Identification Through Sparsity in Factor Models

Simon Freyaldenhoven
Federal Reserve Bank of Philadelphia

This version: June 16, 2020

Online Appendix

List of Tables

1 Minimum overlap required for Assumption 3 to hold on average. . . . . . . . . . . . 4
2 Minimum overlap required for Assumption 3 to hold in 95% of all cases. . . . . . . 5
3 List of all stocks included in financial application. . . . . . . . . . . . . . . . . . . . . . 15

List of Figures

1 $\beta_k$ as a function of group size $|A_k|$. . . . . . . . . . . . . . . . . . . . . . . . . . . 3
2 Objective function for different rotation criteria. . . . . . . . . . . . . . . . . . . . . 7
3 Comparison of loading matrices across different rotation criteria. . . . . . . . . . . 8
4 Boxplot depicting maximum cosine similarity under a four-factor DGP. . . . . . . 10
5 PCA loadings in macroeconomic application. . . . . . . . . . . . . . . . . . . . . . . . . 13
6 Rotated loadings in macroeconomic application. . . . . . . . . . . . . . . . . . . . . . . 14

*Email: simon.freyaldenhoven@phil.frb.org*
A Further Discussion of Assumption 3 and Its Implications

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

$$
\beta^k(v_\star k) = \sum_{i \in A_k} |v_{ik}| 1\{\lambda_{ik}^* v_{ik} \geq 0\} - \sum_{i \in A_k} |v_{ik}| 1\{\lambda_{ik}^* v_{ik} < 0\} \\
= N^+ \frac{1}{N} \sum_{i \in A_k} |v_{ik}| 1\{\lambda_{ik}^* v_{ik} \geq 0\} - N^- \frac{1}{N} \sum_{i \in A_k} |v_{ik}| 1\{\lambda_{ik}^* v_{ik} < 0\} \\
= N^- (|v^+_{ik}| - |v^-_{ik}|) + (N^+ - N^-) |v^+_{ik}|
$$

(1)

Thus, the difference in Equation (1) above and Equation (9) of the main text can be decomposed into the difference between two conditional means multiplied by $N^-$, and the difference between the number of sum terms used in each sum multiplied by a constant. While we treat $\Lambda^*$ as fixed throughout, suppose $\Lambda^*$ was random and assume $r = 2$. Further suppose $\lambda_{ik}$ is distributed symmetrically and independently if $i \in A_k$, and equal to zero otherwise. Then, both of terms of the decomposition in (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 15 of the main text). Thus, $\beta^1(v_\star 1)$ will also be $O_p(\sqrt{n})$. Treating $\Lambda$ as a fixed parameter, we avoid making any distributional assumptions on $\Lambda$, but instead simply define the above difference as $\beta^k(v_\star k)$.

We next illustrate the behavior of $\beta^k = \max_{v_\star k \in V_k} \beta^k(v_\star k)$ in finite sample for a hypothetical $\Lambda^*$. We create a $n \times 2$ loading matrix $\Lambda^*$ with entries $\lambda_{ik}^* \overset{i.i.d.}{\sim} N(0, 1)$ if $i \in A_k$, and $\lambda_{ik}^* = 0$ otherwise. Further, $|A_2| = n$, such that $A_1 \subset A_2$. Online Appendix Figure 1 then depicts how $\beta^1$ changes as we increase the size of the set $A_1$. Online Appendix Figure 1 confirms that $\beta^1$ grows proportionally to the square root of the size of the set $A_1$.

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

$$
\|v_{\star k}^{A_1^c}\|_1 > \beta^k(v_\star k),
$$

(2)
is plausible, we further need to consider $\|v_{\star k}^{A_1^c}\|_1$ and, intuitively, under what conditions $\|v_{\star k}^{A_1^c}\|_1 > \sqrt{n}$.

1Changing the covariance structure such that $\text{Cov}(\lambda_{11}, \lambda_{12}) \neq 0$ does not affect the results below.

2Specifically, we orthonormalize $\Lambda^*$ using Gram-Schmidt, such that $\frac{U^T U}{n} = I_2$, and write $U = (u_1, u_2) = (\lambda_1^*, v)$ (Redefining $\lambda_1^*$ to have unit length is in line with the setup of the paper, see Section 2). Note that, with $r = 2$, the set $V_1$ only contains two vectors that are identical up to a sign indeterminacy.
Online Appendix Figure 1: Illustration of $\beta^1 = \max_{v_i \in V_1} \beta^1(v_1)$ as a function of group size $|A_1|$. $\lambda^{*}_{ik} \overset{i.i.d.}{\sim} N(0, 1)$ if $i \in A_k$, $\lambda^{*}_{ik} = 0$ otherwise, and $A_1 \subset A_2$. Figure based on 1000 simulations.

Recall that, for constants $q_1$ and $q_2$, $v_{*1} = q_1 \lambda^{*}_{11} + q_2 \lambda^{*}_{21}$, and thus $\|v_{*1}\|_1 = q_2 \|\lambda^{*}_{21}\|_1$, a constant times the sum of the absolute values of $\lambda^{*}_{21}$ on $A_{1}^c$. It can be shown that $\|v_{*1}\|_2 = 1$ and $\lambda^{*}_{k1} \perp v_{*k}$ implies $q_2^2 = (1 - [\lambda^{*1}_{k1}, \lambda^{*1}_{k2}]^T)^{-1}$. Hence $q_2 \geq 1$, and $\|\lambda^{*}_{21}\|_1 > \beta^k$ is sufficient for (2) to hold. In general, the sum of the absolute values of $\lambda^{*}_{21}$ on $A_{1}^c$ 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 $\lambda^{*}_{11}$ considered above, we can infer the minimum value needed for $\|v_{*1}\|_1$ to fulfill the condition in (2) for a given group size from Online Appendix Figure 1 (e.g., at $|A_1| = 400$, around 10), we next directly report the number of outcomes affected by $F_2$, but not $F_1$ that is needed for condition (2) to hold across a number of different DGPs for $\Lambda^*$. In particular, let

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

We create a $500 \times 2$ loading matrix $\Lambda^*$ with $A_1 \subset A_2$ and $|A_1| = 300$. We draw $\lambda^{*}_{i1} \overset{i.i.d.}{\sim} N(\mu, \Sigma)$ if $i \in A_1$, $\lambda^{*}_{i2} \overset{i.i.d.}{\sim} N(\mu_2, 1)$ if $i \in A_{1}^c \cap A_2$, and $\lambda^{*}_{ik} = 0$ otherwise. To get a better sense of how demanding Assumption 3 is in practice, we then vary both $\mu$ and $\rho$. The results are depicted in Online Appendix Tables 1 and 2. Each entry depicts the minimum size of $A_{1}^c \cap A_2$ for $\|v_{*k}\|_1 > \beta^k(v_{*k})$ to hold for a given combination of $\mu$ and $\rho$.

Online Appendix Table 1 depicts the minimum number of entries in $A_{1}^c \cap A_2$ needed for condition (2) to hold for $k = 1$ on average across repeated realizations. Note that, with $\mu = (0, 0)$
| Means $\mu_k$ | Correlation $\rho$ |
|--------------|-----------------|
| $\mu_1$ 0 | $\rho$ -0.4, -0.2, 0, 0.2, 0.4 |
| $\mu_2$ 1 |  |
| 0 10 | 11 | 11 | 11 | 10 |
| 0 10 | 10 | 10 | 10 | 10 |
| 1 36 | 31 | 27 | 21 | 16 |
| 1 26 | 21 | 16 | 21 | 11 |
| 2 62 | 53 | 44 | 36 | 27 |
| 2 36 | 27 | 36 | 44 | 53 |

**Online Appendix Table 1:** Smallest number of entries in $A_1^c \cap A_2$ needed for condition (2) to hold on average for $k = 1$. $|A_1| = |A_1 \cap A_2| = 300$. Table varies the distribution of $\lambda_{ik}^*$ through the parameters $\mu$ and $\rho$. In particular, $\lambda_{ik}^* \sim \text{i.i.d. } N(\mu, \Sigma)$ if $i \in A_1$, $\lambda_{i2}^* \sim \text{i.i.d. } N(\mu_2, 1)$ if $i \in A_1^c \cap A_2$, and $\lambda_{ik}^* = 0$ otherwise. Table based on 1000 simulations.

and $\rho = 0$, this DGP is among those considered in Online Appendix Figure 1 (at $|A_1| = 300$ 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 (2) to hold on average. In other words: on average, $F_1 \in F^{\text{exact}}$, and is 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 (2) to hold in 95% of all realizations. Again focusing on the case $\mu = (0, 0)$ and $\rho = 0$, it states that, if $|A_1^c \cap A_2| \geq 28$, $F_1 \in F^{\text{exact}}$ in 95% of all simulations. We note that, while for some combinations of $\mu$ and $\rho$, $|A_1^c \cap A_2|$ is smaller, we need a fairly large number of outcomes affected by $F_2$, but not by $F_1$ in other regions of the parameter space. On the other hand, it still allows for significant overlap between $A_1$ and $A_2$. Further, as we showed in Online Appendix Figure 1, the required size for $|A_1^c \cap A_2|$ will be smaller in cases where the “joint active set” $A_1 \cap A_2$ is smaller than 300, which is frequently the case in applications.

In conclusion, we impose condition (2) as a high-level assumption in Assumption 3. This section outlined how one could alternatively treat the loadings as random, and use restrictions on the distribution of $\lambda_{ik}$ to prove an upper bound on $\beta_k$ directly.
| Means $\mu_k$ | Correlation $\rho$ |
|---------------|-------------------|
| $\mu_1$ $\mu_2$ | -0.4 -0.2 0 0.2 0.4 |
| 0 0 | 25 27 28 26 25 |
| 0 1 | 26 27 27 26 24 |
| 1 1 | 58 51 45 37 32 |
| 1 -1 | 32 38 45 52 59 |
| 2 2 | 77 65 55 46 36 |
| 2 -2 | 35 46 56 65 75 |

**Online Appendix Table 2:** Smallest number of entries in $A_1 \cap A_2$ needed for condition (2) to hold in 95% of all realizations for $k = 1$. $|A_1| = |A_1 \cap A_2| = 300$. Table varies the distribution of $\lambda_{ik}^*$ through the parameters $\mu$ and $\rho$. In particular, $\lambda_{i}^* \sim i.i.d. \mathcal{N}(\mu, \Sigma)$ if $i \in A_1$, $\lambda_{i2}^* \sim i.i.d. \mathcal{N}(\mu_2, 1)$ if $i \in A_1^c \cap A_2$, and $\lambda_{ik}^* = 0$ otherwise. Table based on 1000 simulations.
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 $\ell_1$-norm of a loading vector across all rotations of the Principal Component estimate $\Lambda^0$ in Online Appendix Figure 2a. Online Appendix Figure 2a is identical to the left half of Figure 4 in the main text.

Recall from Equation 3 in the main text that the objective function for the Varimax criterion can be expressed as

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

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

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

as the contribution to the objective function $Q(\Lambda)$ by an individual column $\lambda_k$. The restriction $R_{ik} R_{ik} = 1$, combined with the choice of an initial estimate $\Lambda^0$ that is orthonormal, implies that the second part of (4) is constant. Maximizing (3) is therefore equivalent to maximizing the columnwise $\ell_4$-norm, with the added restriction that the resulting loading vectors are orthonormal.

In the two-dimensional case ($r = 2$), the unit length restriction on $\lambda_k$ means we can write $\lambda_k = \sin(\theta_k) \lambda_{1k} + \cos(\theta_k) \lambda_{2k}$ for some $\theta_k$, making it easy to visualize rotations. We depict the $\ell_4$-norm across rotations $\theta_k$ for a single vector $\lambda_k$ in Online Appendix Figure 2b below.

Further note that at any solution $\tilde{\Lambda}$ to (3), $\tilde{\Lambda}^t \tilde{\Lambda} = I$ implies that

$$\tilde{\lambda}_{1k} \tilde{\lambda}_{2k} = 0 \iff \left[ \sin(\theta_1) \lambda_{1k}^0 + \cos(\theta_1) \lambda_{2k}^0 \right] \left[ \sin(\theta_2) \lambda_{1k}^0 + \cos(\theta_2) \lambda_{2k}^0 \right] = 0 \iff \sin(\theta_1) \sin(\theta_2) + \cos(\theta_1) \cos(\theta_2) = 0 \iff \cos(\theta_1 - \theta_2) = 0 \iff \theta_1 - \theta_2 = \frac{\pi}{2} + g\pi, g \in \mathbb{Z}.$$
$\cos(\theta_k)\lambda_{k2}^0\|^4_4$, subject to the constraint that $(\theta_1 - \theta_2)$ fulfills the condition stated in (5) above. We see in Online Appendix Figure 2b that, without the orthogonality restriction, the nonsingular rotations with the largest $\ell_4$-norm correspond to $\theta_1^{\ell_4}$ and $\theta_2^{\ell_4}$, marked by the dashed black lines. The restriction in (5) forces the difference between any solutions $\tilde{\theta}_1$ and $\tilde{\theta}_2$ (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 would expect this to frequently be the case, and gave an intuitive explanation for this in Section 3.

(a) $\|\lambda_{k}\|_1 = \|\sin(\theta)\lambda_{k1}^0 + \cos(\theta)\lambda_{k2}^0\|_1$ as a function of $\theta$. Minimum achieved at $\tilde{\theta}$.

(b) $\|\lambda_{k}\|_4^4 = \|\sin(\theta)\lambda_{k1}^0 + \cos(\theta)\lambda_{k2}^0\|_4^4$ as a function of $\theta$. Maximum achieved at $\theta^{\ell_4}$. Maximum under the constraint in (5) achieved at $\tilde{\theta}$.

Online Appendix Figure 2: Comparison of objective functions across different criteria. Figure depicts the value of the respective objective function ($\ell_1$- and $\ell_4$-norm) across all rotations in the space spanned by the initial estimate $\Lambda^0$.

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 2 for the Promax criterion.

However, in order to visually assess the performance of the different rotation criteria, we next depict the estimated loading matrix for all four of the above criteria in Online Appendix Figure 3. For reference, Panel 3a repeats the true loading matrix $\Lambda^*$. Panel 3b depicts $\tilde{\Lambda}$, the estimate
Online Appendix Figure 3: Comparison of loading matrices across different rotation criteria. Each panel depicts the loadings associated with all 207 outcomes, where the top diagram depicts $\lambda_{1}$ and bottom panel $\lambda_{2}$. Panels 3b-3e differ in the criterion that determines $\theta_k$ in $\lambda_{k} = sin(\theta_k)\lambda_{1} + cos(\theta_k)\lambda_{2}$. 

(a) True loading matrix $\Lambda^*$. 

(b) The non-singular “rotated” matrix with smallest $\ell_1$-norm $\|\Lambda\|_1$, $\tilde{\Lambda}$.

(c) The “rotated” matrix $\tilde{\Lambda}$ that maximizes the Varimax criterion.

(d) The non-singular “rotated” matrix with largest $\ell_4$-norm $\|\Lambda\|_4$, $\Lambda_{\ell_4}$.

(e) The “rotated” matrix $\Lambda^{PROMAX}$ that is the result of the Promax rotation.
that minimizes the $\ell_1$-norm of the loadings across rotations of $\Lambda^0$, which corresponds to the linear combinations at $\tilde{\theta} = [\tilde{\theta}_1, \tilde{\theta}_2]$ in Online Appendix Figure 2a. Panel 3c depicts $\hat{\Lambda}$, the estimate that maximizes the Varimax criterion across rotations of $\Lambda^0$, which corresponds to the linear combinations at $\hat{\theta} = [\hat{\theta}_1, \hat{\theta}_2]$ in Online Appendix Figure 2b. Panel 3d depicts $\Lambda^{\ell_4}$, the estimate that maximizes the $\ell_4$-norm of the loadings across rotations of $\Lambda^0$, which corresponds to the linear combinations at $\theta^{\ell_4} = [\theta_{1}^{\ell_4}, \theta_{2}^{\ell_4}]$ in Online Appendix Figure 2b. Panel 3e depicts $\Lambda^{PROMAX}$, the estimate that maximizes the Promax criterion across rotations of $\Lambda^0$.

The differences between the four estimates in Online Appendix Figure 3 are small and all four appear to be good estimates of $\Lambda^*$. Since the DGP we considered so far is very stylized, and Online Appendix Figure 3 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 7 in the main text. Recall that under this DGP, there are four factors. Of these four factors, one affects all outcomes, while the remaining three are local. Online Appendix Figure 4 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 100 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 $i \in A_k^c$ (on the left), and one in which $\lambda_{ik} \sim_i.d. N(0, \sigma^2)$, $\sigma^2 = \frac{1}{n}$ for all $i \in A_k^c$ (on the right).

Panels 4a and 4b are identical to Figure 7c-7d 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 $\lambda^*_2$ in particular (the local factor affecting the most outcomes) is significantly lower than that of $\hat{\Lambda}$. Maximizing the $\ell_4$-norm directly (Panels 4e-4f) performs even worse, in particular for $\lambda^*_2$.

Thus, we conclude that our proposed criterion based on the $\ell_1$-norm outperforms all of the considered quartic criteria in this simulation exercise.
Online Appendix Figure 4: Maximum cosine similarity for all four factor loadings $\lambda^*_k$, $k = 1, \ldots, 4$. The first factor is global, while factors 2-4 are local. Boxplots based on 100 realizations.
C Algorithmic Implementation

Recall the minimization problem we consider throughout to identify $\Lambda^*$ (or more precisely, an individual column $\lambda^*_k$):

$$\min_{R_{\cdot k}} \left\| \sum_{l=1}^{r} \lambda^o_l R_{lk} \right\|_1 \text{ such that } R_{\cdot k}' R_{\cdot k} = 1 \forall k. \quad (6)$$

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 $\Lambda^0$ as an initial estimate for $\Lambda^*$ that fulfills $\Lambda^0' \Lambda^0 = I$.

2. Next, we find all local minima of (6). We achieve this by first drawing a random grid of starting points $R_{\cdot k}^{0j}$, $j = 1, \ldots, J$, where $R_{lk}^{0j} = \frac{x_l}{\|x\|} \sim N(0, I_r)$.

   For each starting point $R_{\cdot k}^{0j}$, we find the argmin of (6), denoted by $R_{\cdot k}^{1j}$.

   At the end of this step, we have $J$ candidate solutions $R_{\cdot k}^{1j}$.

3. In general, many of the $J$ 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, such that we are left with $P$ unique rotation vectors $R_{\cdot k}^p$.

   We sort these unique rotation columns $R_{\cdot k}^p$, $p = 1, \ldots, P$ in descending order according to their occurrences.

   At the end of this step, we have $P$ candidate solutions $R_{\cdot k}^p$.

4. Each candidates $R_{\cdot k}^p$ correspond to a local minimum of (6). In this last step, we combine these solutions into the rotation matrix $\tilde{R}$.

   With $R_{\cdot k}^p$ sorted according to their occurrences, we initialize $\tilde{R} = R_{\cdot k}^1$ and iteratively append $R_{\cdot k}^p$, $p = 2, \ldots, P$ whenever the resulting matrix does not become (close to) singular. Denote the resulting $r \times \tilde{r}$ matrix by $\tilde{R} = [\tilde{R}_{\cdot 1}, \ldots, \tilde{R}_{\cdot \tilde{r}}]$.

   Finally, we distinguish two cases:

   The number of random starting points $J$ increases with the number of factors $r$. In particular, for $r = (2, 3, 4, 5, 6)$, we find that grid sizes $J = (150, 300, 1000, 2000, 3000, 5000)$ work well in practice. The constrained minimization in (6) becomes numerically more difficult as $r$ increases. We find that the performance of the MATLAB solver starts to deteriorate for $r > 6$ and leave a refinement of the optimization routine for future research.

   We use $fmincon$, a native optimization routine included in MATLAB. We also implemented our algorithm using $fminsearch$ and $fminunc$, and found both to be significantly slower than $fmincon$. 

---

3The number of random starting points $J$ increases with the number of factors $r$. In particular, for $r = (2, 3, 4, 5, 6)$, we find that grid sizes $J = (150, 300, 1000, 2000, 3000, 5000)$ work well in practice. The constrained minimization in (6) becomes numerically more difficult as $r$ increases. We find that the performance of the MATLAB solver starts to deteriorate for $r > 6$ and leave a refinement of the optimization routine for future research.

4We use $fmincon$, a native optimization routine included in MATLAB. We also implemented our algorithm using $fminsearch$ and $fminunc$, and found both to be significantly slower than $fmincon$. 

---

11
• $\bar{r} \geq r$: We have more candidate solutions than the number of factors. In this case, we simply keep the most frequently found $r$ rotation columns in $\bar{R}$, and $\bar{R} = \bar{R}_{\bullet,1:r}$. The result, $\bar{\Lambda} = \Lambda^0\bar{R}$, is our proposed estimate for the loading matrix $\Lambda^*$.

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

---

5We choose the entry $d$ to pick out the loading vector $\lambda^0_d$ that maximizes the minimum singular value of the combined matrix $[\bar{\Lambda}, \lambda^0_d]$. Intuitively, this is the column in $\Lambda^0$ that is furthest away from any linear combination of the columns in $\bar{\Lambda}$.
Online Appendix Figure 5: Illustration of PCA loadings $\lambda_{*k}^0$ for $k = 1, \ldots, 8$. Bars correspond to the loadings associated with the individual macroeconomic indicators from Section 6.2.
Online Appendix Figure 6: Illustration of additional rotated loadings $\tilde{\lambda}_k$ for $k = 6, 7, 8$. Bars correspond to the loadings associated with the individual macroeconomic indicators from Section 6.2.
E Data Appendix

E.1 Financial Data

The following procedure was used to obtain the dataset:

1. To obtain the stock symbols, the Wikipedia page for the respective stock index was scraped on April 23, 2015.

2. The corresponding stock prices were extracted from Yahoo! Finance and converted to daily returns.

3. 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 only kept the primary listing for stocks listed on multiple stock exchanges, and only those days were kept 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. The complete list of stocks included is provided in Online Appendix Table 3 below. Online Appendix Table 3 also links each stock symbol to its corresponding company and the respective primary industry.

| 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 | BBI    | Beiersdorf | Consumer goods    |
| Frankfurt | BMW    | BMW     | Manufacturing                 |
| Frankfurt | CBK    | Commerzbank | Banking               |
| Frankfurt | CSN    | 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 | EOGAN  | EON     | 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 | VOV3   | Volkswagen Group | Manufacturing |
| London    | AAL    | Anglo American plc | Mining               |
| London    | ABF    | Associated British Foods | Food |

15
| Traded in | Ticker | Company | Prime Standard industry group |
|-----------|--------|---------|-------------------------------|
| 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    | FLC    | Friends Life Group | Investment                |
| London    | FRES   | Fresnillo plc | Mining                      |
| London    | GP    | GAS | Security                       |
| London    | GKN    | GKN | Manufacturing                  |
| London    | GSK    | GlaxoSmithKline | Pharmaceuticals          |
| London    | HL     | Hargreaves Lansdown | Finance                        |
| London    | HMSO   | Hammerson | Property                      |
| London    | HSBA   | HSBC | Banking                       |
| London    | I&O    | International Consolidated Airlines | Transport air |
| London    | IHG    | InterContinental Hotels Group | Hotels                  |
| London    | IML    | IMI plc | Engineering                   |
| London    | IMT    | Imperial Tobacco Group | Tobacco                   |
| London    | ITRK   | Interkt Group | Product testing            |
| 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    | LLFY   | Lloyds Banking Group | Banking                     |
| London    | MGTT   | 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    | Reexam | Packaging                      |
| London    | RIO    | Rio Tinto Group | Mining                       |
| London    | RR     | Rolls-Royce Group | Manufacturing             |
| London    | RRS    | Ratcliff Resources | Mining                       |
| London    | RSA    | RSA Insurance Group | Insurance                      |
| London    | SAB    | SABMiller | Beverages                     |
| London    | SBSY   | J Sainsbury plc | Supermarket               |
| London    | SDN    | 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                         |
| Traded in | Ticker | Company | Prime Standard industry group |
|-----------|--------|---------|-------------------------------|
| London    | TLW    | Tullow Oil | Oil and gas |
| London    | TSCO   | Tesco | Supermarket |
| London    | ULYR   | Unilever | Consumer goods |
| London    | UU     | United Utilities | Water |
| London    | VED    | Vedanta Resources | Mining |
| London    | VOD    | Vodafone Group | Telecoms |
| 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  | AG    | 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      | 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 |
| 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 |
| Traded in | Ticker | Company                                    | Prime Standard industry group |
|-----------|--------|-------------------------------------------|-------------------------------|
| New York  | MMM    | 3M Company                                 | Conglomerate                  |
| New York  | MEO    | 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                                    | 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                                 | Airlines                      |
| Paris     | AI     | Air Liquide                                | commodity chemicals           |
| Paris     | AJR    | Airbus Group                               | aerospace                     |
| Paris     | AKE    | Arkema chemicals                           | chemicals                     |
| Paris     | ALO    | Alstom                                     | industrial machinery          |
| Paris     | ALLU   | 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     | Ettior                                     | medical supplies              |
| Paris     | EN     | Bouygues                                   | heavy construction            |
| Paris     | FP     | Total                                      | integrated oil and gas        |
| Paris     | GLG    | 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    | Vedia Environnement                        | water                         |
| Traded in | Ticker | Company | Prime Standard industry group |
|-----------|--------|---------|-------------------------------|
| 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    | Gazi-Globe Ltd. | Real Estate |
| Tel Aviv  | HARL   | Harel Insurance Ins. & 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 Mizrachi-Tlalot 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  | PROG   | 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.