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

Simon Freyaldenhoven
Federal Reserve Bank of Philadelphia*

May 3, 2023

Online Appendix

Contents

*Email: simon.freyaldenhoven@phil.frb.org

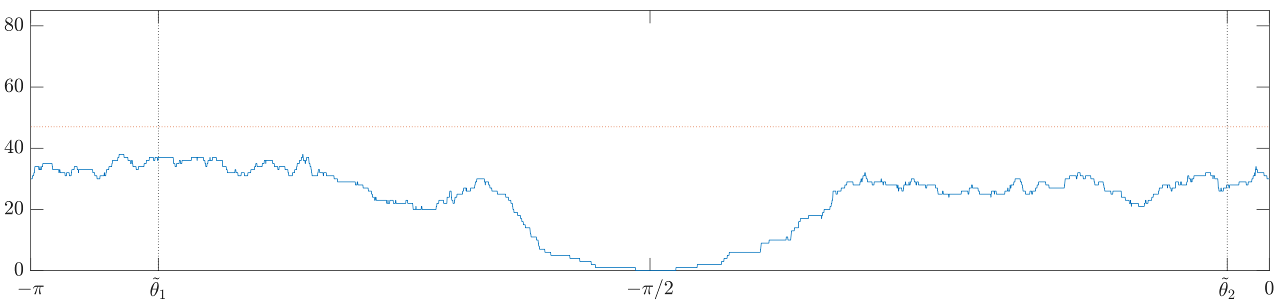
List of Tables

List of Figures

A A Stylized Example with Dense Loading Vectors

Suppose that in our baseline DGP from Section 3.1 we instead set $m_1 = m_2 = 207$, such that there are no zeroes in the loading matrix Λ^*. Then, 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 $i = 1, \ldots, n$. But this also implies that for all weights w_1 and w_2, $\lambda_{\bullet k} = w_1\lambda^0_{\bullet 1} + w_2\lambda^0_{\bullet 2}$ will generally be non-zero everywhere, such that there exists no linear combination of $\lambda^0_{\bullet 1}$ and $\lambda^0_{\bullet 2}$ that is sparse. Online Appendix Figure 1 illustrates. It depicts the number of loadings with an absolute value less than h_n.

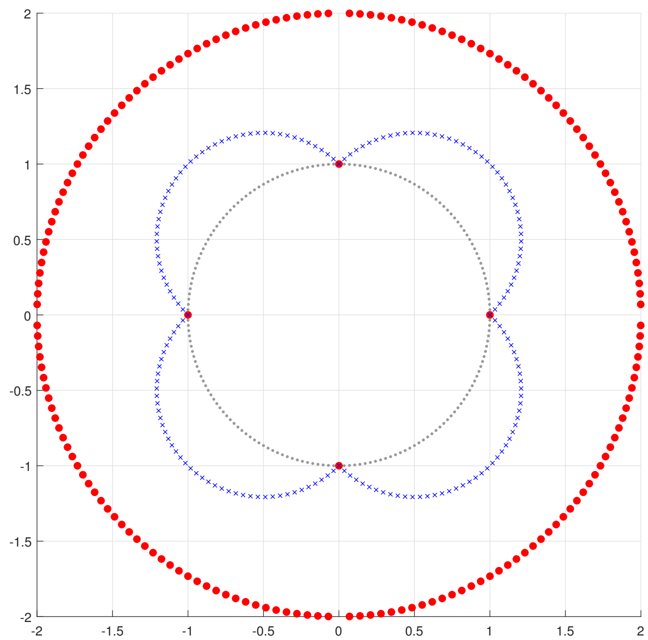
Online Appendix Figure 1 also illustrates how our rotation criterion can be used to test for the presence of local factors in a given dataset. In Section 4.1.2 we introduce a test that effectively consists of counting the number of small loadings in the most sparse estimated loading vector $\tilde{\lambda}_{ik}$ and comparing it to the number of small loadings that would be expected if the loading vector was "dense" with normally distributed loadings. In Online Appendix Figure 1, this corresponds to checking whether there exists a rotation above the red dashed line. Since the number of small loadings is less than critical value for all rotations, our testing procedure correctly concludes that there are no local factors in the data.



Online Appendix Figure 1: Depicted is the number of small loadings, $\sum_{i=1}^{n} \mathbf{1}_{|\lambda_{ik}|<h_n}$, where $\lambda_{\bullet k} = sin(\theta)\lambda^0_{\bullet l1} + cos(\theta)\lambda^0_{\bullet l2}$ as a function of the angle θ for a DGP with no local factors. Dashed red line represents critical value for testing whether there are local factors in the data.

B Geometric Illustration

To provide some further geometric intuition for our proposed rotation criterion, consider rotating the unit vector $i_1 = (1, 0)$ in the x,y-plane, analogous to rotating a loading vector for $n = 2$, where one can think of i_1 as a hypothetical $\lambda^*_{\bullet 1}$. Keeping its Euclidean length equal to one, this rotation produces the unit circle, depicted by the small gray dots in Online Appendix Figure 2, where the distance from the origin represents the ℓ_2-norm of each rotation (which is constant and equal to one).



Online Appendix Figure 2: Comparison of (pseudo-)norms across rotations of the unit vector in two dimensions. The size of each norm is illustrated by its distance from the origin. Depicted are ℓ_0-norm (large, red circles), ℓ_1-norm (blue crosses), and ℓ_2-norm (small, grey circles).

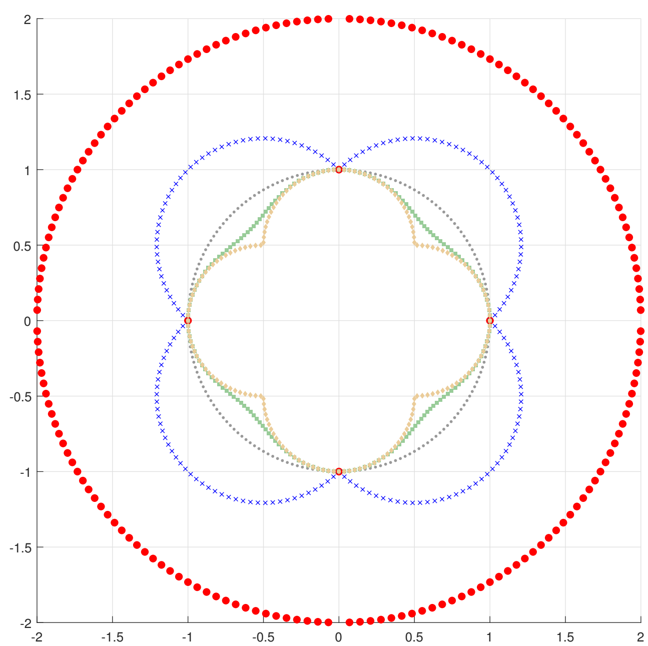
For each of these rotations, we next consider the corresponding ℓ_0-(pseudo)norm. As we rotate the unit vector, we scale up each rotation such that its Euclidean distance from the origin equals the value of its ℓ_0-(pseudo)norm. This corresponds to the outermost series, depicted by large red circles in Online Appendix Figure 2. We see that almost all rotations have two non-zero elements, such that $\|\cdot\|_0 = 2$, with the exceptions of the four vectors that align with either of the coordinate axes, where $\|\cdot\|_0 = 1$. Online Appendix Figure 2 also illustrates one reason why minimizing the ℓ_0-(pseudo)norm directly is infeasible in higher dimensions: the objective function is flat across almost all rotations with discontinuities at its minima.

We therefore propose to instead minimize the ℓ_1-norm across rotations. In Online Appendix Figure 2, the value of this norm across rotations is depicted by blue crosses. Specifically, for each rotation, the size of the ℓ_1-norm is represented by the distance of each cross from the origin. Two things are worth noting. First, we note the kink points that occur whenever one of the entries is equal to exactly zero, which is reminiscent of the ℓ_1-penalty term in high-dimensional linear regressions. We exploit the presence of these kink points in Section 4 to establish that the ℓ_1-norm is indeed minimized at sparse rotations. Second, the ℓ_1-norm is continuous and decreases toward a local minimum in its neighborhood, which makes it computationally appealing (we discuss the algorithmic implementation of our criterion in more detail in Online Appendix E).

B.1 Connection to Existing Rotation Criteria

As we noted in Section 3.2 of the paper, most existing rotation criteria use quartic functions of the loadings and are thus closely related to maximizing $\|\lambda_{\bullet k}\|_4^4 = \sum_{i=1}^n \lambda_{ik}^4$, subject to a constant ℓ_2-norm.

To compare our proposal and existing rotation criteria graphically, consider the trivial case of only two outcomes (and two factors) again. In Online Appendix Figure 3, we added the ℓ_4-norm $\|\cdot\|_4$, represented by the green squares, as well as the ℓ_∞-norm, represented by orange diamonds, to Online Appendix Figure 2.



Online Appendix Figure 3: Comparison of (pseudo-)norms across rotations of the unit vector in two dimensions. The size of each norm is illustrated by their distance from the origin. Depicted are ℓ_0-norm (large, red circles), ℓ_1-norm (blue crosses), ℓ_2-norm (small, grey circles), ℓ_4-norm (green squares), and ℓ_∞-norm (yellow diamonds).

Intuitively, maximizing the ℓ_∞-norm identifies the rotation with the largest entry. In contrast, minimizing the ℓ_0-norm essentially identifies the rotation with the smallest entries. While minimizing the ℓ_1-norm is a relaxation of the ℓ_0-norm, maximizing the ℓ_4-norm is a relaxation of the ℓ_∞-norm.

With just two outcomes, we note that maximizing the ℓ_4-norm across rotations produces four maxima, which coincide with the minima for the ℓ_0- and ℓ_1-norms. More generally, given the "resource-constraint" of a constant ℓ_2-norm, the solutions to the two optimization problems (maximizing ℓ_4, minimizing ℓ_1) will often look similar. However, 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.

C Further Discussion of Assumptions 3 and 5

To gain intuition for the value of $\beta^k(v_{\bullet k})$, let $N^+ = \sum_{i \in \mathcal{A}_k} \mathbf{1}\{\lambda^*_{ik} v_{ik} \geq 0\}$, $N^- = \sum_{i \in \mathcal{A}_k} \mathbf{1}\{\lambda^*_{ik} v_{ik} < 0\}$ and suppose that $N^- \leq N^+$ (the same result will hold for the opposite case). Then,

$$\begin{aligned}
\beta^k(v_{\bullet k}) &= \sum_{i \in \mathcal{A}_k} |v_{ik}| \, \mathbf{1}\{\lambda^*_{ik} v_{ik} \geq 0\} - \sum_{i \in \mathcal{A}_k} |v_{ik}| \, \mathbf{1}\{\lambda^*_{ik} v_{ik} < 0\} \\
&= N^+ \underbrace{\frac{1}{N^+} \sum_{i \in \mathcal{A}_k} |v_{ik}| \, \mathbf{1}\{\lambda^*_{ik} v_{ik} \geq 0\}}_{\equiv \overline{|v^+_{ik}|}} - N^- \underbrace{\frac{1}{N^-} \sum_{i \in \mathcal{A}_k} |v_{ik}| \, \mathbf{1}\{\lambda^*_{ik} v_{ik} < 0\}}_{\equiv \overline{|v^-_{ik}|}} \\
&= N^- \left(\overline{|v^+_{ik}|} - \overline{|v^-_{ik}|} \right) + (N^+ - N^-) \overline{|v^+_{ik}|}.
\end{aligned} \tag{OA.1}$$

Thus, the difference in Equation (OA.1) above and Equation (11) of the main text can be decomposed into the difference between two conditional means multiplied by N^-, and the difference between the number of terms used in each sum multiplied by a constant. While we treat Λ^* as fixed throughout, suppose Λ^* was random and assume $r = 2$. Further suppose λ_{ik} is distributed symmetrically and independently if $i \in \mathcal{A}_k$, and equal to zero otherwise. Then, both terms of the decomposition in (OA.1) would be $O_p(\sqrt{n})$ under some regularity conditions (e.g., $\lambda_{ik} \overset{i.i.d.}{\sim} N(0, \sigma)$, as in Example 3 on Page 16 of the main text). Thus, $\beta^1(v_{\bullet 1})$ will also be $O_p(\sqrt{n})$. Treating Λ as a fixed parameter, we avoid making any distributional assumptions on Λ, but instead simply define the above difference as $\beta^k(v_{\bullet k})$.

We next illustrate the behavior of $\beta^k = \max_{v_{\bullet k} \in V_k} \beta^k(v_{\bullet k})$ in finite sample for a hypothetical Λ^*. We create a $n \times 2$ loading matrix Λ^* with entries $\lambda^*_{ik} \overset{i.i.d.}{\sim} N(0, 1)$ if $i \in \mathcal{A}_k$, and $\lambda^*_{ik} = 0$ otherwise. Further, $|\mathcal{A}_2| = n$, such that $\mathcal{A}_1 \subset \mathcal{A}_2$.[1] Online Appendix Figure 4 then depicts how β^1 changes as we increase the size of the set $\mathcal{A}_1$.[2] Online Appendix Figure 4 confirms that β^1 grows proportionally to the square root of the size of the set $\mathcal{A}_1$.

As we have just argued theoretically and showed in simulation, in many cases $\beta^k \asymp \sqrt{n}$. To determine whether the condition in Assumption 3,

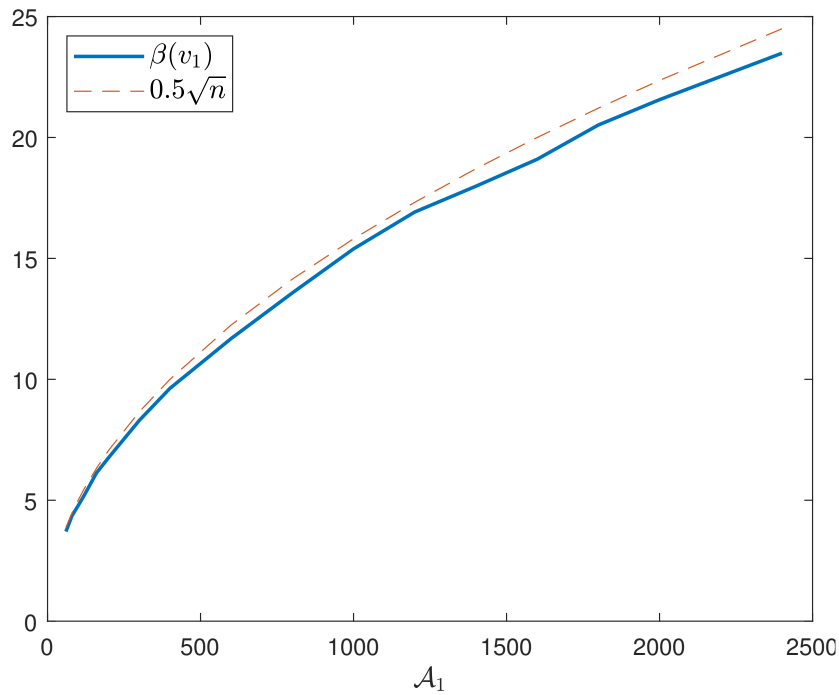
$$\|v^{\mathcal{A}^c_k}_{\bullet k}\|_1 > \beta^k(v_{\bullet k}), \tag{OA.2}$$

is plausible, we further need to consider $\|v^{\mathcal{A}^c_k}_{\bullet k}\|_1$ and, intuitively, under what conditions $\|v^{\mathcal{A}^c_k}_{\bullet k}\|_1 > \sqrt{n}$.

Recall that, for constants q_1 and q_2, $v_{\bullet 1} = q_1 \lambda^*_{\bullet 1} + q_2 \lambda^*_{\bullet 2}$, and thus $\|v^{\mathcal{A}^c_1}_{\bullet 1}\|_1 = q_2 \|\lambda^{*\mathcal{A}^c_1}_{\bullet 2}\|_1$, a constant

[1] Changing the covariance structure such that $Cov(\lambda_{i1}, \lambda_{i2}) \neq 0$ does not affect the results below.

[2] Specifically, we orthonormalize Λ^* using Gram-Schmidt, such that $\frac{U'U}{n} = I_2$, and write $U = (u_1, u_2) = (\lambda^*_1, v)$ (redefining λ^*_1 to have unit length is in line with the setup of the paper, see Section 2). Note that, with $r = 2$, the set V_1 contains only two vectors that are identical up to a sign indeterminacy.



Online Appendix Figure 4: Illustration of $\beta^1 = \max_{v_1 \in V_1} \beta^1(v_1)$ as a function of group size $|\mathcal{A}_1|$. $\lambda^*_{ik} \overset{i.i.d.}{\sim} N(0,1)$ if $i \in \mathcal{A}_k$, $\lambda^*_{ik} = 0$ otherwise, and $\mathcal{A}_1 \subset \mathcal{A}_2$. Depicted is the average value of β^1 across 10,000 simulations.

times the sum of the absolute values of $\lambda^*_{\bullet 2}$ on $\mathcal{A}_1^c$. It can be shown that $\|v_{\bullet 1}\|_2^2 = 1$ and $\lambda^*_{\bullet k} \perp v_{\bullet k}$ implies $q_2^2 = (1 - [\frac{\lambda^{*\prime}_{\bullet 1}\lambda^*_{\bullet 2}}{n}]^2)^{-1}$. Hence $q_2 \geq 1$, and $\|\lambda^{*\mathcal{A}_1^c}_{\bullet 2}\|_1 > \beta^k$ is sufficient for (OA.2) to hold. In general, the sum of the absolute values of λ^*_{2i} on $\mathcal{A}_1^c$ will be proportional to the number of outcomes affected by F_2, but not F_1. This suggests that $F_1 \in F^{exact}$ if there are proportionally more than $\sqrt{n}$ outcomes that are affected by F_2, but not F_1.

While, under the distribution of λ^*_{i1} considered above, we can infer the minimum value needed for $\|v^{\mathcal{A}_1^c}_{\bullet 1}\|_1$ to fulfill the condition in (OA.2) for a given group size from Online Appendix Figure 4 (e.g., at $|\mathcal{A}_1| = 500$, around 10), we next directly report the number of outcomes affected by F_2, but not F_1 that is needed for condition (OA.2) to hold across a number of different DGPs for Λ^*. In particular, let

$$\mu = (\mu_1, \mu_2) \qquad \text{and} \qquad \Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

We create a 500×2 loading matrix Λ^* with $\mathcal{A}_1 \subset \mathcal{A}_2$ and $|\mathcal{A}_1| = 300$. We draw $\lambda^*_{i\bullet} \overset{i.i.d.}{\sim} N(\mu, \Sigma)$ if $i \in \mathcal{A}_1$, $\lambda^*_{i2} \overset{i.i.d.}{\sim} N(\mu_2, 1)$ if $i \in \mathcal{A}_1^c \cap \mathcal{A}_2$, and $\lambda^*_{ik} = 0$ otherwise. To get a better sense of how demanding Assumption 3 is in practice, we then vary both μ and ρ. The results are depicted in Online Appendix Tables 1 and 2. Each entry depicts the minimum size of $\mathcal{A}_1^c \cap \mathcal{A}_2$ for $\|v^{\mathcal{A}_k^c}_{\bullet k}\|_1 > \beta^k(v_{\bullet k})$ to hold for a given combination of μ and ρ.

Means μ_k		Correlation ρ				
μ_1	μ_2	-0.4	-0.2	0	0.2	0.4
0	0	10	11	11	11	10
0	1	10	11	11	11	10
1	1	36	31	26	21	16
1	-1	16	21	26	31	36
2	2	62	53	44	35	27
2	-2	27	35	44	53	62

Online Appendix Table 1: Smallest number of entries in $\mathcal{A}_1^c \cap \mathcal{A}_2$ needed for condition (OA.2) to hold on average for $k = 1$. $|\mathcal{A}_1| = |\mathcal{A}_1 \cap \mathcal{A}_2| = 300$. Table varies the distribution of λ_{ik}^* through the parameters μ and ρ. In particular, $\lambda_{ik}^* \overset{i.i.d.}{\sim} N(\mu, \Sigma)$ if $i \in \mathcal{A}_1$, $\lambda_{i2}^* \overset{i.i.d.}{\sim} N(\mu_2, 1)$ if $i \in \mathcal{A}_1^c \cap \mathcal{A}_2$, and $\lambda_{ik}^* = 0$ otherwise. Table based on 1,000 simulations.

Online Appendix Table 1 depicts the minimum number of entries in $\mathcal{A}_1^c \cap \mathcal{A}_2$ needed for condition (OA.2) to hold for $k = 1$ on average across repeated realizations. Note that, with $\mu = (0, 0)$ and $\rho = 0$, this DGP is among those considered in Online Appendix Figure 4 (at $|\mathcal{A}_1| = 300$ in the figure). It states that, under this DGP, it is sufficient if 11 outcomes are affected by F_2, but not by F_1 for (OA.2) to hold on average. In other words: on average, $F_1 \in F^{exact}$, and thus identified, whenever more than 11 outcomes are affected by F_2, but not by F_1.

Online Appendix Table 2 depicts the number of entries needed for condition (OA.2) to hold in 95% of all realizations. For example, for $\mu = (1, 1)$ and $\rho = 0.4$, it states that, if $|\mathcal{A}_1^c \cap \mathcal{A}_2| \geq 32$, $F_1 \in F^{exact}$ in 95% of all simulations.

Means μ_k		Correlation ρ				
μ_1	μ_2	-0.4	-0.2	0	0.2	0.4
0	0	24	27	28	27	25
0	1	25	27	25	26	25
1	1	59	51	45	38	32
1	-1	31	38	46	52	58
2	2	76	66	55	46	35
2	-2	35	46	56	66	75

Online Appendix Table 2: Smallest number of entries in $\mathcal{A}_1^c \cap \mathcal{A}_2$ needed for condition (OA.2) to hold in 95% of all realizations for $k = 1$. $|\mathcal{A}_1| = |\mathcal{A}_1 \cap \mathcal{A}_2| = 300$. Table varies the distribution of λ_{ik}^* through the parameters μ and ρ. In particular, $\lambda_{ik}^* \overset{i.i.d.}{\sim} N(\mu, \Sigma)$ if $i \in \mathcal{A}_1$, $\lambda_{i2}^* \overset{i.i.d.}{\sim} N(\mu_2, 1)$ if $i \in \mathcal{A}_1^c \cap \mathcal{A}_2$, and $\lambda_{ik}^* = 0$ otherwise. Table based on 1,000 simulations.

Consistent with our discussion in the main text, we find that the required number of entries in $\mathcal{A}_1^c \cap \mathcal{A}_2$ is larger for combinations of μ and ρ for which $\lambda_{\bullet 1}^*$ and $\lambda_{\bullet 2}^*$ are further from orthogonality.

Two things are worth noting. First, we still allow for significant overlap between $\mathcal{A}_1$ and $\mathcal{A}_2$ in all specifications. Second, in all depicted simulations $\mathcal{A}_1 \subset \mathcal{A}_2$. This means that all numbers are conservative. If $\mathcal{A}_1 \cap \mathcal{A}_2 \neq \mathcal{A}_1$, $\beta^1(v_{\bullet 1})$ will tend to be smaller, and the number of entries in $\mathcal{A}_1^c \cap \mathcal{A}_2$ required for identification of $\lambda^*_{\bullet k}$ will also be smaller than those depicted in Online Appendix Tables 1 and 2.

In conclusion, in the paper we treat the loadings as fixed and impose condition (OA.2) as a high-level assumption. This section outlined how plausible this assumption is under random loadings.

D Alternative Criteria

In this section we illustrate the behavior of a number of alternative rotation criteria. We start with our illustrative two-factor DGP from Section 3.1. For reference, we first repeat an illustration of the ℓ_1-norm of a loading vector across all rotations of the Principal Component estimate Λ^0 in Online Appendix Figure 5a. Online Appendix Figure 5a is identical to Figure 3 in the main text.

Recall from Equation 6 in the main text that the objective function for the Varimax criterion can be expressed as

$$\max_{R:R'R=I} Q(\Lambda^0 R) = Q(\Lambda) = \sum_{k=1}^{r} \left[\sum_{i=1}^{n} \lambda_{ik}^4 - \frac{1}{n} \left(\sum_{i=1}^{n} \lambda_{ik}^2 \right)^2 \right] \tag{OA.3}$$

and denote the argmin to (OA.3) by $\ddot{R}$, and the corresponding loading matrix by $\ddot{\Lambda} = \Lambda^0 \ddot{R}$. Noting that $Q(\Lambda)$ is additively separable in $\lambda_{\cdot k}$, $k = 1, \ldots, r$ (and ignoring the orthogonality constraint on R for now), we obtain

$$\sum_{i=1}^{n} \lambda_{ik}^4 - \frac{1}{n} \left(\sum_{i=1}^{n} \lambda_{ik}^2 \right)^2 \tag{OA.4}$$

as the contribution to the objective function $Q(\Lambda)$ by an individual column $\lambda_{\bullet k}$. The restriction $R'_{\bullet k} R_{\bullet k} = 1$, combined with the choice of an initial estimate Λ^0 that is orthonormal, implies that the second part of (OA.4) is constant. Maximizing (OA.3) is therefore equivalent to maximizing the columnwise ℓ_4-norm, with the added restriction that the resulting loading vectors are orthonormal.

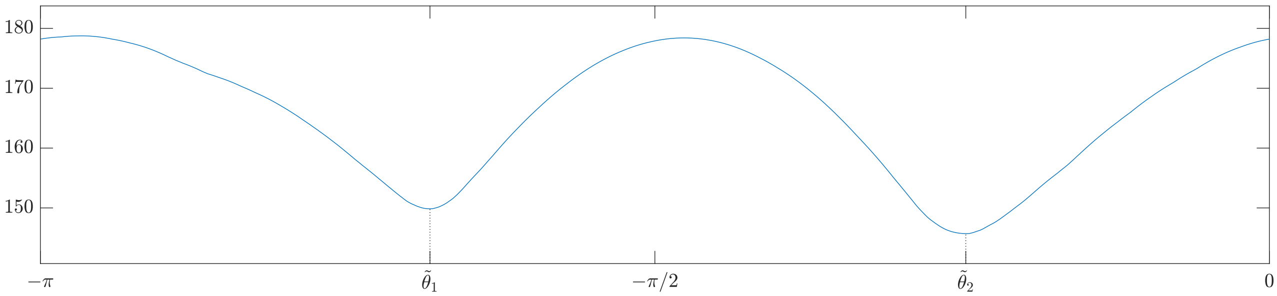
In the two-dimensional case ($r = 2$), the unit length restriction on $\lambda_{\bullet k}$ means we can write $\lambda_{\bullet k} = sin(\theta_k)\lambda_{\bullet 1}^0 + cos(\theta_k)\lambda_{\bullet 2}^0$ for some θ_k, making it easy to visualize rotations. We depict the ℓ_4-norm across rotations θ_k for a single vector $\lambda_{\bullet k}$ in Online Appendix Figure 5b below.

Further note that at any solution $\ddot{\Lambda}$ to (OA.3), $\frac{\ddot{\Lambda}'\ddot{\Lambda}}{n} = I$ implies that

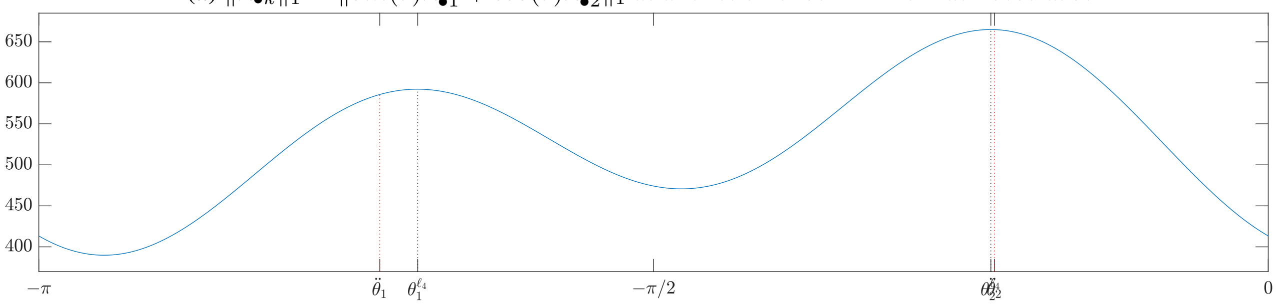
$$\begin{aligned}
& & \ddot{\lambda}'_{\bullet 1}\ddot{\lambda}_{\bullet 2} &= 0 \\
\Leftrightarrow & & \left[sin(\theta_1)\lambda_{\bullet 1}^0 + cos(\theta_1)\lambda_{\bullet 2}^0\right]' \left[sin(\theta_2)\lambda_{\bullet 1}^0 + cos(\theta_2)\lambda_{\bullet 2}^0\right] &= 0 \\
\Leftrightarrow & & sin(\theta_1)sin(\theta_2) + cos(\theta_1)cos(\theta_2) &= 0 \\
\Leftrightarrow & & cos(\theta_1 - \theta_2) &= 0 \\
\Leftrightarrow & & \theta_1 - \theta_2 &= \frac{\pi}{2} + g\pi, g \in \mathbb{Z}.
\end{aligned} \tag{OA.5}$$

Thus the Varimax criterion will maximize $\|\lambda_{\bullet 1}\|_4^4 + \|\lambda_{\bullet 2}\|_4^4$, where $\|\lambda_{\bullet k}\|_4^4 = \|sin(\theta_k)\lambda_{\bullet 1}^0 + cos(\theta_k)\lambda_{\bullet 2}^0\|_4^4$, subject to the constraint that $(\theta_1 - \theta_2)$ fulfills the condition stated in (OA.5) above. We see in Online Appendix Figure 5b that, without the orthogonality restriction, the nonsingular rotations with the largest ℓ_4-norm correspond to $\theta_1^{\ell_4}$ and $\theta_2^{\ell_4}$, marked by the dashed black lines. The restriction in (OA.5) forces the difference between any solutions $\ddot{\theta}_1$ and $\ddot{\theta}_2$ (at the red dashed lines) to be slightly larger than that found between $\theta_1^{\ell_4}$ and $\theta_2^{\ell_4}$.

We further note that $\theta_1^{\ell_4}$ and $\tilde{\theta}_1$ are close, but not identical. We expect this to frequently be the case and give an intuitive explanation for this in Online Appendix B.



(a) $\|\lambda_{\bullet k}\|_1 = \|sin(\theta)\lambda^0_{\bullet 1} + cos(\theta)\lambda^0_{\bullet 2}\|_1$ as a function of θ. Minimum achieved at $\tilde{\theta}$.

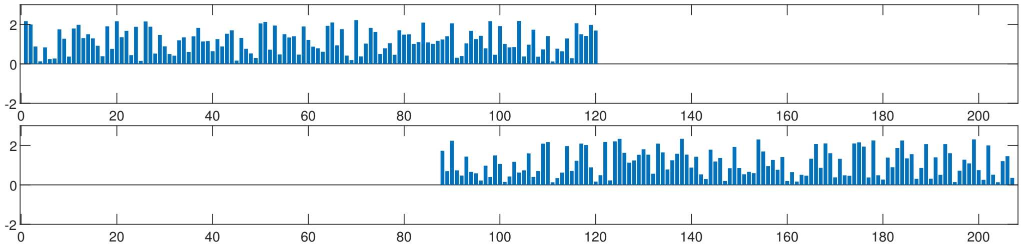


(b) $\|\lambda_{\bullet k}\|_4^4 = \|sin(\theta)\lambda^0_{\bullet 1} + cos(\theta)\lambda^0_{\bullet 2}\|_4^4$ as a function of θ. Maximum achieved at θ^{ℓ_4}. Maximum under the constraint in (OA.5) achieved at $\ddot{\theta}$.

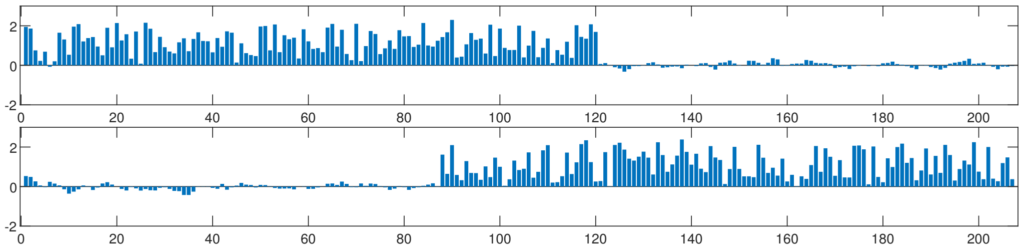
Online Appendix Figure 5: Comparison of objective functions across different criteria. Figure depicts the value of the respective objective function (ℓ_1- and ℓ_4-norm) across all rotations in the space spanned by the initial estimate Λ^0.

Finally, we also consider Promax (Hendrickson and White 1964). Promax is one of the most commonly used oblique rotations in the literature, and a native implementation of it is included in many statistical software including MATLAB. The Promax rotation consists of two steps. The first step computes the Varimax rotation and raises all its entries to the fourth power to define a target matrix. In the second step, the Promax estimate is then obtained by computing a least-square fit from the Varimax solution to the previously defined target matrix. Due to the nature of this criterion, there is no obvious equivalent to Online Appendix Figure 5 for the Promax criterion.

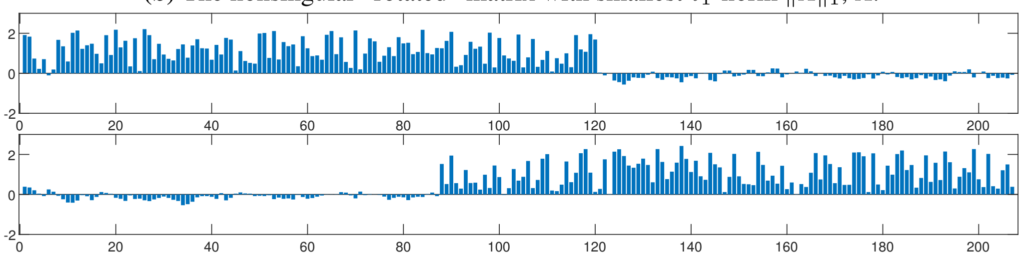
However, in order to visually assess the performance of the different rotation criteria (ℓ_1, ℓ_4, Varimax, Promax), we next depict the estimated loading matrix for all four criteria based on a single realization in Online Appendix Figure 6. For reference, Panel 6a repeats the true loading matrix Λ^*. Panel 6b depicts $\tilde{\Lambda}$, the estimate that minimizes the ℓ_1-norm of the loadings across rotations of Λ^0, which corresponds to the linear combinations at $\tilde{\theta} = [\tilde{\theta}_1, \tilde{\theta}_2]$ in Online Appendix Figure 5a.



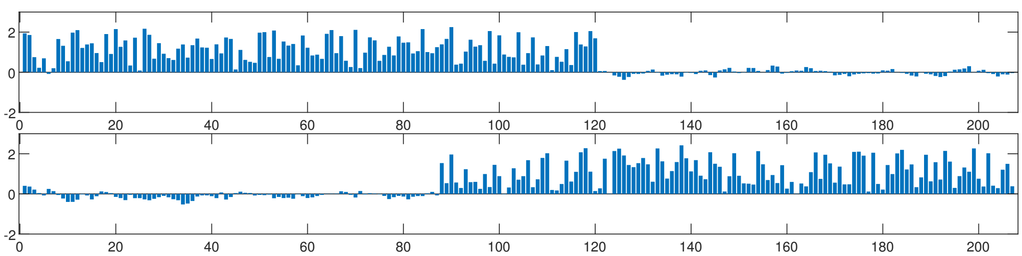
(a) True loading matrix Λ^*.



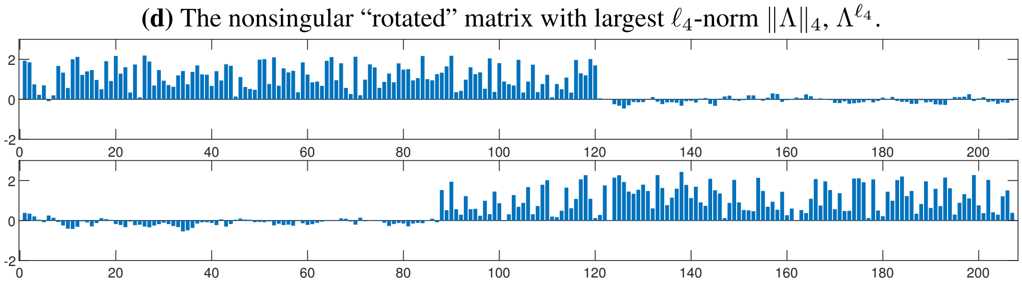
(b) The nonsingular "rotated" matrix with smallest ℓ_1-norm $\|\Lambda\|_1$, $\tilde{\Lambda}$.



(c) The "rotated" matrix $\ddot{\Lambda}$ that maximizes the Varimax criterion.



(d) The nonsingular "rotated" matrix with largest ℓ_4-norm $\|\Lambda\|_4$, Λ^{ℓ_4}.



(e) The "rotated" matrix Λ^{PROMAX} that is the result of the Promax rotation.

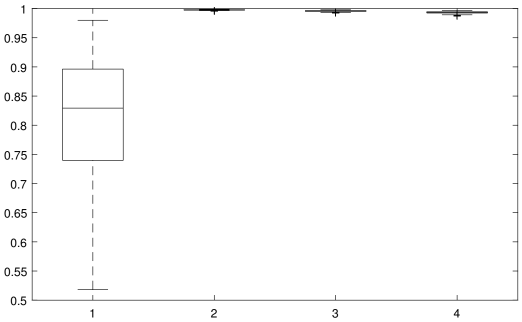
Online Appendix Figure 6: Comparison of loading matrices across different rotation criteria. Each panel depicts the loadings associated with all 207 outcomes, where the top diagram depicts $\lambda'_{\bullet 1}$ and bottom panel $\lambda'_{\bullet 2}$. Panels 6b-6e differ in the criterion that determines θ_k in $\lambda_{\bullet k} = sin(\theta_k)\lambda^0_{\bullet 1} + cos(\theta_k)\lambda^0_{\bullet 2}$.

Panel 6c depicts $\ddot{\Lambda}$, the estimate that maximizes the Varimax criterion across rotations of Λ^0, which corresponds to the linear combinations at $\ddot{\theta} = [\ddot{\theta}_1, \ddot{\theta}_2]$ in Online Appendix Figure 5b. Panel 6d depicts Λ^{ℓ_4}, the estimate that maximizes the ℓ_4-norm of the loadings across rotations of Λ^0, which corresponds to the linear combinations at $\theta^{\ell_4} = [\theta_1^{\ell_4}, \theta_2^{\ell_4}]$ in Online Appendix Figure 5b. Panel 6e depicts Λ^{PROMAX}, the estimate that maximizes the Promax criterion across rotations of Λ^0.

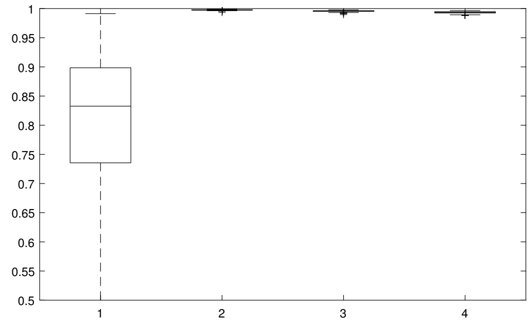
In the single realization depicted here, the differences between the four estimates in Online Appendix Figure 6 appear small and all four appear to be good estimates of Λ^*. Since the DGP we considered so far is perhaps unrealistically simple, and Online Appendix Figure 6 is based on a single realization, we next turn to repeated simulations, and repeat the exercise from Section 5.1 by using the same DGP underlying Figure 9 in the main text. Recall that under this DGP, there are four factors. Of these four factors, the first affects all outcomes, while the remaining three are local. Online Appendix Figure 7 uses a boxplot to visualize the performance of the different rotation criteria. It depicts the maximum cosine similarity as defined in Section 5 for each factor across 500 realizations of the DGP. In line with the main text, we consider two versions of this DGP: one in which $\lambda_{ik} = 0$ for all $i \in \mathcal{A}_k^c$ (on the left), and one in which $\lambda_{ik} \overset{i.i.d.}{\sim} N(0, \sigma^2)$, $\sigma^2 = \frac{1}{n}$ for all $i \in \mathcal{A}_k^c$ (on the right).

Panels 7a and 7b are identical to Figure 9b-9d in the main text, and demonstrate that the three local factors can be successfully recovered by our proposed criterion. The other three rotation methods perform significantly worse. While the estimators based on both Varimax and Promax still consistently achieve a similarity of above 0.9 for the local factors, the similarity with $\lambda_{\cdot 2}^*$ in particular (the local factor affecting the most outcomes) is significantly lower than that of $\tilde{\Lambda}$. Maximizing the ℓ_4-norm directly (Panels 7e-7f) performs even worse, in particular for $\lambda_{\cdot 2}^*$.

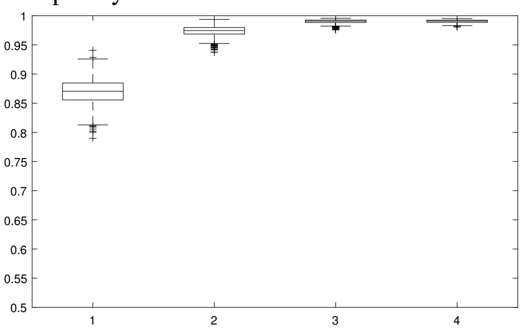
Thus, we conclude that our proposed criterion based on the ℓ_1-norm outperforms all of the considered quartic criteria in this simulation exercise.



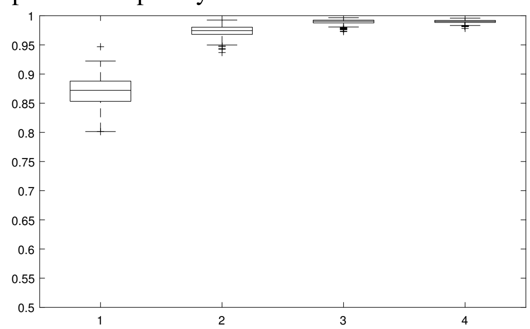
(a) $MC_k(\tilde{\Lambda})$ of rotated estimator (ℓ_1-norm) under exact sparsity



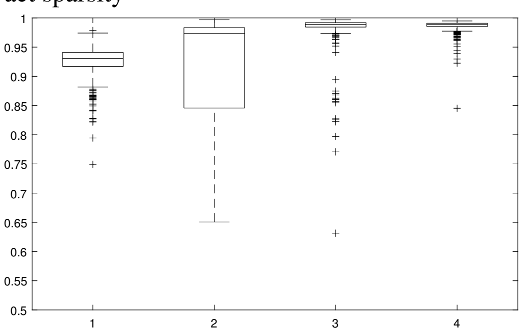
(b) $MC_k(\tilde{\Lambda})$ of rotated estimator (ℓ_1-norm) under approximate sparsity



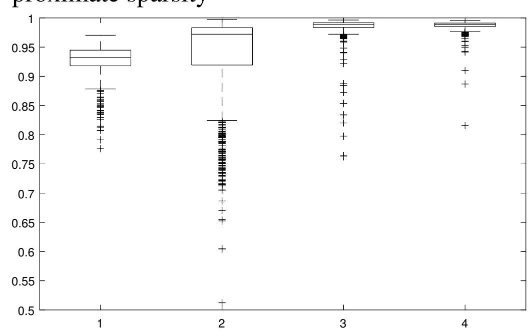
(c) $MC_k(\ddot{\Lambda})$ of rotated estimator (Varimax) under exact sparsity



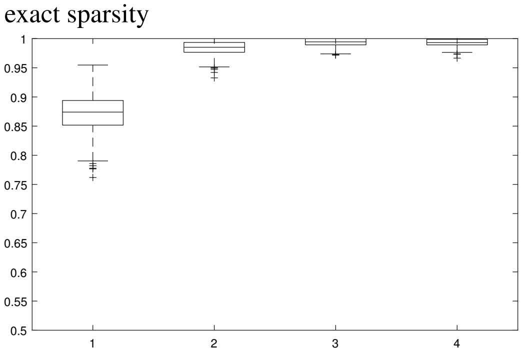
(d) $MC_k(\ddot{\Lambda})$ of rotated estimator (Varimax) under approximate sparsity



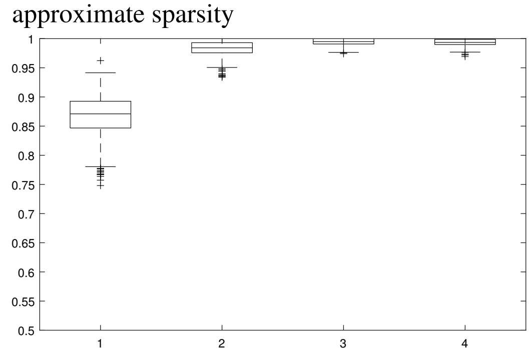
(e) $MC_k(\Lambda^{\ell_4})$ of rotated estimator (ℓ_4-norm) under exact sparsity



(f) $MC_k(\Lambda^{\ell_4})$ of rotated estimator (ℓ_4-norm) under approximate sparsity



(g) $MC_k(\Lambda^{PROMAX})$ of rotated estimator (Promax) under exact sparsity



(h) $MC_k(\Lambda^{PROMAX})$ of rotated estimator (Promax) under approximate sparsity

Online Appendix Figure 7: Maximum cosine similarity for all four factor loadings λ_k^*, k=$1, \ldots, 4$. The first factor is global, while factors 2-4 are local. Boxplots based on 100 realizations.

E Algorithmic Implementation

Recall the minimization problem we consider throughout to identify Λ^* (or more precisely, an individual column $\lambda^*_{\bullet k}$):

$$\min_{R_{\bullet k}} \|\sum_{l=1}^{r} \lambda^0_{\bullet l} R_{lk}\|_1 \qquad \text{such that } R'_{\bullet k} R_{\bullet k} = 1. \tag{OA.6}$$

We next discuss how to implement our estimator in practice. Our implementation consists of the following four steps:

1. We first compute the Principal Component estimator Λ^0 as an initial estimate for Λ^* that fulfills $\frac{\Lambda^{0\prime}\Lambda^0}{n} = I$.

2. Next, we draw a random grid of G starting points $R^{0g}_{\bullet k}, g = 1, \ldots, G$, where $R^{0j}_{lk} = \frac{x_l}{\|x\|}, x \overset{i.i.d.}{\sim} N(0, I_r)$.[3] For each starting point $R^{0g}_{\bullet k}$, we then find the argmin of (OA.6), denoted by $R^{1g}_{\bullet k}$ as follows.

   First, convert $R_{\bullet k}$ to spherical coordinates:

$$\begin{aligned} R_{1k} &= \tilde{r}\cos(\theta_{1k}) \\ R_{2k} &= \tilde{r}\sin(\theta_{1k})\cos(\theta_{2k}) \\ &\vdots \\ R_{r-1,k} &= \tilde{r}\sin(\theta_{1k})\ldots\sin(\theta_{r-2,k})\cos(\theta_{r-1,k}) \\ R_{rk} &= \tilde{r}\sin(\theta_{1k})\ldots\sin(\theta_{r-2,k})\sin(\theta_{r-1,k}), \end{aligned} \tag{OA.7}$$

   denoting by $\theta_k = [\theta_{1k}, \theta_{2k}, \ldots, \theta_{r-1,k}]$ and $\tilde{r}_k$ angles and radius respectively. The constraint in (OA.6) is equivalent to setting $\tilde{r}_k = 1$, such that the constraint minimization problem in OA.6 simply becomes

$$\min_{\theta_k} \left\| \lambda^0_{\bullet 1}\cos(\theta_{1k}) + \lambda^0_{\bullet r}\Pi_{p=1}^{r-1}\sin(\theta_{pk}) + \sum_{l=2}^{r-1} \lambda^0_{\bullet l}\cos(\theta_l)\Pi_{p=2}^{l-1}\sin(\theta_{p-1,k}) \right\|_1, \tag{OA.8}$$

   an unconstrained optimization over the angles θ_k. For a solution θ^g_k, corresponding to starting point $R^{0g}_{\bullet k}$, we then use (OA.7) to obtain the corresponding cartesian coordinates $R^{1g}_{\bullet k}$.

   At the end of this step, we have G candidate solutions $R^{1g}_{\bullet k}$.

[3]The number of random starting points G increases with the number of factors r. In particular, we set $G = \{300, 500, 1000, 2000\}$ for $r = \{2, 3, 4, 5\}$ respectively, $G = 3000$ for $r \in \{6, 7, 8\}$, and $G = 5000$ for $r \geq 9$. We implement our algorithm using $fminsearch$, a native optimization routine included in MATLAB.

3. Many of the G candidate solutions will be (close to) identical (because many starting points will converge to the same local minimum). In this step, we consolidate identical solutions into a single candidate.

   To do so, we sort all G candidate solutions according to the ℓ_1-norm of the corresponding candidate estimate $\tilde{\lambda}^g_{\bullet k} = \Lambda^0 R^{1g}_{\bullet k}$, such that, with slight abuse of notation, $\|\tilde{\lambda}^1_{\bullet k}\|_1 \leq \|\tilde{\lambda}^2_{\bullet k}\|_1 \leq \ldots \leq \|\tilde{\lambda}^G_{\bullet k}\|_1$. We then drop all candidates $R^{1g}_{\bullet k}$ if there exists a $R^{1g'}_{\bullet k}$, $g' < g$ such that $\|R^{1g}_{\bullet k} - R^{1g'}_{\bullet k}\|_2 / r < 0.05$.

   At the end of this step, we have P candidate solutions $R^p_{\bullet k}$, sorted in ascending order according to the ℓ_1-norm of the corresponding candidate estimate $\tilde{\lambda}^p_{\bullet k} = \Lambda^0 R^p_{\bullet k}$.

4. Each candidate $R^p_{\bullet k}$ corresponds to a local minimum of (OA.6). In this last step, we combine these solutions into the rotation matrix $\tilde{R}$.

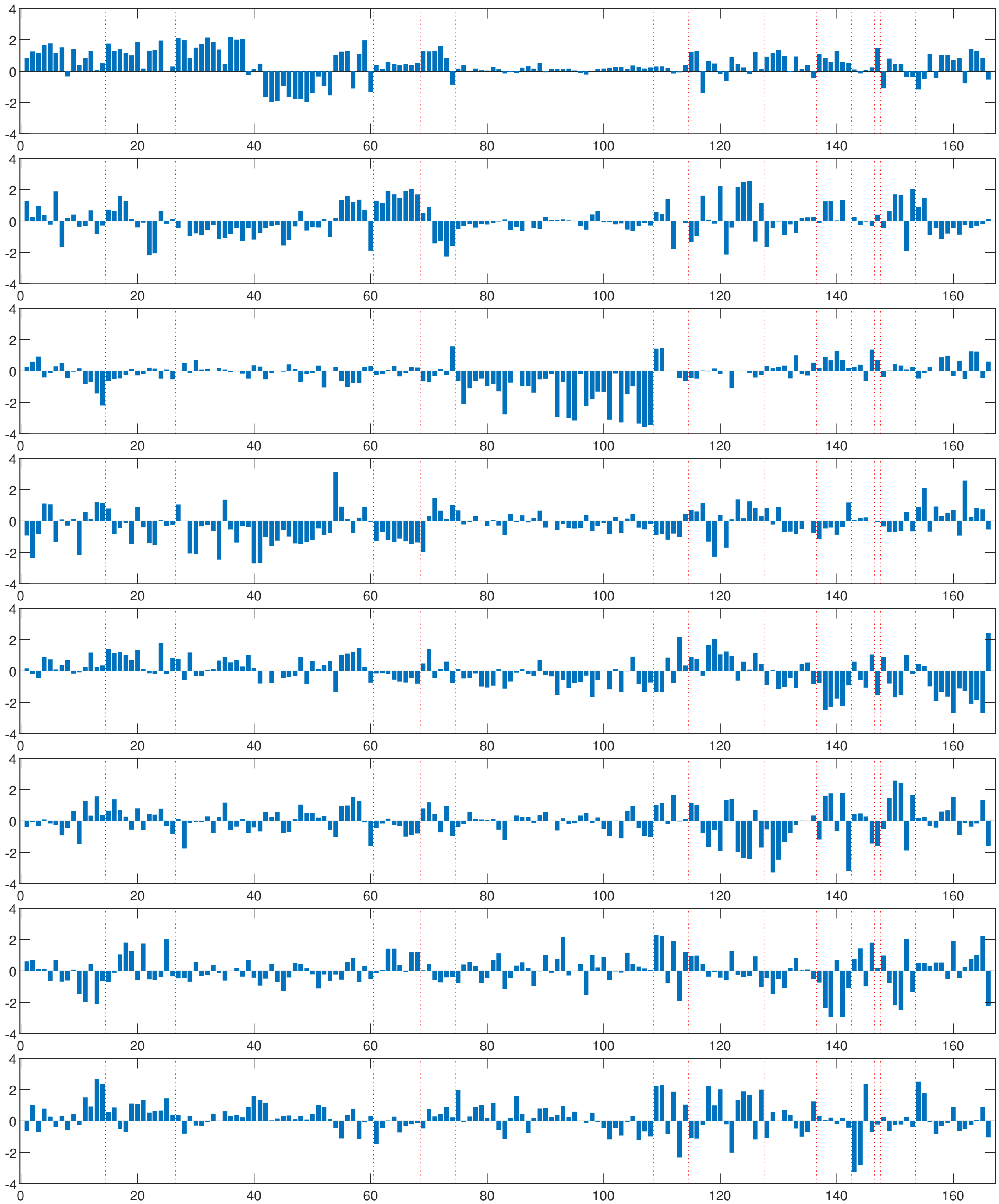
   We initialize $\ddot{R} = R^1_{\bullet k}$ and iteratively append $R^p_{\bullet k}, p = 2, \ldots, P$ whenever the resulting matrix does not become (close to) singular. Denote the resulting $r \times \ddot{r}$ matrix by $\ddot{R} = [\ddot{R}_{\bullet 1}, \ldots, \ddot{R}_{\bullet \ddot{r}}]$.

   Finally, we distinguish two cases:

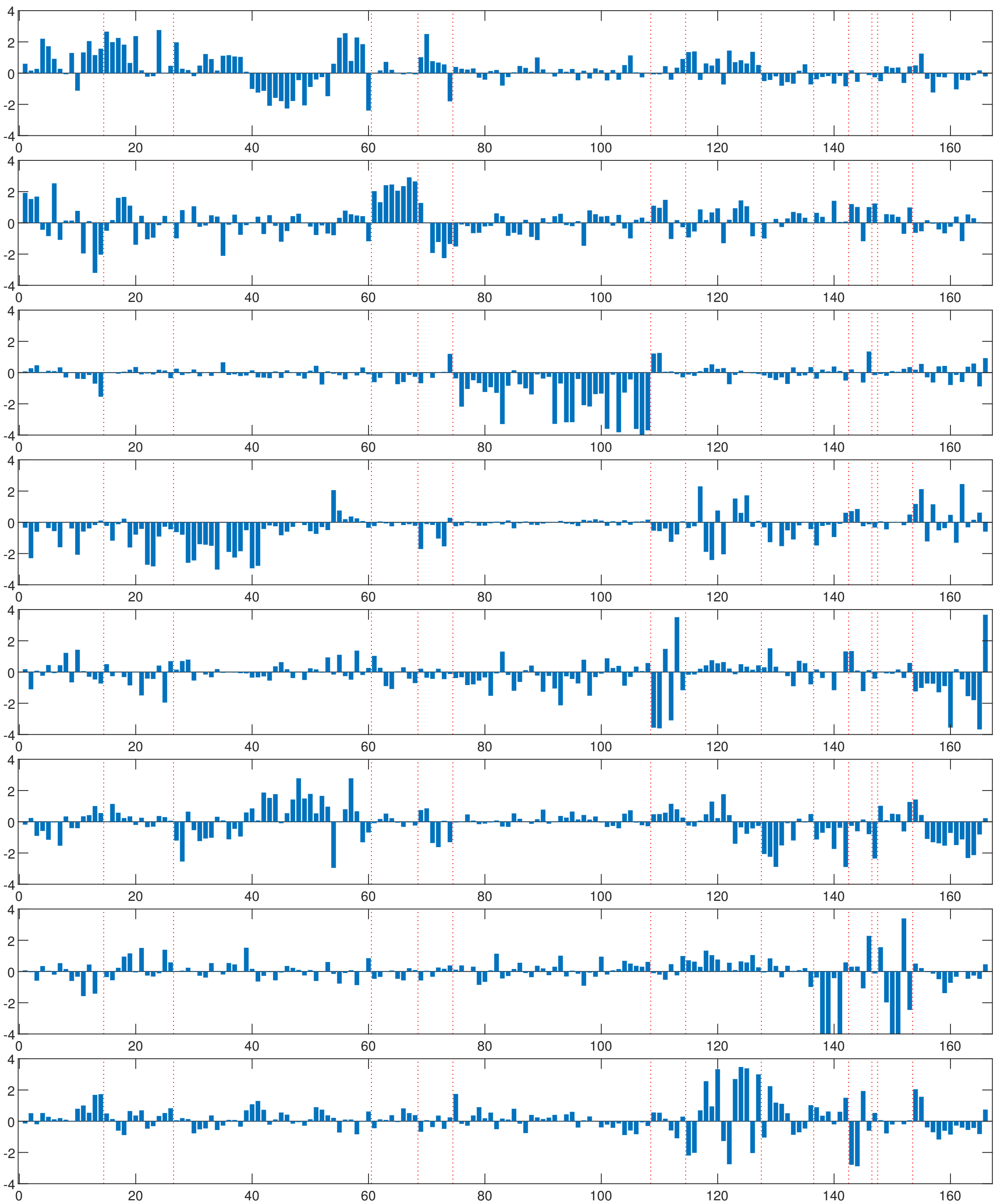
   - $\ddot{r} = r$: The number of candidate solutions is equal to the number of factors. Then, we simply set $\tilde{R} = \ddot{R}$. The result, $\tilde{\Lambda} = \Lambda^0 \tilde{R}$, is our proposed estimate for the loading matrix Λ^*.
   - $\ddot{r} < r$: There are fewer candidate solutions than the number of factors. In this case, we iteratively append vectors e^d to $\ddot{R}$, where e^d denotes an $r \times 1$ vector with dth entry $e^d_d = 1$, and zeros everywhere else. Note that this is equivalent to adding $(r - \ddot{r})$ columns of Λ^0 to $\ddot{\Lambda} = \Lambda^0 \ddot{R}$ directly.[4] The result, $\tilde{\Lambda}$, is our proposed estimate for the loading matrix Λ^*.

[4] We choose the entry d to pick out the loading vector $\lambda^0_{\bullet d}$ that maximizes the minimum singular value of the combined matrix $[\ddot{\Lambda}, \lambda^0_{\bullet d}]$. Intuitively, this is the column in Λ^0 that is furthest away from any linear combination of the columns in $\ddot{\Lambda}$.

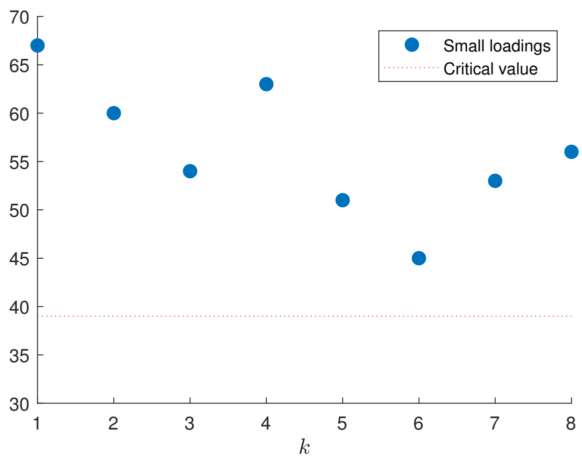
F Additional Figures for Macroeconomic Application



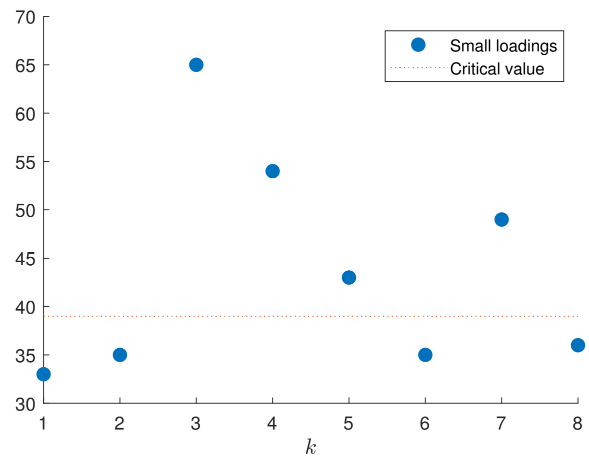
Online Appendix Figure 8: Illustration of the loading vectors $\lambda^0_{\bullet k}$, $k = 1, \ldots, 8$, in Principal Component estimate Λ^0 for panel of macroeconomic indicators. Bars correspond to the loadings associated with the individual macroeconomic indicators from Section 6.2.



Online Appendix Figure 9: Illustration of the loading vectors $\ddot{\lambda}_{\bullet k}$, $k = 1, \ldots, 8$, in Varimax estimate $\ddot{\Lambda}$ for 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).



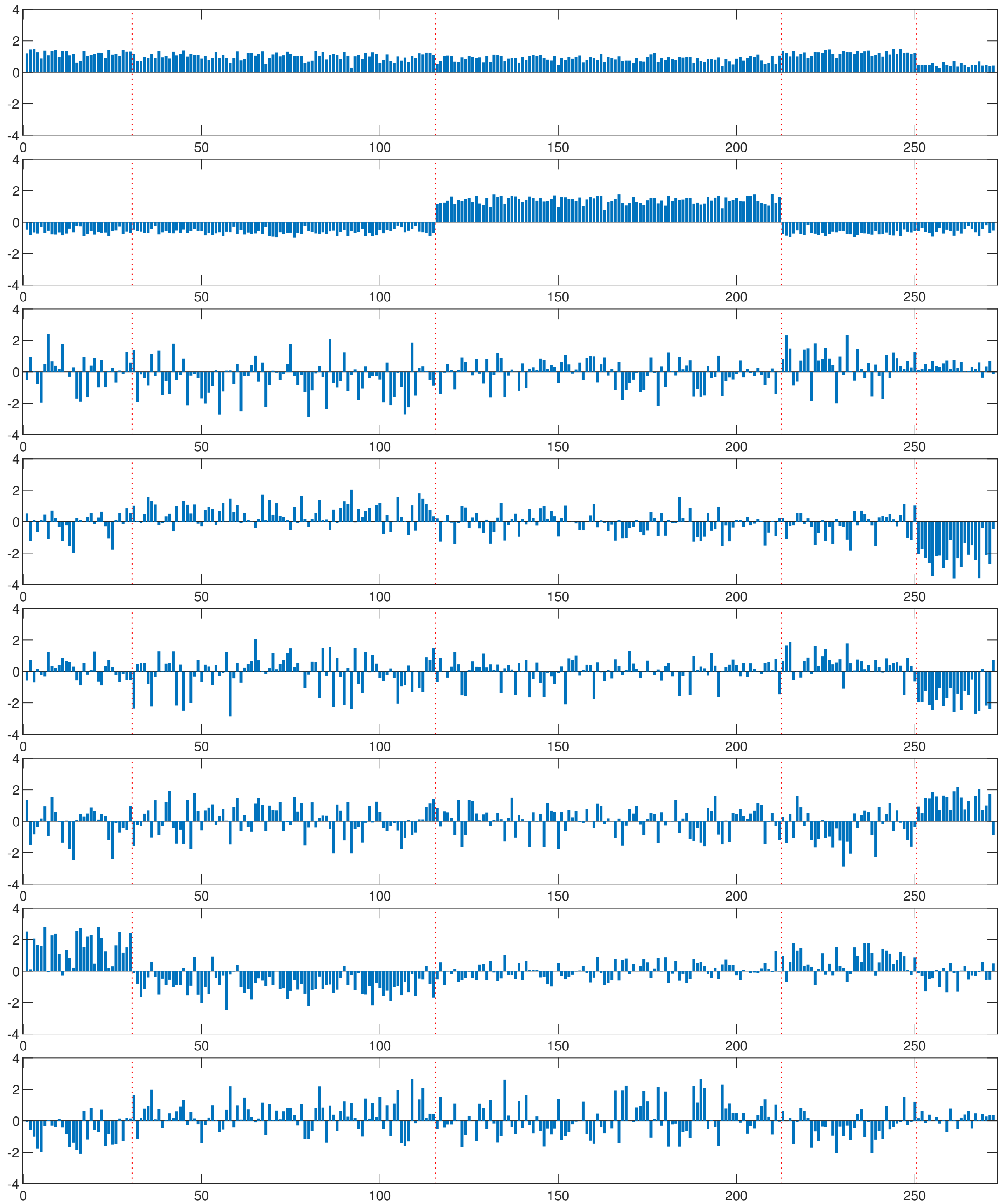
(a) Number of small elements in the rotation of Λ^0 that minimizes the ℓ_1-norm, $\tilde{\Lambda}$.



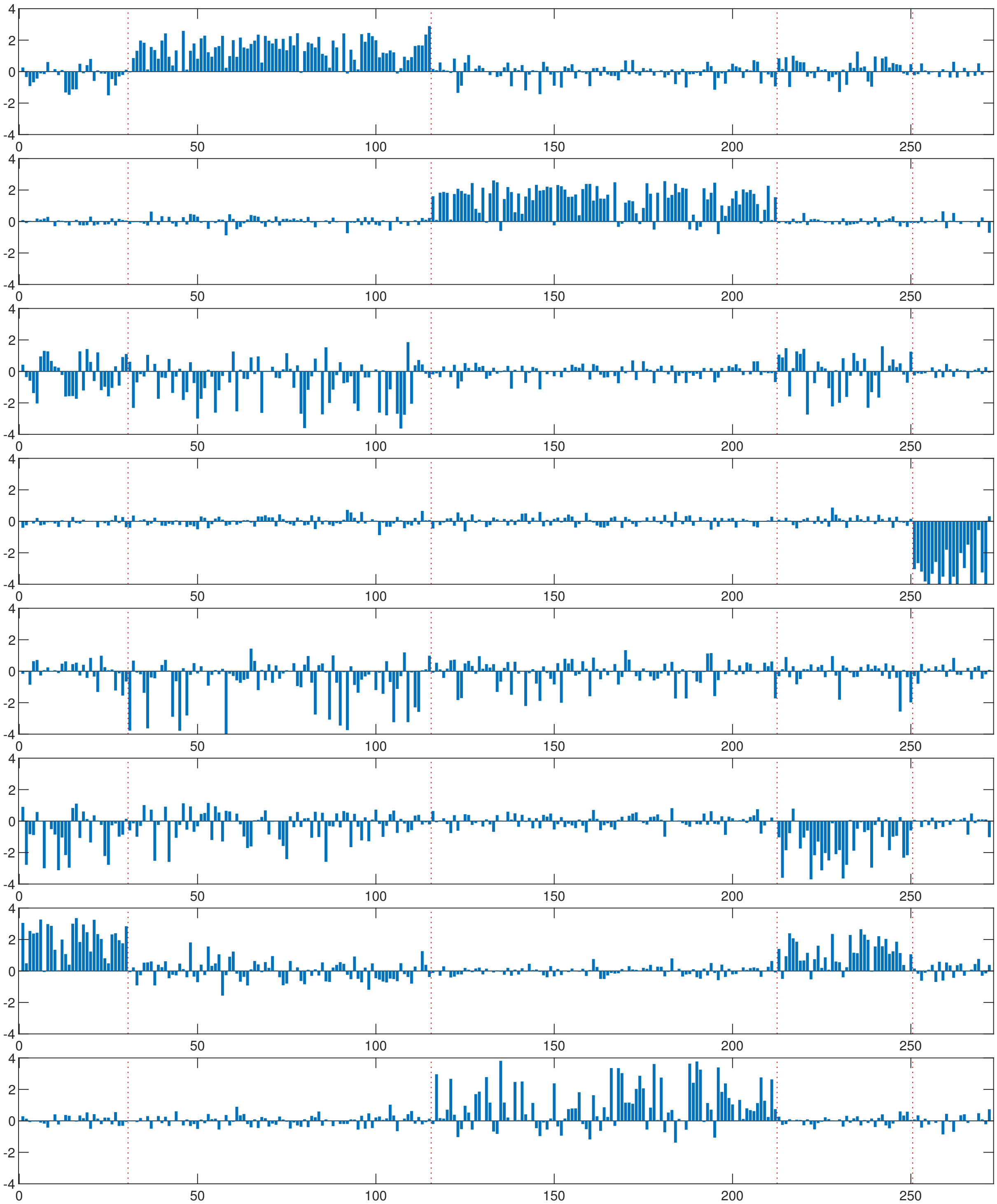
(b) Number of small elements in the rotation of Λ^0 that maximizes the Varimax criterion, $\ddot{\Lambda}$.

Online Appendix Figure 10: For each $k = 1, \ldots, 8$, the points above represent number of small elements in $\lambda_{\bullet k}$, the kth column in Λ, for a panel of US macroeconomic indicators. Dotted red line indicates critical value for $\hat{\mathcal{L}}_0$.

G Additional Results for Financial Application



Online Appendix Figure 11: Illustration of the loading vectors $\lambda^0_{\bullet k}$, $k = 1, \ldots, 8$, in Principal Component estimate Λ^0 for panel of international asset returns. Bars correspond to the loadings of the 272 individual stocks. Geographical groups are Germany, UK, US, France, and Middle East, separated by dashed lines.



Online Appendix Figure 12: Illustration of the loading vectors $\ddot{\lambda}_{\bullet k}$, $k = 1, \ldots, 8$, in Varimax estimate $\ddot{\Lambda}$ for 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.

G.1 The Data

Traded in	Ticker	Company	Prime Standard industry group
Frankfurt	ADS	Adidas	Clothing
Frankfurt	ALV	Allianz	Insurance
Frankfurt	BAS	BASF	Chemicals
Frankfurt	BAYN	Bayer	Pharmaceuticals and Chemicals
Frankfurt	BEI	Beiersdorf	Consumer goods
Frankfurt	BMW	BMW	Manufacturing
Frankfurt	CBK	Commerzbank	Banking
Frankfurt	CON	Continental	Manufacturing
Frankfurt	DAI	Daimler	Manufacturing
Frankfurt	DBK	Deutsche Bank	Banking
Frankfurt	DB1	Deutsche Börse	Securities
Frankfurt	LHA	Deutsche Lufthansa	Transport Aviation
Frankfurt	DPW	Deutsche Post	Communications
Frankfurt	DTE	Deutsche Telekom	Communications
Frankfurt	EOAN	E.ON	Energy
Frankfurt	FRE	Fresenius	Medical
Frankfurt	FME	Fresenius Medical Care	Medical
Frankfurt	HEI	HeidelbergCement	Building
Frankfurt	HEN3	Henkel	Consumer goods
Frankfurt	IFX	Infineon Technologies	Manufacturing
Frankfurt	SDF	K+S	Chemicals
Frankfurt	LXS	Lanxess	Chemicals
Frankfurt	LIN	Linde	Industrial gases
Frankfurt	MRK	Merck	Pharmaceuticals
Frankfurt	MUV2	Munich Re	Insurance
Frankfurt	RWE	RWE	Energy
Frankfurt	SAP	SAP	IT
Frankfurt	SIE	Siemens	Industrial, electronics
Frankfurt	TKA	ThyssenKrupp	Industrial, manufacturing
Frankfurt	VOW3	Volkswagen Group	Manufacturing
London	AAL	Anglo American plc	Mining
London	ABF	Associated British Foods	Food
London	ADM	Admiral Group	Insurance
London	ADN	Aberdeen Asset Management	Fund management
London	AGK	Aggreko	Generator hire
London	ANTO	Antofagasta	Mining
London	ARM	ARM Holdings	Engineering
London	AV	Aviva	Insurance
London	AZN	AstraZeneca	Pharmaceuticals
London	BA	BAE Systems	Military
London	BAB	Babcock International	Consulting
London	BARC	Barclays	Banking
London	BG	BG Group	Oil and gas
London	BLND	British Land Co	Property
London	BLT	BHP Billiton	Mining
London	BNZL	Bunzl	Industrial products
London	BP	BP	Oil and gas
London	BRBY	Burberry Group	Fashion
London	BT-A	BT Group	Telecomms
London	CNA	Centrica	Energy
London	CPG	Compass Group	Food
London	CPI	Capita	Support Services
London	CRDA	Croda International	Chemicals
London	CRH	CRH plc	Building materials
London	DGE	Diageo	Beverages
London	EXPN	Experian	Information
London	FLG	Friends Life Group	Investment
London	FRES	Fresnillo plc	Mining
London	GFS	G4S	Security
London	GKN	GKN	Manufacturing
London	GSK	GlaxoSmithKline	Pharmaceuticals
London	HL	Hargreaves Lansdown	Finance
London	HMSO	Hammerson	Property
London	HSBA	HSBC	Banking
London	IAG	International Consolidated Airlines	Transport air
London	IHG	InterContinental Hotels Group	Hotels
London	IMI	IMI plc	Engineering
London	IMT	Imperial Tobacco Group	Tobacco
London	ITRK	Intertek Group	Product testing

Traded in	Ticker	Company	Prime Standard industry group
London	ITV	ITV plc	Media
London	JMAT	Johnson Matthey	Chemicals
London	KGF	Kingfisher plc	Retail homeware
London	LAND	Land Securities Group	Property
London	LGEN	Legal & General	Insurance
London	LLOY	Lloyds Banking Group	Banking
London	MGGT	Meggitt	Engineering
London	MKS	Marks & Spencer Group	Retailer
London	MRO	Melrose plc	Engineering
London	MRW	Morrison Supermarkets	Supermarket
London	NG	National Grid plc	Energy
London	NXT	Next plc	Retail clothing
London	OML	Old Mutual	Insurance
London	PFC	Petrofac	Oil and gas
London	PRU	Prudential plc	Finance
London	RB	Reckitt Benckiser	Consumer goods
London	RBS	Royal Bank of Scotland Group	Banking
London	RDSA	Royal Dutch Shell	Oil and gas
London	REL	Reed Elsevier	Publishing
London	REX	Rexam	Packaging
London	RIO	Rio Tinto Group	Mining
London	RR	Rolls-Royce Group	Manufacturing
London	RRS	Randgold Resources	Mining
London	RSA	RSA Insurance Group	Insurance
London	SAB	SABMiller	Beverages
London	SBRY	J Sainsbury plc	Supermarket
London	SDR	Schroders	Fund management
London	SGE	Sage Group	IT
London	SL	Standard Life	Fund management
London	SMIN	Smiths Group	Engineering
London	SRP	Serco	Outsourced services
London	SSE	SSE plc	Energy
London	STAN	Standard Chartered	Banking
London	SVT	Severn Trent	Water
London	TATE	Tate & Lyle	Food
London	TLW	Tullow Oil	Oil and gas
London	TSCO	Tesco	Supermarket
London	ULVR	Unilever	Consumer goods
London	UU	United Utilities	Water
London	VED	Vedanta Resources	Mining
London	VOD	Vodafone Group	Telecomms
London	WEIR	Weir Group	Engineering
London	WG	Wood Group	Oil and gas
London	WOS	Wolseley plc	Building materials
London	WPP	WPP plc	Media
London	WTB	Whitbread	Retail hospitality
New York	AAPL	Apple Inc.	Consumer electronics
New York	ABT	Abbott Laboratories	Pharmaceuticals
New York	ACN	Accenture plc	Professional services
New York	AIG	American International Group Inc.	Insurance
New York	ALL	Allstate Corp.	Insurance
New York	AMGN	Amgen Inc.	Biotechnology
New York	AMZN	Amazon.com	Internet
New York	APA	Apache Corp.	Oil and Gas
New York	APC	Anadarko Petroleum Corporation	Oil and Gas
New York	AXP	American Express Inc.	Consumer finance
New York	BA	Boeing Co.	Aerospace and defense
New York	BAC	Bank of America Corp	Banking
New York	BAX	Baxter International Inc	medical supplies
New York	BIIB	Biogen Idec	Biotechnology
New York	BK	Bank of New York	Banking
New York	BMY	Bristol-Myers Squibb	Pharmaceuticals
New York	BRK.B	Berkshire Hathaway	Conglomerate
New York	C	Citigroup Inc	Banking
New York	CAT	Caterpillar Inc	Construction and Mining Equipment
New York	CL	Colgate-Palmolive Co.	Personal Care
New York	CMCSA	Comcast Corporation	Telecommunications
New York	COF	Capital One Financial Corp.	Financial Services
New York	COP	ConocoPhillips	Oil and Gas
New York	COST	Costco	Retail
New York	CSCO	Cisco Systems	Networking equipment
New York	CVS	CVS Caremark	Health Care
New York	CVX	Chevron	Oil and gas

Traded in	Ticker	Company	Prime Standard industry group
New York	DD	DuPont	Chemical industry
New York	DIS	The Walt Disney Company	Broadcasting and Entertainment
New York	DOW	Dow Chemical	Chemicals
New York	DVN	Devon Energy	Energy
New York	EBAY	eBay Inc.	Internet
New York	EMC	EMC Corporation	Computer storage
New York	EMR	Emerson Electric Co.	Electrical equipment
New York	EXC	Exelon	Energy
New York	F	Ford Motor	Manufacturing
New York	FCX	Freeport-McMoran	Mining
New York	FDX	FedEx	Courier
New York	FOXA	Twenty-First Century Fox, Inc	Media
New York	GD	General Dynamics	Aerospace and Defense
New York	GE	General Electric Co.	Conglomerate
New York	GILD	Gilead Sciences	Biotechnology
New York	GM	General Motors	Manufacturing
New York	GS	Goldman Sachs	Banking
New York	HAL	Halliburton	Oilfield services
New York	HD	Home Depot	Retail
New York	HON	Honeywell	Conglomerate
New York	HPQ	Hewlett Packard Co	Computer and IT
New York	IBM	International Business Machines	Computers and Technology
New York	INTC	Intel Corporation	Semiconductors
New York	JNJ	Johnson & Johnson Inc	Pharmaceuticals
New York	JPM	JP Morgan Chase & Co	Banking
New York	KO	The Coca-Cola Company	Beverages
New York	LLY	Eli Lilly and Company	Pharmaceuticals
New York	LMT	Lockheed-Martin	Aerospace and Defense
New York	LOW	Lowe's	Retail
New York	MA	Masterclass Inc	Banking
New York	MCD	McDonald's Corp	Fast Food
New York	MDLZ	Mondelēz International	Food processing
New York	MDT	Medtronic Inc.	Medical equipment
New York	MET	Metlife Inc.	Financial Services
New York	MMM	3M Company	Conglomerate
New York	MO	Altria Group	Tobacco
New York	MON	Monsanto	Agribusiness
New York	MRK	Merck & Co.	Pharmaceuticals
New York	MS	Morgan Stanley	Banking
New York	MSFT	Microsoft	Software
New York	NKE	Nike	Apparel
New York	NOV	National Oilwell Varco	Oilfield services
New York	NSC	Norfolk Southern Corp	Transportation (Railway)
New York	ORCL	Oracle Corporation	Software
New York	OXY	Occidental Petroleum Corp.	Oil and Gas
New York	PEP	Pepsico Inc.	Beverages
New York	PFE	Pfizer Inc	Pharmaceuticals
New York	PG	Procter & Gamble Co	Consumer goods
New York	PM	Phillip Morris International	Tobacco
New York	QCOM	Qualcomm Inc.	Semiconductors, Telecommunications
New York	RTN	Raytheon Co (NEW)	Aerospace and Defense
New York	SBUX	Starbucks Corporation	Coffee shop
New York	SLB	Schlumberger	Oilfield services
New York	SO	Southern Company	Energy and Telecommunications
New York	SPG	Simon Property Group, Inc.	Real estate
New York	T	AT&T Inc	Telecommunications
New York	TGT	Target Corp.	Retail
New York	TWX	Time Warner Inc.	Media
New York	TXN	Texas Instruments	Semiconductors
New York	UNH	UnitedHealth Group Inc.	Health Care
New York	UNP	Union Pacific Corp.	Transportation (Railway)
New York	UPS	United Parcel Service Inc	Courier
New York	USB	US Bancorp	Banking
New York	UTX	United Technologies Corp	Conglomerate
New York	V	Visa Inc.	Banking
New York	VZ	Verizon Communications Inc	Telecommunications
New York	WBA	Walgreens Boots Alliance	Pharmaceuticals, Retail
New York	WFC	Wells Fargo	Banking
New York	WMT	Wal-Mart	Retail
New York	XOM	Exxon Mobil Corp	Oil and Gas
Paris	AC	Accor	hotels
Paris	ACA	Crédit Agricole	banks
Paris	AF	Air France	Airline

Traded in	Ticker	Company	Prime Standard industry group
Paris	AI	Air Liquide	commodity chemicals
Paris	AIR	Airbus Group	aerospace
Paris	AKE	Arkema chemicals	chemicals
Paris	ALO	Alstom	industrial machinery
Paris	ALU	Alcatel-Lucent	telecommunications
Paris	BN	Groupe Danone	food products
Paris	BNP	BNP Paribas	banks
Paris	CA	Carrefour	food retailers and wholesalers
Paris	CAP	Capgemini	computer services
Paris	CS	AXA	full line insurance
Paris	DG	Vinci	heavy construction
Paris	EDF	EDF	electricity
Paris	EI	Essilor	medical supplies
Paris	EN	Bouygues	heavy construction
Paris	FP	Total	integrated oil and gas
Paris	GLE	Société Générale	banks
Paris	GSZ	GDF Suez	gas distribution
Paris	KER	Kering	retail business
Paris	LG	Lafarge	building materials and fixtures
Paris	LR	Legrand	electrical components and equipment
Paris	MC	LVMH	clothing and accessories
Paris	ML	Michelin	tires
Paris	OR	L'Oréal	personal products
Paris	ORA	Orange	telecommunications
Paris	PUB	Publicis	media agencies
Paris	RI	Pernod Ricard	distillers and vintners
Paris	RNO	Renault	automobiles
Paris	SAF	Safran	aerospace and defence
Paris	SGO	Saint-Gobain	building materials and fixtures
Paris	STM	STMicroelectronics	semiconductors
Paris	SU	Schneider Electric	electrical components and equipment
Paris	TEC	Technip	oil equipment and services
Paris	VIE	Veolia Environnement	water
Paris	VIV	Vivendi	broadcasting and entertainment
Paris	VK	Vallourec	industrial machinery
Tel Aviv	BEZQ	Bezeq The Israel Telecommunication Corp, Ltd.	Telecommunication
Tel Aviv	CEL	Cellcom (Israel)	Telecommunication
Tel Aviv	CLIS	Clal Insurance Enterprises Holdings Ltd.	Insurance
Tel Aviv	DLEKG	Delek Group	Oil and Gas
Tel Aviv	DSCT	Israel Discount Bank Ltd	Banks
Tel Aviv	ESLT	Elbit Systems	Aerospace and Defence
Tel Aviv	FRUT	Frutarom Industries, Ltd.	Chemicals
Tel Aviv	GZT	Gazit-Globe Ltd.	Real Estate
Tel Aviv	HARL	Harel Insurance Inv. & Fin. Services Ltd	Insurance
Tel Aviv	ICL	Israel Chemicals Ltd.	Chemicals
Tel Aviv	LUMI	Bank Leumi Ltd.	Banks
Tel Aviv	MGDL	Migdal Insurance and Financial Holdings Ltd.	Insurance
Tel Aviv	MZTF	Bank Mizrahi-Tfahot Ltd	Banks
Tel Aviv	NICE	NICE Systems Ltd.	Technology
Tel Aviv	ORL	BAZAN - Oil Refineries Ltd	Oil & Gas Producers
Tel Aviv	ORMT	Ormat Industries	Alternative Energy
Tel Aviv	OSEM	Osem	Food Producers
Tel Aviv	POLI	Bank Hapoalim Ltd.	Banks
Tel Aviv	PRGO	Perrigo Company	Pharmaceuticals
Tel Aviv	PTNR	Partner Communications Company Ltd.	Telecommunication
Tel Aviv	PZOL	Paz Oil Company Ltd.	Oil & Gas Services
Tel Aviv	TEVA	Teva Pharmaceutical Industries Ltd.	Pharmaceuticals

Online Appendix Table 3: Complete list of all stocks included in the analysis of Section 6.1 of the main paper. The following procedure was used to obtain the dataset: First, to obtain the stock symbols, the Wikipedia page for the respective stock index was scraped on April 23, 2015. Second, the corresponding stock prices were extracted from Yahoo! Finance and converted to daily returns. The data ranges from 01/01/2011 until 03/20/2015. To avoid missing values, we dropped all stocks that were not publicly listed during the entire timespan. We kept only the primary listing for stocks listed on multiple stock exchanges, and only those days that were active trading days on all five stock exchanges. After consolidating the data to correct for missing values, 272 stocks remained in the dataset spanning 687 observations.

H Mathematical Appendix

H.1 Proofs

Proof of Theorem 1. Suppose Theorem 1 did not hold. Then, there exists an $l^* \in 1, \ldots, r^*$, such that for all k with $R_{l^*,k} \neq 0$ there exists a set $\mathcal{L}_k \ni l^*, |\mathcal{L}_k| > 1$, such that not only $R_{l^*,k} \neq 0$, but also $R_{l',k} \neq 0$ for all $l' \in \mathcal{L}_k, l' \neq l^*$. Now consider

$$\|\tilde{\lambda}_{\bullet k}\|_0 = \| \sum_{l=1}^{r} \lambda^*_{\bullet l} R_{lk} \|_0 = \| \sum_{l \in \mathcal{L}_k} \lambda^*_{\bullet l} R_{lk} \|_0.$$

Let the vector w^k be defined as the $(r-1) \times 1$ vector obtained when deleting the entry with index l^* from $\tilde{w}^k = \frac{R_{\bullet k}}{-R_{l^*,k}}$, where we use the fact that $R_{l^*,k} \neq 0$. Then,

$$\| \sum_{l \in \mathcal{L}_k} \lambda^*_{\bullet l} R_{lk} \|_0 = \| \sum_{l \in \mathcal{L}_k} \lambda^*_{\bullet l} \frac{R_{lk}}{-R_{l^*,k}} \|_0 = |\mathcal{A}_{w^k,-l^*} \cup \mathcal{A}_{l^*}| - |\mathcal{C}| = |\mathcal{A}_{l^*}| + |\mathcal{A}_{w^k,-l^*} \cap \mathcal{A}^c_{l^*}| - |\mathcal{C}| \,,$$

where $\mathcal{C}$ is defined as the set of indices i, such that

$$\sum_{l \neq l^*, l \in \mathcal{L}_k} -\frac{R_{lk}}{R_{l^*,k}} \lambda^*_{il} = \lambda^*_{i,l^*} \quad \text{and} \quad \lambda^*_{i,l^*} \neq 0.$$

By Assumption 2, $|\mathcal{A}_{w^k,-l^*} \cap \mathcal{A}^c_{l^*}| > |\mathcal{C}|$, and therefore

$$|\mathcal{A}_{l^*}| + |\mathcal{A}_{w^k,-l^*} \cap \mathcal{A}^c_{l^*}| - |\mathcal{C}| > |\mathcal{A}_{l^*}|.$$

Now consider setting $R_{l',k}$, $l' \in \mathcal{L}_k$ to zero, and denote the corresponding matrix by R^*, with $R^*_{l^*,k}$ being the remaining non-zero entry in the kth column. Note that this will always be possible while preserving the nonsingularity of R^* until for every $l = 1, \ldots, r^*$, there exists an index k, such that $R_{lk} \neq 0$ and $R_{l',k} = 0 \; \forall l' \neq l$.[5] It immediately follows that

$$\|\tilde{\lambda}_{\bullet k}\|_0 = \| \sum_{l \in \mathcal{L}_k} \lambda^*_{\bullet l} R_{lk} \|_0 > |\mathcal{A}_{l^*}| = \|\lambda^*_{\bullet l^*} R^*_{l^*,k}\|_0 = \|\Lambda^* R^*\|_0,$$

which contradicts that R is a solution to the minimization problem (9).

We conclude that for every $l = 1, \ldots, r^*$, there exists an index k, such that $R_{lk} \neq 0$ and $R_{l',k} = 0$ $\forall l' \neq l$, and that the corresponding index l is distinct for each k. This completes the proof. □

Proof of Proposition 1. **Part a)**. Since $\breve{\Lambda} = \Lambda^* R$ for some nonsingular matrix R, it follows from

[5] To see this, note that Λ^* and $\tilde{\Lambda}$ span the same space. Setting $R_{l',k}$, $l' \neq l^*$ equal to zero simply amounts to rotating $\tilde{\lambda}_{\bullet k}$ (one of the vectors spanning the space spanned by $\tilde{\Lambda}$) to align with $\lambda^*_{\bullet r}$.

Lemma 1 that $|\breve{\mathcal{A}}_k| \geq n - b^*$ for $k = 1, \ldots, r$, where $|\breve{\mathcal{A}}_k| = \|\breve{\lambda}_{\bullet k}\|_0 = \|\tilde{\lambda}_{\bullet k}\|_0$. Thus, it follows that

$$\mathcal{L}_0(\breve{\Lambda}) = \max_k(\sum_{i=1}^{n} \mathbf{1}\{\left|\breve{\lambda}_{ik}\right| = 0\}) = \max_k(|\breve{\mathcal{A}}_k^c|) \leq b^* = o(n).$$

Part b). If there exists a factor F_k with $|\mathcal{A}_k| < (1-\gamma)n$ for some $\gamma \in (0, 1]$, then clearly $\mathcal{L}_0(\Lambda^*) \geq \gamma n$. Further, since $\mathcal{L}_0(\breve{\Lambda}) \geq \mathcal{L}_0(\Lambda^*)$, it also immediately follows that $\mathcal{L}_0(\breve{\Lambda}) \geq \gamma n$. □

Proof of Theorem 2. Consider $\lambda_{\bullet k} = w_1 \lambda^*_{\bullet k} + w_2 v_{\bullet k}$, where $v_{\bullet k}$ is an arbitrary linear combination of $\lambda^*_{\bullet l}$, $l = 1, \ldots, r$, such that $\|v_{\bullet k}\|_2^2 = n$, $\lambda^*_{\bullet k} \perp v_{\bullet k}$ and $w_1^2 + w_2^2 = 1$. This can be thought of as considering all rotations of $\lambda^*_{\bullet k}$ in the subspace spanned by Λ^* without changing its length (in the standard ℓ_2-sense). Next, note that $v_{\bullet k} = v_{\bullet k}^{\mathcal{A}_k} + v_{\bullet k}^{\mathcal{A}_k^c}$, and therefore

$$\begin{aligned} \|\lambda_{\bullet k}\|_1 = \|w_1\lambda^*_{\bullet k} + w_2 v_{\bullet k}\|_1 &= \|w_1 \lambda^{*\mathcal{A}_k}_{\bullet k} + w_2 v^{\mathcal{A}_k}_{\bullet k}\|_1 + |w_2|\, \|v^{\mathcal{A}_k^c}_{\bullet k}\|_1 \\ &= |w_1|\, \|\lambda^{*\mathcal{A}_k}_{\bullet k}\|_1 - |w_2|\, \beta^k(v_{\bullet k}) + |w_2|\, \|v^{\mathcal{A}_k^c}_{\bullet k}\|_1 \qquad \text{(OA.9)} \\ &= |w_1|\, \|\lambda^*_{\bullet k}\|_1 + |w_2| \left(\|v^{\mathcal{A}_k^c}_{\bullet k}\|_1 - \beta^k(v_{\bullet k})\right), \end{aligned}$$

where (OA.9) follows from Lemma 3 when we consider a small neighborhood around $\lambda^*_{\bullet k}$. To make this explicit, set $w_1 = \sqrt{1 - w_2^2}$ and $|w_2| = \epsilon$. We want to show that

$$\begin{aligned} & \sqrt{1-\epsilon^2}\|\lambda^*_{\bullet k}\|_1 + \epsilon \left(\|v^{\mathcal{A}_k^c}_{\bullet k}\|_1 - \beta^k(v_{\bullet k})\right) > \|\lambda^*_{\bullet k}\|_1 \\ \Leftrightarrow \quad & \left(\sqrt{1-\epsilon^2} - 1\right) \|\lambda^*_{\bullet k}\|_1 + \epsilon \left(\|v^{\mathcal{A}_k^c}_{\bullet k}\|_1 - \beta^k(v_{\bullet k})\right) > 0 \\ \Leftrightarrow \quad & \frac{\sqrt{1-\epsilon^2} - 1}{\epsilon} + \frac{\|v^{\mathcal{A}_k^c}_{\bullet k}\|_1 - \beta^k(v_{\bullet k})}{\|\lambda^*_{\bullet k}\|_1} > 0. \end{aligned}$$

By Lemma 2, the first part can be made arbitrarily small for small enough values of ϵ. On the other hand, the second part is larger than zero by Assumption 3 and does not depend on ϵ. We therefore conclude that for a small enough neighborhood, $\|\lambda_{\bullet k}\|_1 > \|\lambda^*_{\bullet k}\|_1$ for all rotations of $\lambda^*_{\bullet k}$ in the subspace spanned by Λ^* and hence $\|\lambda^*_{\bullet k}\|_1$ is a local minimum of (14). □

Proof of Theorem 3. By Lemma 4 there exists a $(1 \times r)$ vector $\psi_{i\bullet}$ with $\psi_{il} = O_p(\frac{1}{\sqrt{n}})$ for $l = 1, \ldots, r$, such that

$$\lambda^0_{ik} - \sum_{l=1}^{r} \lambda^*_{il} H_{lk} = \sum_{l=1}^{r} \psi_{il} H_{lk}.$$

Thus,

$$\lambda^0_{ik} - (\lambda^*_{i\bullet} - \psi_{i\bullet}) H_{\bullet k} = 0$$

$$\Leftrightarrow \qquad \lambda_{ik}^0 - \dot{\lambda}_{i\bullet} H_{\bullet k} = 0, \tag{OA.10}$$

by defining $\dot{\lambda}_{i\bullet} \equiv (\lambda_{i\bullet}^* - \psi_{i\bullet})$. Equation (OA.10) states that Λ^0 is an exact rotation of $\dot{\Lambda}$, where $\dot{\lambda}_{ik} - \lambda_{ik}^* = O_p(\frac{1}{\sqrt{n}})$ for all i, k.

Further, note that, under Assumption 4(a), $\sum_{i \notin \mathcal{A}_k} \left|\dot{\lambda}_{ik}\right| = O(\sqrt{n})$ for all $F_k \in \mathcal{F}$, since $\sum_{i \notin \mathcal{A}_k} \left|\dot{\lambda}_{ik}\right| = \sum_{i \notin \mathcal{A}_k} \left|(\lambda_{ik}^* + (\dot{\lambda}_{ik} - \lambda_{ik}^*)\right| \leq \sum_{i \notin \mathcal{A}_k} \left|\lambda_{ik}^*\right| + \sum_{i \notin \mathcal{A}_k} \left|\dot{\lambda}_{ik} - \lambda_{ik}^*\right| = O(\sqrt{n})$. Similarly, Assumptions 4(b)-4(c) hold for $\dot{\Lambda}$ if they hold for Λ^* and thus the conditions in Assumption 4 hold for $\dot{\Lambda}$ if they are satisfied for Λ^*.

We next show that the conditions in Assumption 5 also hold for $\dot{\Lambda}$ if they are satisfied for Λ^*. Suppose $F_k \in \mathcal{F}$ and let $\dot{v}_{\bullet k}$ be an arbitrary linear combination of $\dot{\lambda}_{\bullet l}$, $l = 1, \ldots, r$, such that $\|\dot{v}_{\bullet k}\|_2^2 = n$ and $\dot{\lambda}_{\bullet k} \perp \dot{v}_{\bullet k}$. Then, there exists a $v_{\bullet k} \in V_k$, such that

$$\begin{aligned}
\beta^k(\dot{v}_{\bullet k}) =& \left| \sum_{i \in \mathcal{A}_k} |\dot{v}_{ik}| \, \mathbf{1}\{\dot{\lambda}_{ik}\dot{v}_{ik} \geq 0\} - \sum_{i \in \mathcal{A}_k} |\dot{v}_{ik}| \, \mathbf{1}\{\dot{\lambda}_{ik}\dot{v}_{ik} < 0\} \right| \\
=& \left| \sum_{i \in \mathcal{A}_k} |\dot{v}_{ik}| \, \mathbf{1}\{\lambda_{ik}^*\dot{v}_{ik} \geq 0\} - \sum_{i \in \mathcal{A}_k} |\dot{v}_{ik}| \, \mathbf{1}\{\lambda_{ik}^*\dot{v}_{ik} < 0\} \right| \qquad \text{(OA.11)} \\
=& \left| \left(\sum_{i \in \mathcal{A}_k} |v_{ik}| \, \mathbf{1}\{\lambda_{ik}^*\dot{v}_{ik} \geq 0\} - \sum_{i \in \mathcal{A}_k} |v_{ik}| \, \mathbf{1}\{\lambda_{ik}^*\dot{v}_{ik} < 0\} \right) \right. \\
& \left. + \left(\sum_{i \in \mathcal{A}_k} (|\dot{v}_{ik}| - |v_{ik}|) \mathbf{1}\{\lambda_{ik}^*\dot{v}_{ik} \geq 0\} - \sum_{i \in \mathcal{A}_k} (|\dot{v}_{ik}| - |v_{ik}|) \mathbf{1}\{\lambda_{ik}^*\dot{v}_{ik} < 0\} \right) \right| \\
=& \left| \sum_{i \in \mathcal{A}_k} |v_{ik}| \, \mathbf{1}\{\lambda_{ik}^*\dot{v}_{ik} \geq 0\} - \sum_{i \in \mathcal{A}_k} |v_{ik}| \, \mathbf{1}\{\lambda_{ik}^*\dot{v}_{ik} < 0\} + O(\sqrt{n}) \right| \qquad \text{(OA.12)} \\
=& \left| \sum_{i \in \mathcal{A}_k} |v_{ik}| \, \mathbf{1}\{\lambda_{ik}^* v_{ik} \geq 0\} - \sum_{i \in \mathcal{A}_k} |v_{ik}| \, \mathbf{1}\{\lambda_{ik}^* v_{ik} < 0\} \right| + O(\sqrt{n}) \qquad \text{(OA.13)} \\
\leq& \, \beta^k(v_{\bullet k}) + O(\sqrt{n}),
\end{aligned}$$

where (OA.11) follows from the fact that the signs of $\dot{\lambda}_{ik}$ and λ_{ik}^* will be identical on $\mathcal{A}_k$, because of the lower bound on λ_{ik}^* from Assumption 4(c). (OA.12) follows from Lemma 5, which states that, if two vectors are "close," their orthogonal complements are also close. Finally, (OA.13) follows from Lemma 6.

Thus, if $F_k \in \mathcal{F}$, and thus for some $c_{min} > 0$ and $N < \infty$, whenever $n > N$,

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

the following must also be true:

$$\|\dot{v}_{\bullet k}^{\mathcal{A}_k^c}\|_1 - \beta^k(\dot{v}_{\bullet k}) > c_{min} n^{\frac{3}{4}} \qquad \forall \dot{v}_{\bullet k} \in \dot{V}_k,$$

where $\dot{V}_k$ denotes the set of all linear combinations $\dot{v}_{\bullet k}$ of $\dot{\lambda}_{\bullet l}$, $l = 1, \ldots, r$, such that $\|\dot{v}_{\bullet k}\|_2^2 = n$ and $\dot{\lambda}_{\bullet k} \perp \dot{v}_{\bullet k}$. Therefore, Assumption 5 is satisfied for $\dot{\Lambda}$ if it is satisfied for Λ^* and we can invoke Lemma 7 to yield $\bar{\lambda}_{ik} = \dot{\lambda}_{ik} + O(n^{-1/4})$. Using the triangle inequality, it then immediately also follows that

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

Similarly, by Lemma 7, $\|\dot{\lambda}_{\bullet k} - \bar{\lambda}_{\bullet k}\| = O(n^{\frac{1}{4}})$. Using the triangle inequality again we obtain $\|\lambda_{\bullet k}^* - \bar{\lambda}_{\bullet k}\| = O_p(n^{\frac{1}{4}})$, and thus $\frac{1}{n}\|\lambda_{\bullet k}^* - \bar{\lambda}_{\bullet k}\|^2 = O_p(n^{-\frac{1}{2}})$. □

H.2 Auxiliary Lemmata

Lemma 1. *Suppose* $|\mathcal{A}_k| = n$ *for* $k = 1, \ldots, r$, *and define* $b^* = \max_k \max_z b_k(z)$, *where* $b_k(z)$ *is defined in Assumption 2. Then, for any nonsingular matrix* R, $|\tilde{\mathcal{A}}_k| \geq n - b^*$ *for* $k = 1, \ldots, r$, *where* $\tilde{\Lambda} = \Lambda^* R$ *and* $|\tilde{\mathcal{A}}_k| = \|\tilde{\lambda}_{\bullet k}\|_0$.

Proof. Suppose Lemma 1 does not hold, and there exists a linear combination $R_{\bullet k}$, $\|R_{\bullet k}\|_0 > 0$, such that $\|\Lambda^* R_{\bullet k}\|_0 = \|\tilde{\lambda}_{\bullet k}\|_0 < n - b^*$. Then, there must exist a set $\mathcal{D}$, with $|\mathcal{D}| > b^*$ such that

$$\sum \lambda_{il}^* R_{lk} = 0 \ \forall i \in \mathcal{D}.$$

Since $|\mathcal{A}_k| = n$ for $k = 1, \ldots, r$, this also implies that, for some $l^* \in \{1, \ldots, r\}$, $R_{l^*,k} \neq 0$

$$-\sum_{l \neq l^*} \lambda_{il}^* \frac{R_{lk}}{R_{l^*,k}} = \lambda_{i,l^*}^* \ \forall i \in \mathcal{D}.$$

But since, by definition of b^*, $|\mathcal{D}| \leq b^*$, this leads to the desired contradiction. □

Lemma 2.

$$\lim_{\epsilon \to 0} \frac{1 - \sqrt{1 - \epsilon^2}}{\epsilon} = 0.$$

Proof. This immediately follows from L'Hospital's rule:

$$\lim_{\epsilon \to 0} \frac{1 - \sqrt{1 - \epsilon^2}}{\epsilon} = \lim_{\epsilon \to 0} \frac{2\epsilon}{2\sqrt{1 - \epsilon^2}} = 0.$$

$\square$

Lemma 3. *Under Assumption 3 and Assumption 4 (a) and (c), there exists a $c_0 > 0$, such that for all $\epsilon \in (0, c_0)$:*

$$\|\sqrt{1-\epsilon^2}\lambda_{\bullet k}^{*\mathcal{A}_k} + \epsilon v_{\bullet k}^{\mathcal{A}_k}\|_1 = \sqrt{1-\epsilon^2}\|\lambda_{\bullet k}^{*\mathcal{A}_k}\|_1 - \epsilon\, \beta^k(v_{\bullet k}).$$

Proof. Since $\left|\lambda_{ik}^*\right| > c\ \forall i \in \mathcal{A}_k$, for small enough $\epsilon > 0$:

$$\begin{aligned}
&\|\sqrt{1-\epsilon^2}\lambda_{\bullet k}^{*\mathcal{A}_k} + \epsilon v_{\bullet k}^{\mathcal{A}_k}\|_1 \\
&= \sum_{i\in\mathcal{A}_k}\left(\sqrt{1-\epsilon^2}|\lambda_{ik}^*| + \epsilon\left|v_{ik}^{\mathcal{A}_k}\right|\right)\mathbf{1}\{\lambda_{ik}^* v_{ik}^{\mathcal{A}_k} \geq 0\} + \sum_{i\in\mathcal{A}_k}\left(\sqrt{1-\epsilon^2}|\lambda_{ik}^*| - \epsilon\left|v_{ik}^{\mathcal{A}_k}\right|\right)\mathbf{1}\{\lambda_{ik}^* v_{ik}^{\mathcal{A}_k} < 0\} \\
&= \sqrt{1-\epsilon^2}\|\lambda_{\bullet k}^{*\mathcal{A}_k}\|_1 + \epsilon\left(\sum_{i\in\mathcal{A}_k}\left|v_{ik}^{\mathcal{A}_k}\right|\mathbf{1}\{\lambda_{ik}^* v_{ik}^{\mathcal{A}_k} \geq 0\} - \sum_{i\in\mathcal{A}_k}\left|v_{ik}^{\mathcal{A}_k}\right|\mathbf{1}\{\lambda_{ik}^* v_{ik}^{\mathcal{A}_k} < 0\}\right) \\
&= \sqrt{1-\epsilon^2}\|\lambda_{\bullet k}^{*\mathcal{A}_k}\|_1 + \epsilon\left(\sum_{i\in\mathcal{A}_k}|v_{ik}|\,\mathbf{1}\{\lambda_{ik}^* v_{ik} \geq 0\} - \sum_{i\in\mathcal{A}_k}|v_{ik}|\,\mathbf{1}\{\lambda_{ik}^* v_{ik} < 0\}\right) \\
&= \sqrt{1-\epsilon^2}\|\lambda_{\bullet k}^{*\mathcal{A}_k}\|_1 - \epsilon\, \beta^k(v_{\bullet k}),
\end{aligned}$$

by definition of $\beta^k(v_{\bullet k})$. Note that Assumption 1 implies Assumption 4 (a) and (c), and Lemma 3 therefore also holds under Assumptions 1 and 3. $\square$

Lemma 4. *Under Assumption 6 there exists a $(1 \times r)$ vector $\psi_{i\bullet}$, such that*

$$\lambda_{ik}^0 - \sum_{l=1}^{r}\lambda_{il}^* H_{lk} = \sum_{l=1}^{r}\psi_{il}H_{lk},$$

where $\psi_{il} = O_p(\frac{1}{\sqrt{n}})$ for $l = 1, \ldots, r$.

Proof. Define $\xi_{ik} = \lambda_{ik}^0 - \sum_{l=1}^r \lambda_{il}^* H_{lk}$ and note that, by Assumption 6, $\xi_{ik} = O_p(\frac{1}{\sqrt{n}})$. Since H is full rank, we can decompose $\xi_{i\bullet}$ as $\xi_{i\bullet} = H\psi_{i\bullet}$ for a $r \times 1$ vector $\psi_{i\bullet}$, where $\psi_{i\bullet} = H^{-1}\xi_{i\bullet}$. Thus, $\psi_{il} = \sum_{k=1}^r H_{kl}^{-1}\xi_{ik}$, where H_{kl}^{-1} denotes the corresponding element of H^{-1}. By Assumption 6, $\psi_{il} = \sum_{k=1}^r H_{kl}^{-1}\xi_{ik} < C\sum_{k=1}^r \xi_{ik} = O_p(\frac{1}{\sqrt{n}})$, which completes the proof.

$\square$

Lemma 5. *Suppose Assumption 4 holds and let $\dot{\lambda}_{\bullet k}$ be a vector with $\|\dot{\lambda}_{ik} - \lambda_{ik}^*\| = O_p(\frac{1}{\sqrt{n}})\ \forall i$ as $n \to \infty$ and let $\dot{v}_{\bullet k}$ be a linear combination of $\dot{\lambda}_{\bullet l}$, $l = 1, \ldots, r$, such that $\|\dot{v}_{\bullet k}\|_2^2 = n$ and $\dot{\lambda}_{\bullet k} \perp \dot{v}_{\bullet k}$.*
Then, there exists a vector $v_{\bullet k}$ with

1. $\lambda_{\bullet k}^{*\prime} v_{\bullet k} = 0$,

2. $(\dot{v}_{ik} - v_{ik}) = O_p(\frac{1}{\sqrt{n}})$.

Proof. Define $S_k = \sum_i^{\mathcal{A}_k^c} \big(\lambda_{ik}^* - \dot{\lambda}_{ik}\big)\dot{v}_{ik}$ and note that for some constant $V < \infty$, $S_k \leq V \sum_i^{\mathcal{A}_k^c} \big(\lambda_{ik}^* - \dot{\lambda}_{ik}\big) = O(\sqrt{n})$.

Let $s_{ik} = -\frac{S_k}{\lambda_{ik}^* |\mathcal{A}_k|}$ for $i \in \mathcal{A}_k$ and

$$v_{ik} = \begin{cases} \dot{v}_{ik} - \frac{\dot{v}_{ik}}{\lambda_{ik}^*}(\lambda_{ik}^* - \dot{\lambda}_{ik}) + s_{ik} & \text{if } i \in \mathcal{A}_k \\ \dot{v}_{ik} & \text{otherwise.} \end{cases}$$

Then,

$$\begin{aligned}
\sum_{i=1}^{n} \lambda_{ik}^* v_{ik} &= \sum_i^{\mathcal{A}_k} \lambda_{ik}^* v_{ik} + \sum_i^{\mathcal{A}_k^c} \lambda_{ik}^* v_{ik} = \sum_i^{\mathcal{A}_k} \lambda_{ik}^* \Big(\dot{v}_{ik} - \frac{\dot{v}_{ik}}{\lambda_{ik}^*}(\lambda_{ik}^* - \dot{\lambda}_{ik}) + s_{ik}\Big) + \sum_i^{\mathcal{A}_k^c} \lambda_{ik}^* \dot{v}_{ik} \\
&= \sum_i^{\mathcal{A}_k} \Big(\lambda_{ik}^* \dot{v}_{ik} - \dot{v}_{ik}(\lambda_{ik}^* - \dot{\lambda}_{ik}) + \lambda_{ik}^* s_{ik}\Big) + \sum_i^{\mathcal{A}_k^c} \lambda_{ik}^* \dot{v}_{ik} \\
&= \sum_i^{\mathcal{A}_k} \Big(\dot{\lambda}_{ik} \dot{v}_{ik} + \lambda_{ik}^* s_{ik}\Big) + \sum_i^{\mathcal{A}_k^c} \lambda_{ik}^* \dot{v}_{ik} \\
&= \sum_i^{n} \dot{\lambda}_{ik} \dot{v}_{ik} - \sum_i^{\mathcal{A}_k^c} \dot{\lambda}_{ik} \dot{v}_{ik} + \sum_i^{\mathcal{A}_k} \lambda_{ik}^* s_{ik} + \sum_i^{\mathcal{A}_k^c} \lambda_{ik}^* \dot{v}_{ik} \\
&= -\sum_i^{\mathcal{A}_k^c} \dot{\lambda}_{ik} \dot{v}_{ik} + \sum_i^{\mathcal{A}_k} \lambda_{ik}^* s_{ik} + \sum_i^{\mathcal{A}_k^c} \lambda_{ik}^* \dot{v}_{ik} = \sum_i^{\mathcal{A}_k} \lambda_{ik}^* s_{ik} + \sum_i^{\mathcal{A}_k^c} \big(\lambda_{ik}^* - \dot{\lambda}_{ik}\big)\dot{v}_{ik} \\
&= -\sum_i^{\mathcal{A}_k} \lambda_{ik}^* \frac{S_k}{\lambda_{ik}^* |\mathcal{A}_k|} + S_k = 0.
\end{aligned}$$

Further, since by Assumption 4, $|\mathcal{A}_k| > c_0 n$ for some $c_0 > 0$, and λ_{ik}^* is bounded away from zero on $\mathcal{A}_k$, $s_{ik} = O_p(\frac{1}{\sqrt{n}})$ for $i \in \mathcal{A}_k$. Since also $\frac{v_{ik}}{\dot{\lambda}_{ik}}(\dot{\lambda}_{ik} - \lambda_{ik}^*) = O_p(\frac{1}{\sqrt{n}})$ for $i \in \mathcal{A}_k$, we therefore conclude that $\dot{v}_{ik} - v_{ik} = O_p(\frac{1}{\sqrt{n}})$, which completes the proof. □

Intuitively, Lemma 5 states that, if two vectors are "close," their orthogonal complements are also close.

Lemma 6. *Suppose Assumptions 4-5 hold and $F_k \in \mathcal{F}$. Assume $\dot{\lambda}_{ik} - \lambda_{ik}^* = O(\frac{1}{\sqrt{n}})$ $\forall i$ and let $\dot{v}_{\bullet k}$ be a linear combination of $\dot{\lambda}_{\bullet l}$, $l = 1, \ldots, r$, such that $\|\dot{v}_{\bullet k}\|_2^2 = n$ and $\dot{\lambda}_{\bullet k} \perp \dot{v}_{\bullet k}$. Then,*

$$\begin{aligned}
&\sum_{i \in \mathcal{A}_k} |v_{ik}| \, \mathbf{1}\{\lambda_{ik}^* \dot{v}_{ik} \geq 0\} - \sum_{i \in \mathcal{A}_k} |v_{ik}| \, \mathbf{1}\{\lambda_{ki}^* \dot{v}_{ik} < 0\} \\
&\qquad = \sum_{i \in \mathcal{A}_k} |v_{ik}| \, \mathbf{1}\{\lambda_{ik}^* v_{ik} \geq 0\} - \sum_{i \in \mathcal{A}_k} |v_{ik}| \, \mathbf{1}\{\lambda_{ik}^* v_{ik} < 0\} + O(\sqrt{n}).
\end{aligned}$$

Proof. Let $\mathcal{S} = \{i : v_{ik} = O_p(\frac{1}{\sqrt{n}})\}$. Then,

$$
\begin{aligned}
&\sum_{i\in\mathcal{A}_k} |v_{ik}|\,\mathbf{1}\{\lambda^*_{ik}\dot{v}_{ik} \geq 0\} - \sum_{i\in\mathcal{A}_k} |v_{ik}|\,\mathbf{1}\{\lambda^*_{ik}\dot{v}_{ik} < 0\} \\
&= \Big(\sum_{i\in\mathcal{A}_k\cap\mathcal{S}} |v_{ik}|\,\mathbf{1}\{\lambda^*_{ik}\dot{v}_{ik} \geq 0\} + \sum_{i\in\mathcal{A}_k\cap\mathcal{S}^c} |v_{ik}|\,\mathbf{1}\{\lambda^*_{ik}\dot{v}_{ik} \geq 0\}\Big) \\
&\qquad\qquad - \Big(\sum_{i\in\mathcal{A}_k\cap\mathcal{S}} |v_{ik}|\,\mathbf{1}\{\lambda^*_{ik}\dot{v}_{ik} < 0\} + \sum_{i\in\mathcal{A}_k\cap\mathcal{S}^c} |v_{ik}|\,\mathbf{1}\{\lambda^*_{ik}\dot{v}_{ik} < 0\}\Big) \\
&= \Big(\sum_{i\in\mathcal{A}_k\cap\mathcal{S}} |v_{ik}|\,\mathbf{1}\{\lambda^*_{ik}\dot{v}_{ik} \geq 0\} + \sum_{i\in\mathcal{A}_k\cap\mathcal{S}^c} |v_{ik}|\,\mathbf{1}\{\lambda^*_{ik}v_{ik} \geq 0\}\Big) \\
&\qquad\qquad - \Big(\sum_{i\in\mathcal{A}_k\cap\mathcal{S}} |v_{ik}|\,\mathbf{1}\{\lambda^*_{ik}\dot{v}_{ik} < 0\} + \sum_{i\in\mathcal{A}_k\cap\mathcal{S}^c} |v_{ik}|\,\mathbf{1}\{\lambda^*_{ik}v_{ik} < 0\}\Big) \\
&= \Big(O_p(\sqrt{n}) + \sum_{i\in\mathcal{A}_k\cap\mathcal{S}^c} |v_{ik}|\,\mathbf{1}\{\lambda^*_{ik}v_{ik} \geq 0\}\Big) - \Big(O_p(\sqrt{n}) + \sum_{i\in\mathcal{A}_k\cap\mathcal{S}^c} |v_{ik}|\,\mathbf{1}\{\lambda^*_{ik}v_{ik} < 0\}\Big) \\
&= \sum_{i\in\mathcal{A}_k\cap\mathcal{S}^c} |v_{ik}|\,\mathbf{1}\{\lambda^*_{ik}v_{ik} \geq 0\} - \sum_{i\in\mathcal{A}_k\cap\mathcal{S}^c} |v_{ik}|\,\mathbf{1}\{\lambda^*_{ik}v_{ik} < 0\} + O_p(\sqrt{n}) \\
&= \sum_{i\in\mathcal{A}_k} |v_{ik}|\,\mathbf{1}\{\lambda^*_{ik}v_{ik} \geq 0\} - \sum_{i\in\mathcal{A}_k} |v_{ik}|\,\mathbf{1}\{\lambda^*_{ik}v_{ik} < 0\} + O_p(\sqrt{n}),
\end{aligned}
$$

using that $(\dot{v}_{ik} - v_{ik}) = O_p(\frac{1}{\sqrt{n}})$ by Lemma 5. □

Lemma 7. *Suppose $n \to \infty$, Assumptions 4 and 5 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 $F_k \in \mathcal{F}$, there exists a local minimum of* (OA.6) *at $\bar{R}_{\bullet k}$, with $\bar{\lambda}_{\bullet k} = \Lambda^0 \bar{R}_{\bullet k}$, such that*

$$\bar{\lambda}_{ik} = \lambda^*_{ik} + O(n^{-1/4}) \tag{OA.14}$$

and

$$\frac{1}{n}\|\lambda^*_{\bullet k} - \bar{\lambda}_{\bullet k}\|^2 = O(n^{-\frac{1}{2}}). \tag{OA.15}$$

Proof. Consider $\lambda_{\bullet k} = w_1\lambda^*_{\bullet k} + w_2 v_{\bullet k}$, where $v_{\bullet k}$ is an arbitrary linear combination of $\lambda^*_{\bullet l}$, $l = 1, \ldots, r$, such that $\|v_{\bullet k}\|_2^2 = n$, $\lambda^*_{\bullet k} \perp v_{\bullet k}$, and $w_1^2 + w_2^2 = 1$. This can be thought of as considering all rotations of $\lambda^*_{\bullet k}$ in the subspace spanned by Λ^* without changing its length (in the standard ℓ_2-sense). Then,

$$
\begin{aligned}
\|\lambda_{\bullet k}\|_1 = \|w_1\lambda^*_{\bullet k} + w_2 v_{\bullet k}\|_1 &= \|w_1\lambda^{*\mathcal{A}_k}_{\bullet k} + w_2 v^{\mathcal{A}_k}_{\bullet k}\|_1 + \|w_1\lambda^{*\mathcal{A}^c_k}_{\bullet k} + w_2 v^{\mathcal{A}^c_k}_{\bullet k}\|_1 \\
&\geq \|w_1\lambda^{*\mathcal{A}_k}_{\bullet k} + w_2 v^{\mathcal{A}_k}_{\bullet k}\|_1 + \|w_2 v^{\mathcal{A}^c_k}_{\bullet k}\|_1 - |w_1|\,\|\lambda^{*\mathcal{A}^c_k}_{\bullet k}\|_1 \\
&= |w_1|\,\|\lambda^{*\mathcal{A}_k}_{\bullet k}\|_1 - |w_2|\,\beta^k(v_{\bullet k}) + w_2\|v^{\mathcal{A}^c_k}_{\bullet k}\|_1 - |w_1|\,\|\lambda^{*\mathcal{A}^c_k}_{\bullet k}\|_1
\end{aligned} \tag{OA.16}
$$

$$= |w_1|\, \|\lambda^{*\mathcal{A}_k}_{\bullet k}\|_1 + |w_2| \left(\|v^{\mathcal{A}^c_k}_{\bullet k}\|_1 - \beta^k(v_{\bullet k}) \right) - |w_1|\, \|\lambda^{*\mathcal{A}^c_k}_{\bullet k}\|_1,$$

where (OA.16) follows from Lemma 3 when we consider a small neighborhood around λ^*_k. To make this explicit, set $w_1 = \sqrt{1 - w_2^2}$ and $|w_2| = \epsilon$. We want to show that

$$\begin{aligned}
& \sqrt{1-\epsilon^2}\|\lambda^{*\mathcal{A}_k}_{\bullet k}\|_1 + \epsilon \left(\|v^{\mathcal{A}^c_k}_{\bullet k}\|_1 - \beta^k(v_{\bullet k}) \right) - \sqrt{1-\epsilon^2}\|\lambda^{*\mathcal{A}^c_k}_{\bullet k}\|_1 > \|\lambda^*_{\bullet k}\|_1 \\
\Leftrightarrow \quad & \left(\sqrt{1-\epsilon^2}\right) \|\lambda^{*\mathcal{A}_k}_{\bullet k}\|_1 - \|\lambda^*_{\bullet k}\|_1 + \epsilon \left(\|v^{\mathcal{A}^c_k}_{\bullet k}\|_1 - \beta^k(v_{\bullet k}) \right) - \sqrt{1-\epsilon^2}\|\lambda^{*\mathcal{A}^c_k}_{\bullet k}\|_1 > 0 \\
\Leftrightarrow \quad & \left(\sqrt{1-\epsilon^2} - 1\right) \|\lambda^{*\mathcal{A}_k}_{\bullet k}\|_1 + \left(\|\lambda^{*\mathcal{A}_k}_{\bullet k}\|_1 - \|\lambda^*_{\bullet k}\|_1 \right) \\
& \qquad + \epsilon \left(\|v^{\mathcal{A}^c_k}_{\bullet k}\|_1 - \beta^k(v_{\bullet k}) \right) - \sqrt{1-\epsilon^2}\|\lambda^{*\mathcal{A}^c_k}_{\bullet k}\|_1 > 0 \\
\Leftrightarrow \quad & \left(\sqrt{1-\epsilon^2} - 1\right) \|\lambda^{*\mathcal{A}_k}_{\bullet k}\|_1 + \epsilon \left(\|v^{\mathcal{A}^c_k}_{\bullet k}\|_1 - \beta^k(v_{\bullet k}) \right) - \left(\sqrt{1-\epsilon^2} + 1\right) \|\lambda^{*\mathcal{A}^c_k}_{\bullet k}\|_1 > 0.
\end{aligned}$$

Note that

$$\begin{aligned}
& \left(\sqrt{1-\epsilon^2} - 1\right)\|\lambda^{*\mathcal{A}_k}_{\bullet k}\|_1 + \epsilon \left(\|v^{\mathcal{A}^c_k}_{\bullet k}\|_1 - \beta^k(v_{\bullet k}) \right) - \left(\sqrt{1-\epsilon^2} + 1\right) \|\lambda^{*\mathcal{A}^c_k}_{\bullet k}\|_1 \\
& \qquad > \underbrace{(\sqrt{1-\epsilon^2} - 1)\|\lambda^{*\mathcal{A}_k}_{\bullet k}\|_1}_{\equiv \mathbf{I}} + \underbrace{\epsilon(\|v^{\mathcal{A}^c_k}_{\bullet k}\|_1 - \beta^k(v_{\bullet k}))}_{\equiv \mathbf{II}} - \underbrace{2\|\lambda^{*\mathcal{A}^c_k}_{\bullet k}\|_1}_{\equiv \mathbf{III}}.
\end{aligned}$$

With c_{min} defined in Assumption 3, choose $\epsilon = \frac{\bar{C}}{c_{min}} n^{-\frac{1}{4}}$ for some constant $\bar{C}$. Then, we have the following bounds on the three terms above:

$$\begin{aligned}
\mathbf{I} &= \left(\sqrt{1-\epsilon^2} - 1\right) \|\lambda^{*\mathcal{A}_k}_{\bullet k}\|_1 = O(\frac{1}{\sqrt{n}})O(n) = O(\sqrt{n}), \\
\mathbf{II} &= \epsilon(\|v^{\mathcal{A}^c_k}_{\bullet k}\|_1 - \beta^k(v_{\bullet k})) > n^{-\frac{1}{4}}\bar{C}n^{\frac{3}{4}} = \bar{C}\sqrt{n}, \\
\mathbf{III} &= 2\|\lambda^{*\mathcal{A}^c_k}_{\bullet k}\|_1 = O(\sqrt{n}),
\end{aligned}$$

where the rate in $\mathbf{I}$ follows from the fact that for $0 < \epsilon < 1$, $\left|\sqrt{1-\epsilon^2} - 1\right| < \epsilon^2 = O(\frac{1}{\sqrt{n}})$. The inequality for $\mathbf{II}$ follows from Assumption 5, and the stated rate in $\mathbf{III}$ follows from Assumption 4 (a). Thus, for large enough $\bar{C}$, the second term dominates and is positive.

This guarantees that $\|\lambda_{\bullet k}\|_1 > \|\lambda^*_{\bullet k}\|_1$. Since Λ^* is a feasible solution to (14), and by continuity of the ℓ_1-norm, there must therefore exist a $\bar{\lambda}_{\bullet k} = \bar{w}_1\lambda^*_{\bullet k} + \bar{w}_2 v_{\bullet k}$, $\bar{w}_2 = O(n^{-\frac{1}{4}})$ with $\bar{w}_1^2 + \bar{w}_2^2 = 1$, that is a local minimum of (14). Rewrite $(\lambda^*_{\bullet k} - \bar{\lambda}_{\bullet k})$ as follows:

$$\lambda^*_{\bullet k} - \bar{\lambda}_{\bullet k} = \lambda^*_{\bullet k} - (w_1\lambda^*_{\bullet k} + w_2 v_{\bullet k})\| = \lambda^*_{\bullet k} - \sqrt{1 - w_2^2}\lambda^*_{\bullet k} - w_2 v_{\bullet k} \tag{OA.17}$$

$$= w_2 \left(\frac{1 - \sqrt{1 - w_2^2}}{w_2} \lambda^*_{\bullet k} - v_{\bullet k} \right), \tag{OA.18}$$

where we note that $\frac{1-\sqrt{1-w_2^2}}{w_2} \leq 1$ for $0 < w_2 < 1$. We can use (OA.18) for two results.

First, by orthogonality of $\lambda^*_{\bullet k}$ and $v_{\bullet k}$:

$$w_2 \| \frac{1-\sqrt{1-w_2^2}}{w_2} \lambda^*_{\bullet k} - v_{\bullet k} \| = w_2 \| \frac{1-\sqrt{1-w_2^2}}{w_2} \lambda^*_{\bullet k} - v_{\bullet k} \| < w_2 \| \lambda^*_{\bullet k} - v_{\bullet k} \|,$$

Further, since $\|\lambda^*_{\bullet k}\| = \|v_{\bullet k}\| = \sqrt{n}$, it follows that $w_2 \|\lambda^*_{\bullet k} - v_{\bullet k}\| = w_2\sqrt{2n}$. We therefore conclude that, with $w_2 = O(n^{-\frac{1}{4}})$,

$$\|\lambda^*_{\bullet k} - \bar{\lambda}_{\bullet k}\| < w_2 \|\lambda^*_{\bullet k} - v_{\bullet k}\| = O(n^{-\frac{1}{4}})\sqrt{2n} = O(n^{\frac{1}{4}}),$$

and thus $\frac{1}{n}\|\lambda^*_{\bullet k} - \bar{\lambda}_{\bullet k}\|^2 = O(n^{-\frac{1}{2}})$.

Second, we can write down (OA.18) elementwise to obtain

$$\lambda^*_{ik} - \bar{\lambda}_{ik} = w_2 \left(\frac{1-\sqrt{1-w_2^2}}{w_2} \lambda^*_{ik} - v_{ik} \right)$$

and it immediately follows that $\bar{\lambda}_{ik} = \lambda^*_{ik} + O(n^{-1/4})$. $\square$