APPENDIX

A BASIC FACTS

In this section, we present some basic facts that are used in proofs presented in this paper. We use notation $[A : B]$ to denotes the concatenation of matrices A and B that have the same number of rows.

Lemma A.1. *A $T \times T$ matrix S_η defined as $S_\eta = M_1 + \eta P_1$ for a real number η satisfies that*

$$rank(S_\eta) = \begin{cases} T-1 & \text{if} \quad \eta = 0 \\ T & \text{if} \quad \eta \neq 0. \end{cases}$$

Proof. Since P_1 and M_1 are idempotent matrices and $M_1 = I_T - P_1$, we have that $rank(P_1) = tr(P_1) = 1$ and that $rank(M_1) = T - rank(P_1) = T - 1$ (Lütkepohl, 1996).

If $\eta = 0$, $rank(S_\eta) = rank(M_1) = T - 1$. Now, suppose that $\eta \neq 0$. Since M_1 is a symmetric matrix, its eigen-decomposition can be written as

$$M_1 = ULU^\top$$

where the orthonormal matrix $U = [u_1 : u_2 : \cdots : u_{T-1}] \in \mathbb{R}^{T\times(T-1)}$ consists of the eigenvectors of M_1 in its columns and $L = diag(\lambda_1, \lambda_2, \cdots, \lambda_{T-1}) \in \mathbb{R}^{(T-1)\times(T-1)}$ is a diagonal matrix of the associated (nonzero) eigenvalues. It is clear that $\mathbb{1}_T^\top u_t = 0$ for all $t \in \{1, 2, \cdots, T-1\}$. Thus, we can write the eigen-decomposition of the matrix S_η as

$$S_\eta = M_1 + \eta P_1 = ULU^\top + \eta \left(\frac{1}{\sqrt{T}}\mathbb{1}_T\right)\left(\frac{1}{\sqrt{T}}\mathbb{1}_T\right)^\top = \tilde{U}\tilde{L}\tilde{U}$$

where $\tilde{U}$ and $\tilde{\Lambda}$ can be written as

$$\tilde{U} = \left[U : \frac{1}{\sqrt{T}}\mathbb{1}_T\right] \in \mathbb{R}^{T\times T} \quad \text{and} \quad \tilde{L} = diag(\lambda_1, \lambda_2, \cdots, \lambda_{T-1}, \eta) \in \mathbb{R}^{T\times T}.$$

$\tilde{U}$ is orthonormal and all of the diagonal entries of $\tilde{L}$ are nonzero, implying that $rank(S_\eta) = T$. □

Lemma A.2. *For any $m, n \in \mathbb{N}$ and $A \in \mathbb{R}^{m\times n}$, we have*

$$\text{rowsp}(A^\top A) \overset{(a)}{=} \text{colsp}(A^\top A) \overset{(b)}{=} \text{colsp}(A^\top) \overset{(c)}{=} \text{rowsp}(A).$$

Proof. (a) Let $x \in \text{colsp}(A^\top A)$ be given. Then, there exists an $y \in \mathbb{R}^n$ such that $x = A'Ay$. Now, we have that $x^\top = y^\top A^\top A$, which implies that $x \in \text{rowsp}(A^\top A)$. Thus, $\text{colsp}(A^\top A) \subset \text{rowsp}(A^\top A)$. Similarly, it can be proved that $\text{rowsp}(A^\top A) \subset \text{colsp}(A^\top A)$. (b) We refer to Magnus & Neudecker (2019, Chapter 1.7). (c) is clear from the definition of column spaces and row spaces. □

Lemma A.3. *Let $m, n \in \mathbb{N}$, $A \in \mathbb{R}^{m\times n}$ and $V \in \mathbb{R}^{m\times m}$. If V is symmetric and non-singular, then*

$$\text{colsp}(A^\top VA) = \text{colsp}(A^\top A).$$

Proof. To see $\text{colsp}(A^\top VA) \subset \text{colsp}(A^\top A)$, let $x \in \text{colsp}(A^\top VA)$ be given. $\Rightarrow \exists y \in \mathbb{R}^n$ such that $x = (A^\top VA)y = A^\top(VAy)$. $\Rightarrow x \in \text{colsp}(A^\top)$. $\Rightarrow$ By Lemma A.2, $x \in \text{colsp}(A^\top A)$.

To see $\text{colsp}(A^\top VA) \supset \text{colsp}(A^\top A)$, consider the eigen-decomposition of $V = ULU^\top$. Put $Q = L^{1/2}U^\top$. $\Rightarrow Q$ is an $m \times m$ invertible matrix and satisfies $V = Q^\top Q$. Now, let $x \in \text{colsp}(A^\top A)$. $\Rightarrow$ By Lemma A.2, $x \in \text{colsp}(A^\top)$, *i.e.*, $\exists z \in \mathbb{R}^m$ such that

$$x = A^\top z = A^\top Q^\top\left((Q^\top)^{-1}z\right).$$

$\Rightarrow x \in \text{colsp}(A^\top Q^\top)$. By Lemma A.2, $x \in \text{colsp}(A^\top Q^\top QA) = \text{colsp}(A^\top VA)$. □

Lemma A.4. *Let $m, n \in \mathbb{N}$, $A \in \mathbb{R}^{m\times n}$ and $V \in \mathbb{R}^{m\times m}$. If V is symmetric, then*

$$\text{colsp}(A^\top VA) = \text{colsp}(A^\top V).$$

Proof. Let $rank(V) = r \leq m$. Without loss of generality, we can assume that $r \geq 1$, as the case when $r = 0$ is clear. Consider the eigen-decomposition of $V = ULU^\top$, where $L \in \mathbb{R}^{r\times r}$ is a diagonal matrix of real eigenvalues and $U \in \mathbb{R}^{m\times r}$ consists of r orthogonal eigenvectors, *i.e.*, $U^\top U = I_r$. Then, we have that

$$\begin{aligned}
\mathrm{colsp}(A^\top VA) &= \mathrm{colsp}(A^\top ULU^\top A) = \mathrm{colsp}(A^\top UL^{1/2}L^{1/2}U^\top A)\\
&\overset{(a)}{=} \mathrm{colsp}(A^\top UL^{1/2}LL^{1/2}U^\top A) = \mathrm{colsp}(A^\top ULLU^\top A)\\
&\overset{(b)}{=} \mathrm{colsp}(A^\top ULU^\top ULU^\top A) = \mathrm{colsp}(A^\top VVA)\\
&\overset{(c)}{=} \mathrm{colsp}(A^\top V),
\end{aligned}$$

where (a) and (b) are true since L and $U^\top U = I_r$ are symmetric non-singular matrices in conjunction with Lemma A.3. (c) holds true due to Lemma A.2. □

B Derivations of the First-Order Optimality Conditions

In this section, we derive the first-order optimality conditions for our proposed optimization problem ($\mathcal{P}$) based on basic facts of matrix calculus (Lütkepohl, 1996; Magnus & Neudecker, 2019). We use the symbol $\mathsf{d}f$ to denote the differential of a function f of matrices. In this section, we assume that an $N \times N$ matrix V is symmetric and do not assume singularity nor positive-definiteness.

Reformulation of the proposed objective function To simplify notations, we multiply the objective function ϕ by NT and rewrite it as follows.

$$\begin{aligned}
&NT\phi(\Lambda, F)\\
&= \left\|\tilde{X} - \tilde{F}\Lambda^\top\right\|_F^2 + T\eta\left\|\bar{X} - \Lambda\bar{F}\right\|_V^2\\
&= \left\|M_1\left(X - F\Lambda^\top\right)\right\|_F^2 + T\eta\left\|X^\top\left(\frac{1}{T}\mathbb{1}_T\right) - \Lambda F^\top\left(\frac{1}{T}\mathbb{1}_T\right)\right\|_V^2\\
&= \left\|M_1\left(X - F\Lambda^\top\right)\right\|_F^2 + T\eta\left\|(X - F\Lambda^\top)^\top\left(\frac{1}{T}\mathbb{1}_T\right)\right\|_V^2\\
&= tr\left(M_1(X - F\Lambda^\top)(X - F\Lambda^\top)^\top M_1\right) + T\eta\, tr\left(\frac{1}{T}\mathbb{1}_T^\top(X - F\Lambda^\top)V(X - F\Lambda^\top)^\top\frac{1}{T}\mathbb{1}_T\right)\\
&= tr\left(M_1(X - F\Lambda^\top)(X - F\Lambda^\top)^\top M_1\right) + \eta\, tr\left(P_1(X - F\Lambda^\top)V(X - F\Lambda^\top)^\top P_1\right).
\end{aligned} \tag{13}$$

Differentials of the proposed objective function In order to compute the derivatives of ϕ, we first derive differentials of a matrix-valued function $(\Lambda, F) \mapsto (X - F\Lambda^\top)W(X - F\Lambda^\top)^\top$, where W is assumed to be an $N \times N$ symmetric matrix, as follows.

$$\begin{aligned}
&\mathsf{d}\left((X - F\Lambda^\top)W(X - F\Lambda^\top)^\top\right)\\
&\overset{(a)}{=} \left(\mathsf{d}(X - F\Lambda^\top)\right)W(X - F\Lambda^\top)^\top + (X - F\Lambda^\top)W\left(\mathsf{d}(X - F\Lambda^\top)\right)^\top\\
&\overset{(b)}{=} \left[\left(-(\mathsf{d}F)\Lambda^\top - F(\mathsf{d}\Lambda)^\top\right)W(X - F\Lambda^\top)^\top\right] + (X - F\Lambda^\top)W\left(-(\mathsf{d}F)\Lambda^\top - F(\mathsf{d}\Lambda^\top)\right)^\top\\
&= -\left[(X - F\Lambda^\top)W^\top\Lambda(\mathsf{d}F)^\top + (X - F\Lambda^\top)W^\top(\mathsf{d}\Lambda)F^\top\right]^\top\\
&\qquad - \underbrace{\left[(X - F\Lambda^\top)W\Lambda(\mathsf{d}F)^\top + (X - F\Lambda^\top)W(\mathsf{d}\Lambda)F^\top\right]}_{=A_W}\\
&= -A_W^\top - A_W,
\end{aligned} \tag{14}$$

where $A_W := (X - F\Lambda^\top)W\Lambda(\mathsf{d}F)^\top + (X - F\Lambda^\top)W(\mathsf{d}\Lambda)F^\top$. Here, (a) holds true since $\mathsf{d}(UWV) = (\mathsf{d}U)WV + U(\mathsf{d}W)V + UW(\mathsf{d}V)$ and $\mathsf{d}W = 0$ for arbitrary variables U, V and an arbitrary constant W, and (b) holds since

$$\mathsf{d}(X - F\Lambda^\top) = -(\mathsf{d}F)\Lambda^\top - F(\mathsf{d}\Lambda^\top) = -(\mathsf{d}F)\Lambda^\top - F(\mathsf{d}\Lambda)^\top.$$

The last two are just rearrangement that uses the assumption that W is symmetric.

Using Eq. (14), we derive the differential of each term of $NT\phi$ in Eq. (13) as follows. The first term:

$$\begin{aligned}
\mathsf{d}\left(\left\|\tilde{X}-\tilde{F}\Lambda^\top\right\|_F^2\right) &= \mathsf{d}tr\left(M_1(X-F\Lambda^\top)(X-F\Lambda^\top)^\top M_1\right)\\
&\overset{(a)}{=} tr\left(M_1\mathsf{d}\left((X-F\Lambda^\top)(X-F\Lambda^\top)^\top\right)M_1\right)\\
&\overset{(b)}{=} -tr\left(M_1\left(A_{I_N}^\top + A_{I_N}\right)M_1\right)\\
&= -2tr\left(M_1 A_{I_N} M_1\right)\\
&= -2tr\left(M_1\left((X-F\Lambda^\top)\Lambda(\mathsf{d}F)^\top + (X-F\Lambda^\top)(\mathsf{d}\Lambda)F^\top\right)M_1\right)\\
&= -2tr\left(M_1(X-F\Lambda^\top)\Lambda(\mathsf{d}F)^\top M_1\right) - 2tr\left(M_1(X-F\Lambda^\top)(\mathsf{d}\Lambda)F^\top M_1\right)\\
&= -2tr\left(M_1(X-F\Lambda^\top)\Lambda(\mathsf{d}F)^\top\right) - 2tr\left(F^\top M_1(X-F\Lambda^\top)(\mathsf{d}\Lambda)\right).
\end{aligned} \tag{15}$$

Here, (a) holds since $\mathsf{d}tr(AXB) = tr(A(\mathsf{d}X)B)$ for an arbitrary variable X and constants A and B, (b) is due to Eq. (14), and other equalities are clear from properties of the trace operator, namely, $tr(A) = tr(A^\top)$ and $tr(AB) = tr(BA)$ for matrices A, B with appropriate order and the fact that M_1 is a symmetric and idempotent matrix.

The differential of the second term of Eq. (13) is derived similarly to the derivation for the first term:

$$\begin{aligned}
\mathsf{d}\left(T\left\|\bar{X}-\Lambda\bar{F}\right\|_V^2\right) &= \mathsf{d}tr\left(P_1(X-F\Lambda^\top)V(X-F\Lambda^\top)^\top P_1\right)\\
&= tr\left(P_1\mathsf{d}\left((X-F\Lambda^\top)V(X-F\Lambda^\top)^\top\right)P_1\right)\\
&= -tr\left(P_1\left(A_V^\top + A_V\right)P_1\right)\\
&= -2tr\left(P_1 A_V P_1\right)\\
&= -2tr\left(P_1\left((X-F\Lambda^\top)V\Lambda(\mathsf{d}F)^\top + (X-F\Lambda^\top)V(\mathsf{d}\Lambda)F^\top\right)P_1\right)\\
&= -2tr\left(P_1(X-F\Lambda^\top)V\Lambda(\mathsf{d}F)^\top P_1\right) - 2tr\left(P_1(X-F\Lambda^\top)V(\mathsf{d}\Lambda)F^\top P_1\right)\\
&= -2tr\left(P_1(X-F\Lambda^\top)V\Lambda(\mathsf{d}F)^\top\right) - 2tr\left(F^\top P_1(X-F\Lambda^\top)V(\mathsf{d}\Lambda)\right).
\end{aligned} \tag{16}$$

Combining Eqs. (15) and (16), we obtain the differential of $NT\phi$ as follows.

$$\begin{aligned}
\mathsf{d}(NT\phi(\Lambda,F)) = &-2tr\left(M_1(X-F\Lambda^\top)\Lambda(\mathsf{d}F)^\top\right) - 2tr\left(F^\top M_1(X-F\Lambda^\top)(\mathsf{d}\Lambda)\right)\\
&- 2\eta\, tr\left(P_1(X-F\Lambda^\top)V\Lambda(\mathsf{d}F)^\top\right) - 2\eta\, tr\left(F^\top P_1(X-F\Lambda^\top)V(\mathsf{d}\Lambda)\right).
\end{aligned} \tag{17}$$

The first-order optimality condition underlying the update step for pricing factors Based on the differential of $NT\phi$ in Eq. (17), we can write the first-order optimality condition for F, i.e., $\nabla_F\phi(\Lambda,F) = 0$, as follows.

$$\begin{aligned}
& M_1(X-F\Lambda^\top)\Lambda + \eta P_1(X-F\Lambda^\top)V\Lambda = 0\\
\Leftrightarrow\quad & M_1X\Lambda + \eta P_1XV\Lambda = M_1F\Lambda^\top\Lambda + \eta P_1F\Lambda^\top V\Lambda\\
\Leftrightarrow\quad & \left[X + P_1X(\eta V - I_N)\right]\Lambda = \left[F\Lambda^\top + P_1F\Lambda^\top(\eta V - I_N)\right]\Lambda\\
\Leftrightarrow\quad & \left[\Lambda^\top\otimes M_1 + \eta(\Lambda^\top V)\otimes P_1\right]\mathrm{vec}(X) = \left[\Lambda^\top\otimes M_1 + \eta(\Lambda^\top V\otimes P_1)\right]\mathrm{vec}(F\Lambda^\top)\\
& \qquad = \left[\Lambda^\top\otimes M_1 + \eta(\Lambda^\top V\otimes P_1)\right](\Lambda\otimes I_T)\mathrm{vec}(F)\\
\Leftrightarrow\quad & (\Lambda^\top\otimes I_T)\left[I_N\otimes M_1 + \eta(V\otimes P_1)\right]\mathrm{vec}(X)\\
& \qquad = (\Lambda^\top\otimes I_T)\left[I_N\otimes M_1 + \eta(V\otimes P_1)\right](\Lambda\otimes I_T)\mathrm{vec}(F),
\end{aligned} \tag{18}$$

where the first three lines are simple reformulations from Eq. (17), and the last four lines are due to the assumption that V is a symmetric matrix and the fact that $\mathrm{vec}(ABC) = (C^\top\otimes A)\mathrm{vec}(B)$ for matrices A, B and C of proper orders.

The optimality condition for F when $V = I_N$ and $\eta > 0$ For the case when $V = I_N$ and $\eta > 0$, we have that the relation

$$F = X\Lambda^\top(\Lambda^\top\Lambda)^{-1}$$

derived from the first-order optimality condition that allows substituting for F. Indeed, substituting $V = I_N$ to the second line in Eq. (18) leads to

$$\begin{aligned} & M_1 F\Lambda^\top\Lambda + \eta P_1 F\Lambda^\top\Lambda = M_1 X\Lambda + \eta P_1 X\Lambda \\ \Leftrightarrow\quad & (M_1 + \eta P_1)F\Lambda^\top\Lambda = (M_1 + \eta P_1)X\Lambda \\ \Leftrightarrow\quad & F = X\Lambda(\Lambda^\top\Lambda)^{-1} \end{aligned} \tag{19}$$

where second line is a simple reformulation of the equation and the last line is satisfied due to the assumption that $\eta > 0$ and Lemma A.1.

The first-order optimality condition underlying the update step for factor loadings Similarly to the way of deriving the optimality condition for pricing factors, the first-order optimality condition for Λ, i.e., $\nabla_\Lambda \phi(\Lambda, F) = 0$, is given by

$$\begin{aligned} & F^\top M_1(X - F\Lambda^\top) + \eta F^\top P_1(X - F\Lambda^\top)V = 0 \\ \Leftrightarrow & F^\top M_1 X + \eta F^\top P_1 X V = F^\top M_1 F\Lambda^\top + \eta F^\top P_1 F\Lambda^\top V \\ \Leftrightarrow & F^\top\left[X + P_1 X(\eta V - I_N)\right] = F^\top\left[F\Lambda^\top + P_1 F\Lambda^\top(\eta V - I_N)\right] \\ \Leftrightarrow & \left[I_N \otimes F^\top M_1 + \eta(V \otimes F^\top P_1)\right]\mathrm{vec(X)} = \left[I_N \otimes F^\top M_1 + \eta(V \otimes F^\top P_1)\right]\mathrm{vec}(F\Lambda^\top) \\ & \qquad = \left[I_N \otimes F^\top M_1 + \eta(V \otimes F^\top P_1)\right](I_N \otimes F)\mathrm{vec}(\Lambda^\top) \\ \Leftrightarrow & (I_N \otimes F^\top)\left[I_N \otimes M_1 + \eta(V \otimes P_1)\right]\mathrm{vec(X)} \\ & \qquad = (I_N \otimes F^\top)\left[I_N \otimes M_1 + \eta(V \otimes P_1)\right](I_N \otimes F)\mathrm{vec}(\Lambda^\top). \end{aligned} \tag{20}$$

C Proof of Proposition 4.1

The goal of this section is to prove Proposition 4.1. To this end, we first prove the following lemma.

Lemma C.1. *Assume that $V \in \mathbb{R}^{N\times N}$ is symmetric and has the eigen-decomposition $V = \tilde{U}\tilde{L}\tilde{U}^\top$ where $\tilde{U} \in \mathbb{R}^{N\times r'}$, $\tilde{L} = diag(\lambda_1, \cdots, \lambda_{r'}) \in \mathbb{R}^{r'\times r'}$ with $\lambda_1, \cdots, \lambda_{r'} \neq 0$, and $rank(V) = r' \leq N$. Let η be any real number. Then, the following matrix diagonalization is true.*

$$\begin{aligned} P &:= I_N \otimes M_1 + \eta(V \otimes P_1) \\ &= (U \otimes U_1)D(U^\top \otimes U_1^\top). \end{aligned} \tag{21}$$

Here, D is an $NT \times NT$ diagonal matrix consisting of entries of $0, 1,$ and $\eta\lambda_i$ on its diagonal. Specifically, the number of 0s is $N - r'$, the number of 1s is $N(T-1)$, and the number of $\eta\lambda_i$ is r'. $U = [\tilde{U} : \tilde{U}^\perp]$ for some matrix $\tilde{U}^\perp \in \mathbb{R}^{N\times(N-r')}$ such that $U^\top U = I_N$ and $U_1 = \left[\frac{1}{\sqrt{T}}\mathbb{1}_T : \tilde{U}_1^\perp\right]$ for some matrix $\tilde{U}_1^\perp \in \mathbb{R}^{T\times(T-1)}$ such that $U_1^\top U_1 = I_T$.

Proof. Without loss of generality, we assume that $r' < N$. We can choose an orthogonal basis of the null space of V to construct $\tilde{U}^\perp$. Then we have that $U^\top U = I_N$. In the same way, from the fact that $\frac{1}{\sqrt{T}}\mathbb{1}_T$ is the eigenvector of the rank-1 matrix P_1 corresponding to the eigenvalue of 1, we can find $\tilde{U}_1^\perp$ that satisfies $U_1^\top U_1 = I_T$. Then, the following equalities are satisfied.

$$V = ULU^\top, \quad P_1 = U_1 L_1 U_1^\top, \quad UU^\top = U^\top U = I_N, \quad U_1 U_1^\top = U_1^\top U_1 = I_T,$$

where $L = diag(\lambda_1, \cdots, \lambda_{r'}, 0, \cdots, 0) \in \mathbb{R}^{N\times N}$, $\lambda_1, \cdots, \lambda_{r'} > 0$, and $L_1 = diag(1, 0, \cdots, 0) \in \mathbb{R}^{T\times T}$. Then, we have that

$$\begin{aligned}
P &= I_N \otimes M_1 + \eta(V \otimes P_1) \\
&= I_N \otimes I_T - I_N \otimes P_1 + \eta(V \otimes P_1) \\
&= I_N \otimes I_T + (\eta V - I_N) \otimes P_1 \\
&= (UU^\top) \otimes (U_1 U_1^\top) + (\eta U L U^\top - UU^\top) \otimes (U_1 L_1 U_1^\top) \\
&= (U \otimes U_1)(U^\top \otimes U_1^\top) + (U(\eta L - I_N)U^\top) \otimes (U_1 L_1 U_1^\top) \\
&= (U \otimes U_1)(I_N \otimes I_T)(U^\top \otimes U_1^\top) + (U \otimes U_1)((\eta L - I_N) \otimes L_1)(U^\top \otimes U_1^\top) \\
&= (U \otimes U_1)\left[(I_N \otimes I_T) + ((\eta L - I_N) \otimes L_1)\right](U^\top \otimes U_1^\top).
\end{aligned}$$

Now, define an $NT \times NT$ diagonal matrix D as

$$D = (I_N \otimes I_T) + ((\eta L - I_N) \otimes L_1). \tag{22}$$

It is diagonal since a Kronecker product of diagonal matrices is diagonal and sum of diagonal matrices is diagonal. It has N blocks of $T \times T$ diagonal matrices on its diagonal, and the i-th block, for $i \in \{1, \cdots, N\}$, is

$$\begin{aligned}
I_T + (\eta\lambda_i - 1)L_1 &= diag(\eta\lambda_i, 1, \cdots, 1) \qquad && \text{if } i \le r', \\
I_T - L_1 &= diag(0, 1, \cdots, 1) && \text{otherwise.}
\end{aligned}$$

□

Proposition 4.1 which is given below is proved using Lemmas A.4 and C.1.

Proposition C.2 (Proposition 4.1 in the main text)**.** *Suppose that $V \in \mathbb{S}_+^N$. Then, there exist solutions to equations (9) and (10). If it is additionally assumed that V is positive-definite, $\eta > 0$ and Λ_* and F_* have full column rank, i.e., $rank(\Lambda_*) = rank(F_*) = K$, then the solutions are unique.*

Proof. Eqs. (9) and (10) can be rewritten as

$$\begin{aligned}
A^\top P \,\mathrm{vec}(X) &= A^\top P A \,\mathrm{vec}(F), \\
B^\top P \,\mathrm{vec}(X) &= B^\top P B \,\mathrm{vec}(\Lambda^\top)
\end{aligned} \tag{23}$$

where P is defined in Eq. (21), $A := \Lambda_* \otimes I_T$ and $B := I_N \otimes F_*$. Clearly, P is a symmetric matrix. Using Lemma A.4, we have that $A^\top P \,\mathrm{vec}(X) \in \mathrm{colsp}(A^\top P) = \mathrm{colsp}(A^\top P A)$ and $B^\top P \,\mathrm{vec}(X) \in \mathrm{colsp}(B^\top P) = \mathrm{colsp}(B^\top P B)$, which, in turn, implies that solutions to Eq. (23) exist.

Let us additionally assume that V is positive-definite, $\eta > 0$ and Λ_* and F_* have full column rank. Note that, if V is a symmetric positive-definite matrix and $\eta > 0$, then Eq. (21) in Lemma C.1 is the eigen-decomposition of P whose eigenvalue is either 1 or $\sqrt{\eta\lambda_i}$, both of which are positive. This implies that P is positive-definite. Furthermore, if Λ_* and F_* have full column rank, then so are A and B. This implies that $A^\top P A \in \mathbb{R}^{KT\times KT}$ and $B^\top P B \in \mathbb{R}^{KT\times KT}$ have full rank, *i.e.*, they are invertible, implying that the solutions are unique. □

Furthermore, Lemma C.1 implies the following corollary that is not used in the paper, but might be useful for sanity checks when implementing Algorithm 1.

Corollary C.3. *Assume $V \in \mathbb{R}^{N\times N}$ is a symmetric matrix and $rank(V) = r' \le N$. Then, the rank of the matrix $P \in \mathbb{R}^{NT\times NT}$ defined in Eq. (21) satisfies*

$$rank(P) = \begin{cases} N(T-1), & \text{if } \eta = 0. \\ N(T-1) + r', & \text{if } \eta \ne 0, \end{cases} \tag{24}$$

Furthermore, P is non-singular if and only if V is non-singular and $\eta \ne 0$.

Proof. By counting the number of nonzero entries on the diagonal of D in Eq. (22), we can see that the equality in Eq. (24) is true.

Next, suppose that P is non-singular. Assume, to arrive at a contradiction, that V is singular or $\eta = 0$. First, suppose that V is singular. $\Rightarrow r' < N$. Then, we have that $rank(P) \leq max\{N(T-1), N(T-1)+r'\} = N(T-1)+r' = NT-(N-r') < NT$, which contradicts that P is non-singular. Second, suppose that $\eta = 0$. $\Rightarrow rank(P) = N(T-1) < NT$, which contradicts that P is non-singular. Conversely, suppose that V is non-singular and $\eta \neq 0$. Then, we have that $r' = N$, implying $rank(P) = N(T-1)+r' = NT$. Thus, P is non-singular. $\square$

D NUMERICAL PROPERTIES FOR LARGER DATA SETS

We present convergence results of the algorithm for $K = 4$ and $\eta = 10$ when applied to larger real-world data sets in Figures 6 and 7. We consider two cross-sectional dimension $N \in \{25, 370\}$ and three time-series dimension $T \in \{60, 240, 600\}$. Figure 6 illustrates the results for the data set with $N = 25$, which consists of the 5x5 Size-B/M portfolios. In Figure 7, we present the results for the data set with $N = 370$, consisting of portfolios formed from sorts on 37 anomalies widely utilized in the finance literature, *e.g.*, (Kelly et al., 2019; Lettau & Pelger, 2020a). Our findings demonstrate that the algorithm performs effectively across a range of realistic scenarios as shown for the smaller data set in the main text.

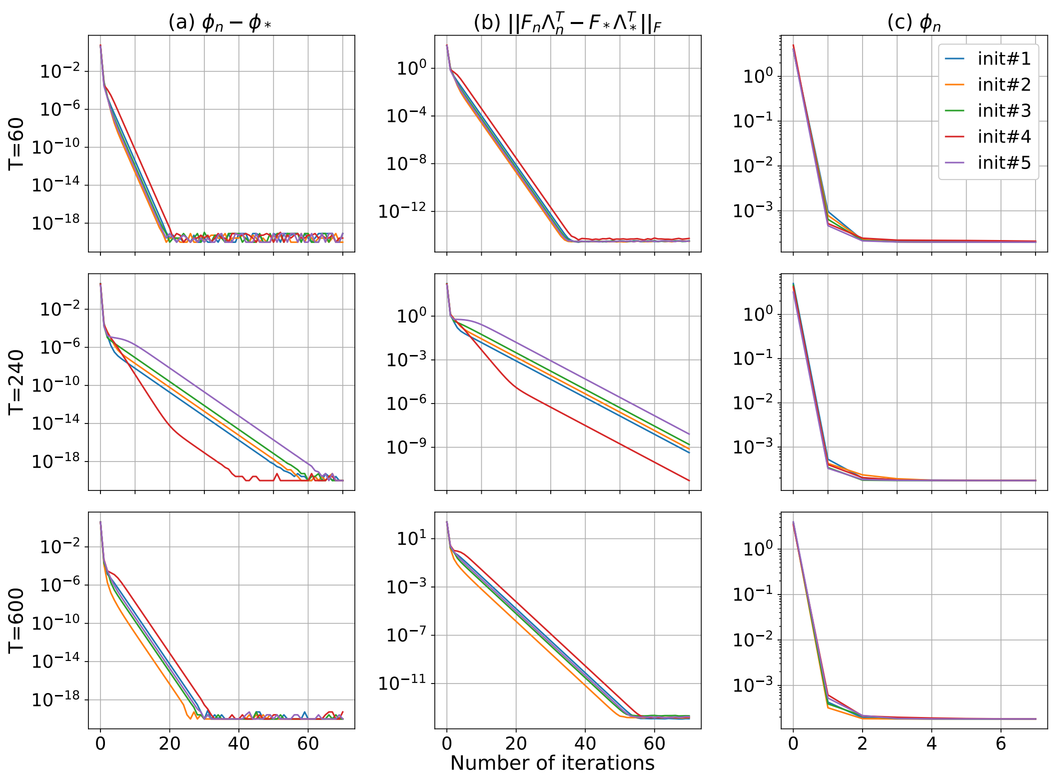

Figure 6: Convergence of Algorithm 1 when $N = 25$. Each curve represents one random initialization for (F_0, Λ_0).

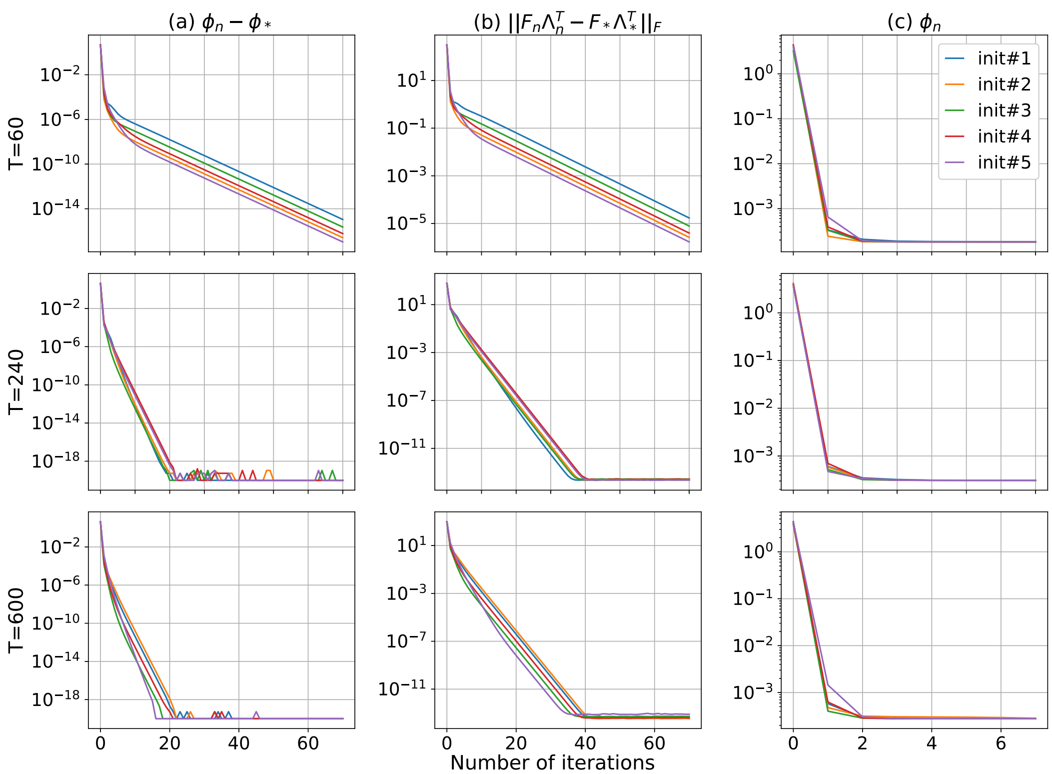

Figure 7: Convergence of Algorithm 1 when $N = 370$. Each curve represents one random initialization for (F_0, Λ_0).