

Factor Models: Identification and Consistency

Sangmyung Ha

January 2025

1 Factor Models

Consider the factor model

$$x_{ti} = c_{ti} + e_{ti} = f_t' \alpha_i + e_{ti}, \quad i = 1, \dots, N, \quad t = 1, \dots, T,$$

where $c_{ti} = f_t' \alpha_i$ is the common component, $f_t = (f_{t1}, \dots, f_{tK})'$ is a K -dimensional vector of factors, $\alpha_i = (\alpha_{i1}, \dots, \alpha_{iK})'$ is a K -dimensional vector of factor loadings, and e_{ti} is the idiosyncratic error. The loadings may be fixed or random. Throughout, K is fixed and known, and both N and T tend to infinity. We write $\|v\|$ for the Euclidean norm of a vector and $\|B\|_F$ and $\|B\|_{\text{op}}$ for the Frobenius and operator norms of a matrix, respectively.

In matrix form, the model is

$$X = C + E = FA' + E,$$

where $X = (x_{ti})$ and $E = (e_{ti})$ are $T \times N$ matrices, $F = (f_1, \dots, f_T)'$ is a $T \times K$ matrix, and $A = (\alpha_1, \dots, \alpha_N)'$ is an $N \times K$ matrix. We first consider conditions under which the common component can be asymptotically separated from the idiosyncratic component, and then study estimation of the factors and loadings up to a nonsingular transformation.

1.1 Identification of Common Component

Suppose that

$$Q_F := \frac{F'F}{T} = \frac{1}{T} \sum_{t=1}^T f_t f_t' \xrightarrow{p} M_F > 0, \quad (1)$$

$$Q_A := \frac{A'A}{N} = \frac{1}{N} \sum_{i=1}^N \alpha_i \alpha_i' \xrightarrow{p} M_A > 0, \quad (2)$$

where M_F and M_A are finite, nonrandom, positive-definite matrices. Consider the singular value decompositions

$$\begin{aligned} C &= U^{(c)} D^{(c)} V^{(c)'}, \\ E &= U^{(id)} D^{(id)} V^{(id)'}, \end{aligned}$$

where the columns of the left and right singular-vector matrices are orthonormal. The matrices $U^{(c)}$ and $V^{(c)}$ have K columns, while $U^{(id)}$ and $V^{(id)}$ have $L = \min(N, T)$ columns. Write $\delta_j^{(c)}$ and $\delta_j^{(id)}$ for the singular values of C and E , respectively, in decreasing order.

The total variation in the common component satisfies

$$\frac{\|C\|_F^2}{NT} = \frac{\text{tr}(A'AF'F)}{NT} = \text{tr}(Q_A Q_F) \xrightarrow{p} \text{tr}(M_A M_F) > 0.$$

More importantly, every one of its K signal singular values has order $\sqrt{NT}$ with a strictly positive normalized limit. To see this, let $\lambda_j(B)$ denote the j th largest eigenvalue of a symmetric matrix B . The K largest eigenvalues of $CC'/(NT)$ equal those of $Q_A^{1/2} Q_F Q_A^{1/2}$, so

$$\frac{(\delta_j^{(c)})^2}{NT} = \lambda_j(Q_A^{1/2} Q_F Q_A^{1/2}) \xrightarrow{p} \lambda_j(M_A^{1/2} M_F M_A^{1/2}) =: \lambda_j^* > 0 \quad (3)$$

for $j = 1, \dots, K$. In particular, $\|C\|_{\text{op}} = O_p(\sqrt{NT})$ and $\delta_K^{(c)}/\sqrt{NT}$ is bounded away from zero in probability.

The idiosyncratic component may have total variation of the same order. For example, the condition

$$\sup_{N,T} \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N \mathbb{E}(e_{ti}^2) < \infty$$

implies $\|E\|_F^2 = O_p(NT)$ by Markov's inequality. If, in addition, $(NT)^{-1} \sum_{t=1}^T \sum_{i=1}^N e_{ti}^2 \xrightarrow{p} \sigma_e^2 \in (0, \infty)$, then the common and idiosyncratic components have comparable total variation. Neither statement, however, guarantees that the largest singular value of E is small. For example, if $E = z\iota'_N$ with $z \in \mathbb{R}^T$, $z_t \in \{-1, 1\}$, then $\|E\|_F^2 = NT$ but $\|E\|_{\text{op}} = \sqrt{NT}$.

We therefore impose the separate restriction

$$\|E\|_{\text{op}} = \max_{1 \leq j \leq L} \delta_j^{(id)} = O_p\left(\sqrt{\max(N, T)}\right). \quad (4)$$

It follows that

$$\frac{\|E\|_{\text{op}}}{\sqrt{NT}} = O_p\left(\frac{1}{\sqrt{\min(N, T)}}\right) = o_p(1).$$

Thus, the largest idiosyncratic singular value is asymptotically negligible relative to the smallest signal singular value. The distinction between the components is not necessarily their total variation, but the concentration of the common variation in a fixed number of large singular values.

Sufficient conditions on the errors. If the entries of E are independent, have mean zero, and satisfy

$$\sup_{N,T} \max_{1 \leq t \leq T, 1 \leq i \leq N} \mathbb{E}|e_{ti}|^4 < \infty,$$

then Latała's inequality gives¹

$$\mathbb{E}\|E\|_{\text{op}} = O\left(\sqrt{N} + \sqrt{T} + (NT)^{1/4}\right) = O(\sqrt{N} + \sqrt{T}).$$

Markov's inequality then establishes (4). The fourth-moment bound must hold uniformly as the panel dimensions increase; finiteness of each individual fourth moment is not sufficient.

¹Latała, R. (2005), "Some estimates of norms of random matrices," *Proceedings of the American Mathematical Society*, 133, 1273–1282.

Independence can be relaxed. As one sufficient example, suppose that

$$E = \Gamma_T^{1/2} U \Sigma_N^{1/2},$$

where Γ_T and Σ_N are deterministic, symmetric, positive-semidefinite matrices, and the entries of U are independent with mean zero, variance one, and uniformly bounded fourth moments. Use symmetric positive-semidefinite square roots and assume

$$\sup_T \|\Gamma_T\|_{\text{op}} < \infty, \quad \sup_N \|\Sigma_N\|_{\text{op}} < \infty.$$

Then

$$\|E\|_{\text{op}} \leq \|\Gamma_T\|_{\text{op}}^{1/2} \|U\|_{\text{op}} \|\Sigma_N\|_{\text{op}}^{1/2} = O_p\left(\sqrt{\max(N, T)}\right).$$

In this construction,

$$\text{Cov}(e_{ti}, e_{sj}) = (\Gamma_T)_{ts} (\Sigma_N)_{ij}.$$

Hence, (4) permits both serial and cross-sectional dependence.

Non-pervasiveness. Condition (4) rules out a fixed-dimensional strong-factor representation of E itself. More precisely, E cannot admit a representation

$$e_{ti} = \tilde{f}'_t \tilde{\alpha}_i \tag{5}$$

with a fixed positive number of factors for which

$$\frac{\tilde{F}'\tilde{F}}{T} \xrightarrow{p} \widetilde{M}_F > 0, \quad \frac{\tilde{A}'\tilde{A}}{N} \xrightarrow{p} \widetilde{M}_A > 0.$$

Indeed, the argument in (3) would imply that $\|E\|_{\text{op}}/\sqrt{NT}$ is bounded away from zero in probability, contradicting (4).

The identification discussed here is asymptotic spectral separation and recovery of C . It is not exact uniqueness of the decomposition $X = C + E$ in a finite sample. A precise average consistency result for the estimated common component is established below.

1.2 Identification of the Factors

Even when the common component is recoverable, the factors and loadings are not separately identified without a normalization. For any nonsingular $K \times K$ matrix H ,

$$FA' = (FH)(AH^{-1})', \quad H^{-1} := (H')^{-1}.$$

Thus, replacing f_t by $H'f_t$ and α_i by $H^{-1}\alpha_i$ leaves c_{ti} unchanged.

When the limiting signal eigenvalues are distinct, a common asymptotic normalization is

$$\frac{F'F}{T} \xrightarrow{p} I_K, \quad \frac{A'A}{N} \xrightarrow{p} D_A, \tag{6}$$

where $D_A = \text{diag}(d_1, \dots, d_K)$ and $d_1 > \dots > d_K > 0$. These restrictions fix asymptotic scale and ordering, but simultaneous sign changes of corresponding factor and loading columns remain possible. A sign convention is needed to select a uniquely oriented representation. If limiting eigenvalues are repeated, additional orthogonal rotations within the corresponding eigenspaces also remain possible.

The limiting restrictions in (6) should be distinguished from exact finite-sample normalizations. The principal-component estimators below satisfy exact sample restrictions. The consistency results, however, are stated relative to a sample-dependent nonsingular transformation and require neither (6) nor distinct limiting signal eigenvalues.

1.3 Estimation

Write the singular value decomposition of X as

$$X = UDV' = U_{1:K}D_{1:K}V'_{1:K} + U_{K+1:L}D_{K+1:L}V'_{K+1:L},$$

with singular values in decreasing order. The principal-component estimators are

$$\hat{F} = \sqrt{T}U_{1:K}, \quad \hat{A} = \frac{1}{\sqrt{T}}V_{1:K}D_{1:K} = \frac{X'\hat{F}}{T}.$$

Their rows are denoted by $\hat{f}'_t$ and $\hat{\alpha}'_i$. Define

$$\hat{C} = \hat{F}\hat{A}' = U_{1:K}D_{1:K}V'_{1:K}, \quad \hat{E} = X - \hat{C}.$$

The truncated singular value decomposition is a best rank- K approximation to X in the Frobenius norm. Equivalently, $(\hat{F}, \hat{A})$ solves the least-squares problem

$$\min_{\tilde{F}, \tilde{A}: \tilde{F}'\tilde{F}/T = I_K} \frac{1}{NT} \|X - \tilde{F}\tilde{A}'\|_F^2.$$

The singular value decomposition selects a solution with $\hat{A}'\hat{A}$ diagonal and ordered. The estimators satisfy

$$\frac{\hat{F}'\hat{F}}{T} = I_K, \quad \frac{\hat{A}'\hat{A}}{N} = \frac{D_{1:K}^2}{NT}, \quad \hat{F}'\hat{E} = 0, \quad \hat{E}\hat{A} = 0.$$

The residual orthogonality conditions are sample identities. They do not imply corresponding assumptions on the true factors, loadings, and errors.

For the sharper factor and loading rates derived below, impose the following additional averaged-score conditions:

$$\frac{1}{N} \sum_{i=1}^N \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^T f_t e_{ti} \right\|^2 = \frac{\|E'F\|_F^2}{NT} = O_p(1), \quad (7)$$

$$\frac{1}{T} \sum_{t=1}^T \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N \alpha_i e_{ti} \right\|^2 = \frac{\|EA\|_F^2}{NT} = O_p(1). \quad (8)$$

These assumptions directly control averages over increasing numbers of units or periods. Pointwise statements such as $N^{-1} \sum_i \alpha_i e_{ti} = O_p(N^{-1/2})$ for each fixed t do not, by themselves, justify averaging the squared scores over a growing number of periods.

For example, sufficient conditions for (7) and (8) are

$$\sup_{N,T} \max_{1 \leq i \leq N} \mathbb{E} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^T f_t e_{ti} \right\|^2 < \infty,$$

$$\sup_{N,T} \max_{1 \leq t \leq T} \mathbb{E} \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N \alpha_i e_{ti} \right\|^2 < \infty.$$

Markov's inequality then establishes the required averaged bounds.

The separable-error construction above also provides a sufficient example when U is independent of (F, A) . In that case,

$$\begin{aligned}\mathbb{E}(\|EA\|_F^2 \mid F, A) &= \text{tr}(\Gamma_T) \text{tr}(A' \Sigma_N A) \leq T \|\Gamma_T\|_{\text{op}} \|\Sigma_N\|_{\text{op}} \|A\|_F^2, \\ \mathbb{E}(\|E'F\|_F^2 \mid F, A) &= \text{tr}(\Sigma_N) \text{tr}(F' \Gamma_T F) \leq N \|\Sigma_N\|_{\text{op}} \|\Gamma_T\|_{\text{op}} \|F\|_F^2.\end{aligned}$$

Conditions (1) and (2), together with conditional Markov inequalities, imply (7) and (8). Independence of U from (F, A) is an additional sufficient condition for this example; it is not needed when (7) and (8) are imposed directly.

1.4 Theory: Consistency

Suppose that (1), (2), (4), (7), and (8) hold. We establish average consistency of the factors and loadings up to a compatible transformation, without imposing a restriction on the relative growth rates of N and T .²

Eigenvalue separation. Let

$$V_{NT} := \frac{D_{1:K}^2}{NT}.$$

Since $X = C + E$, submultiplicativity of the operator norm gives

$$\begin{aligned}\left\| \frac{XX'}{NT} - \frac{CC'}{NT} \right\|_{\text{op}} &\leq \frac{2\|C\|_{\text{op}}\|E\|_{\text{op}} + \|E\|_{\text{op}}^2}{NT} \\ &= O_p\left(\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{T}}\right) = o_p(1).\end{aligned}$$

By the eigenvalue perturbation inequality and (3),

$$V_{NT} \xrightarrow{p} \text{diag}(\lambda_1^*, \dots, \lambda_K^*), \quad \lambda_K^* > 0.$$

Consequently,

$$\|V_{NT}\|_{\text{op}} + \|V_{NT}^{-1}\|_{\text{op}} = O_p(1). \tag{9}$$

All inverses below are understood on events whose probability tends to one.

Factor estimation. Define

$$B_{NT} := \frac{F' \hat{F}}{T}, \quad H_{NT} := Q_A B_{NT} V_{NT}^{-1}. \tag{10}$$

Because $\|F\|_{\text{op}} = O_p(\sqrt{T})$ and $\|\hat{F}\|_{\text{op}} = \sqrt{T}$, both B_{NT} and H_{NT} are bounded in probability. The principal-component eigenvalue equation is

$$\frac{XX'}{NT} \hat{F} = \hat{F} V_{NT}.$$

²For a related matrix-norm approach to principal-component estimation, see Bai, J., and S. Ng (2023), ‘‘Approximate factor models with weaker loadings.’’ The strong-factor case corresponds to the normalized Gram-matrix conditions used here.

Substituting $X = FA' + E$ and rearranging yields the exact identity

$$\Delta_F := \hat{F} - FH_{NT} = \left(\frac{FA'E'\hat{F}}{NT} + \frac{EAF'\hat{F}}{NT} + \frac{EE'\hat{F}}{NT} \right) V_{NT}^{-1}. \quad (11)$$

Condition (8) gives $\|EA\|_F = O_p(\sqrt{NT})$. Hence,

$$\begin{aligned} \frac{1}{\sqrt{T}} \left\| \frac{FA'E'\hat{F}}{NT} \right\|_F &\leq \frac{\|F\|_{\text{op}} \|EA\|_F \|\hat{F}\|_{\text{op}}}{NT^{3/2}} = O_p(N^{-1/2}), \\ \frac{1}{\sqrt{T}} \left\| \frac{EAF'\hat{F}}{NT} \right\|_F &\leq \frac{\|EA\|_F \|F\|_{\text{op}} \|\hat{F}\|_{\text{op}}}{NT^{3/2}} = O_p(N^{-1/2}), \\ \frac{1}{\sqrt{T}} \left\| \frac{EE'\hat{F}}{NT} \right\|_F &\leq \frac{\|E\|_{\text{op}}^2 \|\hat{F}\|_F}{NT^{3/2}} = O_p\left(\frac{1}{N} + \frac{1}{T}\right), \end{aligned}$$

where the last line uses (4) and $\|\hat{F}\|_F = \sqrt{TK}$. Combining these inequalities with (9) gives

$$\frac{\|\Delta_F\|_F}{\sqrt{T}} = O_p\left(\frac{1}{\sqrt{N}} + \frac{1}{T}\right). \quad (12)$$

Therefore,

$$\frac{1}{T} \|\hat{F} - FH_{NT}\|_F^2 = \frac{1}{T} \sum_{t=1}^T \|\hat{f}_t - H'_{NT} f_t\|^2 = O_p\left(\frac{1}{N} + \frac{1}{T^2}\right). \quad (13)$$

This factor-rate argument uses (8), but does not require (7).

Nonsingularity of the transformation. Since $\|\Delta_F\|_F/\sqrt{T} = o_p(1)$ and $\hat{F}'\hat{F}/T = I_K$,

$$H'_{NT} Q_F H_{NT} = \frac{(\hat{F} - \Delta_F)'(\hat{F} - \Delta_F)}{T} = I_K + o_p(1).$$

Together with $Q_F \xrightarrow{P} M_F > 0$, this implies that H_{NT} is nonsingular with probability tending to one and that $\|H_{NT}^{-1}\|_{\text{op}} = O_p(1)$. Since $B_{NT} = Q_A^{-1} H_{NT} V_{NT}$, the matrix B_{NT} is also nonsingular with probability tending to one, and

$$\|B_{NT}^{-1}\|_{\text{op}} = O_p(1).$$

Loading estimation with the same transformation. Define

$$W_{NT} := \frac{E'\hat{F}}{T}.$$

The loading estimator satisfies

$$\hat{A} = \frac{X'\hat{F}}{T} = AB_{NT} + W_{NT}. \quad (14)$$

Using $\hat{F} = FH_{NT} + \Delta_F$, condition (7), and (12), we obtain

$$\begin{aligned} \frac{\|W_{NT}\|_F}{\sqrt{N}} &\leq \frac{\|E'F\|_F}{T\sqrt{N}} \|H_{NT}\|_{\text{op}} + \frac{\|E\|_{\text{op}}}{\sqrt{NT}} \frac{\|\Delta_F\|_F}{\sqrt{T}} \\ &= O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left[\left(\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{T}}\right) \left(\frac{1}{\sqrt{N}} + \frac{1}{T}\right)\right] \\ &= O_p\left(\frac{1}{\sqrt{T}} + \frac{1}{N}\right). \end{aligned}$$

To relate B_{NT} to the inverse transpose of H_{NT} , use $V_{NT} = \hat{A}'\hat{A}/N$ and define

$$R_{NT} := V_{NT} - B'_{NT}Q_AB_{NT} = \frac{B'_{NT}A'W_{NT} + W'_{NT}AB_{NT} + W'_{NT}W_{NT}}{N}.$$

Then

$$\begin{aligned} \|R_{NT}\|_{\text{op}} &\leq 2\|B_{NT}\|_{\text{op}} \frac{\|A\|_{\text{op}}}{\sqrt{N}} \frac{\|W_{NT}\|_F}{\sqrt{N}} + \frac{\|W_{NT}\|_F^2}{N} \\ &= O_p\left(\frac{1}{\sqrt{T}} + \frac{1}{N}\right). \end{aligned}$$

The definition of H_{NT} gives the exact relation

$$H_{NT}^{-'} - B_{NT} = Q_A^{-1}B_{NT}^{-'}(V_{NT} - B'_{NT}Q_AB_{NT}) = Q_A^{-1}B_{NT}^{-'}R_{NT}.$$

It follows that

$$\|H_{NT}^{-'} - B_{NT}\|_{\text{op}} = O_p\left(\frac{1}{\sqrt{T}} + \frac{1}{N}\right).$$

Consequently, (14) implies

$$\begin{aligned} \frac{1}{\sqrt{N}}\|\hat{A} - AH_{NT}^{-'}\|_F &\leq \frac{\|W_{NT}\|_F}{\sqrt{N}} + \frac{\|A\|_F}{\sqrt{N}}\|B_{NT} - H_{NT}^{-'}\|_{\text{op}} \\ &= O_p\left(\frac{1}{\sqrt{T}} + \frac{1}{N}\right). \end{aligned}$$

Therefore,

$$\frac{1}{N}\|\hat{A} - AH_{NT}^{-'}\|_F^2 = \frac{1}{N} \sum_{i=1}^N \|\hat{\alpha}_i - H_{NT}^{-1}\alpha_i\|^2 = O_p\left(\frac{1}{T} + \frac{1}{N^2}\right). \quad (15)$$

Thus, the factor and loading rates hold with compatible transformations H_{NT} and $H_{NT}^{-'}$.

Estimation of the common component. An average consistency bound for $\hat{C}$ follows directly from the truncated singular value decomposition and does not require (7) or (8). Since $\text{rank}(C) \leq K$, the best rank- K approximation property in operator norm gives

$$\|X - \hat{C}\|_{\text{op}} = \delta_{K+1}(X) \leq \|X - C\|_{\text{op}} = \|E\|_{\text{op}}.$$

Hence, $\|\hat{C} - C\|_{\text{op}} \leq 2\|E\|_{\text{op}}$. Since $\text{rank}(\hat{C} - C) \leq 2K$,

$$\frac{\|\hat{C} - C\|_F^2}{NT} \leq \frac{2K\|\hat{C} - C\|_{\text{op}}^2}{NT} \leq \frac{8K\|E\|_{\text{op}}^2}{NT} = O_p\left(\frac{1}{N} + \frac{1}{T}\right). \quad (16)$$

These are average squared-error results. By themselves, they do not establish consistency for every fixed entry or uniform consistency over all entries. Also, the rates in (13) and (15) are relative to the sample-dependent transformation; replacing that transformation by a fixed limiting matrix requires additional rate control.

1.5 Remarks

The error operator-norm condition can be weakened when only consistency is required. Suppose that (4) is replaced by

$$\|E\|_{\text{op}} = O_p(\sqrt{\delta_{NT}}), \quad \delta_{NT} > 0, \quad \frac{\delta_{NT}}{NT} \rightarrow 0, \quad (17)$$

where δ_{NT} is a deterministic sequence. Conditions (1) and (2) continue to guarantee signal singular values of order $\sqrt{NT}$, while (17) gives $\|E\|_{\text{op}} = o_p(\sqrt{NT})$. The eigenvalue separation and bounded inverse arguments therefore remain valid.

If the averaged loading-error score condition (8) is retained, the same expansion in (11) gives

$$\frac{1}{\sqrt{T}}\|\hat{F} - FH_{NT}\|_F = O_p\left(\frac{1}{\sqrt{N}} + \frac{\delta_{NT}}{NT}\right),$$

and hence

$$\frac{1}{T}\|\hat{F} - FH_{NT}\|_F^2 = O_p\left(\frac{1}{N} + \frac{\delta_{NT}^2}{N^2T^2}\right).$$

The N^{-1} term relies on (8); it does not follow from the weaker operator-norm restriction alone.

Without (8), the available bound is

$$\|EA\|_F \leq \|E\|_{\text{op}}\|A\|_F = O_p(\sqrt{N\delta_{NT}}).$$

Substituting this bound into (11) yields

$$\frac{1}{\sqrt{T}}\|\hat{F} - FH_{NT}\|_F = O_p\left(\sqrt{\frac{\delta_{NT}}{NT}} + \frac{\delta_{NT}}{NT}\right) = O_p\left(\sqrt{\frac{\delta_{NT}}{NT}}\right).$$

Therefore, under (1), (2), and (17) alone,

$$\frac{1}{T}\|\hat{F} - FH_{NT}\|_F^2 = O_p\left(\frac{\delta_{NT}}{NT}\right), \quad \frac{1}{NT}\|\hat{C} - C\|_F^2 = O_p\left(\frac{\delta_{NT}}{NT}\right).$$

Spectral separation is therefore sufficient for average consistency, whereas sharper factor and loading rates depend on additional restrictions on errors averaged against the true loadings and factors.