) = Q(\Lambda) = \sum_{k=1}^{r}\left[\sum_{i=1}^{n}\lambda_{ik}^4 - \frac{c}{n}\left(\sum_{i=1}^{n}\lambda_{ik}^2\right)^2\right]. \tag{6}$$

For example, setting c to 0, 1, and $r/2$, respectively, results in the Quartimax (Carroll 1953), Varimax (Kaiser 1958), and Equamax (Saunders 1962) rotation criteria. Considering one loading vector at a time, it becomes clear that these are closely related to maximizing $\|\lambda_{\bullet k}\|_4^4 = \sum_{i=1}^n \lambda_{ik}^4$, subject to a constant ℓ_2-norm. In contrast, we propose to minimize $\|\lambda_{\bullet k}\|_1 = \sum_{i=1}^n |\lambda_{ik}|$, subject to a constant ℓ_2-norm.

To gain an intuition for the difference between the two approaches (maximizing ℓ_4, minimizing ℓ_1), it is instructive to first consider maximizing the ℓ_∞-norm and contrast this with minimizing the ℓ_0-norm across rotations. Intuitively, maximizing the ℓ_∞-norm identifies the rotation with the largest entry, while minimizing the ℓ_0-norm essentially identifies the rotation with the smallest entries. Minimizing the ℓ_1-norm is a relaxation of minimizing the ℓ_0-norm, while maximizing the ℓ_4-norm is a relaxation of the maximizing the ℓ_∞-norm. Our formal sparsity assumptions have direct implications for the behavior of the ℓ_0- and ℓ_1-norms, but not the ℓ_4- or ℓ_∞-norms. We conjecture this is the reason why, under sparsity assumptions, formal results have been difficult to achieve using existing rotation criteria that are quartic functions of the loadings.

We discuss the connection between our proposed method and a variety of quartic criteria, including criteria that result in oblique factor rotations (e.g., Hendrickson and White [1964]), further in Online Appendices B and D.

testing criterion correctly suggests that no local factors are present in the data. Intuitively, if both $\lambda_{\bullet 1}^*$ and $\lambda_{\bullet 2}^*$ are non-zero everywhere, no linear combination of the two exists with a significant sparsity pattern.

4 Identification

4.1 Under Exact Sparsity

We start by assuming an exact sparsity pattern in the loading matrix.

Assumption 1. *For each factor k, we can partition the set of indices $i = 1, 2, \ldots, n$ into a set of indices $\mathcal{A}_k$ with cardinality $|\mathcal{A}_k|$ and its complement $\mathcal{A}_k^c$, such that:*

(a) $\lambda^*_{ik} \neq 0$ *and* $\left|\lambda^*_{ik}\right| < C \; \forall i \in \mathcal{A}_k$ *and a constant* C.

(b) $\lambda^*_{ik} = 0 \; \forall i \notin \mathcal{A}_k$.

(c) $\exists c > 0$, *such that* $\left|\lambda^*_{ik}\right| > c \; \forall i \in \mathcal{A}_k$.

Parts (a)-(b) define $\mathcal{A}_k$ as the support of $\lambda^*_{\bullet k}$, and we may think of $\mathcal{A}_k$ as the "active set" for a given factor or loading vector: it collects the indices of all outcomes affected by that factor. Some results additionally require Assumption 1(c), which will be relaxed in Section 4.2, where we allow for approximate sparsity in the loading matrix.

Assumption 2. *Define $\Lambda^*_{\bullet,-m}$ as the n by $(r-1)$ submatrix of Λ^* obtained by deleting the mth column in Λ^*, and $\mathcal{A}_{z,-m}$ as the support of a linear combination $\Lambda^*_{\bullet,-m} z$ for a given $(r-1)$ vector of finite weights z and let $b_k(z) = \max|\mathcal{B}|$, such that*

$$\left[\Lambda^{*\mathcal{A}_k}_{\bullet,-k}\right]^{\mathcal{B}} z = \left[\lambda^{*\mathcal{A}_k}_{\bullet,k}\right]^{\mathcal{B}}. \tag{7}$$

Then, there exists a set of factors $\mathcal{F}^b$, such that $\forall F_k \in \mathcal{F}^b$, $\left|\mathcal{A}_k^c \cap \mathcal{A}_{z,-k}\right| - b_k(z) > 0$ for all $z \neq 0$.

Different versions of the set $\mathcal{F}^b$ will appear throughout the paper. $\mathcal{F}^b$ approximately defines a group of factors whose associated active sets are not supersets of another factor's active set (see the discussion below). We will generally be able to show identification for the loading vectors of factors in varying versions of $\mathcal{F}^b$. For instance, in a two-factor model, with a global factor affecting all outcomes and a local factor affecting only a subset of the outcomes, only the latter will be in $\mathcal{F}^b$. In such a model, the loading vector of the local factor is identified, whereas the loading vector of the global factor is not. In practice, these "local" factors, affecting only a subset of the outcomes, are often precisely those that are economically interesting because we may find them interpretable. To ease notation we note that Assumption 2 reduces to the following simple condition if $r = 2$ before further discussion.

Assumption 2'. *Let $b = \max |\mathcal{B}|$, such that $\mathcal{B} \subseteq (\mathcal{A}_1 \cap \mathcal{A}_2)$, and for all $i \in \mathcal{B}$*

$$c^* \lambda^*_{i,1} = \lambda^*_{i,2} \tag{8}$$

for some constant c^*.

Then, there exists a set of factors $\mathcal{F}^b$, such that $\forall F_k \in \mathcal{F}^b$, $\left|\mathcal{A}_k^c \cap \mathcal{A}_l\right| - b > 0$ for $l \neq k$.

Thus, b is defined as the size of the largest set of non-zero entries in the loading vectors such that the two loading vectors are perfectly collinear on that set. The inequality above then states that $F_1 \in \mathcal{F}^b$ if more than b outcomes are affected by F_2, but not F_1. This is slightly more restrictive than the condition that $\mathcal{A}_1$ may not be a superset of $\mathcal{A}_2$.

To gain further intuition, consider the following three specific examples (with $r = 2$).

1. Suppose $\mathcal{A}_1 \cap \mathcal{A}_2 = \varnothing$ (The two factors affect different, non-overlapping groups of outcomes).
   Then, $b = 0$, while $\left|\mathcal{A}_k^c \cap \mathcal{A}_l\right| = |\mathcal{A}_l| > 0$. Therefore, $\mathcal{A}_1 \cap \mathcal{A}_2 = \varnothing$ implies that $F_1, F_2 \in \mathcal{F}^b$.

2. Suppose $\mathcal{A}_2 \subseteq \mathcal{A}_1$ (The second factor F_2 affects a subset of the outcomes affected by F_1).
   Then, $\mathcal{A}_1^c \cap \mathcal{A}_2 = 0$, and it immediately follows that $F_1 \notin \mathcal{F}^b$. Thus, whenever $\mathcal{A}_k$ is a superset of another active set $\mathcal{A}_l$, F_k cannot be a member of the set $\mathcal{F}^b$.

3. Suppose $|\mathcal{A}_k| \asymp n$ for $k = 1, 2$ as $n \to \infty$, and $\lambda^*_{ik} \overset{i.i.d.}{\sim} N(0, \sigma)$ if $i \in \mathcal{A}_k$, and $\lambda^*_{ik} = 0$ otherwise.
   Then, with probability 1, $\frac{\lambda_{i1}}{\lambda_{i2}} \neq \frac{\lambda_{j1}}{\lambda_{j2}}$ for all $i, j \in \mathcal{A}_1 \cap \mathcal{A}_2$, and therefore $b = 1$. It follows that $F_1 \in \mathcal{F}^b$ if there are at least two outcomes affected by F_2, but not F_1.

More generally, a larger value of b means that the two loading vectors are more similar on the intersection of their supports. Since we generally treat the factor loadings as parameters rather than random variables, we rely on this high level assumption without specifying b further. More primitive conditions can be derived if we treat the loadings as random instead. For instance, under the third example above, we saw that $b = 1$ almost surely.

This intuition also holds for $r \geq 3$. Define $b_k = max_z b_k(z)$ as the size of the largest set such that we can represent the (non-zero part of the) loading vector $\lambda^*_{\bullet k}$ as an exact linear combination of the remaining loading vectors on this set. A small value for b_k (e.g., $b_k = r - 1$) means this set is small, and intuitively states that the different loading vectors are further from collinearity. The restriction $\left|\mathcal{A}_k^c \cap \mathcal{A}_{z,-k}\right| > b_k(z)$ is then again slightly more restrictive than the assumption that the active set for factor F_k is not a superset of any other factor, and becomes more restrictive for larger values of b_k.

Assumption 2 therefore implies a trade-off between the similarity in the supports of different loading vectors, and the similarity in the loadings on their joint support. The closer to collinearity the loading vectors are on their joint support (corresponding to a large value of b_k), the more distinct we require their active sets to be to achieve identification. Throughout, we refer to factors in the various versions of the set $\mathcal{F}^b$ as local factors.

Remark 2. It will generally also be possible to identify the subspace spanned by a given subset of loading vectors, even if their active sets are closely related, as long as they are sufficiently distinct from the active sets of all other factors. In order not to further complicate our notation, we will ignore this and simply consider such factors "unidentifiable" in the remainder of this paper.

4.1.1 Minimizing the ℓ_0-norm

Let R be a $r \times r$ matrix and consider the following minimization problem

$$\tilde{R} = \arg\min_{R} \|\Lambda^* R\|_0, \qquad \text{s.t. } R \text{ is nonsingular.} \tag{9}$$

Define $\tilde{\Lambda} = \Lambda^* \tilde{R}$ as the rotation of Λ^* corresponding to this mimimum. We start with the simplest case in which all factors are local:

Corollary 1. *Suppose Assumptions 1(a)-(b) and 2 hold, and $F_k \in \mathcal{F}^b$ for $k = 1, \dots, r$. Then $\tilde{\Lambda} = \Lambda^* P^*$ for some generalized permutation matrix P^*.*

Corollary 1 is a direct consequence of Theorem 1 below. It states that, in a model with only local factors, any rotation of the true loading matrix will be less sparse than the truth. Thus the rotation with the highest degree of sparsity identifies the individual loading vectors, up to an arbitrary relabeling of the factors and arbitrary scaling constants.

In most settings of economic interest, there will be at least some factor F_k, such that $F_k \notin \mathcal{F}^b$ (for instance, a global factor). In such a case, suppose $F_k \in \mathcal{F}^b$ for $k = 1, \dots, r^*$ and $F_k \notin \mathcal{F}^b$ for $k = r^* + 1, \dots, r$, and partition Λ^* accordingly: $\Lambda^* = [\Lambda^*_{\bullet,1:r^*} \Lambda^*_{\bullet,r^*+1:r}]$. The partitioning of the factors as above is without loss of generality, since it can always be achieved by a simple relabeling of the factors.

Theorem 1. *Suppose Assumptions 1(a)-(b) and 2 hold. Then for every $l = 1, \dots, r^*$, there exists an index k (which depends on l), such that $\tilde{R}_{l,k} \neq 0$ and $\tilde{R}_{l',k} = 0 \; \forall l' \neq l$.*

Theorem 1 establishes the following: If the true DGP includes local factors ($\mathcal{F}^b$ is non-empty), the loading vectors for such local factors can be identified by maximizing the degree of sparsity in the loading matrix across rotations. The intuition is that, of all possible rotations of local factors, none will be as sparse as the truth, Λ^*. Note that Theorem 1 does not say anything about factors that are not in $\mathcal{F}^b$. For instance, if there are global factors with a corresponding loading vector that has non-zero entries everywhere, identification of such factors based on a sparsity criterion will clearly be impossible.

Remark 3. Note that Theorem 1 still holds if we replace Λ^* in (9) with any rotation of Λ^*, $\Lambda^0 = \Lambda^* H$, where H is a nonsingular matrix.

4.1.2 Determining the Existence of Local Factors

We next ask whether the existence of local factors is testable. To this end, define

$$\mathcal{L}_0(\Lambda) = \max_k \Big(\sum_{i=1}^{n} \mathbf{1}\{|\lambda_{ik}| = 0\} \Big).$$

Thus, for a loading matrix Λ, $\mathcal{L}_0(\Lambda)$ is equal to the largest number of loadings equal to zero across loading vectors $\lambda_{\bullet k}$, $k = 1, \ldots, r$. We obtain the following result:

Proposition 1. *Suppose Assumptions 1 and 2 hold. Further suppose that* $b^* = \max_k \max_z b_k(z) = o(n)$*, where* $b_k(z)$ *is defined in Assumption 2. Let* $\tilde{\Lambda}$ *be a solution to* (9)*, and denote the column-normalized version of* $\tilde{\Lambda}$ *by* $\breve{\Lambda}$*:* $\breve{\lambda}_{\bullet k} = \frac{\tilde{\lambda}_{\bullet k}}{\|\tilde{\lambda}_{\bullet k}\|}$*. Then, as* $n \to \infty$*,*

a) $\mathcal{L}_0(\breve{\Lambda}) = o(n)$ *if* $|\mathcal{A}_k| = n$ *for* $k = 1, \ldots, r$.

b) For any $\gamma \in (0, 1)$*,* $\mathcal{L}_0(\breve{\Lambda}) \geq \gamma n$ *if there exist a factor* F_k *with* $|\mathcal{A}_k| < (1 - \gamma)n$.

Recall that $\tilde{\Lambda}$ represents the "most sparse rotation" in the space spanned by the true loading matrix. Intuitively, Proposition 1 states that, if all factors affect all outcomes ($|\mathcal{A}_k| = n$ for $k = 1, \ldots, r$), even the most sparse rotation in the space spanned by the true loading matrix will not have a significant sparsity pattern. On the other hand, if there exists a factor F_k for which least a constant fraction of its associated loadings are equal to zero, then clearly such a sparse rotation exists.

Thus, $\mathbf{1}\{\mathcal{L}_0(\breve{\Lambda}) \geq \gamma n\}$ can be used as a criterion to determine whether any factors are present that affect less than $(1 - \gamma)$ of the outcomes.

Remark 4. While Proposition 1 assumes no estimation error and requires exact sparsity, we show that it performs well in simulations in Section 5 and discuss its implementation in practice below. Size and power derivations of our testing criterion can presumably be done under additional assumptions on the distribution of the loadings (e.g. under normality, as in the third example of the previous section). We leave this as an interesting venue for future research.

In practice, since generally none of the estimated loadings will be exactly zero, we use

$$\mathbf{1}\{\hat{\mathcal{L}}_0(\breve{\Lambda})) \geq \gamma n\}, \quad \text{where } \hat{\mathcal{L}}_0(\breve{\Lambda}) = \max_k (\sum_{i=1}^{n} \mathbf{1}\{\left|\breve{\lambda}_{ik}\right| < h_n\}). \tag{10}$$

Thus, $\hat{\mathcal{L}}_0(\breve{\Lambda})$ is equal to the largest number of loadings smaller than the threshold h_n across the rotated loading vectors $\breve{\lambda}_{\bullet k}$, $k = 1, \ldots, r$.[6]

We propose the following tuning parameters h_n and γ.

- $h_n = \frac{1}{log(n)}$,
- $\gamma = \gamma_0 + \tilde{\gamma} = \gamma^0 + \left(p + c^{\alpha_\gamma}\sqrt{\frac{p(1-p)}{n}}\right)$, where $p = \phi(h_n) - \phi(-h_n)$,

where c^{α_γ} is a critical value for chosen significance level α_γ, and $\phi(\cdot)$ denotes the cdf of the standard normal distribution.

The intuition for this choice of parameters is the following. Asymptotically, $h_n = o_p(1)$ implies that $\tilde{\gamma} = 0$ and thus $\gamma = \gamma_0$, so that we are simply counting whether more than the fraction γ_0 of loadings is (close to) zero. For finite n, the additional term $\tilde{\gamma}$ reflects an upper bound on the proportion of loadings one would expect to fall inside the set $[-h_n, h_n]$ in finite sample if the loadings λ^*_{ik} are normally distributed. Our simulations suggest that setting $\gamma_0 = 0.03$ and $\alpha_\gamma = 0.05$ works well.

4.1.3 Minimizing the ℓ_1-norm

Minimizing the ℓ_0-norm directly is infeasible in practice for two reasons. First, minimizing the ℓ_0-norm directly will generally be computationally prohibitive. One can compare this to high-dimensional sparse linear regression models, where optimal subset selection is similarly infeasible. On the other hand, a vast body of literature exists documenting both the theoretical and practical appeal of using the ℓ_1-norm instead as a regularization in linear regression models (e.g., Bühlmann and Van De Geer 2011). We similarly propose to minimize the

[6] We reemphasize that counting the number of small loadings in an arbitrary rotation, e.g. using the principal component estimator Λ^0, would not work. In general, the number of small loadings will be small even under sparsity in Λ^* (cf. Figures 1 and 2). It is therefore crucial to first find the most sparse rotation $\tilde{\Lambda}$.

ℓ_1-norm of Λ to make our approach computationally feasible. Second, minimizing the ℓ_0-norm generally requires exact rather than approximate sparsity in the loadings. In fact, even under exact sparsity of the true loading matrix, any estimation error in the initial estimate of the loading space will generally mean that there are no exact zeros in any rotations of the estimated loading vectors. We therefore turn our attention to the ℓ_1-norm of the loading matrix next.

Assumption 3. *Let V_k denote the set of all linear combination $v_{\bullet k}$ of $\lambda^*_{\bullet l}$, $l = 1, \ldots, r$, such that $\|v_{\bullet k}\|_2^2 = n$ and $\lambda^*_{\bullet k} \perp v_{\bullet k}$ and define*

$$\beta^k(v_{\bullet k}) = \left| \sum_{i \in \mathcal{A}_k} |v_{ik}| \, \mathbf{1}\{\lambda^*_{ik} v_{ik} \geq 0\} - \sum_{i \in \mathcal{A}_k} |v_{ik}| \, \mathbf{1}\{\lambda^*_{ik} v_{ik} < 0\} \right|. \tag{11}$$

Then, there exists a set of factors $\mathcal{F}^{exact}$, such that, $\forall F_k \in \mathcal{F}^{exact}$,

$$\|v_{\bullet k}^{\mathcal{A}_k^c}\|_1 > \beta^k(v_{\bullet k}) \qquad \forall v_{\bullet k} \in V_k. \tag{12}$$

Assumption 3 is similar to Assumption 2, with $\mathcal{F}^{exact}$ approximately defining a group of factors whose associated active sets are not supersets of another factor's active set. To again ease notation we note that Assumption 3 can be simplified to the following if $r = 2$ before further discussion. With two factors, $v_{\bullet 1} = q_1 \lambda^*_{\bullet 1} + q_2 \lambda^*_{\bullet 2}$ for some constants q_1, q_2. Further, by definition of $\mathcal{A}_1$, $\|v_{\bullet 1}^{\mathcal{A}_1^c}\|_1 = q_2 \|\lambda_{\bullet 2}^{*\mathcal{A}_1^c}\|_1$, a constant times the sum of the absolute values of λ^*_{2i} on $\mathcal{A}_1^c$. Intuitively, the expression in (12) then states that $F_1 \in \mathcal{F}^{exact}$ iff $\|\lambda_{\bullet 2}^{*\mathcal{A}_1^c}\|_1 > \frac{1}{q_2}\beta^k(v_{\bullet k})$. Thus, Assumption 3 requires a lower bound on the total (absolute) value of the loadings $\lambda_{\bullet 2}$ outside of the set $\mathcal{A}_1$ for F_1 to be in $\mathcal{F}^{exact}$.

To gain intuition for Assumption 3, we consider our three examples from earlier again (with $r = 2$).

1. Suppose $\mathcal{A}_1 \cap \mathcal{A}_2 = \varnothing$ (The two factors affect different, non-overlapping groups of outcomes).
   Then, $\lambda^*_{\bullet 1} \perp v_{\bullet 1}$ implies $v_{\bullet 1} = \lambda^*_{\bullet 2}$, and thus $\beta^1(v_{\bullet 1}) = \beta^1(\lambda^*_{\bullet 2}) = 0$. Further, note that $\|\lambda_{\bullet 2}^{*\mathcal{A}_1^c}\|_1 = \|\lambda^*_{\bullet 2}\|_1$. It follows from $rank(\Lambda^*) = r$ that $\|\lambda_{\bullet 2}\|_1 > 0$, and therefore $\mathcal{A}_1 \cap \mathcal{A}_2 = \varnothing$ implies that (12) holds and $F_1 \in \mathcal{F}^{exact}$. Clearly, the same reasoning can be applied to F_2. Therefore, $\mathcal{A}_1 \cap \mathcal{A}_2 = \varnothing$ implies that $F_1, F_2 \in \mathcal{F}^{exact}$.

2. Suppose $\mathcal{A}_2 \subseteq \mathcal{A}_1$ (The second factor F_2 affects a subset of the outcomes affected by F_1).

Then, $\lambda^{*\mathcal{A}_1^c}_{\bullet 2} = 0$, and it immediately follows that $F_1 \notin \mathcal{F}^{exact}$. Thus, whenever $\mathcal{A}_k$ is a superset of another active set $\mathcal{A}_l$, F_k cannot be a member of the set $\mathcal{F}^{exact}$.

3. Suppose $|\mathcal{A}_k| \asymp n$ for $k = 1, 2$ as $n \to \infty$, and $\lambda^*_{ik} \overset{i.i.d.}{\sim} N(0, \sigma)$ if $i \in \mathcal{A}_k$, and $\lambda^*_{ik} = 0$ otherwise.
   Then, it can be shown (see Online Appendix C) that $\beta^k(v_{\bullet k}) = O_p(\sqrt{n})$ and thus that $F_1 \in \mathcal{F}^{exact}$ if $\sqrt{n} = o_p(\|\lambda^{*\mathcal{A}_1^c}_{\bullet 2}\|_1)$. Intuitively, this states that $F_1 \in \mathcal{F}^{exact}$ if more than $\sqrt{n}$ outcomes are affected by F_2, but not F_1.

We further discuss our high level Assumption 3, as well as low level assumptions sufficient for our results, in Online Appendix C.

In what follows, we will work with an initial rotation of Λ^*, rather than with Λ^* directly. We will denote this as $\Lambda^0 = \Lambda^* H$, where H is nonsingular. Λ^0 has the property that its columns have equal length and are orthogonal, such that $\frac{\Lambda^{0\prime}\Lambda^0}{n} = I$. While Λ^0 is not unique, clearly such a rotation always exists. Intuitively, one can think of Λ^0 as the rotation of Λ^* that is estimated by the Principal Component estimator, at this point ignoring any estimation error. Importantly, $\frac{\Lambda^{0\prime}\Lambda^0}{n} = I$ implies $\|\lambda_{\bullet k}\|_2 = \|\Lambda^0 \Upsilon\|_2 = \sqrt{n}$ for any $(r \times 1)$ vector Υ with $\|\Upsilon\|_2 = 1$. When considering the l_1-norm of $\lambda_{\bullet k}$ for different linear combinations Υ, we therefore hold the l_2-norm of $\lambda_{\bullet k}$ constant across those combinations. To this end, consider a the following optimization problem:

$$\min_R \|\Lambda^0 R\|_1 \qquad \text{such that } R \text{ is nonsingular and } \|R_{\bullet k}\|_2 = 1 \ \forall k. \tag{13}$$

Noting that $\|\Lambda^0 R\|_1 = \sum_{k=1}^r \| \sum_{l=1}^r \lambda^0_{\bullet l} R_{lk}\|_1$, we see that (13) is separable in k and consists of k identical parts up to the nonsingularity constraint. We thus consider one part at a time:

$$\min_{R_{\bullet k}} \| \sum_{l=1}^r \lambda^0_{\bullet l} R_{lk}\|_1 \qquad \text{such that } R'_{\bullet k} R_{\bullet k} = 1. \tag{14}$$

Defining $R^*_{\bullet k}$ as the vector that gives $\lambda^*_{\bullet k} = \Lambda^0 R^*_{\bullet k}$, we obtain obtain the following result.

Theorem 2. *Suppose Assumptions 1 and 3 hold and we have access to a rotation of the true loading matrix, $\Lambda^0 = \Lambda^* H$, where H is nonsingular and $\frac{\Lambda^{0\prime}\Lambda^0}{n} = I$. If $F_k \in \mathcal{F}^{exact}$, the minimization in* (14) *has a local minimum at $R^*_{\bullet k}$.*

Theorem 2 states that the ℓ_1-norm of λ^*_k is a minimum of (14) if F_k is a local factor. Note that any set of local minima of (14) for $k = 1, \ldots, r$ is also a local minimum of (13). By imposing the additional constraint that R is nonsingular, we rule out that multiple columns in

R lead to the same $\lambda^*_{\bullet k}$ and ensure that any solution $\tilde{\Lambda} = \Lambda^0 \tilde{R}$ to (13) spans the same space as Λ^0.

4.2 Under Approximate Sparsity and Estimation Error

Theorems 1 and 2 required exact sparsity, which is quite restrictive. We therefore next redefine the sets $\mathcal{A}_k$ to allow for approximate sparsity in the loading matrix.

Assumption 4. *For each factor F_k, we can partition the set of indices $i = 1, 2, \ldots, n$ into a set of indices $\mathcal{A}_k$ with cardinality $|\mathcal{A}_k|$ and its complement, such that as $n \to \infty$,*

(a) $\sum_{i \notin \mathcal{A}_k} \left|\lambda^*_{ik}\right| = O(\sqrt{n})$.

(b) $|\mathcal{A}_k| > c_0 n$ *for some* $c_0 > 0$.

(c) $\left|\lambda^*_{ik}\right| > c \ \forall i \in \mathcal{A}_k$ *and* $\left|\lambda^*_{ik}\right| < C \ \forall i$ *for constants* $0 < c, C < \infty$.

Assumption 4(a) relaxes the definition of $\mathcal{A}_k$ to allow for approximate sparsity. We may still think of $\mathcal{A}_k$ as the active (or important) set for a given factor F_k, but F_k may now also affect other outcomes, with Assumption 4(a) restricting how much. Assumption 4(b) can be thought of as a pervasiveness assumption. Together with Assumption 4(c), it states that each factor affects a constant fraction of all outcomes, which is commonly maintained in the literature. For our main result (Theorem 3), we require access to a $\sqrt{n}$ consistent estimate of the space spanned by Λ^* and the ability to obtain such a $\sqrt{n}$ consistent estimate generally implies that factors must be pervasive (Freyaldenhoven 2022).

Assumption 5. *Let V_k denote the set of all linear combinations $v_{\bullet k}$ of $\lambda^*_{\bullet l}$, $l = 1, \ldots, r$, such that $\|v_{\bullet k}\|_2^2 = n$ and $\lambda^*_{\bullet k} \perp v_{\bullet k}$ and define*

$$\beta^k(v_{\bullet k}) = \left| \sum_{i \in \mathcal{A}_k} |v_{ik}| \, \mathbf{1}\{\lambda^*_{ik} v_{ik} \geq 0\} - \sum_{i \in \mathcal{A}_k} |v_{ik}| \, \mathbf{1}\{\lambda^*_{ik} v_{ik} < 0\} \right|. \tag{15}$$

Then, there exists a set of factors $\mathcal{F}$, such that, for any $F_k \in \mathcal{F}$, and some $c_{min} > 0$ and $N < \infty$, whenever $n > N$:

$$\|v_{\bullet k}^{\mathcal{A}_k^c}\|_1 - \beta^k(v_{\bullet k}) > c_{min} n^{\frac{3}{4}} \qquad \forall v_{\bullet k} \in V_k. \tag{16}$$

Assumption 5 slightly strengthens Assumption 3 in order to accommodate non-zero entries of $\lambda^*_{\bullet k}$ on $\mathcal{A}_k^c$, with a trade-off similar to the one we observed in Assumptions 2 and

3.[7]

So far, we assumed access to an initial rotation of Λ^*, $\Lambda^0 = \Lambda^* H$. In practice, we will only have access to an estimate of such a rotation. We remain agnostic about where such an initial estimate may come from but simply require $\sqrt{n}$ consistency.

Assumption 6. *We have access to an initial estimate Λ^0 with $\frac{\Lambda^{0\prime}\Lambda^0}{n} = I$, such that $\|\lambda^0_{ik} - \lambda^*_{i\bullet}H_{\bullet k}\| = O_p(\frac{1}{\sqrt{n}})$, where H is nonsingular and the elements in H^{-1} are bounded above by some constant $C < \infty$.*

A large literature exists detailing various conditions on the primitives of the model that allows an estimate with this rate. An obvious candidate that achieves $\sqrt{n}$ consistency under some regularity conditions would be the Principal Component estimator (Stock and Watson 2002, Bai and Ng 2002, Bai 2003)[8]. This is the estimator we use in our simulations and applications. The main result of the paper follows.

Theorem 3. *Suppose $n \to \infty$, Assumptions 4-6 hold, and $F_k \in \mathcal{F}$. Then, there exists a local minimum of* (14) *at $\bar{R}_{\bullet k}$, with $\bar{\lambda}_{\bullet k} = \Lambda^0 \bar{R}_{\bullet k}$, such that*

$$\bar{\lambda}_{ik} = \lambda^*_{ik} + O_p(n^{-1/4}) \tag{17}$$

and

$$\frac{1}{n}\|\lambda^*_{\bullet k} - \bar{\lambda}_{\bullet k}\|^2 = O_p(n^{-\frac{1}{2}}). \tag{18}$$

Theorem 3 shows that the minimization problem in (14) can be used to consistently estimated the loadings (and the individual loading vectors) of local factors, even under approximate sparsity and when allowing for estimation error in the initial estimate Λ^0. Since the number of elements in Λ^* increases with n, Theorem 3 also establishes consistency of the estimated loadings in terms of an averaged norm (similar to the results in, e.g., Bai and Ng 2002, Ando and Bai 2017).

In Theorems 1, 2, and 3 we established identification for the loadings corresponding to factors in $\mathcal{F}^b, \mathcal{F}^{exact}$, and $\mathcal{F}$ respectively. Before we conclude this section, we briefly summarize the similarities and differences between these sets in a simple example. To this end, suppose $r = 2$ and $\lambda_{ik} \sim N(0,1)$ if $i \in \mathcal{A}_k$ for $k = 1, 2$. We consider three different cases.

[7] In fact, the only difference between Assumptions 3 and 5 is the lower bound on the RHS of (16). We again point the reader to Online Appendix C for further discussion of this high level assumption.

[8] Also see Bai and Ng [2021] for a more detailed discussion of the rotation matrix H.

1. Under exact sparsity, no estimation error of the loading space, and using the (infeasible) rotation of Λ that minimizes the ℓ_0-norm:
   $\lambda^*_{\bullet k}$ will be identified if $\left|\mathcal{A}_k^c \cap \mathcal{A}_l\right| > 1$.

2. Under exact sparsity, no estimation error of the loading space, and using the rotation of Λ that minimizes the ℓ_1-norm:
   $\lambda^*_{\bullet k}$ will be identified if $\left|\mathcal{A}_k^c \cap \mathcal{A}_l\right| > C\sqrt{n}$ for all $C < \infty$.

3. Under approximate sparsity, a $\sqrt{n}$-consistent estimate of the loading space, and using the rotation of Λ that minimizes the ℓ_1-norm:
   $\lambda^*_{\bullet k}$ will be identified if $\left|\mathcal{A}_k^c \cap \mathcal{A}_l\right| > c_{min} n^{\frac{3}{4}}$ for some $c_{min} > 0$.

The last case is the most relevant in practice. Since it uses the ℓ_1-norm, it is feasible to implement, and it allows for both estimation error in an initial estimate of the loading space and approximate sparsity. The "price to pay" to still guarantee identification is that we need the active sets $\mathcal{A}_1$ and $\mathcal{A}_2$ to be more different compared to, for instance, the infeasible setup in the first case.

Theorem 3 suggests the following simple two-step algorithm to consistently estimate any loading vectors that correspond to local factors.

1. Obtain a $\sqrt{n}$ consistent estimate Λ^0 that forms an orthonormal basis of the loading space (e.g. by extracting the leading r principal components).

2. Find the rotation $\tilde{R}$ that minimizes the l_1-norm of the loadings:

$$\min_R \| \sum_{l=1}^{r} \lambda^0_{\bullet l} R_{lk} \|_1 \qquad \text{such that } R'_{\bullet k} R_{\bullet k} = 1, R \text{ nonsingular.} \tag{19}$$

By Theorem 3, if there are local factors present in the data, their true loading vectors will be (close to) a local argmin of (19).

Remark 5. The optimization problem in (19) is still computationally challenging as it involves finding multiple minima of a non-convex function over the surface of an r-dimensional sphere. In practice, we translate the problem into spherical coordinates, which turns the constraint optimization in (19) into an unconstrained optimization problem that is much easier to solve.

We discuss this transformation, and our algorithmic implementation in general, in more detail in Online Appendix E.

Remark 6. Throughout, all results concerned the loadings Λ^*. A natural question is to what extent our consistency results for the individual loading vectors translate into consistency results for the corresponding individual factors. Perhaps surprisingly, even the local factors will generally not be identified. In other words, knowing a loading vector $\lambda^*_{\bullet k}$, and thus how the corresponding factor F_k affects all outcomes, is *not* sufficient to identify the corresponding factor F_k without further assumptions.

For intuition, suppose we were to form estimates for the factors at each time period by a cross-sectional regression of the outcomes on the estimated factor loadings, such that

$$F_t = (\tilde{\Lambda}'\tilde{\Lambda})^{-1}\tilde{\Lambda}'X_t \qquad \text{for } t = 1, \ldots, T. \tag{20}$$

Intuitively, consistency of F_{kt} requires knowledge of *all* loading vectors $\lambda^*_{\bullet k}$, $k = 1, \ldots, r$. However, a setting in which all factors are local (in which case the entire loading matrix Λ is identified, such that $\tilde{\Lambda} \approx \Lambda^*$) appears unlikely in most economic applications. We also conjecture that one could achieve identification of the individual factors under additional restrictions (such as orthogonality of the factors or the loadings).

Remark 7. Our results establish identification for the loading vectors of the local factors in $\mathcal{F}$. In Section 6, we discuss some heuristics on how to determine which factors are in this set. Formally identifying which loading vectors are identified would be another interesting avenue for future research.

5 Simulations

This section presents results from Monte Carlo simulations to evaluate the performance of our proposals in finite sample. We start by revisiting the baseline DGP from our stylized example in Section 3.1 and provide some more details about this DGP. The factors $F_k, k = 1, 2$ are generated jointly normal with a correlation of 0.3, unit variances, and are $i.i.d.$ over time. The error terms have the following correlation structure:

$$\begin{aligned} e_{ti} &= \rho e_{t-1,i} + (1-\rho^2)^{1/2} v_{it}, \\ v_{ti} &= \beta v_{t,i-1} + (1-\beta^2)^{1/2} u_{it}, \quad u_{it} \overset{i.i.d.}{\sim} N(0,1), \end{aligned}$$

with $(\rho, \beta) = (0.3, 0.1)$, which Onatski [2010] argues are good approximations to many financial datasets.

First, we test whether local factors are present. To this end, we simulate 2000 realizations of our baseline DGP. For each realization, we simulate new loadings in Λ^*. Using $\mathbf{1}\{\hat{\mathcal{L}}_0(\tilde{\Lambda}) >$

$\gamma n\}$ to test for the presence of local factors, we successfully detect the existence of local factors in all 2000 simulation runs for this DGP.

Our next goal is to recover Λ^*. To summarize the performance of an estimator across simulation runs, we use the cosine similarity between the columns in Λ^* and an estimate $\hat{\Lambda}$. Because the factors can always be reordered, for each true loading vector $\lambda^*_{\bullet l}$, we use the maximum cosine similarity with any estimated loading vector to measure how closely we are able to recover $\lambda^*_{\bullet l}$. Formally, define the maximum cosine similarity $MC_l(\hat{\Lambda})$ between the true loading vector $\lambda^*_{\bullet l}$ and an estimate $\hat{\Lambda}$ as

$$MC_l(\hat{\Lambda}) = \max_k \frac{\hat{\lambda}'_{\bullet k}\lambda^*_{\bullet l}}{\|\hat{\lambda}_{\bullet k}\|\|\lambda^*_{\bullet l}\|} \quad \text{for } l = 1, \dots, r. \tag{21}$$

Thus, a value of MC_l close to one means that one of the estimated loading vectors $\hat{\lambda}_{\bullet k}$, $k = 1, \dots, r$, is close to $\lambda^*_{\bullet l}$.

The maximum cosine similarity corresponding to Figures 1-4 in Section 3.1 is depicted in the first two columns of Table 1. The first column confirms that the Principal Component estimator does not successfully recover either of the two loading vectors. On the other hand, consistent with Figure 4, our proposed estimator can successfully identify the true loading matrix Λ^*. Because both factors symmetrically affect the same number of outcomes in our baseline DGP, the two rows look similar. While $\lambda^*_{ik} \overset{i.i.d.}{\sim} U(0.1, 2.9)$ for $i \in \mathcal{A}_k$ was chosen to

For $i \in \mathcal{A}_k$:	$\lambda^*_{ik} \sim U(0.1, 1.9)$		$\lambda^*_{ik} \sim N(1, 1)$	
Estimator $\hat{\Lambda}$	Λ^0	$\tilde{\Lambda}$	Λ^0	$\tilde{\Lambda}$
MC_1	0.777	0.990	0.773	0.994
MC_2	0.780	0.990	0.774	0.994

Table 1: Maximum cosine similarity $MC_l(\hat{\Lambda})$ across DGPs and estimators. Λ^0 refers to the Principal Component estimator, while $\tilde{\Lambda}$ represents our proposed rotation that minimizes the ℓ_1-norm across all rotations. Depicted are averages based on 2000 realizations.

satisfy the upper and lower bounds assumed on Λ^* in the previous section, we also consider $\lambda^*_{ik} \overset{i.i.d.}{\sim} N(1, 1)$ for $i \in \mathcal{A}_k$ in the third and fourth column of Table 1. Column 4 demonstrates that changing the distribution of the loadings λ_{ik} on $\mathcal{A}_k$ has no meaningful impact on our results.

The previous results confirm that our proposed ℓ_1-rotation and testing criterion work well in our baseline DGP: we can reliably detect the presence of local factors, and can correctly

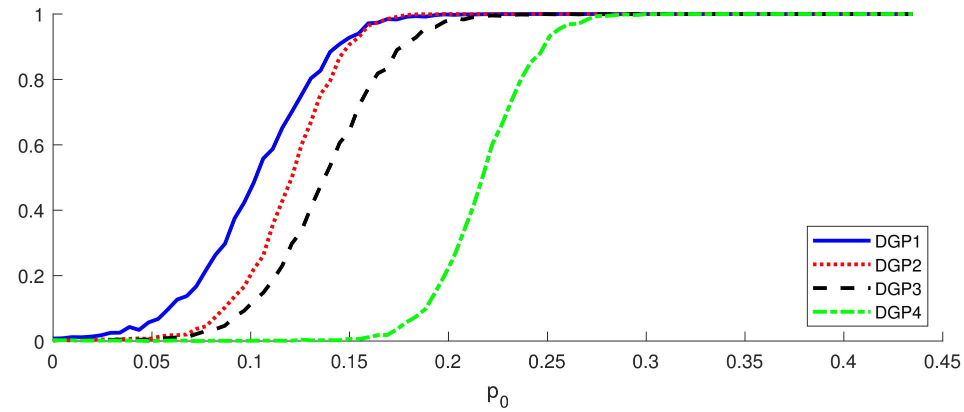


Figure 6: Power curves of testing criterion $\hat{\mathcal{L}}_0$ that tests whether all factors are global. Each line corresponds to a different distribution of the non-zero loadings on Λ^* and depicts the empirical rejection frequency for the null that all factors are global across 2000 simulations as a function of p_0, the fraction of loadings that are equal to zero.

recover the sparsity pattern in the loading matrix, thereby identifying the individual loading vectors. We next consider a variety of data-generating processes to approximate a range of situations a practitioner might encounter in practice.

5.1 Results for a Variety of Data-Generating Processes

We start by varying the degree of sparsity in Λ^*. We use the criterion $\mathbf{1}\{\hat{\mathcal{L}}_0(\tilde{\Lambda}) > \gamma n\}$ introduced in Section 4.1.2 to determine whether local factors are present in the data. We continue to use tuning parameters $\gamma_0 = 0.03$ and $\alpha_\gamma = 0.05$. Figure 6 depicts the empirical frequency with which $\hat{\mathcal{L}}_0(\tilde{\Lambda}) > \gamma n$ when we vary the number of loadings that are equal to zero in Λ^*. $\tilde{\Lambda}$ refers to the estimate for Λ^* obtained by our proposed ℓ_1-rotation.

We consider four variants of our baseline DGP. In DGP1, the fraction of loadings that is equal to zero is p_0 for both factors and $\lambda^*_{ik} \sim N(0,1)$ if $i \in \mathcal{A}_k$. In DGP2, the fraction of loadings that is equal to zero is p_0 for both factors and $\lambda^*_{ik} \sim N(1,1)$ if $i \in \mathcal{A}_k$. In DGP3, one factor affects all outcomes. The fraction of loadings that is equal to zero is p_0 for the second factor and $\lambda^*_{ik} \sim N(1,1)$ if $i \in \mathcal{A}_k$. In DGP4, the fraction of loadings that is equal to zero is p_0 for both factors and $\lambda^*_{ik} \sim U(0.1, 1.9)$ if $i \in \mathcal{A}_k$.

Under DGP1, we achieve a size of 0.01, and correctly detect local factors in more than 95% of realizations if more than 16% of the loadings are equal to zero. With the non-zero entries of Λ^* less concentrated around zero in the remaining three DGPs, our test becomes more conservative, and we obtain sizes of 0.00 for the remaining three DGPs. Similarly, a larger fraction of zero loadings is generally necessary to reliably detect the existence of local factors using our test statistic in these DGPs. However, we conclude that our testing procedure

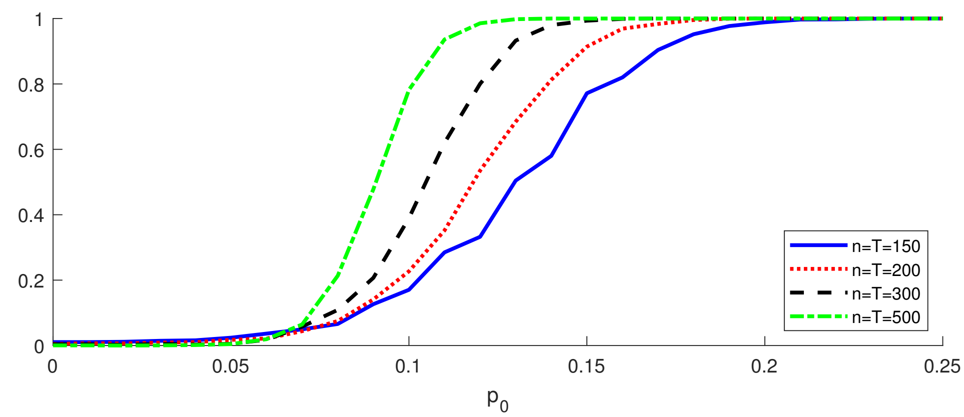


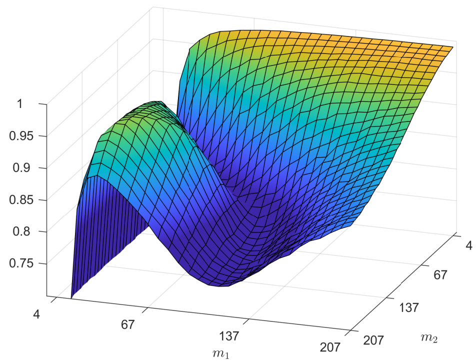
Figure 7: Power curves of testing criterion $\hat{\mathcal{L}}_0$ that tests whether all factors are global for different sample sizes. Each line depicts the empirical rejection frequency for the null that all factors are global across 2000 simulations as a function of p_0, the fraction of loadings that are equal to zero for both factors. For all non-zero loadings, $\lambda^*_{ik} \sim N(1, 1)$.

successfully detects the presence of local factors in DGPs with a significant sparsity pattern in Λ^*: across all DGPs we can detect the existence of local factors in almost all realization whenever at least one loading vector has more than 25% of its entries equal to zero.

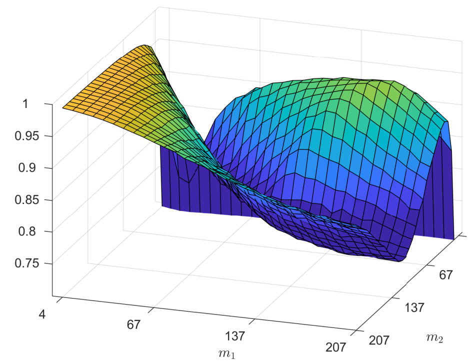
Figure 7 illustrates how the power curve changes for varying sample sizes. It again depicts the empirical rejection frequency (the fraction of simulations with which $\hat{\mathcal{L}}_0(\tilde{\Lambda}) > \gamma n$) as a function of p_0, the fraction of loadings that are equal to zero. The underlying DGP is DGP2, with $\lambda^*_{ik} \sim N(1, 1)$ if $i \in \mathcal{A}_k$. We set $T = n$, varying both n and T together.[9] With $T = n = 150$ our test detects the presence of local factors in more than 95% of realizations when at least 18% of the loadings are equal to zero. As the sample increases, the performance of our testing criterion improves further: With $T = n = 500$ our test detects the presence of local factors in more than 95% of realizations when as little as 12% of the loadings are equal to zero.

We next turn to the estimation of the individual loading vectors again, varying the degree of sparsity in the loading matrix by varying the values of $m_1 = |\mathcal{A}_1|$ and $m_2 = |\mathcal{A}_2|$. We maintain that $\lambda^*_{ik} \overset{i.i.d.}{\sim} N(1, 1)$, for $i \in \mathcal{A}_k$. All other parameters remain unchanged from our baseline DGP. The corresponding result is depicted in Figure 8. Figure 8 depicts how well we are able to estimate the true factor loadings λ^*_1 and λ^*_2 as a function of m_1 and m_2. Panels 8a and 8b depict the performance of the Principal Component estimator for λ^*_1 and λ^*_2. Unsurprisingly, the maximum cosine similarities are generally significantly below one. The exception to this are cases in which one factor is extremely weak. In such cases, the data

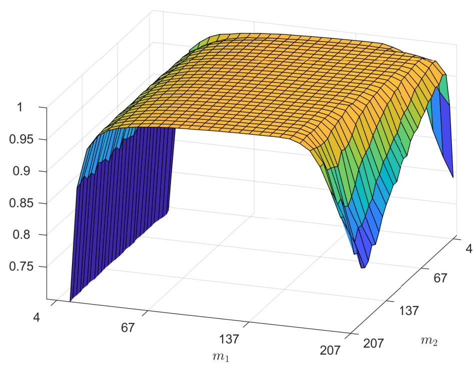
[9]The results are qualitatively similar if we hold T fixed.



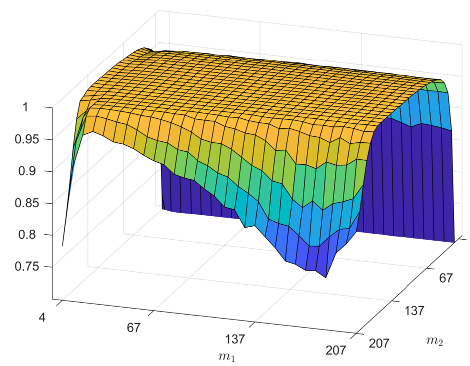
(a) PC Estimator: $MC_1(\Lambda^0)$



(b) PC Estimator: $MC_2(\Lambda^0)$



(c) Rotated Estimator: $MC_1(\tilde{\Lambda})$



(d) Rotated Estimator: $MC_2(\tilde{\Lambda})$

Figure 8: Maximum cosine similarity of estimators with each of the true loading vectors $\lambda^*_{\bullet k}$ as a function of the degree of sparsity in the loading matrix. m_k refers to the number of non-zero entries in the kth column of Λ^*. Depicted are averages over 500 realizations.

effectively has a factor structure with a single factor, there is no rotational indeterminacy, and the sole strong factor is identified.

Panels 8c and 8d depict the maximum cosine similarity for our proposed estimate $\tilde{\Lambda}$. We are able to separately identify the two loading vectors throughout most of the parameter space using our ℓ_1-criterion. The exception occurs in the regions of the parameter space where a factor becomes either "global" or very weak. For example, along the right edge of Figure 8c, F_1 affects all observables. Since only the loading vectors corresponding to factors in $\mathcal{F}$ are identified, and clearly $F_1 \notin \mathcal{F}$ in this region, this is not surprising. On the opposite side of Figure 8c only a handful of outcomes are affected by F_1. λ^*_1 is therefore only weakly identified, and our initial estimate of the loading space is poor, resulting in a maximum cosine similarity less than one. We further conclude from panels 8c-8d that an identification failure for one of the loading vectors does not imply identification failure for the other.

We next increase the size of the model and consider a DGP with $(T, n) = (500, 300)$ and $r = 4$, with a small amount of correlation between the factors. Specifically, let $F_t \sim N(0, \Sigma_F)$, $i.i.d$ over time, with

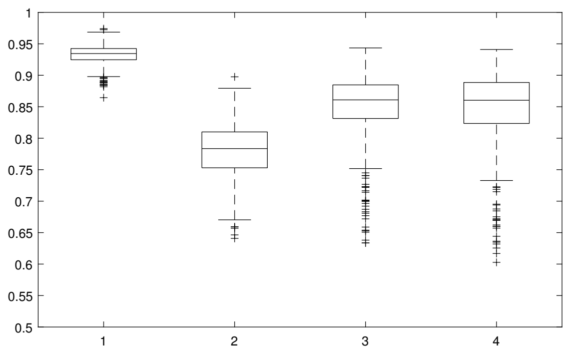
$$\Sigma_F = \begin{bmatrix} 1.0 & 0.3 & 0.0 & 0.0 \\ 0.3 & 1.0 & 0.3 & 0.0 \\ 0.0 & 0.3 & 1.0 & 0.3 \\ 0.0 & 0.0 & 0.3 & 1.0 \end{bmatrix}.$$

The first factor in this DGP is "global," while the remaining three are local to varying degrees. Specifically, the 300-by-4 loading matrix Λ^* has entries $\lambda^*_{ik} \overset{i.i.d.}{\sim} N(1, 1)$ if $i \in \mathcal{A}_k$, and $\lambda^*_{ik} = 0$ otherwise. The subsets $\mathcal{A}_k$ will be of varying size and dictate which variables are affected by each factor k, with the sequence of group sizes given by $\{|\mathcal{A}_k|\}_{k=1}^4 = \{300, 170, 96, 72\}$ for the four factors. The idiosyncratic component e_{it} is created the same way it was in our baseline DGP. Finally, we consider a variant of this DGP in which there is no exact sparsity, but rather an approximate version thereof. Here, $\lambda^*_{ik} \overset{i.i.d.}{\sim} N(0, \frac{1}{n})$, for all $i \in \mathcal{A}^c_k$.

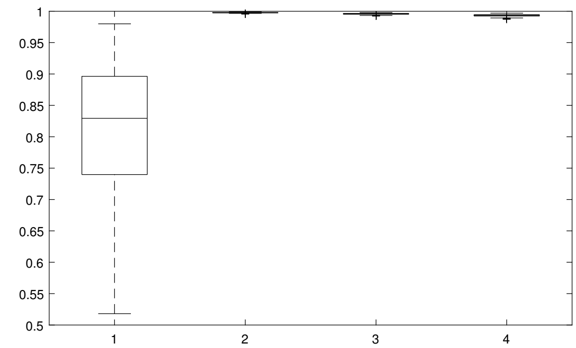
First, we test for the existence of local factors. For both DGPs (exact and approximate sparsity), our proposed testing criterion correctly detects the presence of local factors in all simulation runs.

Figure 9 then uses a boxplot to visualize the performance of Λ^0 and $\tilde{\Lambda}$. It depicts the maximum cosine similarity for each factor across 500 realizations. The data underlying Figures 9a and 9b has an exact sparsity pattern ($\lambda^*_{ik} = 0$ if $i \in \mathcal{A}^c_k$). As expected, we do not consistently recover the true loadings using the Principal Component estimator Λ^0 (cf. Figure 9a). On the other hand, Figure 9b depicts the similarity between the truth, Λ^*, and our proposed estimate $\tilde{\Lambda}$. Since the first factor does not exhibit any sparsity, there is no information in the ℓ_1-norm that could help identify the corresponding loading vector. As a consequence, the similarity is below one, and identification fails for this loading vector. On the other hand, the loading vectors of the three local factors exhibit maximum cosine similarities that are visually indistinguishable from one in all realizations. Underlying Figures 9c and 9d is the variant of our DGP with approximate sparsity in the loading matrix. Based on Figures 9c and 9d, the above conclusions are unchanged. Our proposed estimator $\tilde{\Lambda}$ recovers the loading vectors associated with the three local factors in all realizations.

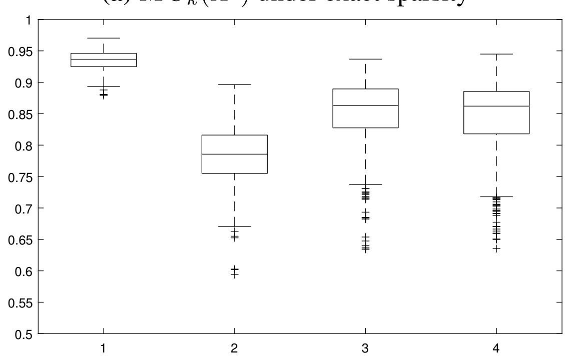
In Online Appendix D, we compare the performance of our proposed estimator to a number of existing heuristics that are currently widely used to simplify the loading matrix, including some of the quartic criteria discussed in Section 3.2. We find that our ℓ_1-rotation



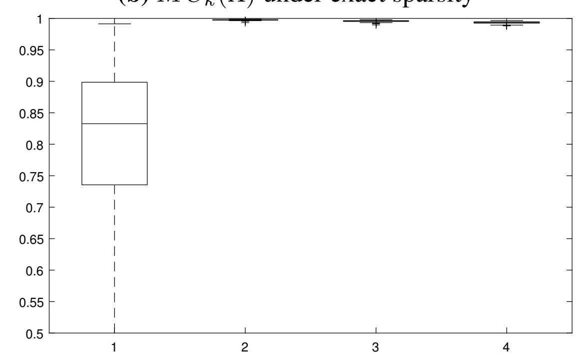
(a) $MC_k(\Lambda^0)$ under exact sparsity



(b) $MC_k(\tilde{\Lambda})$ under exact sparsity



(c) $MC_k(\Lambda^0)$ under approximate sparsity



(d) $MC_k(\tilde{\Lambda})$ under approximate sparsity

Figure 9: Each panel depicts the maximum cosine similarity of an estimator with all four of the true loading vectors $\lambda^*_{\bullet k}$. Λ^0 denotes Principal Component estimator, while $\tilde{\Lambda}$ denotes estimate after proposed rotation. The first factor is global, factors 2-4 are local. Boxplots based on 500 realizations.

performs better than these alternative methods.

6 Applications

We next apply our rotation criterion to two economic applications in which factor models have been widely used, chosen to capture two scenarios a practitioner might encounter. First, we consider a dataset of international stock returns. Because of the global nature of this dataset, we expect the presence of region-specific factors in this dataset. We are therefore interested whether our method can detect these local factors and recover the geographic structure of the data. Second, we consider a large panel of US macroeconomic indicators, where it is less clear a priori whether local factors are present.

6.1 Identifying Common Shocks in International Asset Returns

Let R_{it} denote the return of asset i at time t. We assume that the excess returns, $R_{it} - \mathbb{E}(R_i)$, follow a factor structure, such that

$$x_{it} = R_{it} - \mathbb{E}(R_i) = \lambda_i F_t + e_{it}. \tag{22}$$

We treat the common factors as unobserved, so we need to replace F_t and λ_i by their estimates $\hat{F}_t$ and $\hat{\lambda}_i$. In financial economics, these estimates are commonly obtained by Principal Component Analysis (Connor and Korajczyk 1986, Ludvigson and Ng 2007). We propose to identify the individual loading vectors using our ℓ_1-criterion.

Our dataset consists of daily returns for a large number of stocks from different parts of the world. In particular, it includes individual stock returns for companies that were part of the *DAX30* (Germany), the *FTSE100* (UK), the *S&P100* (US), the *CAC40* (France), or the *TA100* (Middle East) on April 23, 2015.[10] In total, the data covers 272 stocks spanning 687 observations from 01/01/2011 until 03/20/2015. We determine the number of factors to be eight using Bai and Ng [2002]'s Information Criterion, and will accordingly use $r = 8$ in what follows.

To estimate the space spanned by these eight factors, we then estimate the leading eight principal components. Unsurprisingly, we find that each of the eight principal components loads on most of the 272 individual stocks. The estimated loadings corresponding to the Principal Component estimator Λ^0 can be found in Online Appendix Figure 11.

In contrast, Figure 10 depicts our proposed estimator $\tilde{\Lambda}$. The thin dashed lines separate the geographical groups as described above, in the order of Frankfurt, London, New York, Paris, and Tel Aviv. In contrast to the Principal Component estimate, we see that its loading vectors are highly concentrated on a subset of outcomes.[11] It reveals strong regional dependencies in asset returns as illustrated in Table 2. For example, $\tilde{\lambda}_{\bullet 1}$ is almost entirely concentrated on stocks in the Middle East, $\tilde{\lambda}_{\bullet 2}$ and $\tilde{\lambda}_{\bullet 3}$ are concentrated on stocks in the US, and $\tilde{\lambda}_{\bullet 8}$ is concentrated on stocks in the UK. The exception is $\tilde{\lambda}_{\bullet 4}$, whose large entries are dispersed geographically. However, all 25 stocks with a loading larger than two for this factor belong

[10] We further restrict the stocks in the TA100 to those with a weight by market capitalization in the TA100 of at least 0.5 %. This makes the remaining stocks comparable in size to the rest of the sample. For a more detailed discussion of the data, see Online Appendix G.

[11] We again stress that there is no "shrinkage" involved in our estimator, such that our sparse representation of the factors fits the data exactly as well as a rotation with dense loadings. This also implies that none of the estimated loadings in $\tilde{\Lambda}$ will be exactly equal to zero. A further regularization step is beyond the scope of this paper. See Pelger and Xiong [2021] for a potential approach to such regularization.

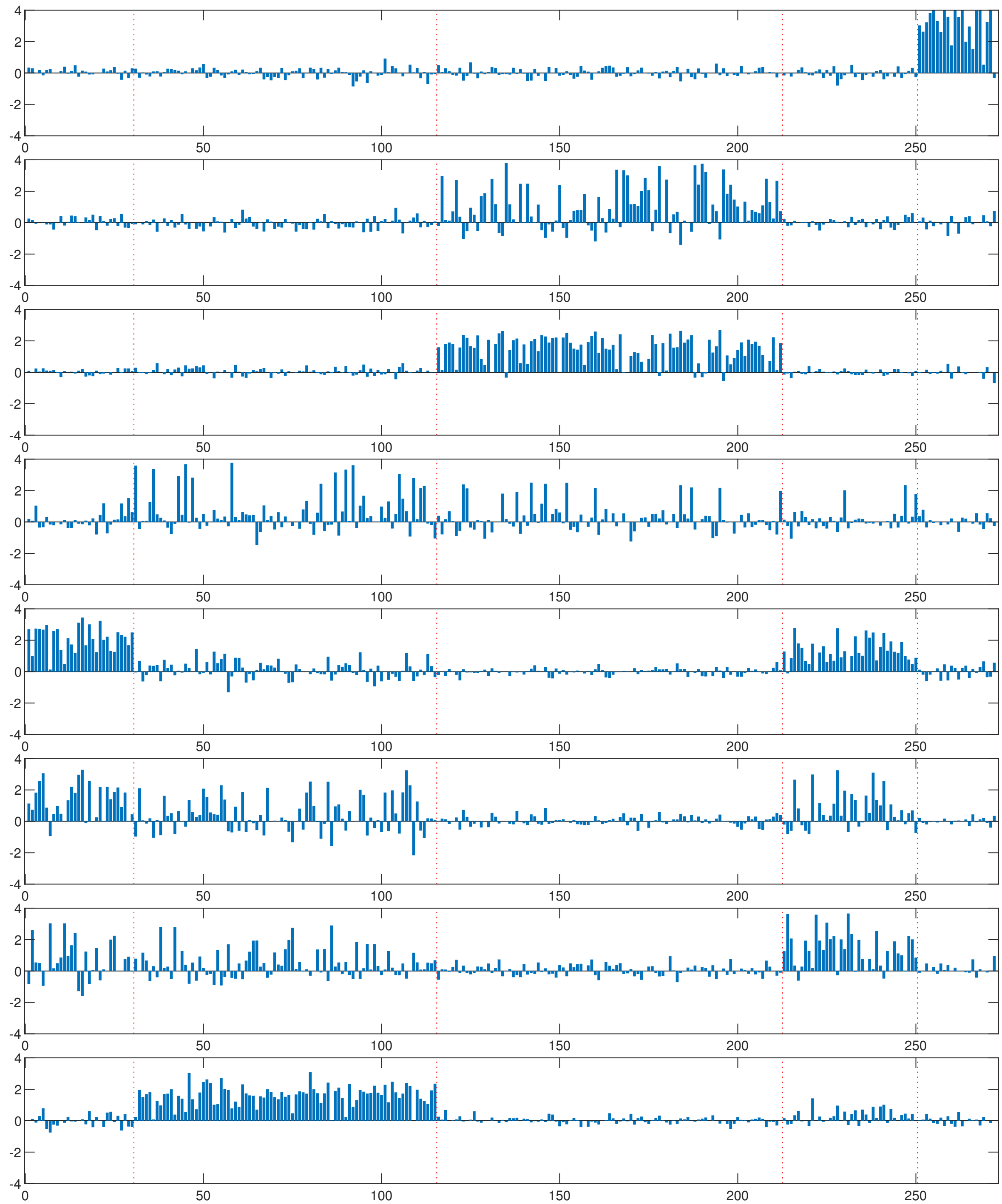


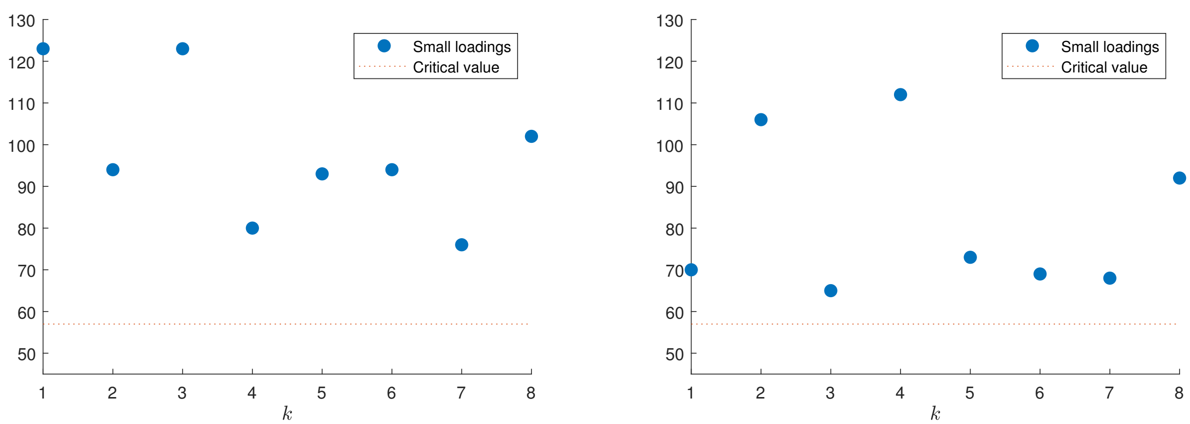
Figure 10: Illustration of the rotated loading vectors $\tilde{\lambda}_{\bullet k}$ for $k = 1, \ldots, 8$ in panel of international asset returns. Bars correspond to the loadings of the 272 individual stocks for the kth estimated loading vector. Geographical groups are Germany, UK, US, France, and Middle East, separated by dashed lines.

to the *Oil & Gas* or the *Mining* sector, enabling us to clearly label this a sector-specific factor.

w Factor	Region	Sector
1	Middle East	
2	US	
3	US	
4	Global	Natural Resources (Oil and Mining)
5	Germany, France	
6	Germany, France, UK	
7	Germany, France, UK	
8	UK	

Table 2: Interpretation of individual factors in panel of international asset returns, based on estimated loading matrix $\tilde{\Lambda}$.

Figure 11a depicts the number of coefficients in each estimated loading vector $\tilde{\lambda}_{\bullet k}$ that are "small", where small is defined as in Section 4.1.2. It illustrates how, in this dataset, all columns of $\tilde{\Lambda}$ exhibit significant sparsity, and our testing criterion therefore finds the existence of local factors in this dataset. Note that the fact that some factors have very similar sets of non-zero loadings (e.g. $\mathcal{A}_2 \approx \mathcal{A}_3$) , means that the corresponding loading vectors are identified only jointly and not separately.



(a) Number of small elements in the rotation of Λ^0 that minimizes the ℓ_1-norm, $\tilde{\Lambda}$.

(b) Number of small elements in the rotation of Λ^0 that maximizes the Varimax criterion, $\ddot{\Lambda}$.

Figure 11: For each $k = 1, \ldots, 8$, the points above represent number of small elements in $\lambda_{\bullet k}$, the kth column in Λ, for daily excess returns of an international sample of stock returns. Dotted red line indicates critical value for $\hat{\mathcal{L}}_0$.

Alternatively, one could use the widely used Varimax criterion (Kaiser [1958]) to simplify the estimated loading matrix Λ^0. The result is depicted in Online Appendix G. In order to compare the performance of our ℓ_1-rotation with the Varimax rotation, Figure 11b depicts the number of coefficients that are small in each vector of $\ddot{\Lambda}$, the Varimax rotation. We conclude that the rotation of Λ^0 that minimizes the ℓ_1-norm, $\tilde{\Lambda}$, has significantly more small loadings than the rotation of Λ^0 that maximizes the Varimax criterion, $\ddot{\Lambda}$.

6.2 Local Factors and the Illusion of Sparsity in Macroeconomic Forecasting

We next apply our identification strategy to a large panel of US macroeconomic indicators. In particular, we use the FRED-QD data collected and maintained by Michael W. McCracken.[12] Our final sample contains 206 quarterly observations of 166 macroeconomic variables, primarily for the US economy.

Two papers that have looked into the nature of the optimal forecasting model in the context of a very similar dataset are De Mol et al. [2008] and Giannone et al. [2021]. Both papers investigate how forecasts that use sparsity inducing regularization compare to regularization methods that do not lead to a sparsity pattern in the predictors (such as ridge regressions or factor augmented regressions). In a Bayesian framework, Giannone et al. [2021] find a significant sparsity pattern for the predictors (more than 75% of their regressors have a coefficient of zero), but note that there is substantial uncertainty about the identity of the relevant predictors. Specifically, they note that the selection of the relevant predictors varies across posterior draws. They posit that these findings may reflect the fact that many predictors contain similar information. Similarly, De Mol et al. [2008] make the following two observations:

1. "The high correlation of the Lasso forecast with the PC forecast suggests that our data is highly collinear: Under collinearity, when appropriately selected, a few variables should capture the essence of the covariation of the data and, as principal components, span approximately the space of the common factors."

2. "The selection [of variables by the Lasso] is different at different points in the sample, although selected variables generally belong to the same economic category."

These observations can be rationalized by the presence of local factors, with each factor affecting only a subset of the observed indicators (which will generally belong to the same

[12]Data are available at `https://research.stlouisfed.org/econ/mccracken/fred-databases`. Versions of this dataset have been used extensively in the literature on macroeconomic forecasting (De Mol et al. 2008, Stock and Watson 2016). For a full description of the data, we refer the reader to McCracken [2019]. We use data from 1967Q1-2019Q1 and follow the transformations of the raw data as outlined in McCracken and Ng [2016] to achieve stationarity and remove a small number of outliers. We only use the disaggregated time series in our estimation of the factor structure and disregard the aggregates (Boivin and Ng 2006, Stock and Watson 2016). We also drop a small number of series with missing observations.

economic category). Figure 12 illustrates. It depicts a scenario with 4 local factors F_{kt}, $k = 1, \ldots, 4$. Each factor F_{kt} affects the outcomes in $\mathcal{A}_k$, denoted by $X_t^{\mathcal{A}_k}$. The optimal forecast is based on the factors.

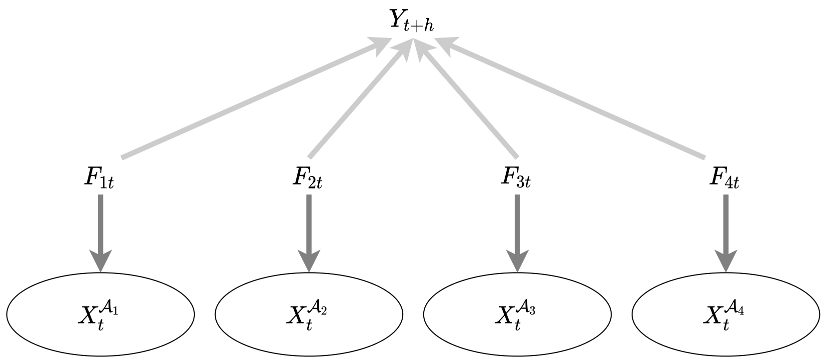


Figure 12: Illustration of local factors in a forecasting exercise. A factor F_{kt} affects only outcomes in $\mathcal{A}_k$, denoted by $X_t^{\mathcal{A}_k}$.

In such a setting, a regularized estimator that induces sparsity in the individual components of X_t will tend to select a single variable from each group $\mathcal{A}_k$ as a noisy proxy for F_{kt}. The selected set of regressors will thus approximately span the space of the common factors. However, selection within groups will be unstable and sensitive to minor perturbations of the data, thereby leading to varying variable selection from the same groups across subsamples or posterior draws.

To determine whether a "group structure" with local factors like the one depicted in Figure 12 is present in our panel of macroeconomic indicators, we first determine the number of factors to be eight, using the Information Criterion of Bai and Ng [2002], and accordingly use $r = 8$ in what follows. To estimate the space spanned by these eight factors, we then estimate the leading eight principal components. Unsurprisingly, these load on most of the 166 observed outcomes. The estimated loadings using the Principal Component estimator Λ^0 can be found in Online Appendix Figure 8. In contrast, Figure 13 depicts our proposed estimator $\tilde{\Lambda}$. In order to gain an understanding of the factors, Table 3 reproduces the grouping of variables as suggested in McCracken [2019], which is in turn based on Stock and Watson [2012]. The corresponding groups of variables are separated by dashed lines in Figure 13.

The first factor almost exclusively drives all price variables (group 6), allowing an easy interpretation as an aggregated price index of which we observe multiple measurements. The second factor is mainly associated with household balance sheets and stock markets (groups 10 and 13). This captures the intuitive notion that an increase in asset prices will be associated with an improvement in household balance sheets. Accordingly, almost all of those indicators

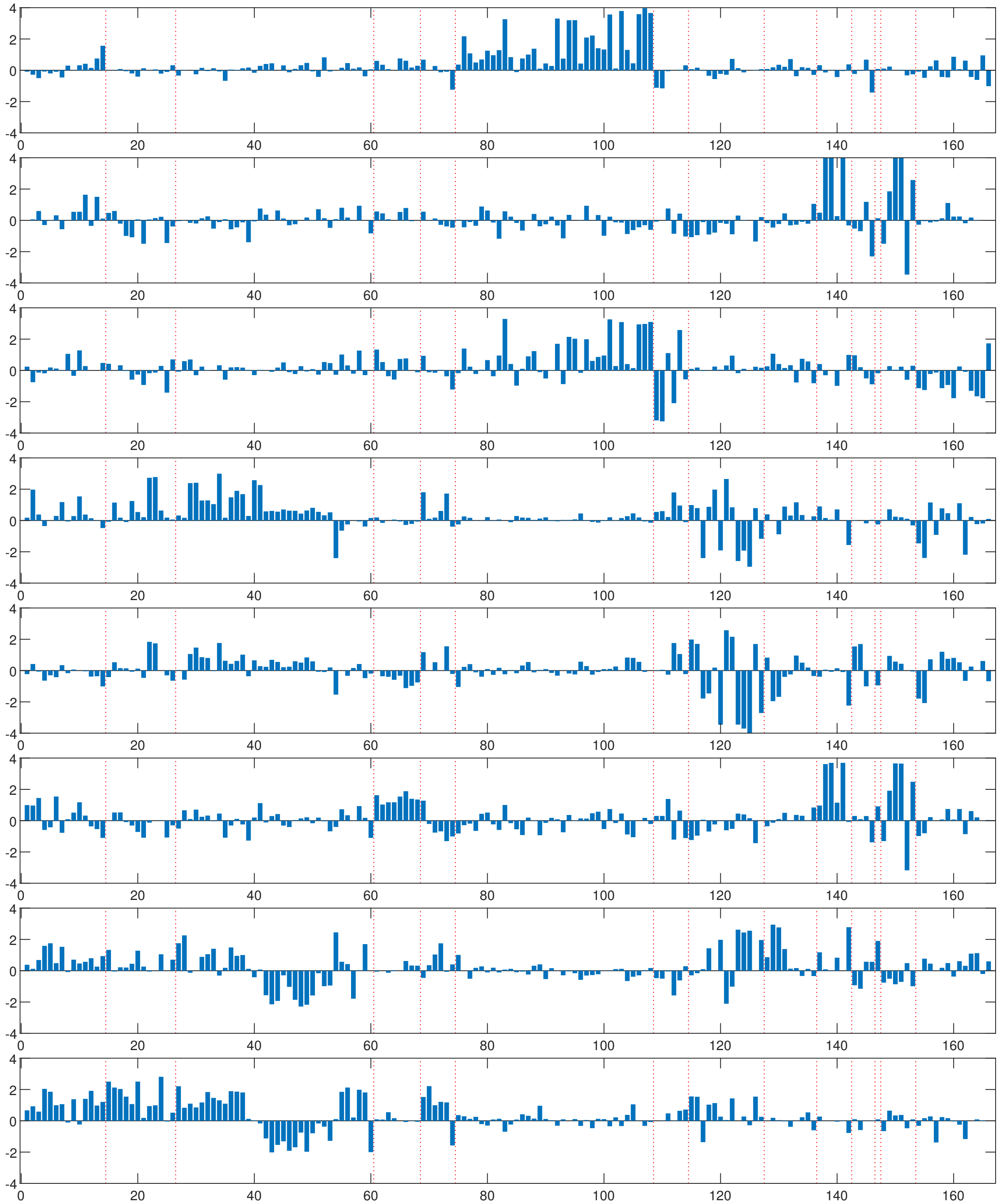


Figure 13: Illustration of the rotated loading vectors $\tilde{\lambda}_{\bullet k}$ for $k = 1, \ldots, 8$ in panel of macroeconomic indicators. Bars correspond to the 166 individual indicators for the kth estimated loading vector. Groups of variables are separated by dashed lines (see Table 3).

Group	Category	Associated variables
1	National Income and Product Accounts (NIPA)	1-14
2	Industrial Production	15-26
3	Employment and Unemployment	27-60
4	Housing	61-68
5	Inventories, Orders, and Sales	69-74
6	Prices	75-108
7	Earnings and Productivity	109-114
8	Interest Rates	115-127
9	Money and Credit	128-136
10	Household Balance Sheets	137-142
11	Exchange Rates	143-146
12	Other	147
13	Stock Markets	148-153
14	Non-Household Balance Sheets	154-166

Table 3: Grouping of variables in panel of US macroeconomic indicators.

are associated with positive loadings, with the exception of the dividend yield, which has a large negative loading.

The third factor mainly affects Price variables, Earnings and Productivity indicators, and Non-Household Balance Sheet variables. The fourth factor mainly affects interest rates, employment indicators, and industrial production. However, while the first two loading vectors exhibit a clear sparsity pattern, in line with our Assumptions, the picture is less clear for subsequent factors. For example, the outcomes affected by the third factor are approximately a superset of those affected by the first factors. As discussed in Section 4, $\mathcal{A}_3 \supseteq \mathcal{A}_1$ would rule out identification of the loading vector for the third factor. Similarly, the set of affected outcomes are very similar for the fourth and fifth factors, suggesting their loadings may not be separately identified. We therefore refrain from interpreting additional factors.

Since there exist multiple rotations of Λ^0 that have a significant sparsity pattern, our testing criterion again suggests the existence of local factors in this dataset. Since the conclusions are similar, we relegate the equivalent to Figure 11 to Online Appendix F. However, we again find that our $\ell 1$-rotation criterion performs better than the Varimax rotation criterion: the rotation of Λ^0 that minimizes the ℓ_1-norm, $\tilde{\Lambda}$, has significantly more small loadings than the rotation of Λ^0 that maximizes the Varimax criterion, $\ddot{\Lambda}$.

In conclusion, we find strong evidence that there are indeed local factors present in the data (e.g. a price index), which can serve as an explanation for the "Illusion of Sparsity" found by Giannone et al. [2021].

7 Conclusion

We introduce a new rotation criterion to simplify the loading matrix in factor models. Our rotation criterion minimizes the ℓ_1-norm of the loadings and is theoretically appealing. Unlike existing heuristics, such as the Varimax criterion (Kaiser [1958]), we derive theoretical guarantees for our rotation criterion if the true loading matrix is sparse: under (approximate) sparsity in the loading matrix, our ℓ_1-rotation can be used to identify the individual loading vectors. We further introduce a way to determine whether local factors are present in a given dataset.

Our ℓ_1-rotation criterion performs well across simulations and two economic applications, where it outperforms existing and widely used rotation criteria. In our two applications, we find strong evidence that local factors are indeed present in the data in both cases. In both applications our method estimates sensible economic objects, which a researcher would not be able to recover otherwise.

References

Seung C Ahn and Alex R Horenstein. Eigenvalue ratio test for the number of factors. *Econometrica*, 81(3):1203–1227, 2013.

Tomohiro Ando and Jushan Bai. Clustering huge number of financial time series: A panel data approach with high-dimensional predictors and factor structures. *Journal of the American Statistical Association*, 112(519):1182–1198, 2017.

Jushan Bai. Inferential theory for factor models of large dimensions. *Econometrica*, 71(1): 135–171, 2003.

Jushan Bai and Serena Ng. Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221, 2002.

Jushan Bai and Serena Ng. Principal components estimation and identification of static factors. *Journal of Econometrics*, 176(1):18–29, 2013.

Jushan Bai and Serena Ng. Approximate factor models with weaker loadings. arxiv preprint arxiv:2109.03773, 2021.

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

Peter Bühlmann and Sara Van De Geer. *Statistics for high-dimensional data: Methods, theory and applications*. Springer Science & Business Media, 2011.

John B Carroll. An analytical solution for approximating simple structure in factor analysis. *Psychometrika*, 18:23–38, 1953.

In Choi, Dukpa Kim, Yun Jung Kim, and Noh-Sun Kwark. A multilevel factor model: Identification, asymptotic theory and applications. *Journal of Applied Econometrics*, 33(3): 355–377, 2018.

Gregory Connor and Robert A Korajczyk. Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of Financial Economics*, 15(3): 373–394, 1986.

Christine De Mol, Domenico Giannone, and Lucrezia Reichlin. Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics*, 146(2):318–328, 2008.

Simon Freyaldenhoven. Factor models with local factors—determining the number of relevant factors. *Journal of Econometrics*, 229(1):80–102, 2022.

Domenico Giannone, Michele Lenza, and Giorgio E Primiceri. Economic predictions with big data: The illusion of sparsity. *Econometrica*, 89(5):2409–2437, 2021.

Alan E Hendrickson and Paul Owen White. Promax: A quick method for rotation to oblique simple structure. *British Journal of Statistical Psychology*, 17(1):65–70, 1964.

Robert I Jennrich. Rotation to simple loadings using component loss functions: The oblique case. *Psychometrika*, 71:173–191, 2006.

Ian T Jolliffe, Nickolay T Trendafilov, and Mudassir Uddin. A modified principal component technique based on the lasso. *Journal of Computational and Graphical Statistics*, 12(3): 531–547, 2003.

Henry F Kaiser. The Varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23:187–200, 1958.

Jeffrey Owen Katz and F James Rohlf. Functionplane – a new approach to simple structure rotation. *Psychometrika*, 39:37–51, 1974.

Sylvia Kaufmann and Christian Schumacher. Bayesian estimation of sparse dynamic factor models with order-independent and ex-post mode identification. *Journal of Econometrics*, 210(1):116–134, 2019.

Johannes T Kristensen. Diffusion indexes with sparse loadings. *Journal of Business & Economic Statistics*, 35(3):434–451, 2017.

Sydney C Ludvigson and Serena Ng. The empirical risk–return relation: A factor analysis approach. *Journal of Financial Economics*, 83(1):171–222, 2007.

Sydney C Ludvigson and Serena Ng. Macro factors in bond risk premia. *Review of Financial Studies*, 22(12):5027–5067, 2009.

Michael W McCracken. FRED-QD updated appendix. Working paper, Federal Reserve Bank of St. Louis, 2019.

Michael W McCracken and Serena Ng. FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34(4):574–589, 2016.

Emanuel Moench, Serena Ng, and Simon Potter. Dynamic hierarchical factor models. *The Review of Economics and Statistics*, 95(5):1811–1817, 2013.

Alexei Onatski. Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics*, 92(4):1004–1016, 2010.

Markus Pelger and Ruoxuan Xiong. Interpretable sparse proximate factors for large dimensions. *Journal of Business & Economic Statistics*, pages 1–23, 2021.

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

William W Rozeboom. Theory & practice of analytic hyperplane optimization. *Multivariate Behavioral Research*, 26(1):179–197, 1991.

David R Saunders. Trans-Varimax-some properties of the ratiomax and Equamax criteria for blind orthogonal rotation. *American Psychologist*, 17(6):395–396, 1962.

James H Stock and Mark W Watson. Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–162, 2002.

James H Stock and Mark W Watson. Disentangling the channels of the 2007—09 recession. *Brookings Papers on Economic Activity*, (1):81–135, 2012.

James H Stock and Mark W Watson. Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. *Handbook of Macroeconomics*, 2:415–525, 2016.

Yoshimasa Uematsu and Takashi Yamagata. Estimation of sparsity-induced weak factor models. *Journal of Business & Economic Statistics*, pages 1–15, 2022.

Hui Zou, Trevor Hastie, and Robert Tibshirani. Sparse principal component analysis. *Journal of Computational and Graphical Statistics*, 15(2):265–286, 2006.