Asymptotic Theory for Factor Models

Seunghyun Kim

June 22, 2023

Asymptotics for the Baseline Model

Bai and Ng (2002), Bai (2003)

Suppose there are T time periods and N macroeconomic variables available in the sample, and denote the ith variable at time t by $\underbrace{X_{it}}_{(1\times1)}$. Let there exist $r < \min(N,T)$ common factors $\underbrace{F_t}_{(r\times1)}$ that determine these variables, with the loadings of each factor on the ith variable being given as $\lambda_i \in \mathbb{R}^{r\times1}$. The model is then given as

$$X_{it} = \lambda_i' F_t + e_{it},$$

where e_{it} is the idiosyncratic element of X_{it}, to whom we have yet to impose any time series properties.

Let $\underline{X}_i = (X_{i1}, \cdots, X_{iT})'$, $F^0 = (F_1^0, \cdots, F_T^0)'$, $\underline{e}_i = (e_{i1}, \cdots, e_{iT})'$, where the superscript 0 represents the true values. Then,

$$\underbrace{\underline{X}_i}_{(T\times1)} = \underbrace{F^0}_{(T\times r)} \underbrace{\lambda_i^0}_{(r\times1)} + \underbrace{\underline{e}_i}_{(T\times1)},$$

and collecting $X = (\underline{X}_1, \cdots, \underline{X}_N)$, $\Lambda^{0\prime} = (\lambda_1^0, \cdots, \lambda_N^0)$ and $e = (\underline{e}_1, \cdots, \underline{e}_N)$,

$$X = F^0 \Lambda^{0\prime} + e.$$

Alternatively, denoting $X_t = (X_{1t}, \cdots, X_{Nt})'$ and $e_t = (e_{1t}, \cdots, e_{Nt})'$, we can organize the data as

$$\underbrace{X_t}_{(N\times1)} = \underbrace{\Lambda^{0\prime}}_{(N\times r)} \underbrace{F_t^0}_{(r\times1)} + \underbrace{e_t}_{(N\times1)}.$$

1.1 Derivation of Estimators of Factors and their Loadings

The factor model defined above is estimated using the technique of asymptotic principal components, which just means that Λ and F are found as the solution to the following least squares minimization problem (we assume that k factors are estimated):

$$\min_{\Lambda,F} S(k,\Lambda,F) = \min_{\Lambda,F} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it}-\lambda_i' F_t)^2.$$

In other words, Λ and F are found as the values that minimize the mean squared deviation from the dependent variables in the given sample.

To simplify this problem, note that we can write

$$\begin{aligned} S(k,\Lambda,F) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it}-\lambda_i' F_t)^2 \\ &= \frac{1}{NT}\sum_{t=1}^{T}(X_t-\Lambda F_t)'(X_t-\Lambda F_t) \\ &= \frac{1}{NT}\operatorname{tr}\left((X-F\Lambda')(X-F\Lambda')'\right). \end{aligned}$$

The model can now be estimated in two different ways depending on whether we solve the first order condition for Λ or F first.

i) **Solving for Λ**

In this case, the first derivative of the above objective function for $\operatorname{vec}(\Lambda)$ is

$$\begin{aligned} \frac{\partial}{\partial \operatorname{vec}(\Lambda)'} S(k,\Lambda,F) &= \frac{\partial \operatorname{tr}\left((X'-\Lambda F')'(X'-\Lambda F')\right)}{\partial \operatorname{vec}\left((X'-\Lambda F')\right)'} \cdot \frac{\partial \operatorname{vec}\left((X'-\Lambda F')\right)}{\partial \operatorname{vec}(\Lambda)'} \\ &= -2\operatorname{vec}\left((X'-\Lambda F')\right)' \cdot (F \bigotimes I_N) = -2\operatorname{vec}\left((X'-\Lambda F')F\right)', \end{aligned}$$

where we used the fact that

$$\frac{\partial \operatorname{tr}(A'A)}{\partial A} = 2A$$

and

$$\begin{aligned} \frac{\partial \operatorname{vec}\left((X'-\Lambda F')\right)}{\partial \operatorname{vec}(\Lambda)'} &= -\frac{\partial \operatorname{vec}(\Lambda F')}{\partial \operatorname{vec}(\Lambda)} \\ &= -(F \bigotimes I_N) \cdot \frac{\partial \operatorname{vec}(\Lambda)}{\partial \operatorname{vec}(\Lambda)'} = -(F \bigotimes I_N). \end{aligned}$$

Therefore, the estimator of Λ given F is

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

Substituting this into the objective function yields the concentrated form

$$\begin{aligned} V(k,F) = S(k,\Lambda(F),F) &= \frac{1}{NT}\operatorname{tr}\left((X - F\Lambda(F)')(X - F\Lambda(F)')'\right) \\ &= \frac{1}{NT}\operatorname{tr}\left((X - F(F'F)^{-1}F'X)'(X - F(F'F)^{-1}F'X)\right) \\ &= \frac{1}{NT}\operatorname{tr}\left(X'\left(I_T - F(F'F)^{-1}F'\right)X\right). \end{aligned}$$

ii) **Solving for** F

In this case, the first derivative of the above objective function for $\operatorname{vec}(F)$ is

$$\begin{aligned} \frac{\partial}{\partial \operatorname{vec}(F)'} S(k,\Lambda,F) &= \frac{\partial \operatorname{tr}\left((X - F\Lambda')'(X - F\Lambda')\right)}{\partial \operatorname{vec}\left((X - F\Lambda')\right)'} \cdot \frac{\partial \operatorname{vec}\left((X - F\Lambda')\right)}{\partial \operatorname{vec}(F)'} \\ &= -2\operatorname{vec}\left((X - F\Lambda')\right)' \cdot (\Lambda \bigotimes I_N) = -2\operatorname{vec}\left((X - F\Lambda')\Lambda\right)'. \end{aligned}$$

Therefore, the estimator of F given Λ is

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

Substituting this into the objective function yields the concentrated form

$$\begin{aligned} \tilde{V}(k,\Lambda) = S(k,\Lambda,F(\Lambda)) &= \frac{1}{NT}\operatorname{tr}\left((X - F(\Lambda)\Lambda')(X - F(\Lambda)F')'\right) \\ &= \frac{1}{NT}\operatorname{tr}\left((X - X\Lambda(\Lambda'\Lambda)^{-1}\Lambda')(X - X\Lambda(\Lambda'\Lambda)^{-1}\Lambda')'\right) \\ &= \frac{1}{NT}\operatorname{tr}\left(X\left(I_N - \Lambda(\Lambda'\Lambda)^{-1}\Lambda'\right)X\right). \end{aligned}$$

It remains to derive the estimator of F and Λ under the above specifications. Let us first derive the estimator of F.

The problem is now to minimize the concentrated objective function

$$V(k,F) = \frac{1}{NT}\operatorname{tr}\left(X'\left(I_T - F(F'F)^{-1}F'\right)X\right) = \frac{1}{NT}\operatorname{tr}(X'X) - \frac{1}{NT}\operatorname{tr}\left(X'F(F'F)^{-1}F'X\right)$$
$$= \frac{1}{NT}\operatorname{tr}(X'X) - \frac{1}{NT}\operatorname{tr}\left((F'F)^{-1}F'XX'F\right)$$

with respect to F, which reduces to the problem of maximizing

$$\operatorname{tr}\left((F'F)^{-1}F'XX'F\right)$$

with respect to F. Imposing the normalization $\frac{F'F}{T} = I_k$ implies that we must find the F that solves the constrained maximization problem

$$\max_{F\in\mathbb{R}^{T\times k}} \quad \operatorname{tr}(F'(XX')F)$$
$$\text{subject to} \quad \frac{1}{T}F'F = I_k.$$

Let us now proceed step by step. To begin with, we find the upper bound of the expression $\operatorname{tr}(A'MA)$ for any $A \in \mathbb{R}^{T\times k}$ such that $A'A = I_k$ and positive semidefinite $M \in \mathbb{R}^{T\times T}$.

(1) **Expanding the expression of** $\operatorname{tr}(A'MA)$

Because M is symmetric, by the principal axis theorem there exist an orthogonal matrix $P \in \mathbb{R}^{T\times T}$ and a diagonal matrix $D = \operatorname{diag}(\mu_1,\cdots,\mu_T) \in \mathbb{R}^{T\times T}$ with $\mu_1 \geq \cdots \geq \mu_T \geq 0$ equal to the ordered eigenvalues of M such that $M = PDP'$. As such, defining $B = P'A$ and denoting the columns of B by $B_1,\cdots,B_k \in \mathbb{R}^T$,

$$\operatorname{tr}(A'MA) = \operatorname{tr}(B'DB) = \operatorname{tr}\left[\begin{pmatrix} B_1'DB_1 & \cdots & B_1'DB_k \\ \vdots & \ddots & \vdots \\ B_k'DB_1 & \cdots & B_k'DB_k \end{pmatrix}\right] = \sum_{j=1}^{k} B_j'DB_j$$
$$= \sum_{j=1}^{k}\sum_{i=1}^{T}\mu_i B_{ij}^2 = \sum_{i=1}^{T}\mu_i\left(\sum_{j=1}^{k}B_{ij}^2\right) = \sum_{i=1}^{T}\mu_i\left|\tilde{B}_i\right|^2,$$

where $\tilde{B}_1,\cdots,\tilde{B}_T \in \mathbb{R}^k$ are the rows of B.

(2) **Finding an Upper Bound for** $\left|\tilde{B}_1\right|^2,\cdots,\left|\tilde{B}_T\right|^2$

Because

$$B'B = \begin{pmatrix} B_1'B_1 & \cdots & B_1'B_k \\ \vdots & \ddots & \vdots \\ B_k'B_1 & \cdots & B_k'B_k \end{pmatrix}$$
$$= A'PP'A = A'A = I_k,$$

$\{B_1, \cdots, B_k\}$ is an orthonormal set and thus a collection of linearly independent T-dimensional vectors with norm 1. Letting V be the subspace of $\mathbb{R}^T$ spanned by $\{B_1, \cdots, B_k\}$, there exists an orthogonal complement $V_\perp$ of V, and $V_\perp$ satisfies $\mathbb{R}^T = V \bigoplus V_\perp$. Let $\{\beta_1, \cdots, \beta_{T-k}\}$ be an orthonormal basis of $V_\perp$; then, $\mathbb{R}^T = V \bigoplus V_\perp$ implies that

$$\{B_1, \cdots B_k, \beta_1, \cdots, \beta_{T-k}\}$$

is an orthonormal basis of $\mathbb{R}^T$. Define

$$B_\perp = \begin{pmatrix} \beta_1 & \cdots & \beta_{T-k} \end{pmatrix} \in \mathbb{R}^{T \times (T-k)},$$

and let $C = \begin{pmatrix} B & B_\perp \end{pmatrix} \in \mathbb{R}^{T \times T}$. Because the columns of B form an orthonormal basis of $\mathbb{R}^T$, C is an orthogonal matrix, and as such the norm of the rows of C are all equal to 1. Letting $C_1, \cdots, C_T$ be the rows of C, and denoting $\beta_j = (\beta_{1j}, \cdots, \beta_{Tj})'$ for any $1 \leq j \leq T-k$, this implies that

$$0 \leq \left|\tilde{B}_i\right|^2 = \sum_{j=1}^{k} B_{ij}^2 \leq \sum_{j=1}^{k} B_{ij}^2 + \sum_{j=1}^{T-k} \beta_{ij}^2 = |C_i|^2 = 1$$

for any $1 \leq i \leq T$, or that the squared norm of the rows of B are bounded above by 1.

(3) **Finding the Upper Bound of** $\text{tr}(A'MA)$

Note that

$$\sum_{i=1}^{T} \left|\tilde{B}_i\right|^2 = \sum_{i=1}^{T}\sum_{j=1}^{r} B_{ij}^2 = \sum_{j=1}^{k}\sum_{i=1}^{T} B_{ij}^2 = \sum_{j=1}^{k} |B_j|^2 = k,$$

since the columns of B all have norm equal to 1. Therefore, denoting $a_i = \left|\tilde{B}_i\right|^2$ for $1 \leq i \leq T$, $a_1, \cdots, a_T$ are values such that

i) $0 \leq a_i \leq 1$ for all $1 \leq i \leq T$, and

ii) $a_1 + \cdots + a_T = k$.

With this in mind, we can see that

$$
\begin{aligned}
\operatorname{tr}(A'MA) &= \sum_{i=1}^{T} \mu_i \left|\tilde{B}_i\right|^2 = \sum_{i=1}^{T} \mu_i a_i \\
&= \sum_{i=1}^{T} \mu_i a_i + \mu_k \sum_{i=1}^{k} (1-a_i) - \mu_k \sum_{i=1}^{k} (1-a_i) \\
&= \left[\sum_{i=1}^{k} \mu_i a_i + \mu_k \sum_{i=1}^{k} (1-a_i)\right] + \sum_{i=k+1}^{T} \mu_i a_i - \mu_k \sum_{i=1}^{k} (1-a_i) \\
&\leq \left[\sum_{i=1}^{k} \mu_i a_i + \sum_{i=1}^{k} \mu_i (1-a_i)\right] + \mu_k \left[\sum_{i=k+1}^{T} a_i - \sum_{i=1}^{k} (1-a_i)\right] \\
&\qquad\qquad (\mu_1 \geq \cdots \geq \mu_k, \text{ while } \mu_k \geq \mu_{k+1} \geq \cdots \geq \mu_T) \\
&= \sum_{i=1}^{k} \mu_i + \mu_k \left[\sum_{i=1}^{T} a_i - k\right] \\
&= \sum_{i=1}^{k} \mu_i. \qquad\qquad (\textstyle\sum_{i=1}^{T} a_i = k)
\end{aligned}
$$

Therefore, $\operatorname{tr}(A'MA)$ is bounded above by the sum of the k largest eigenvalues of M.

It is now a simple matter to solve the maximization problem. Letting $v_1, \cdots, v_k \in \mathbb{R}^T$ be an orthonormal set of eigenvectors of M corresponding $\mu_1, \cdots, \mu_k$, denote $V = \begin{pmatrix} v_1 & \cdots & v_k \end{pmatrix} \in \mathbb{R}^{T \times k}$ and note that

$$
MV = V \begin{pmatrix} \mu_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mu_k \end{pmatrix} = V\overline{D}.
$$

Then, $V'V = I_k$ and

$$
\operatorname{tr}(V'MV) = \operatorname{tr}(V'V\overline{D}) = \operatorname{tr}(\overline{D}) = \sum_{i=1}^{k} \mu_i.
$$

Putting $A = V$ allows the function $\operatorname{tr}(A'MA)$ to attain its upper bound, and therefore it is the maximizer of $\operatorname{tr}(A'MA)$ over the set of all $T \times k$ matrices A such that $A'A = I_k$.

As such, letting the columns of $\frac{1}{\sqrt{T}}\tilde{F}^k$ be orthonormal eigenvectors of XX' corresponding to the k largest eigenvalues $\mu_1 \geq \cdots \geq \mu_k \geq 0$ of XX',

$$
\operatorname{tr}\left(\frac{1}{T}\tilde{F}^{k\prime}(XX')\tilde{F}^k\right) = \sum_{i=1}^{k} \mu_i
$$

and $\tilde{F}^k$ is a solution to the stated maximization problem; we put the superscript k to emphasize the fact that k factors have been estimated.

This is why the method is called asympototic "principal components"; in effect, the estimated factors are the first k principal components of XX', found successively by searching for the linear combination of the data that yields the largest empirical variance. Here, the linear combination is across time, so that the factors are the k collection of weights that best explain the variation in the data across time.

The maximized value of $\text{tr}(F'(XX')F)$ becomes

$$\begin{aligned}\text{tr}(\tilde{F}^{k\prime}(XX')\tilde{F}^k) &= T\,\text{tr}\left(\frac{1}{\sqrt{T}}\tilde{F}^{k\prime}(XX')\frac{1}{\sqrt{T}}\tilde{F}^k\right)\\ &= T\,\text{tr}\left(\frac{1}{T}\tilde{F}^{k\prime}\tilde{F}^k\cdot D\right) = T\,\text{tr}(D) = T\sum_{i=1}^{k}\mu_i,\end{aligned}$$

where D is the $k\times k$ diagonal matrix collecting $\mu_1,\cdots,\mu_k$. Therefore, the minimized value of the objective function is

$$V(k,\tilde{F}^k) = \frac{1}{NT}\text{tr}(XX') - \frac{1}{NT}\sum_{i=1}^{k}\mu_i,$$

and the estimator of the factor loadings Λ is

$$\tilde{\Lambda}^k = \frac{1}{T}X'\tilde{F}^k.$$

Likewise, the estimator of Λ that minimizes the concentrated objective function

$$\tilde{V}(k,\Lambda) = \frac{1}{NT}\operatorname{tr}(XX') - \frac{1}{NT}\operatorname{tr}\left((\Lambda'\Lambda)^{-1}\Lambda'(X'X)\Lambda\right)$$

subject to the normalization $\frac{\Lambda'\Lambda}{N}$ is given as $\sqrt{N}$ times a set of k orthonormal eigenvectors corresponding to the k largest eigenvalues of $X'X$. Denoting this estimator by $\overline{\Lambda}^k$, the estimator of F is now given as

$$\overline{F}^k = \frac{1}{N}X\overline{\Lambda}^k,$$

and the minimized value of the objective function is

$$\tilde{V}(k,\overline{\Lambda}^k) = \frac{1}{NT}\operatorname{tr}(XX') - \frac{1}{NT}\sum_{i=1}^{k} v_i,$$

where $v_1 \geq \cdots \geq v_k$ are the k largest eigenvalues of $X'X$.

Note that, if the k largest eigenvalues $\mu_1, \cdots, \mu_k$ of XX' are positive, then they are the same as those of $X'X$; this indicates that, if the k largest eigenvalues of XX' are positive, then either of the above approaches yields the same minimum value of the objective function. In other words, $\tilde{F}^k$ and $\tilde{\Lambda}^k$ are not unique solutions to the least squares problem.
We can also see that, for any nonsingular $k \times k$ matrix P, $\tilde{F}^k P$ also solves the minimization problem, since

$$\begin{aligned} V(k,\tilde{F}^k P) &= \frac{1}{NT}\operatorname{tr}\left(X'\left(I_T - \tilde{F}^k P(P'\tilde{F}^{k\prime}\tilde{F}^k P)^{-1}P'\tilde{F}^{k\prime}\right)X\right) \\ &= \frac{1}{NT}\operatorname{tr}\left(X'\left(I_T - \tilde{F}^k(\tilde{F}^{k\prime}\tilde{F}^k)^{-1}\tilde{F}^{k\prime}\right)X\right) = V(k,\tilde{F}^k). \end{aligned}$$

1.2 Asymptotic Properties of the Estimated Factors: Assumptions and Notation

We now show, under the same set of assumptions, that the panel information criteria introdued in Bai and Ng (2002) consistently estimate the number of factors, and that the estimators derived above possess the asymptotic properties laid out in Bai (2003).

To make the discussion slightly more formal, we let $(\Omega, \mathcal{H}, \mathbb{P})$ be our probability space and assume that every random element in the following exposition is defined on Ω and $\mathcal{H}$-measurable.

1.2.1 The Trace Norm on $\mathbb{R}^{m\times n}$

Throughout, we will treat the matrix space $\mathbb{R}^{m\times n}$ as a metric space under the metric induced by the trace norm $\|\cdot\|$ on $\mathbb{R}^{m\times n}$, defined as

$$\|A\| = \mathrm{tr}(A'A)^{\frac{1}{2}}$$

for any $A \in \mathbb{R}^{m\times n}$. It is very easy to see that $\|A\|^2$ is simply the sum of the squares of all the entries of A, and it follows that

$$\|A\| = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}^2\right)^{\frac{1}{2}} \leq \sum_{i=1}^{m}\sum_{j=1}^{n} |A_{ij}|.$$

We will now show that $\|\cdot\|$ possesses the properties that a matrix norm such as the operator norm should possess.

Recall that, for any $n \in N_+$, defining $S^{n\times n}$ as the set of all symmetric $n\times n$ matrices, $S^{n\times n}$ is a linear subspace of the real vector space $\mathbb{R}^{n\times n}$: this can be seen easily, since the zero $n\times n$ matrix is symmetric and, for any $a \in \mathbb{R}$ and $A, B \in S^{n\times n}$, $(aA+B)' = aA'+B' = aA+B$ and thus $aA+B \in S^{n\times n}$.
In addition, the operation $\langle\cdot,\cdot\rangle : S^{n\times n}\times S^{n\times n} \to \mathbb{R}$ defined as

$$\langle A, B\rangle = \mathrm{tr}(A'B)$$

for any $A, B \in S^{n\times n}$ is an inner product defined on $S^{n\times n}$:

- For any $a \in \mathbb{R}$ and $A, B, C \in S^{n\times n}$,

  $$\langle aA+B, C\rangle = \mathrm{tr}((aA+B)'C) = \mathrm{tr}(a\cdot A'C + B'C) = a\cdot \mathrm{tr}(A'C) + \mathrm{tr}(B'C) = a\cdot\langle A, C\rangle + \langle B, C\rangle,$$

  so that $\langle\cdot,\cdot\rangle$ is linear in its first argument.

- For any $A, B \in S^{n\times n}$,

$$\langle A, B\rangle = \mathrm{tr}(A'B) = \mathrm{tr}(BA') = \mathrm{tr}(B'A) = \langle B, A\rangle,$$

  where we used both the commutativity property of the trace operation and the symmetry of A and B.

- For any $A \in S^{n\times n}$,

$$\langle A, A\rangle = \mathrm{tr}(A'A) = \mathrm{tr}(A^2).$$

  Letting $A = PDP'$ be the eigendecomposition of A (which exists because A is real and symmetric), $A = O$ if and only if all the diagonal entries of D are 0. Lettng $\mu_1, \cdots, \mu_n$ be the diagonal entries of D, since $A^2 = PD^2P'$ and $\mathrm{tr}(A^2) = \mathrm{tr}(D^2)$, we can see that

$$\langle A, A\rangle = \mathrm{tr}(D^2) = \sum_{i=1}^{n} \mu_i^2 \geq 0,$$

  where the inequality holds as an equality if and only if $\mu_1 = \cdots = \mu_n = 0$, or $D = O$. Therefore, $\langle A, A\rangle > 0$ if $A \neq O$.

We have just shown that $(S^{n\times n}, \langle\cdot,\cdot\rangle)$ is a real inner product space; denote by $\|\cdot\|_{tr}$ the norm induced by $\langle\cdot,\cdot\rangle$. Since

$$\|A\|_{tr} = (\langle A, A\rangle)^{\frac{1}{2}} = \mathrm{tr}(A'A)^{\frac{1}{2}}$$

for any $A \in S^{n\times n}$, we can see that $\|\cdot\|_{tr}$ equals the trace norm $\|\cdot\|$ on $S^{n\times n}$.
By the Cauchy-Schwarz inequality,

$$\left|\mathrm{tr}(A'B)\right| = |\langle A, B\rangle| \leq \|A\|_{tr}\|B\|_{tr}$$

for any $A, B \in S^{n\times n}$.
In particular, for any positive semidefinite $A \in S^{n\times n}$, letting $A = PDP'$ be its eigendecomposition and $\mu_1, \cdots, \mu_n$ be the diagonal entries of D (the eigenvalues of A), $\mu_1, \cdots, \mu_n \geq 0$. Therefore,

$$\|A\|_{tr} = \mathrm{tr}(A^2)^{\frac{1}{2}} = \left(\sum_{i=1}^{n} \mu_i^2\right)^{\frac{1}{2}} \leq \sum_{i=1}^{n} \mu_i = \mathrm{tr}(A),$$

which tells us that the trace norm of a positive semidefinite matrix is majorized by its trace.

Returning to the general setting of the space of all real $m \times n$ matrices $\mathbb{R}^{m\times n}$, we can now see that the trace norm $\|\cdot\|$ on $\mathbb{R}^{m\times n}$ has the following properties:

- $\|AB\| \leq \|A\|\|B\|$

For any $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times p}$,

$$\begin{aligned}
\|AB\|^2 = \operatorname{tr}(B'A'AB) &= \operatorname{tr}((A'A)(BB')) \\
&= \langle A'A, BB'\rangle && (A'A, BB' \text{ are } n\times n \text{ symmetric matrices}) \\
&\le \|A'A\|_{tr}\cdot\|BB'\|_{tr} && \text{(The Cauchy-Schwarz Inequality)} \\
&\le \operatorname{tr}(A'A)\cdot\operatorname{tr}(BB') && (A'A, BB' \text{ are positive semidefinite}) \\
&= \|A\|^2\cdot\|B\|^2.
\end{aligned}$$

Therefore,

$$\|AB\| \le \|A\|\cdot\|B\|.$$

- $\|a\cdot A\| = |a|\cdot\|A\|$
  Let $a \in \mathbb{R}$ and $A \in \mathbb{R}^{m\times n}$. Then,

$$\|aA\| = \operatorname{tr}(a^2A'A)^{\frac{1}{2}} = |a|\cdot\operatorname{tr}(A'A)^{\frac{1}{2}} = |a|\cdot\|A\|.$$

- $\|A+B\| \le \|A\|+\|B\|$
  Let $A, B \in \mathbb{R}^{m\times n}$;

$$\|A+B\|^2 = \operatorname{tr}((A+B)'(A+B)) = \operatorname{tr}(A'A)+\operatorname{tr}(B'B)+\operatorname{tr}(B'A)+\operatorname{tr}(A'B).$$

Letting the (i,j)th entry of A,B be denoted A_{ij}, B_{ij} for any $1\le i\le m$, $1\le j\le n$, note that

$$\operatorname{tr}(B'A) = \operatorname{tr}(A'B) = \sum_{j=1}^{n}\sum_{i=1}^{m} A_{ij}B_{ij},$$

and by the Cauchy-Schwarz inequality,

$$\sum_{i=1}^{m} A_{ij}B_{ij} \le \sum_{i=1}^{m}|A_{ij}B_{ij}| \le \left(\sum_{i=1}^{m} A_{ij}^2\right)^{\frac{1}{2}}\left(\sum_{i=1}^{m} B_{ij}^2\right)^{\frac{1}{2}}$$

for any $1\le j\le n$, so that another application of the Cauchy-Schwarz inequality yields

$$\begin{aligned}
\sum_{j=1}^{n}\sum_{i=1}^{m} A_{ij}B_{ij} &\le \sum_{j=1}^{n}\left(\sum_{i=1}^{m} A_{ij}^2\right)^{\frac{1}{2}}\left(\sum_{i=1}^{m} B_{ij}^2\right)^{\frac{1}{2}} \\
&\le \left(\sum_{j=1}^{n}\sum_{i=1}^{m} A_{ij}^2\right)^{\frac{1}{2}}\left(\sum_{j=1}^{n}\sum_{i=1}^{m} B_{ij}^2\right)^{\frac{1}{2}} = \|A\|\|B\|.
\end{aligned}$$

Therefore,

$$\begin{aligned}\|A+B\|^2 &= \operatorname{tr}(A'A)+\operatorname{tr}(B'B)+\operatorname{tr}(B'A)+\operatorname{tr}(A'B) \\ &\leq \|A\|^2+\|B\|^2+2\cdot\|A\|\|B\| = (\|A\|+\|B\|)^2.\end{aligned}$$

- **$\|A\|=0$ if and only if $A=O$**
  Let $A\in\mathbb{R}^{m\times n}$. Suppose that $\|A\|=0$. Then,

$$0=\operatorname{tr}(A'A)=\sum_{i=1}^{m}\sum_{j=1}^{n}A_{ij}^2,$$

so that $A_{ij}=0$ for any $1\leq i\leq m$, $1\leq j\leq n$. It follows that $A=O$. It is obvious that $\|A\|=0$ if $A=O$.
We have now shown that $\|\cdot\|$ is a norm on $\mathbb{R}^{m\times n}$. Therefore, we can induce a metric d on $\mathbb{R}^{m\times n}$ by defining

$$d(A,B)=\|A-B\|$$

for any $A,B\in\mathbb{R}^{m\times n}$.

- **For any $x\in\mathbb{R}^n$, $|x|=\|x\|$**
  Let x be an n-dimensional real valued vector whose euclidean norm is $|x|$. Then, $\|x\|$ is well-defined as the norm of the $n\times 1$ matrix x. It is easy to see that

$$\|x\|^2=\operatorname{tr}(x'x)=|x|^2.$$

By implication, for some $A\in\mathbb{R}^{m\times n}$ and $x\in\mathbb{R}^n$,

$$|Ax|=\|Ax\|\leq\|A\|\cdot\|x\|=\|A\|\cdot|x|.$$

- **Inversion is Continuous under $\|\cdot\|$**
  The proofs here follow those in chapter 9 of PMA for the operator norm.
  For any $n\in N_+$, let Ω^o be the space of all invertible $n\times n$ matrices. We first show that Ω^o is open under the metric induced by the trace norm $\|\cdot\|$.

Choose any $A\in\Omega^o$. Because $A^{-1}\neq O$, $\|A^{-1}\|>0$. Let $B\in\mathbb{R}^{n\times n}$ be an element in the open ball $B(A,1/\|A\|)$ around A, that is,

$$\|A-B\|<\frac{1}{\|A^{-1}\|}.$$

Choose any $x \in \mathbb{R}^n$, and suppose that $x \neq \mathbf{0}$. Then,

$$\begin{aligned} |x| = \left|A^{-1}Ax\right| \leq \left\|A^{-1}\right\| \cdot |Ax - Bx + Bx| \\ \leq \left\|A^{-1}\right\| \cdot (\|A-B\||x| + |Bx|). \end{aligned}$$

Because $|x| > 0$, we have

$$\left\|A^{-1}\right\| \cdot \|A-B\||x| < |x|,$$

so that

$$|x| < |x| + |Bx|,$$

which implies $|Bx| > 0$, or $Bx \neq \mathbf{0}$. By contraposition, if $Bx = \mathbf{0}$, then $x = \mathbf{0}$. This tells us that the null space of B consists only of the zero vector $\mathbf{0}$, and as such that B is an invertible matrix.

This holds for any $B \in B(A, 1/\|A\|)$, so $B(A, 1/\|A\|) \subset \Omega^o$. This in turn holds for any $A \in \Omega^o$, so Ω^o is open with respect to the metric induced by the trace norm.

Now we can easily show that matrix inversion is continuous.

Define $f : \Omega^o \to \Omega^o$ as

$$f(A) = A^{-1} \quad \text{for any } A \in \mathbb{R}^{n \times n}.$$

Choose any $A \in \Omega^o$, and $B \in \mathbb{R}^{n \times n}$ such that $\|A - B\| < \delta$. Then, $B \in \Omega^o$ by the above result, and because $\|A^{-1}\| \cdot \|A - B\| < 1$,

$$\begin{aligned} \left\|B^{-1}\right\| = \left\|A^{-1}AB^{-1}\right\| \leq \left\|A^{-1}\right\| \cdot \left\|(A-B)B^{-1} + I_n\right\| \\ \leq \left\|A^{-1}\right\| \cdot \|A-B\| \cdot \left\|B^{-1}\right\| + \sqrt{n} \cdot \left\|A^{-1}\right\| \end{aligned}$$

implies

$$\left\|B^{-1}\right\| \leq \frac{\sqrt{n} \cdot \|A^{-1}\|}{1 - \|A^{-1}\| \cdot \|A - B\|}.$$

It then follows that

$$\begin{aligned} \|f(A) - f(B)\| = \left\|A^{-1}(A-B)B^{-1}\right\| \leq \|A-B\| \cdot \left\|A^{-1}\right\| \cdot \left\|B^{-1}\right\| \\ \leq \frac{\sqrt{n} \cdot \|A^{-1}\| \cdot \|A-B\|}{1 - \|A^{-1}\| \cdot \|A-B\|}. \end{aligned}$$

The right hand side goes to 0 as $\|A - B\| \to 0$, so it follows that $\|f(A) - f(B)\|$ also goes to 0 as $\|A - B\| \to 0$. This shows us that f is a continuous function on Ω^o.

1.2.2 Big O Notation in Probability

Let $\{X_n\}_{n \in N_+}$ be a sequence of random elements taking values in some matrix space (E, d) whose metric d is induced by the matrix norm $\|\cdot\|$ on E. Let E be equipped with a Borel σ-algebra $\mathcal{E}$ generated by the metric topology on E induced by the metric d. Note that the relationship $\|\langle A, B\rangle\| \leq \|A\|\|B\|$ holds for any matrix under either the operator or trace norm.

We say that $\{X_n\}_{n \in N_+}$ is a $O_p(a_n)$ process for a positive real sequence $\{a_n\}_{n \in N_+}$ if $\{\frac{X_n}{a_n}\}_{n \in N_+}$ is bounded in probability, that is,

$$\text{For any } \epsilon > 0 \text{, there exists an } M > 0 \text{ and } N \in N_+ \text{ such that } \mathbb{P}\left(\left\|\frac{X_n}{a_n}\right\| > M\right) < \epsilon \text{ for any } n \geq N.$$

On the other hand, we say that $\{X_n\}_{n \in N_+}$ is an $o_p(a_n)$ process if $\{\frac{X_n}{a_n}\}_{n \in N_+}$ converges in probability to 0; to state the definition explicitly,

$$\text{For any } \delta > 0 \text{ and } \epsilon > 0 \text{, there exists an } N \in N_+ \text{ such that } \mathbb{P}\left(\left\|\frac{X_n}{a_n}\right\| > \delta\right) < \epsilon \text{ for any } n \geq N.$$

If $\{X_n\}_{n \in N_+}$ is $O_p(a_n)$ $(o_p(a_n))$, then $\{\frac{X_n}{a_n}\}_{n \in N_+}$ is $O_p(1)$ $(o_p(1))$, so we will mostly deal with processes of $O_p(1)$ and $o_p(1)$.

The following are some important properties of $o_p(1)$ and $O_p(1)$ processes:

1) **If $\{X_n\}_{n \in N_+}$ is $o_p(1)$, then it is also $O_p(1)$**

   This follows almost immediately from the definition. Suppose $\{X_n\}_{n \in N_+}$ is $o_p(1)$, and choose any $\epsilon > 0$; then, by definition,

$$\lim_{n \to \infty} \mathbb{P}(\|X_n\| > 1) = 0,$$

   so there exists an $N \in N_+$ such that $\mathbb{P}(\|X_n\| > 1) < \epsilon$ for any $n \geq N$. We can put $M = 1$; this holds for any $\epsilon > 0$, so by definition $\{X_n\}_{n \in N_+}$ is $O_p(1)$.

2) **If $\{X_n\}_{n \in N_+}$ converges in probability or distribution, then $\{X_n\}_{n \in N_+}$ is $O_p(1)$**

   First suppose $X_n \xrightarrow{d} X$ for some random element X taking values in E, and let the distributions of $\|X_n\|$, $\|X\|$ be the measures μ_n, μ on the real line $\mathbb{R}$ for all $n \in N_+$. By the Portmanteau theorem,

$$\lim_{n \to \infty} \mu_n(A) = \mu(A)$$

   for any Borel sets A on the real line whose boundary has measure 0 under μ, that is,

$\mu(\partial A) = 0$.
To prove our result, fix $\epsilon > 0$.

We first prove that the set defined as

$$\mathcal{M} = \{x \in \mathbb{R} \mid \mu(\{x\}) > 0\}$$

is at most countable. For any $n \in N_+$, define

$$\mathcal{M}_n = \{x \in \mathbb{R} \mid \mu(\{x\}) > \frac{1}{n}\};$$

then, $\mathcal{M} = \bigcup_n \mathcal{M}_n$. Suppose $\mathcal{M}_n$ contains more than n elements. Then, letting J be a finite subset of $\mathcal{M}_n$ with $n+1$ elements, we can see that J is a measurable subset of $\mathbb{R}$ and

$$1 = \mu(\mathbb{R}) \geq \mu(J) = \mu\left(\bigcup_{x \in J} \{x\}\right) = \sum_{x \in J} \mu(\{x\}) \geq \sum_{x \in J} \frac{1}{n} = \frac{|J|}{n} \geq \frac{n+1}{n} > 1,$$

which is a contradiction. Therefore, $\mathcal{M}_n$ contains at most n elements and is a finite set. Since $\mathcal{M}$ is the countable union of finite sets, it is at most countable.

Now note that $\mathbb{P}(\|X\| > n) = \mu((n, +\infty)) \to 0$ as $n \to \infty$; this is because $\{(n, +\infty)\}_{n \in N_+}$ is a sequence of subsets of $\mathbb{R}$ decreasing to $\emptyset$, which implies, by sequential continuity, that

$$\lim_{n \to \infty} \mu((n, +\infty)) = \mu(\emptyset) = 0.$$

Therefore, there exists an $N \in N_+$ such that

$$\mathbb{P}(\|X\| > n) < \frac{\epsilon}{2}$$

for any $n \geq N$. If $\mu(\{N\}) > 0$, then there exists an $M > N$ such that $\mu(\{M\}) = 0$, since otherwise the uncountable set $[N, +\infty)$ must be contained in the countable set $\mathcal{M}$, a contradiction. This implies that

$$\mu((M, +\infty)) = \mathbb{P}(\|X\| > M) \leq \mathbb{P}(\|X\| > N) < \frac{\epsilon}{2}.$$

Since $(M, +\infty)$ is a Borel set whose boundary has measure 0 under μ, by the Portmanteau theorem

$$\lim_{n \to \infty} \mu_n((M, +\infty)) = \mu((M, +\infty)).$$

As such, there exists an $N_0 \in N_+$ such that

$$|\mu_n((M, +\infty)) - \mu((M, +\infty))| < \frac{\epsilon}{2}$$

for any $n \geq N_0$. Note that $\mu_n((M,+\infty)) = \mathbb{P}(\|X_n\| > M)$ for any $n \in N_+$, so that

$$|\mathbb{P}(\|X_n\| > M) - \mathbb{P}(\|X\| > M)| < \frac{\epsilon}{2}$$

for any $n \geq N_0$.

As such, we have the string of inequalities

$$\mathbb{P}(\|X_n\| > M) \leq |\mathbb{P}(\|X_n\| > M) - \mathbb{P}(\|X\| > M)| + \mathbb{P}(\|X\| > M) < \epsilon$$

for any $n \geq N_0$. This holds for any $\epsilon > 0$, so by definition $\{X_n\}_{n \in N_+}$ is bounded in probability and thus $O_p(1)$.

Now suppose that $X_n \xrightarrow{p} X$ to some random element X taking values in E. Then, because convergence in probability implies convergence in distribution, $X_n \xrightarrow{d} X$ and $\{X_n\}_{n \in N_+}$ is again $O_p(1)$.

3) **If $\{X_n\}_{n \in N_+}$ is $o_p(1)$ and $\{Y_n\}_{n \in N_+}$ is $O_p(1)$, then $\{X_n Y_n\}_{n \in N_+}$ is $o_p(1)$**

The property in question is surprisingly easy to show:
Choose any $\delta > 0$ and $\epsilon > 0$. By boundedness in probability, there exists an $M > 0$ and $N_0 \in N_+$ such that

$$\mathbb{P}(\|Y_n\| > M) < \frac{\epsilon}{2}$$

for any $n \geq N_0$. Note that M and N_0 depend only on ϵ.
Then, by convergence in probabiliy, there exists an $N_1 \geq N_0$ such that

$$\mathbb{P}\left(\|X_n\| > \frac{\delta}{M}\right) < \frac{\epsilon}{2}$$

for any $n \geq N_1$. Here, N_1 depends on N_0, M, ϵ and δ, and therefore only on ϵ and δ.
Now note that, for any $n \geq N_1$,

$$\begin{aligned}\mathbb{P}(\|X_n Y_n\| > \delta) &\leq \mathbb{P}\left(\|X_n\| > \frac{\delta}{M}\right) + \mathbb{P}(\|Y_n\| > M) \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.\end{aligned}$$

This holds for any $\epsilon > 0$, so

$$\lim_{n \to \infty} \mathbb{P}(\|X_n Y_n\| > \delta) = 0,$$

and this again holds for any $\delta > 0$, so $X_n Y_n \xrightarrow{p} 0$ and by definition $\{X_n Y_n\}_{n\in N_+}$ is $o_p(1)$.

4) **If $\{X_n\}_{n\in N_+}$ and $\{Y_n\}_{n\in N_+}$ are $O_p(1)$, then $\{X_n Y_n\}_{n\in N_+}$ is also $O_p(1)$**

Choose some $\epsilon > 0$. Then, by definition, there exist $M_1, M_2 > 0$ and $N \in N_+$ such that

$$\mathbb{P}(\|X_n\| > M_1) < \frac{\epsilon}{2} \quad \text{and} \quad \mathbb{P}(\|Y_m\| > M_2) < \epsilon 2$$

for any $n \geq N$. Then,

$$\mathbb{P}(\|X_n Y_n\| > M_1 M_2) \leq \mathbb{P}(\|X_n\| > M_1) + \mathbb{P}(\|Y_m\| > M_2) < \epsilon$$

for any $n \geq N$, so by definition, $\{X_n Y_n\}_{n\in N_+}$ is also $O_p(1)$.

5) **If $\{X_n\}_{n\in N_+}$ and $\{Y_n\}_{n\in N_+}$ are $O_p(1)$, then $\{X_n + Y_n\}_{n\in N_+}$ is also $O_p(1)$**

For any $\epsilon > 0$, there exist $M_1, M_2 > 0$ and $N \in N_+$ such that

$$\mathbb{P}(\|X_n\| > M_1) < \frac{\epsilon}{2} \quad \text{and} \quad \mathbb{P}(\|Y_m\| > M_2) < \epsilon 2$$

for any $n \geq N$. Then,

$$\mathbb{P}(\|X_n + Y_n\| > M_1 + M_2) \leq \mathbb{P}(\|X_n\| > M_1) + \mathbb{P}(\|Y_m\| > M_2) < \epsilon$$

for any $n \geq N$, so by definition, $\{X_n + Y_n\}_{n\in N_+}$ is also $O_p(1)$.

6) **If $\{X_n\}_{n\in N_+}$ is $O_p(1)$ and $\{Y_n\}_{n\in N_+}$ is $o_p(1)$, then $\{X_n + Y_n\}_{n\in N_+}$ is $O_p(1)$**

Since being $o_p(1)$ implies being $O_p(1)$, $\{X_n + Y_n\}_{n\in N_+}$ is the sum of two $O_p(1)$ processes and is thus $O_p(1)$.

7) **If $\{X_n\}_{n\in N_+}$ is $O_p(a_n)$ and $\{Y_n\}_{n\in N_+}$ is $O_p(b_n)$, where $\frac{a_n}{b_n} \to 0$, then $\{X_n + Y_n\}_{n\in N_+}$ is $O_p(b_n)$ and $\{X_n Y_n\}_{n\in N_+}$ is $O_p(a_n b_n)$**

By definition, $\{\frac{X_n}{a_n}\}_{n\in N_+}$ and $\{\frac{Y_n}{b_n}\}_{n\in N_+}$ are $O_p(1)$. Furthermore, since

$$\frac{X_n}{b_n} = \frac{X_n}{a_n} \cdot \frac{a_n}{b_n}$$

for any $n \in N_+$, where $\{\frac{a_n}{b_n}\}_{n\in N_+}$ is a sequence of degenerate random variables that is $o_p(1)$, it follows that $\{\frac{X_n}{b_n}\}_{n\in N_+}$ is $o_p(1)$.
Therefore, the process $\{\frac{X_n}{b_n} + \frac{Y_n}{b_n}\}_{n\in N_+}$ is the sum of an $o_p(1)$ process and an $O_p(1)$ process, so that it is itself $O_p(1)$. It follows that $\{X_n + Y_n\}_{n\in N_+}$ is $O_p(b_n)$.
In contrast, we showed that the product of two $O_p(1)$ processes is also $O_p(1)$, so that the process $\{\frac{X_nY_n}{a_nb_n}\}_{n\in N_+}$ is $O_p(1)$; it follows that $\{X_nY_n\}_{n\in N_+}$ is $O_p(a_nb_n)$.

8) **If $\{X_n\}_{n\in N_+}$ is $O_p(1)$, then for any real positive sequence $\{a_n\}_{n\in N_+}$ that increases to $+\infty$, $\{X_n\}_{n\in N_+}$ is $o_p(a_n)$**

Note that $\{\frac{1}{a_n}\}_{n\in N_+}$ can be considered a sequence of degenerate random variables that converges to 0 in probability. Thus, $\frac{1}{a_n} = o_p(1)$, so that, by the result above,

$$\frac{X_n}{a_n} = \frac{1}{a_n} X_n = o_p(1)$$

as well. This implies that $\{X_n\}_{n\in N_+}$ is $o_p(a_n)$.

9) **If there exists an $M > 0$ and $N \in N_+$ such that $\mathbb{E}[\|X_n\|] < M$ for any $n \geq N$, then $\{X_n\}_{n\in N_+}$ is $O_p(1)$**

Suppose that there exists an $M > 0$ and $N \in N_+$ such that $\mathbb{E}[\|X_n\|] < M$ for any $n \geq N$. Then, for any $\epsilon > 0$,

$$\mathbb{P}\left(\|X_n\| > \frac{M}{\epsilon}\right) \leq \frac{\epsilon}{M}\mathbb{E}[\|X_n\|] < \epsilon$$

for any $n \geq N$, so that $\{X_n\}_{n\in N_+}$ is $O_p(1)$ by definition.

10) **If there exists an convergent positive real sequence $\{a_n\}_{n\in N_+}$ and $N \in N_+$ such that $\mathbb{E}[\|X_n\|] < a_n$ for any $n \geq N$, then $\{X_n\}_{n\in N_+}$ is $O_p(1)$**

Because any convergent real sequence is also bounded, there exists an $M > 0$ such that $a_n < M$ for any $n \in N_+$. By impliciation, $\mathbb{E}[\|X_n\|] < a_n < M$ for any $n \geq N$, so by the previous result, $\{X_n\}_{n \in N_+}$ is $O_p(1)$.

11) **If $\|X_n\| \leq \|Y_n\|$ for any $n \in N_+$ and $\{Y_n\}_{n \in N_+}$ is $O_p(1)$, then $\{X_n\}_{n \in N_+}$ is also $O_p(1)$**

For any ϵ, there exists an $M > 0$ and $N \in N_+$ such that

$$\mathbb{P}(\|Y_n\| > M) < \epsilon$$

for any $n \geq N$. Then, for any $n \geq N$,

$$\mathbb{P}(\|X_n\| > M) \leq \mathbb{P}(\|Y_n\| > M) < \epsilon,$$

so that $\{X_n\}_{n \in N_+}$ is also $O_p(1)$.

In general, a heuristic used to understand the big and small O notations is that a process $\{X_n\}_{n \in N_+}$ that is $O_p(a_n)$ converges at speed a_n provided that $a_n \to 0$ as $n \to \infty$, so that $\frac{X_n}{a_n}$ becomes stable. Therefore, the sum of two convergent sequences converges at the rate of the slower sequence, while their product converges at the product of their respective rates, which explains fact 7). Furthermore, a stable sequence (a sequence of $O_p(1)$) converges when multiplied by a sequence that converges to 0, which explains fact 8).

We reiterate the above findings for future reference:

1) **An** $o_p(1)$ **process is also** $O_p(1)$

2) **Any process converging in probability or distribution is** $O_p(1)$

3) $o_p(1)O_p(1) = o_p(1)$

4) $O_p(1)O_p(1) = O_p(1)$

5) $O_p(1) + O_p(1) = O_p(1)$

6) $O_p(1) + o_p(1) = O_p(1)$

7) **If** $\frac{a_n}{b_n} \to 0$**, then** $O_p(a_n) + O_p(b_n) = O_p(b_n)$ **and** $O_p(a_n)O_p(b_n) = O_p(a_n b_n)$

8) **If** $a_n \nearrow +\infty$**, then** $O_p(1) = o_p(a_n)$

9) **Any sequence whose first moments are bounded is** $O_p(1)$

10) **Any sequence whose first moments are majorized by a convergent sequence is** $O_p(1)$

11) **Any sequence majorized by an** $O_p(1)$ **sequence is also** $O_p(1)$

1.2.3 Continuity of Eigenvalues and Eigenvectors

Another property that will be used extensively throughout the proofs is the continuity of eigenvalues and eigenvectors. This allows us to infer the convergence of eigenvalues and eigenvectors from the convergence of the associated matrices via the continuous mapping theorem. Seeing as how the factor and factor loading estimators are constructed as eigenvectors of a random matrix, this will come in handy later on.

The first problem we must deal with is the uniqueness of eigenvalues and eigenvectors. Given an arbitrary square matrix $A \in \mathbb{R}^{n\times n}$, A has n (possibly complex) eigenvalues, and thus a mapping from $\mathbb{R}^{n\times n}$ to the eigenvalues of $n \times n$ matrices must be a mapping from a matrix to a set of $n!$ permutations of its eigenvalues in order to be a function. Likewise, if a diagonalizable matrix A has repeated eigenvalues, then the eigenspace corresponding to that eigenvalue does not have a unique orthonormal basis, which means that we cannot recover a unique eigenbasis of A.

Continuity of Ordered Eigenvalues when the Eigenvalues are Real

To deal with the eigenvalue uniqueness problem, we first define the set $\mathcal{M}_n$ of $n \times n$ matrices with real eigenvalues, and the subset Λ_n of $\mathbb{R}^n$ defined as

$$\Lambda_n = \{(x_1, \cdots, x_n) \in \mathbb{R}^n \mid x_1 \geq \cdots \geq x_n\}.$$

We let the ordered eigenvalues of any $A \in \mathcal{M}_n$ be the eigenvalues of A ordered from largest to smallest. That is, $\lambda_1, \cdots, \lambda_n \in \mathbb{R}$ are the ordered eigenvalues of A if they are eigenvalues of A and $\lambda_1 \geq \cdots \geq \lambda_n$.
The function $eig_n : \mathcal{M}_n \to \Lambda_n$ is defined as

$$eig_n(A) = \text{The vector of ordered eigenvalues of } A$$

for any $A \in \mathcal{M}_n$. We can then show that eig is a continuous function under the matrix norm $\|\cdot\|$ on $\mathcal{M}_n$ and the euclidean norm $|\cdot|$ on $\mathbb{R}^n$. To this end, recall that, for any matrix $A \in \mathcal{M}_n$ and an eigenvalue $\lambda \in \mathbb{R}$ of A, letting $v \in \mathbb{R}^n$ be an eigenvector of λ_i,

$$|\lambda_i||v| = |\lambda_i v| = |Av| \leq \|A\||v|.$$

Because v is non-zero by the definition of an eigenvector, $|v| > 0$ and therefore

$$|\lambda_i| \leq \|A\| < +\infty.$$

The vector $(\lambda_1, \cdots, \lambda_n) \in \Lambda_n$ of ordered eigenvalues of A is now bounded above as follows:

$$|(\lambda_1, \cdots, \lambda_n)| \leq \sum_{i=1}^{n} |\lambda_i| \leq n \cdot \|A\| < +\infty.$$

Now choose any $A \in \mathcal{M}_n$ and let $\{A_k\}_{k \in N_+}$ be a sequence in $\mathcal{M}_n$ converging to A in the operator norm. For any $k \in N_+$, let $\lambda^{(k)} = (\lambda_1^{(k)}, \cdots, \lambda_n^{(k)}) \in \Lambda_n$ be the collection of ordered eigenvalues of A_k, and likewise define $\lambda = (\lambda_1, \cdots, \lambda_n) \in \Lambda_n$ for A. Then, from the preceding result we know that

$$\left|\lambda^{(k)}\right| \leq n \cdot \|A_k\|$$

for any $k \in N_+$. Since $\|A_k - A\|$ converges to 0 as $k \to \infty$, the sequence $\{n \cdot \|A_k\|\}_{k \in N_+}$ is bounded, and by the above inequality, so is $\{\lambda^{(k)}\}_{k \in N_+} \subset \Lambda_n$.

Λ_n is a subset of $\mathbb{R}^n$, so $\{\lambda^{(k)}\}_{k \in N_+}$ is a bounded subset of $\mathbb{R}^n$, which implies by the Bolzano-Weierstrass theorem that $\{\lambda^{(k)}\}_{k \in N_+}$ has at least one convergent subsequence. It remains to show that every convergent subsequence of $\{\lambda^{(k)}\}_{k \in N_+}$ converges to λ to complete the proof.

Suppose $\{\lambda^{(k_m)}\}_{m \in N_+}$ is a convergent subsequence of $\{\lambda^{(k)}\}_{k \in N_+}$, with limit $\lambda^* = (\lambda_1^*, \cdots, \lambda_n^*) \in \mathbb{R}^n$. It can easily be shown that Λ_n is a closed subset of $\mathbb{R}^n$, so $\lambda^* \in \Lambda_n$, that is, $\lambda_1^* \geq \cdots \geq \lambda_n^*$. For any $1 \leq i \leq n$ and $m \in N_+$, $\lambda_i^{(k_m)}$ solves the equation

$$\left|\lambda_i^{(k_m)} I_n - A_{k_m}\right| = 0,$$

and by the continuity of the determinant and the fact that $A_{k_m} \to A$ as $m \to \infty$,

$$|\lambda_i^* I_n - A| = \lim_{m \to \infty} \left|\lambda_i^{(k_m)} I_n - A_{k_m}\right| = 0.$$

It follows that $\lambda_1^*, \cdots, \lambda_n^*$ are eigenvalues of A such that $\lambda_1^* \geq \cdots \geq \lambda_n^*$; by the uniqueness of ordered eigenvalues, $\lambda^* = \lambda$ and $\{\lambda^{(k_m)}\}_{m \in N_+}$ converges to λ. This holds for any convergent subsequence of $\{\lambda^{(k)}\}_{k \in N_+}$, so

$$\lim_{k \to \infty} \lambda^{(k)} = \lambda.$$

This can be rewritten as

$$\lim_{k \to \infty} eig_n(A_k) = eig_n(A),$$

so that eig_n is a continuous function.

The Continuity of Eigenvectors when the Eigenvalues are Distinct

Now that we have established the continuity of ordered eigenvalues of matrices whose eigenvalues are real, we can establish the continuity of normalized eigenvectors for a subset of $\mathcal{M}_n$ via similar methods.

Formally, define the subset $\mathcal{M}_n^d$ of $\mathcal{M}_n$ as the set of $n \times n$ matrices with real **and distinct** eigenvalues.
Let $A \in \mathcal{M}_n^d$ with distinct ordered eigenvalues $\lambda_1 > \cdots > \lambda_n$. Then, because the eigenvalues are all distinct, the eigenspace corresponding to each eigenvalue has dimension exactly 1. This means that, for any $1 \leq i \leq n$, there exists exactly two orthonormal bases for the eigenspace corresponding to λ_i: specifically, for any eigenvector v_i of λ_i with norm 1, $\{v_i\}$ and $\{-v_i\}$ are the only orthonormal bases for the eigenspace of λ_i.
Suppose that we are given a set of n signs $s = (s_1, \cdots, s_n)$. Then, the above result means that, for any $1 \leq i \leq n$, there exists exactly one eigenvector v_i of λ_i with norm 1 and first entry with the sign s_i. As such, collecting the unique normalized eigenvectors $v_1, \cdots, v_n \in \mathbb{R}^n$ of $\lambda_1, \cdots, \lambda_n$ whose first entries have the signs $s_1, \cdot, s_n$ into the matrix $V = \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix}$, V is a nonsingular matrix (eigenspaces of different eigenvalues are linearly independent) with columns of norm 1 such that

$$AV = VD,$$

where D is the diagonal matrix collecting the ordered eigenvalues $\lambda_1, \cdots, \lambda_n$ of A.

Based on the above observation, given a vector of n signs s we can define the function $eigvec_n^s : \mathcal{M}_n^d \to \mathbb{R}^{n \times n}$ as

$$\begin{aligned} eigvec_n^s(A) = &\text{ The unique } n \times n \text{ nonsingular matrix with columns of norm 1 and signs } s \text{ such that} \\ &AV = VD, \text{ where } D \text{ is the diagonal matrix collecting the ordered eigenvalues of } A \end{aligned}$$

for any $A \in \mathcal{M}_n^d$. Note that, because the columns of $eigvec_n^s(A)$ have norm 1 and $\|eigvec_n^s(A)\|^2$ is bounded above by the sum of the squared norms of the columns of $eigvec_n^s(A)$, it follows that

$$\|eigvec_n^s(A)\|^2 \leq n, \text{ or } \|eigvec_n^s(A)\| \leq \sqrt{n}.$$

Note that, if A is symmetric, then $eigvec_n^s(A)$ is an orthogonal matrix, since the eigenvectors in different eigenspaces are orthogonal in this case.
It is now easy to show the continuity of $eigvec_n^s$ on $\mathcal{M}_n^d$.

Let $A \in \mathcal{M}_n^d$ and let $\{A_k\}_{k \in N_+}$ be a sequence in $\mathcal{M}_n^d$ converging to A in the operator norm. For any $k \in N_+$, let $\lambda^{(k)} = (\lambda_1^{(k)}, \cdots, \lambda_n^{(k)}) \in \Lambda_n$ be the collection of ordered eigenvalues of A_k, and likewise define $\lambda = (\lambda_1, \cdots, \lambda_n) \in \Lambda_n$ for A. Define

$$D_k = \begin{pmatrix} \lambda_1^{(k)} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n^{(k)} \end{pmatrix} \quad \text{and} \quad D = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{pmatrix}$$

for any $k \in N_+$. Then, letting $V_k = eigvec_n^s(A_k)$ for any $k \in N_+$ and $V = eigvec_n^s(A)$, we have

$$A_k V_k = V_k D_k \quad \text{and} \quad AV = VD$$

for any $k \in N_+$.
In addition, $\|V_k\| \leq \sqrt{n}$ for any $k \in N_+$, so the sequence $\{V_k\}_{k \in N_+}$ is bounded in the operator norm. Since matrix spaces can be seen as extensions of euclidean spaces under the operator norm, by the Bolzano-Weierstrass theorem $\{V_k\}_{k \in N_+}$ has a convergent subsequence. As before, it now remains to see that every convergent subsequence of $\{V_k\}_{k \in N_+}$ converges to V.

To see this, first observe that, because $\{A_k\}_{k \in N_+}$ is a sequence in $\mathcal{M}_n$ and A and element of $\mathcal{M}_n$, by the result proved above

$$\lambda^{(k)} = eig_n(A_k) \to eig_n(A) = \lambda$$

as $k \to \infty$. Thus, $D_k \to D$ in the operator norm.
Let $\{V_{k_m}\}_{m \in N_+}$ be any convergent subsequence of $\{V_k\}_{k \in N_+}$ with limit V^*. Since the convergence $V_{k_m} \to V$ in the operator norm implies element-wise convergence, the columns of each V_{k_m} have norm 1, and the unit circle on $\mathbb{R}^n$ is closed, the columns of V^* must also have norm 1. In addition, because signs are preserved across limits, the first entries of each column of V^* have the signs assigned in s. Note also that, because

$$A_{k_m} V_{k_m} = V_{k_m} D_{k_m}$$

for any $m \in N_+$, where $A_{k_m} \to A$ and $D_{k_m} \to D$ in the operator norm, taking $m \to \infty$ on both sides yields

$$AV^* = V^* D.$$

By definition, $V^* = eigvec_n^s(A) = V$, and as such $V_{k_m} \to V$ as $m \to \infty$. This holds for any subsequence of $\{V_k\}_{k \in N_+}$, so

$$\lim_{k \to \infty} V_k = V,$$

or in other words,

$$\lim_{k\to\infty} eigvec_n^s(A_k) = eigvec_n^s(A).$$

1.2.4 Assumptions and Preliminaries

Bai and Ng (2002) and Bai (2003) both prove their results under an approximate factor model framework with time-series serial correlation and heteroskedasticity, that is, they assume that the errors e_{it} still retain correlation across both the cross-sectional and time dimensions and that the distributions of the time series $\{e_{it}\}_{t\in\mathbb{Z}}$ for $i \in N_+$ may not be identical. They also do not specify whether these series should be stationary or non-stationary, imbuing the model with the utmost generality.

The same results will be proved here under the stronger assumptions of an exact factor model and stationarity, that is, we will assume that $\{e_{it}\}_{t\in\mathbb{Z}}$ are i.i.d. and weakly stationary time series for all $i \in N_+$. We will also make the following assumptions:

(1) **Non-triviality of Scaled Factors**
We assume that there exists a $k_{max} \in N_+$ such that $r < k_{max}$ and the k_{max} largest eigenvalues of XX' are always positive. This implies that the k largest eigenvalues of XX' are always positive for $1 \leq k \leq k_{max}$, and as such that, when we use the scaled factors $\hat{F}^k = \frac{1}{NT} XX'\tilde{F}^k$ later on, the scaled factors are non-zero, or non-trivial.
Additionally, we assume the true number of factors r satisfies $r < k_{max}$.

(2) **Second Moment Convergence of True Factors and Factor Loadings**
We assume that the factor loadings $\lambda_1^0, \cdots, \lambda_N^0$ are nonrandom, and that there exists an $M > 0$ such that

$$\sup_{t\in N_+} \mathbb{E}\left|F_t^0\right|^2 \leq M,$$

$$\sup_{i\in N_+} \left|\lambda_i^0\right|^2 \leq M.$$

In addition, we assume that

$$\frac{F^{0\prime}F^0}{T} \xrightarrow{p} \Sigma_F \quad \text{and} \quad \frac{\Lambda^{0\prime}\Lambda^0}{T} \rightarrow \Sigma_\Lambda$$

for some positive definite matrices $\Sigma_F.\Sigma_\Lambda \in \mathbb{R}^{r\times r}$.

(3) **Exact Factor Model**
We assume that the processes $\{e_{it}\}_{t\in\mathbb{Z}}$ are independent and identically distributed for any $i \in N_+$.

(4) **Stationarity of Errors**
We assume that $\{e_{it}\}_{t\in\mathbb{Z}}$ is weakly stationary with mean 0 and autocovariance function

$\gamma : \mathbb{Z} \to \mathbb{R}$.

In addition, we assume that the autocovariances are absolutely summable and that the time series has bounded fourth moments, that is, there exists an $\mu_4 < +\infty$ such that $\mathbb{E}\left[e_{it}^4\right] < \mu_4$ for any $t \in N_+$.

(5) **Weak Dependence between Factors and Errors**

There exists an $M > 0$ such that

$$\mathbb{E}\left[\frac{1}{NT}\sum_{i=1}^{N}\left|\sum_{t=1}^{T} F_t^0 e_{it}\right|^2\right] \leq M$$

$$\mathbb{E}\left|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1}^{T} F_s^0 \left(e_{it}e_{is} - \gamma(t-s)\right)\right|^2 \leq M \qquad \text{(for any } t \in N_+\text{)}$$

$$\mathbb{E}\left\|\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\sum_{i=1}^{N} F_t^0 \lambda_i^{0\prime} e_{it}\right\|^2 \leq M$$

for any $N, T \in N_+$.

(6) **CLT for Time Dimension**

For any $i \in N_+$,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} F_t^0 e_{it} \xrightarrow{d} N\left[\mathbf{0}, \Phi_i\right],$$

for the positive definite matrix

$$\Phi_i = \operatorname*{plim}_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} e_{it}^2 F_t^0 F_t^{0\prime}.$$

(7) **Sufficient Conditions for Factor Identification**

The k_{max} largest eigenvalues of XX' are distinct for any $N, T \in N_+$ such that $T \geq k_{max}$. Likewise, the $r \times r$ matrix $\Sigma_\Lambda \Sigma_F$ has distinct eigenvalues.

(8) **The Probability Limit of $\frac{F^{0\prime}\tilde{F}^k}{T}$**

We assume that, for any $1 \leq k \leq k_{max}$, there exists an $r \times k$ matrix Q^k of full rank such that

$$\frac{F^{0\prime}\tilde{F}^k}{T} \xrightarrow{p} Q^k.$$

This assumption greatly simplifies the proofs below, and, together with assumption 7, we can express Q^r in terms of quantities related to the matrices Σ_Λ, Σ_F.

The following are some implications of the above assumptions:

- **Implications of Assumption 2**
  Letting $\Sigma_F^{\frac{1}{2}}$ be the Cholesky factor of Σ_F, the $r \times r$ matrix $\Sigma_F^{\frac{1}{2}\prime}\Sigma_\Lambda\Sigma_F^{\frac{1}{2}}$ is positive definite and thus has r positive eigenvalues. By implication, the eigenvalues of $\Sigma_\Lambda\Sigma_F$ are equal to those of $\Sigma_F^{\frac{1}{2}\prime}\Sigma_\Lambda\Sigma_F^{\frac{1}{2}}$ and are thus all positive.

- **Implications of Assumption 5**
  The first two statements of the assumption tells us that the errors e_{it} and their cross products $e_{it}e_{is} - \gamma(s-t)$ are only weakly dependent on the factors F_t^0. This imposes some kind of exogeneity on the factors and is standard in much of the time series literature.
  Meanwhile, the third statement in the assumption says that the common component $F_t^0\lambda_i^{0\prime}$ itself is only weakly correlated with the errors.

- **Implications of Assumption 7**
  Recall that the factor estimate $\tilde{F}^k$ was derived as

  $$\tilde{F}^k = \sqrt{T} \times \text{The orthonormal eigenvectors of } XX' \text{ corresponding to its } k \text{ largest eigenvalues}$$

  for any $1 \leq k \leq k_{max}$. Since the k_{max} largest eigenvalues of XX' are assumed to be distinct, so are its k largest eigenvalues. By implication, the eigenspaces corresponding to the k largest eigenvalues of XX' have dimension 1, meaning that each eigenspace has an orthonormal eigenbasis consisting of a single vector of norm 1 that is unique up to sign changes. Since eigenvectors corresponding to different eigenvalues are orthogonal to one another, the columns of $\tilde{F}^k$ are determined uniquely up to sign changes.
  By extension, the factor loadings $\tilde{\Lambda}^k = \frac{1}{T}X'\tilde{F}^k$ are also determined uniquely up to sign changes.

We first prove some preliminary results about the rate of convergence of some factors before moving onto the actual proof. The results are as below:

- $\frac{1}{T}\sum_{t=1}^{T}\left|\tilde{F}_t^k\right|^2 = k$

  To see this, note simply that

$$\frac{1}{T}\sum_{t=1}^{T}\left|\tilde{F}_t^k\right|^2 = \mathrm{tr}\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{F}_t^k\tilde{F}_t^{k\prime}\right) = \mathrm{tr}\left(\frac{\tilde{F}^{k\prime}\tilde{F}_t^k}{T}\right) = k$$

  for any $T \in N_+$ and $1 \leq k \leq k_{max}$.

- $\frac{1}{NT}\sum_{t=1}^{T}\left|\sum_{i=1}^{N}\lambda_i^0 e_{it}\right|^2 = O_p(1)$

  This is the counterpart to the first assumption in assumption (5), and it implies that the term in the expectations is $O_p(1)$. This can be derived directly from the fact that the true factor loadings are nonrandom and the stationarity of $\{e_{it}\}_{t\in\mathbb{Z}}$.

  Note that, for any $t \in N_+$,

$$\begin{aligned}\mathbb{E}\left|\sum_{i=1}^{N}\lambda_i^0 e_{it}\right|^2 &\leq \sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}\left[e_{it}e_{jt}\lambda_i^{0\prime}\lambda_j^0\right] \\ &= \sum_{i=1}^{N}\lambda_i^{0\prime}\lambda_i^0\cdot\mathbb{E}\left[e_{it}^2\right] = \gamma(0)\cdot\left(\sum_{i=1}^{N}\lambda_i^{0\prime}\lambda_i^0\right) \\ &= \gamma(0)\cdot\mathrm{tr}\left(\Lambda^{0\prime}\Lambda^0\right)\end{aligned}$$

  by the assumed exact factor structure and weak stationarity of the errors.
  As such, we can see that

$$\mathbb{E}\left[\frac{1}{NT}\sum_{t=1}^{T}\left|\sum_{i=1}^{N}\lambda_i^0 e_{it}\right|^2\right] \leq \gamma(0)\cdot\mathrm{tr}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right).$$

  Because $\frac{\Lambda^{0\prime}\Lambda^0}{N} \to \Sigma_\Lambda$ by assumption, the expectation of $\frac{1}{NT}\sum_{t=1}^{T}\left|\sum_{i=1}^{N}\lambda_i^0 e_{it}\right|^2$ is majorized by a convergent real sequence, which implies that $\frac{1}{NT}\sum_{t=1}^{T}\left|\sum_{i=1}^{N}\lambda_i^0 e_{it}\right|^2$ is $O_p(1)$.

- $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}|X_{it}|^2 = O_p(1)$

The above process can be bounded above as follows:

$$\begin{aligned}
\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}|X_{it}|^2 &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\lambda_i^{0\prime}F_t^0 + e_{it})^2 \\
&= \mathrm{tr}\left(\frac{1}{N}\Lambda^0\frac{F^{0\prime}F^0}{T}\Lambda^{0\prime}\right) + 2\frac{1}{N}\sum_{i=1}^{N}\lambda_i^{0\prime}\left(\frac{1}{T}\sum_{t=1}^{T}F_t^0 e_{it}\right) + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}^2.
\end{aligned}$$

The first term is clearly $O_p(1)$, since

$$\mathrm{tr}\left(\frac{1}{N}\Lambda^0\frac{F^{0\prime}F^0}{T}\Lambda^{0\prime}\right) = \mathrm{tr}\left[\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)\left(\frac{F^{0\prime}F^0}{T}\right)\right]$$

and the matrices $\frac{\Lambda^{0\prime}\Lambda^0}{N}$, $\frac{F^{0\prime}F^0}{T}$ converge in probability to positive definite matrices.
As for the second term, because

$$\mathbb{E}\left[\frac{1}{NT}\sum_{i=1}^{N}\left|F^{0\prime}\underline{e}_i\right|^2\right] = \mathbb{E}\left[\frac{1}{NT}\sum_{i=1}^{N}\left|\sum_{t=1}^{T}F_t^0 e_{it}\right|^2\right] < M$$

for some $M > 0$ by assumption 5 and

$$\begin{aligned}
\left|\frac{1}{N}\sum_{i=1}^{N}\lambda_i^{0\prime}\left(\frac{1}{T}\sum_{t=1}^{T}F_t^0 e_{it}\right)\right| &\leq \frac{1}{NT}\sum_{i=1}^{N}\left|\lambda_i^0\right|\left|\sum_{t=1}^{T}F_t^0 e_{it}\right| \\
&\leq \frac{1}{NT}\left(\sum_{i=1}^{N}\left|\lambda_i^0\right|^2\right)^{\frac{1}{2}}\left(\sum_{i=1}^{N}\left|\sum_{t=1}^{T}F_t^0 e_{it}\right|^2\right)^{\frac{1}{2}} = \mathrm{tr}\left(\frac{1}{T}\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{\frac{1}{2}}\left(\frac{1}{NT}\sum_{i=1}^{N}\left|\sum_{t=1}^{T}F_t^0 e_{it}\right|^2\right)
\end{aligned}$$

by the Cauchy-Schwarz inequality, we can see that

$$\mathbb{E}\left[\left|\frac{1}{N}\sum_{i=1}^{N}\lambda_i^{0\prime}\left(\frac{1}{T}\sum_{t=1}^{T}F_t^0 e_{it}\right)\right|\right] \leq M\cdot\mathrm{tr}\left(\frac{1}{T}\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{\frac{1}{2}},$$

where the last term goes to 0 as $N,T \to \infty$, so the second term is $o_p(1)$.
Finally,

$$\mathbb{E}\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}^2\right] = \gamma(0) < +\infty,$$

so the last term is $O_p(1)$. Therefore,

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}|X_{it}|^2 = O_p(1) + o_p(1) + O_p(1) = O_p(1).$$

- $\frac{1}{\sqrt{N}}\Lambda^{0\prime}e_t \xrightarrow{d} N\left(\mathbf{0}, \gamma(0)\cdot\Sigma_\Lambda\right)$

  $\{\lambda_i^0 e_{it}\}_{i\in N_+}$ is a sequence of independent random vectors by the exact factor model assumption with a finite mean $\mathbf{0}$ and covariance matrix $\gamma(0)\cdot\lambda_i^0\lambda_i^{0\prime}$ for any $i \in N_+$. Note that

$$\frac{1}{N}\sum_{i=1}^{N}\left(\gamma(0)\cdot\lambda_i^0\lambda_i^{0\prime}\right) = \gamma(0)\frac{\Lambda^{0\prime}\Lambda^0}{N} \to \gamma(0)\Sigma_\Lambda$$

  as $N \to \infty$, and

$$\begin{aligned}\frac{1}{N^3}\sum_{i=1}^{N}\mathbb{E}\left[\left|\lambda_i^0 e_{it}\right|^4\right] &\leq \mu_4\cdot\left(\frac{1}{N^3}\sum_{i=1}^{N}\left|\lambda_i^0\right|^4\right)\\ &\leq \mu_4\frac{1}{N}\left(\frac{1}{N}\sum_{i=1}^{N}\left|\lambda_i^0\right|^2\right)^2 = \mu_4\frac{1}{N}\left(\text{tr}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)\right)^2,\end{aligned}$$

  Thus,

$$\lim_{N\to\infty}\frac{1}{N^3}\sum_{i=1}^{N}\mathbb{E}\left[\left|\lambda_i^0 e_{it}\right|^4\right] = 0.$$

  We can now apply Lyapunov's CLT to see that

$$\frac{1}{\sqrt{N}}\Lambda^{0\prime}e_t = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i^0 e_{it} \xrightarrow{d} N\left[\mathbf{0}, \gamma(0)\cdot\Sigma_\Lambda\right]$$

  as $N, T \to \infty$.

1.3 Consistency of the Estimated Factors

The consistency of the factors is established by first normalizing them in an appropriate way. Recall that the columns of $\frac{1}{\sqrt{T}}\tilde{F}^k$ form an orthonormal set of k eigenvectors of XX' corresponding to its k largest eigenvalues. Let the k largest eigenvalues of $\frac{1}{NT}XX'$ be collected in the diagonal matrix $V_{NT}^k \in \mathbb{R}^{k\times k}$; then, the diagonal entries of V_{NT}^k are $\frac{1}{NT}$ times the k largest eigenvalues of XX', which implies that

$$\left(\frac{1}{NT}XX'\right)\frac{1}{\sqrt{T}}\tilde{F}^k = \frac{1}{\sqrt{T}}\tilde{F}^k V_{NT}^k.$$

By implication,

$$V_{NT}^k = \left(\frac{1}{T}\tilde{F}^{k\prime}\tilde{F}^k\right)V_{NT}^k = \frac{1}{T}\left[\tilde{F}^{k\prime}\left(\frac{1}{NT}XX'\right)\tilde{F}^k\right].$$

Our normalizaiton of the estimated factors is

$$\hat{F}^k = \tilde{F}^k V_{NT}^k.$$

Because the diagonal entries of V_{NT}^k are assumed to be positive, V_{NT}^k is non-singular and thus

$$V(k,\hat{F}^k) = V(k,\tilde{F}^k).$$

Effectively, $\hat{F}^k$ scales each factor estimate by the corresponding eigenvalue.

We can expand $\hat{F}^k$ as

$$\begin{aligned}
\hat{F}^k &= \left(\frac{1}{NT}XX'\right)\tilde{F}^k \\
&= \left[\frac{1}{NT}\left(F^0\Lambda^{0\prime}+e\right)\left(F^0\Lambda^{0\prime}+e\right)'\right]\tilde{F}^k \\
&= F^0\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)\left(\frac{F^{0\prime}\tilde{F}^k}{T}\right) + \frac{1}{NT}e\Lambda^0F^{0\prime}\tilde{F}^k + \frac{1}{NT}F^0\Lambda^{0\prime}e'\tilde{F}^k + \frac{1}{NT}ee'\tilde{F}^k.
\end{aligned}$$

Defining $H^k = \left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)\left(\frac{F^{0\prime}\tilde{F}^k}{T}\right)$, which is a $r\times k$-matrix valued random element with rank $\min(k,r)$, we can see that

$$\left\|\hat{F}^k - F^0H^k\right\| \le \left\|\frac{1}{NT}e\Lambda^0F^{0\prime}\tilde{F}^k\right\| + \left\|\frac{1}{NT}F^0\Lambda^{0\prime}e'\tilde{F}^k\right\| + \left\|\frac{1}{NT}ee'\tilde{F}^k\right\|.$$

Note that

$$\left\|H^k\right\| = \left\|\left(\frac{\tilde{F}'F^0}{T}\right)\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)\right\| \leq \frac{1}{T}\left\|\tilde{F}^k\right\|\left\|F^0\right\|\left\|\frac{\Lambda^{0\prime}\Lambda^0}{N}\right\|$$
$$\leq \frac{1}{T}\left\|\tilde{F}'\tilde{F}^k\right\|^{\frac{1}{2}}\left\|F^{0\prime}F^0\right\|^{\frac{1}{2}}\left\|\frac{\Lambda^{0\prime}\Lambda^0}{N}\right\|$$
$$= \left\|\frac{\tilde{F}'\tilde{F}}{T}\right\|^{\frac{1}{2}}\left\|\frac{F^{0\prime}F^0}{T}\right\|^{\frac{1}{2}}\left\|\frac{\Lambda^{0\prime}\Lambda^0}{N}\right\|.$$

All three matrices on the right hand side are $O_p(1)$, so H^k is also $O_p(1)$ for any $k \in N_+$.

For any real numbers $x_1, \cdots, x_n$, by the Cauchy-Schwarz inequality,

$$\left|\sum_{i=1}^n x_i\right| \leq \sum_{i=1}^n |x_i \cdot 1| \leq \left(\sum_{i=1}^n x_i^2\right)^{\frac{1}{2}}\left(\sum_{i=1}^n 1\right)^{\frac{1}{2}} = \sqrt{n}\cdot\left(\sum_{i=1}^n x_i^2\right)^{\frac{1}{2}},$$

so we have

$$\left(\sum_{i=1}^n x_i\right)^2 \leq n\cdot\left(\sum_{i=1}^n x_i^2\right).$$

Using this inequality, we can see that

$$\frac{1}{T}\left\|\hat{F}^k - F^0H^k\right\| \leq 3\frac{1}{T}\left[\left\|\frac{1}{NT}e\Lambda^0F^{0\prime}\tilde{F}^k\right\|^2 + \left\|\frac{1}{NT}F^0\Lambda^{0\prime}e'\tilde{F}^k\right\|^2 + \left\|\frac{1}{NT}ee'\tilde{F}^k\right\|^2\right].$$

We examine each term in turn:

1) $\frac{1}{T}\left\|\frac{1}{NT}e\Lambda^0F^{0\prime}\tilde{F}^k\right\|^2$
The rows of the term inside the norm are given by

$$e\Lambda^0F^{0\prime}\tilde{F}^k = \begin{pmatrix} e_1'\Lambda^0F^{0\prime}\tilde{F}^k \\ \vdots \\ e_T'\Lambda^0F^{0\prime}\tilde{F}^k \end{pmatrix}.$$

Therefore,

$$\begin{aligned}\left\|e\Lambda^0F^{0\prime}\tilde{F}^k\right\|^2 &= \operatorname{tr}\left[\left(e\Lambda^0F^{0\prime}\tilde{F}^k\right)\left(e\Lambda^0F^{0\prime}\tilde{F}^k\right)'\right] \\ &= \sum_{t=1}^{T}\left|e_t'\Lambda^0F^{0\prime}\tilde{F}^k\right|^2,\end{aligned}$$

and this term can be further majorized as

$$\begin{aligned}\sum_{t=1}^{T}\left|e_t'\Lambda^0F^{0\prime}\tilde{F}^k\right|^2 &\leq \left(\sum_{t=1}^{T}\left|e_t'\Lambda^0\right|^2\right)\cdot\left\|F^0\right\|^2\cdot\left\|\tilde{F}^k\right\|^2 \\ &= \left(\sum_{t=1}^{T}\left|\sum_{i=1}^{N}\lambda_i^0e_{it}\right|^2\right)\cdot\left\|F^0\right\|^2\cdot\left\|\tilde{F}^k\right\|^2.\end{aligned}$$

As such,

$$\begin{aligned}\frac{1}{T}\left\|\frac{1}{NT}e\Lambda^0F^{0\prime}\tilde{F}^k\right\|^2 &= \frac{1}{N^2T^3}\left\|e\Lambda^0F^{0\prime}\tilde{F}^k\right\|^2 \\ &\leq \frac{1}{N}\left(\frac{1}{NT}\sum_{t=1}^{T}\left|\sum_{i=1}^{N}\lambda_i^0e_{it}\right|^2\right)\cdot\operatorname{tr}\left(\frac{F^{0\prime}F^0}{T}\right)\cdot\operatorname{tr}\left(\frac{\tilde{F}^{k\prime}\tilde{F}^k}{T}\right) \\ &= \frac{k}{N}\left(\frac{1}{NT}\sum_{t=1}^{T}\left|\sum_{i=1}^{N}\lambda_i^0e_{it}\right|^2\right)\cdot\operatorname{tr}\left(\frac{F^{0\prime}F^0}{T}\right),\end{aligned}$$

where the last equality follows because $\frac{\tilde{F}^{k\prime}\tilde{F}^k}{T} = I_k$. All terms on the right hand side, aside from $\frac{k}{N}$, is $O_p(1)$, so

$$\frac{1}{T}\left\|\frac{1}{NT}e\Lambda^0F^{0\prime}\tilde{F}^k\right\|^2 = O_p\left(\frac{1}{N}\right).$$

2) $\frac{1}{T}\left\|\frac{1}{NT}F^0\Lambda^{0\prime}e'\tilde{F}^k\right\|^2$

The rows of the matrix in within the norm are given by

$$F^0\Lambda^{0\prime}e'\tilde{F}^k = \begin{pmatrix} F_1^{0\prime}\Lambda^{0\prime}e'\tilde{F}^k \\ \vdots \\ F_T^{0\prime}\Lambda^{0\prime}e'\tilde{F}^k \end{pmatrix},$$

so by the definition of the trace norm,

$$\begin{aligned}\left\|F^0\Lambda^{0\prime}e'\tilde{F}^k\right\|^2 &= \sum_{t=1}^{T}\left|F_t^{0\prime}\Lambda^{0\prime}e'\tilde{F}^k\right|^2 \\ &\leq \left(\sum_{t=1}^{T}\left|F_t^0\right|^2\right)\cdot\left\|\Lambda^{0\prime}e'\right\|^2\cdot\left\|\tilde{F}^k\right\|^2 \\ &= \operatorname{tr}(F^{0\prime}F^0)\cdot\left\|\Lambda^{0\prime}e'\right\|^2\cdot\operatorname{tr}(\tilde{F}^{k\prime}\tilde{F}^k) = kT\cdot\operatorname{tr}(F^{0\prime}F^0)\cdot\left\|\Lambda^{0\prime}e'\right\|^2.\end{aligned}$$

The columns of $\Lambda^{0\prime}e'$ are given by

$$\Lambda^{0\prime}e' = \begin{pmatrix}\Lambda^{0\prime}e_1 & \cdots\Lambda^{0\prime}e_T\end{pmatrix},$$

so the trace norm once again tells us that

$$\left\|\Lambda^{0\prime}e'\right\|^2 = \sum_{t=1}^{T}\left|\Lambda^{0\prime}e_t\right|^2 = \sum_{t=1}^{T}\left|\sum_{i=1}^{N}\lambda_i^0 e_{it}\right|^2.$$

Therefore,

$$\begin{aligned}\frac{1}{T}\left\|\frac{1}{NT}F^0\Lambda^{0\prime}e'\tilde{F}^k\right\|^2 &= \frac{1}{N^2T^3}\left\|F^0\Lambda^{0\prime}e'\tilde{F}^k\right\|^2 \\ &\leq \frac{k}{N}\cdot\operatorname{tr}\left(\frac{F^{0\prime}F^0}{T}\right)\cdot\left(\frac{1}{NT}\sum_{t=1}^{T}\left|\sum_{i=1}^{N}\lambda_i^0 e_{it}\right|^2\right).\end{aligned}$$

Once again, the two rightmost terms are $O_p(1)$, so

$$\frac{1}{T}\left\|\frac{1}{NT}F^0\Lambda^{0\prime}e'\tilde{F}^k\right\|^2 = O_p\left(\frac{1}{N}\right).$$

3) $\frac{1}{T}\left\|\frac{1}{NT}ee'\tilde{F}^k\right\|^2$
Note that

$$ee'\tilde{F}^k = \begin{pmatrix} e_1'e'\tilde{F}^k \\ \vdots \\ e_T'e'\tilde{F}^k, \end{pmatrix}$$

so that

$$\left\|ee'\tilde{F}^k\right\|^2 = \sum_{t=1}^{T}\left|e_t'e'\tilde{F}^k\right|^2.$$

For any $t \in N_+$,

$$\begin{aligned}
e_t'e'\tilde{F}^k &= e_t'\sum_{s=1}^{T} e_s\tilde{F}_s^{k\prime} \\
&= \sum_{s=1}^{T}\left(\sum_{i=1}^{N} e_{it}e_{is}\right)\tilde{F}_s^{k\prime} \\
&= \sum_{s=1}^{T}\sum_{i=1}^{N}(e_{it}e_{is} - \gamma(t-s))\tilde{F}_s^{k\prime} + \sum_{s=1}^{T}\sum_{i=1}^{N}\gamma(t-s)\tilde{F}_s^{k\prime} \\
&= \sum_{s=1}^{T}\sum_{i=1}^{N}(e_{it}e_{is} - \gamma(t-s))\tilde{F}_s^{k\prime} + N\cdot\sum_{s=1}^{T}\gamma(t-s)\tilde{F}_s^{k\prime}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\left|e_t'e'\tilde{F}^k\right|^2 &\leq 2\left[\left|\sum_{s=1}^{T}\sum_{i=1}^{N}\tilde{F}_s^k(e_{it}e_{is} - \gamma(t-s))\right|^2 + N^2\cdot\left|\sum_{s=1}^{T}\gamma(t-s)\tilde{F}_s^{k\prime}\right|^2\right] \\
&\leq 2\left[\left|\sum_{s=1}^{T}\sum_{i=1}^{N}\tilde{F}_s^k(e_{it}e_{is} - \gamma(t-s))\right|^2 + N^2\cdot\left(\sum_{s=1}^{T}\gamma(t-s)^2\right)\left(\sum_{s=1}^{T}\left|\tilde{F}_s^k\right|^2\right)\right] \\
&\leq 2\left[\left|\sum_{s=1}^{T}\sum_{i=1}^{N}\tilde{F}_s^k(e_{it}e_{is} - \gamma(t-s))\right|^2 + kN^2T\cdot Z\right],
\end{aligned}$$

where we used the Cauchy-Schwarz inequality to justify the second inequality, and we write

$$Z = \sum_{z=-\infty}^{\infty}\gamma(z)^2,$$

which is finite because the autocovariances $\gamma(\cdot)$ are absolutely summable and absolute summability implies square summability.

This holds for any $t \in N_+$, so

$$\frac{1}{T}\left\|\frac{1}{NT}ee'\tilde{F}^k\right\|^2 = \frac{1}{N^2T^3}\left\|ee'\tilde{F}^k\right\| \le \frac{1}{N^2T^3}\sum_{t=1}^{T}\left|e_t'e'\tilde{F}^k\right|^2$$
$$\le 2\frac{1}{N^2T^3}\sum_{t=1}^{T}\left|\sum_{s=1}^{T}\sum_{i=1}^{N}\tilde{F}_s^k(e_{it}e_{is}-\gamma(t-s))\right|^2 + 2\frac{1}{N^2T^3}\sum_{t=1}^{T}kN^2T\cdot Z$$
$$\le 2\frac{1}{N^2T^3}\sum_{t=1}^{T}\left(\sum_{s=1}^{T}\left|\tilde{F}_s^k\right|\cdot\left|\sum_{i=1}^{N}(e_{it}e_{is}-\gamma(t-s))\right|\right)^2 + 2\cdot\frac{1}{T}kZ.$$

By the Cauchy-Schwarz inequality,

$$\frac{1}{NT^2}\left(\sum_{s=1}^{T}\left|\tilde{F}_s^k\right|\left|\sum_{i=1}^{N}(e_{it}e_{is}-\gamma(t-s))\right|\right)^2 \le \left(\frac{1}{T}\sum_{s=1}^{T}\left|\tilde{F}_s^k\right|^2\right)\left(\frac{1}{NT}\sum_{s=1}^{T}\left|\sum_{i=1}^{N}(e_{it}e_{is}-\gamma(t-s))\right|^2\right)$$
$$= k\cdot\frac{1}{NT}\sum_{s=1}^{T}\left|\sum_{i=1}^{N}(e_{it}e_{is}-\gamma(t-s))\right|^2.$$

Since

$$\mathbb{E}\left[\frac{1}{NT}\sum_{s=1}^{T}\left|\sum_{i=1}^{N}(e_{it}e_{is}-\gamma(t-s))\right|^2\right] = \frac{1}{NT}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}\left[(e_{it}e_{is}-\gamma(s-t))(e_{jt}e_{js}-\gamma(s-t))\right]$$
$$= \frac{1}{NT}\sum_{s=1}^{T}\sum_{i=1}^{N}\mathbb{E}\left[(e_{it}e_{is}-\gamma(s-t))^2\right]$$
$$= \frac{1}{NT}\sum_{s=1}^{T}\sum_{i=1}^{N}\left[\mathbb{E}\left[e_{it}^2e_{is}^2\right]-\gamma(s-t)^2\right] \le \frac{1}{NT}\sum_{s=1}^{T}\sum_{i=1}^{N}\mu_4 = \mu_4,$$

we can see that

$$\mathbb{E}\left[\frac{1}{N^2T^3}\sum_{t=1}^{T}\left(\sum_{s=1}^{T}\left|\tilde{F}_s^k\right|\cdot\left|\sum_{i=1}^{N}(e_{it}e_{is}-\gamma(t-s))\right|\right)^2\right]$$
$$= \frac{1}{NT}\sum_{t=1}^{T}\mathbb{E}\left[\frac{1}{NT^2}\left(\sum_{s=1}^{T}\left|\tilde{F}_s^k\right|\left|\sum_{i=1}^{N}(e_{it}e_{is}-\gamma(t-s))\right|\right)^2\right] \le \frac{k\cdot\mu_4}{N},$$

so

$$\frac{1}{N^2T^3}\sum_{t=1}^{T}\left(\sum_{s=1}^{T}\left|\tilde{F}_s^k\right|\cdot\left|\sum_{i=1}^{N}(e_{it}e_{is}-\gamma(t-s))\right|\right)^2 = O_p\left(\frac{1}{N}\right).$$

We can finally see that

$$\frac{1}{T}\left\|\frac{1}{NT}ee'\tilde{F}^k\right\|^2 \le O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{T}\right).$$

It follows that

$$\frac{1}{T}\sum_{t=1}^{T}\left|\hat{F}_t^k - H^{k\prime}F_t^0\right|^2 = \frac{1}{T}\left\|\hat{F}^k - F^0\cdot H^k\right\|^2$$
$$\leq O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{T}\right),$$

and as such,

$$\frac{1}{T}\sum_{t=1}^{T}\left|\hat{F}_t^k - H^{k\prime}F_t^0\right|^2 = \frac{1}{T}\left\|\hat{F}^k - F^0\cdot H^k\right\|^2 = O_p\left(\delta_{NT}^{-1}\right),$$

where $\delta_{NT} = \min(N,T)$.

Therefore, even if the true number of factors is unknown, the mean squared deviation of the estimated factors from some linear combination of the true factors converges at rate $\min(N,T)$, provided that the factors are suitably scaled.

1.4 Information Criteria for the Number of Factors

As the number of factors k increases, the value of the minimzed objective function decreases. To see this, for any $T \in N_+$ let $\mu_1 \geq \cdots \geq \mu_T \geq 0$ be the ordered eigenvalues of XX'. Then, for any $1 \leq k < k_{max}$,

$$V(k,\tilde{F}^k) = \frac{1}{NT}\operatorname{tr}(XX') - \frac{1}{NT}\sum_{i=1}^{k}\mu_i = \frac{1}{NT}\sum_{i=k+1}^{T}\mu_i$$
$$> \frac{1}{NT}\sum_{i=k+2}^{T}\mu_i = \frac{1}{NT}\operatorname{tr}(XX') - \frac{1}{NT}\sum_{i=1}^{k+1}\mu_i = V(k+1,\tilde{F}^{k+1}),$$

where the inequality is strict because $\mu_{k+1} > 0$ by assumption 7. This indicates that, the larger the difference in the kth and $k+1$th eigenvalues of XX', the smaller the difference between $V(k,\tilde{F}^k)$ and $V(k+1,\tilde{F}^{k+1})$, which in turn implies that it is likely that there are k true factors. As such, an intuitive approach to determining the number of factors could be to choose the $1 \leq k < k_{max}$ such that the ratio $\frac{\mu_k}{\mu_{k+1}}$; indeed, this is the approach taken in Ahn and Hornstein (2013).

In contrast, Bai and Ng suggest using $V(k,\tilde{F}^k)$ in a role similar to the estimated error variance in traditional information criteria such as the AIC or BIC when constructing their informatio criteria, since it can be interpreted as the sum of squared residuals in a traditional least squares context. Specifically, they suggest using the criterion

$$PC(k) = V(k,\tilde{F}^k) + kg(N,T),$$

for $1 \leq k \leq k_{max}$, where $g(N,T)$ is a penalty term representing the inefficiency that arises as more factors are included into the model.

Because V^k_{NT} is nonsingular under our assumptions, substituting $\hat{F}^k = \tilde{F}^k V^k_{NT}$ into the function $V(k,\cdot)$ also yields the same value as $\tilde{F}^k$, that is,

$$V(k,\hat{F}^k) = V(k,\tilde{F}^k).$$

Thus, the information criterion can be reformulated as

$$PC(k) = V(k,\hat{F}^k) + kg(N,T).$$

This representation will help us derive the asymptotic properties of the information criterion more effectively, since our asymptotic results above were formulated in terms of $\hat{F}^k$ instead of $\tilde{F}^k$.

We now show the conditions that the penalty function $g(N,T)$ must satisfy in order for the number of factors derived using the above criterion to consistently estimate the true number of factors r; that is, we show the conditions under which

$$\mathbb{P}(k^* \neq r) \to 0$$

as $N,T \to \infty$, where k^* is the number of factors chosen by $PC(k)$.

We first show some preliminary results before moving onto the proof.

1.4.1 Preliminary Result 1

We want to show that:

$$\textbf{If } 1 \leq k \leq r\textbf{, then } V(k,\hat{F}^k) - V(k,F^0H^k) = O_p(\delta_{NT}^{-1/2})\textbf{.}$$

Let $1 \leq k \leq r$. Note that

$$\begin{aligned} V(k,\hat{F}^k) &= \frac{1}{NT}\,\mathrm{tr}\left(X'\left[I_T - \hat{F}^k\left(\hat{F}^{k\prime}\hat{F}^k\right)^{-1}\hat{F}^{k\prime}\right]X\right) \\ &= \frac{1}{NT}\sum_{i=1}^{N}\underline{X}_i'\left(I_T - P_{\hat{F}}^k\right)\underline{X}_i, \end{aligned}$$

where $P_{\hat{F}}^k = \hat{F}^k\left(\hat{F}^{k\prime}\hat{F}^k\right)^{-1}\hat{F}^{k\prime}$. Likewise, because

$$\begin{aligned} V(k,F^0H^k) &= \frac{1}{NT}\,\mathrm{tr}\left(X'\left[I_T - F^0H^k\left(H^{k\prime}F^{0\prime}F^0H^k\right)^{-1}H^{k\prime}F^{0\prime}\right]X\right) \\ &= \frac{1}{NT}\sum_{i=1}^{N}\underline{X}_i'\left(I_T - P_{FH}^k\right)\underline{X}_i, \end{aligned}$$

where $P_{FH}^k = F^0H^k\left(H^{k\prime}F^{0\prime}F^0H^k\right)^{-1}H^{k\prime}F^{0\prime}$, it follows that

$$V(k,\hat{F}^k) - V(k,F^0H^k) = \frac{1}{NT}\sum_{i=1}^{N}\underline{X}_i'\left(P_{FH}^k - P_{\hat{F}}^k\right)\underline{X}_i.$$

Define $D_k = \frac{\hat{F}^{k\prime}\hat{F}^k}{T}$ and $D_0 = \frac{H^{k\prime}F^{0\prime}F^0H^k}{T}$. Since $\hat{F}^k = \tilde{F}^kV_{NT}^k$ and $\frac{\tilde{F}^{k\prime}\tilde{F}^k}{T} = I_k$, we have

$$D_k = \frac{\hat{F}^{k\prime}\hat{F}^k}{T} = \left(V_{NT}^k\right)^2,$$

and by the definition of H^k,

$$D_0 = \left(\frac{\tilde{F}^{k\prime}F^0}{T}\right)\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)\left(\frac{F^{0\prime}F^0}{T}\right)\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)\left(\frac{F^{0\prime}\tilde{F}^k}{T}\right).$$

Note that D_0 is $O_p(1)$, since H^k, $\frac{F^{0\prime}F^0}{T}$ and $\frac{F^{0\prime}\tilde{F}^k}{T}$ are $O_p(1)$ by assumption. Specifically,

$$\frac{F^{0\prime}\tilde{F}^k}{T} \xrightarrow{p} Q^k,$$

where Q^k has rank $k = \min(r,k)$ by assumption. and $\frac{\Lambda^{0\prime}\Lambda^0}{N}$, $\frac{F^{0\prime}F^0}{T}$ converge in probability to positive definite matrices Σ_Λ and Σ_F, so

$$D_0 \xrightarrow{p} L = Q^{k\prime}\Sigma_\Lambda\Sigma_F\Sigma_\Lambda Q^k,$$

where the $k \times k$ matrix on the right hand side has full rank and is thus positive definite.

Note that $D_k - D_0$ can be decomposed as

$$\begin{aligned}D_k - D_0 &= \frac{\hat{F}^{k\prime}\hat{F}^k}{T} - \frac{H^{k\prime}F^{0\prime}F^0H^k}{T} \\ &= \frac{1}{T}(\hat{F}^k - F^0H^k)'(\hat{F}^k - F^0H^k) + \frac{1}{T}(\hat{F}^k - F^0H^k)'F^0H^k + \frac{1}{T}H^{k\prime}F^{0\prime}(\hat{F}^k - F^0H^k),\end{aligned}$$

so that, by the triangle inequality,

$$\begin{aligned}\|D_k - D_0\| &\leq \frac{1}{T}\left\|\hat{F}^k - F^0H^k\right\|^2 + 2\left(\frac{1}{\sqrt{T}}\left\|\hat{F}^k - F^0H^k\right\|\right)\cdot\left(\frac{1}{\sqrt{T}}\left\|F^0\right\|\right)\cdot\left\|H^k\right\| \\ &\leq \frac{1}{T}\left\|\hat{F}^k - F^0H^k\right\|^2 + 2\left(\frac{1}{\sqrt{T}}\left\|\hat{F}^k - F^0H^k\right\|\right)\cdot \mathrm{tr}\left(\frac{F^{0\prime}F^0}{T}\right)^{\frac{1}{2}}\cdot\left\|H^k\right\|.\end{aligned}$$

From the results studied earlier, we can now say that

$$\|D_k - D_0\| \leq O_p\left(\frac{1}{\min(N,T)}\right) + O_p\left(\frac{1}{\min(\sqrt{N},\sqrt{T})}\right) = O_p(\delta_{NT}^{-1/2}).$$

It follows from this discovery that, since

$$\|D_k - L\| \leq \|D_0 - D_k\| + \|D_0 - L\|,$$

where $\|D_0 - D_k\| = O_p(\delta_{NT}^{-1/2})$ and $\|D_0 - L\| = o_p(1)$, it follows that $\|D_k - L\| = o_p(1)$, that is, $D_k \xrightarrow{p} L$. Since L is positive definite, by the CMT we have $D_0^{-1} \xrightarrow{p} L^{-1}$ and $D_k^{-1} \xrightarrow{p} L^{-1}$; $D_0^{-1} = O_p(1)$ and $D_k^{-1} = O_p(1)$.
Furthermore, the relationship

$$\left\|D_k^{-1} - D_0^{-1}\right\| = \left\|D_k^{-1}(D_k - D_0)D_0^{-1}\right\| \leq \left\|D_k^{-1}\right\|\left\|D_0^{-1}\right\|\|D_k - D_0\|$$

tells us that $\left\|D_k^{-1} - D_0^{-1}\right\|$ is $O_p(\delta_{NT}^{-1/2})$.

Expanding the kernel $\left(P^k_{FH} - P^k_{\hat{F}}\right)$, we can see that

$$\begin{aligned}
P^k_{FH} - P^k_{\hat{F}} &= F^0 H^k \left(H^{k\prime} F^{0\prime} F^0 H^k\right)^{-1} H^{k\prime} F^{0\prime} - \hat{F}^k \left(\hat{F}^{k\prime}\hat{F}^k\right)^{-1} \hat{F}^{k\prime} \\
&= \frac{1}{T} F^0 H^k D_0^{-1} H^{k\prime} F^{0\prime} - \frac{1}{T} \hat{F}^k D_k^{-1} \hat{F}^{k\prime} \\
&= \frac{1}{T} F^0 H^k D_0^{-1} H^{k\prime} F^{0\prime} - \frac{1}{T} \left[\hat{F}^k - F^0 H^k + F^0 H^k\right] D_k^{-1} \left[\hat{F}^k - F^0 H^k + F^0 H^k\right]' \\
&= \frac{1}{T} F^0 H^k D_0^{-1} H^{k\prime} F^{0\prime} - \frac{1}{T} \left(\hat{F}^k - F^0 H^k\right) D_k^{-1} \left(\hat{F}^k - F^0 H^k\right)' - \frac{1}{T} \left(\hat{F}^k - F^0 H^k\right) D_k^{-1} H^{k\prime} F^{0\prime} \\
&\quad - \frac{1}{T} F^0 H^k D_k^{-1} \left(\hat{F}^k - F^0 H^k\right)' - \frac{1}{T} F^0 H^k D_k^{-1} H^{k\prime} F^{0\prime} \\
&= \frac{1}{T} F^0 H^k \left(D_0^{-1} - D_k^{-1}\right) H^{k\prime} F^{0\prime} - \frac{1}{T} \left(\hat{F}^k - F^0 H^k\right) D_k^{-1} \left(\hat{F}^k - F^0 H^k\right)' \\
&\quad - \frac{1}{T} \left(\hat{F}^k - F^0 H^k\right) D_k^{-1} H^{k\prime} F^{0\prime} - \frac{1}{T} F^0 H^k D_k^{-1} \left(\hat{F}^k - F^0 H^k\right)'.
\end{aligned}$$

Therefore,

$$\begin{aligned}
V(k, \hat{F}^k) - V(k, F^0 H^k) &= \frac{1}{NT} \sum_{i=1}^{N} \underline{X}_i' \left(P^k_{FH} - P^k_{\hat{F}}\right) \underline{X}_i \\
&= \frac{1}{NT^2} \sum_{i=1}^{N} \underline{X}_i' F^0 H^k \left(D_0^{-1} - D_k^{-1}\right) H^{k\prime} F^{0\prime} \underline{X}_i \\
&\quad - \frac{1}{NT^2} \sum_{i=1}^{N} \underline{X}_i' \left(\hat{F}^k - F^0 H^k\right) D_k^{-1} \left(\hat{F}^k - F^0 H^k\right)' \underline{X}_i \\
&\quad - \frac{1}{NT^2} \sum_{i=1}^{N} \underline{X}_i' \left(\hat{F}^k - F^0 H^k\right) D_k^{-1} H^{k\prime} F^{0\prime} \underline{X}_i \\
&\quad - \frac{1}{NT^2} \sum_{i=1}^{N} \underline{X}_i' F^0 H^k D_k^{-1} \left(\hat{F}^k - F^0 H^k\right)' \underline{X}_i,
\end{aligned}$$

so that

$$\begin{aligned}
\left|V(k, \hat{F}^k) - V(k, F^0 H^k)\right| \leq \left(\frac{1}{NT} \sum_{i=1}^{N} |\underline{X}_i|^2\right) [&\frac{1}{T} \left\|F^0\right\|^2 \cdot \left\|H^k\right\|^2 \cdot \left\|D_0^{-1} - D_k^{-1}\right\| \\
&+ \frac{1}{T} \left\|\hat{F}^k - F^0 H^k\right\|^2 \cdot \left\|D_k^{-1}\right\| \\
&+ 2 \cdot \left(\frac{1}{\sqrt{T}} \left\|\hat{F}^k - F^0 H^k\right\|\right) \left(\frac{1}{\sqrt{T}} \left\|F^0\right\|\right) \cdot \left\|H^k\right\|].
\end{aligned}$$

Since

$$
\begin{aligned}
\frac{1}{\sqrt{T}}\left\|\hat{F}^k - F^0 H^k\right\| &= O_p\left(\delta_{NT}^{-1/2}\right), \\
\left\|D_0^{-1} - D_k^{-1}\right\| &= O_p\left(\delta_{NT}^{-1/2}\right) \\
\frac{1}{\sqrt{T}}\left\|F^0\right\| = \operatorname{tr}\left(\frac{F^{0\prime}F^0}{T}\right)^{\frac{1}{2}} &= O_p(1) \\
\left\|D_k^{-1}\right\| &= O_p(1) \\
\frac{1}{NT}\sum_{i=1}^{N}|\underline{X}_i|^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}^2 &= O_p(1),
\end{aligned}
$$

we can see that

$$
\left|V(k,\hat{F}^k) - V(k,F^0H^k)\right| \leq O_p\left(\delta_{NT}^{-1/2}\right) + O_p\left(\delta_{NT}^{-1}\right) = O_p\left(\delta_{NT}^{-1/2}\right),
$$

and as such

$$
V(k,\hat{F}^k) - V(k,F^0H^k) = O_p\left(\delta_{NT}^{-1/2}\right).
$$

1.4.2 Preliminary Result 2

Here we want to show that:

$$\textbf{For any } 1 \le k < r\textbf{, there exists a } \tau_k > 0 \textbf{ such that}$$
$$\lim_{N,T\to\infty} \mathbb{P}\left(V(k, F^0 H^k) - V(r, F^0) \ge \tau_k\right) = 1.$$

Let $1 \le k < r$, and define $P_F^r = F^0(F^{0\prime}F^0)^{-1}F^{0\prime}$.
Since $k < r$, by assumption $\frac{F^{0\prime}\tilde{F}^k}{T} \xrightarrow{p} Q^k$ to a $r \times k$ matrix Q^k with rank $k < r$, and as such

$$H^{k\prime} = \left(\frac{\tilde{F}^{k\prime}F^0}{T}\right)\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right) \xrightarrow{p} Q^{k\prime}\Sigma_\Lambda,$$

where $Q^{k\prime}\Sigma_\Lambda$ has rank $k < r$. Note that

$$\begin{aligned}\left\|P_F^r - P_{FH}^k\right\| &= \left\|\frac{1}{\sqrt{T}}F^0\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\frac{1}{\sqrt{T}}F^{0\prime} - \frac{1}{\sqrt{T}}F^0H^k\left(H^{k\prime}\frac{F^{0\prime}F^0}{T}H^k\right)^{-1}H^{k\prime}\frac{1}{\sqrt{T}}F^{0\prime}\right\| \\ &\le \frac{1}{T}\left\|F^0\right\|^2 \cdot \left\|\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\right\| \cdot \left\|I_r - \left(\frac{F^{0\prime}F^0}{T}\right)^{\frac{1}{2}\prime} H^k\left(H^{k\prime}\frac{F^{0\prime}F^0}{T}H^k\right)^{-1}H^{k\prime}\left(\frac{F^{0\prime}F^0}{T}\right)^{\frac{1}{2}}\right\|,\end{aligned}$$

where $\left(\frac{F^{0\prime}F^0}{T}\right)^{\frac{1}{2}}$ is the Cholesky factor of $\frac{F^{0\prime}F^0}{T}$.
Since

$$H^{k\prime}\frac{F^{0\prime}F^0}{T}H^k \xrightarrow{p} Q^{k\prime}\Sigma_\Lambda\Sigma_F\Sigma_\Lambda Q^k = L,$$

where L has rank k,

$$\left(H^{k\prime}\frac{F^{0\prime}F^0}{T}H^k\right)^{-1} \xrightarrow{p} L^{-1}$$

and

$$B_{NT} = \left(\frac{F^{0\prime}F^0}{T}\right)^{\frac{1}{2}\prime} H^k\left(H^{k\prime}\frac{F^{0\prime}F^0}{T}H^k\right)^{-1}H^{k\prime}\left(\frac{F^{0\prime}F^0}{T}\right)^{\frac{1}{2}} \xrightarrow{p} \Sigma_F^{\frac{1}{2}\prime}\Sigma_\Lambda Q^k L^{-1}Q^{k\prime}\Sigma_\Lambda\Sigma_F^{\frac{1}{2}},$$

where $\Sigma_F^{\frac{1}{2}}$ is the Cholesky factor of Σ_F and $\left(\frac{F^{0\prime}F^0}{T}\right)^{\frac{1}{2}} \xrightarrow{p} \Sigma_F^{\frac{1}{2}}$ by the CMT and the continuity of the Cholesky operation.
Defining the $r \times k$ matrix $P = \Sigma_F^{\frac{1}{2}\prime}\Sigma_\Lambda Q^k$, we can see that

$$B_{NT} \xrightarrow{p} P(P'P)^{-1}P'.$$

Letting $M = I_r - P(P'P)^{-1}P'$, because M is symmetric and idempotent, its trace equals its rank; therefore,

$$\operatorname{rank}(M)=\operatorname{tr}(M)=r-\operatorname{tr}(P(P'P)^{-1}P')=r-\operatorname{tr}(I_k)=r-k>0$$

by the assumption that $1\leq k<r$. This means that

$$I_r-B_{NT}\xrightarrow{p}M,$$

and by implication

$$\|I_k-B_{NT}\|\xrightarrow{p}\|M\|=\operatorname{tr}(M'M)^{\frac{1}{2}}=\operatorname{tr}(M)^{\frac{1}{2}}=\sqrt{r-k}.$$

Since

$$\frac{1}{T}\left\|F^0\right\|^2=\operatorname{tr}\left(\frac{F^{0\prime}F^0}{T}\right)\xrightarrow{p}\operatorname{tr}(\Sigma_F)$$

by the CMT and the continuity of the trace operation,

$$\left(\frac{1}{T}\left\|F^0\right\|^2\right)\cdot\left\|\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\right\|\cdot\|I_r-B_{NT}\|\xrightarrow{p}\operatorname{tr}(\Sigma_F)\left\|\Sigma_F^{-1}\right\|\sqrt{r-k}>0.$$

and as such $\left\|P_F^r-P_{FH}^k\right\|=O_p(1)$.

$V(k,F^0H^k)-V(r,F^0)$ can be decomposed as

$$\begin{aligned}V(k,F^0H^k)-V(r,F^0)&=\frac{1}{NT}\sum_{i=1}^{N}\underline{X}_i'(I_T-P_{FH}^k)\underline{X}_i-\frac{1}{NT}\sum_{i=1}^{N}\underline{X}_i'(I_T-P_F^r)\underline{X}_i\\&=\frac{1}{NT}\operatorname{tr}\left(X'\left(P_F^r-P_{FH}^k\right)X\right)\\&=\frac{1}{NT}\operatorname{tr}\left(\left(F^0\Lambda^{0\prime}+e\right)'\left(P_F^r-P_{FH}^k\right)\left(F^0\Lambda^{0\prime}+e\right)\right)\\&=\operatorname{tr}\left(\frac{1}{NT}\Lambda^0F^{0\prime}\left(P_F^r-P_{FH}^k\right)F^0\Lambda^{0\prime}\right)\\&\quad+2\operatorname{tr}\left(\frac{1}{NT}\Lambda^0F^{0\prime}\left(P_F^r-P_{FH}^k\right)e\right)+\operatorname{tr}\left(\frac{1}{NT}e'\left(P_F^r-P_{FH}^k\right)e\right).\end{aligned}$$

For any $N,T\in N_+$,

$$\operatorname{tr}\left(\frac{1}{NT}e'\left(P_F^r-P_{FH}^k\right)e\right)=\frac{1}{NT}\sum_{i=1}^{N}\underline{e}_i'\left(P_F^r-P_{FH}^k\right)\underline{e}_i\geq 0,$$

since $P_F^r-P_{FH}^k$ is positive semidefinite. Thus,

$$V(k,F^0H^k)-V(r,F^0)\geq\operatorname{tr}\left(\frac{1}{NT}\Lambda^0F^{0\prime}\left(P_F^r-P_{FH}^k\right)F^0\Lambda^{0\prime}\right)+2\operatorname{tr}\left(\frac{1}{NT}\Lambda^0F^{0\prime}\left(P_F^r-P_{FH}^k\right)e\right).$$

We examine each term in turn:

1) $\operatorname{tr}\left(\frac{1}{NT}\Lambda^0 F^{0\prime}\left(P_F^r - P_{FH}^k\right) F^0 \Lambda^{0\prime}\right)$

Note that

$$\operatorname{tr}\left(\frac{1}{NT}\Lambda^0 F^{0\prime}\left(P_F^r - P_{FH}^k\right) F^0 \Lambda^{0\prime}\right) = \operatorname{tr}\left(\left[\frac{1}{T}F^{0\prime}\left(P_F^r - P_{FH}^k\right) F^0\right] \cdot \frac{\Lambda^{0\prime}\Lambda^0}{N}\right).$$

We now have

$$\begin{aligned}\frac{1}{T}F^{0\prime}\left(P_F^r - P_{FH}^k\right) F^0 &= \frac{F^{0\prime}F^0}{T} - \frac{F^{0\prime}F^0}{T}H^k\left(\frac{H^{k\prime}F^{0\prime}F^0H^k}{T}\right)^{-1}H^{k\prime}\frac{F^{0\prime}F^0}{T} \\ &\xrightarrow{p} \Sigma_F - \Sigma_F\Sigma_\Lambda Q^k L^{-1} Q^{k\prime}\Sigma_\Lambda\Sigma_F,\end{aligned}$$

where

$$\begin{aligned}\Sigma_F - \Sigma_F\Sigma_\Lambda Q^k L^{-1} Q^{k\prime}\Sigma_\Lambda\Sigma_F &= \Sigma_F^{\frac{1}{2}}\left(I_r - \Sigma_F^{\frac{1}{2}\prime}\Sigma_\Lambda Q^k L^{-1} Q^{k\prime}\Sigma_\Lambda\Sigma_F^{\frac{1}{2}}\right)\Sigma_F^{\frac{1}{2}\prime} \\ &= \Sigma_F^{\frac{1}{2}} M \Sigma_F^{\frac{1}{2}\prime}.\end{aligned}$$

By the CMT and the continuity of the trace operation,

$$\operatorname{tr}\left(\frac{1}{NT}\Lambda^0 F^{0\prime}\left(P_F^r - P_{FH}^k\right) F^0 \Lambda^{0\prime}\right) \xrightarrow{p} \operatorname{tr}\left(\Sigma_F^{\frac{1}{2}} M \Sigma_F^{\frac{1}{2}\prime}\Sigma_\Lambda\right).$$

Denoting the above limit by τ_k (the limit depends on k through M), we can see that

$$\tau_k = \operatorname{tr}\left(\Sigma_F^{\frac{1}{2}} M \Sigma_F^{\frac{1}{2}\prime}\Sigma_\Lambda\right) = \operatorname{tr}\left(\Sigma_\Lambda^{\frac{1}{2}\prime}\Sigma_F^{\frac{1}{2}} M \Sigma_F^{\frac{1}{2}\prime}\Sigma_\Lambda^{\frac{1}{2}}\right).$$

The matrix $\Sigma_\Lambda^{\frac{1}{2}\prime}\Sigma_F^{\frac{1}{2}} M \Sigma_F^{\frac{1}{2}\prime}\Sigma_\Lambda^{\frac{1}{2}}$ is symmetric positive semidefinite, so its trace τ_k is the sum of all its (non-negative) eigenvalues (this follows by using the eigendecomposition of the matrix).
We now establish the rank of $\Sigma_\Lambda^{\frac{1}{2}\prime}\Sigma_F^{\frac{1}{2}} M \Sigma_F^{\frac{1}{2}\prime}\Sigma_\Lambda^{\frac{1}{2}}$.
For any $\alpha \in \mathbb{R}^r$, if

$$\Sigma_\Lambda^{\frac{1}{2}\prime}\Sigma_F^{\frac{1}{2}} M \Sigma_F^{\frac{1}{2}\prime}\Sigma_\Lambda^{\frac{1}{2}}\alpha = \mathbf{0},$$

then

$$\left(\Sigma_F^{\frac{1}{2}\prime}\Sigma_\Lambda^{\frac{1}{2}}\alpha\right)' M \left(\Sigma_F^{\frac{1}{2}\prime}\Sigma_\Lambda^{\frac{1}{2}}\alpha\right) = 0$$

as well. Denoting $\beta = \Sigma_F^{\frac{1}{2}\prime}\Sigma_\Lambda^{\frac{1}{2}}\alpha \neq \mathbf{0}$, because M is symmetric and idempotent, the above equality implies that

$$(M\beta)'(M\beta) = 0,$$

or that $M\beta = M\Sigma_F^{\frac{1}{2}\prime}\Sigma_\Lambda^{\frac{1}{2}}\alpha = \mathbf{0}$.
Conversely, if $M\Sigma_F^{\frac{1}{2}\prime}\Sigma_\Lambda^{\frac{1}{2}}\alpha = \mathbf{0}$, then

$$\Sigma_\Lambda^{\frac{1}{2}\prime}\Sigma_F^{\frac{1}{2}}M\Sigma_F^{\frac{1}{2}\prime}\Sigma_\Lambda^{\frac{1}{2}}\alpha = \mathbf{0}$$

trivially, so that the null spaces of $\Sigma_\Lambda^{\frac{1}{2}\prime}\Sigma_F^{\frac{1}{2}}M\Sigma_F^{\frac{1}{2}\prime}\Sigma_\Lambda^{\frac{1}{2}}$ and $M\Sigma_F^{\frac{1}{2}\prime}\Sigma_\Lambda^{\frac{1}{2}}$ are equivalent. The two matrices are both $r \times r$ matrices, so this implies that they have the same rank.

Since a matrix and its transpose has the same rank, $\Sigma_\Lambda^{\frac{1}{2}\prime}\Sigma_F^{\frac{1}{2}}M\Sigma_F^{\frac{1}{2}\prime}\Sigma_\Lambda^{\frac{1}{2}}$ has the same rank as $\Sigma_\Lambda^{\frac{1}{2}\prime}\Sigma_F^{\frac{1}{2}}M$. By the same line of reasoning as above, we can show that $\Sigma_\Lambda^{\frac{1}{2}\prime}\Sigma_F^{\frac{1}{2}}M$ has the same rank as M, and by extension $\Sigma_\Lambda^{\frac{1}{2}\prime}\Sigma_F^{\frac{1}{2}}M\Sigma_F^{\frac{1}{2}\prime}\Sigma_\Lambda^{\frac{1}{2}}$. Therefore,

$$\text{rank}\left(\Sigma_\Lambda^{\frac{1}{2}\prime}\Sigma_F^{\frac{1}{2}}M\Sigma_F^{\frac{1}{2}\prime}\Sigma_\Lambda^{\frac{1}{2}}\right) = \text{rank}(M) = r - k > 0,$$

which implies that it has at least one non-zero eigenvalue and therefore that its trace τ_k is positive.

Therefore,

$$\text{tr}\left(\frac{1}{NT}\Lambda^0 F^{0\prime}\left(P_F^r - P_{FH}^k\right)F^0\Lambda^{0\prime}\right) \xrightarrow{p} \tau_k > 0.$$

2) $\operatorname{tr}\left(\frac{1}{NT}\Lambda^0 F^{0\prime}\left(P_F^r - P_{FH}^k\right)e\right)$

Note that

$$\operatorname{tr}\left(\frac{1}{NT}\Lambda^0 F^{0\prime}\left(P_F^r - P_{FH}^k\right)e\right) = \operatorname{tr}\left(\frac{1}{NT}e\Lambda^0 F^{0\prime}\left(P_F^r - P_{FH}^k\right)\right).$$

Since

$$e\Lambda^0 = \begin{pmatrix} e_1'\Lambda^0 \\ \vdots \\ e_T'\Lambda^0 \end{pmatrix}$$

and

$$e_t'\Lambda^0 = \sum_{i=1}^{N} e_{it}\lambda_i^{0\prime}$$

for any $t \in N_+$, we have

$$\left\|e\Lambda^0\right\|^2 = \sum_{t=1}^{T}\left|e_t'\Lambda^0\right|^2 = \sum_{t=1}^{T}\left|\sum_{i=1}^{N}\lambda_i^0 e_{it}\right|^2.$$

This allows us to see that

$$\begin{aligned}\left\|\frac{1}{NT}e\Lambda^0 F^{0\prime}\right\|^2 &= \frac{1}{N^2T^2}\sum_{t=1}^{T}\left|e\Lambda^0 F_t^0\right|^2 \\ &\leq \left(\frac{1}{T}\sum_{t=1}^{T}\left|F_t^0\right|^2\right)\left(\frac{1}{N^2T}\left\|e\Lambda^0\right\|^2\right) \\ &= \frac{1}{N}\operatorname{tr}\left(\frac{F^{0\prime}F^0}{T}\right)\left(\frac{1}{NT}\sum_{t=1}^{T}\left|\sum_{i=1}^{N}\lambda_i^0 e_{it}\right|^2\right).\end{aligned}$$

The two rightmost terms are $O_p(1)$, so

$$\frac{1}{NT}e\Lambda^0 F^{0\prime} = O_p\left(\frac{1}{\sqrt{N}}\right).$$

Meanwhile,

$$P_F^r - P_{FH}^k = O_p(1),$$

so we can see that

$$\frac{1}{NT}e\Lambda^0 F^{0\prime}\left(P_F^r - P_{FH}^k\right) = O_p\left(\frac{1}{\sqrt{N}}\right),$$

and as such that

$$\frac{1}{NT} e \Lambda^0 F^{0\prime} \left(P_F^r - P_{FH}^k \right) \xrightarrow{p} O.$$

By the CMT and the continuity of the trace operation,

$$\operatorname{tr}\left(\frac{1}{NT} \Lambda^0 F^{0\prime} \left(P_F^r - P_{FH}^k \right) e \right) = o_p(1).$$

From the above results, we can conclude that

$$A_{NT}^k = \operatorname{tr}\left(\frac{1}{NT} \Lambda^0 F^{0\prime} \left(P_F^r - P_{FH}^k \right) F^0 \Lambda^{0\prime} \right) + 2 \cdot \operatorname{tr}\left(\frac{1}{NT} \Lambda^0 F^{0\prime} \left(P_F^r - P_{FH}^k \right) e \right) \xrightarrow{p} \tau_k > 0.$$

Therefore,

$$\begin{aligned} 0 \leq \mathbb{P}\left(V(k, F^0 H^k) - V(r, F^0) < \frac{\tau_k}{2} \right) &\leq \mathbb{P}\left(A_{NT}^k < \tau_k - \frac{\tau_k}{2} \right) \\ &= \mathbb{P}\left(\tau_k - A_{NT}^k > \frac{\tau_k}{2} \right) \leq \mathbb{P}\left(\left| \tau_k - A_{NT}^k \right| > \frac{\tau_k}{2} \right), \end{aligned}$$

where the latter term converges to 0 as $N, T \to \infty$ by the definition of convergence in probability and the fact that $\frac{\tau_k}{2} > 0$. It follows that

$$\lim_{N,T \to \infty} \mathbb{P}\left(V(k, F^0 H^k) - V(r, F^0) < \frac{\tau_k}{2} \right) = 0,$$

and redefining τ_k as $\frac{\tau_k}{2}$ yields the desired result.

1.4.3 Preliminary Result 3

Here we show that:

$$\textbf{For any } k \geq r\textbf{,}\ V(k,\hat{F}^k) - V(r,\hat{F}^r) = O_p(\delta_{NT}^{-1}).$$

Suppose that $r \leq k \leq k_{max}$. Then,

$$H^{k\prime} \xrightarrow{p} Q^{k\prime}\Sigma_\Lambda,$$

where Q^k has rank r, so that $Q^{k\prime}\Sigma_\Lambda$ also has rank r, and $H^{k\prime}$ itself has rank r. As such, $H^k H^{k\prime}$ has full rank r and is nonsingular; defining

$$H^{k+} = H^{k\prime}\left(H^k H^{k\prime}\right)^{-1},$$

$H^k H^{k+} = I_r$, so that H^{k+} is the right generalized inverse of H^k.
Since

$$H^k H^{k\prime} \xrightarrow{p} \Sigma_\Lambda Q^k Q^{k\prime}\Sigma_\Lambda,$$

where the limt is nonsingular,

$$\left(H^k H^{k\prime}\right)^{-1} \xrightarrow{p} \left[\Sigma_\Lambda Q^k Q^{k\prime}\Sigma_\Lambda\right]^{-1}$$

and

$$H^{k+} \xrightarrow{p} Q^{k\prime}\Sigma_\Lambda\left[\Sigma_\Lambda Q^k Q^{k\prime}\Sigma_\Lambda\right]^{-1}.$$

By implication, $\left\|H^{k+}\right\| = O_p(1)$.

Note that we can express X as

$$\begin{aligned}X &= F^0\Lambda^{0\prime} + e = F^0 H^k H^{k+}\Lambda^{0\prime} + e\\ &= \hat{F}^k H^{k+}\Lambda^{0\prime} - \left(\hat{F}^k - F^0 H^k\right)H^{k+}\Lambda^{0\prime} + e.\end{aligned}$$

Therefore, defining

$$M_{\hat{F}}^k = I_T - P_{\hat{F}}^k,$$

we have

$$\begin{aligned}
V(k,\hat{F}^k) &= \frac{1}{NT}\operatorname{tr}\left(X' M_{\hat{F}}^k X\right) \\
&= \frac{1}{NT}\operatorname{tr}\left(\left(e - \left(\hat{F}^k - F^0 H^k\right) H^{k+}\Lambda^{0\prime}\right)' M_{\hat{F}}^k \left(e - \left(\hat{F}^k - F^0 H^k\right) H^{k+}\Lambda^{0\prime}\right)\right) \\
&= \frac{1}{NT}\operatorname{tr}\left(e' M_{\hat{F}}^k e\right) + 2\frac{1}{NT}\operatorname{tr}\left(\Lambda^0 H^{k+\prime}\left(\hat{F}^k - F^0 H^k\right)' M_{\hat{F}}^k e\right) \\
&\quad + \frac{1}{NT}\operatorname{tr}\left(\Lambda^0 H^{k+\prime}\left(\hat{F}^k - F^0 H^k\right)' M_{\hat{F}}^k \left(\hat{F}^k - F^0 H^k\right) H^{k+}\Lambda^{0\prime}\right),
\end{aligned}$$

since $M_{\hat{F}}^k \hat{F}^k = O$. Likewise, defining

$$M_F^r = I_T - P_F^r,$$

we have

$$\begin{aligned}
V(r,F^0) &= \frac{1}{NT}\operatorname{tr}\left(X' M_F^r X\right) \\
&= \frac{1}{NT}\operatorname{tr}\left(\left(F^0\Lambda^{0\prime} + e\right)' M_F^r \left(F^0\Lambda^{0\prime} + e\right)\right) \\
&= \frac{1}{NT}\operatorname{tr}\left(e' M_F^r e\right).
\end{aligned}$$

Therefore, the difference between the two can be written as

$$\begin{aligned}
V(k,\hat{F}^k) - V(r,F^0) &= \frac{1}{NT}\operatorname{tr}\left(e'\left(P_F^r - P_{\hat{F}}^k\right)e\right) \\
&\quad + 2\frac{1}{NT}\operatorname{tr}\left(\Lambda^0 H^{k+\prime}\left(\hat{F}^k - F^0 H^k\right)' M_{\hat{F}}^k e\right) \\
&\quad + \frac{1}{NT}\operatorname{tr}\left(\Lambda^0 H^{k+\prime}\left(\hat{F}^k - F^0 H^k\right)' M_{\hat{F}}^k \left(\hat{F}^k - F^0 H^k\right) H^{k+}\Lambda^{0\prime}\right).
\end{aligned}$$

As usual, we will examine each term in turn:

1) $\frac{1}{NT}\operatorname{tr}\left(\Lambda^0 H^{k+\prime}\left(\hat{F}^k - F^0 H^k\right)' M^k_{\hat{F}}\left(\hat{F}^k - F^0 H^k\right) H^{k+}\Lambda^{0\prime}\right)$

Because $P^k_{\hat{F}}$ is symmetric and idempotent,

$$v' P^k_{\hat{F}} v = v' P^{k\prime}_{\hat{F}} P^k_{\hat{F}} v = \left(P^k_{\hat{F}} v\right)' \left(P^k_{\hat{F}} v\right),$$

which tells us that $P^k_{\hat{F}}$ is positive semidefinite. Therefore, for any $v \in \mathbb{R}^T$,

$$v' P^k_{\hat{F}} v = v'\left(P^k_{\hat{F}} - I_T\right) + v'v \geq 0;$$

$P^k_{\hat{F}} - I_T = M^k_{\hat{F}}$, so

$$v' M^k_{\hat{F}} v \leq v'v.$$

Note that

$$\Lambda^0 H^{k+\prime}\left(\hat{F}^k - F^0 H^k\right)' M^k_{\hat{F}}\left(\hat{F}^k - F^0 H^k\right) H^{k+}\Lambda^{0\prime}$$

is a positive semidefinite matrix, and therefore that its trace is non-negative. In addition, due to the property of $M^k_{\hat{F}}$ shown above,

$$\begin{aligned}
0 &\leq \frac{1}{NT}\operatorname{tr}\left(\Lambda^0 H^{k+\prime}\left(\hat{F}^k - F^0 H^k\right)' M^k_{\hat{F}}\left(\hat{F}^k - F^0 H^k\right) H^{k+}\Lambda^{0\prime}\right) \\
&\leq \frac{1}{NT}\operatorname{tr}\left(\Lambda^0 H^{k+\prime}\left(\hat{F}^k - F^0 H^k\right)'\left(\hat{F}^k - F^0 H^k\right) H^{k+}\Lambda^{0\prime}\right) \\
&= \frac{1}{NT}\operatorname{tr}\left(\left(\hat{F}^k - F^0 H^k\right)'\left(\hat{F}^k - F^0 H^k\right)\left(H^{k+}\Lambda^{0\prime}\Lambda^0 H^{k+\prime}\right)\right) \\
&\leq \frac{1}{NT}\left\|\left(\hat{F}^k - F^0 H^k\right)'\left(\hat{F}^k - F^0 H^k\right)\right\| \cdot \left\|H^{k+}\Lambda^{0\prime}\Lambda^0 H^{k+\prime}\right\| \\
&\leq \left(\frac{1}{T}\left\|\hat{F}^k - F^0 H^k\right\|^2\right) \cdot \left(\frac{1}{N}\left\|\Lambda^0\right\|^2\right) \cdot \left\|H^{k+}\right\|^2,
\end{aligned}$$

where we once again utilized the Cauchy-Schwarz inequality applied to the trace inner product on the space of all real symmetric $k \times k$ matrices. The two rightmost terms are $O_p(1)$, so

$$\frac{1}{NT}\operatorname{tr}\left(\Lambda^0 H^{k+\prime}\left(\hat{F}^k - F^0 H^k\right)' M^k_{\hat{F}}\left(\hat{F}^k - F^0 H^k\right) H^{k+}\Lambda^{0\prime}\right) = O_p\left(\delta^{-1}_{NT}\right).$$

2) $\frac{1}{NT}\operatorname{tr}\left(\Lambda^0 H^{k+\prime}\left(\hat{F}^k - F^0 H^k\right)' M_{\hat{F}}^k e\right)$

Since $\hat{F}^{k\prime}\hat{F}^k = T\cdot\left(V_{NT}^k\right)^2$,

$$\left\|P_{\hat{F}}^k\right\| = \left\|\frac{1}{T}\hat{F}^k\left(V_{NT}^k\right)^{-2}\hat{F}^{k\prime}\right\| = \left\|\frac{1}{T}\tilde{F}^k\tilde{F}^{k\prime}\right\| \\ \leq \frac{1}{T}\left\|\tilde{F}^k\right\|^2 = \operatorname{tr}\left(\frac{\tilde{F}^{k\prime}\tilde{F}^k}{T}\right) = k < +\infty,$$

so $P_{\hat{F}}^k = O_p(1)$. It follows that

$$\left\|M_{\hat{F}}^k\right\| \leq \|I_k\| + \left\|P_{\hat{F}}^k\right\| \leq k + \sqrt{k} < +\infty,$$

so that $M_{\hat{F}}^k = O_p(1)$ as well.
Thus,

$$\left|\frac{1}{NT}\operatorname{tr}\left(\Lambda^0 H^{k+\prime}\left(\hat{F}^k - F^0 H^k\right)' M_{\hat{F}}^k e\right)\right| = \left|\operatorname{tr}\left(\frac{1}{NT}H^{k+\prime}\left(\hat{F}^k - F^0 H^k\right)' M_{\hat{F}}^k e'\Lambda^0\right)\right| \\ \leq \sqrt{r}\cdot\left\|\frac{1}{NT}e'\Lambda^0 H^{k+\prime}\left(\hat{F}^k - F^0 H^k\right)' M_{\hat{F}}^k\right\| \\ \leq \frac{\sqrt{r}}{\sqrt{N}}\cdot\left\|\frac{1}{\sqrt{NT}}e'\Lambda^0\right\|\cdot\left\|H^{k+}\right\|\cdot\left(\frac{1}{\sqrt{T}}\left\|\hat{F}^k - F^0 H^k\right\|\right)\cdot\left\|M_{\hat{F}}^k\right\|,$$

where we used the result that, for any $r\times r$ matrix A,

$$|\operatorname{tr}(A)| \leq \sum_{i=1}^r |A_{ii}| \leq \left(\sum_{i=1}^r A_{ii}^2\right)^{\frac{1}{2}}\cdot\sqrt{r} \leq \sqrt{r}\cdot\|A\|$$

by the Cauchy-Schwarz inequality.
Since

$$\left\|e'\Lambda^0\right\|^2 = \sum_{t=1}^T\left|\Lambda^{0\prime}e_t\right|^2 = \sum_{t=1}^T\left|\sum_{i=1}^N \lambda_i^0 e_{it}\right|^2,$$

it follows that

$$\frac{1}{NT}\left\|e'\Lambda^0\right\|^2 = \frac{1}{NT}\sum_{t=1}^T\left|\sum_{i=1}^N \lambda_i^0 e_{it}\right|^2 = O_p(1).$$

Therefore,

$$\frac{1}{NT}\operatorname{tr}\left(\Lambda^0 H^{k+\prime}\left(\hat{F}^k - F^0 H^k\right)' M_{\hat{F}}^k e\right) = O_p\left(\frac{1}{\min(N,\sqrt{NT})}\right).$$

3) $\frac{1}{NT}\operatorname{tr}\left(e'\left(P_F^r - P_{\hat{F}}^k\right)e\right)$

This final term can be decomposed as

$$\frac{1}{NT}\operatorname{tr}\left(e'\left(P_F^r - P_{\hat{F}}^k\right)e\right) = \frac{1}{NT}\operatorname{tr}\left(e'P_F^r e\right) - \frac{1}{NT}\operatorname{tr}\left(e'P_{\hat{F}}^k e\right).$$

Focusing on the first term, note that

$$\begin{aligned} 0 \leq \frac{1}{NT}\operatorname{tr}\left(e'P_F^r e\right) &= \frac{1}{NT}\operatorname{tr}\left(e'F^0\left(F^{0\prime}F^0\right)^{-1}F^{0\prime}e\right) \\ &= \frac{1}{NT^2}\operatorname{tr}\left(\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}F^{0\prime}ee'F^0\right) \\ &\leq \frac{1}{NT^2}\left\|\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\right\|\cdot\left\|F^{0\prime}ee'F^0\right\| \\ &\leq \frac{1}{T}\cdot\left\|\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\right\|\cdot\left(\frac{1}{NT}\left\|F^{0\prime}e\right\|^2\right) \end{aligned}$$

by the Cauchy-Schwarz inequality applied to the trace inner product on the space of $r \times r$ real symmetric matrices. Since

$$F^{0\prime}e = \begin{pmatrix} F^{0\prime}\underline{e}_1 & \cdots & F^{0\prime}\underline{e}_N \end{pmatrix},$$

we have

$$\frac{1}{NT}\left\|F^{0\prime}e\right\|^2 = \frac{1}{NT}\sum_{i=1}^N\left|F^{0\prime}\underline{e}_i\right|^2 = \frac{1}{NT}\sum_{i=1}^N\left|\sum_{t=1}^T F_t^0 e_{it}\right|^2 = O_p(1).$$

Therefore,

$$\frac{1}{NT}\operatorname{tr}\left(e'P_F^r e\right) = O_p\left(\frac{1}{T}\right).$$

As for the second term, since $P_{\hat{F}}^k = \frac{1}{T}\tilde{F}^k\tilde{F}^{k\prime}$,

$$\begin{aligned} \frac{1}{NT}\operatorname{tr}\left(e'P_{\hat{F}}^k e\right) &= \frac{1}{NT^2}\operatorname{tr}\left(e'\tilde{F}^k\tilde{F}^{k\prime}e\right) \\ &= \frac{1}{NT}\operatorname{tr}\left(\left(\frac{1}{\sqrt{T}}\tilde{F}^k\right)'ee'\left(\frac{1}{\sqrt{T}}\tilde{F}^k\right)\right). \end{aligned}$$

Since $\frac{1}{\sqrt{T}}\tilde{F}^k$ is a $T \times k$ matrix such that $\frac{\tilde{F}^{k\prime}\tilde{F}^k}{T} = I_k$ and ee' a $T \times T$ positive semidefinite matrix, as derived all the way back in page 8, we have

$$\operatorname{tr}\left(\left(\frac{1}{\sqrt{T}}\tilde{F}^k\right)'ee'\left(\frac{1}{\sqrt{T}}\tilde{F}^k\right)\right) \leq \sum_{i=1}^k \mu_i,$$

where $\mu_1 \geq \cdots \geq \mu_k \geq 0$ are the k largest eigenvalues of ee'.
Denote by $\rho(ee') = \mu_1$, the largest eigenvalue of ee'. We will show that

$$\frac{1}{T}\rho(ee') = O_p(1).$$

ee' is not a zero matrix, so its largest eigenvalue must always be positive. The non-zero eigenvalues of ee' and $e'e$ are identical, so it follows that $\rho(ee') = \rho(e'e)$. Because $e'e$ is symmetric, using the principal axis theorem we can find a $v^* \in \mathbb{R}^N$ such that $|v^*| = 1$ and $\rho(e'e) = v^{*\prime}e'ev^* = |ev^*|^2$. Since

$$|ev^*|^2 = \left|\sum_{i=1}^{N} \underline{e}_i v_i^*\right|^2 = \sum_{i=1}^{N}\sum_{j=1}^{N} \underline{e}_i'\underline{e}_j v_i^* v_j^*,$$

we can see that

$$\begin{aligned}
\mathbb{E}\left[\rho(e'e)\right] = \mathbb{E}\left[|ev^*|^2\right] &= \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N} \mathbb{E}\left[\underline{e}_i'\underline{e}_j\right] v_i^* v_j^* \\
&= \sum_{i=1}^{N} \mathbb{E}\left[\underline{e}_i'\underline{e}_i\right](v_i^*)^2 \\
&= \sum_{i=1}^{N}(v_i^*)^2 \left(\sum_{t=1}^{T} \mathbb{E}\left[e_{it}^2\right]\right) \\
&= T\gamma(0) \cdot \left(\sum_{i=1}^{N}(v_i^*)^2\right) \\
&= T\gamma(0),
\end{aligned}$$

where the last equality follows because $\sum_{i=1}^{N}(v_i^*)^2 = |v^*|^2 = 1$. Therefore,

$$\mathbb{E}\left[\frac{1}{T}\rho(ee')\right] \leq \gamma(0) < +\infty,$$

which implies that $\frac{1}{T}\rho(ee') = O_p(1)$.

We can finally see that

$$\frac{1}{NT}\operatorname{tr}\left(e'P_{\hat{F}}^k e\right) \leq \frac{1}{NT}\sum_{i=1}^{k}\mu_i \leq \frac{k}{N}\left(\frac{1}{T}\rho(ee')\right),$$

which tells us that

$$\frac{1}{NT}\operatorname{tr}\left(e'P_{\hat{F}}^k e\right) = O_p\left(\frac{1}{N}\right).$$

Putting the above results together, we have

$$\frac{1}{NT}\operatorname{tr}\left(e'\left(P_F^r - P_{\hat{F}}^k\right)e\right) = O_p\left(\frac{1}{T}\right) + O_p\left(\frac{1}{N}\right) = O_p\left(\delta_{NT}^{-1}\right).$$

The results above reveal that

$$V(k,\hat{F}^k) - V(r,F^0) = O_p\left(\delta_{NT}^{-1}\right) + O_p\left(\frac{1}{\min(N,\sqrt{NT})}\right) = O_p\left(\delta_{NT}^{-1}\right).$$

Now note that, for any $r \leq k \leq k_{max}$,

$$\left|V(k,\hat{F}^k) - V(r,\hat{F}^r)\right| \leq \left|V(k,\hat{F}^k) - V(r,F^0)\right| + \left|V(r,\hat{F}^r) - V(r,F^0)\right| \leq 2 \max_{r\leq k\leq k_{max}} \left|V(k,\hat{F}^k) - V(r,F^0)\right|.$$

Since $\left|V(k,\hat{F}^k) - V(r,F^0)\right|$ is $O_p(\delta_{NT}^{-1})$ for all $r \leq k \leq k_{max}$, we can see that $V(k,\hat{F}^k) - V(r,\hat{F}^r)$ is also $O_p(\delta_{NT}^{-1})$ for any $r \leq k \leq k_{max}$.

1.4.4 The Consistency of $PC(k)$

Letting $\hat{k}$ be the number of factors chosen by $PC(k)$, we hope to show that

$$\lim_{N,T\to\infty} \mathbb{P}\left(\hat{k} \neq r\right) = 0.$$

Since the inclusion

$$\{\hat{k} \neq r\} \subset \bigcup_{1\leq k\leq k_{max},\ \ k\neq r} \{PC(k) < PC(r)\}$$

holds, it suffices to prove that

$$\lim_{N,T\to\infty} \mathbb{P}\left(PC(k) < PC(r)\right) = 0$$

for all $1 \leq k \leq k_{max}$ such that $k \neq r$.
Note that

$$PC(k) - PC(r) = V(k, \hat{F}^k) - V(r, \hat{F}^r) + (k-r)g(N,T)$$

for any $1 \leq k \leq k_{max}$; as such, $PC(k) < PC(r)$ if and only if $V(k, \hat{F}^k) - V(r, \hat{F}^r) < (k-r)g(N,T)$.
We study two distinct cases:

1) **The Case** $k < r$

Initially, let $k < r$. Then, we can decompose

$$V(k,\hat{F}^k) - V(r,\hat{F}^r) = \left(V(k,\hat{F}^k) - V(k,F^0H^k)\right) + \left(V(k,F^0H^k) - V(r,F^0H^r)\right) + \left(V(r,F^0H^r) - V(r,\hat{F}^r)\right)$$

By premilinary result 1, the first and third terms are $O_p(\delta_{NT}^{-1/2})$, which implies that they are $o_p(1)$. As for the second term, since H^r is nonsingular $r \times r$ matrix valued,

$$V(r, F^0H^r) = V(r, F^0).$$

Thus, by preliminary result 2,

$$\lim_{N,T\to\infty} \mathbb{P}\left(V(k,F^0H^k) - V(r,F^0) \geq \tau_k\right) = 1.$$

for some $\tau_k > 0$. Denoting

$$A_{NT} = -\left(V(k,\hat{F}^k) - V(k,F^0H^k)\right) - \left(V(r,F^0H^r) - V(r,\hat{F}^r)\right) = o_p(1),$$

we can now see that

$$\mathbb{P}\left(V(k,F^0H^k)-V(r,F^0)\geq\tau_k\right)\leq\mathbb{P}\left(V(k,\hat{F}^k)-V(r,\hat{F}^r)\geq\frac{\tau_k}{2}\right)+\mathbb{P}\left(A_{NT}\geq\frac{\tau_k}{2}\right),$$

so taking $N,T\to\infty$ on both sides yields

$$1\leq\liminf_{N,T\to\infty}\mathbb{P}\left(V(k,\hat{F}^k)-V(r,\hat{F}^r)\geq\frac{\tau_k}{2}\right)\leq\limsup_{N,T\to\infty}\mathbb{P}\left(V(k,\hat{F}^k)-V(r,\hat{F}^r)\geq\frac{\tau_k}{2}\right)\leq 1,$$

and therefore implies

$$\lim_{N,T\to\infty}\mathbb{P}\left(V(k,\hat{F}^k)-V(r,\hat{F}^r)\geq\frac{\tau_k}{2}\right)=1.$$

Since

$$\begin{aligned}\mathbb{P}(PC(k)<PC(r))&\leq\mathbb{P}\left(PC(k)-PC(r)<\frac{\tau_k}{4}\right)=\mathbb{P}\left(\left(V(k,\hat{F}^k)-V(r,\hat{F}^r)\right)-(r-k)g(N,T)<\frac{\tau_k}{4}\right)\\&\leq\mathbb{P}\left(V(k,\hat{F}^k)-V(r,\hat{F}^r)<\frac{\tau_k}{2}\right)+\mathbb{P}\left(g(N,T)>\frac{\tau_k}{4(r-k)}\right),\end{aligned}$$

if $g(N,T)\to 0$ as $N,T\to\infty$, then

$$\lim_{N,T\to\infty}\mathbb{P}(PC(k)<PC(r))=0.$$

2) **The Case** $k>r$

Now suppose that $r<k\leq k_{max}$. It holds that

$$\begin{aligned}\mathbb{P}(PC(k)<PC(r))&=\mathbb{P}\left(V(k,\hat{F}^k)-V(r,\hat{F}^r)+(k-r)g(N,T)<0\right)\\&=\mathbb{P}\left(V(r,\hat{F}^r)-V(k,\hat{F}^k)>(k-r)g(N,T)\right)=\mathbb{P}\left(\frac{V(r,\hat{F}^r)-V(k,\hat{F}^k)}{g(N,T)}>k-r\right),\end{aligned}$$

where $k-r\geq 1$. From preliminary result 3, we have the relation

$$V(k,\hat{F}^k)-V(r,\hat{F}^r)=O_p(\delta_{NT}^{-1}),$$

indicating that $V(r,\hat{F}^r)-V(k,\hat{F}^k)$ converges at the same rate as δ_{NT}^{-1}.
If $g(N,T)$ converges to 0 at a rate faster than δ_{NT}^{-1}, then $\frac{V(r,\hat{F}^r)-V(k,\hat{F}^k)}{g(N,T)}$ will diverge to $+\infty$ in probability, meaning that $\mathbb{P}(PC(k)<PC(r))\to 1$ as $N,T\to\infty$.
On the other hand, if $g(N,T)$ converges to 0 at the same rate as δ_{NT}^{-1}, $\frac{V(r,\hat{F}^r)-V(k,\hat{F}^k)}{g(N,T)}$ may converge to a level larger than $k-r$, at which point $\mathbb{P}(PC(k)<PC(r))\to 0$ as $N,T\to\infty$ is not guaranteed.

Therefore, in order for $\frac{V(r,\hat{F}^r)-V(k,\hat{F}^k)}{g(N,T)}$ to converge to 0, or for $\mathbb{P}(PC(k) < PC(r)) \to 0$ as $N,T \to \infty$, it must be the case that $g(N,T)$ goes to 0 at a rate slower than δ_{NT}^{-1}. What this means is that

$$\delta_{NT} \cdot g(N,T) = \frac{g(N,T)}{\delta_{NT}^{-1}} \to +\infty$$

as $N,T \to \infty$, since the denominator goes to 0 at a faster rate than the numerator.

We can now see that the penalty term must satisfy two conditions for $\mathbb{P}(PC(k) < PC(r)) \to 0$ as $N,T \to \infty$ for any $1 \leq k \neq r \leq k_{max}$:

i) $g(N,T) \to 0$ in order for $\mathbb{P}(PC(k) < PC(r)) \to 0$ for $k < r$, while

ii) $\delta_{NT} \cdot g(N,T) \to +\infty$ as $N,T \to \infty$ in order for $\mathbb{P}(PC(k) < PC(r)) \to 0$ for $k > r$.

Note that these conditions mirror those ensuring consistency of traditional information criteria used to determine lag orders, such as the AIC and BIC; the penalty term must converge to 0 in order to rule out lag orders below the true lag order but not too fast in order to rule out lag orders higher than the true lag order.

Bai and Ng propose threee specific information criteria that satisfy the constraints above, which are given as

$$\begin{aligned}
IC_{p_1}(k) &= \log\left(V(k,\hat{F}^k)\right) + k\left(\frac{N+T}{NT}\right)\log\left(\frac{NT}{N+T}\right) \\
IC_{p_2}(k) &= \log\left(V(k,\hat{F}^k)\right) + k\left(\frac{N+T}{NT}\right)\log(\delta_{NT}) \\
IC_{p_3}(k) &= \log\left(V(k,\hat{F}^k)\right) + k\frac{\log(\delta_{NT})}{\delta_{NT}},
\end{aligned}$$

where the choice $\frac{NT}{N+T}$ was considered because $\delta_{NT} \approx \frac{NT}{N+T}$ for large N,T.

1.5 The Asymptotic Distribution of the Estimated Factors

Suppose now that the true number of factors r is known, and that $k = r$ factors are estimated. Let the estimators without the superscript k denote the estimators under $k = r$ factors. Specifically, we let $\tilde{F}$ be the estimated factors, which were obtained as $\sqrt{T}$ times the orthonormal eigenvectors corresponding to the r largest eigenvalues of XX', which are assumed to be positive, and

$$H = \left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)\left(\frac{F^{0\prime}\tilde{F}}{T}\right).$$

We now proceed in steps:

1.5.1 Decomposing $\tilde{F}_t$

Return to equation (1), namely

$$\hat{F} - F^0 H = \frac{1}{NT} e\Lambda^0 F^{0\prime}\tilde{F} + \frac{1}{NT} F^0\Lambda^{0\prime}e'\tilde{F} + \frac{1}{NT} ee'\tilde{F}.$$

By implication, for any $t \in N_+$, choosing a T greater than t, we have

$$\hat{F}_t - H'F_t^0 = \frac{1}{NT}\tilde{F}'F^0\Lambda^{0\prime}e_t + \frac{1}{NT}\tilde{F}'e\Lambda^0 F_t^0 + \frac{1}{NT}\tilde{F}'e \cdot e_t.$$

We already saw earlier that, under our assumptions,

$$\frac{1}{T}\sum_{t=1}^{T}\left|\hat{F}_t - H'F_t^0\right|^2 = \frac{1}{T}\left\|\hat{F} - F^0 H\right\|^2 = O_p\left(\delta_{NT}^{-1}\right).$$

Since V_{NT} is a diagonal matrix with positive diagonal entries, it is nonsingular, so that, by $V_{NT}\tilde{F}_t = \hat{F}_t$,

$$\frac{1}{T}\sum_{t=1}^{T}\left|\tilde{F}_t - \tilde{H}'F_t^0\right|^2 = V_{NT}^{-1} \cdot \frac{1}{T}\sum_{t=1}^{T}\left|\hat{F}_t - H'F_t^0\right|^2,$$

where

$$\tilde{H}' = V_{NT}^{-1}H' = V_{NT}^{-1}\left(\frac{\tilde{F}'F^0}{T}\right)\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right).$$

Likewise,

$$\tilde{F}_t - \tilde{H}'F_t^0 = V_{NT}^{-1}\left[\frac{1}{NT}\tilde{F}'F^0\Lambda^{0\prime}e_t + \frac{1}{NT}\tilde{F}'e\Lambda^0 F_t^0 + \frac{1}{NT}\tilde{F}'e \cdot e_t\right].$$

1.5.2 The Probability Limit of V_{NT}

We now show that V_{NT} converges in probability to a positive definite matrix. This is easy to show under the assumption that $\frac{F^{0\prime}\tilde{F}}{T}$ converges in probability to some nonsingular $r \times r$ matrix Q.

From the consistency result above, we found that

$$\frac{1}{T}\left\|\hat{F} - F^0 H\right\|^2 = \frac{1}{T}\sum_{t=1}^{T}\left|\hat{F}_t - H' F_t^0\right|^2 = O_p\left(\delta_{NT}^{-1}\right),$$

or equivalently, that

$$\frac{1}{\sqrt{T}}\left(\hat{F} - F^0 H\right) = \frac{1}{\sqrt{T}}\left(\tilde{F} V_{NT} - F^0 H\right) = O_p\left(\frac{1}{\min(\sqrt{N}, \sqrt{T})}\right) = o_p(1).$$

Premultiplying both sides by $\frac{\tilde{F}'}{\sqrt{T}}$ and using the fact that $\frac{\tilde{F}'\tilde{F}}{T} = I_r$ implies that

$$V_{NT} - \frac{\tilde{F}' F^0}{T}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)\frac{F^{0\prime}\tilde{F}}{T} = o_p(1),$$

and because

$$\frac{\tilde{F}' F^0}{T}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)\frac{F^{0\prime}\tilde{F}}{T} \xrightarrow{p} Q'\Sigma_\Lambda Q,$$

it follows that

$$V_{NT} \xrightarrow{p} V = Q'\Sigma_\Lambda Q,$$

where V is positive definite because Q has full rank and Σ_Λ is positive definite, and is diagonal because V_{NT} is diagonal for any N,T. In addition, the diagonal entries of V are ordered because the diagonal entries of V_{NT} are ordered.
By the continuous mapping theorem, we now have

$$V_{NT}^{-1} \xrightarrow{p} V^{-1}.$$

The Form of V

To derive the specific form of V, we proceed as follows.
Premultiplying both sides of the equation $\frac{1}{\sqrt{T}}\left(\tilde{F}V_{NT} - F^0 H\right) = o_p(1)$ by $\frac{F^{0\prime}}{\sqrt{T}} = O_p(1)$ yields the equation

$$\frac{F^{0\prime}\tilde{F}}{T}V_{NT} - \frac{F^{0\prime}F^0}{T}H = o_p(1).$$

Since $H = \left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)\left(\frac{F^{0\prime}\tilde{F}}{T}\right)$, rearranging terms we have

$$\frac{F^{0\prime}\tilde{F}}{T}V_{NT} - \left[\left(\frac{F^{0\prime}F^0}{T}\right)\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)\right]\frac{F^{0\prime}\tilde{F}}{T} = o_p(1).$$

We can see that

$$\left[\left(\frac{F^{0\prime}F^0}{T}\right)\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)\right]\frac{F^{0\prime}\tilde{F}}{T} \xrightarrow{p} \Sigma_F\Sigma_\Lambda Q,$$

or equivalently,

$$\left[\left(\frac{F^{0\prime}F^0}{T}\right)\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)\right]\frac{F^{0\prime}\tilde{F}}{T} - \Sigma_F\Sigma_\Lambda Q = o_p(1).$$

Likewise, we have

$$\frac{F^{0\prime}\tilde{F}}{T}V_{NT} - QV = o_p(1).$$

By implication,

$$\Sigma_F\Sigma_\Lambda Q - QV = o_p(1),$$

and because the left hand side is deterministic, this means

$$\Sigma_F\Sigma_\Lambda Q = QV.$$

By definition, V is a diagonal matrix with diagonal entries equal to the eigenvalues of $\Sigma_F\Sigma_\Lambda$. Because Σ_F and Σ_Λ are positive definite, the eigenvalues of $\Sigma_F\Sigma_\Lambda$ are exactly those of $\Sigma_\Lambda\Sigma_F$.

The Probability Limit of $\frac{F^{0\prime}\tilde{F}}{T}$

We now derive the expression for the probability limit Q of $\frac{F^{0\prime}\tilde{F}}{T}$ by relying on the assumption that the eigenvalues of $\Sigma_F\Sigma_\Lambda$ are distinct.
Letting $\Sigma_\Lambda^{\frac{1}{2}}$ be the Cholesky factor of Σ_Λ, since

$$QV = \Sigma_F\Sigma_\Lambda Q = \Sigma_F\Sigma_\Lambda^{\frac{1}{2}}\Sigma_\Lambda^{\frac{1}{2}\prime}Q,$$

premultiplying both sides by $\Sigma_\Lambda^{\frac{1}{2}\prime}$ implies

$$\Sigma_\Lambda^{\frac{1}{2}\prime}\Sigma_F\Sigma_\Lambda^{\frac{1}{2}}\left[\Sigma_\Lambda^{\frac{1}{2}\prime}Q\right] = \left[\Sigma_\Lambda^{\frac{1}{2}\prime}Q\right]V.$$

Denoting

$$\Gamma = \Sigma_\Lambda^{\frac{1}{2}\prime}Q,$$

it is clear that Γ is a nonsingular $r \times r$ matrix whose columns are (non-normalized) eigenvectors of $\Sigma_\Lambda^{\frac{1}{2}\prime}\Sigma_F\Sigma_\Lambda^{\frac{1}{2}}$, which shares eigenvalues with $\Sigma_F\Sigma_\Lambda$. Since the eigenvalues in question are all distinct, the columns of Γ are orthogonal to one another.
Let V^* be the $r \times r$ diagonal matrix with the ith diagonal entry equal to the norm of the ith column of Γ. Then, defining

$$\Gamma^* = \Gamma(V^*)^{-1},$$

the norms of the columns of $\Gamma*$ are normalized to 1. Furthermore, by the orthogonality of the columns of Γ, we have

$$(V^*)^2 = \Gamma'\Gamma = Q'\Sigma_\Lambda Q = V,$$

so that the squared norm of each column of Γ is precisely the eigenvalue to which it corresponds. It follows that

$$\Gamma^* = \Gamma V^{-\frac{1}{2}} = \Sigma_\Lambda^{\frac{1}{2}\prime}QV^{-\frac{1}{2}},$$

so that Q is now recovered as

$$Q = \Sigma_\Lambda^{-\frac{1}{2}\prime}\Gamma^*V^{\frac{1}{2}}.$$

Note that, by the distinctness of the eigenvalues of $\Sigma_F\Sigma_\Lambda$, the matrix of orthonormal eigenvectors Γ^* is unique up to sign changes, and as such, the probability limit Q is also unique up to sign changes.

1.5.3 The Rate of Convergence of $\tilde{F}_t - \tilde{H}'F_t^0$

We know that

$$\frac{1}{T}\sum_{t=1}^{T}\left|\tilde{F}_t - \tilde{H}'F_t^0\right|^2 = V_{NT}^{-1}\cdot\frac{1}{T}\sum_{t=1}^{T}\left|\hat{F}_t - H'F_t^0\right|^2,$$

where

$$\tilde{H}' = V_{NT}^{-1}H' = V_{NT}^{-1}\left(\frac{\tilde{F}'F^0}{T}\right)\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right).$$

Since we showed above that $V_{NT}^{-1} = O_p(1)$, and $\frac{1}{T}\sum_{t=1}^{T}\left|\hat{F}_t - H'F_t^0\right|^2 = O_p(\delta_{NT}^{-1})$ by the result of section 2, we can see that

$$\frac{1}{T}\sum_{t=1}^{T}\left|\tilde{F}_t - \tilde{H}'F_t^0\right|^2 = \frac{1}{T}\left\|\tilde{F} - F^0\tilde{H}\right\|^2 = O_p(\delta_{NT}^{-1})$$

as well.

Also recall that

$$\tilde{F}_t - \tilde{H}'F_t^0 = V_{NT}^{-1}\left[\frac{1}{NT}\tilde{F}'F^0\Lambda^{0\prime}e_t + \frac{1}{NT}\tilde{F}'e\Lambda^0F_t^0 + \frac{1}{NT}\tilde{F}'e\cdot e_t\right].$$

for any $t \in N_+$. Now we investigate the rate of convergence of each term on the right:

1) $\frac{1}{NT}\tilde{F}'F^0\Lambda^{0\prime}e_t$

   Note that

$$\begin{aligned}
\left|\frac{1}{NT}\tilde{F}'F^0\Lambda^{0\prime}e_t\right|^2 &\leq 2\frac{1}{N^2T^2}\left|\left(\tilde{F} - F^0\tilde{H}\right)'F^0\Lambda^{0\prime}e_t\right|^2 + 2\frac{1}{N^2T^2}\left|\tilde{H}'F^{0\prime}F^0\Lambda^{0\prime}e_t\right|^2 \\
&\leq 2\frac{1}{N^2T^2}\left\|\tilde{F} - F^0\tilde{H}\right\|^2\cdot\left\|F^0\right\|^2\cdot\left|\Lambda^{0\prime}e_t\right|^2 + 2\frac{1}{N^2T^2}\left\|\tilde{H}\right\|^2\cdot\left\|F^0\right\|^4\cdot\left|\Lambda^{0\prime}e_t\right|^2 \\
&\leq 2\frac{1}{N}\cdot\left(\frac{1}{T}\left\|\tilde{F} - F^0\tilde{H}\right\|^2\right)\cdot\mathrm{tr}\left(\frac{F^{0\prime}F^0}{T}\right)\cdot\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i^0e_{it}\right|^2 \\
&\quad + 2\frac{1}{N}\cdot\left\|\tilde{H}\right\|^2\cdot\mathrm{tr}\left(\frac{F^{0\prime}F^0}{T}\right)^2\cdot\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i^0e_{it}\right|^2.
\end{aligned}$$

   We saw above that

$$\frac{1}{T}\left\|\tilde{F} - F^0\tilde{H}\right\|^2 = O_p\left(\frac{1}{\min(N,T)}\right),$$

and $\tilde{H}, \frac{F^{0\prime}F^0}{T}$ are $O_p(1)$. Due to the nonrandomness of the factor loadings,

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i^0 e_{it} \xrightarrow{d} N\left(\mathbf{0}, \gamma(0)\cdot\Sigma_\Lambda\right).$$

Therefore,d $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i^0 e_{it}$ are $O_p(1)$, and we have

$$\left|\frac{1}{NT}\tilde{F}'F^0\Lambda^{0\prime}e_t\right|^2 \le O_p\left(\frac{1}{N\cdot\min(N,T)}\right) + O_p\left(\frac{1}{N}\right) = O_p\left(\frac{1}{N}\right).$$

By implication,

$$\frac{1}{NT}\tilde{F}'F^0\Lambda^{0\prime}e_t = O_p\left(\frac{1}{\sqrt{N}}\right).$$

2) $\frac{1}{NT}\tilde{F}'e\Lambda^0 F_t^0$
Once again, note that

$$\begin{aligned}
\left|\frac{1}{NT}\tilde{F}'e\Lambda^0F_t^0\right|^2 &\le 2\frac{1}{N^2T^2}\left|\left(\tilde{F}-F^0\tilde{H}\right)'e\Lambda^0F_t^0\right|^2 + 2\frac{1}{N^2T^2}\left|\tilde{H}'F^{0\prime}e\Lambda^0F_t^0\right|^2 \\
&\le 2\frac{1}{N^2T^2}\left\|\tilde{F}-F^0\tilde{H}\right\|^2\cdot\left\|e\Lambda^0\right\|^2\cdot\left|F_t^0\right|^2 + 2\frac{1}{N^2T^2}\left\|\tilde{H}\right\|^2\cdot\left\|F^0\right\|^2\cdot\left\|e\Lambda^0\right\|^2\cdot\left|F_t^0\right|^2 \\
&\le 2\frac{1}{N}\cdot\left(\frac{1}{T}\left\|\tilde{F}-F^0\tilde{H}\right\|^2\right)\cdot\left(\frac{1}{NT}\sum_{s=1}^{T}\left|\Lambda^{0\prime}e_s\right|^2\right)\cdot\left|F_t^0\right|^2 \\
&\quad + 2\frac{1}{NT}\cdot\left\|\tilde{H}\right\|^2\cdot\operatorname{tr}\left(\frac{F^{0\prime}F^0}{T}\right)\cdot\left(\frac{1}{NT}\sum_{s=1}^{T}\left|\Lambda^{0\prime}e_s\right|^2\right)\cdot\left|F_t^0\right|^2.
\end{aligned}$$

Since $\left|F_t^0\right|^2, \frac{F^{0\prime}F^0}{T}$ and

$$\frac{1}{NT}\sum_{s=1}^{T}\left|\Lambda^{0\prime}e_s\right|^2 = \frac{1}{NT}\sum_{s=1}^{T}\left|\sum_{i=1}^{N}\lambda_i^0 e_{is}\right|^2$$

are all $O_p(1)$,

$$\left|\frac{1}{NT}\tilde{F}'e\Lambda^0F_t^0\right|^2 = O_p\left(\frac{1}{N\cdot\min(N,T)}\right) + O_p\left(\frac{1}{NT}\right).$$

$\frac{1}{N\cdot\min(N,T)}$ converges to 0 at a slower rate than $\frac{1}{NT}$, so

$$\left|\frac{1}{NT}\tilde{F}'e\Lambda^0F_t^0\right|^2 = O_p\left(\frac{1}{N\cdot\min(N,T)}\right)$$

and as such

$$\frac{1}{NT}\tilde{F}'e\Lambda^0 F_t^0 = O_p\left(\frac{1}{\min(N,\sqrt{NT})}\right).$$

3) $\frac{1}{NT}\tilde{F}'e\cdot e_t$

We can expand the final term as

$$\begin{aligned}
\left|\frac{1}{NT}\tilde{F}'e\cdot e_t\right| &\le \frac{1}{NT}\left|\sum_{s=1}^{T}\tilde{F}_s\cdot e_s' e_t\right| \\
&\le \frac{1}{NT}\left|\sum_{s=1}^{T}\sum_{i=1}^{N}\tilde{F}_s\left(e_{it}e_{is}-\gamma(t-s)\right)\right| + \frac{1}{T}\left|\sum_{s=1}^{T}\tilde{F}_s\gamma(t-s)\right| \\
&\le \frac{1}{\sqrt{N}}\left|\frac{1}{T\cdot\sqrt{N}}\sum_{s=1}^{T}\sum_{i=1}^{N}\tilde{F}_s\left(e_{it}e_{is}-\gamma(t-s)\right)\right| \\
&\quad + \frac{1}{T}\left|\sum_{s=1}^{T}\left(\tilde{F}_s-\tilde{H}'F_s^0\right)\gamma(t-s)\right| + \frac{1}{T}\left|\sum_{s=1}^{T}\tilde{H}'F_s^0\gamma(t-s)\right| \\
&\le \frac{1}{\sqrt{N}}\left|\frac{1}{T\cdot\sqrt{N}}\sum_{s=1}^{T}\sum_{i=1}^{N}\tilde{F}_s\left(e_{it}e_{is}-\gamma(t-s)\right)\right| \\
&\quad + \frac{1}{\sqrt{T}}\left(\frac{1}{T}\sum_{s=1}^{T}\left|\tilde{F}_s-\tilde{H}'F_s^0\right|^2\right)^{\frac{1}{2}}\left(\sum_{z=-\infty}^{\infty}\gamma(z)^2\right)^{\frac{1}{2}} + \frac{1}{T}\left\|\tilde{H}\right\|\cdot\left(\sum_{s=1}^{T}\left|F_s^0\right|\cdot|\gamma(s-t)|\right),
\end{aligned}$$

where the last inequality uses the Cauchy-Schwarz inequality.

Expanding the first term, we find that

$$\begin{aligned}
\left|\frac{1}{T\cdot\sqrt{N}}\sum_{s=1}^{T}\sum_{i=1}^{N}\tilde{F}_s\left(e_{it}e_{is}-\gamma(t-s)\right)\right| &\le \frac{1}{T\cdot\sqrt{N}}\sum_{s=1}^{T}\left|\tilde{F}_s-\tilde{H}'F_s^0\right|\cdot\left|\sum_{i=1}^{N}\left(e_{it}e_{is}-\gamma(t-s)\right)\right| \\
&\quad + \frac{1}{\sqrt{T}}\left\|\tilde{H}\right\|\cdot\left|\frac{1}{\sqrt{NT}}\sum_{s=1}^{T}\sum_{i=1}^{N}F_s^0\left(e_{it}e_{is}-\gamma(t-s)\right)\right| \\
&\le \left(\frac{1}{T}\sum_{s=1}^{T}\left|\tilde{F}_s-\tilde{H}'F_s^0\right|^2\right)^{\frac{1}{2}}\left(\frac{1}{NT}\sum_{s=1}^{T}\left|\sum_{i=1}^{N}\left(e_{it}e_{is}-\gamma(t-s)\right)\right|^2\right)^{\frac{1}{2}} \\
&\quad + \frac{1}{\sqrt{T}}\left\|\tilde{H}\right\|\cdot\left|\frac{1}{\sqrt{NT}}\sum_{s=1}^{T}\sum_{i=1}^{N}F_s^0\left(e_{it}e_{is}-\gamma(t-s)\right)\right|,
\end{aligned}$$

where the last inequality follows once again from the Cauchy-Schwarz inequality.

By assumption,

$$\frac{1}{\sqrt{NT}}\sum_{s=1}^{T}\sum_{i=1}^{N}F_s^0\left(e_{it}e_{is}-\gamma(t-s)\right) = O_p(1),$$

and we showed that

$$\frac{1}{NT}\sum_{s=1}^{T}\left|\sum_{i=1}^{N}(e_{it}e_{is}-\gamma(t-s))\right|^2$$

has bounded first moments and is thus $O_p(1)$ in section 3, so

$$\left|\frac{1}{T\cdot\sqrt{N}}\sum_{s=1}^{T}\sum_{i=1}^{N}\tilde{F}_s(e_{it}e_{is}-\gamma(t-s))\right| = O_p\left(\frac{1}{\min(\sqrt{N},\sqrt{T})}\right)+O_p\left(\frac{1}{\sqrt{T}}\right)=O_p\left(\frac{1}{\min(\sqrt{N},\sqrt{T})}\right).$$

Meanwhile,

$$\begin{aligned}\mathbb{E}\left[\sum_{s=1}^{T}\left|F_s^0\right|\cdot|\gamma(s-t)|\right] &= \sum_{s=1}^{T}\mathbb{E}\left[\left|F_s^0\right|\cdot|\gamma(s-t)|\right]\\ &\le \sum_{s=1}^{T}\left(\mathbb{E}\left|F_s^0\right|^2\right)^{\frac{1}{2}}\cdot|\gamma(s-t)| \qquad \text{(Hölder's inequality)}\\ &\le \left(\sup_{s\in N_+}\mathbb{E}\left|F_s^0\right|^2\right)^{\frac{1}{2}}\cdot\left(\sum_{z=-\infty}^{\infty}|\gamma(z)|\right).\end{aligned}$$

By the absolute summability of $\gamma(\cdot)$ and the finiteness of $\sup_{s\in N_+}\mathbb{E}|F_s^0|^2$, the term on the right hand side is finite, so that

$$\sum_{s=1}^{T}\left|F_s^0\right|\cdot|\gamma(s-t)| = O_p(1).$$

Therefore,

$$\left|\frac{1}{NT}\tilde{F}'e\cdot e_t\right| \le O_p\left(\frac{1}{\min(N,\sqrt{NT})}\right)+O_p\left(\frac{1}{\min(T,\sqrt{NT})}\right)+O_p\left(\frac{1}{T}\right).$$

Since $\frac{1}{\min(N,\sqrt{NT})}$ and $\frac{1}{\min(T,\sqrt{NT})}$ converge to 0 at a slower rate than $\frac{1}{T}$,

$$\frac{1}{NT}\tilde{F}'e\cdot e_t = O_p\left(\frac{1}{\min(N,\sqrt{NT})}\right)+O_p\left(\frac{1}{\min(T,\sqrt{NT})}\right).$$

From the above results, we can infer that

$$\tilde{F}_t - \tilde{H}' F_t^0 = O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\min(N, \sqrt{NT})}\right) + O_p\left(\frac{1}{\min(N, \sqrt{NT})}\right) + O_p\left(\frac{1}{\min(T, \sqrt{NT})}\right).$$

Clearly, $\frac{1}{\sqrt{N}}$ converges to 0 at a slower rate than $\frac{1}{\sqrt{N}\cdot\min(\sqrt{N},\sqrt{T})}$ as $N,T \to \infty$, so

$$\tilde{F}_t - \tilde{H}' F_t^0 = O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\min(T, \sqrt{NT})}\right).$$

Suppose that $\frac{\sqrt{N}}{T} \to 0$ as $N,T \to \infty$. Then,

$$\frac{\frac{1}{\min(T,\sqrt{NT})}}{\frac{1}{\sqrt{N}}} = \frac{1}{\min\left(\frac{T}{\sqrt{N}}, \sqrt{T}\right)} \to 0$$

as $N,T \to \infty$, implying that $\frac{1}{\min(T,\sqrt{NT})}$ converges to 0 fasterthan $\frac{1}{\sqrt{N}}$. Therefore,

$$\tilde{F}_t - \tilde{H}' F_t^0 = O_p\left(\frac{1}{\sqrt{N}}\right),$$

and because the $O_p\left(\frac{1}{\sqrt{N}}\right)$ term corresponds to $\frac{1}{NT}\tilde{F}' F^0 \Lambda^{0\prime} e_t$, we have

$$\begin{aligned}
\sqrt{N}\left(\tilde{F}_t - \tilde{H}' F_t^0\right) &= V_{NT}^{-1} \frac{1}{\sqrt{N}T} \tilde{F}' F^0 \Lambda^{0\prime} e_t + o_p(1) \\
&= V_{NT}^{-1} \left(\frac{\tilde{F}' F^0}{T}\right) \frac{\Lambda^{0\prime} e_t}{\sqrt{N}} + o_p(1).
\end{aligned}$$

1.5.4 The Asymptotic Distribution of $\sqrt{N}\left(\tilde{F}_t - \tilde{H}'F_t^0\right)$

We showed above that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i^0 e_{it} \xrightarrow{d} N\left[\mathbf{0}, \gamma(0)\cdot\Sigma_\Lambda\right]$$

under assumptions 1 to 6. It follows by Slutsky's theorem that

$$V_{NT}^{-1}\left(\frac{\tilde{F}'F^0}{T}\right)\frac{\Lambda^{0\prime}e_t}{\sqrt{N}} \xrightarrow{d} N\left[\mathbf{0}, \gamma(0)\cdot V^{-1}Q'\Sigma_\Lambda Q V^{-1}\right],$$

and again that

$$\sqrt{N}\left(\tilde{F}_t - \tilde{H}'F_t^0\right) = V_{NT}^{-1}\left(\frac{\tilde{F}'F^0}{T}\right)\frac{\Lambda^{0\prime}e_t}{\sqrt{N}} + o_p(1) \xrightarrow{d} N\left[\mathbf{0}, \gamma(0)\cdot V^{-1}Q'\Sigma_\Lambda Q V^{-1}\right].$$

To further simplify the expression above, recall that the probability limit Q of $\frac{F^{0\prime}\tilde{F}}{T}$ is given as

$$Q = \Sigma_\Lambda^{-\frac{1}{2}\prime}\Gamma^* V^{\frac{1}{2}},$$

where Γ^* collects the unique (up to sign changes) orthonormal eigenbasis of $\Sigma_\Lambda^{\frac{1}{2}\prime}\Sigma_F\Sigma_\Lambda^{\frac{1}{2}}$. Thus,

$$\Sigma_\Lambda Q V^{-1} = \Sigma_\Lambda\Sigma_\Lambda^{-\frac{1}{2}\prime}\Gamma^* V^{\frac{1}{2}}V^{-1} = \Sigma_\Lambda^{\frac{1}{2}}\Gamma^* V^{-\frac{1}{2}} = Q'^{-1},$$

using the property that $\Gamma^{*-1} = \Gamma^{*\prime}$. It follows that

$$V^{-1}Q'\Sigma_\Lambda Q V^{-1} = V^{-1}Q'Q'^{-1}V^{-1} = V^{-2},$$

and the asymptotic distribution of $\tilde{F}_t$ becomes

$$\sqrt{N}\left(\tilde{F}_t - \tilde{H}'F_t^0\right) \xrightarrow{d} N\left[\mathbf{0}, \gamma(0)\cdot V^{-2}\right].$$

In other words, the factors at time t are asymptotically independent with the asymptotic variance of the ith factor being equal to the error variance $\gamma(0)$ divided by the square of the ith largest eigenvalue of $\Sigma_\Lambda\Sigma_F$. By implication, the more overall cross-sectional variation is explained by a factor, the more precisely it is estimated.

1.6 Asymptotic Theory for the Estimated Factor Loadings

The estimated factor loadings $\tilde{\Lambda}$ were seen to be given as

$$\begin{aligned}\tilde{\Lambda}' = \frac{1}{T}\tilde{F}'X &= \frac{1}{T}\tilde{F}'(F^0\Lambda^{0\prime} + e) \\ &= \frac{1}{T}\tilde{F}'F^0\Lambda^{0\prime} + \frac{1}{T}\tilde{F}'e.\end{aligned}$$

This implies that, for any $i \in N_+$,

$$\tilde{\lambda}_i = \frac{1}{T}\tilde{F}'F^0\lambda_i^0 + \frac{1}{T}\tilde{F}'\underline{e}_i,$$

given that $N > i$. Writing $F^0 - \tilde{F}\tilde{H}^{-1} + \tilde{F}\tilde{H}^{-1}$ in place of F^0, the above expression can be further expanded as

$$\begin{aligned}\tilde{\lambda}_i &= \frac{1}{T}\tilde{F}'\left(F^0 - \tilde{F}\tilde{H}^{-1}\right)\lambda_i^0 + \tilde{H}^{-1}\lambda_i^0 + \frac{1}{T}\tilde{F}'\underline{e}_i \\ &= \tilde{H}^{-1}\lambda_i^0 + \frac{1}{T}\tilde{F}'\left(F^0 - \tilde{F}\tilde{H}^{-1}\right)\lambda_i^0 + \frac{1}{T}\left(\tilde{F} - F^0\tilde{H}\right)'\underline{e}_i + \frac{1}{T}\tilde{H}'F^{0\prime}\underline{e}_i,\end{aligned}$$

so that

$$\tilde{\lambda}_i - \tilde{H}^{-1}\lambda_i^0 = -\frac{1}{T}\tilde{F}'\left(\tilde{F} - F^0\tilde{H}\right)\tilde{H}^{-1}\lambda_i^0 + \frac{1}{T}\left(\tilde{F} - F^0\tilde{H}\right)'\underline{e}_i + \frac{1}{T}\tilde{H}'F^{0\prime}\underline{e}_i.$$

We will study the rate of convergence of each term on the right.
The easiest term to deal with is the rightmost term:

$$\left|\frac{1}{T}\tilde{H}'F^{0\prime}\underline{e}_i\right| \leq \frac{1}{\sqrt{T}}\left\|\tilde{H}\right\| \cdot \left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}F_t^0 e_{it}\right|,$$

where

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}F_t^0 e_{it} = O_p(1)$$

by assumption 6, so

$$\frac{1}{T}\tilde{H}'F^{0\prime}\underline{e}_i = O_p\left(\frac{1}{\sqrt{T}}\right).$$

1.6.1 The Rate of Convergence of $\frac{1}{T}\left(\tilde{F}-F^0\tilde{H}\right)'\underline{e}_i$

Using equation (1), we can decompose the term $\frac{1}{T}\left(\tilde{F}-F^0\tilde{H}\right)'\underline{e}_i$ as

$$\begin{aligned}\frac{1}{T}\left(\tilde{F}-F^0\tilde{H}\right)'\underline{e}_i &= V_{NT}^{-1}\frac{1}{T}\left(\hat{F}-F^0H\right)'\underline{e}_i \\ &= V_{NT}^{-1}\left[\frac{1}{NT^2}\tilde{F}'F^0\Lambda^{0\prime}e'\cdot\underline{e}_i+\frac{1}{NT^2}\tilde{F}'e\Lambda^0F^{0\prime}\underline{e}_i+\frac{1}{NT^2}\tilde{F}'ee'\cdot\underline{e}_i\right].\end{aligned}$$

We study each term one by one:

1) $\frac{1}{NT^2}\tilde{F}'F^0\Lambda^{0\prime}e'\cdot\underline{e}_i$

As in section 5, we decompose this term as follows:

$$\begin{aligned}\left|\frac{1}{NT^2}\tilde{F}'F^0\Lambda^{0\prime}e'\cdot\underline{e}_i\right|^2 &\leq 2\frac{1}{N^2T^4}\left|\left(\tilde{F}-F^0\tilde{H}\right)'F^0\Lambda^{0\prime}e'\cdot\underline{e}_i\right|^2+2\frac{1}{N^2T^4}\left|\tilde{H}'F^{0\prime}F^0\Lambda^{0\prime}e'\cdot\underline{e}_i\right|^2 \\ &\leq 2\frac{1}{N}\left(\frac{1}{T}\left\|\tilde{F}-F^0\tilde{H}\right\|^2\right)\cdot\operatorname{tr}\left(\frac{F^{0\prime}F^0}{T}\right)\cdot\left(\frac{1}{NT}\sum_{t=1}^{T}\left|\Lambda^{0\prime}e_t\right|^2\right)\cdot\left(\frac{1}{T}\sum_{t=1}^{T}e_{it}^2\right) \\ &\quad+2\frac{1}{NT}\left\|\tilde{H}\right\|^2\cdot\operatorname{tr}\left(\frac{F^{0\prime}F^0}{T}\right)^2\cdot\left(\frac{1}{NT}\sum_{t=1}^{T}\left|\Lambda^{0\prime}e_t\right|^2\right)\cdot\left(\frac{1}{T}\sum_{t=1}^{T}e_{it}^2\right).\end{aligned}$$

Note that

$$\frac{1}{NT}\sum_{t=1}^{T}\left|\Lambda^{0\prime}e_t\right|^2=\frac{1}{NT}\sum_{t=1}^{T}\left|\sum_{j=1}^{N}\lambda_i^0e_{it}\right|^2=O_p(1),$$

and that

$$\mathbb{E}\left|\frac{1}{T}\sum_{t=1}^{T}e_{it}^2\right|=\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T}e_{it}^2\right]=\gamma(0)<+\infty,$$

implying that $\frac{1}{T}\sum_{t=1}^{T}e_{it}^2=O_p(1)$. As such,

$$\left|\frac{1}{NT^2}\tilde{F}'F^0\Lambda^{0\prime}e'\cdot\underline{e}_i\right|^2\leq O_p\left(\frac{1}{N\cdot\min(N,T)}\right)+O_p\left(\frac{1}{NT}\right)=O_p\left(\frac{1}{N\cdot\min(N,T)}\right),$$

which implies that

$$\frac{1}{NT^2}\tilde{F}'F^0\Lambda^{0\prime}e'\cdot\underline{e}_i=O_p\left(\frac{1}{\min(N,\sqrt{NT})}\right).$$

2) $\frac{1}{NT^2}\tilde{F}' e\Lambda^0 F^{0\prime}\underline{e}_i$

As is routine by now, we can see that

$$\begin{aligned}\left|\frac{1}{NT^2}\tilde{F}' e\Lambda^0 F^{0\prime}\underline{e}_i\right|^2 &\leq 2\frac{1}{N^2T^4}\left|\left(\tilde{F}-F^0\tilde{H}\right)' e\Lambda^0 F^{0\prime}\underline{e}_i\right|^2 + 2\frac{1}{N^2T^4}\left|\tilde{H}' F^{0\prime} e\Lambda^0 F^{0\prime}\underline{e}_i\right|^2 \\ &\leq 2\frac{1}{N}\cdot\left(\frac{1}{T}\left\|\tilde{F}-F^0\tilde{H}\right\|^2\right)\cdot\left(\frac{1}{NT}\sum_{t=1}^{T}\left|e_t'\Lambda^0\right|^2\right)\cdot\mathrm{tr}\left(\frac{F^{0\prime}F^0}{T}\right)\cdot\left(\frac{1}{T}\sum_{t=1}^{T}e_{it}^2\right) \\ &\quad + 2\frac{1}{NT}\left\|\tilde{H}\right\|^2\cdot\mathrm{tr}\left(\frac{F^{0\prime}F^0}{T}\right)\cdot\left(\frac{1}{NT}\sum_{t=1}^{T}\left|e_t'\Lambda^0\right|^2\right)\cdot\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}F_t^0 e_{it}\right|^2.\end{aligned}$$

By assumption 6, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}F_t^0 e_{it} = O_p(1)$, and

$$\frac{1}{NT}\sum_{t=1}^{T}\left|e_t'\Lambda^0\right|^2 = \frac{1}{NT}\sum_{t=1}^{T}\left|\sum_{j=1}^{N}\lambda_i^0 e_{it}\right|^2 = O_p(1),$$

so

$$\left|\frac{1}{NT^2}\tilde{F}' e\Lambda^0 F^{0\prime}\underline{e}_i\right|^2 \leq O_p\left(\frac{1}{N\cdot\min(N,T)}\right) + O_p\left(\frac{1}{NT}\right) = O_p\left(\frac{1}{N\cdot\min(N,T)}\right).$$

This tells us that

$$\frac{1}{NT^2}\tilde{F}' e\Lambda^0 F^{0\prime}\underline{e}_i = O_p\left(\frac{1}{\min(N,\sqrt{NT})}\right),$$

so that the two terms examined so far have the same rate of convergence.

3) $\frac{1}{NT^2}\tilde{F}' ee'\cdot\underline{e}_i$

Note that

$$\begin{aligned}\tilde{F}' ee'\cdot\underline{e}_i &= \sum_{t=1}^{T}\tilde{F}' e\cdot e_t e_{it} = \sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{F}_s e_s' e_t e_{it} \\ &= \sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{j=1}^{N}\tilde{F}_s\left(e_{jt}e_{js}-\gamma(t-s)\right)e_{it} + N\cdot\sum_{t=1}^{T}\sum_{s=1}^{T}\gamma(t-s)\tilde{F}_s e_{it},\end{aligned}$$

so that

$$\left|\frac{1}{NT^2}\tilde{F}' ee'\cdot\underline{e}_i\right| \leq \frac{1}{NT^2}\left|\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{j=1}^{N}\tilde{F}_s\left(e_{jt}e_{js}-\gamma(t-s)\right)e_{it}\right| + \frac{1}{T^2}\left|\sum_{t=1}^{T}\sum_{s=1}^{T}\gamma(t-s)\tilde{F}_s e_{it}\right|.$$

We start with the easier second term.

The Second Term

Note that

$$\left|\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\gamma(t-s)\tilde{F}_s e_{it}\right| \leq \left|\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}(\tilde{F}_s-\tilde{H}'F_s^0)\gamma(t-s)e_{it}\right| + \left\|\tilde{H}\right\|\cdot\left|\frac{1}{T^2}\sum_{t=1}^{T}F_s^0\gamma(t-s)e_{it}\right|.$$

By repeated applications of the Cauchy-Schwarz inequality,

$$\begin{aligned}
\left|\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}(\tilde{F}_s-\tilde{H}'F_s^0)\gamma(t-s)e_{it}\right| &\leq \frac{1}{T^2}\sum_{s=1}^{T}\left|\tilde{F}_s-\tilde{H}'F_s^0\right|\cdot\left|\sum_{t=1}^{T}\gamma(t-s)e_{it}\right| \\
&\leq \left(\frac{1}{T}\sum_{s=1}^{T}\left|\tilde{F}_s-\tilde{H}'F_s^0\right|^2\right)^{\frac{1}{2}}\left(\frac{1}{T^3}\sum_{s=1}^{T}\left|\sum_{t=1}^{T}\gamma(t-s)e_{it}\right|^2\right)^{\frac{1}{2}} \\
&\leq \left(\frac{1}{T}\left\|\tilde{F}-F^0\tilde{H}\right\|^2\right)^{\frac{1}{2}}\left(\frac{1}{T^3}\sum_{s=1}^{T}\left(\sum_{t=1}^{T}\gamma(t-s)^2\right)\left(\sum_{t=1}^{T}e_{it}^2\right)\right)^{\frac{1}{2}} \\
&\leq \left(\frac{1}{T}\left\|\tilde{F}-F^0\tilde{H}\right\|^2\right)^{\frac{1}{2}}\left(\left(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\gamma(t-s)^2\right)\left(\frac{1}{T}\sum_{t=1}^{T}e_{it}^2\right)\right)^{\frac{1}{2}} \\
&\leq \sqrt{\frac{Z}{T}}\cdot\left(\frac{1}{T}\left\|\tilde{F}-F^0\tilde{H}\right\|^2\right)^{\frac{1}{2}}\left(\frac{1}{T}\sum_{t=1}^{T}e_{it}^2\right)^{\frac{1}{2}},
\end{aligned}$$

where $Z=\sum_{z=-\infty}^{\infty}\gamma(z)^2<+\infty$ as usual. It follows that

$$\left|\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}(\tilde{F}_s-\tilde{H}'F_s^0)\gamma(t-s)e_{it}\right| = O_p\left(\frac{1}{\min(T,\sqrt{NT})}\right).$$

Meanwhile,

$$\begin{aligned}
\mathbb{E}\left|\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}F_s^0\gamma(t-s)e_{it}\right| &\leq \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\gamma(t-s)\mathbb{E}\left(\left|F_s^0\right||e_{it}|\right) \\
&\leq \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\gamma(t-s)\left(\mathbb{E}\left|F_s^0\right|^2\right)^{\frac{1}{2}}\left(\mathbb{E}|e_{it}|^2\right)^{\frac{1}{2}} \\
&\leq \left(\sup_{s\in N_+}\mathbb{E}\left|F_s^0\right|^2\right)^{\frac{1}{2}}\cdot\gamma(0)^{\frac{1}{2}}\cdot\left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\gamma(t-s)\right) \\
&\leq \left(\sup_{s\in N_+}\mathbb{E}\left|F_s^0\right|^2\right)^{\frac{1}{2}}\cdot\gamma(0)^{\frac{1}{2}}\cdot\frac{1}{T}Z.
\end{aligned}$$

All the terms on the right are finite constants except for $\frac{1}{T}$, so the term inside the expectations is $O_p\left(\frac{1}{T}\right)$.
It follows that

$$\left|\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\gamma(t-s)\tilde{F}_s e_{it}\right| = O_p\left(\frac{1}{\min(T,\sqrt{NT})}\right)+O_p\left(\frac{1}{T}\right) = O_p\left(\frac{1}{\min(T,\sqrt{NT})}\right).$$

The First Term

As for the first term, note that

$$\left|\frac{1}{NT^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{j=1}^{N}\tilde{F}_s\left(e_{jt}e_{js}-\gamma(t-s)\right)e_{it}\right|$$

$$\leq \frac{1}{NT^2}\sum_{s=1}^{T}\left|\tilde{F}_s-\tilde{H}'F_s^0\right|\cdot\left|\sum_{t=1}^{T}\sum_{j=1}^{N}\left(e_{jt}e_{js}-\gamma(t-s)\right)e_{it}\right|$$

$$+\left\|\tilde{H}\right\|\cdot\left(\frac{1}{NT^2}\sum_{t=1}^{T}\left|\sum_{s=1}^{T}\sum_{j=1}^{N}F_s^0\left(e_{jt}e_{js}-\gamma(t-s)\right)\right|\cdot|e_{it}|\right)$$

$$\leq \frac{1}{\sqrt{N}}\left(\frac{1}{T}\sum_{s=1}^{T}\left|\tilde{F}_s-\tilde{H}'F_s^0\right|^2\right)^{\frac{1}{2}}\cdot\left(\frac{1}{NT^3}\sum_{s=1}^{T}\left|\sum_{t=1}^{T}\sum_{j=1}^{N}\left(e_{jt}e_{js}-\gamma(t-s)\right)e_{it}\right|^2\right)^{\frac{1}{2}}$$

$$+\frac{1}{\sqrt{NT}}\left\|\tilde{H}\right\|\cdot\left(\frac{1}{T}\sum_{t=1}^{T}\left|\frac{1}{\sqrt{NT}}\sum_{s=1}^{T}\sum_{j=1}^{N}F_s^0\left(e_{jt}e_{js}-\gamma(t-s)\right)\right|^2\right)^{\frac{1}{2}}\cdot\left(\frac{1}{T}\sum_{t=1}^{T}e_{it}^2\right)^{\frac{1}{2}}.$$

By assumption, there exists an $M>0$ such that

$$\mathbb{E}\left|\frac{1}{\sqrt{NT}}\sum_{s=1}^{T}\sum_{j=1}^{N}F_s^0\left(e_{jt}e_{js}-\gamma(t-s)\right)\right|^2<M$$

for any $t\in N_+$ and $N,T\in N_+$, so

$$\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T}\left|\frac{1}{\sqrt{NT}}\sum_{s=1}^{T}\sum_{j=1}^{N}F_s^0\left(e_{jt}e_{js}-\gamma(t-s)\right)\right|^2\right]\leq M$$

and therefore the rightmost term is $O_p\left(\frac{1}{\sqrt{NT}}\right)$.

On the other hand, for any $s\in N_+$,

$$\frac{1}{NT^2}\left|\sum_{t=1}^{T}\sum_{j=1}^{N}\left(e_{jt}e_{js}-\gamma(t-s)\right)e_{it}\right|^2\leq\left(\frac{1}{NT}\sum_{t=1}^{T}\left|\sum_{j=1}^{N}\left(e_{jt}e_{js}-\gamma(s-t)\right)\right|^2\right)\left(\frac{1}{T}\sum_{t=1}^{T}e_{it}^2\right)$$

by the Cauchy-Schwarz inequality, so that

$$\frac{1}{NT^3}\sum_{s=1}^{T}\left|\sum_{t=1}^{T}\sum_{j=1}^{N}\left(e_{jt}e_{js}-\gamma(t-s)\right)e_{it}\right|^2\leq\left(\frac{1}{T}\sum_{s=1}^{T}\left[\frac{1}{NT}\sum_{t=1}^{T}\left|\sum_{j=1}^{N}\left(e_{jt}e_{js}-\gamma(s-t)\right)\right|^2\right]\right)\cdot\left(\frac{1}{T}\sum_{t=1}^{T}e_{it}^2\right).$$

We saw in section 3 that

$$\mathbb{E}\left[\frac{1}{NT}\sum_{t=1}^{T}\left|\sum_{j=1}^{N}\left(e_{jt}e_{js}-\gamma(s-t)\right)\right|^2\right]\leq\mu_4,$$

for any $s \in N_+$ and $N,T \in N_+$, so $\frac{1}{T}\sum_{s=1}^{T}\left[\frac{1}{NT}\sum_{t=1}^{T}\left|\sum_{j=1}^{N}\left(e_{jt}e_{js}-\gamma(s-t)\right)\right|^2\right]$ has bounded first moments for any $N,T \in N_+$ as well, which tells us that the term inside the expectation is $O_p(1)$. It follows that

$$\frac{1}{NT^3}\sum_{s=1}^{T}\left|\sum_{t=1}^{T}\sum_{j=1}^{N}\left(e_{jt}e_{js}-\gamma(t-s)\right)e_{it}\right|^2 = O_p(1),$$

so that

$$\left|\frac{1}{NT^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{j=1}^{N}\tilde{F}_s\left(e_{jt}e_{js}-\gamma(t-s)\right)e_{it}\right| = O_p\left(\frac{1}{\min(N,\sqrt{NT})}\right)+O_p\left(\frac{1}{\sqrt{NT}}\right) = O_p\left(\frac{1}{\min(N,\sqrt{NT})}\right).$$

Putting the two results together, we have

$$\frac{1}{NT^2}\tilde{F}'ee'\cdot\underline{e}_i = O_p\left(\frac{1}{\min(T,\sqrt{NT})}\right)+O_p\left(\frac{1}{\min(N,\sqrt{NT})}\right).$$

Results 1) to 3) reveal that

$$\frac{1}{T}\left(\tilde{F}-F^0\tilde{H}\right)'\underline{e}_i = O_p\left(\frac{1}{\min(T,\sqrt{NT})}\right)+O_p\left(\frac{1}{\min(N,\sqrt{NT})}\right).$$

Since

$$\min(\min(T,\sqrt{NT}),\min(N,\sqrt{NT})) = \min(N,T),$$

we have

$$\frac{1}{T}\left(\tilde{F}-F^0\tilde{H}\right)'\underline{e}_i = O_p\left(\frac{1}{\min(N,T)}\right).$$

1.6.2 The Rate of Convergence of $\frac{1}{T}\left(\tilde{F}-F^0\tilde{H}\right)'F^0$

We once again turn to the decomposition

$$\frac{1}{T}\left(\tilde{F}-F^0\tilde{H}\right)'F^0 = V_{NT}^{-1}\left[\frac{1}{NT^2}\tilde{F}'F^0\Lambda^{0\prime}e'F^0 + \frac{1}{NT^2}\tilde{F}'e\Lambda^0F^{0\prime}F^0 + \frac{1}{NT^2}\tilde{F}'ee'F^0\right].$$

We study each term one by one:

1) $\frac{1}{NT^2}\tilde{F}'F^0\Lambda^{0\prime}e'F^0$

Note that

$$\begin{aligned}\left\|\frac{1}{NT^2}\tilde{F}'F^0\Lambda^{0\prime}e'F^0\right\|^2 &\leq 2\frac{1}{N^2T^4}\left\|\tilde{F}-F^0\tilde{H}\right\|^2\cdot\left\|F^0\right\|^2\cdot\left\|\Lambda^{0\prime}e'F^0\right\|^2 + 2\frac{1}{N^2T^4}\left\|\tilde{H}\right\|^2\cdot\left\|F^{0\prime}F^0\Lambda^{0\prime}e'F^0\right\|^2\\ &\leq 2\frac{1}{NT}\cdot\left(\frac{1}{T}\left\|\tilde{F}-F^0\tilde{H}\right\|^2\right)\cdot\mathrm{tr}\left(\frac{F^{0\prime}F^0}{T}\right)\cdot\left(\frac{1}{NT}\left\|\Lambda^{0\prime}e'F^0\right\|^2\right)\\ &\quad+2\frac{1}{NT}\left\|\tilde{H}\right\|^2\cdot\mathrm{tr}\left(\frac{F^{0\prime}F^0}{T}\right)^2\cdot\left(\frac{1}{NT}\left\|\Lambda^{0\prime}e'F^0\right\|^2\right).\end{aligned}$$

$\Lambda^{0\prime}e'F^0$ can be expressed as

$$\Lambda^{0\prime}e'F^0 = \begin{pmatrix}\Lambda^{0\prime}e_1 & \cdots & \Lambda^{0\prime}e_T\end{pmatrix}\begin{pmatrix}F_1^{0\prime}\\ \vdots \\ F_T^{0\prime}\end{pmatrix} = \sum_{t=1}^{T}\Lambda^{0\prime}e_tF_t^{0\prime},$$

and since $\Lambda^{0\prime}e_t = \sum_{i=1}^N\lambda_i^0e_{it}$ for each $t\in N_+$, we have

$$\Lambda^{0\prime}e'F^0 = \sum_{t=1}^{T}\sum_{i=1}^{N}\lambda_i^0F_t^{0\prime}e_{it}.$$

By assumption 5,

$$\frac{1}{NT}\left\|\Lambda^{0\prime}e'F^0\right\|^2 = \frac{1}{NT}\left\|\sum_{t=1}^{T}\sum_{i=1}^{N}\lambda_i^0F_t^{0\prime}e_{it}\right\|^2$$

is $O_p(1)$, so

$$\left\|\frac{1}{NT^2}\tilde{F}'F^0\Lambda^{0\prime}e'F^0\right\|^2 \leq O_p\left(\frac{1}{NT\cdot\min(N,T)}\right)+O_p\left(\frac{1}{NT}\right) = O_p\left(\frac{1}{NT}\right).$$

By implication,

$$\frac{1}{NT^2}\tilde{F}'F^0\Lambda^{0\prime}e'F^0 = O_p\left(\frac{1}{\sqrt{NT}}\right).$$

2) $\frac{1}{NT^2}\tilde{F}'e\Lambda^0 F^{0\prime}F^0$

As usual, we majorize the above term as follows:

$$\begin{aligned}\left\|\frac{1}{NT^2}\tilde{F}'e\Lambda^0 F^{0\prime}F^0\right\|^2 &\leq 2\frac{1}{N^2T^4}\left\|\tilde{F}-F^0\tilde{H}\right\|^2\cdot\left\|e\Lambda^0\right\|^2\cdot\left\|F^0\right\|^4+2\frac{1}{N^2T^4}\left\|\tilde{H}\right\|^2\cdot\left\|F^{0\prime}e\Lambda^0\right\|^2\cdot\left\|F^0\right\|^4\\&\leq 2\frac{1}{N}\left(\frac{1}{T}\left\|\tilde{F}-F^0\tilde{H}\right\|^2\right)\cdot\left(\frac{1}{NT}\sum_{t=1}^{T}\left|\sum_{i=1}^{N}\lambda_i^0 e_{it}\right|^2\right)\cdot\mathrm{tr}\left(\frac{F^{0\prime}F^0}{T}\right)^2\\&\quad+2\frac{1}{NT}\left\|\tilde{H}\right\|^2\cdot\left(\frac{1}{NT}\left\|F^{0\prime}e\Lambda^0\right\|^2\right)\cdot\mathrm{tr}\left(\frac{F^{0\prime}F^0}{T}\right)^2.\end{aligned}$$

Since $F^{0\prime}e\Lambda^0=\left(\Lambda^{0\prime}e'F^0\right)'$, all the components on the rightmost term except for $\frac{1}{NT}$ is $O_p(1)$. By implication,

$$\left\|\frac{1}{NT^2}\tilde{F}'e\Lambda^0 F^{0\prime}F^0\right\|^2\leq O_p\left(\frac{1}{N\cdot\min(N,T)}\right)+O_p\left(\frac{1}{NT}\right)=O_p\left(\frac{1}{N\cdot\min(N,T)}\right),$$

and as such

$$\frac{1}{NT^2}\tilde{F}'e\Lambda^0 F^{0\prime}F^0=O_p\left(\frac{1}{\min(N,\sqrt{NT})}\right).$$

3) $\frac{1}{NT^2}\tilde{F}'ee'F^0$

The proof for this component almost exactly mirrors that of the previous subsection for $\frac{1}{NT^2}\tilde{F}'ee'\underline{e}_i$. We re-state it for the sake of completeness:

Note that

$$\begin{aligned}\tilde{F}'ee'F^0&=\sum_{t=1}^{T}\tilde{F}'e\cdot e_tF_t^{0\prime}=\sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{F}_se_s'e_tF_t^{0\prime}\\&=\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{j=1}^{N}\tilde{F}_s\left(e_{jt}e_{js}-\gamma(t-s)\right)F_t^{0\prime}+N\cdot\sum_{t=1}^{T}\sum_{s=1}^{T}\gamma(t-s)\tilde{F}_sF_t^{0\prime},\end{aligned}$$

so that

$$\left\|\frac{1}{NT^2}\tilde{F}'ee'F^0\right\|\leq\frac{1}{NT^2}\left\|\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{j=1}^{N}\tilde{F}_s\left(e_{jt}e_{js}-\gamma(t-s)\right)F_t^{0\prime}\right\|+\frac{1}{T^2}\left\|\sum_{t=1}^{T}\sum_{s=1}^{T}\gamma(t-s)\tilde{F}_sF_t^{0\prime}\right\|.$$

Once again, we start with the easier second term.

The Second Term

Note that

$$\left\|\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\gamma(t-s)\tilde{F}_sF_t^{0\prime}\right\| \leq \left\|\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}(\tilde{F}_s-\tilde{H}'F_s^0)\gamma(t-s)F_t^{0\prime}\right\| + \left\|\tilde{H}\right\|\cdot\left\|\frac{1}{T^2}\sum_{t=1}^{T}F_s^0\gamma(t-s)F_t^{0\prime}\right\|.$$

By repeated applications of the Cauchy-Schwarz inequality,

$$\begin{aligned}
\left\|\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}(\tilde{F}_s-\tilde{H}'F_s^0)\gamma(t-s)F_t^{0\prime}\right\| &\leq \frac{1}{T^2}\sum_{s=1}^{T}\left|\tilde{F}_s-\tilde{H}'F_s^0\right|\cdot\left|\sum_{t=1}^{T}\gamma(t-s)F_t^0\right| \\
&\leq \left(\frac{1}{T}\sum_{s=1}^{T}\left|\tilde{F}_s-\tilde{H}'F_s^0\right|^2\right)^{\frac{1}{2}}\left(\frac{1}{T^3}\sum_{s=1}^{T}\left|\sum_{t=1}^{T}\gamma(t-s)F_t^0\right|^2\right)^{\frac{1}{2}} \\
&\leq \left(\frac{1}{T}\left\|\tilde{F}-F^0\tilde{H}\right\|^2\right)^{\frac{1}{2}}\left(\frac{1}{T^3}\sum_{s=1}^{T}\left(\sum_{t=1}^{T}\gamma(t-s)^2\right)\left(\sum_{t=1}^{T}\left|F_t^0\right|^2\right)\right)^{\frac{1}{2}} \\
&\leq \sqrt{\frac{Z}{T}}\cdot\left(\frac{1}{T}\left\|\tilde{F}-F^0\tilde{H}\right\|^2\right)^{\frac{1}{2}}\operatorname{tr}\left(\frac{F^{0\prime}F^0}{T}\right)^{\frac{1}{2}},
\end{aligned}$$

where $Z=\sum_{z=-\infty}^{\infty}\gamma(z)^2<+\infty$. It follows that

$$\left\|\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}(\tilde{F}_s-\tilde{H}'F_s^0)\gamma(t-s)F_t^{0\prime}\right\| = O_p\left(\frac{1}{\min(T,\sqrt{NT})}\right).$$

Meanwhile,

$$\begin{aligned}
\mathbb{E}\left\|\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}F_s^0\gamma(t-s)F_t^{0\prime}\right\| &\leq \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\gamma(t-s)\mathbb{E}\left(\left|F_s^0\right|\left|F_t^0\right|\right) \\
&\leq \left(\sup_{s\in N_+}\mathbb{E}\left|F_s^0\right|^2\right)\cdot\gamma(0)^{\frac{1}{2}}\cdot\left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\gamma(t-s)\right) \\
&\leq \left(\sup_{s\in N_+}\mathbb{E}\left|F_s^0\right|^2\right)\cdot\frac{1}{T}Z.
\end{aligned}$$

All the terms on the right are finite constants except for $\frac{1}{T}$, so the term inside the expectations is $O_p\left(\frac{1}{T}\right)$.

It follows that

$$\left\|\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\gamma(t-s)\tilde{F}_sF_t^{0\prime}\right\| = O_p\left(\frac{1}{\min(T,\sqrt{NT})}\right)+O_p\left(\frac{1}{T}\right) = O_p\left(\frac{1}{\min(T,\sqrt{NT})}\right).$$

The First Term

As for the first term, note that

$$\left\| \frac{1}{NT^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{j=1}^{N} \tilde{F}_s \left(e_{jt}e_{js} - \gamma(t-s)\right) F_t^{0\prime} \right\|$$
$$\leq \frac{1}{NT^2} \sum_{s=1}^{T} \left| \tilde{F}_s - \tilde{H}' F_s^0 \right| \cdot \left| \sum_{t=1}^{T} \sum_{j=1}^{N} \left(e_{jt}e_{js} - \gamma(t-s)\right) F_t^0 \right|$$
$$+ \left\| \tilde{H} \right\| \cdot \left(\frac{1}{NT^2} \sum_{t=1}^{T} \left| \sum_{s=1}^{T} \sum_{j=1}^{N} F_s^0 \left(e_{jt}e_{js} - \gamma(t-s)\right) \right| \cdot \left| F_t^0 \right| \right)$$
$$\leq \frac{1}{\sqrt{N}} \left(\frac{1}{T} \sum_{s=1}^{T} \left| \tilde{F}_s - \tilde{H}' F_s^0 \right|^2 \right)^{\frac{1}{2}} \cdot \left(\frac{1}{NT^3} \sum_{s=1}^{T} \left| \sum_{t=1}^{T} \sum_{j=1}^{N} \left(e_{jt}e_{js} - \gamma(t-s)\right) F_t^0 \right|^2 \right)^{\frac{1}{2}}$$
$$+ \frac{1}{\sqrt{NT}} \left\| \tilde{H} \right\| \cdot \left(\frac{1}{T} \sum_{t=1}^{T} \left| \frac{1}{\sqrt{NT}} \sum_{s=1}^{T} \sum_{j=1}^{N} F_s^0 \left(e_{jt}e_{js} - \gamma(t-s)\right) \right|^2 \right)^{\frac{1}{2}} \cdot \mathrm{tr} \left(\frac{F^{0\prime} F^0}{T} \right)^{\frac{1}{2}}.$$

We saw in the previous subsection that

$$\frac{1}{T} \sum_{t=1}^{T} \left| \frac{1}{\sqrt{NT}} \sum_{s=1}^{T} \sum_{j=1}^{N} F_s^0 \left(e_{jt}e_{js} - \gamma(t-s)\right) \right|^2 = O_p(1),$$

so the rightmost term is $O_p\left(\frac{1}{\sqrt{NT}}\right)$.

On the other hand, for any $s \in N_+$,

$$\frac{1}{NT^2} \left| \sum_{t=1}^{T} \sum_{j=1}^{N} \left(e_{jt}e_{js} - \gamma(t-s)\right) F_t^0 \right|^2 \leq \left(\frac{1}{NT} \sum_{t=1}^{T} \left| \sum_{j=1}^{N} \left(e_{jt}e_{js} - \gamma(s-t)\right) \right|^2 \right) \mathrm{tr} \left(\frac{F^{0\prime} F^0}{T} \right)$$

by the Cauchy-Schwarz inequality, so that

$$\frac{1}{NT^3} \sum_{s=1}^{T} \left| \sum_{t=1}^{T} \sum_{j=1}^{N} \left(e_{jt}e_{js} - \gamma(t-s)\right) F_t^0 \right|^2 \leq \left(\frac{1}{T} \sum_{s=1}^{T} \left[\frac{1}{NT} \sum_{t=1}^{T} \left| \sum_{j=1}^{N} \left(e_{jt}e_{js} - \gamma(s-t)\right) \right|^2 \right] \right) \cdot \mathrm{tr} \left(\frac{F^{0\prime} F^0}{T} \right).$$

Again, we saw in the previous subsection that

$$\frac{1}{T} \sum_{s=1}^{T} \left[\frac{1}{NT} \sum_{t=1}^{T} \left| \sum_{j=1}^{N} \left(e_{jt}e_{js} - \gamma(s-t)\right) \right|^2 \right] = O_p(1),$$

which tells us that

$$\frac{1}{NT^3} \sum_{s=1}^{T} \left| \sum_{t=1}^{T} \sum_{j=1}^{N} \left(e_{jt}e_{js} - \gamma(t-s)\right) F_t^0 \right|^2 = O_p(1).$$

Therefore,

$$\left\| \frac{1}{NT^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{j=1}^{N} \tilde{F}_s \left(e_{jt} e_{js} - \gamma(t-s) \right) F_t^{0\prime} \right\| = O_p\left(\frac{1}{\min(N, \sqrt{NT})} \right) + O_p\left(\frac{1}{\sqrt{NT}} \right) = O_p\left(\frac{1}{\min(N, \sqrt{NT})} \right).$$

Putting the two results together, we have

$$\frac{1}{NT^2} \tilde{F}' e e' F^0 = O_p\left(\frac{1}{\min(T, \sqrt{NT})} \right) + O_p\left(\frac{1}{\min(N, \sqrt{NT})} \right).$$

Results 1) to 3) reveal that, as in the previous subsection,

$$\begin{aligned} \frac{1}{T} \left(\tilde{F} - F^0 \tilde{H} \right)' F^0 &= O_p\left(\frac{1}{\min(T, \sqrt{NT})} \right) + O_p\left(\frac{1}{\min(N, \sqrt{NT})} \right) + O_p\left(\frac{1}{\sqrt{NT}} \right) \\ &= O_p\left(\frac{1}{\min(T, \sqrt{NT})} \right) + O_p\left(\frac{1}{\min(N, \sqrt{NT})} \right). \end{aligned}$$

Once again, this implies

$$\frac{1}{T} \left(\tilde{F} - F^0 \tilde{H} \right)' F^0 = O_p\left(\frac{1}{\min(N, T)} \right).$$

We can now easily recover the rate of convergence of $\frac{1}{T} \left(\tilde{F} - F^0 \tilde{H} \right)' \tilde{F}$;

$$\frac{1}{T} \left(\tilde{F} - F^0 \tilde{H} \right)' \tilde{F} = \frac{1}{T} \left(\tilde{F} - F^0 \tilde{H} \right)' \left(\tilde{F} - F^0 \tilde{H} \right) + \frac{1}{T} \left(\tilde{F} - F^0 \tilde{H} \right)' F^0 \tilde{H},$$

so

$$\left\| \frac{1}{T} \left(\tilde{F} - F^0 \tilde{H} \right)' \tilde{F} \right\| \leq \frac{1}{T} \left\| \tilde{F} - F^0 \tilde{H} \right\|^2 + \frac{1}{T} \left\| \left(\tilde{F} - F^0 \tilde{H} \right)' F^0 \right\| \left\| \tilde{H} \right\|.$$

We already know that

$$\begin{aligned} \frac{1}{T} \left\| \tilde{F} - F^0 \tilde{H} \right\|^2 &= O_p\left(\frac{1}{\min(N, T)} \right) \\ \frac{1}{T} \left\| \left(\tilde{F} - F^0 \tilde{H} \right)' F^0 \right\| &= O_p\left(\frac{1}{\min(N, T)} \right); \end{aligned}$$

therefore,

$$\frac{1}{T} \left(\tilde{F} - F^0 \tilde{H} \right)' \tilde{F} = O_p\left(\frac{1}{\min(N, T)} \right)$$

as well.

1.6.3 The Asymptotic Distribution of $\sqrt{T}\left(\tilde{\lambda}_i - \tilde{H}^{-1}\lambda_i\right)$

Return to the equation

$$\tilde{\lambda}_i - \tilde{H}^{-1}\lambda_i^0 = -\frac{1}{T}\tilde{F}'\left(\tilde{F} - F^0\tilde{H}\right)\tilde{H}^{-1}\lambda_i^0 + \frac{1}{T}\left(\tilde{F} - F^0\tilde{H}\right)'\underline{e}_i + \frac{1}{T}\tilde{H}'F^{0\prime}\underline{e}_i.$$

We have shown that:

$$\frac{1}{T}\tilde{F}'\left(\tilde{F} - F^0\tilde{H}\right) = O_p\left(\frac{1}{\min(N,T)}\right),$$
$$\frac{1}{T}\left(\tilde{F} - F^0\tilde{H}\right)'\underline{e}_i = O_p\left(\frac{1}{\min(N,T)}\right),$$
$$\frac{1}{T}\tilde{H}'F^{0\prime}\underline{e}_i = O_p\left(\frac{1}{\sqrt{T}}\right).$$

Therefore,

$$\tilde{\lambda}_i - \tilde{H}^{-1}\lambda_i^0 = O_p\left(\frac{1}{\min(N,T)}\right) + O_p\left(\frac{1}{\sqrt{T}}\right).$$

If $\frac{\sqrt{T}}{N} \to 0$ as $N,T \to \infty$, then

$$\frac{\frac{1}{\min(N,T)}}{\frac{1}{\sqrt{T}}} = \frac{1}{\min\left(\frac{N}{\sqrt{T}}, \sqrt{T}\right)} \to 0$$

as $N,T \to \infty$, meaning that $\frac{1}{\min(N,T)}$ converges to 0 faster than $\frac{1}{\sqrt{T}}$. Thus, in this case,

$$\tilde{\lambda}_i - \tilde{H}^{-1}\lambda_i^0 = O_p\left(\frac{1}{\sqrt{T}}\right),$$

and because the rightmost term in the original equation is the unique $O_p\left(\frac{1}{\sqrt{T}}\right)$ term, we can write

$$\sqrt{T}\left(\tilde{\lambda}_i - \tilde{H}^{-1}\lambda_i^0\right) = \frac{1}{\sqrt{T}}\tilde{H}'F^{0\prime}\underline{e}_i + o_p(1).$$

By assumption,

$$\frac{1}{\sqrt{T}}F^{0\prime}\underline{e}_i = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}F_t^0 e_{it} \xrightarrow{d} N\left[\mathbf{0}, \Phi_i\right],$$

where

$$\Phi_i = \operatorname*{plim}_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}e_{it}^2 F_t^0 F_t^{0\prime},$$

and because

$$\tilde{H} = \left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)\left(\frac{F^{0\prime}\tilde{F}}{T}\right)V_{NT}^{-1},$$

we have

$$\tilde{H} \xrightarrow{p} \Sigma_\Lambda Q V^{-1}.$$

By Slutsky's theorem,

$$\sqrt{T}\left(\tilde{\lambda}_i - \tilde{H}^{-1}\lambda_i^0\right) \xrightarrow{d} N\left[\mathbf{0}, \Sigma_\Lambda Q V^{-1}\Phi_i V^{-1} Q' \Sigma_\Lambda\right].$$

Again, we can further simplify the asymptotic variance above by noting that the probability limit Q of $\frac{F^{0\prime}\tilde{F}}{T}$ is given as

$$Q = \Sigma_\Lambda^{-\frac{1}{2}\prime}\Gamma^* V^{\frac{1}{2}},$$

where Γ^* collects the unique (up to sign changes) orthonormal eigenbasis of $\Sigma_\Lambda^{\frac{1}{2}\prime}\Sigma_F\Sigma_\Lambda^{\frac{1}{2}}$. Thus,

$$\Sigma_\Lambda Q V^{-1} = \Sigma_\Lambda \Sigma_\Lambda^{-\frac{1}{2}\prime}\Gamma^* V^{\frac{1}{2}} V^{-1} = \Sigma_\Lambda^{\frac{1}{2}}\Gamma^* V^{-\frac{1}{2}} = Q'^{-1},$$

using the property that $\Gamma^{*-1} = \Gamma^{*\prime}$. The asymptotic distribution of $\tilde{\lambda}_i$ now becomes

$$\sqrt{T}\left(\tilde{\lambda}_i - \tilde{H}^{-1}\lambda_i^0\right) \xrightarrow{d} N\left[\mathbf{0}, Q'^{-1}\Phi_i Q^{-1}\right].$$

1.7 Asymptotic Theory: Summary

We now summarize our findings to this point.
Consider the factor model given as

$$X_{it} = \lambda_i^{0\prime} F_t^0 + e_{it}$$

for any $1 \leq i \leq N$, $1 \leq t \leq T$. Organizing the data into the full panel

$$X = F^0 \Lambda^{0\prime} + e$$

as above, the estimator of F and Λ found by minimizing the objective function

$$\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(X_{it} - \lambda_i' F_t\right)^2 = \frac{1}{NT} \operatorname{tr}\left((X - F\Lambda')(X - F\Lambda)'\right)$$

with respect to F and Λ.

The following assumptions are made:

(1) **Non-triviality of Scaled Factors**
We assume that there exists a $k_{max} \in N_+$ such that $r < k_{max}$ and the k_{max} largest eigenvalues of XX' are always positive. This implies that the k largest eigenvalues of XX' are always positive for $1 \leq k \leq k_{max}$, and as such that, when we use the scaled factors $\hat{F}^k = \frac{1}{NT} XX' \tilde{F}^k$ later on, the scaled factors are non-zero, or non-trivial.
Additionally, we assume the true number of factors r satisfies $r < k_{max}$.

(2) **Second Moment Convergence of True Factors and Factor Loadings**
We assume that there exists an $M > 0$ such that

$$\sup_{t \in N_+} \mathbb{E} \left| F_t^0 \right|^2 \leq M,$$

and that the factor loadings $\lambda_1^0, \cdots, \lambda_N^0$ are nonrandom.
In addition, we assume that

$$\frac{F^{0\prime} F^0}{T} \xrightarrow{p} \Sigma_F \quad \text{and} \quad \frac{\Lambda^{0\prime} \Lambda^0}{T} \to \Sigma_\Lambda$$

for some positive definite matrices $\Sigma_F . \Sigma_\Lambda \in \mathbb{R}^{r \times r}$.

(3) **Exact Factor Model**
We assume that the processes $\{e_{it}\}_{t \in \mathbb{Z}}$ are independent and identically distributed for any $i \in N_+$.

(4) **Stationarity of Errors**

We assume that $\{e_{it}\}_{t\in\mathbb{Z}}$ is weakly stationary with mean 0 and autocovariance function $\gamma : \mathbb{Z} \to \mathbb{R}$.

In addition, we assume that the autocovariances are absolutely summable and that the time series has bounded fourth moments, that is, there exists an $\mu_4 < +\infty$ such that $\mathbb{E}\left[e_{it}^4\right] < \mu_4$ for any $t \in N_+$.

(5) **Weak Dependence between Factors and Errors**

There exists an $M > 0$ such that

$$\mathbb{E}\left[\frac{1}{NT}\sum_{i=1}^{N}\left|\sum_{t=1}^{T}F_t^0 e_{it}\right|^2\right] \leq M$$

$$\mathbb{E}\left|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1}^{T}F_s^0\left(e_{it}e_{is} - \gamma(t-s)\right)\right|^2 \leq M \qquad \text{(for any } t \in N_+\text{)}$$

$$\mathbb{E}\left\|\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\sum_{i=1}^{N}F_t^0\lambda_i^{0\prime}e_{it}\right\|^2 \leq M$$

for any $N, T \in N_+$.

(6) **CLT for Time Dimension**

For any $i \in N_+$,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}F_t^0 e_{it} \xrightarrow{d} N\left[\mathbf{0}, \Phi_i\right],$$

for the positive definite matrix

$$\Phi_i = \operatorname*{plim}_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}e_{it}^2 F_t^0 F_t^{0\prime}.$$

(7) **Sufficient Conditions for Factor Identification**

The k_{max} largest eigenvalues of XX' are distinct for any $N, T \in N_+$ such that $T \geq k_{max}$. Likewise, the $r \times r$ matrix $\Sigma_\Lambda \Sigma_F$ has distinct eigenvalues.

(8) **The Probability Limit of $\frac{F^{0\prime}\tilde{F}^k}{T}$**

We assume that, for any $1 \leq k \leq k_{max}$, there exists an $r \times k$ matrix Q^k of full rank such that

$$\frac{F^{0\prime}\tilde{F}^k}{T} \xrightarrow{p} Q^k.$$

1.7.1 Estimation and Determination of Number of Factors

Suppose we estimate $1 \leq k \leq k_{max}$ factors. Solving for Λ first, we obtain the concentrated objective function

$$V(k,F) = \frac{1}{NT}\operatorname{tr}(XX') - \frac{1}{T}\operatorname{tr}\left(X'F(F'F)^{-1}F'X\right),$$

and finding the $T \times k$ matrix F that minimizes this expression and satisfies $\frac{F'F}{T}$ is given as

$$\tilde{F}^k = \sqrt{T} \times \text{The orthonormal eigenvectors of } XX' \text{ corresponding to its } k \text{ largest eigenvalues,}$$

and the associated factor loading estimators are $\tilde{\Lambda}^k = \frac{1}{T}X'\tilde{F}^k$.

On the other hand, solving for F first, we obtain the concentrated objective function

$$\tilde{V}(k,\Lambda) = \frac{1}{NT}\operatorname{tr}(XX') - \frac{1}{T}\operatorname{tr}\left(X\Lambda(\Lambda'\Lambda)^{-1}\Lambda'X'\right),$$

and finding the $N \times k$ matrix Λ that minimizes this expression and satisfies $\frac{\Lambda'\Lambda}{N}$ is given as

$$\bar{\Lambda}^k = \sqrt{N} \times \text{The orthonormal eigenvectors of } X'X \text{ corresponding to its } k \text{ largest eigenvalues,}$$

and the associated factor estimators are $\bar{F}^k = \frac{1}{N}X\bar{\Lambda}^k$.

To choose the number of factors, we can make use of information criteria of the form

$$PC(k) = V(k,\tilde{F}^k) + kg(N,T) = \frac{1}{NT}\operatorname{tr}\left(XX'\right) - \frac{1}{NT}\sum_{i=1}^{k}\mu_i + kg(N,T),$$

where $\mu_1 \geq \cdots \geq \mu_k > 0$ are the k largest eigenvalues of $\frac{1}{NT}XX'$ and $g(N,T)$ is a penalty term. Under the assumption that $\frac{\tilde{F}^{k\prime}F^0}{T}$ converges in probability to some $k \times r$ matrix of full rank, it can be shown that the value k^* that minimizes the above information criterion consistently estimates the true number of factors r if, as $N,T \to \infty$,

- $g(N,T) \to 0$, and
- $\min(N,T) \cdot g(N,T) \to +\infty$; $g(N,T)$ goes to 0 at a rate slower than $\frac{1}{\min(N,T)}$.

The first condition ensures that k is not chosen to be smaller than r, while the second condition is needed for k to not be chosen as a value larger than r.

To control for scale effects, the information criterion

$$IC(k) = \log\left(V(k,\hat{F}^k)\right) + k\frac{\log(\delta_{NT})}{\delta_{NT}}$$

is proposed, where each δ_{NT} can be replaced by $\frac{NT}{N+T}$.

1.7.2 Asymptotic Distributions of $\tilde{F}_t$ and $\bar{\lambda}_i$

Now suppose that the true number of factors r is known, and let $\tilde{F} = \tilde{F}^r$ and $\bar{\Lambda} = \bar{\Lambda}^r$. Under the above assumptions, we can show that

$$V_{NT} \xrightarrow{p} V,$$

where V_{NT} collects the r largest eigenvalues of $\frac{1}{NT}XX'$, or equivalently $\frac{1}{NT}X'X$, and V the eigenvalues of $\Sigma_\Lambda \Sigma_F$.

In addition, we obtain the following consisteny results:

$$\frac{1}{T}\sum_{t=1}^{T}\left|\tilde{F}_t - \tilde{H}'F_t^0\right|^2 = \frac{1}{T}\left\|\tilde{F} - F^0\tilde{H}\right\|^2 = O_p(\delta_{NT}^{-1}),$$

where

$$\tilde{H} = \left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)\left(\frac{F^{0\prime}\tilde{F}}{T}\right)V_{NT}^{-1}$$

and $\delta_{NT} = \min(N,T)$. This tells us that the estimated factors are consistent up to a rotation of the true factors and factor loadings, where the rotation is given by $\tilde{H}$. The failure of exact consistency makes sense because the estimator $\tilde{F}$ is itself a normalized rotation out of an infinite number of possible minimizers of the concentrated objective function.

The above consistency result can now be used to derive a tractable expansion of individual factors. For any $t \in N_+$, we can see that

$$\sqrt{N}\left(\tilde{F}_t - \tilde{H}'F_t^0\right) = V_{NT}^{-1}\left(\frac{\tilde{F}'F^0}{T}\right)\left(\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\lambda_j^0 e_{jt}\right) + o_p(1)$$

if $\frac{\sqrt{N}}{T} \to 0$ as $N,T \to \infty$.

Likewise for the factor loadings, for any $i \in N_+$,

$$\sqrt{T}\left(\tilde{\lambda}_i - \tilde{H}^{-1}\cdot\lambda_i^0\right) = \frac{1}{\sqrt{T}}\tilde{H}'F^{0\prime}\underline{e}_i + o_p(1)$$

if $\frac{\sqrt{T}}{N} \to 0$ as $N,T \to \infty$.

It then follows that

$$\sqrt{N}\left(\tilde{F}_t - \tilde{H}'F_t^0\right) \xrightarrow{d} N\left[\mathbf{0}, \gamma(0)\cdot V^{-2}\right]$$
$$\sqrt{T}\left(\bar{\lambda}_i - \bar{G}'\lambda_i^0\right) \xrightarrow{d} N\left[\mathbf{0}, Q'^{-1}\Phi_i Q^{-1}\right],$$

where

$$\Phi_i = \operatorname*{plim}_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} e_{it}^2 F_t^0 F_t^{0\prime}.$$

Note the Relative Growth Rate of N and T. In order for the above asymptotic results to hold, N and T must satisfy

$$\frac{\sqrt{N}}{T}, \frac{\sqrt{T}}{N} \to 0$$

as $N,T \to \infty$. This tells us that N must not grow faster than T^2, nor should T grow faster than N^2. In other words, for the asymptotic results to hold N must not be significantly larger than T, and vice versa.

1.7.3 Asymptotic Principal Components as OLS

Using the fact that

$$\tilde{H}'\left(\frac{\Lambda^{0\prime}\Lambda^{0}}{N}\right)^{-1}=V_{NT}^{-1}\left(\frac{\tilde{F}'F^{0}}{T}\right),$$

we can rewrite equation (1) as

$$\begin{aligned}\sqrt{N}\left(\tilde{F}_t-\tilde{H}'F_t^0\right)&=\tilde{H}'\left(\frac{\Lambda^{0\prime}\Lambda^{0}}{N}\right)^{-1}\frac{1}{\sqrt{N}}\Lambda^{0\prime}e_t+o_p(1)\\&=\sqrt{N}\left[\left(\Lambda^0\tilde{H}'^{-1}\right)'\left(\Lambda^0\tilde{H}'^{-1}\right)\right]^{-1}\left(\Lambda^0\tilde{H}'^{-1}\right)'e_t+o_p(1).\end{aligned}$$

Since

$$X_t=\Lambda^0F_t^0+e_t=\Lambda^0\tilde{H}'^{-1}\left(\tilde{H}'F_t^0\right)+e_t,$$

and the estimator $\tilde{F}_t$ can be written as

$$\begin{aligned}\tilde{F}_t&=\tilde{H}'F_t^0+\left[\left(\Lambda^0\tilde{H}'^{-1}\right)'\left(\Lambda^0\tilde{H}'^{-1}\right)\right]^{-1}\left(\Lambda^0\tilde{H}'^{-1}\right)'e_t+o_p(1)\\&=\left[\left(\Lambda^0\tilde{H}'^{-1}\right)'\left(\Lambda^0\tilde{H}'^{-1}\right)\right]^{-1}\left(\Lambda^0\tilde{H}'^{-1}\right)'X_t+o_p(1),\end{aligned}$$

$\tilde{F}_t$ can be interpreted as the OLS coefficient estimator from the regression of the dependent variable X_t on the regressors $\Lambda^0\tilde{H}'^{-1}$. This is intuitively appealing because, as even Bai and Ng (2002) pointed out, given the factor loadings Λ^0 the equation

$$X_t=\Lambda^0F_t^0+e_t$$

looks like your typical linear regression equation with true coefficients F_t^0. Since we can only consistently estimate the rotation $\tilde{H}'F_t^0$ of the true coefficients, the OLS estimator $\tilde{F}_t$ consistently estimates the coefficients of the equation with rotated regressors

$$X_t=\Lambda^0\tilde{H}'^{-1}\left(\tilde{H}'F_t^0\right)+e_t.$$

Asymptotics for Multilevel Models

Choi et al. (2018)

Now we turn to multilevel factor models, in which the observations are clustered into groups and there exist factors affecting every observation and those that only affect observations in some cluster. This model is mostly used to study cross-country models of the macroeconomy, in which there are global factors affecting every obseration and country factors that affect only the observations in each country.
Analysis of these kinds of models was pioneered in Choi et al. (2018), and involves much of the same principal components machinery as in the usual factor models, except with the inclusion of canonical correlation analysis when deriving the initial global factor estimate.

The model is formulated as follows. Let X_{mit} be the observation of individual i in country m at time t, and assume that there are N_m individuals in country m, so that the total number of cross-sectional observations is $N = N_1 + \cdots + N_m$. Letting G_t be the collection of r global factors at time t and F_{mt} the collection of r_m country-specific factors for country m, with respective factor loadings $\Gamma_{mi} \in \mathbb{R}^r$ and $\lambda_{mi} \in \mathbb{R}^{r_m}$, we assume X_{mit} is determined as

$$X_{mit} = \gamma_{mi}' G_t + \lambda_{mi}' F_{mt} + e_{mit},$$

where e_{mit} is an idiosyncratic error term.
Collecting $X_{mt} = (X_{m1t}, \cdots, X_{m,N_m,t})'$, $\Gamma_m = (\gamma_{m1}, \cdots, \gamma_{m,N_m})'$, $\Lambda_m = (\lambda_{m1}, \cdots, \lambda_{m,N_m})'$ and $e_{mt} = (e_{m1t}, \cdots, e_{m,N_m,t})'$, we can collect the observations for each country m into

$$\begin{aligned}
\underbrace{X_{mt}}_{N_m \times 1} &= \underbrace{\Gamma_m}_{N_m \times r} \cdot \underbrace{G_t}_{r \times 1} + \underbrace{\Lambda_m}_{N_m \times r_m} \cdot \underbrace{F_{mt}}_{r_m \times 1} + \underbrace{e_{mt}}_{N_m \times 1} \\
&= \begin{pmatrix} \Gamma_m & \Lambda_m \end{pmatrix} \begin{pmatrix} G_t \\ F_{mt} \end{pmatrix} + e_{mt} \\
&= \underbrace{\Theta_m}_{N_m \times r + r_m} \cdot \underbrace{K_{mt}}_{r + r_m \times 1} + e_{mt}.
\end{aligned}$$

Further defining $X_m = (X_{m1}, \cdots, X_{mT})'$, $G = (G_1, \cdots, G_T)'$, $F_m = (F_{m1}, \cdots, F_{mT})'$, and $e_m =$

$(e_{m1}, \cdots, e_{mT})'$, we have

$$\begin{aligned}
\underbrace{X_m}_{T \times N_m} &= \underbrace{G}_{T \times r} \cdot \Gamma'_m + \underbrace{F_m}_{T \times r_m} \cdot \Lambda'_m + \underbrace{e_m}_{T \times N_m} \\
&= \begin{pmatrix} G & F_m \end{pmatrix} \cdot \begin{pmatrix} \Gamma'_m \\ \Lambda'_m \end{pmatrix} + e_m \\
&= \underbrace{K_m}_{T \times r + r_m} \cdot \underbrace{\Theta'_m}_{r + r_m \times N_m} + e_m,
\end{aligned}$$

where $K_m = (K_{m1}, \cdots, K_{mT})'$.

2.1 Canonical Correlation Analysis

If PCA is aimed at recovering the linear combinations of a given set of data that best explains the variation contained in that set of data, Canonical Correlation Analysis (CCA) is designed to recover the linear combinations of two different sets of data that best captures the correlation between them.

2.1.1 Population CCA

Formally, suppose that there exist two random vectors X and Y, each of dimension n_1 and n_2. Let the covariance matrix of the n_1+n_2-dimensional random vector $(Y,X)'$ be defined as

$$\Sigma = \begin{pmatrix} \Sigma_Y & \Sigma_{YX} \\ \Sigma_{XY} & \Sigma_X \end{pmatrix},$$

where $\Sigma_{YX} = \mathrm{Cov}(Y,X)$ and Σ_Y, Σ_X are the covariance matrices of X and Y, and Σ, Σ_X, Σ_Y are assumed to be of full rank.
Choosing $k \le \min(n_1,n_2)$, the goal is to find sets $H = (h_1,\cdots,h_k) \in \mathbb{R}^{n_1\times k}$ and $R = (r_1,\cdots,r_k) \in \mathbb{R}^{n_2\times k}$ of weights such that:

$$\begin{aligned} H'\Sigma_Y H &= \mathrm{Cov}(H'Y) = I_k \\ R'\Sigma_X R &= \mathrm{Cov}(R'X) = I_k \\ H'\Sigma_{YX} R &= \mathrm{Cov}(H'Y, R'X) = \begin{pmatrix} \rho_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \rho_k \end{pmatrix} \quad \text{where } \rho_1^2 \ge \cdots \ge \rho_k^2 \end{aligned}$$

and $\sum_{i=1}^k \rho_i^2$ is maximized. Since $\rho_1,\cdots,\rho_k$ can be interpreted as the correlation coefficients of $h_i'Y$, $r_i'X$ for each $1 \le i \le k$ given the normalization of their variances to 0, $H'Y$ and $R'X$ can be viewed as the collection of orthonormal linear combinations of Y and X that have the highest correlations. The values $\rho_1,\cdots,\rho_k$ are therfore called the first k canonical correlations.

We now search for the weights H,R that satisfy the above conditions.
Assuming that $n_2 \le n_1$, we first study some properties of the eigenvalues of the matrix $\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\Sigma_X^{-1}$; let $\lambda \in \mathbb{C}$ be an eigenvalue of $\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\Sigma_X^{-1}$.

- **If $\lambda \neq 0$, then $\lambda \in (0,1)$**
  Suppose that $\lambda \neq 0$. Then, there exists a nonzero $v \in \mathbb{R}^{n_2}$ such that

$$\left(\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\Sigma_X^{-1}\right) v = \lambda \cdot v.$$

Then,

$$\left(\Sigma_X^{-\frac{1}{2}}\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\Sigma_X^{-\frac{1}{2}\prime}\right)\left(\Sigma_X^{-\frac{1}{2}}v\right) = \lambda\cdot\left(\Sigma_X^{-\frac{1}{2}}v\right),$$

where $\Sigma_X^{-\frac{1}{2}}v \neq \mathbf{0}$ because otherwise, $\left(\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\Sigma_X^{-1}\right)v = \mathbf{0} = \lambda\cdot v$, a contradiction. Therefore, λ is also an eigenvalue of the positive semidefinite matrix $\Sigma_X^{-\frac{1}{2}}\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\Sigma_X^{-\frac{1}{2}}$ and thus a real non-negative value.
By the same line of logic, $\lambda > 0$ is also an eigenvalue of $\Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}\Sigma_Y^{-1}$.

We can also see that $\lambda \leq 1$. Note first that the determinant of Σ is

$$|\Sigma| = \left|\begin{pmatrix}\Sigma_Y & \Sigma_{YX}\\ \Sigma_{XY} & \Sigma_X\end{pmatrix}\right| = |\Sigma_X|\cdot\left|\Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\right| \neq 0,$$

and because $|\Sigma_X| \neq 0$, it follows that $\Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}$ is a non-singular matrix. It is also positive definite, since it is the inverse of the $(1,1)$ block in Σ^{-1}, which must be positive definite by the positive definiteness of Σ^{-1}. λ now satisfies

$$\begin{aligned}\left|\lambda\cdot I_{n_2} - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\Sigma_X^{-1}\right| &= \left|\Sigma_X^{-1}\right|\cdot\left|\lambda\cdot\Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\right|\\ &= \left|\Sigma_X^{-1}\right|\cdot\left|(\lambda-1)\cdot\Sigma_X + \left(\Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\right)\right|\\ &= \left|\Sigma_X^{-1}\right||\Sigma_X|\cdot\left|(\lambda-1)\cdot I_{n_2} + \Sigma_X^{-\frac{1}{2}}\left(\Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\right)\Sigma_X^{-\frac{1}{2}\prime}\right|\\ &= \left|(\lambda-1)\cdot I_{n_2} + \Sigma_X^{-\frac{1}{2}}\left(\Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\right)\Sigma_X^{-\frac{1}{2}\prime}\right| = 0,\end{aligned}$$

so that $\lambda-1$ is an eigenvalue of the negative definite matrix $-\Sigma_X^{-\frac{1}{2}}\left(\Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\right)\Sigma_X^{-\frac{1}{2}\prime}$.
It follows that $\lambda - 1 < 0$, or that $\lambda < 1$.
We have thus seen that $\lambda \in (0,1)$.

- **$\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\Sigma_X^{-1}$ and $\Sigma_X^{-\frac{1}{2}}\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\Sigma_X^{-\frac{1}{2}\prime}$ have the same number of eigenvalues equal to 0**

  λ is a non-zero eigenvalue of $\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\Sigma_X^{-1}$ if and only if it is also a non-zero eigenvalue of $\Sigma_X^{-\frac{1}{2}}\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\Sigma_X^{-\frac{1}{2}\prime}$. Since the two matrices are of the same dimensions, it follows that they have the same number of eigenvalues equal to 0.

By implication, the matrices $\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\Sigma_X^{-1}$ and $\Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}\Sigma_Y^{-1}$ share the exact same set of eigenvalues, of which the non-zero ones lie in the interval $(0,1)$.
If $0 \leq r \leq n_2$ eigenvalues of $\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\Sigma_X^{-1}$ are equal to 0, then this means the rank of Σ_{XY} is $n_2 - r \geq 0$, and as such, the matrix $\Sigma_Y^{-\frac{1}{2}}\Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}\Sigma_Y^{-\frac{1}{2}\prime}$ has exactly $n_1 - (n_2 - r)$ eigenvalues equal to 0. Since the non-zero eigenvalues of $\Sigma_Y^{-\frac{1}{2}}\Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}\Sigma_Y^{-\frac{1}{2}\prime}$ and $\Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}\Sigma_Y^{-1}$ are

the same, this means that $\Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}\Sigma_Y^{-1}$ also has $n_1-(n_2-r)$ eigenvalues equal to 0.

In conclusion,

$$eig_{n_2}\left(\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\Sigma_X^{-1}\right) = eig_{n_2}\left(\Sigma_X^{-\frac{1}{2}}\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\Sigma_X^{-\frac{1}{2}\prime}\right) \in [0,1]^{n_2}$$

and

$$\begin{aligned} eig_{n_1}\left(\Sigma_Y^{-\frac{1}{2}}\Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}\Sigma_Y^{-\frac{1}{2}\prime}\right) &= eig_{n_1}\left(\Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}\Sigma_Y^{-1}\right) \\ &= \left(eig_{n_2}\left(\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\Sigma_X^{-1}\right)', \underbrace{0,\cdots,0}_{n_1-n_2}\right)' \in [0,1]^{n_1}. \end{aligned}$$

Now let us solve the maximization problem

$$\begin{aligned} \max_{H\in\mathbb{R}^{n_1\times k}, R\in\mathbb{R}^{n_2\times k}} \quad & \sum_{i=1}^{k}\left(h_i'\Sigma_{YX}r_i\right)^2 \\ \text{subject to} \quad & h_i'\Sigma_Y h_i = r_i'\Sigma_X r_i = 1 \text{ for any } 1\le i\le k, \end{aligned}$$

to find the weights $h_1,\cdots,h_k\in\mathbb{R}^{n_1\times 1}$ and $r_1,\cdots,r_k\in\mathbb{R}^{n_2\times 1}$ normalized to $h_i'\Sigma_Y h_i = r_i'\Sigma_X r_i = 1$ such that $\sum_{i=1}^{k}\left(h_i'\Sigma_{YX}r_i\right)^2$ is maximized.
The Lagrangian to this problem is defined as

$$\mathcal{L}(H,R,\lambda,\mu) = \sum_{i=1}^{k}\left(h_i'\Sigma_{YX}r_i\right)^2 + \sum_{i=1}^{k}\left[\lambda_i\left(1-h_i'\Sigma_Y h_i\right)+\mu_i\left(1-r_i'\Sigma_X r_i\right)\right].$$

Suppose $H\in\mathbb{R}^{n_1\times k}$ and $R\in\mathbb{R}^{n_2\times k}$ form a solution to the problem. The first order condition for maximization tells us that

$$\begin{aligned} \frac{\partial\mathcal{L}}{\partial h_i} &= 2\left(h_i'\Sigma_{YX}r_i\right)\cdot\Sigma_{YX}r_i - 2\lambda_i\cdot\Sigma_Y h_i = 0, \\ \frac{\partial\mathcal{L}}{\partial r_i} &= 2\left(h_i'\Sigma_{YX}r_i\right)\cdot\Sigma_{XY}h_i - 2\mu_i\cdot\Sigma_X r_i = 0, \\ h_i'\Sigma_Y h_i &= r_i'\Sigma_X r_i = 1 \end{aligned}$$

for any $1\le i\le k$. Rearranging the f.o.c.s above yields

$$\begin{aligned} \left(h_i'\Sigma_{YX}r_i\right)\cdot\Sigma_{YX}r_i &= \lambda_i\cdot\Sigma_Y h_i \\ \left(h_i'\Sigma_{YX}r_i\right)\cdot\Sigma_{XY}h_i &= \mu_i\cdot\Sigma_X r_i, \end{aligned}$$

so premultiplying each equation by h_i and r_i gives us

$$\left(h_i'\Sigma_{YX}r_i\right)^2 = \lambda_i \quad \text{and} \quad \left(h_i'\Sigma_{YX}r_i\right)^2 = \mu_i,$$

implying that $\lambda_i = \mu_i \geq 0$ for any $1 \leq i \leq k$.
It follows that

$$\lambda_i \cdot \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}r_i = \lambda_i \left(h_i'\Sigma_{YX}r_i\right) \cdot \Sigma_{XY}h_i = \lambda_i^2\Sigma_X r_i$$
$$\mu_i \cdot \Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}h_i = \mu_i \left(h_i'\Sigma_{YX}r_i\right) \cdot \Sigma_{YX}r_i = \mu_i^2\Sigma_Y h_i,$$

so we have

$$\lambda_i \cdot \left(\lambda_i\Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\right) r_i = \mathbf{0}$$
$$\mu_i \cdot \left(\mu_i\Sigma_Y - \Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}\right) h_i = \mathbf{0}.$$

We now investigate two distinct cases:

- $\lambda_i = \mu_i \neq 0$
  Suppose $\lambda_i = \mu_i \neq 0$. Then, the above reduece s to

$$\left(\lambda_i\Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\right) r_i = \mathbf{0}$$
$$\left(\mu_i\Sigma_Y - \Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}\right) h_i = \mathbf{0},$$

  for non-zero vectors r_i and h_i, so that

$$\left|\lambda_i\Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\right| = \left|\lambda_i \cdot I_{n_2} - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\Sigma_X^{-1}\right| \cdot |\Sigma_X| = 0$$
$$\left|\mu_i\Sigma_Y - \Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}\right| = \left|\mu_i \cdot I_{n_1} - \Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}\Sigma_Y^{-1}\right| \cdot |\Sigma_Y| = 0$$

  and therefore $\mu_i = \lambda_i$ is an eigenvalue of the matrices $\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\Sigma_X^{-1}$ and $\Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}\Sigma_Y^{-1}$.
  By the results shown above, $\lambda_i = \mu_i \in (0, 1)$.
  Since

$$\left(\lambda_i I_{n_2} - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\Sigma_X^{-1}\right)(\Sigma_X r_i) = \mathbf{0},$$

  $\Sigma_X r_i \in \mathbb{R}^{n_2 \times 1}$ is an eigenvector of $\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\Sigma_X^{-1}$ corresponding to λ_i, and likewise, $\Sigma_Y h_i \in \mathbb{R}^{n_1 \times 1}$ is an eigenvector of $\Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}\Sigma_Y^{-1}$ corresponding to μ_i.

  In summation, if $\lambda_i = \mu_i > 0$, then $\lambda_i = \mu_i \in (0, 1)$ and r_i, h_i are Σ_X^{-1}, Σ_Y^{-1} times eigenvectors of

$$\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\Sigma_X^{-1} \quad \text{and} \quad \Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}\Sigma_Y^{-1}$$

  corresponding to the eigenvalue $\lambda_i = \mu_i$.
  Equivalently, we are able to see that $\Sigma_X^{\frac{1}{2}\prime} r_i$, $\Sigma_Y^{\frac{1}{2}\prime} h_i$ are orthonormal eigenvectors of the symmetric matrices

$$\Sigma_X^{-\frac{1}{2}}\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\Sigma_X^{-\frac{1}{2}\prime} \quad \text{and} \quad \Sigma_Y^{-\frac{1}{2}}\Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}\Sigma_Y^{-\frac{1}{2}\prime}$$

corresponding to the eigenvalue $\lambda_i = \mu_i$.

- $\lambda_i = \mu_i = 0$
  Suppose now that $\lambda_i = \mu_i = 0$. It follows that the k largest eigenvalues of $\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\Sigma_X^{-1}$ contain 0; otherwise, we can increase the value of the objective function $\sum_{i=1}^{k}\lambda_i$ by choosing $\Sigma_X r_i$ to be an eigenvalue of some non-negative eigenvalue of $\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\Sigma_X^{-1}$. This in turn implies that $\Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}\Sigma_Y^{-1}$ has the $n_1 - n_2$ more eigenvalues equal to 0 than $\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\Sigma_X^{-1}$, and that its k largest eigenvalues also contain 0.

  When $\lambda_i = \mu_i = 0$, then the first order conditions always hold, and we are able to choose r_i, h_i as any vectors that satisfy $r_i'\Sigma_X r_i = h_i'\Sigma_Y h_i = 1$. Therefore, to maintain consistency with the case above, we choose r_i and h_i as vectors satisfying

$$\left(\lambda_i\Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\right) r_i = \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX} r_i = \mathbf{0}$$
$$\left(\mu_i\Sigma_Y - \Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}\right) h_i = \Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY} h_i = \mathbf{0},$$

  where the existence of r_i and h_i are guranteed by the observation above.
  This means that $\Sigma_X r_i$, $\Sigma_Y h_i$ are again orthonormal eigenvectors of

$$\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\Sigma_X^{-1} \quad \text{and} \quad \Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}\Sigma_Y^{-1}$$

  corresponding to the 0 eigenvalue of the above matrices.

We have thus seen that r_i, h_i are chosen so that $\Sigma_X r_i$, $\Sigma_Y h_i$ are eigenvectors of $\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\Sigma_X^{-1}$ and $\Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}\Sigma_Y^{-1}$ corresponding to the eigenvalue λ_i for any $1 \le i \le k$, where

$$\lambda_i = (h_i'\Sigma_{YX} r_i)^2$$

for any $1 \le i \le k$. Therefore, $h_1, \cdots, h_k$ and $r_1, \cdots, r_k$ are solutions to the maximization problem only if $\lambda_1 \ge \cdots \ge \lambda_k$ are the k largest ordered eigenvalues of $\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\Sigma_X^{-1}$ (or $\Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}\Sigma_Y^{-1}$).

Note that $h_1, \cdots, h_k$ and $r_1, \cdots, r_k$ can be chosen so that

$$\{\Sigma_Y^{\frac{1}{2}\prime} h_1, \cdots, \Sigma_Y^{\frac{1}{2}\prime} h_k\} \quad \text{and} \quad \{\Sigma_X^{\frac{1}{2}\prime} r_1, \cdots, \Sigma_X^{\frac{1}{2}\prime} r_k\}$$

are orthonormal sets of eigenvectors corresponding to $\lambda_1 \ge \cdots \ge \lambda_k$ for the symmetric positive semidefinite matrices

$$\Sigma_Y^{-\frac{1}{2}}\Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}\Sigma_Y^{-\frac{1}{2}\prime} \quad \text{and} \quad \Sigma_X^{-\frac{1}{2}}\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\Sigma_X^{-\frac{1}{2}\prime}.$$

It follows that

$$H'\Sigma_Y H = \begin{pmatrix} h_1'\Sigma_X^{\frac{1}{2}} \\ \vdots \\ h_k'\Sigma_X^{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} \Sigma_X^{\frac{1}{2}\prime} h_1 & \cdots & \Sigma_X^{\frac{1}{2}\prime} h_k \end{pmatrix} = I_k,$$

and likewise, $R'\Sigma_X R = I_k$.

Finally, for any $1 \leq i \leq k$, we denote

$$h_i'\Sigma_{YX} r_i = \rho_i = \sqrt{\lambda_i} \text{ or } -\sqrt{\lambda_i},$$

which represents the ith largest possible correlation between linear combinations of X and Y. For any $1 \leq i \neq j \leq k$ such that $\lambda_i > 0$, since

$$\rho_i \cdot \Sigma_{YX} r_i = \lambda_i \cdot \Sigma_Y h_i$$

by the first order conditions, premultiplying both sides by h_j' yields

$$\rho_i \cdot h_j'\Sigma_{YX} r_i = \lambda_i \cdot h_j'\Sigma_Y h_i = 0,$$

where the last equality follows because $H'\Sigma_Y H = I_k$. Since $\lambda_i > 0$, we have

$$h_j'\Sigma_{YX} r_i = 0.$$

Meanwhile, if $\lambda_i = 0$, then note that r_i is chosen so that

$$\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX} r_i = \mathbf{0};$$

premultiplying the above equation by r_i' yields

$$(\Sigma_{YX} r_i)'\Sigma_Y^{-1}(\Sigma_{YX} r_i) = 0,$$

and since Σ_Y^{-1} is positive definite, $\Sigma_{YX} r_i = \mathbf{0}$ and

$$h_j'\Sigma_{YX} r_i = 0.$$

The above analysis applies to the case where $n_2 \geq n_1$ as well, since we have restricted $k \leq \min(n_1, n_2)$.

In summary, for any $k \le \min(n_1, n_2)$, if we choose $H = (h_1, \cdots, h_k) \in \mathbb{R}^{n_1 \times k}$ and $R = (r_1, \cdots, r_k) \in \mathbb{R}^{n_2 \times k}$ such that:

- $1 \ge \lambda_1 \ge \cdots \lambda_k \ge 0$ are the k largest solutions to the equations

$$\left|\lambda \cdot \Sigma_Y - \Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}\right| = 0,$$

  or equivalently,

$$\left|\lambda \cdot \Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\right| = 0,$$

- $\{\Sigma_Y^{\frac{1}{2}\prime} h_1, \cdots, \Sigma_Y^{\frac{1}{2}\prime} h_k\}$ is a set of orthonormal eigenvectors of the matrix

$$\Sigma_Y^{-\frac{1}{2}}\Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}\Sigma_Y^{-\frac{1}{2}\prime}$$

  corresponding to $\lambda_1, \cdots, \lambda_k$

- $\{\Sigma_X^{\frac{1}{2}\prime} r_1, \cdots, \Sigma_X^{\frac{1}{2}\prime} r_k\}$ is a set of orthonormal eigenvectors of the matrix

$$\Sigma_X^{-\frac{1}{2}}\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\Sigma_X^{-\frac{1}{2}\prime}$$

  corresponding to $\lambda_1, \cdots, \lambda_k$,

then we have

$$H'\Sigma_Y H = \begin{pmatrix} h_1'\Sigma_Y h_1 & \cdots & h_1'\Sigma_Y h_k \\ \vdots & \ddots & \vdots \\ h_k'\Sigma_Y h_1 & \cdots & h_k'\Sigma_Y h_k \end{pmatrix} = I_k$$

$$R'\Sigma_X R = \begin{pmatrix} r_1'\Sigma_X r_1 & \cdots & r_1'\Sigma_X r_k \\ \vdots & \ddots & \vdots \\ r_k'\Sigma_X r_1 & \cdots & r_k'\Sigma_X r_k \end{pmatrix} = I_k$$

$$H'\Sigma_{YX} R = \begin{pmatrix} h_1'\Sigma_{YX} r_1 & \cdots & h_1'\Sigma_{YX} r_k \\ \vdots & \ddots & \vdots \\ h_k'\Sigma_{YX} r_1 & \cdots & h_k'\Sigma_{YX} r_k \end{pmatrix} = \begin{pmatrix} \rho_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \rho_k, \end{pmatrix} \quad \text{where } \rho_1^2 = \lambda_1, \cdots, \rho_k^2 = \lambda_k.$$

$\rho_1, \cdots, \rho_k \in [-1, 1]$ can be interpreted as the k largest correlations (in magnitude) achievable between linear combinations of X and Y.

2.1.2 Sample CCA

The problem we studied above was canonical correlation analysis for the population. To conduct the same analysis for the sample, suppose $\mathbb{X} \in \mathbb{R}^{T \times n_2}$ and $\mathbb{Y} \in \mathbb{R}^{T \times n_1}$ are data matrices collecting T sample observations of the random vectors X and Y such that

$$
\begin{aligned}
S_X &= \frac{1}{T}\mathbb{X}'\mathbb{X} \xrightarrow{p} \Sigma_X, \\
S_Y &= \frac{1}{T}\mathbb{Y}'\mathbb{Y} \xrightarrow{p} \Sigma_Y, \quad \text{and} \\
S_{YX} &= \frac{1}{T}\mathbb{Y}'\mathbb{X} \xrightarrow{p} \Sigma_{YX},
\end{aligned}
$$

as $T \to \infty$, where we define $S_{XY} = S_{YX}'$. Then, the first $k \le \min(n_1, n_2)$ sample canonical correlations can be found as the k largest eigenvalues $1 \ge \hat{\lambda}_1 \ge \cdots \ge \hat{\lambda}_k \ge 0$ that solve the equation

$$\left|\lambda \cdot S_Y - S_{YX} S_X^{-1} S_{XY}\right| = 0,$$

or equivalently

$$\left|\lambda \cdot S_X - S_{XY} S_Y^{-1} S_{YX}\right| = 0.$$

The sample weights $\hat{H}$ and $\hat{R}$ that take values in $\mathbb{R}^{n_1 \times k}$ and $\mathbb{R}^{n_2 \times k}$ are found as $S_Y^{-\frac{1}{2}\prime}$ and $S_X^{-\frac{1}{2}\prime}$ times a set of orthonormal eigenvectors of

$$S_Y^{-\frac{1}{2}} S_{YX} S_X^{-1} S_{XY} S_Y^{-\frac{1}{2}\prime}$$

and

$$S_X^{-\frac{1}{2}} S_{XY} S_Y^{-1} S_{YX} S_X^{-\frac{1}{2}\prime}$$

corresponding to $\hat{\lambda}_1 \ge \cdots \ge \hat{\lambda}_k$.

Then, we have

$$
\begin{aligned}
\hat{H}' S_Y \hat{H} &= I_k \\
\hat{R}' S_X \hat{R} &= I_k \\
\hat{H}' S_{YX} \hat{R} &= \begin{pmatrix} \hat{\rho}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \hat{\rho}_k \end{pmatrix}, \quad \text{where } \hat{\rho}_1^2 = \hat{\lambda}_1, \cdots, \hat{\rho}_k^2 = \hat{\lambda}_k,
\end{aligned}
$$

by much the same process as above.

By the continuity of ordered eigenvalues and the probability limits stated above, we can see that

$$\hat{\lambda}_i \xrightarrow{p} \lambda_i$$

for $1 \leq i \leq k$, so that the squared sample canonical correlations are consistent for their population counterparts.

2.2 Assumptions and Preliminaries

We retain much of the same assumptions as in the unilevel factor model studied above, tailored for the presence of global and country-specific factors.
As usual, the superscript 0 indicates true values. This time, the assumptions will be given in blocks:

I) **Block 1: Relationship between Global and Country Factors**

(1) **Relationship between Global and Country-Specific Factors**
The processes $\{G_t^0\}_{t\in\mathbb{Z}}, \{F_{1t}^0\}_{t\in\mathbb{Z}}, \cdots, \{F_{Mt}^0\}_{t\in\mathbb{Z}}$ are independent processes.

(2) **Time Series Properties of Global and Country-Specific Factors**
We assume that $\{G_t^0\}_{t\in\mathbb{Z}}, \{F_{1t}^0\}_{t\in\mathbb{Z}}, \cdots, \{F_{Mt}^0\}_{t\in\mathbb{Z}}$ are weakly stationary mean zero processes such that the covariance matrix of $K_{mt}^0 = (G_t^{0\prime}, F_{mt}^{0\prime})'$ is given as

$$\Sigma_m = \begin{pmatrix} \Sigma_G & O \\ O & \Sigma_{F,m} \end{pmatrix},$$

where both Σ_G and $\Sigma_{F,m}$ are positive definite.

(3) **Rate of Convergence of Cross Products**
For any two countries m, n, we assume that

$$\frac{1}{T}\sum_{t=1}^{T} K_{mt}^0 K_{nt}^{0\prime} - \mathbb{E}\left[K_{mt}^0 K_{nt}^{0\prime}\right] = O_p\left(\frac{1}{\sqrt{T}}\right).$$

(4) **Magnitude of Country-Specific Cross Sectional Observations**
We assume that $N = N_1 + \cdots + N_M$ and each N_m is of the same magnitude ($\frac{N}{N_m} = O_p(1)$ for each m). In other words, the number of cross-sectional observations from one country does not dominate those of other countries.

II) **Block 2: Asymptotics of the Factor Estimators of Each Country**

(1) **Non-Triviality of Global Factors**

For any country m, we assume that the $r+r_m$ largest eigenvalues of $X_m X_m'$ are always positive. This implies that the $r+r_m$ largest eigenvalues of $X_m' X_m$ are the same as those of $X_m X_m'$ and thus also positive. We collect the $r+r_m$ largest eigenvalues of $\frac{1}{N_m T} X_m X_m'$ in $V_{N_m,T}$.

(2) **Second Moment Convergence of True Factors and Factor Loadings**

Define

$$\theta_{mi}^{0\prime} = \begin{pmatrix} \gamma_{mi}^{0\prime} & \lambda_{mi}^{0\prime} \end{pmatrix}$$

for any $1 \leq m \leq M$ and $1 \leq i \leq N_m$, which is the ith row of Θ_m^0.

We assume that, for any $1 \leq m \leq M$, the factor loadings $\theta_{m1}^0, \cdots, \theta_{m,N_m}^0$ are non-random, and that there exists a constant $K > 0$ such that

$$\sup_{t \in N_+} \mathbb{E}\left|K_{mt}^0\right|^2 \leq K$$

$$\sup_{i \in N_+} \left|\theta_{mi}^0\right|^2 \leq K.$$

In addition, for any country m, we assume that

$$\frac{\Theta_m^{0\prime}\Theta_m^0}{N_m} \xrightarrow{p} \Sigma_{\Theta,m} \quad \text{and} \quad \frac{K_m^{0\prime}K_m^0}{T} \xrightarrow{p} \Sigma_{K,m},$$

where $\Sigma_{\Theta,m}, \Sigma_{K,m} \in \mathbb{R}^{(r+r_m)\times(r+r_m)}$ are positive definite matrices.

(3) **Exact Factor Model**

We assume that the processes $\{e_{mit}\}_{t\in\mathbb{Z}}$ are independent and identically distributed across m and i.

(4) **Stationarity of Errors**

For any m and i, we assume that the process $\{e_{mit}\}_{t\in\mathbb{Z}}$ is weakly stationary with mean 0 and autocovariance function $\gamma : \mathbb{Z} \to \mathbb{R}$.

In addition, we assume absolutely summable autocovariances ($\sum_{z=-\infty}^{\infty} |\gamma(z)| < +\infty$), and that the process has bounded fourth moments, that is, $\sup_{t\in\mathbb{Z}} \mathbb{E}\left[e_{mit}^4\right] \leq \mu_4 < +\infty$.

(5) **Weak Dependence between Factors and Errors**

For any $1 \leq m \leq M$, we assume that there exists a constant $K > 0$ such that

$$\mathbb{E}\left[\frac{1}{N_m T}\sum_{i=1}^{N_m}\left|\sum_{t=1}^{T} K_{mt}^{0} e_{mit}\right|^2\right] \leq K$$

$$\mathbb{E}\left|\frac{1}{\sqrt{N_m T}}\sum_{s=1}^{T}\sum_{i=1}^{N_m} K_{mt}^{0}\left(e_{mit}e_{mis} - \gamma(t-s)\right)\right|^2 \leq K \qquad \text{(for any } t \in N_+\text{)}$$

$$\mathbb{E}\left\|\frac{1}{\sqrt{N_m T}}\sum_{t=1}^{T}\sum_{i=1}^{N_m} K_{mit}^{0}\theta_{mi}^{0\prime} e_{mit}\right\|^2 \leq K$$

for any $N_m, T \in N_+$.

(6) **CLT for Time Dimension**

For any $1 \leq m \leq M$ and $i \in N_+$,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} K_{mt}^{0} e_{mit} \xrightarrow{d} N\left[\mathbf{0}, \Phi_{mi}\right]$$

where

$$\Phi_{mi} = \operatorname*{plim}_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} e_{mit}^2 K_{mt}^{0} K_{mt}^{0\prime}.$$

(7) **Sufficient Conditions for Factor Identification**

We assume that the $r + r_m$ eigenvalues collected in $V_{N_m,T}$, as well as those of $\Sigma_{\Theta,m}\Sigma_{K,m}$, are distinct.

(8) **The Probability Limit of $\frac{K_m^{0\prime}\tilde{K}_m}{T}$**

For any $1 \leq m \leq M$ and the estimator $\tilde{K}_m = \left(\tilde{K}_{m1}', \cdots, \tilde{K}_{mT}'\right)'$ of the factors K_m defined below, there exists a nonsingular $r \times r$ matrix Q_m such that

$$\frac{K_m^{0\prime}\tilde{K}_m}{T} \xrightarrow{p} Q_m.$$

2.3 Step 1: The Initial Estimation of Global Factors

The estimation of the global and country-specific factors proceeds in steps. In the initial step, data on only two countries are used to derive rudimentary estimators of the global factors, and in the subsequent step the country-specific factors are estimated by treating the estimators of the global factors derived in the previous step as if they were the true global factors.
The factors estimated as such rely only on data on two countries, so to make use of the full data, the global factors are re-estimated, this time using the country-specific factor estimators derived in the second step in place of the true country-specific factors. Finally, the country-specific factors are re-estimated using the newly estimated global factors in place of the true global factors.
In this section we focus on the first step of the estimation, which uses canonical correlation analysis to estiamte the global factors. The next three steps all rely on familiar principal components analysis used in Bai and Ng (2002) and Bai (2003).

We now proceed in steps:

2.3.1 Estimation of Factors Affecting Each Country

Recall the concatenated model

$$X_{mt} = \Theta_m^0 \cdot K_{mt}^0 + e_{mt},$$

where

$$\Theta_m^0 = \begin{pmatrix} \Gamma_m^0 & \Lambda_m^0 \end{pmatrix} \quad \text{and} \quad K_{mt}^0 = \begin{pmatrix} G_t^0 \\ F_{mt}^0 \end{pmatrix}.$$

For any country m,

$$X_{mt} = \Theta_m^0 \cdot K_{mt}^0 + e_{mt}$$

for $1 \leq t \leq T$ and N_m cross-sectional observations resembles the usual unilevel factor model studied in previous sections. Therefore, we can estimate the factors K_{mt} and factor loadings Θ_m by minimizing the objective function

$$\frac{1}{N_m T} \sum_{i=1}^{N_m} \sum_{t=1}^{T} \left[X_{mit} - \begin{pmatrix} \gamma_{mi}' & \lambda_{mi}' \end{pmatrix} \begin{pmatrix} G_t \\ F_{mt} \end{pmatrix} \right]^2 = \frac{1}{N_m T} \operatorname{tr}\left(\left(X_m - K_m \Theta_m' \right) \left(X_m - K_m \Theta_m' \right)' \right)$$

with respect to $K_m = (K_{m1}, \cdots, K_{mT})'$ and Θ_m. Denote these estimates by $\bar{\Theta}_m$ and $\tilde{K}_m$.
From what we know of the unilevel factor model, assuming that we concentrate out Θ_m first and then estimate the factors K_m, under the usual factor identification condition $\frac{\tilde{K}_m' \tilde{K}_m}{T} = I_r$,

the estimators $\tilde{K}_m$ and $\tilde{\Theta}_m$ are given by:

$\tilde{K}_m = \sqrt{T} \times$ The orthonormal eigenvectors of $X_m X_m'$ corresponding to its r largest eigenvalues
$\tilde{\Theta}_m = \frac{1}{T} X_m' \tilde{K}_m$.

From the second block of assumptions we made, the result on unilevel factor models that we derived earlier tells us that:

- **Consistency of the Factors**

$$\frac{1}{T}\sum_{t=1}^{T}\left|\tilde{K}_{mt} - \tilde{H}_m' K_{mt}^0\right|^2 = \frac{1}{T}\left\|\tilde{K}_m - K_m^0 \tilde{H}_m\right\|^2 = O_p\left(\frac{1}{\min(N_m, T)}\right),$$

  where

$$\tilde{H}_m = \left(\frac{\Theta_m^{0\prime}\Theta_m^0}{N_m}\right)\left(\frac{K_m^{0\prime}\tilde{K}_m}{T}\right) V_{N_m,T}^{-1}$$

- **Consistency of $V_{N_m,T}$ and $\frac{K_m^{0\prime}\tilde{K}_m}{T}$**

$$V_{N_m,T} \xrightarrow{p} V_m,$$

  where V_m collects the r eigenvalues of $\Omega_{\Theta,m}\Omega_{K,m}$, which are assumed to be all positive and distinct. By implication,

$$\tilde{H}_m \xrightarrow{p} \Sigma_{\Theta,m} Q_m V_m^{-1} := H_m^0.$$

- **Rate of Convergence of the Factor Estimator**
  For any $t \in T$,

$$\tilde{K}_{mt} - \tilde{H}_m' K_{mt}^0 = O_p\left(\frac{1}{\sqrt{N_m}}\right) + O_p\left(\frac{1}{\min(N_m, T)}\right) = O_p\left(\frac{1}{\min(\sqrt{N_m}, T)}\right).$$

- **Rate of Convergence of the Factor Loading Estimator**
  For any $i \in N_+$,

$$\tilde{\theta}_{mi} - \tilde{H}_m^{-1}\theta_{mi}^0 = O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{\min(N_m, T)}\right) = O_p\left(\frac{1}{\min(N_m, \sqrt{T})}\right).$$

From the consistency result we have

$$\frac{1}{\sqrt{T}}\left(\tilde{K}_m - K_m^0\tilde{H}_m\right) = O_p\left(\frac{1}{\min(\sqrt{N_m},\sqrt{T})}\right).$$

Meanwhile, since

$$\left\|\frac{1}{\sqrt{T}}K_m^0\tilde{H}_m\right\|^2 \leq \left(\frac{1}{T}\left\|K_m^0\right\|^2\right)\left(\left\|\tilde{H}_m\right\|\right) = \operatorname{tr}\left(\frac{K_m^{0\prime}K_m}{T}\right)\cdot\left\|\tilde{H}_m\right\|,$$

where both terms on the right hand side are $O_p(1)$, we can tell that

$$\frac{1}{\sqrt{T}}K_m^0\tilde{H}_m = O_p(1).$$

2.3.2 Canonical Correlation Analysis with Two Countries

Select two countries, with indices $m = 1,2$ for convenience.

Once the above estimation has been completed for $m = 1,2$, we now note that the global factors G_t^0 form the first r elements of each combined factor K_{1t}^0 and K_{2t}^0, while the latter r_1 and r_2 elements, F_{1t}^0 and F_{2t}^0, are assumed to be uncorrelated. As such, the first r canonical correlations of K_{1t}^0 and K_{2t}^0 must be 1, with the weights being the first r standard basis vectors in $\mathbb{R}^{r+r_1}$ and $\mathbb{R}^{r+r_2}$, while the remaning $\min(r_1, r_2)$ canonical correlations must be 0. We now confirm this intuition.

Population CCA

By assumption, the covariance matrix of K_{mt}^0 is

$$\Sigma_m = \begin{pmatrix} \Sigma_G & O \\ O & \Sigma_{F,m} \end{pmatrix},$$

where both Σ_G and $\Sigma_{F,m}$ are positive definite. Moreover,

$$\Sigma_{12} = \mathbb{E}\left[K_{1t}^0 K_{2t}^{0\prime}\right] = \begin{pmatrix} \Sigma_G & O \\ O & O \end{pmatrix} \in \mathbb{R}^{(r+r_1)\times(r+r_2)}.$$

Letting $r_1 = \min(r_1, r_2)$ without loss of generality, the $r + r_1$ squared canonical correlations $\rho_1^2 \geq \cdots \geq \rho_{r+r_1}^2$ of K_{1t}^0 and K_{2t}^0 are the first $r + r_1$ ordered eigenvalues solving the equation

$$\left|\rho_i^2 \cdot \Sigma_1 - \Sigma_{12}\Sigma_2^{-1}\Sigma_{12}'\right| = 0,$$

or equivalently,

$$\left|\rho_i^2 \cdot \Sigma_2 - \Sigma_{12}'\Sigma_1^{-1}\Sigma_{12}\right| = 0,$$

for $1 \leq i \leq r + r_1$.

Note that, for any $1 \leq i \leq r + r_1$,

$$\rho_i^2 \cdot \Sigma_1 - \Sigma_{12}\Sigma_2^{-1}\Sigma_{12}' = \begin{pmatrix} \rho_i^2 \cdot \Sigma_G - \Sigma_G & O \\ O & \rho_i^2 \cdot \Sigma_{F,1} \end{pmatrix},$$

so that

$$\begin{aligned} \left|\rho_i^2 \cdot \Sigma_1 - \Sigma_{12}\Sigma_2^{-1}\Sigma_{12}'\right| &= \left|(\rho_i^2 - 1)\Sigma_G\right| \cdot \left|\rho_i^2 \cdot \Sigma_{F,1}\right| \\ &= (\rho_i^2 - 1)^r (\rho_i^2)^{r_1} |\Sigma_G| \cdot |\Sigma_{F,1}|. \end{aligned}$$

It follows that $\rho_1^2 = \cdots = \rho_r^2 = 1$ and $\rho_{r+1}^2 = \cdots = \rho_{r+r_1}^2 = 0$.

Sample Estimators of the Covariance Matrices

Define

$$\tilde{S}_m = \frac{1}{T}\sum_{t=1}^{T}\tilde{K}_{mt}\tilde{K}'_{mt} = \frac{1}{T}\tilde{K}'_m\tilde{K}_m$$

for $m = 1,2$, and let

$$\tilde{S}_{12} = \frac{1}{T}\sum_{t=1}^{T}\tilde{K}_{1t}\tilde{K}'_{2t} = \frac{1}{T}\tilde{K}'_1\tilde{K}_2.$$

Since

$$\frac{1}{\sqrt{T}}\tilde{K}_m = \frac{1}{\sqrt{T}}K^0_m\tilde{H}_m + O_p\left(\frac{1}{\min(\sqrt{N_m},\sqrt{T})}\right)$$

and

$$\frac{1}{\sqrt{T}}K^0_m\tilde{H}_m = O_p(1),$$

for any $m = 1,2$ we have

$$\begin{aligned}
\tilde{S}_m &= \frac{1}{T}\sum_{t=1}^{T}\tilde{K}_{mt}\tilde{K}'_{mt} \\
&= \left(\frac{1}{\sqrt{T}}\tilde{K}_m\right)'\left(\frac{1}{\sqrt{T}}\tilde{K}_m\right) \\
&= \left(\frac{1}{\sqrt{T}}K^0_m\tilde{H}_m + O_p\left(\frac{1}{\min(\sqrt{N_m},\sqrt{T})}\right)\right)'\left(\frac{1}{\sqrt{T}}K^0_m\tilde{H}_m + O_p\left(\frac{1}{\min(\sqrt{N_m},\sqrt{T})}\right)\right) \\
&= \tilde{H}'_m\left(\frac{K^{0\prime}_m K^0_m}{T}\right)\tilde{H}_m + O_p\left(\frac{1}{\min(\sqrt{N_m},\sqrt{T})}\right).
\end{aligned}$$

Likewise,

$$\tilde{S}_{12} = \tilde{H}'_1\left(\frac{1}{T}K^{0\prime}_1K^0_2\right)\tilde{H}_2 + O_p\left(\frac{1}{\min(\sqrt{N_1},\sqrt{N_2},\sqrt{T})}\right).$$

Defining

$$S_m = \frac{1}{T}\sum_{t=1}^{T}K^0_{mt}K^{0\prime}_{mt} = \frac{1}{T}K^{0\prime}_mK^0_m$$

$$S_{12} = \frac{1}{T}\sum_{t=1}^{T}K^0_{1t}K^{0\prime}_{2t} = \frac{1}{T}K^{0\prime}_1K^0_2$$

for $m = 1, 2$, we can see that

$$\tilde{S}_m = \tilde{H}_m' S_m \tilde{H}_m + O_p\left(\frac{1}{\min(\sqrt{N_m}, \sqrt{T})}\right)$$
$$\tilde{S}_{12} = \tilde{H}_1' S_{12} \tilde{H}_2 + O_p\left(\frac{1}{\min(\sqrt{N_1}, \sqrt{N_2}, \sqrt{T})}\right).$$

By implication,

$$\tilde{S}_m - \tilde{H}_m' S_m \tilde{H}_m \xrightarrow{p} 0$$

and

$$\tilde{H}_m' S_m \tilde{H}_m \xrightarrow{p} H_m^{0\prime} \Sigma_{K,m} H_m^0,$$

where the limit is an $r \times r$ matrix of full rank, we have

$$\tilde{S}_m \xrightarrow{p} H_m^{0\prime} \Sigma_{K,m} H_m^0$$

and therefore

$$\left(\tilde{S}_m\right)^{-1}, \left(\tilde{H}_m' S_m \tilde{H}_m\right)^{-1} \xrightarrow{p} \left(H_m^{0\prime} \Sigma_{K,m} H_m^0\right)^{-1}$$

by the CMT. This tells us that $\left(\tilde{S}_m\right)^{-1}, \left(\tilde{H}_m' S_m \tilde{H}_m\right)^{-1}$ are $O_p(1)$, and as such, the decomposition

$$\left\| \tilde{S}_m^{-1} - \left(\tilde{H}_m' S_m \tilde{H}_m\right)^{-1} \right\| \leq \left\| \tilde{S}_m - \tilde{H}_m' S_m \tilde{H}_m \right\| \cdot \left\| \tilde{S}_m^{-1} \right\| \cdot \left\| \left(\tilde{H}_m' S_m \tilde{H}_m\right)^{-1} \right\|$$

implies that

$$\tilde{S}_m^{-1} - \tilde{H}_m^{-1} S_m^{-1} \tilde{H}_m'^{-1} = O_p\left(\frac{1}{\min(\sqrt{N_m}, \sqrt{T})}\right).$$

The above results imply that

$$\tilde{S}_1^{-1} \tilde{S}_{12} \tilde{S}_2^{-1} \tilde{S}_{12}' = \tilde{H}_1^{-1} \left(S_1^{-1} S_{12} S_2^{-1} S_{12}'\right) \tilde{H}_1 + O_p\left(\frac{1}{\min(\sqrt{N_1}, \sqrt{N_2}, \sqrt{T})}\right).$$

It remains to find the rate of convergence of $\tilde{H}_1^{-1} \left(S_1^{-1} S_{12} S_2^{-1} S_{12}'\right) \tilde{H}_1$.

Because

$$S_m - \Sigma_m = O_p\left(\frac{1}{\sqrt{T}}\right)$$

by assumption, and

$$\left\|S_m^{-1} - \Sigma_m^{-1}\right\| \leq \left\|S_m^{-1}\right\|\left\|\Sigma_m^{-1}\right\| \cdot \|S_m - \Sigma_m\|,$$

where $S_m^{-1} \xrightarrow{p} \Sigma_m^{-1}$ implies that $S_m^{-1} = O_p(1)$, we can see that

$$S_m^{-1} - \Sigma_m^{-1} = O_p\left(\frac{1}{\sqrt{T}}\right).$$

In addition, by assumption

$$S_{12} - \Sigma_{12} = O_p\left(\frac{1}{\sqrt{T}}\right),$$

so that

$$\begin{aligned} &S_1^{-1} S_{12} S_2^{-1} S_{12}' \\ &\quad = \left(\Sigma_1^{-1} + \left(S_1^{-1} - \Sigma_1^{-1}\right)\right)\left(\Sigma_{12} + (S_{12} - \Sigma_{12})\right)\left(\Sigma_2^{-1} + \left(S_2^{-1} - \Sigma_2^{-1}\right)\right)\left(\Sigma_{12} + (S_{12} - \Sigma_{12})\right)' \\ &\qquad = \Sigma_1^{-1}\Sigma_{12}\Sigma_2^{-1}\Sigma_{12}' + O_p\left(\frac{1}{\sqrt{T}}\right). \end{aligned}$$

Therefore,

$$\begin{aligned} \tilde{H}_1^{-1}\left(S_1^{-1} S_{12} S_2^{-1} S_{12}'\right)\tilde{H}_1 &= \tilde{H}_1^{-1}\left(S_1^{-1} S_{12} S_2^{-1} S_{12}' - \Sigma_1^{-1}\Sigma_{12}\Sigma_2^{-1}\Sigma_{12}'\right)\tilde{H}_1 + \tilde{H}_1^{-1}\Sigma_1^{-1}\Sigma_{12}\Sigma_2^{-1}\Sigma_{12}'\tilde{H}_1 \\ &= \tilde{H}_1^{-1}\Sigma_1^{-1}\Sigma_{12}\Sigma_2^{-1}\Sigma_{12}'\tilde{H}_1 + O_p\left(\frac{1}{\sqrt{T}}\right), \end{aligned}$$

which implies that

$$\begin{aligned} \tilde{S}_1^{-1}\tilde{S}_{12}\tilde{S}_2^{-1}\tilde{S}_{12}' - \tilde{H}_1^{-1}\Sigma_1^{-1}\Sigma_{12}\Sigma_2^{-1}\Sigma_{12}'\tilde{H}_1 &= O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{\min(\sqrt{N_1}, \sqrt{N_2}, \sqrt{T})}\right) \\ &= O_p\left(\frac{1}{\min(\sqrt{N_1}, \sqrt{N_2}, \sqrt{T})}\right). \end{aligned}$$

This also tells us that

$$\tilde{S}_1^{-1}\tilde{S}_{12}\tilde{S}_2^{-1}\tilde{S}_{12}' \xrightarrow{p} H_1^{0-1}\Sigma_1^{-1}\Sigma_{12}\Sigma_2^{-1}\Sigma_{12}' H_1^0.$$

Let $\tilde{\mu}$ and $\tilde{v}$ be the first r sample canonical correlation weights of $\tilde{K}_1$ and $\tilde{K}_2$, that is,

$$\tilde{\mu}'\tilde{S}_1\tilde{\mu} = I_r$$
$$\tilde{v}'\tilde{S}_2\tilde{v} = I_r$$
$$\tilde{\mu}'\tilde{S}_{12}\tilde{v} = \begin{pmatrix} \tilde{\rho}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \tilde{\rho}_r, \end{pmatrix} = D_{N_1,N_2,T}$$
$$\left(\tilde{S}_{12}\tilde{S}_2^{-1}\tilde{S}_{12}' - \tilde{\rho}_i^2 \cdot \tilde{S}_1\right)\tilde{\mu}_i = \mathbf{0} \qquad \text{for any } 1 \le i \le r$$
$$\left(\tilde{S}_{12}'\tilde{S}_1^{-1}\tilde{S}_{12} - \tilde{\rho}_i^2 \cdot \tilde{S}_2\right)\tilde{v}_i = \mathbf{0} \qquad \text{for any } 1 \le i \le r$$

where $\tilde{\rho}_1^2 \ge \cdots \ge \tilde{\rho}_r^2$ are the r largest eigenvalues that solve the equation

$$\left|\lambda \cdot \tilde{S}_1 - \tilde{S}_{12}\tilde{S}_2^{-1}\tilde{S}_{12}'\right| = 0,$$

or equivalently,

$$\left|\lambda \cdot \tilde{S}_2 - \tilde{S}_{12}'\tilde{S}_1^{-1}\tilde{S}_{12}\right| = 0.$$

Since

$$\tilde{S}_1^{-1}\tilde{S}_{12}\tilde{S}_2^{-1}\tilde{S}_{12}' \xrightarrow{p} H_1^{0-1}\Sigma_1^{-1}\Sigma_{12}\Sigma_2^{-1}\Sigma_{12}'H_1^0,$$

by the continuity of ordered eigenvalues, $\tilde{\rho}_1^2 \ge \cdots \ge \tilde{\rho}_r^2$ converge in probability to the r largest eigenvalues that solve the equation

$$\left|H_1^{0-1}\Sigma_1^{-1}\Sigma_{12}\Sigma_2^{-1}\Sigma_{12}'H_1^0 - \lambda \cdot I_{r+r_1}\right| = 0.$$

Equivalently,

$$\left|\Sigma_{12}\Sigma_2^{-1}\Sigma_{12}' - \lambda \cdot \Sigma_1\right| = 0.$$

We saw above that the $r + r_1$ solutions to this problem are $\underbrace{1, \cdots, 1}_{r}, \underbrace{0, \cdots, 0}_{r_1}$, so

$$\tilde{\rho}_i^2 \xrightarrow{p} 1$$

for $1 \le i \le r$.

We can also study the rate of convergence of $\tilde{\mu}$, $\tilde{\mu}$ is $O_p(1)$ but not $o_p(1)$. Its $O_p(1)$ nature follows from the fact that

$$\begin{aligned}
\|\tilde{\mu}\| &= \left\|\tilde{\mu}'\tilde{S}_1^{\frac{1}{2}}\tilde{S}_1^{-\frac{1}{2}}\right\| \\
&\leq \left\|\tilde{S}_1^{-\frac{1}{2}}\right\| \cdot \left\|\tilde{\mu}'\tilde{S}_1^{\frac{1}{2}}\right\| \\
&\leq \left\|\tilde{S}_1^{-1}\right\|^{\frac{1}{2}} \left\|\tilde{\mu}'\tilde{S}_1\tilde{\mu}\right\| \\
&= \left\|\tilde{S}_1^{-1}\right\|^{\frac{1}{2}},
\end{aligned}$$

where $\tilde{S}_1 \xrightarrow{p} H_1^{0\prime}\Sigma_1 H_1^0$, implies that $\tilde{S}_1^{-1} = O_p(1)$.
On the other hand, if $\tilde{\mu} = o_p(1)$, then taking limits on both sides of $\tilde{\mu}'\tilde{S}_1\tilde{\mu} = I_r$, we have

$$I_r = O'\Sigma_1 O = O,$$

a contradiction, so $\tilde{\mu}$ must be $O_p(1)$ but not $o_p(1)$.

2.3.3 The Initial Estimator of the Global Factors

Our initial estimator $\hat{G}^{(1)} = \begin{pmatrix} \hat{G}_1^{(1)} & \cdots & \hat{G}_T^{(1)} \end{pmatrix}'$ of the global factors G is defined as

$$\hat{G}^{(1)} = \tilde{K}_1 \tilde{\mu},$$

which is the collection of linear combinations of the columns of $\tilde{K}_1$ with weights assigned by $\tilde{\mu}$. Heuristically, the columns of $\tilde{K}_1\tilde{\mu}$ are the r linear combinations of the columns of $\tilde{K}_1$ that yield the highest correlation with similar combinations of the columns of $\tilde{K}_2$, that is, our estimator of G isolates the parts of $\tilde{K}_1$ that are most highly correlated with $\tilde{K}_2$. Since the columns of $\tilde{K}_1$ represent the estimated time series of each factor in $K_{1t}^0 = (G_t^{0\prime}, F_{1t}^{0\prime})'$, which shares the global factor G_t^0 with K_{2t}^0, out choice of $\hat{G}^{(1)}$ as the estimator of G means that we identify the global factor as that the part of K_{1t}^0 and K_{2t}^0 that is correlated. It follows then that the country-specific factors in K_{1t}^0 and K_{2t}^0 are identified as the parts of K_{1t}^0 and K_{2t}^0 that are uncorrelated with one another.

Consistency of Mean Squared Global Factor Estimates

To show that these global factor estimators are consistent, note that

$$\left(\tilde{S}_1^{-1}\tilde{S}_{12}\tilde{S}_2^{-1}\tilde{S}_{12}'\right)\tilde{\mu} = \tilde{\mu} D_{N_1,N_2,T},$$

so that

$$\hat{G}^{(1)} D_{N_1,N_2,T}, = \tilde{K}_1 \tilde{\mu} D_{N_1,N_2,T}, = \tilde{K}_1 \left(\tilde{S}_1^{-1}\tilde{S}_{12}\tilde{S}_2^{-1}\tilde{S}_{12}'\right)\tilde{\mu}$$

Since

$$\tilde{S}_1^{-1}\tilde{S}_{12}\tilde{S}_2^{-1}\tilde{S}_{12}' - \tilde{H}_1^{-1}\Sigma_1^{-1}\Sigma_{12}\Sigma_2^{-1}\Sigma_{12}'\tilde{H}_1 = O_p\left(\frac{1}{\min(\sqrt{T},\sqrt{N_1},\sqrt{N_2})}\right),$$

we can write

$$\begin{aligned}\frac{1}{\sqrt{T}}\hat{G}^{(1)} D_{N_1,N_2,T}, &= \left(\frac{1}{\sqrt{T}}\tilde{K}_1\right)\left(\tilde{S}_1^{-1}\tilde{S}_{12}\tilde{S}_2^{-1}\tilde{S}_{12}'\right)\tilde{\mu} \\ &= \left(\frac{1}{\sqrt{T}}\tilde{K}_1\right)\left(\tilde{H}_1^{-1}\Sigma_1^{-1}\Sigma_{12}\Sigma_2^{-1}\Sigma_{12}'\tilde{H}_1\right)\tilde{\mu} + O_p\left(\frac{1}{\min(\sqrt{T},\sqrt{N_1},\sqrt{N_2})}\right),\end{aligned}$$

owing to the fact that $\frac{1}{\sqrt{T}}\tilde{K}_1$ is $O_p(1)$.

Note that

$$\frac{1}{\sqrt{T}}\tilde{K}_1 = \frac{1}{\sqrt{T}}K_1^0\tilde{H}_1 + O_p\left(\frac{1}{\min(\sqrt{N_1},\sqrt{T})}\right)$$

and that $\tilde{H}_1$, $\tilde{\mu}$ are $O_p(1)$. We then have

$$\frac{1}{\sqrt{T}}\hat{G}^{(1)}D_{N_1,N_2,T}, = \left(\frac{1}{\sqrt{T}}\tilde{K}_1\right)\left(\tilde{H}_1^{-1}\Sigma_1^{-1}\Sigma_{12}\Sigma_2^{-1}\Sigma_{12}'\tilde{H}_1\right)\tilde{\mu}+O_p\left(\frac{1}{\min(\sqrt{T},\sqrt{N_1},\sqrt{N_2})}\right)$$
$$= \frac{1}{\sqrt{T}}K_1^0\left(\Sigma_1^{-1}\Sigma_{12}\Sigma_2^{-1}\Sigma_{12}'\right)\tilde{H}_1\tilde{\mu}+O_p\left(\frac{1}{\min(\sqrt{T},\sqrt{N_1},\sqrt{N_2})}\right).$$

We know that

$$D_{N_1,N_2,T}\xrightarrow{p} I_r$$

by the convergence of the canonical correlations. Defining

$$\tilde{Q} = \left(\Sigma_1^{-1}\Sigma_{12}\Sigma_2^{-1}\Sigma_{12}'\right)\tilde{H}_1\tilde{\mu}\cdot D_{N_1,N_2,T}^{-1},$$

which is a random matrix taking values in $\mathbb{R}^{(r+r_1)\times r}$, it follows that

$$\frac{1}{\sqrt{T}}\hat{G}^{(1)} = \frac{1}{\sqrt{T}}K_1^0\tilde{Q}+O_p\left(\frac{1}{\min(\sqrt{T},\sqrt{N_1},\sqrt{N_2})}\right)$$

because $D_{N_1,N_2,T} = O_p(1)$. Therefore,

$$\frac{1}{\sqrt{T}}\left\|\hat{G}^{(1)}-K_1^0\tilde{Q}\right\| = O_p\left(\frac{1}{\min(\sqrt{T},\sqrt{N_1},\sqrt{N_2})}\right),$$

and we have

$$\frac{1}{T}\left\|\hat{G}^{(1)}-K_1^0\tilde{Q}\right\|^2 = O_p\left(\frac{1}{\min(N_1,N_2,T)}\right),$$

which is the familiar consistency result.

Note now that

$$\Sigma_1^{-1}\Sigma_{12}\Sigma_2^{-1}\Sigma_{12}' = \begin{pmatrix} \Sigma_G^{-1} & O \\ O & \Sigma_{F,1}^{-1} \end{pmatrix} \begin{pmatrix} \Sigma_G & O \\ O & O \end{pmatrix} = \begin{pmatrix} I_r & O \\ O & O \end{pmatrix}.$$

Therefore,

$$\tilde{Q} = \cdot \begin{pmatrix} I_r & O \\ O & O \end{pmatrix} \cdot \tilde{H}_1\tilde{\mu} \cdot D_{N_1,N_2,T}^{-1} = \begin{pmatrix} \tilde{Q}^{(1)} \\ O \end{pmatrix},$$

where $\tilde{Q}^{(1)}$ collects the first r rows of $\tilde{Q}$, and is defined as

$$\tilde{Q}^{(1)} = \left(\frac{\Gamma_1^{0\prime}\Theta_1^0}{N_1}\right)\left(\frac{K_1^{0\prime}\tilde{K}_1}{T}\right) V_{N_1,T}^{-1}\tilde{\mu} D_{N_1,N_2,T}^{-1},$$

which tells us that $\tilde{Q}^{(1)} = O_p(1)$.

Since K_1^0 can be decomposed as,

$$K_1^0 = \begin{pmatrix} G^0 & F_1^0 \end{pmatrix},$$

we have

$$\frac{1}{T}\left\|\hat{G}^{(1)} - G^0\tilde{Q}^{(1)}\right\|^2 = O_p\left(\frac{1}{\min(N_1, N_2, T)}\right),$$

or equivalently,

$$\frac{1}{T}\sum_{t=1}^{T}\left|\hat{G}_t^{(1)} - \tilde{Q}^{(1)\prime}G_t^0\right|^2 = O_p\left(\frac{1}{\min(N_1, N_2, T)}\right).$$

This result tells us that the initial global factor estimator is consistent for a rotation of the true global factors, and that the global factors are thus identified.

Consistency of Individual Global Factor Estimates

So far we have seen that the mean square of the factor estimates is consistent for some rotation of the global factors. We can now show a stronger result that says the individual factor estimates are consistent for some rotation of the corresponding global factor. The derivation follows similarly to the above.

For any $t \in N_+$, note as before that

$$\hat{G}_t^{(1)} = D_{N_1,N_2,T}^{-1} \cdot \tilde{\mu}' \left(\tilde{S}_1^{-1}\tilde{S}_{12}\tilde{S}_2^{-1}\tilde{S}_{12}'\right)' \tilde{K}_{1t}.$$

We already know that

$$\tilde{K}_{1t} - \tilde{H}_1' K_{1t}^0 = O_p\left(\frac{1}{\min(\sqrt{N_1}, T)}\right),$$

and because this implies that $\tilde{K}_{1t} \xrightarrow{p} H_1^{0\prime} K_{1t}^0$ as $N_1, T \to \infty$, $\tilde{K}_{1t} = O_p(1)$.
Therefore,

$$\begin{aligned}
\hat{G}_t^{(1)} &= D_{N_1,N_2,T}^{-1} \cdot \tilde{\mu}' \left(\tilde{S}_1^{-1}\tilde{S}_{12}\tilde{S}_2^{-1}\tilde{S}_{12}'\right)' \tilde{K}_{1t} \\
&= D_{N_1,N_2,T}^{-1} \cdot \tilde{\mu}' \left(\tilde{H}_1^{-1}\Sigma_1^{-1}\Sigma_{12}\Sigma_2^{-1}\Sigma_{12}'\tilde{H}_1\right)' \tilde{K}_{1t} + O_p\left(\frac{1}{\min(\sqrt{T}, \sqrt{N_1}, \sqrt{N_2})}\right) \\
&= D_{N_1,N_2,T}^{-1} \cdot \tilde{\mu}' \left(\tilde{H}_1^{-1}\Sigma_1^{-1}\Sigma_{12}\Sigma_2^{-1}\Sigma_{12}'\tilde{H}_1\right)' \tilde{H}_1' K_{1t}^0 + O_p\left(\frac{1}{\min(\sqrt{N_1}, T)}\right) + O_p\left(\frac{1}{\min(\sqrt{T}, \sqrt{N_1}, \sqrt{N_2})}\right) \\
&= D_{N_1,N_2,T}^{-1} \cdot \tilde{\mu}' \tilde{H}_1' \left(\Sigma_1^{-1}\Sigma_{12}\Sigma_2^{-1}\Sigma_{12}'\right)' K_{1t}^0 + O_p\left(\frac{1}{\min(\sqrt{T}, \sqrt{N_1}, \sqrt{N_2})}\right) \\
&= \tilde{Q}' \cdot K_{1t}^0 + O_p\left(\frac{1}{\min(\sqrt{T}, \sqrt{N_1}, \sqrt{N_2})}\right).
\end{aligned}$$

Since

$$K_{1t}^0 = \begin{pmatrix} G_t^0 \\ F_{1t}^0 \end{pmatrix} \quad \text{and} \quad \tilde{Q} = \begin{pmatrix} \tilde{Q}^{(1)} \\ O \end{pmatrix},$$

we have $\tilde{Q}' K_{1t}^0 = \tilde{Q}^{(1)\prime} G_t^0$ and

$$\hat{G}_t^{(1)} - \tilde{Q}^{(1)\prime} G_t^0 = O_p\left(\frac{1}{\min(\sqrt{T}, \sqrt{N_1}, \sqrt{N_2})}\right).$$

This tells us that, for any $t \in N_+$, the estimate of the global factor $\hat{G}_t^{(1)}$ at time t is consistent for some rotation of the true factors G_t^0.
In contrast to factor estimates of unilevel factor models, the rate of convergence depends on $\sqrt{T}$ instead of T.

2.4 Step 2: The Initial Estimation of Country-Specific Factors

Now that we have obtained initial estimators of the global factors $\hat{G}^{(1)}$, we use them as proxies for the true global factors G^0 in the factor equation

$$X_{mt} = \Gamma_m^0 \cdot G_t + \Lambda_m^0 \cdot F_{mt}^0 + e_{mt}.$$

The estimators of F_{1t} and F_{2t} are extracted from this model by solving the usual asymptotic principal components problem

$$\begin{aligned}\min_{F_m,\Lambda_m,\Gamma_m} \frac{1}{N_m T} \sum_{i=1}^{N_m} \sum_{t=1}^{T} \left(X_{it} - \Gamma_m \cdot \hat{G}_t^{(1)} - \Lambda_m \cdot F_{mt} \right)^2 & \\ = \frac{1}{N_m T} \operatorname{tr} \left(\left(X_m - \hat{G}^{(1)} \Gamma_m' - F_m \cdot \Lambda_m' \right) \left(X_m - \hat{G}^{(1)} \Gamma_m' - F_m \cdot \Lambda_m' \right)' \right).\end{aligned}$$

Given F_m and Λ_m, the minimizer $\Gamma_m(F_m, \Lambda_m)$ of the above function becomes

$$\Gamma_m(F_m, \Lambda_m) = (X_m - F_m \cdot \Lambda_m')' \hat{G}^{(1)} \left(\hat{G}^{(1)\prime} \hat{G}^{(1)} \right)^{-1},$$

in analogy with the case of unilevel factor models.

The concentrated objective function becomes

$$\begin{aligned}\frac{1}{N_m T} \operatorname{tr} \left(\left(X_m - \hat{G}^{(1)} \Gamma_m(F_m, \Lambda_m)' - F_m \cdot \Lambda_m' \right) \left(X_m - \hat{G}^{(1)} \Gamma_m(F_m, \Lambda_m)' - F_m \cdot \Lambda_m' \right)' \right) & \\ = \frac{1}{N_m T} \operatorname{tr} \left((X_m - F_m \cdot \Lambda_m')' M_{\hat{G}^{(1)}} (X_m - F_m \cdot \Lambda_m') \right),\end{aligned}$$

where $M_{\hat{G}^{(1)}} = I_T - \hat{G}^{(1)} \left(\hat{G}^{(1)\prime} \hat{G}^{(1)} \right)^{-1} \hat{G}^{(1)\prime}$ is the residual maker corresponding to $\hat{G}^{(1)}$. Given F_m, the minimizer $\Lambda_m(F_m)$ of the concentrated function becomes

$$\Lambda_m(F_m) = X_m' M_{\hat{G}^{(1)}} F_m \left(F_m' M_{\hat{G}^{(1)}} F_m \right)^{-1},$$

again in analogy with the case of unilevel factor models.

The finalized concentrated objective function becomes

$$\begin{aligned}\frac{1}{N_m T} \operatorname{tr} \left((X_m - F_m \cdot \Lambda_m(F_m)')' M_{\hat{G}^{(1)}} (X_m - F_m \cdot \Lambda_m(F_m)') \right) & \\ = \frac{1}{N_m T} \operatorname{tr} \left(X_m' M_{\hat{G}^{(1)}} M_F M_{\hat{G}^{(1)}} X_m \right),\end{aligned}$$

where $M_F = I_T - M_{\hat{G}^{(1)}} F_m \left(F_m' M_{\hat{G}^{(1)}} F_m \right)^{-1} F_m' M_{\hat{G}^{(1)}}$ is the residual maker corresponding to $M_{\hat{G}^{(1)}} F_m$.

Defining $X_m^G = M_{\hat{G}^{(1)}} X_m$ and $F_m^G = M_{\hat{G}^{(1)}} F_m$, this function can be rewritten as

$$\frac{1}{N_m T} \operatorname{tr}\left(X_m^{G\prime} X_m^G\right) - \frac{1}{N_m T} \operatorname{tr}\left(X_m^{G\prime} F_m^G \left(F_m^{G\prime} F_m^G\right)^{-1} F_m^{G\prime} X_m^G\right),$$

so normalizing $\frac{F_m^{G\prime} F_m^G}{T} = I_{r_m}$, our analysis of unilevel factor models tells us that the estimator $\tilde{F}_m^G$ of F_m^G is

$$\begin{aligned}\tilde{F}_m^G = \sqrt{T} \times & \text{The collection of } r_m \text{ orthonormal eigenvectors of } X_m^G X_m^{G\prime} \\ & \text{corresponding to its } r_m \text{ largest eigenvalues,}\end{aligned}$$

where we collect the r_m largest eigenvalues of $\frac{1}{N_m T} X_m^G X_m^{G\prime}$ in the matrix $V_{N_m,T}^G$.
Our estimator of the country specific factors F_m is now given by

$$\tilde{F}_m^{(1)} = \tilde{F}_m^G.$$

Note that this is not an estimator of F_m per se, but rather the quantity $M_{\hat{G}(1)} F_m$. We will show below that, nevertheless, $\tilde{F}_m^{(1)}$ consistently estimates a rotation of the true country specific factors F_m^0.

Before moving on, we note an observation that will make our lives much easier:

$$\begin{aligned}\hat{G}^{(1)\prime} \hat{G}^{(1)} &= \tilde{\mu}' \tilde{K}_1' \tilde{K}_1 \tilde{\mu} \\ &= T \cdot \tilde{\mu}' \left(\frac{1}{T} \tilde{K}_1' \tilde{K}_1\right) \tilde{\mu} \\ &= T \cdot \tilde{\mu}' \tilde{S}_1 \tilde{\mu} = T,\end{aligned}$$

since $\tilde{\mu}$ is chosen so that $\tilde{\mu}' \tilde{S}_1 \tilde{\mu} = I_r$.
This means that

$$M_{\hat{G}^{(1)}} = I_T - \frac{1}{T} \hat{G}^{(1)} \hat{G}^{(1)\prime},$$

where

$$\frac{1}{T} \hat{G}^{(1)} \hat{G}^{(1)\prime} = \frac{1}{T} G^0 \tilde{Q}^{(1)} \tilde{Q}^{(1)\prime} G^{0\prime} + O_p\left(\frac{1}{\min(N_1, N_2, T)}\right).$$

Since

$$\left\| \frac{1}{T} G^0 \tilde{Q}^{(1)} \tilde{Q}^{(1)\prime} G^{0\prime} \right\| \leq \operatorname{tr}\left(\frac{G^{0\prime} G^0}{T}\right) \cdot \left\| \tilde{Q}^{(1)} \right\|^2,$$

$\frac{1}{T} G^0 \tilde{Q}^{(1)} \tilde{Q}^{(1)\prime} G^{0\prime} = O_p(1)$, and by implication, so is $\frac{1}{T} \hat{G}^{(1)} \hat{G}^{(1)\prime}$.

2.4.1 Transforming the Model

We can transform the true model as follows:

$$\begin{aligned} X_m &= G^0\Gamma_m^{0\prime} + F_m^0\Lambda_m^{0\prime} + e_m \\ &= G^0\tilde{Q}^{(1)}\left(\tilde{Q}^{(1)}\right)^{-1}\Gamma_m^{0\prime} + F_m^0\Lambda_m^{0\prime} + e_m \\ &= \hat{G}^{(1)}\left(\tilde{Q}^{(1)}\right)^{-1}\Gamma_m^{0\prime} + F_m^0\Lambda_m^{0\prime} + e_m - \left(\hat{G}^{(1)} - G^0\tilde{Q}^{(1)}\right)\left(\tilde{Q}^{(1)}\right)^{-1}\Gamma_m^{0\prime}. \end{aligned}$$

Premultiplying both sides by $M_{\hat{G}^{(1)}}$ yields

$$X_m^G = \underbrace{M_{\hat{G}^{(1)}}F_m^0}_{F_m^{G0}}\Lambda_m^{0\prime} + M_{\hat{G}^{(1)}}e_m + \underbrace{M_{\hat{G}^{(1)}}\left(\hat{G}^{(1)} - G^0\tilde{Q}^{(1)}\right)\left(\tilde{Q}^{(1)}\right)^{-1}\Gamma_m^{0\prime}}_{a_m}.$$

The original paper uses this expansion to prove that $\tilde{F}_m^G$ is consistent for some rotation of F_m^0. However, because the eigenvectors $\tilde{\mu}$ do not converge to some quantity in this case (due to the non-uniqueness of eigenvalues at the limit), we cannot establish that $\left(\tilde{Q}^{(1)}\right)^{-1}$ is $O_p(1)$, which means that the proof in the original paper falls apart.

Fortunately, given that

$$\tilde{F}_{mt}^G - \tilde{\Omega}_m' F_{mt}^{G0} = o_p(1)$$

for some $O_p(1)$ rotation $\tilde{\Omega}_m$, we can easily establish that $\tilde{F}_{mt}^G$ is consistent for some rotation of F_{mt}^0 as well.
To see this, first note that $F_m^{G0} = M_{\hat{G}^{(1)}}F_m^0 = F_m^0 - \frac{1}{T}\hat{G}^{(1)}\hat{G}^{(1)\prime}F_m^0$ by definition, so

$$F_{mt}^{G0} = F_{mt}^0 - \frac{1}{T}F_m^{0\prime}\hat{G}^{(1)}\hat{G}_t^{(1)}.$$

Since

$$F_m^{0\prime}\hat{G}^{(1)} = F_m^{0\prime}\left(\hat{G}^{(1)} - G^0\tilde{Q}^{(1)}\right) + F_m^{0\prime}G^0\tilde{Q}^{(1)},$$

we have

$$\frac{1}{T}F_m^{0\prime}\hat{G}^{(1)} - \frac{F_m^{0\prime}G^0}{T}\tilde{Q}^{(1)} = \left(\frac{1}{\sqrt{T}}F_m^{0\prime}\right)\left(\frac{1}{\sqrt{T}}\left(\hat{G}^{(1)} - G^0\tilde{Q}^{(1)}\right)\right) = o_p(1).$$

Furthermore,

$$\frac{F_m^{0\prime}G^0}{T} \xrightarrow{p} O,$$

so that

$$\frac{1}{T}F_m^{0\prime}\hat{G}^{(1)}\hat{G}_t^{(1)} = o_p(1).$$

As such,

$$F_{mt}^{G0} - F_{mt}^0 = o_p(1),$$

and because $\tilde{\Omega}_m = O_p(1)$, we have

$$\tilde{F}_{mt}^G - \tilde{\Omega}_m' F_{mt}^0 = o_p(1).$$

Factor-Augmented Vector Autoregressions

Bernanke et al. (2005)

In Bernanke et al. (2005), the authors introduce a Factor-Augmented Vector Autoregression (FAVAR) model, in which the factors, which include both observable and unobservable variables, are assumed to follow a VAR and serve as the common factors in a unilevel factor model.

3.1 Motivation for Factor Augmentation

To motivate the use of a factor-augmented version of the VAR model instead of the usual SVAR model, the authors of the paper cite three shortcomings of traditional SVAR analyses:

- **Discrepancies in Information Sets**
  While policymakers usually make their decision based on a multitude of macroeconomic variables, SVAR models only include a select few of those variables, necessarily leading to omitted variable bias. In other words, the information set implied by the VAR model and used in actual policymaking are vastly different.

- **Measurement Errors**
  It is often unclear whether single variables such as GDP or inflation can sufficiently represent economic concepts such as real activity or the nominal side of the economy. In addition, even if they can represent such concepts, there are a multitude of measurement errors associated with these variables, which can distort analysis. For this reason, it seems inappropriate to conduct VAR analysis exclusively with observable variables, as in traditional models.

- **Limitations in Analysis**
  In traditional SVAR models, we can conduct impulse response analysis or variance decomposition for those variables that are included in the VAR system. Thus, a model that enables us to study the responses of a variety of macroeconomic variables to certain

shocks, such as monetary policy shocks, will serve as an improvement upon traditional SVAR models.

In light of the above limitations of the standard SVAR model, the authors suggest estimating a factor-augmented VAR model, in which unobserved factors affecting a plethora of macro variables are included in the VAR system alongside the usual observable variables.
To motivate this specific setup, a simple backward-looking model of the economy is introduced as follows:

$$\pi_t = \delta\pi_{t-1} + \kappa(y_{t-1} - y^n_{t-1}) + s_t \qquad \text{(Phillips Curve)}$$
$$y_t = y^n_{t-1} - \psi\,(i_{t-1} - \pi_{t-1}) + d_t \qquad \text{(IS Curve)}$$
$$y^n_t = \rho y^n_{t-1} + \eta_t \qquad \text{(Evolution of Natural Output)}$$
$$s_t = \alpha s_{t-1} + \upsilon_t \qquad \text{(Evolution of Cost-push Shocks)}$$
$$i_t = \beta\pi_t + \gamma(y_t - y^n_t) + \varepsilon_t. \qquad \text{(Monetary Policy Rule)}$$

This model can be seen as the solution to a traditional NK model with shocks evolving according to an AR(1) process.
In this model, there are five endogenous variables, the inflation rate π_t, output y_t, natural output y^n_t, the cost-push shock s_t, and the nominal interest rate i_t, that jointly follow a VAR process. Of these, natural output and the cost-push shock are unobserved, while the other variables are observed, so it stands to reason that we should estimate a VAR model with two unobsered factors y^n_t, s_t and three observed variables y_t, π_t and i_t.
However, even this might be insufficient, since y_t and π_t often represent the real and nominal sides of the economy in a NK model, and there is reason to believe that GDP and inflation might be insufficient proxies for these economic concepts. Therefore, it is reasonable to think of the model as a VAR with four unobserved variables y^n_t, s_t, y_t, π_t and a single observed variable, the central bank's policy instrument i_t. This is the main FAVAR specification used in the paper.

3.2 The FAVAR Model

Formally, consider k unobservable factors F_t and m observable macro variables Y_t that are jointly generated by a VAR process specified as follows:

$$\Phi(L) \cdot \begin{pmatrix} F_t \\ Y_t \end{pmatrix} = v_t,$$

where $\Phi(L) = I_{k+m} - \Phi_1 L - \cdots - \Phi_p L^p$ is an AR lag polynomial of some lag order p, and v_t is a white noise process with covariance matrix Σ.

F_t and Y_t are then assumed to be common factors affecting a wide variety of macroeconomic variables. Specifically, let there be N "informational" macro variables, and that the ith such variable is determined as

$$X_{it} = \lambda_i^{f\prime} \cdot F_t + \lambda_i^{y\prime} \cdot Y_t + \varepsilon_{it},$$

where $\lambda_i^f \in \mathbb{R}^k$ and $\lambda_i^y \in \mathbb{R}^m$ are the factor loadings of x_{it} on F_t and Y_t. Heuristically, $X_{1t}, \cdots, X_{Nt}$ may be taken to be "noisy measures of the unobserved factors".
Defining $X_t = (X_{1t}, \cdots, X_{Nt})'$, $\Lambda^f = (\lambda_1^f, \cdots, \lambda_N^f)' \in \mathbb{R}^{N\times k}$, $\Lambda^y = (\lambda_1^y, \cdots, \lambda_m^y)' \in \mathbb{R}^{N\times m}$, and $\varepsilon_t = (\varepsilon_{1t}, \cdots, \varepsilon_{Nt})'$, we now have the concatenated model

$$X_t = \Lambda^f \cdot F_t + \Lambda^y \cdot Y_t + \varepsilon_t.$$

Note that this setup is virtually identical to the unilevel factor model studied above, except that now we explicitly specify the dynamics of the factors, namely that they follow a VAR process.

The FAVAR model is now determined by the following equations (where the lag order is set to 1 for notational simplicity):

$$X_t = \Lambda^f \cdot F_t + \Lambda^y \cdot Y_t + \varepsilon_t \qquad \text{(Measurement equation)}$$

$$\begin{pmatrix} F_t \\ Y_t \end{pmatrix} = \Phi \cdot \begin{pmatrix} F_{t-1} \\ Y_{t-1} \end{pmatrix} + v_t \qquad \text{(Transition Equation)}$$

3.3 Estimating the FAVAR Model

The authors propose two ways to estimate this model:

The first estimation method is a two-step approach, in which the factors F_t, Y_t are first estimated via principal components as if the measurement equation represented a unilevel factor model, as in Bai (2003), and then the estimated factors are plugged into the transition equation to recover the VAR parameters. This has the advantage of being semi-parametric and thus applicable to more general settings, but suffers from the problem of generated regressors in the second step. Furthermore, we do not use the fact that Y_t is observable in the first step.
As noted above when discussing the models in Bai and Ng (2002) and Bai (2003), the principal components estimator of the factors in the measurement equations consistenly estimate a rotation of the true factors F_t, Y_t under regularity assumptions; therefore, in order to recover the estimator of F_t from the estimator of F_t, Y_t, we must devise a means of extracting the part of the latter that is independent of Y_t. This is done by imposing identification restrictions in the second step of the estimation procedure.

The second estimation method is a one-step, or joint estimation, approach, in which the factors F_t and the parameters of the model are estimated at once by estimating the state-space model given by

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} \Lambda^f & \Lambda^y \\ O & I_m \end{pmatrix} \begin{pmatrix} F_t \\ Y_t \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ \mathbf{0} \end{pmatrix}, \quad \text{where} \quad \varepsilon_t \sim iidN(\mathbf{0}, \Omega) \qquad \text{(Measurement equation)}$$

$$\begin{pmatrix} F_t \\ Y_t \end{pmatrix} = \Phi \cdot \begin{pmatrix} F_{t-1} \\ Y_{t-1} \end{pmatrix} + v_t \quad \text{where} \quad v_t \sim iidN(\mathbf{0}, \Sigma), \qquad \text{(Transition Equation)}$$

which explicitly imposes the restriction that Y_t are observed and that the errors are iid normal, unlike in the two-step approach. Because this model does not suffer from the generated regressors problem, it is more robust in terms of estimation, but because the dimension N of X_t is large, we must rely on Bayesian priors to smooth the likelihood function, which makes computation very costly.

3.4 Identification of the FAVAR Model

Before estimating the model, it is necessary to impose some restrictions on the model parameters. There are two types of these restrctions: restrictions that identify the unobserved component F_t and those that help identify the structural shocks. We discuss each restriction in turn.

3.4.1 Identifying F_t in the One-Step Approach

Because F_t is unobserved, if we do not impose any restrictions on the model, we are unable to obtain a unique estimate of F_t. Specifically, consider a non-singular matrix $H \in \mathbb{R}^{k\times k}$ and $B \in \mathbb{R}^{k\times m}$, and define

$$F_t^* = HF_t + BY_t.$$

Then, F_t^* is also a k-vector of unobserved factors such that $F_t = H^{-1}F_t^* - H^{-1}BY_t$ and thus

$$\begin{aligned}
X_t &= \Lambda^f \cdot F_t + \Lambda^y \cdot Y_t + \varepsilon_t \\
&= \Lambda^f \cdot \left(H^{-1}F_t^* - H^{-1}BY_t\right) + \Lambda^y \cdot Y_t + \varepsilon_t \\
&= \left(\Lambda^f \cdot H^{-1}\right)F_t^* + \left(\Lambda^y - \Lambda^f H^{-1}B\right)Y_t + \varepsilon_t \\
&= \Lambda^{f*} \cdot F_t^* + \Lambda^{y*} \cdot Y_t + \varepsilon_t,
\end{aligned}$$

so that X_t retains the same linear factor model structure as before.
Furthermore, since

$$\begin{pmatrix} F_t \\ Y_t \end{pmatrix} = \begin{pmatrix} H^{-1}F_t^* - H^{-1}BY_t \\ Y_t \end{pmatrix} = \begin{pmatrix} H^{-1} & -H^{-1}B \\ O & I_m \end{pmatrix} \begin{pmatrix} F_t^* \\ Y_t \end{pmatrix},$$

defining

$$\begin{aligned}
\Phi^* &= \begin{pmatrix} H^{-1} & -H^{-1}B \\ O & I_m \end{pmatrix}^{-1} \Phi \begin{pmatrix} H^{-1} & -H^{-1}B \\ O & I_m \end{pmatrix}, \\
u_t &= \begin{pmatrix} H^{-1} & -H^{-1}B \\ O & I_m \end{pmatrix}^{-1} v_t \quad \text{and} \quad \Sigma^* = \begin{pmatrix} H^{-1} & -H^{-1}B \\ O & I_m \end{pmatrix}^{-1} \Sigma \begin{pmatrix} H^{-1} & -H^{-1}B \\ O & I_m \end{pmatrix}^{\prime -1}
\end{aligned}$$

tells us that

$$\begin{pmatrix} F_t^* \\ Y_t \end{pmatrix} = \Phi^* \begin{pmatrix} F_{t-1}^* \\ Y_{t-1} \end{pmatrix} + u_t.$$

Therefore, the new factors F_t^* constructed as a linear combination of F_t and Y_t also satisfy the equations

$$X_t = \Lambda^{f*} \cdot F_t^* + \Lambda^{y*} \cdot Y_t + \varepsilon_t,$$
$$\begin{pmatrix} F_t^* \\ Y_t \end{pmatrix} = \Phi^* \begin{pmatrix} F_{t-1}^* \\ Y_{t-1} \end{pmatrix} + u_t,$$

where $u_t \sim WN(\mathbf{0}, \Sigma^*)$.
Because F_t and F_t^* are both unobserved components, the likelihood of the model will be the same under the original parameters $\Lambda^f, \Lambda^y, \Phi, \Sigma, \Omega$ and the new paramters $\Lambda^{f*}, \Lambda^{y*}, \Phi^*, \Sigma^*, \Omega$. This means that the model is unidentified, and that estimators of F_t will not be consistent under the one-step estimation approach. We now survey the restrictions that must be imposed in order for estimates of the factors to consistently estimate the true factors in the one-step approach.

Suppose we do not want to impose any restrictions on the VAR parameters Φ and Σ governing the factor dynamics. Then, Λ^f and Λ^y must be required to satsify some restrictions in a manner such that, if $F_t^* = HF_t + BY_t$ is a transformation of the factors that yields the same value of the likelihood as F_t, then $H = I_n$ and $B = O$.
The authors propose imposing the restriction that

$$\Lambda^f = \begin{pmatrix} I_k \\ \Lambda^{f(2)} \end{pmatrix} \in \mathbb{R}^{N \times k} \quad \text{and} \quad \Lambda^y = \begin{pmatrix} O_{k \times m} \\ \Lambda^{y(2)} \end{pmatrix} \in \mathbb{R}^{N \times m}.$$

If Λ^f, Λ^{f*}, Λ^y and Λ^{y*} satisfy the above restrictions, then $H = I_k$ and $B = O$, so that $F_t^* = F_t$ and the model is identified.
The above restirctions imply that the first k variables included in X_t are determined as

$$\begin{pmatrix} X_{1t} \\ \vdots \\ X_{kt} \end{pmatrix} = F_t + \begin{pmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{kt} \end{pmatrix},$$

so that they are precisely the unobserved factors F_t with additional noise represented by $\varepsilon_{1t}, \cdots, \varepsilon_{kt}$. Therefore, under the proposed restrictions, we are identifying the unobserved factors by assuming that the first k variables in X_t incorporate information about F_t and F_t alone; simply put, F_t and Y_t are distinguished by assuming that the former can affect the first k variables, but the latter cannot.

3.4.2 Identifying the Factors in the Two-Step Approach

In the two-step estimation approach, the difficulty of recovering a unique estimator of the factors can be circumvented by simply estimaing the principal components under the normalization presented in Bai and Ng (2002) and Bai (2003). Specifically, letting $X = (X_1, \cdots, X_T)'$, provided that the k largest eigenvalues of XX' and the probability limit of

$$\left[\frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix}\lambda_i^f \\ \lambda_i^y\end{pmatrix}\begin{pmatrix}\lambda_i^{f\prime} & \lambda_i^{y\prime}\end{pmatrix}\right]\left[\frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix}F_t \\ Y_t\end{pmatrix}\begin{pmatrix}F_t' & Y_t'\end{pmatrix}\right]$$

are distinct for any N, T, the k unobsered factors in F_t can be estimated uniquely up to sign changes by $\hat{C} = (\hat{C}_1, \cdots, \hat{C}_T)'$, where

$$\hat{C} = \sqrt{T} \times \text{The orthonormal eigenvectors corresponding to the } k \text{ largest eigenvalues of } XX',$$

so that $\frac{\hat{C}'\hat{C}}{T} = I_k$.

Furthermore, as we have seen above, under appropriate regularity assumptions $\hat{C}_t$ is consistent for some rotation of the true factors F_t, Y_t.

It is then up to restrictions imposed in the second step of the estimation procedure to separate the part of $\hat{C}_t$ due to Y_t, so that the resulting estimator of F_t truly represents the unobserved factors F_t.

3.4.3 Identifying the Monetary Policy Shock in the One-Step Approach

In this paper, the authors identify the monetary policy shock in the usual recursive manner, by assuming that the factors cannot contemporaneously affect the policy instrument i_t, which is the only variable comprising Y_t. They emphasize that this means the unobserved factors F_t do not have to be separately identified, unlike with other identification approaches.

To achieve identification of the state space model used for one-step estimation under this identification scheme for monetary policy shocks, we need only impose the restriction that B_0 is lower triangular in the structural factor dynamics

$$B_0 \begin{pmatrix} F_t \\ Y_t \end{pmatrix} = B_0 \Phi \begin{pmatrix} F_{t-1} \\ Y_{t-1} \end{pmatrix} + B_0 v_t,$$

where $B_0^{-1} B_0'^{-1} = \Sigma$, alongside the restrictions on Λ^f and Λ^y prposed above.
Under the suggested identificaiton scheme for monetary policy shocks, factors are slow-moving (do not contemporaneously respond to the interest rate, which means that they respond ”slowly” to changes in policy) in contrast to other fast-moving variables, which respond contemporaneously to the interest rate and thus respond relatively ”quickly” to changes in policy. Since the first k variables in X_t are just the unobserved factors F_t with noise, this means that the first k variables in X_t must be slow-moving; otherwise, we would have a model where fast moving variables are determined by slow moving variables plus noise, which is unreasonable.

3.4.4 Identifying the Monetary Policy Shock in the Two-Step Approach

To achieve identification of the model under the two-step estimation approach with the identification scheme for monetary policy shocks proposed above, we must find a way to separate the effects of Y_t from $\hat{C}$ in the initial step. This is because, as is, $\hat{C}$ represents a linear combination of the true factors F_t and i_t.
To make the necessity of this separation clearer, suppose that $\hat{C}_t$ is used as a proxy for F_t to estimate the VAR parameters. Since there exist random matrices H_1, H_2 taking values in $\mathbb{R}^{k\times k}$ and $\mathbb{R}^{k\times m}$ such that

$$\hat{C}_t \approx H_F \cdot F_t + H_y \cdot i_t,$$

for large N, T, in this case $\hat{C}_t$ would respond contemporaneously to changes in i_t through the term $H_y \cdot i_t$. Since F_t are characterized by their slow response to i_t, in this case $\hat{C}_t$ cannot be viewed as a true estimator of F_t.
In addition, $\hat{C}_t$ could not be used instead of F_t to estimate the VAR model. Specifically, since $\hat{C}_t$ responds contemporaneously to i_t, this means that we cannot impose the recursive identification assumption to the VAR system with endogenous variables $\hat{C}_t, i_t$, and therefore that the VAR model in the second step of the estimation procedure remains unidentified.

A possible way of partialling out the effect of i_t in $\hat{C}_t$ is to regress $\hat{C}_t$ on $\hat{C}_t^*$, a measure of the common factors other than i_t, and i_t in the linear model

$$\hat{C}_t = b_c \cdot \hat{C}_t^* + b_i \cdot i_t + e_t.$$

Then, using the OLS estimator $\hat{b}_c$ and $\hat{b}_i$ of b_c and b_i, the residual

$$\hat{F}_t = \hat{C}_t - \hat{b}_i \cdot i_t$$

would be used as the estimator of F_t. This regression in effect represents separating $\hat{C}_t$ into the part depending directly on i_t and the part with no direct dependence on i_t, so that $\hat{C}_t - \hat{b}_i \cdot i_t$ represents the part of $\hat{C}_t$ that is not directly determined by i_t.

The reason we do not directly regress $\hat{C}_t$ on i_t is because, in general, F_t and i_t may be correlated, so that removing every part of $\hat{C}_t$ that is correlated with i_t could end up removing the parts of F_t correlated with i_t as well.
Instead, the above linear model allows us to separate from $\hat{C}_t$ the part that is directly correlated with i_t, while leaving intact the part that is indirectly correlated with i_t through F_t.

As a measure of $\hat{C}_t^*$, the authors propose using the principal components from the truncated version of X_t consisting only of slow-moving variables. This method makes use of the identification assumption that F_t are the slow-moving variables themselves and thus that much of the information in F_t is contained in the slow-moving variables comprising X_t.

The reason $\hat{C}_t*$ is not used as a direct estimator of F_t is because, even though it represents much of the information in F_t, it does not reflect all the information in F_t since it was derived only using slow-moving variables, whereas information on F_t could potentially be contained in the fast-moving variables if their loading on F_t is non-zero.

Interactive Fixed Effects Models

Bai (2009)

Bai introduces the interactive fixed effects model in his 2009 paper, which imposes a factor structure on the unobserved individual/time effects in a fixed effects model and exploits the large N, large T framework to estimate the model.

4.1 Interactive Fixed Effects

Consider a typical panel data model with dependent variable Y_{it}, independent variable X_{it}, individual/time effects δ_{it} and idosyncratic errors e_{it} that are related as

$$Y_{it} = X_{it}'\beta + \delta_{it} + e_{it}$$

for any $1 \leq i \leq N$, $1 \leq t \leq T$. In the fixed effects literature, it is customary to assume large N and small T, and impose an additive structure on the fixed effects δ_{it}, that is, to assume that

$$\delta_{it} = \alpha_i + \varepsilon_t.$$

This allows the individual effects α_i to be removed via within-sample demeaning or first differencing, and the time effects to be controlled for via time dummies.

However, there are cases in which the additive structure may be inappropriate, and instead an interactive effects framework, in which the individual/time effects are represented as

$$\delta_{it} = \lambda_i' F_t$$

for an r-dimensional vector λ_i and F_t, is needed. Note that this specification nests the additive frameworks, since we need only let

$$\lambda_i = \begin{pmatrix} \alpha_i \\ 1 \end{pmatrix} \quad \text{and} \quad F_t = \begin{pmatrix} 1 \\ \varepsilon_t \end{pmatrix}$$

for the additive structure to be represented in terms of interactive effects.

To illustrate the usefulness of the interactive effects structure, consider, for instance, a panel model for wages with time-varying prices for the unobserved components. Specifically, suppose that Y_{it} stands for the wage of individual i with age$t$, X_{it} a vector of exogenous variables that affect the wage Y_{it}, and λ_i an r-vector of unobserved individual characteristics such as ability and social awareness. The traditional fixed effects model, in which

$$Y_{it} = X_{it}'\beta + \lambda_i + e_{it},$$

implicitly assumes that the price of each unobserved component in λ_i does not vary across cohorts and thus is normalized to 1. However, if the price of each unobserved component varies across cohorts and is represented in the vector F_t, then the interactive fixed effects framework

$$Y_{it} = X_{it}'\beta + \lambda_i' F_t + e_{it}$$

becomes more appropriate.

Before moving onto estimation, we first organize the model into a vector/matrix form as we did for the unilevel and multilevel factor models.

Suppose that there are r unobserved factors and k observed exogenous variables, so that λ_i, F_t are r-dimensional vectors and X_{it}, β are k-dimensional.

Defining $X_i = (X_{i1}, \cdots, X_{iT})'$, $Y_i = (Y_{i1}, \cdots, Y_{iT})'$, $e_i = (e_{i1}, \cdots, e_{iT})'$ and $F = (F_1, \cdots, F_T)'$, the model can be written as

$$\underbrace{Y_i}_{T\times 1} = \underbrace{X_i}_{T\times k} \cdot \underbrace{\beta}_{k\times 1} + \underbrace{F}_{T\times r} \cdot \underbrace{\lambda_i}_{r\times 1} + \underbrace{e_i}_{T\times 1}.$$

On the other hand, combining the data by time, so that $Y_t = (Y_{1t}, \cdots, Y_{Nt})$, $X_t = (X_{1t}, \cdots, X_{Nt})$, $e_t = (e_{1t}, \cdots, e_{Nt})$ and $\Lambda = (\lambda_1, \cdots, \lambda_N)'$, we have

$$\underbrace{Y_t}_{N\times 1} = \underbrace{X_t}_{N\times k} \cdot \underbrace{\beta}_{k\times 1} + \underbrace{\Lambda}_{N\times r} \cdot \underbrace{F_t}_{r\times 1} + \underbrace{e_t}_{N\times 1}.$$

4.2 Estimation of the Interactive Effects Model

The estimators of β, Λ and F are found as the minimizers of the average sum of squared errors

$$\begin{aligned} SSR(\beta,F,\Lambda) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(Y_{it}-X_{it}'\beta-\lambda_i'F_t\right)^2 \\ &= \frac{1}{NT}\sum_{i=1}^{N}(Y_i-X_i\beta-F\lambda_i)'(Y_i-X_i\beta-F\lambda_i) \end{aligned}$$

subject to the normalizations $\frac{F'F}{T}=I_r$ and diagonal $\Lambda'\Lambda$.

4.2.1 Concentrating out Λ

We first concentrate out Λ; for any $1\leq i\leq N$, given β and F the minimizer $\lambda_i(\beta,F)$ of the objective function with respect to λ_i satisfies the f.o.c.

$$F'(Y_i-X_i\beta-F\lambda_i(\beta,F))=\mathbf{0}.$$

Therefore, we have

$$\lambda_i(\beta,F)=(F'F)^{-1}F'(Y_i-X_i\beta),$$

and substituting this into the objective function, we have the concentrated version

$$\begin{aligned} V(\beta,F) &= \frac{1}{NT}\sum_{i=1}^{N}\left[Y_i-X_i\beta-F(F'F)^{-1}F'(Y_i-X_i\beta)\right]'\left[Y_i-X_i\beta-F(F'F)^{-1}F'(Y_i-X_i\beta)\right] \\ &= \frac{1}{NT}\sum_{i=1}^{N}(Y_i-X_i\beta)'M_F(Y_i-X_i\beta), \end{aligned}$$

where $M_F=I_T-F(F'F)^{-1}F'$ is the residual maker associated with the $T\times r$ matrix F. We define the projection matrix $P_F=F(F'F)^{-1}F'$ for later use.
Since M_F is symmetric and idempotent, its rank is given by

$$\text{rank}(M_F)=\text{tr}(M_F)=T-r.$$

4.2.2 Estimating β and F

The estimators of β and F are obtained as the minimizers of $V(\beta,F)$.
First, assume that F is known. Then, the minimizer $\beta(F)$ of $V(\beta,F)$ satsifeis the f.o.c.

$$\sum_{i=1}^{N}X_i'M_F(Y_i-X_i\cdot\beta(F))=\mathbf{0},$$

so that $\beta(F)$ is the least squares estimator

$$\beta(F) = \left[\sum_{i=1}^{N} X_i' M_F X_i\right]^{-1} \sum_{i=1}^{N} X_i' M_F Y_i.$$

On the other hand, if β is known, then defining $W_i = Y_i - X_i\beta$ and $W = (W_1, \cdots, W_N)$, the objective function can be written as

$$\begin{aligned}\frac{1}{NT}\sum_{i=1}^{N} W_i' M_F W_i &= \frac{1}{NT}\operatorname{tr}(W' M_F W)\\ &= \frac{1}{NT}\operatorname{tr}(W'W) - \frac{1}{NT}\operatorname{tr}(W'F(F'F)^{-1}F'W).\end{aligned}$$

The minimizer of the above function with respect to F subject to the normalization $\frac{F'F}{T} = I_r$ is the solution to the maximization problem

$$\begin{aligned}\max_{F \in \mathbb{R}^{T\times r}} \quad & \operatorname{tr}(F'WW'F)\\ \text{subject to} \quad & \frac{F'F}{T} = I_r.\end{aligned}$$

We already proved, in our study of unilevel factor models, that $\operatorname{tr}(F'WW'F)$ is bounded above by the sum of the r largest eigenvalues of the positive semidefinite matrix WW', and this upper bound is attained when we set F equal to $\sqrt{T}$ times the r orthonormal eigenvectors corresponding to the r largest eigenvalues of WW'. Therefore, as is par for the course by now in factor models, we set

$$F(\beta) = \sqrt{T} \times \text{Orthonormal eigenvectors of } WW' \text{ corresponding to its } r \text{ largest eigenvalues .}$$

We also denote by V_{NT} the $r \times r$ diagonal matrix collecting the r largest eigenvalues of $\frac{1}{NT}WW'$ as its diagonal elements, as is also customary by now. It follows that

$$\left(\frac{1}{NT}WW'\right)\cdot F(\beta) = F(\beta)V_{NT},$$

and by definition, $F(\beta)$ can also be characterized as the $T \times r$ matrix that satisfies the above equation along with the normalization $\frac{F(\beta)'F(\beta)}{T} = I_r$.

4.2.3 The Estimators of β, F, Λ

So far, we have derived the estimators of β and F assuming that the other was given. Suppose that $\hat{\beta}$ and $\hat{F}$ are minimizers of $V(\beta, F)$. Then, by definition,

$$\hat{\beta} = \beta(\hat{F}) \quad \text{and} \quad \hat{F} = F(\hat{\beta}).$$

To see why this is the case, suppose that $\hat{\beta} \neq \beta(\hat{F})$. Then, by the definition of $\beta(F)$ for an arbitrary $T \times r$ matrix valued random matrix F, it follows that

$$V(\hat{\beta}, \hat{F}) > V(\beta(\hat{F}), \hat{F}),$$

which contradicts the assumption that $\hat{\beta}$ and $\hat{F}$ minimize $V(\beta, F)$.
It follows from a similar line of reasoning that $\hat{F} = F(\hat{\beta})$ must also hold true.
Therefore, $\hat{\beta}$ and $\hat{F}$ must satisfy the nonlinear equations

$$\begin{aligned}
\left[\sum_{i=1}^{N} X_i' M_{\hat{F}} X_i\right]^{-1} \sum_{i=1}^{N} X_i' M_{\hat{F}} Y_i &= \hat{\beta} \\
\left(\frac{1}{NT} \sum_{i=1}^{N} (Y_i - X_i\hat{\beta})(Y_i - X_i\hat{\beta})'\right) \cdot \hat{F} &= \hat{F} V_{NT}, \\
\frac{\hat{F}'\hat{F}}{T} &= I_r.
\end{aligned}$$

The estimator of each λ_i is then given as

$$\hat{\lambda}_i = \lambda_i(\hat{\beta}, \hat{F}) = \frac{1}{T} \hat{F}'(Y_i - X_i\hat{\beta})$$

for every $1 \leq i \leq N$.

We have estimated β, F, Λ by first concentrating out Λ, but of course it is possible to concentrate out F first and then jointly estimate β and Λ. The method chosen in Bai (2009) is because his primary interest is in the common factors F_t instead of the individual unobserved components λ_i.

4.3 Assumptions and Preliminaries

To establish consistency of the estimators $\hat{\beta}, \hat{F}$ above, and later their rates of convergence, we require the following assumptions. They are mostly the same as the assumptions in Bai (2003) used to prove the asymptotic properties of the factor and factor loading estimators, along with some assumptions that takes into consideration the additional term $X_{it}'\beta$ that is not present in pure factor models.

To make the proofs as simple as possible, we assume that the idiosyncratic errors are i.i.d. across both the cross-sectional and time dimensions. The specific assumptions are as follows:

(1) **Bounded Moments of Exogenous Variables**
We assume that there exists an $M < +\infty$ such that

$$\sup_{i \in N_+, t \in N_+} \mathbb{E}|X_{it}|^4 < M.$$

Since $\mathbb{E}|X_{it}|^2 \leq \left(\mathbb{E}|X_{it}|^4\right)^{\frac{1}{2}}$ for any $i, t \in N_+$ by Jensen's inequality, we have

$$\sup_{i \in N_+, t \in N_+} \mathbb{E}|X_{it}|^2 < M$$

as well.

(2) **Identification of β**
Let $\mathcal{F}'$ be the set of all full rank $T \times r$ matrices. For any $F \in \mathcal{F}'$, define

$$D(F) = \frac{1}{NT} \sum_{i=1}^{N} X_i' M_F X_i - \frac{1}{N^2 T} \sum_{i=1}^{N} \sum_{j=1}^{N} X_i' M_F X_i a_{ij},$$

where

$$a_{ij} = \lambda_i^{0\prime} \left(\frac{\Lambda^{0\prime} \Lambda^0}{N} \right)^{-1} \lambda_j^0$$

for any $i, j \in N_+$.
We assume that, there exists a $\rho_{\min} > 0$ such that the minimum eigenvalue of $D(F)$ for any $F \in \mathcal{F}'$ is always greater than or equal to $\rho_{\min}$. This ensures that $D(F)$ is positive definite for any $F \in \mathcal{F}'$.
In addition, we assume that the parameter space of β is bounded.
These ensure that the objective function has a unique minimum at the true value β^0 of β, hence the name of the assumption.

(3) **Non-triviality of Scaled Factors**

We assume that the r largest eigenvalues of XX' are always positive. This implies that the r largest eigenvalues of XX' are always positive, and as such that, when we use the scaled factors $\hat{F} = \frac{1}{NT} XX' \tilde{F}$ later on, the scaled factors are non-zero, or non-trivial.

(4) **Second Moment Convergence of True Factors and Factor Loadings**

We assume that $\{|F_t^0| \mid t \in N_+\}$ and $\{|\lambda_i^0| \mid i \in N_+\}$ are L^2-bounded and that

$$\frac{F^{0\prime} F^0}{T} \xrightarrow{p} \Sigma_F \quad \text{and} \quad \frac{\Lambda^{0\prime} \Lambda^0}{T} \to \Sigma_\Lambda$$

for some positive definite matrices $\Sigma_F, \Sigma_\Lambda \in \mathbb{R}^{r \times r}$.

The factor loadings λ_i are also assumed to be stochastic this time around because they can represent unobserved individual characteristics in the fixed effects model framework.

(5) **I.I.D. Idiosyncratic Errors**

We assume that the process $\{e_{it}\}_{i \in N_+, t \in N_+}$ is independent and identically distributed with finite fourth moment

$$\mathbb{E}\left[e_{it}^4\right] = \mu_4 < +\infty.$$

This implies that the second moment is also finite;

$$\mathbb{E}\left[e_{it}^2\right] = \sigma^2 < +\infty.$$

(6) **Independence of Errors**

We assume that e_{it} is independent of X_{js}, λ_j, F_s for any $j, s \in N_+$.

We can consider the following implications of the above assumptions:

- **The Rates of Convergence of** X_i, X_t
  Note that

$$\frac{1}{T}\|X_i\|^2 \leq \frac{1}{T}\sum_{t=1}^{T}|X_{it}|^2;$$

since

$$\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T}|X_{it}|^2\right] = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}|X_{it}|^2 < M,$$

it follows that $\frac{1}{T}\|X_i\|^2$ is $O_p(1)$.
Likewise, $\frac{1}{N}\|X_t\|^2$ is also $O_p(1)$, and because

$$\frac{1}{NT}\sum_{i=1}^{N}\|X_i\|^2 \leq \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}|X_{it}|^2,$$

it follows that $\frac{1}{NT}\sum_{i=1}^{N}\|X_i\|^2 = O_p(1)$ as well.

- **The Rate of Convergence of the Product of Exogenous Variables and Errors**

It is also true that

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} X_i' e_i = O_p(1).$$

To see this, note that

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} X_i' e_i = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it} e_{it},$$

and recall that

$$\sup_{i,t\in N_+} \mathbb{E}|X_{it}|^2 < M$$

for some $M > 0$. Therefore,

$$\begin{aligned}
\mathbb{E}\left|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} X_i' e_i\right|^2 &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}\left[e_{it}e_{js}X_{it}'X_{js}\right] \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}\left[e_{it}e_{js}\right]\mathbb{E}\left[X_{it}'X_{js}\right] \\
&= \sigma^2\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}|X_{it}|^2 \\
&\leq \sigma^2 M,
\end{aligned}$$

so that

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} X_i' e_i = O_p(1).$$

- **The Rate of Convergence of the Product of Factors and Errors**

Let $\mathcal{F}$ be the set of all $T \times r$ matrices F such that $\frac{F'F}{T} = I_r$. Letting $F_1, \cdots, F_T$ be the rows of any $F \in \mathcal{F}$, we can show that

$$\sup_{F \in \mathcal{F}} \left[\frac{1}{N} \sum_{i=1}^{N} \left| \frac{1}{T} \sum_{t=1}^{T} F_t e_{it} \right|^2 \right] = O_p \left(\frac{1}{\min(\sqrt{N}, T)} \right),$$

that is, $\frac{1}{N} \sum_{i=1}^{N} \left| \frac{1}{T} \sum_{t=1}^{T} F_t e_{it} \right|^2$ converges at a uniform rate in F.

We first expand the above term as

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^{N} \left| \frac{1}{T} \sum_{t=1}^{T} F_t e_{it} \right|^2 &= \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} e_{it} e_{is} F_t' F_s \\ &= \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \left(e_{it} e_{is} - \mathbb{E}\left[e_{it} e_{is} \right] \right) F_t' F_s + \sigma^2 \cdot \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=1}^{T} |F_t|^2, \end{aligned}$$

where the last term follows because $\mathbb{E}[e_{it} e_{is}] = 0$ if $t \neq s$ and $\mathbb{E}[e_{it} e_{is}] = \sigma^2$ if $t = s$.

By the Cauchy-Schwarz inequality, we can see that

$$\begin{aligned} \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \left(e_{it} e_{is} - \mathbb{E}[e_{it} e_{is}] \right) F_t' F_s &\leq \frac{1}{NT} \sum_{t=1}^{T} \sum_{s=1}^{T} \left| \sum_{i=1}^{N} \left(e_{it} e_{is} - \mathbb{E}[e_{it} e_{is}] \right) \right| \left| F_t' F_s \right| \\ &\leq \frac{1}{NT^2} \left[\sum_{t=1}^{T} \sum_{s=1}^{T} \left| \sum_{i=1}^{N} \left(e_{it} e_{is} - \mathbb{E}[e_{it} e_{is}] \right) \right|^2 \right]^{\frac{1}{2}} \cdot \left[\sum_{t=1}^{T} \sum_{s=1}^{T} \left| F_t' F_s \right|^2 \right]^{\frac{1}{2}} \\ &= \frac{1}{\sqrt{N}} \left(\frac{1}{T} \sum_{t=1}^{T} |F_t|^2 \right) \left[\frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \left| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(e_{it} e_{is} - \mathbb{E}[e_{it} e_{is}] \right) \right|^2 \right]^{\frac{1}{2}} \\ &= \frac{r}{\sqrt{N}} \left[\frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \left| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(e_{it} e_{is} - \mathbb{E}[e_{it} e_{is}] \right) \right|^2 \right]^{\frac{1}{2}}, \end{aligned}$$

since

$$\frac{1}{T} \sum_{t=1}^{T} |F_t|^2 = \operatorname{tr} \left(\frac{F'F}{T} \right) = r.$$

We will show below that there exists an $M < +\infty$ such that

$$\mathbb{E} \left| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(e_{it} e_{is} - \mathbb{E}[e_{it} e_{is}] \right) \right|^2 < M$$

for any $t, s \in N_+$; it follows that

$$\mathbb{E}\left[\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(e_{it}e_{is}-\mathbb{E}\left[e_{it}e_{is}\right])\right|^2\right] < M$$

as well, which implies that

$$\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(e_{it}e_{is}-\mathbb{E}\left[e_{it}e_{is}\right]\right)F_t'F_s \leq \frac{r}{\sqrt{N}}O_p(1).$$

The second term can be written as

$$\sigma^2 \cdot \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}|F_t|^2 = \sigma^2 \cdot \frac{r}{T}.$$

Taken together, we have

$$\frac{1}{N}\sum_{i=1}^{N}\left|\frac{1}{T}\sum_{t=1}^{T}F_t e_{it}\right|^2 \leq \sigma^2 \cdot \frac{r}{T} + \frac{r}{\sqrt{N}} \cdot \left[\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(e_{it}e_{is}-\mathbb{E}\left[e_{it}e_{is}\right])\right|^2\right]^{\frac{1}{2}},$$

where the last term is $O_p(1)$. Since none of the terms on the right hand side depend on F, it follows that

$$\sup_{F\in\mathcal{F}}\left[\frac{1}{N}\sum_{i=1}^{N}\left|\frac{1}{T}\sum_{t=1}^{T}F_t e_{it}\right|^2\right] \leq \sigma^2 \cdot \frac{r}{T} + \frac{r}{\sqrt{N}} \cdot \left[\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(e_{it}e_{is}-\mathbb{E}\left[e_{it}e_{is}\right])\right|^2\right]^{\frac{1}{2}},$$

and as such

$$\sup_{F\in\mathcal{F}}\left[\frac{1}{N}\sum_{i=1}^{N}\left|\frac{1}{T}\sum_{t=1}^{T}F_t e_{it}\right|^2\right] = O_p\left(\frac{1}{\min(\sqrt{N}, T)}\right).$$

To show that $\frac{1}{N}\sum_{i=1}^{N}\left|\frac{1}{T}\sum_{t=1}^{T}F_t^0 e_{it}\right|^2 = o_p(1)$, we need only note that

$$\frac{1}{N}\sum_{i=1}^{N}\left|\frac{1}{T}\sum_{t=1}^{T}F_t^0 e_{it}\right|^2 \leq \frac{1}{\sqrt{N}}\operatorname{tr}\left(\frac{F^{0\prime}F^0}{T}\right)\left[\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(e_{it}e_{is}-\mathbb{E}\left[e_{it}e_{is}\right])\right|^2\right]^{\frac{1}{2}}$$
$$+\sigma^2\frac{1}{T}\operatorname{tr}\left(\frac{F^{0\prime}F^0}{T}\right).$$

- **The Rate of Convergence of the Product of Common Component and Errors**

We now investigate the rate of convergence of

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_i^{0\prime}F_t^0 e_{it}.$$

To do so, recall that we assumed $\{|F_t^0| \mid t \in N_+\}$ and $\{|\lambda_i^0| \mid i \in N_+\}$ are L^2-bounded, so that there exists an $M < +\infty$ such that

$$\sup_{t\in N_+}\mathbb{E}\left|F_t^0\right|^2, \sup_{i\in N_+}\mathbb{E}\left|\lambda_i^0\right|^2 < M.$$

It then follows that

$$\begin{aligned}
\mathbb{E}\left|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_i^{0\prime}F_t^0 e_{it}\right|^2 &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}\left[e_{it}e_{js}\lambda_i^{0\prime}F_t^0F_s^{0\prime}\lambda_j\right] \\
&= \sigma^2\cdot\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}\left[\lambda_i^{0\prime}F_t^0F_t^{0\prime}\lambda_i\right] \\
&\le \sigma^2\cdot\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathbb{E}\left|F_t^{0\prime}\right|^2\right)^{\frac{1}{2}}\left(\mathbb{E}\left|\lambda_i^0\right|^2\right)^{\frac{1}{2}} \quad \text{(Hölder's inequality)} \\
&\le \sigma^2 M,
\end{aligned}$$

so that

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_i^{0\prime}F_t^0 e_{it} = O_p(1).$$

- **Rate of Convergnece of the Product of Factor Loadings and Errors**

Here we study the rate of convergence of

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left\|e_i\lambda_i^{0\prime}\right\|.$$

To do so, let λ_{ik}^0 be the kth element of λ_i^0 for $1 \leq k \leq r$, so that

$$\left\|e_i\lambda_i^{0\prime}\right\| = \left\|\begin{pmatrix} e_{i1}\lambda_{i1}^0 & \cdots & e_{i1}\lambda_{ir}^0 \\ \vdots & \ddots & \vdots \\ e_{iT}\lambda_{i1}^0 & \cdots & e_{iT}\lambda_{ir}^0 \end{pmatrix}\right\| \leq \sum_{k=1}^{r}\sum_{t=1}^{T}\left|e_{it}\lambda_{ik}^0\right|.$$

It follows that

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left\|e_i\lambda_i^{0\prime}\right\| \leq \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{k=1}^{r}\sum_{t=1}^{T}\left|e_{it}\lambda_{ik}^0\right|.$$

Recall that

$$\sup_{i\in N_+}\mathbb{E}\left|\lambda_i^0\right|^2 < M$$

for some $M < +\infty$. Now we have

$$\begin{aligned}
\mathbb{E}\left|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{k=1}^{r}\sum_{t=1}^{T}\left|e_{it}\lambda_{ik}^0\right|\right|^2 &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{r}\sum_{l=1}^{r}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}\left[e_{it}e_{js}\lambda_{ik}^0\lambda_{jl}^0\right] \\
&= \sigma^2\cdot\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{k=1}^{r}\sum_{l=1}^{r}\mathbb{E}\left[\lambda_{ik}^0\lambda_{il}^0\right] \\
&\leq \sigma^2\cdot\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{k=1}^{r}\sum_{l=1}^{r}\mathbb{E}\left|\lambda_i^0\right|^2 \qquad \text{(Hölder's inequality)} \\
&\leq \sigma^2 M\cdot\frac{1}{NT}NT\cdot r^2 = \sigma^2 M r^2,
\end{aligned}$$

which implies that

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{k=1}^{r}\sum_{t=1}^{T}\left|e_{it}\lambda_{ik}^0\right| = O_p(1)$$

and therefore that

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left\|e_i\lambda_i^{0\prime}\right\| = O_p(1).$$

- **Rate of Convergence of Squared Errors**

  We will investigate the rate of convergence of

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}^{2};$$

  since

$$\mathbb{E}\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}^{2}\right| = \sigma^2,$$

  $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}^{2} = O_p(1)$.

- **The Rate of Convergence of Error Cross Products**
  For any $t, s \in N_+$, note that

$$\begin{aligned}\mathbb{E}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(e_{it}e_{is} - \mathbb{E}\left[e_{it}e_{is}\right])\right|^2 &= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\text{Cov}\left(e_{it}e_{is}, e_{jt}e_{js}\right) \\ &= \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[(e_{it}e_{is} - \mathbb{E}\left[e_{it}e_{is}\right])^2\right] \\ &= \begin{cases}\sigma^4 & \text{if } t \neq s \\ \mu_4 - \sigma^4 & \text{if } t = s\end{cases} < +\infty.\end{aligned}$$

  Therefore, there exists an $M < +\infty$ such that

$$\mathbb{E}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(e_{it}e_{is} - \mathbb{E}\left[e_{it}e_{is}\right])\right|^2 < M$$

  for any $t, s \in N_+$.

- **The Rates of Convergence of the Covariance of Error Cross Products**

  Choose any $t, s \in N_+$ and $i, j, k, l \in N_+$. Then, we can consider the following cases:

  (1) $t = s$

  In this case,

$$\begin{aligned}\mathrm{Cov}\left(e_{it}e_{jt}, e_{ks}, e_{ls}\right) &= \mathbb{E}\left[e_{it}e_{jt}e_{kt}e_{lt}\right] - \mathbb{E}\left[e_{it}e_{jt}\right] \cdot \mathbb{E}\left[e_{kt}e_{lt}\right] \\ &= \begin{cases} \mu_4 - \sigma^4 & \text{if } i = j = k = l \\ \sigma^4 & \text{if } i = k \neq j = l \text{ or } i = l \neq j = k \\ 0 & \text{otherwise} \end{cases} \cdot\end{aligned}$$

  (2) $t \neq s$

  In this case,

$$\mathrm{Cov}\left(e_{it}e_{jt}, e_{ks}, e_{ls}\right) = \mathbb{E}\left[e_{it}e_{jt}\right] \cdot \mathbb{E}\left[e_{ks}e_{ls}\right] - \mathbb{E}\left[e_{it}e_{jt}\right] \cdot \mathbb{E}\left[e_{ks}e_{ls}\right] = 0$$

  for any i, j, k, l.

  Therefore, we can see that

$$\sum_{1 \le t,s \le T} \sum_{1 \le i,j,k,l \le N} \left|\mathrm{Cov}\left(e_{it}e_{jt}, e_{ks}, e_{ls}\right)\right| \le T(N|\mu_4 - \sigma_4| + 2N(N-1)\sigma^4).$$

  By implication,

$$\frac{1}{TN^2} \sum_{1 \le t,s \le T} \sum_{1 \le i,j,k,l \le N} \left|\mathrm{Cov}\left(e_{it}e_{jt}, e_{ks}, e_{ls}\right)\right| \to 2\sigma^4$$

  as $N, T \to \infty$, meaning that the sequence

$$\{\frac{1}{TN^2} \sum_{1 \le t,s \le T} \sum_{1 \le i,j,k,l \le N} \left|\mathrm{Cov}\left(e_{it}e_{jt}, e_{ks}, e_{ls}\right)\right|\}_{N,T \in N_+}$$

  is bounded.

  By a symmetric argument, it follows that the sequence

$$\{\frac{1}{NT^2} \sum_{1 \le i,j \le N} \sum_{1 \le t,s,u,v \le T} \left|\mathrm{Cov}\left(e_{it}e_{is}, e_{ju}e_{jv}\right)\right|\}_{N,T \in N_+}$$

  is also bounded.

4.4 Consistency of the Least Squares Estimators of β, F

4.4.1 Preliminary Results

As before, let $\mathcal{F}$ be the set of all $T \times r$ matrices F such that $\frac{F'F}{T} = I_r$. Then, we can show the following:

- $\sup_{F\in\mathcal{F}} \left| \frac{1}{NT} \sum_{i=1}^N X_i' M_F e_i \right| = o_p(1)$

Note that, for any $F \in \mathcal{F}$ with rows $F_1, \cdots, F_T$, we can decompose the above expression as

$$\frac{1}{NT}\sum_{i=1}^N X_i' M_F e_i = \frac{1}{NT}\sum_{i=1}^N X_i' e_i - \frac{1}{NT^2}\sum_{i=1}^N X_i' F F' e_i.$$

where we used the fact that

$$F(F'F)^{-1}F' = \frac{1}{T}FF'.$$

Focusing on the second term, we have

$$\begin{aligned}
\left| \frac{1}{NT^2}\sum_{i=1}^N X_i' F F' e_i \right| &\leq \frac{1}{NT^2}\sum_{i=1}^N \|X_i' F\| \cdot \left| \sum_{t=1}^T F_t e_{it} \right| \\
&\leq \|F\| \frac{1}{NT^2} \left(\sum_{i=1}^N \|X_i\|^2 \right)^{\frac{1}{2}} \cdot \left(\sum_{i=1}^N \left| \sum_{t=1}^T F_t e_{it} \right|^2 \right)^{\frac{1}{2}} \\
&= \frac{1}{\sqrt{T}}\|F\| \cdot \left(\frac{1}{NT}\sum_{i=1}^N \|X_i\|^2 \right)^{\frac{1}{2}} \cdot \left(\frac{1}{N}\sum_{i=1}^N \left| \frac{1}{T}\sum_{t=1}^T F_t e_{it} \right|^2 \right)^{\frac{1}{2}} \\
&\leq \sqrt{r} \left(\frac{1}{NT}\sum_{i=1}^N \|X_i\|^2 \right)^{\frac{1}{2}} \cdot \left(\frac{1}{N}\sum_{i=1}^N \left| \frac{1}{T}\sum_{t=1}^T F_t e_{it} \right|^2 \right)^{\frac{1}{2}},
\end{aligned}$$

where we used the Cauchy-Schwarz inequality and the fact that

$$\frac{1}{\sqrt{T}}\|F\| \leq \left[\mathrm{tr}\left(\frac{F'F}{T} \right) \right]^{\frac{1}{2}} = \sqrt{r}.$$

Therefore,

$$\begin{aligned}
\sup_{F\in\mathcal{F}} \left| \frac{1}{NT}\sum_{i=1}^N X_i' M_F e_i \right| &\leq \left| \frac{1}{NT}\sum_{i=1}^N X_i' e_i \right| \\
&\quad + \sqrt{r} \left(\frac{1}{NT}\sum_{i=1}^N \|X_i\|^2 \right)^{\frac{1}{2}} \cdot \left[\sup_{F\in\mathcal{F}} \frac{1}{N}\sum_{i=1}^N \left| \frac{1}{T}\sum_{t=1}^T F_t e_{it} \right|^2 \right]^{\frac{1}{2}}.
\end{aligned}$$

We proved above that

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} X_i' M_F e_i = O_p(1),$$

$$\frac{1}{NT}\sum_{i=1}^{N} \|X_i\|^2 = O_p(1)$$

$$\sup_{F\in\mathcal{F}} \frac{1}{N}\sum_{i=1}^{N} \left| \frac{1}{T}\sum_{t=1}^{T} F_t e_{it} \right|^2 = O_p\left(\frac{1}{\min(\sqrt{N},T)}\right),$$

so

$$\sup_{F\in\mathcal{F}} \left| \frac{1}{NT}\sum_{i=1}^{N} X_i' M_F e_i \right| = o_p(1).$$

- $\sup_{F\in\mathcal{F}}\left|\frac{1}{NT}\sum_{i=1}^{N}\lambda_i{}^0\prime F^{0\prime}M_F e_i\right| = o_p(1)$

  As above, note that

$$\frac{1}{NT}\sum_{i=1}^{N}\lambda_i^{0\prime}F^{0\prime}M_F e_i = \frac{1}{NT}\sum_{i=1}^{N}\lambda_i' F^{0\prime}e_i - \frac{1}{NT^2}\sum_{i=1}^{N}\lambda_i' F^{0\prime}FF'e_i$$

  for any $F\in\mathcal{F}$ with rows $F_1,\cdots,F_T$.
  The second term can further be majorized as

$$\begin{aligned}\left|\frac{1}{NT^2}\sum_{i=1}^{N}\lambda_i^{0\prime}F^{0\prime}FF'e_i\right| &\le \left\|F^{0\prime}F\right\|\frac{1}{NT^2}\sum_{i=1}^{N}\left|\lambda_i^0\right|\cdot\left|\sum_{t=1}^{T}F_t e_{it}\right| \\ &\le \frac{1}{T}\left\|F^0\right\|\|F\|\cdot\left(\frac{1}{N}\sum_{i=1}^{N}\left|\lambda_i^0\right|^2\right)^{\frac{1}{2}}\left(\frac{1}{N}\sum_{i=1}^{N}\left|\sum_{t=1}^{T}F_t e_{it}\right|^2\right)^{\frac{1}{2}} \\ &\le r\cdot\mathrm{tr}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)\cdot\left(\frac{1}{N}\sum_{i=1}^{N}\left|\sum_{t=1}^{T}F_t e_{it}\right|^2\right)^{\frac{1}{2}}.\end{aligned}$$

  Therefore, we have

$$\sup_{F\in\mathcal{F}}\left|\frac{1}{NT}\sum_{i=1}^{N}\lambda_i{}^0\prime F^{0\prime}M_F e_i\right| \le \left|\frac{1}{NT}\sum_{i=1}^{N}\lambda_i' F^{0\prime}e_i\right| + r\cdot\mathrm{tr}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)\cdot\left[\sup_{F\in\mathcal{F}}\frac{1}{N}\sum_{i=1}^{N}\left|\sum_{t=1}^{T}F_t e_{it}\right|^2\right]^{\frac{1}{2}}.$$

  We assumed and proved above that

$$\begin{aligned}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\lambda_i' F^{0\prime}e_i &= O_p(1) \\ \frac{\Lambda^{0\prime}\Lambda^0}{N} &= O_p(1) \\ \sup_{F\in\mathcal{F}}\frac{1}{N}\sum_{i=1}^{N}\left|\sum_{t=1}^{T}F_t e_{it}\right|^2 &= O_p\left(\frac{1}{\min(\sqrt{N},T)}\right),\end{aligned}$$

  so we can conclude that

$$\sup_{F\in\mathcal{F}}\left|\frac{1}{NT}\sum_{i=1}^{N}\lambda_i{}^0\prime F^{0\prime}M_F e_i\right| = o_p(1).$$

- $\sup_{F\in\mathcal{F}}\left|\frac{1}{NT}\sum_{i=1}^{N}e_i'P_Fe_i\right|$

  As is obvious by now, we first choose any $F\in\mathcal{F}$ with rows $F_1,\cdots,F_T$ and note that

$$\frac{1}{NT}\sum_{i=1}^{N}e_i'P_Fe_i=\frac{1}{NT^2}\sum_{i=1}^{N}e_i'FF'e_i.$$

  This term can be simply majorized as

$$\begin{aligned}\left|\frac{1}{NT^2}\sum_{i=1}^{N}e_i'FF'e_i\right| &\leq \frac{1}{NT^2}\sum_{i=1}^{N}|F'e_i|^2\\ &=\frac{1}{N}\sum_{i=1}^{N}\left|\frac{1}{T}\sum_{t=1}^{T}F_te_{it}\right|^2,\end{aligned}$$

  so

$$\sup_{F\in\mathcal{F}}\left|\frac{1}{NT}\sum_{i=1}^{N}e_i'P_Fe_i\right|\leq\sup_{F\in\mathcal{F}}\frac{1}{N}\sum_{i=1}^{N}\left|\frac{1}{T}\sum_{t=1}^{T}F_te_{it}\right|^2,$$

  and because the term on the right is $o_p(1)$, so is

$$\sup_{F\in\mathcal{F}}\left|\frac{1}{NT}\sum_{i=1}^{N}e_i'P_Fe_i\right|.$$

4.4.2 Convergence of the Objective Function

Let $\mathbb{B} \subset \mathbb{R}^k$ be the parameter space for β.

Recall that $(\hat{\beta}, \hat{F})$ is a point in the parameter space $\mathbb{B} \times \mathcal{F}$ that minimizes the objective function

$$V(\beta, F) = \frac{1}{NT} \sum_{i=1}^{N} (Y_i - X_i\beta)' M_F (Y_i - X_i\beta).$$

Since $\frac{1}{NT} \sum_{i=1}^{N} e_i' M_{F^0} e_i$ is a term that does not depend on β, F, we can also say that $(\hat{\beta}, \hat{F})$ minimizes the function

$$\begin{aligned} S(\beta, F) &= V(\beta, F) - \frac{1}{NT} \sum_{i=1}^{N} e_i' M_{F^0} e_i \\ &= \frac{1}{NT} \sum_{i=1}^{N} (Y_i - X_i\beta)' M_F (Y_i - X_i\beta) - \frac{1}{NT} \sum_{i=1}^{N} e_i' M_{F^0} e_i. \end{aligned}$$

We call $S(\beta, F)$ the "centered" objective function.

For any $1 \le i \le N$,

$$Y_i = X_i \beta^0 + F^0 \lambda_i^0 + e_i,$$

so we can decompose the centered SSR $S(\beta, F)$ as

$$\begin{aligned} S(\beta, F) &= \frac{1}{NT} \sum_{i=1}^{N} (Y_i - X_i\beta)' M_F (Y_i - X_i\beta) - \frac{1}{NT} \sum_{i=1}^{N} e_i' M_{F^0} e_i \\ &= \frac{1}{NT} \sum_{i=1}^{N} \left[X_i(\beta^0 - \beta) + F^0\lambda_i^0 + e_i\right]' M_F \left[X_i(\beta^0 - \beta) + F^0\lambda_i^0 + e_i\right] - \frac{1}{NT} \sum_{i=1}^{N} e_i' M_{F^0} e_i \\ &= (\beta - \beta^0)' \frac{1}{NT} \sum_{i=1}^{N} X_i' M_F X_i (\beta - \beta^0) + 2(\beta - \beta^0)' \frac{1}{NT} \sum_{i=1}^{N} X_i' M_F F^0 \lambda_i^0 + 2(\beta - \beta^0)' \frac{1}{NT} \sum_{i=1}^{N} X_i' M_F e_i \\ &\quad + \frac{1}{NT} \sum_{i=1}^{N} \lambda_i^{0\prime} F^{0\prime} M_F F^0 \lambda_i^0 + 2 \frac{1}{NT} \sum_{i=1}^{N} \lambda_i^{0\prime} F^{0\prime} M_F e_i + \frac{1}{NT} e_i' (M_F - M_{F^0}) e_i \end{aligned}$$

for any $\beta \in \mathbb{B}$ and $F \in \mathcal{F}$. Note that we can write

$$\begin{aligned} \frac{1}{NT} \sum_{i=1}^{N} \lambda_i^{0\prime} F^{0\prime} M_F F^0 \lambda_i^0 &= \operatorname{tr}\left(\left[\frac{1}{NT} \sum_{i=1}^{N} \lambda_i^0 \lambda_i^{0\prime} \right] F^{0\prime} M_F F^0 \right) \\ &= \operatorname{tr}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N} \frac{F^{0\prime} M_F F^0}{T} \right). \end{aligned}$$

Defining

$$\tilde{S}(\beta, F) = (\beta - \beta^0)' \frac{1}{NT} \sum_{i=1}^{N} X_i' M_F X_i (\beta - \beta^0) + \operatorname{tr}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N} \frac{F^{0\prime} M_F F^0}{T} \right) + 2(\beta - \beta^0)' \frac{1}{NT} \sum_{i=1}^{N} X_i' M_F F^0 \lambda_i^0,$$

$S(\beta, F)$ can now be expressed as

$$S(\beta, F) = \tilde{S}(\beta, F) + 2(\beta - \beta^0)' \frac{1}{NT} \sum_{i=1}^{N} X_i' M_F e_i$$
$$+ 2 \frac{1}{NT} \sum_{i=1}^{N} \lambda_i^{0\prime} F^{0\prime} M_F e_i + \frac{1}{NT} e_i' (P_{F^0} - P_F) e_i.$$

By assumption, the parameter space $\mathbb{B}$ of β is bounded; that is, there exists an $M < +\infty$ such that, for any $\beta \in \mathbb{B}$, $|\beta| < M$.

From the preliminary results, we can see that

$$\sup_{(\beta,F)\in\mathbb{B}\times\mathcal{F}} \left| (\beta - \beta^0)' \frac{1}{NT} \sum_{i=1}^{N} X_i' M_F e_i \right| \leq 2M \cdot \left[\sup_{F\in\mathcal{F}} \left| \frac{1}{NT} \sum_{i=1}^{N} X_i' M_F e_i \right| \right] = o_p(1),$$
$$\sup_{(\beta,F)\in\mathbb{B}\times\mathcal{F}} \left| \frac{1}{NT} \sum_{i=1}^{N} \lambda_i^{0\prime} F^{0\prime} M_F e_i \right| = \sup_{F\in\mathcal{F}} \left| \frac{1}{NT} \sum_{i=1}^{N} \lambda_i^{0\prime} F^{0\prime} M_F e_i \right| = o_p(1)$$
$$\sup_{(\beta,F)\in\mathbb{B}\times\mathcal{F}} \left| \frac{1}{NT} e_i' (P_{F^0} - P_F) e_i \right| \leq \sup_{F\in\mathcal{F}} \left| \frac{1}{NT} \sum_{i=1}^{N} e_i' P_F e_i \right| + \left| \frac{1}{NT} \sum_{i=1}^{N} e_i' P_{F^0} e_i \right| = o_p(1),$$

where

$$\left| \frac{1}{NT} \sum_{i=1}^{N} e_i' P_{F^0} e_i \right| = o_p(1)$$

because

$$\left| \frac{1}{NT} \sum_{i=1}^{N} e_i' P_{F^0} e_i \right| \leq \left\| \left(\frac{F^{0\prime} F^0}{T} \right)^{-1} \right\| \cdot \left(\frac{1}{NT^2} \sum_{i=1}^{N} \left| F^{0\prime} e_i \right|^2 \right)$$
$$= \left\| \left(\frac{F^{0\prime} F^0}{T} \right)^{-1} \right\| \cdot \left(\frac{1}{N} \sum_{i=1}^{N} \left| \frac{1}{T} \sum_{t=1}^{T} F_t^0 e_{it} \right|^2 \right),$$

where the first term is $O_p(1)$ and the latter $o_p(1)$.
Therefore,

$$\sup_{(\beta,F)\in\mathbb{B}\times\mathcal{F}} \left| S(\beta, F) - \tilde{S}(\beta, F) \right| = o_p(1).$$

4.4.3 The Identification Condition for β and F^0

We now show that the uniform convergence of the objective function defined above implies that $\hat{\beta}$ converges to β. As with the usual consistency result for extremum estimators, we must first show that the true parameters are the unique minimizers of the objective function, in this case $\tilde{S}(\beta, F)$.

It is clear that, for any $r \times r$ nonsingular matrix valued H, because

$$F^0 H (H' F^{0\prime} F^0 H)^{-1} H' F^{0\prime} = F^0 (F^{0\prime} F^0)^{-1} F^{0\prime},$$

we have $M_{F^0} = M_{F^0 H}$ and thus

$$\tilde{S}(\beta^0, F^0 H) = \operatorname{tr}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N} \cdot \frac{F^{0\prime} M_{F^0} F^0}{T} \right) = 0,$$

since $M_{F^0} F^0 = O$.

We can also show the converse, namely that

$$\tilde{S}(\beta, F) = 0$$

implies that $\beta = \beta^0$ and $F = F^0 H$ for some nonsingular $r \times r$ random matrix H.
To this end, note that

$$\begin{aligned}
\operatorname{tr}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N} \cdot \frac{F^{0\prime} M_F F^0}{T} \right) &= \frac{1}{NT} \sum_{i=1}^{N} \lambda_i^{0\prime} F^{0\prime} M_F F^0 \lambda_i^0 \\
&= \frac{1}{NT} \sum_{i=1}^{N} \operatorname{vec}\left(M_F F^0 \lambda_i^0 \right)' \operatorname{vec}\left(M_F F^0 \lambda_i^0 \right) \\
&= \frac{1}{NT} \sum_{i=1}^{N} \operatorname{vec}\left(M_F F^0 \right)' \left(\lambda_i^0 \bigotimes I_T \right) \left(\lambda_i^{0\prime} \bigotimes I_T \right) \operatorname{vec}\left(M_F F^0 \right) \\
&= \operatorname{vec}\left(M_F F^0 \right)' \left[\frac{1}{NT} \sum_{i=1}^{N} \left(\lambda_i^0 \lambda_i^{0\prime} \bigotimes I_T \right) \right] \operatorname{vec}\left(M_F F^0 \right) \\
&= \operatorname{vec}\left(M_F F^0 \right)' \left(\frac{\Lambda^{0\prime}\Lambda^0}{NT} \bigotimes I_T \right) \operatorname{vec}\left(M_F F^0 \right)
\end{aligned}$$

and

$$\begin{aligned}
\frac{1}{NT} \sum_{i=1}^{N} X_i' M_F F^0 \lambda_i^0 &= \frac{1}{NT} \sum_{i=1}^{N} \operatorname{vec}\left(X_i' M_F' \cdot M_F F^0 \lambda_i^0 \right) \\
&= \left[\frac{1}{NT} \sum_{i=1}^{N} \left(\lambda_i^{0\prime} \bigotimes X_i' M_F' \right) \right] \operatorname{vec}\left(M_F F^0 \right),
\end{aligned}$$

defining

$$\begin{aligned}
A &= \frac{1}{NT}\sum_{i=1}^{N} X_i' M_F X_i \\
B &= \frac{\Lambda^{0\prime}\Lambda^0}{NT} \bigotimes I_T \\
C &= \frac{1}{NT}\sum_{i=1}^{N}\left(\lambda_i^0 \bigotimes M_F X_i\right) \\
\eta &= \operatorname{vec}\left(M_F F^0\right),
\end{aligned}$$

we can express $\tilde{S}(\beta, F)$ as

$$\tilde{S}(\beta,F) = (\beta-\beta^0)' A(\beta-\beta^0) + \eta' B\eta + 2(\beta-\beta^0)' C'\eta.$$

The first two terms are immediately recognizable as squares of $(\beta-\beta^0)$ and η, so letting X and Y satisfy

$$\begin{aligned}
\tilde{S}(\beta,F) &= [\eta+X]' B\,[\eta+X] + (\beta-\beta^0)'(A+Y)(\beta-\beta^0) \\
&= (\beta-\beta^0)' A(\beta-\beta^0) + \eta' B\eta + 2(\beta-\beta^0)' C'\eta,
\end{aligned}$$

we can see that $X' B\eta = (\beta-\beta^0)' C'\eta$, so that

$$X = B^{-1}C(\beta-\beta^0),$$

and since

$$X' B X + (\beta-\beta^0)' Y(\beta-\beta^0) = (\beta-\beta^0)'\left[C' B^{-1} C + Y\right](\beta-\beta^0) = 0,$$

we finally have

$$\tilde{S}(\beta,F) = \left[\eta + B^{-1}C(\beta-\beta^0)\right]' B\left[\eta + B^{-1}C(\beta-\beta^0)\right] + (\beta-\beta^0)'(A - C' B^{-1} C)(\beta-\beta^0).$$

Expanding terms, $A - C' B^{-1} C$ is revealed to be

$$\begin{aligned}
A - C' B^{-1} C &= \frac{1}{NT}\sum_{i=1}^{N} X_i' M_F X_i - T\left[\frac{1}{NT}\sum_{i=1}^{N}\left(\lambda_i^0 \bigotimes M_F X_i\right)\right]'\left[\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1} \bigotimes I_T\right]\left[\frac{1}{NT}\sum_{i=1}^{N}\left(\lambda_i^0 \bigotimes M_F X_i\right)\right] \\
&= \frac{1}{NT}\sum_{i=1}^{N} X_i' M_F X_i - \frac{1}{N^2 T}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\lambda_i^{0\prime}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1}\lambda_j^0 \bigotimes X_i' M_F X_i\right) \\
&= \frac{1}{NT}\sum_{i=1}^{N} X_i' M_F X_i - \frac{1}{N^2 T}\sum_{i=1}^{N}\sum_{j=1}^{N} X_i' M_F X_i a_{ij} \\
&= D(F),
\end{aligned}$$

where $a_{ij} = \lambda_i^{0\prime} \left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1} \lambda_j^0$ as defined in assumption 1.
Therefore,

$$\tilde{S}(\beta,F) = (\beta-\beta^0)' D(F)(\beta-\beta^0) + \left[\eta + B^{-1}C(\beta-\beta^0)\right]' B \left[\eta + B^{-1}C(\beta-\beta^0)\right],$$

and because $D(F)$ and B are both positive definite for any $F \in \mathcal{F}'$,

$$\tilde{S}(\beta,F) \geq 0,$$

where equality holds if and only if $\beta = \beta^0$ and $\eta = B^{-1}C(\beta-\beta^0) = \mathbf{0}$.

Suppose that $\tilde{S}(\beta,F) = 0$. Then, by the statement shown above, $\beta = \beta^0$. This implies that $\eta = \mathbf{0}$ and thus

$$M_F F^0 = F^0 - F(F'F)^{-1}F'F^0 = O;$$

because $\frac{F'F}{T} = I_r$, F is of full rank r and thus $F'F^0$ is nonsingular. This implies that

$$F = F^0(F'F^0)^{-1}F'F = F^0H,$$

where $H = (F'F^0)^{-1}F'F$ is a nonsingular $r \times r$ random matrix.

We have thus shown that

$$\tilde{S}(\beta,F) \geq 0$$

for any $(\beta,F) \in \mathbb{B} \times \mathcal{F}'$, and that

$$\tilde{S}(\beta,F) = 0$$

if and only if $\beta = \beta^0$ and there exists a nonsingular $r \times r$ random matrix such that $F = F^0H$. This means that $\tilde{S}(\beta,F)$ is minimized precisely at β^0 and rotations F^0H of the true factors F^0. This proves that $\tilde{S}(\beta,F)$ satisfies the identification condition for extremum estimation.

4.4.4 Consistency of $\hat{\beta}$

Given the uniform convergence of the centered objective function and the identification condition above, we can prove that $\hat{\beta}$ is consistent for β^0.

We currently have the following conditions:

- **(Extremum Estimator)** $(\hat{\beta}, \hat{F})$ is a minimizer of $S(\beta, F)$ on $\mathbb{B} \times \mathcal{F}'$

- **(Identification Condition)** For any $r \times r$ nonsingular H, $(\beta^0, F^0 H)$ is the unique minimizer of $\tilde{S}(\beta, F)$ on $\mathbb{B} \times \mathcal{F}'$

- **(Convergence of Centered Objective Function)** $S(\beta, F)$ converges uniformly in probability to $\tilde{S}(\beta, F)$ on $\mathbb{B} \times \mathcal{F}$, that is,

$$\sup_{\beta \in \mathbb{B}, F \in \mathcal{F}} \left| S(\beta, F) - \tilde{S}(\beta, F) \right| = o_p(1).$$

We now utilize all three conditions above to show that $\hat{\beta}$ is consistent for β^0. We proceed in steps to emphasize the role of each condition in the proof:

Step 1: Using the Identification Condition

Choose any $\delta > 0$, and suppose that

$$\left| \hat{\beta} - \beta^0 \right| > \delta.$$

for some $\beta \in \mathbb{B}$. Then, because $\hat{\beta} \neq \beta^0$ and $\tilde{S}(\beta, F)$ is uniquely minimized at $\beta^0, F^0 H$, we can see that

$$\tilde{S}(\hat{\beta}, \hat{F}) > \tilde{S}(\beta^0, F^0) = 0.$$

To procure a specific positive lower bound for $\tilde{S}(\hat{\beta}, \hat{F})$ that depends on δ, we note that

$$\begin{aligned} \tilde{S}(\hat{\beta}, \hat{F}) &\geq (\hat{\beta} - \beta^0)' D(\hat{F}) (\hat{\beta} - \beta^0) \\ &= \left| \hat{\beta} - \beta^0 \right|^2 \cdot \left(\frac{\hat{\beta} - \beta^0}{\left| \hat{\beta} - \beta^0 \right|} \right)' D(\hat{F}) \left(\frac{\hat{\beta} - \beta^0}{\left| \hat{\beta} - \beta^0 \right|} \right) \\ &\geq \left| \hat{\beta} - \beta^0 \right|^2 \cdot \rho_{\min} > \delta^2 \cdot \rho_{\min} > 0, \end{aligned}$$

where $\rho_{\min} > 0$ was defined as the minimum possible eigenvalue of $D(F)$ for any $F \in \mathcal{F}'$.

Therefore, for any $\delta > 0$,

$$\{\left|\hat{\beta}-\beta^0\right| > \delta\} \subset \{\tilde{S}(\hat{\beta},\hat{F}) > \delta^2 \cdot \rho_{\min}\}.$$

Step 2: Using the Extremum Property of the Estimators

We can easily see that

$$\{\tilde{S}(\hat{\beta},\hat{F}) > \delta^2 \cdot \rho_{\min}\} \subset \left\{\tilde{S}(\hat{\beta},\hat{F}) - S(\hat{\beta},\hat{F}) > \frac{\delta^2 \cdot \rho_{\min}}{2}\right\} \cup \left\{S(\hat{\beta},\hat{F}) > \frac{\delta^2 \cdot \rho_{\min}}{2}\right\}.$$

Note that, because

$$\begin{aligned} S(\beta^0,F^0) &= \frac{1}{NT}\sum_{i=1}^{N}(Y_i - X_i\beta^0)' M_{F^0}(Y_i - X_i\beta^0) - \frac{1}{NT}\sum_{i=1}^{N} e_i' M_{F^0} e_i \\ &= \frac{1}{NT}\sum_{i=1}^{N}(F^0\lambda_i^0 + e_i)' M_{F^0}(F^0\lambda_i^0 + e_i) - \frac{1}{NT}\sum_{i=1}^{N} e_i' M_{F^0} e_i = 0 \end{aligned}$$

and $(\hat{\beta},\hat{F})$ minimizes $S(\beta,F)$ on $\mathbb{B}\times\mathcal{F}'$,

$$S(\hat{\beta},\hat{F}) \leq S(\beta^0,F^0) = 0,$$

meaning that

$$\left\{S(\hat{\beta},\hat{F}) > \frac{\delta^2 \cdot \rho_{\min}}{2}\right\} = \emptyset.$$

Step 3: Using the Uniform Convergence Result

On the other hand, because $\hat{\beta} \in \mathbb{B}$ and $\hat{F} \in \mathcal{F}$,

$$\left|S(\hat{\beta},\hat{F}) - \tilde{S}(\hat{\beta},\hat{F})\right| \leq \sup_{\beta\in\mathbb{B},F\in\mathcal{F}} \left|S(\beta,F) - \tilde{S}(\beta,F)\right|,$$

which implies that

$$\left|S(\hat{\beta},\hat{F}) - \tilde{S}(\hat{\beta},\hat{F})\right| = o_p(1).$$

By definition, this means that

$$\mathbb{P}\left(\left|\tilde{S}(\hat{\beta},\hat{F}) - S(\hat{\beta},\hat{F})\right| > \frac{\delta^2 \cdot \rho_{\min}}{2}\right) \to 0$$

as $N,T \to \infty$.

Putting all the results above together, we can see that

$$\begin{aligned}
\mathbb{P}\left(\left|\hat{\beta}-\beta^0\right|>\delta\right) &\leq \mathbb{P}\left(\tilde{S}(\hat{\beta},\hat{F})>\delta^2\cdot\rho_{\min}\right)\\
&\leq \mathbb{P}\left(\tilde{S}(\hat{\beta},\hat{F})-S(\hat{\beta},\hat{F})>\frac{\delta^2\cdot\rho_{\min}}{2}\right)+\mathbb{P}\left(S(\hat{\beta},\hat{F})>\frac{\delta^2\cdot\rho_{\min}}{2}\right)\\
&\leq \mathbb{P}\left(\tilde{S}(\hat{\beta},\hat{F})-S(\hat{\beta},\hat{F})>\frac{\delta^2\cdot\rho_{\min}}{2}\right)\to 0
\end{aligned}$$

as $N,T\to\infty$. This holds for any $\delta>0$, so by definition,

$$\hat{\beta}\xrightarrow{p}\beta^0.$$

4.4.5 Convergence of $P_{\hat{F}}$

As we have seen multiple times during our study of unilevel and multilevel factor models, because the dimension of $\hat{F}$ increases to infinity as $T \to \infty$, we cannot establish the consistency of $\hat{F}$ through traditional means. Here we establish the convergence of the norm of the difference of the projection matrices $P_{\hat{F}}$ and P_{F^0}; in the next section, we will establish the consistency of $\hat{F}$ akin to that proved in Bai and Ng (2002) and Bai (2003).

First recall that

$$S(\hat{\beta}, \hat{F}) \leq S(\beta^0, F^0) = 0,$$

which we showed above. We also saw that

$$S(\hat{\beta}, \hat{F}) - \tilde{S}(\hat{\beta}, \hat{F}) = o_p(1),$$

where

$$\tilde{S}(\hat{\beta}, \hat{F}) \geq \tilde{S}(\beta^0, F^0) = 0$$

by the identification condition. These three results imply that, for any $\varepsilon > 0$,

$$\begin{aligned}
\mathbb{P}\left(\left|\tilde{S}(\hat{\beta}, \hat{F})\right| > \varepsilon\right) &= \mathbb{P}\left(\tilde{S}(\hat{\beta}, \hat{F}) > \varepsilon\right) \\
&\leq \mathbb{P}\left(\tilde{S}(\hat{\beta}, \hat{F}) - S(\hat{\beta}, \hat{F}) > \frac{\varepsilon}{2}\right) + \mathbb{P}\left(S(\hat{\beta}, \hat{F}) > \frac{\varepsilon}{2}\right) \\
&\leq \mathbb{P}\left(\left|\tilde{S}(\hat{\beta}, \hat{F}) - S(\hat{\beta}, \hat{F})\right| > \frac{\varepsilon}{2}\right) \to 0
\end{aligned}$$

as $N, T \to \infty$. It follows that

$$\tilde{S}(\hat{\beta}, \hat{F}) = o_p(1).$$

By definition,

$$\begin{aligned}
\tilde{S}(\hat{\beta}, \hat{F}) = {} & (\hat{\beta} - \beta^0)' \left(\frac{1}{NT}\sum_{i=1}^{N} X_i' M_{\hat{F}} X_i\right)(\hat{\beta} - \beta^0) \\
& + \operatorname{tr}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\frac{F^{0\prime} M_{\hat{F}} F^0}{T}\right) + 2(\hat{\beta} - \beta^0)' \frac{1}{NT}\sum_{i=1}^{N} X_i' M_{\hat{F}} F^0 \lambda_i^0.
\end{aligned}$$

We can easily show that

$$\frac{1}{NT}\sum_{i=1}^{N} X_i' M_{\hat{F}} X_i \quad \text{and} \quad \frac{1}{NT}\sum_{i=1}^{N} X_i' M_{\hat{F}} F^0 \lambda_i^0$$

are $O_p(1)$; since $\hat{\beta}-\beta^0 = o_p(1)$ by the previous consistency result, we have

$$\begin{aligned}
\operatorname{tr}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\frac{F^{0\prime}M_{\hat{F}}F^0}{T}\right) &= \tilde{S}(\hat{\beta},\hat{F}) - (\hat{\beta}-\beta^0)'\left(\frac{1}{NT}\sum_{i=1}^{N} X_i' M_{\hat{F}} X_i\right)(\hat{\beta}-\beta^0) \\
&\quad - 2(\hat{\beta}-\beta^0)'\frac{1}{NT}\sum_{i=1}^{N} X_i' M_{\hat{F}} F^0 \lambda_i^0 \\
&= o_p(1).
\end{aligned}$$

The matrix $\frac{\Lambda^{0\prime}\Lambda^0}{N}$ converges to Σ_Λ, a nonsingular matrix, so it is $O_p(1)$ but not $o_p(1)$. Therefore, it must be the case that

$$\frac{F^{0\prime}M_{\hat{F}}F^0}{T} = \frac{F^{0\prime}F^0}{T} - \frac{F^{0\prime}\hat{F}\hat{F}F^0}{T^2} = o_p(1).$$

By assumption, $\frac{F^{0\prime}F^0}{T} \xrightarrow{p} \Sigma_F$, an $r \times r$ matrix of full rank, so the above implies that

$$\frac{F^{0\prime}\hat{F}\hat{F}F^0}{T^2} \xrightarrow{p} \Sigma_F$$

as well.

Now we have

$$\begin{aligned}
\left\|P_{\hat{F}} - P_{F^0}\right\|^2 &\le \operatorname{tr}\left((P_{\hat{F}} - P_{F^0})^2\right) \\
&= \operatorname{tr}\left(P_{\hat{F}} + P_{F^0}\right) - 2\operatorname{tr}\left(\frac{\hat{F}' P_{F^0}\hat{F}}{T}\right) \\
&= 2r - 2\operatorname{tr}\left(\frac{\hat{F}'F^0}{T}\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\frac{F^{0\prime}\hat{F}}{T}\right) \\
&= 2r - 2\operatorname{tr}\left(\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\frac{F^{0\prime}\hat{F}\hat{F}'F^0}{T^2}\right) \\
&\xrightarrow{p} 2r - 2\operatorname{tr}(\Sigma_F^{-1}\Sigma_F) = 0.
\end{aligned}$$

4.5 The Rate of Convergence of $\hat{\beta}$

In the last section we proved that the least squares estimator $\hat{\beta}$ is consistent for the true coefficients β^0, and that the projectio matrix $P_{\hat{F}}$ is consistent for P_{F^0}. We are now in a position to investigate the rate at which these estimators converge. The process is long and arduous, so we proceed in small steps.

4.5.1 The Convergence of V_{NT} and Consistency of F

Recall that the least squares estimators $\hat{\beta}$ and $\hat{F}$ are characterized by the equations

$$\left[\sum_{i=1}^{N} X_i' M_{\hat{F}} X_i\right]^{-1} \sum_{i=1}^{N} X_i' M_{\hat{F}} Y_i = \hat{\beta}$$

$$\left(\frac{1}{NT}\sum_{i=1}^{N}(Y_i - X_i\hat{\beta})(Y_i - X_i\hat{\beta})'\right)\hat{F} = \hat{F}V_{NT}$$

and $\frac{\hat{F}'\hat{F}}{T} = I_r$. Using the fact that

$$Y_i - X_i\hat{\beta} = X_i(\beta^0 - \hat{\beta}) + F^0\lambda_i^0 + e_i$$

for any $1 \le i \le N$, we can expand the second equation as follows:

$$\begin{aligned}
\hat{F}V_{NT} &= \left(\frac{1}{NT}\sum_{i=1}^{N}(Y_i - X_i\hat{\beta})(Y_i - X_i\hat{\beta})'\right)\hat{F} \\
&= \left[\frac{1}{NT}\sum_{i=1}^{N}(X_i(\beta^0-\hat{\beta}) + F^0\lambda_i^0 + e_i)(X_i(\beta^0-\hat{\beta}) + F^0\lambda_i^0 + e_i)'\right]\hat{F} \\
&= \underbrace{\frac{1}{NT}\sum_{i=1}^{N}X_i(\beta^0-\hat{\beta})(\beta^0-\hat{\beta})'X_i\hat{F}}_{I1} + \underbrace{\frac{1}{NT}\sum_{i=1}^{N}X_i(\beta^0-\hat{\beta})\lambda_i^{0\prime}F^{0\prime}\hat{F}}_{I2} + \underbrace{\frac{1}{NT}\sum_{i=1}^{N}X_i(\beta^0-\hat{\beta})e_i'\hat{F}}_{I3} \\
&\quad + \underbrace{\frac{1}{NT}\sum_{i=1}^{N}e_i(\beta^0-\hat{\beta})'X_i'\hat{F}}_{I4} + \underbrace{\frac{1}{NT}\sum_{i=1}^{N}e_i\lambda_i^{0\prime}F^{0\prime}\hat{F}}_{I5} + \underbrace{\frac{1}{NT}\sum_{i=1}^{N}e_ie_i'\hat{F}}_{I6} \\
&\quad + \underbrace{\frac{1}{NT}\sum_{i=1}^{N}F^0\lambda_i^0(\beta^0-\hat{\beta})'X_i'\hat{F}}_{I7} + \underbrace{\frac{1}{NT}\sum_{i=1}^{N}F^0\lambda_i^0e_i'\hat{F}}_{I8} + \frac{1}{NT}\sum_{i=1}^{N}F^0\lambda_i^0\lambda_i^{0\prime}F^{0\prime}\hat{F}.
\end{aligned}$$

The last term on the right hand side can be written as

$$\begin{aligned}
\frac{1}{NT}\sum_{i=1}^{N}F^0\lambda_i^0\lambda_i^{0\prime}F^{0\prime}\hat{F} &= F^0\left(\frac{1}{N}\sum_{i=1}^{N}\lambda_i^0\lambda_i^{0\prime}\right)\frac{F^{0\prime}\hat{F}}{T} \\
&= F^0\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)\left(\frac{F^{0\prime}\hat{F}}{T}\right),
\end{aligned}$$

so we can see that

$$\hat{F}V_{NT} - F^0\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)\left(\frac{F^{0\prime}\hat{F}}{T}\right) = I1 + \cdot + I8.$$

We now investigate the rate of convergence of each of the terms $I1, \cdots, I8$.

- $\frac{1}{\sqrt{T}}I1$
  Note that

$$\begin{aligned}\left\|\frac{1}{\sqrt{T}}I1\right\| &= \left\|\frac{1}{NT^{3/2}}\sum_{i=1}^{N}X_i(\beta^0-\hat{\beta})(\beta^0-\hat{\beta})'X_i\hat{F}\right\| \\ &\le \left(\frac{1}{NT}\sum_{i=1}^{N}\|X_i\|^2\right)\left|\beta^0-\hat{\beta}\right|^2\frac{1}{\sqrt{T}}\left\|\hat{F}\right\|.\end{aligned}$$

  Since $\frac{1}{NT}\sum_{i=1}^{N}\|X_i\|^2 = O_p(1)$ and

$$\frac{1}{\sqrt{T}}\left\|\hat{F}\right\| \le \left[\operatorname{tr}\left(\frac{\hat{F}'\hat{F}}{T}\right)\right]^{\frac{1}{2}} = \sqrt{r},$$

  we can see that

$$\frac{1}{\sqrt{T}}I1 = O_p\left(\left|\beta^0-\hat{\beta}\right|^2\right) = o_p\left(\left|\beta^0-\hat{\beta}\right|\right),$$

  where the last equality follows because $\left|\beta^0-\hat{\beta}\right| = o_p(1)$.

- $\frac{1}{\sqrt{T}}I2$
  We can see that

$$\begin{aligned}\left\|\frac{1}{\sqrt{T}}I2\right\| &= \left\|\frac{1}{NT^{3/2}}\sum_{i=1}^{N}X_i(\beta^0-\hat{\beta})\lambda_i^{0\prime}F^{0\prime}\hat{F}\right\| \\ &\le \left(\frac{1}{NT}\sum_{i=1}^{N}\|X_i\|^2\right)^{\frac{1}{2}}\left(\frac{1}{N}\sum_{i=1}^{N}\left|\lambda_i^0\right|^2\right)^{\frac{1}{2}}\left|\beta^0-\hat{\beta}\right|\frac{F^{0\prime}\hat{F}}{T}\end{aligned}$$

  by the Cauchy-Schwarz inequality. All terms except $\left|\beta^0-\hat{\beta}\right|$ on the right hand side are $O_p(1)$, so

$$\frac{1}{\sqrt{T}}I2 = O_p\left(\left|\beta^0-\hat{\beta}\right|\right).$$

- $\frac{1}{\sqrt{T}}I3$

As above, we have

$$\left\|\frac{1}{\sqrt{T}}I3\right\| = \left\|\frac{1}{NT^{3/2}}\sum_{i=1}^{N} X_i(\beta^0-\hat{\beta})e_i'\hat{F}\right\|$$
$$\leq \left(\frac{1}{NT}\sum_{i=1}^{N}\|X_i\|^2\right)^{\frac{1}{2}}\left(\frac{1}{NT}\sum_{i=1}^{N}|e_i|^2\right)^{\frac{1}{2}}\left|\beta^0-\hat{\beta}\right|\cdot\frac{1}{\sqrt{T}}\hat{F}.$$

Again, every term except $\left|\beta^0-\hat{\beta}\right|$ on the right hand side are $O_p(1)$, which means that

$$\frac{1}{\sqrt{T}}I3 = O_p\left(\left|\beta^0-\hat{\beta}\right|\right).$$

- $\frac{1}{\sqrt{T}}I4$

  As should be familiar by now,

$$\left\|\frac{1}{\sqrt{T}}I4\right\| = \left\|\frac{1}{NT^{3/2}}\sum_{i=1}^{N} e_i(\beta^0-\hat{\beta})'X_i'\hat{F}\right\|$$
$$\leq \left(\frac{1}{NT}\sum_{i=1}^{N}\|X_i\|^2\right)^{\frac{1}{2}}\left(\frac{1}{NT}\sum_{i=1}^{N}|e_i|^2\right)^{\frac{1}{2}}\left|\beta^0-\hat{\beta}\right|\cdot\frac{1}{\sqrt{T}}\hat{F}.$$

  The term on the right hand side is exactly the same as the one appearing in the case of $I3$, so

$$\frac{1}{\sqrt{T}}I4 = O_p\left(\left|\beta^0-\hat{\beta}\right|\right).$$

- $\frac{1}{\sqrt{T}}I7$

  Finally, we can see that

$$\left\|\frac{1}{\sqrt{T}}I7\right\| = \left\|\frac{1}{NT^{3/2}}\sum_{i=1}^{N} F^0\lambda_i^0(\beta^0-\hat{\beta})'X_i'\hat{F}\right\|$$
$$\leq \left(\frac{1}{NT}\sum_{i=1}^{N}\|X_i\|^2\right)^{\frac{1}{2}}\left(\frac{1}{N}\sum_{i=1}^{N}\left|\lambda_i^0\right|^2\right)^{\frac{1}{2}}\left|\beta^0-\hat{\beta}\right|\cdot\frac{1}{\sqrt{T}}\left\|\hat{F}\right\|\cdot\frac{1}{\sqrt{T}}\left\|F^0\right\|.$$

  The term on the right hand side is almost exactly the same as the one appearing in the case of $I2$, so

$$\frac{1}{\sqrt{T}}I7 = O_p\left(\left|\beta^0-\hat{\beta}\right|\right).$$

- $\frac{1}{\sqrt{T}} I5$

  This time, we follow the same process as in Bai and Ng (2002). Note that

$$\left\| \frac{1}{\sqrt{T}} I5 \right\| = \left\| \frac{1}{NT^{3/2}} \sum_{i=1}^{N} e_i \lambda_i^{0\prime} F^{0\prime} \hat{F} \right\|$$
$$\leq \frac{1}{\sqrt{N}} \left(\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \left\| e_i \lambda_i^{0\prime} \right\| \right) \cdot \left\| \frac{F^{0\prime} \hat{F}}{T} \right\|.$$

  Since

$$\frac{F^{0\prime} \hat{F}}{T} = O_p(1), \quad \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \left\| e_i \lambda_i^{0\prime} \right\| = O_p(1),$$

  we can see that

$$\frac{1}{\sqrt{T}} I5 = O_p \left(\frac{1}{\sqrt{N}} \right).$$

- $\frac{1}{\sqrt{T}} I8$

  The exposition for this term follows that of the above term almost exactly.

$$\left\| \frac{1}{\sqrt{T}} I8 \right\| = \left\| \frac{1}{NT^{3/2}} \sum_{i=1}^{N} F^0 \lambda_i^0 e_i' \hat{F} \right\|$$
$$\leq \frac{1}{\sqrt{T}} \left\| F^0 \right\| \cdot \frac{1}{\sqrt{T}} \left\| \hat{F} \right\| \cdot \left(\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \left\| \lambda_i^0 e_i' \right\| \right) \cdot \frac{1}{\sqrt{N}},$$

  so as above,

$$\frac{1}{\sqrt{T}} I8 = O_p \left(\frac{1}{\sqrt{N}} \right).$$

- $\frac{1}{\sqrt{T}} I6$

  We now move onto our final and most troublesome term. Note that

$$\left\| \frac{1}{\sqrt{T}} I6 \right\| = \left\| \frac{1}{NT^{3/2}} \sum_{i=1}^{N} e_i e_i' \hat{F} \right\|$$
$$= \left\| \frac{1}{NT^{3/2}} \sum_{i=1}^{N} \sum_{t=1}^{T} e_i e_{it} \hat{F}_t' \right\|$$
$$\leq \left\| \frac{1}{NT^{3/2}} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(e_i e_{it} - \mathbb{E} \left[e_i e_{it} \right] \right) \hat{F}_t' \right\| + \left\| \frac{1}{NT^{3/2}} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbb{E} \left[e_i e_{it} \right] \cdot \hat{F}_t' \right\|.$$

  We study each term in turn.

By the Cauchy-Schwarz inequality,

$$\left\| \frac{1}{NT^{3/2}} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(e_i e_{it} - \mathbb{E}\left[e_i e_{it} \right] \right) \hat{F}_t' \right\| \le \frac{1}{NT^{3/2}} \sum_{t=1}^{T} \left| \sum_{i=1}^{N} \left(e_i e_{it} - \mathbb{E}\left[e_i e_{it} \right] \right) \right| \left| \hat{F}_t' \right|$$

$$\le \frac{1}{\sqrt{N}} \left(\frac{1}{NT^2} \sum_{t=1}^{T} \left| \sum_{i=1}^{N} (e_i e_{it} - \mathbb{E}\left[e_i e_{it} \right]) \right|^2 \right)^{\frac{1}{2}} \left(\frac{1}{T} \sum_{t=1}^{T} \left| \hat{F}_t \right|^2 \right)^{\frac{1}{2}}.$$

Since

$$\sum_{i=1}^{N} (e_i e_{it} - \mathbb{E}\left[e_i e_{it} \right]) = \begin{pmatrix} \sum_{i=1}^{N} (e_{i1} e_{it} - \mathbb{E}\left[e_{i1} e_{it} \right]) \\ \vdots \\ \sum_{i=1}^{N} (e_{iT} e_{it} - \mathbb{E}\left[e_{iT} e_{it} \right]) \end{pmatrix},$$

we have

$$\frac{1}{NT^2} \sum_{t=1}^{T} \left| \sum_{i=1}^{N} (e_i e_{it} - \mathbb{E}\left[e_i e_{it} \right]) \right|^2 = \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \left| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (e_{is} e_{it} - \mathbb{E}\left[e_{is} e_{it} \right]) \right|^2.$$

We showed above that there exists an $M < +\infty$ such that

$$\mathbb{E} \left| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (e_{is} e_{it} - \mathbb{E}\left[e_{is} e_{it} \right]) \right|^2 < M$$

for any $t, s \in N_+$; it follows that

$$\mathbb{E}\left[\frac{1}{NT^2} \sum_{t=1}^{T} \left| \sum_{i=1}^{N} (e_i e_{it} - \mathbb{E}\left[e_i e_{it} \right]) \right|^2 \right] = \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathbb{E} \left| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (e_{is} e_{it} - \mathbb{E}\left[e_{is} e_{it} \right]) \right|^2 < M$$

for any $N, T \in N_+$, so that

$$\frac{1}{NT^2} \sum_{t=1}^{T} \left| \sum_{i=1}^{N} (e_i e_{it} - \mathbb{E}\left[e_i e_{it} \right]) \right|^2 = O_p(1).$$

Therefore,

$$\left\| \frac{1}{NT^{3/2}} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(e_i e_{it} - \mathbb{E}\left[e_i e_{it} \right] \right) \hat{F}_t' \right\| \le \frac{1}{\sqrt{N}} O_p(1) \cdot \mathrm{tr} \left(\frac{\hat{F}' \hat{F}}{T} \right)^{\frac{1}{2}}$$

$$= \frac{1}{\sqrt{N}} O_p(1) \sqrt{r}.$$

As for the second term, we can see that

$$\left\| \frac{1}{NT^{3/2}} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbb{E}\left[e_i e_{it} \right] \hat{F}_t' \right\| \le \frac{1}{\sqrt{T}} \cdot \left(\frac{1}{N^2 T} \sum_{t=1}^{T} \left| \sum_{i=1}^{N} \mathbb{E}\left[e_i e_{it} \right] \right|^2 \right)^{\frac{1}{2}} \left(\frac{1}{T} \sum_{t=1}^{T} \left| \hat{F}_t \right|^2 \right)^{\frac{1}{2}}$$

by applying the Cauchy=Schwarz inequality in the same manner as above. Since

$$\sum_{i=1}^{N} \mathbb{E}\left[e_i e_{it}\right] = \begin{pmatrix} \sum_{i=1}^{N} \mathbb{E}\left[e_{i1} e_{it}\right] \\ \vdots \\ \sum_{i=1}^{N} \mathbb{E}\left[e_{iT} e_{it}\right] \end{pmatrix},$$

by the definition of the euclidean metric we have

$$\begin{aligned} \frac{1}{N^2 T} \sum_{t=1}^{T} \left| \sum_{i=1}^{N} \mathbb{E}\left[e_i e_{it}\right] \right|^2 &= \frac{1}{N^2 T} \sum_{t=1}^{T} \sum_{s=1}^{T} \left| \sum_{i=1}^{N} \mathbb{E}\left[e_{is} e_{it}\right] \right|^2 \\ &= \frac{1}{N^2 T} \sum_{t=1}^{T} \left| \sum_{i=1}^{N} \mathbb{E}\left[e_{it}^2\right] \right|^2 \\ &= \frac{1}{N^2 T} \sigma^2 (N^2 T) = \sigma^2. \end{aligned}$$

Therefore,

$$\begin{aligned} \left\| \frac{1}{N T^{3/2}} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbb{E}\left[e_i e_{it}\right] \hat{F}_t' \right\| &\leq \frac{\sigma^2}{\sqrt{T}} \cdot \operatorname{tr}\left(\frac{\hat{F}' \hat{F}}{T} \right)^{\frac{1}{2}} \\ &= \frac{\sigma^2 \sqrt{r}}{\sqrt{T}}. \end{aligned}$$

Putting the results together,

$$\left\| \frac{1}{\sqrt{T}} I6 \right\| \leq \frac{1}{\sqrt{N}} O_p(1) \sqrt{r} + \frac{\sigma^2 \sqrt{r}}{\sqrt{T}},$$

so that

$$\frac{1}{\sqrt{T}} I6 = O_p\left(\frac{1}{\min(\sqrt{N}, \sqrt{T})} \right).$$

We have seen above that $\frac{1}{\sqrt{T}}$ times $I1,\cdots,I4,I7$ are all $O_p\left(\left|\beta^0-\hat{\beta}\right|\right)$, and that $\frac{1}{\sqrt{T}}$ times the terms $I5,I6,I8$ are $O_p\left(\frac{1}{\min(\sqrt{N},\sqrt{T})}\right)$. Therefore,

$$\frac{1}{\sqrt{T}}(I1+\cdots+I8)=O_p\left(\left|\beta^0-\hat{\beta}\right|\right)+O_p\left(\frac{1}{\min(\sqrt{N},\sqrt{T})}\right),$$

and because $\left|\beta^0-\hat{\beta}\right|$, $\frac{1}{\min(\sqrt{N},\sqrt{T})}$ are $o_p(1)$, it follows that $\frac{1}{\sqrt{T}}(I1+\cdots+I8)=o_p(1)$ as well.

The Probability Limit of V_{NT}

We now assume, as in our study of the unilevel factor model, that there exists a nonsingular $r\times r$ matrix Q such that

$$\frac{F^{0\prime}\hat{F}}{T}\xrightarrow{p}Q.$$

This assumption is for the sake of simplifying the proofs.

Since

$$\hat{F}V_{NT}-F^0\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)\left(\frac{F^{0\prime}\hat{F}}{T}\right)=I1+\cdot+I8,$$

premultiplying both sides by $\frac{\hat{F}'}{\sqrt{T}}$ and using the fact that $\frac{\hat{F}'\hat{F}}{T}=I_r$ implies that

$$V_{NT}-\frac{\hat{F}'F^0}{T}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)\frac{F^{0\prime}\hat{F}}{T}=o_p(1).$$

Because

$$\frac{\hat{F}'F^0}{T}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)\frac{F^{0\prime}\tilde{F}}{T}\xrightarrow{p}Q'\Sigma_\Lambda Q,$$

it follows that

$$V_{NT}\xrightarrow{p}V=Q'\Sigma_\Lambda Q,$$

where V is positive definite because Q has full rank and Σ_Λ is positive definite, and is diagonal because V_{NT} is diagonal for any N,T. In addition, the diagonal entries of V are ordered because the diagonal entries of V_{NT} are ordered.
By the continuous mapping theorem,

$$V_{NT}^{-1}\xrightarrow{p}V^{-1}.$$

To derive the specific form of V, we proceed as follows.
Premultiplying both sides of the equation above by $\frac{F^{0\prime}}{\sqrt{T}} = O_p(1)$ yields the equation

$$\frac{F^{0\prime}\hat{F}}{T}V_{NT} - \left(\frac{F^{0\prime}F^0}{T}\right)\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)\frac{F^{0\prime}\tilde{F}}{T} = o_p(1).$$

From our assumptions, we have

$$\left[\left(\frac{F^{0\prime}F^0}{T}\right)\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)\right]\frac{F^{0\prime}\hat{F}}{T} \xrightarrow{p} \Sigma_F\Sigma_\Lambda Q,$$

or equivalently,

$$\left[\left(\frac{F^{0\prime}F^0}{T}\right)\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)\right]\frac{F^{0\prime}\hat{F}}{T} - \Sigma_F\Sigma_\Lambda Q = o_p(1).$$

Likewise, we have

$$\frac{F^{0\prime}\hat{F}}{T}V_{NT} - QV = o_p(1).$$

By implication,

$$\Sigma_F\Sigma_\Lambda Q - QV = o_p(1),$$

and because the left hand side is deterministic, this means

$$\Sigma_F\Sigma_\Lambda Q - QV = O.$$

By definition, V is a diagonal matrix with diagonal entries equal to the eigenvalues of $\Sigma_F\Sigma_\Lambda$. Because Σ_F and Σ_Λ are positive definite, the eigenvalues of $\Sigma_F\Sigma_\Lambda$ are exactly those of $\Sigma_\Lambda\Sigma_F$.

The Consistency of F

With the above result, we can now establish the mean square consistency of $\hat{F}$ in the usual way. Defining

$$H = \left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)\left(\frac{F^{0\prime}F^0}{T}\right)V_{NT}^{-1},$$

since

$$\hat{F}V_{NT} - F^0 H V_{NT} = I1 + \cdots + I8,$$

we have

$$\frac{1}{\sqrt{T}}\left(\hat{F} - F^0 H\right) = \frac{1}{\sqrt{T}}(I1 + \cdots + I8)V_{NT}^{-1}.$$

Therefore,

$$\frac{1}{\sqrt{T}}\left\|\hat{F} - F^0 H\right\| \leq \left(\left\|\frac{1}{\sqrt{T}}I1\right\| + \cdots + \left\|\frac{1}{\sqrt{T}}I8\right\|\right) \cdot \left\|V_{NT}^{-1}\right\|.$$

By the result established above,

$$V_{NT}^{-1} \xrightarrow{p} V^{-1},$$

so that $V_{NT}^{-1} = O_p(1)$. Since

$$\left\|\frac{1}{\sqrt{T}}I1\right\| + \cdots + \left\|\frac{1}{\sqrt{T}}I8\right\| = O_p\left(\left|\beta^0 - \hat{\beta}\right|\right) + O_p\left(\frac{1}{\min(\sqrt{N}, \sqrt{T})}\right),$$

it stands to reason that

$$\frac{1}{\sqrt{T}}\left\|\hat{F} - F^0 H\right\| = O_p\left(\left|\beta^0 - \hat{\beta}\right|\right) + O_p(\delta_{NT}^{-1/2}),$$

where $\delta_{NT} = \min(N, T)$.

Finally, we are able to see that

$$\frac{1}{T}\left\|\hat{F} - F^0 H\right\|^2 = O_p\left(\left|\beta^0 - \hat{\beta}\right|^2\right) + O_p(\delta_{NT}^{-1}) + O_p\left(\left|\beta^0 - \hat{\beta}\right| \cdot \delta_{NT}^{-1/2}\right);$$

since

$$\min(\left|\beta^0 - \hat{\beta}\right|^2, \delta_{NT}^{-1}) \leq \left|\beta^0 - \hat{\beta}\right| \cdot \delta_{NT}^{-1/2},$$

we can write

$$\frac{1}{T}\left\|\hat{F} - F^0 H\right\|^2 = O_p\left(\left|\beta^0 - \hat{\beta}\right|^2\right) + O_p(\delta_{NT}^{-1}).$$

4.6 Testing Additive Effects against Interactive Effects

One might be interested in testing the additive fixed effect specification

$$Y_{it} = X_{it}'\beta + \alpha_i + \varepsilon_t + e_{it}$$

against the interactive fixed effects specification

$$Y_{it} = X_{it}'\beta + \lambda_i' F_t + e_{it}.$$

Let $\hat{\beta}$ be the iterated least squares estimator of β under the interactive effects model and $\tilde{\beta}$ the LSDV estimator for β in the additive effects model. We saw above that $\hat{\beta}$ is consistent and asymptotically normal when the fixed effects are interactive, and since the additive effects model is a special case of the interactive effects model, $\hat{\beta}$ remains consistent and asymptotically normal under the additive effects model.

On the other hand, since the errors e_{it} are assumed to be homoskedastic, i.i.d. and uncorrelated with the regressors, under the additive effects model the OLS estimator $\tilde{\beta}$ of β is consistent, asymptotically normal, and in fact the asymptotically efficient estimator of β.

In summary,

- Under the additive effects model, $\tilde{\beta}$ is a consistent, asymptotically normal and asymptotically efficient estimator of β.
  $\hat{\beta}$ is a consistent and asymptotically normal estimator of β.

- Under the interactive effects model, $\hat{\beta}$ is a consistent, asymptotically normal estimator of β, but $\tilde{\beta}$ is inconsistent.

Therefore, given that $\tilde{\beta}$ and $\hat{\beta}$ are jointly asymptotically normal under the additive effects model, the conditions for the Hausman test to be asymptotically chi-squared under the null of the additive effects model and consistent under the alternative of the interactive effects model are satisfied. The Hausman test can be formulated as the test with null and alternative hypotheses

$$H_0 : \text{The True Model is the Additive Effects Model}$$
$$H_1 : \text{The True Model is the Interactive Effects Model}$$

and the test statistic

$$\hat{H}_{NT} = NT \cdot (\hat{\beta} - \tilde{\beta})' \left(\hat{V}(\hat{\beta}) - \hat{V}(\tilde{\beta}) \right)^{\dagger} (\hat{\beta} - \tilde{\beta}) \xrightarrow{d} \chi_k^2,$$

where $\hat{V}(\hat{\beta})$ and $\hat{V}(\tilde{\beta})$ are consistent estimators for the asymptotic variances of $\hat{\beta}$ and $\tilde{\beta}$, and

$$\left(\hat{V}(\hat{\beta}) - \hat{V}(\tilde{\beta}) \right)^{\dagger}$$

is a type of pseudoinverse of $\hat{V}(\hat{\beta}) - \hat{V}(\tilde{\beta})$ that equals its inverse when $\hat{V}(\hat{\beta}) - \hat{V}(\tilde{\beta})$ is nonsingular. The specific form of this pseudoinverse will be derived in time.

Below we study the Hausman test in detail, and investigate which assumptions are needed for it to have the desired asymptotic distribution under the null and consistency under the alternative.

4.6.1 The Hausman Test

Assumptions

Let there be two models $\mathcal{M}_1$ and $\mathcal{M}_2$ with the same set of k parameters β, and let $\hat{\beta}$ and $\tilde{\beta}$ be estimators of β in models 1 and 2. Suppose we want to test for the null and alternative hypotheses

$$H_0: \ \mathcal{M}_1 \text{ is the true model} \quad H_0: \ \mathcal{M}_2 \text{ is the true model.}$$

Letting $\{a_N\}_{N \in N_+}$ be some sequence of real numbers increasing to $+\infty$, assume that:

i) **Consistency, Asymptotic Normality and Efficiency Under the Null**
Under $\mathcal{M}_1$,

$$a_N \cdot \begin{pmatrix} \hat{\beta} - \beta \\ \tilde{\beta} - \beta \end{pmatrix} \xrightarrow{d} N \left[\mathbf{0}, \underbrace{\begin{pmatrix} A_1 & A_{12} \\ A_{21} & A_2 \end{pmatrix}}_{A} \right]$$

as $N \to \infty$, where A_2 is the asymptotically efficient covariance matrix.

ii) **Behavior Under the Alternative**
Under $\mathcal{M}_2$,

$$\hat{\beta} \xrightarrow{p} \beta,$$

while

$$\tilde{\beta} \xrightarrow{p} \gamma \neq \beta$$

as $N \to \infty$.

iii) **Consistent Variance Estimators**
Let $\hat{V}(\hat{\beta})$ and $\hat{V}(\tilde{\beta})$ be estimators of the asymptotic variance of $\hat{\beta}$ and $\tilde{\beta}$. We assume that

$$\hat{V}(\hat{\beta}) \xrightarrow{p} A_1, \quad \hat{V}(\tilde{\beta}) \xrightarrow{p} A_2$$

under $\mathcal{M}_1$, and that

$$\hat{V}(\hat{\beta}) - \hat{V}(\tilde{\beta}) \xrightarrow{p} W$$

under $\mathcal{M}_2$, where $W \in \mathbb{R}^{k \times k}$ is a postivie definite matrix.

iv) **Regularity Assumptions for Estimation**
The rank of $\hat{V}(\hat{\beta}) - \hat{V}(\tilde{\beta})$ and $A_1 - A_2$ are constant at $0 < r \leq k$ for any $N \in N_+$. Moreover, the r non-zero eigenvalues of $\hat{V}(\hat{\beta}) - \hat{V}(\tilde{\beta})$ and $A_1 - A_2$ are all distinct.

We will now construct the Hausman test statistic and show that it is asymptotically chi-squared under the null and that it defines a consistent test under the alternative.

The Covariance Structure under the Null

For now, suppose that $\mathcal{M}_1$ is the true model, that is, assume that the null is true. We will show that the asymptotic covariance of the efficient estimator $\tilde{\beta}$ and the difference $\hat{\beta} - \tilde{\beta}$ of the estimators must be zero.

By joint asymptotic normality, under $\mathcal{M}_1$,

$$a_N \left(\hat{\beta} - \tilde{\beta}\right) = a_N \cdot \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} \hat{\beta} - \beta \\ \tilde{\beta} - \beta \end{pmatrix} \xrightarrow{d} \begin{pmatrix} 1 & -1 \end{pmatrix} N(\mathbf{0}, A) = N\left[\mathbf{0}, \underbrace{A_1 - A_{21} - A_{12} + A_2}_{\Sigma}\right].$$

Because $\tilde{\beta}$ is asymptotically efficient, we can now show that the asymptotic covariance A_{12} of $\hat{\beta}$ and $\tilde{\beta}$ is equal to the minimal asymptotic variance A_2.

Define

$$\hat{q} = \hat{\beta} - \tilde{\beta}.$$

By assumption, $\hat{q} \xrightarrow{p} \mathbf{0}$ under $\mathcal{M}_1$.
Suppose that $A_{12} \neq A_2$, and define a new estimator $\check{\beta}$ as

$$\check{\beta}(r) = \tilde{\beta} + rC\hat{q},$$

for any $r \in \mathbb{R}$, where

$$C = -(A_{21} - A_2) \neq O.$$

Then,

$$\begin{aligned}
a_N\left(\check{\beta}(r)-\beta\right) &= a_N\cdot\left(\tilde{\beta}-\beta\right)+a_N\cdot rC\hat{q}\\
&= a_N\cdot\left(\tilde{\beta}-\beta\right)+a_N\cdot rC\left((\hat{\beta}-\beta)-(\tilde{\beta}-\beta)\right)\\
&= a_N\cdot\begin{pmatrix} rC & I_k-rC\end{pmatrix}\begin{pmatrix}\hat{\beta}-\beta\\ \tilde{\beta}-\beta\end{pmatrix}\xrightarrow{d} N\left[\mathbf{0},V(r)\right],
\end{aligned}$$

where

$$V(r)=A_2+r^2C\Sigma C'+r\left(A_{21}-A_2\right)C'+rC\left(A_{12}-A_2\right).$$

As such, $\check{\beta}(r)$ is a consistent and asymptotically normal estimator of β for any $r\in\mathbb{R}$, and by the minimality of A_2, it must be the case that $V(r)-A_2\geq 0$, that is, it must be positive semidefinite.

For any $r\in\mathbb{R}$, using the fact that $C=-(A_{21}-A_2)$, we have

$$V(r)=A_2+r^2C\Sigma C'-2r\cdot CC'.$$

We also assumed that $C'\neq O$, so C' has non-zero rank and thus there exists some non-zero vector $\alpha\in\mathbb{R}^k$ such that $u=C'\alpha\neq\mathbf{0}$. By implication,

$$f(r)=\alpha'V(r)\alpha=\alpha'A_2\alpha+r^2\cdot u'\Sigma u-2r\cdot u'u,$$

where $u'u>0$. Note that $f(0)=\alpha'A_2\alpha$, and that f is differentiable on $\mathbb{R}$ with first order derivative

$$f'(r)=2\left(r\cdot u'\Sigma u-u'u\right)$$

for any $r\in\mathbb{R}$.
$u'\Sigma u\geq 0$ because Σ is a covariance matrix and thus positive semidefinite; if $u'\Sigma u=0$ then $f'(r)=-2u'u<0$ for any $r\in\mathbb{R}$, while if $u'\Sigma u>0$ then $f'(r)<0$ for any $0\leq r<\frac{u'u}{u'\Sigma u}$, where $\frac{u'u}{u'\Sigma u}>0$. In any case, there exists a $x>0$ such that $f'(r)<0$ for any $r\in[0,x)$.
We can now conclude, from the mean value theorem, that there exists a $y\in(0,x)$ such that

$$f(0)-f(x)=f'(y)\cdot x<0,$$

where the last inequality follows because $f'(r)<0$ for any $r\in(0,x)$. By implication,

$$\alpha'A_2\alpha=f(0)<f(x)=\alpha'V(x)\alpha,$$

which contradicts the fact that $A_2-V(x)$ should be positive semidefinite.

Therefore, it must be the case that $A_{12} = A_2$, and as such that

$$\Sigma = A_1 - A_{12} - A_{21} + A_2 = A_1 - A_2.$$

That is, the asymptotic covariance matrix of $\hat{\beta} - \tilde{\beta}$ is simply the difference of their asymptotic covariance matrices.
To obtain an equivalent formulation, note that

$$a_N \cdot \begin{pmatrix} \hat{q} \\ \tilde{\beta} - \beta \end{pmatrix} = a_N \cdot \begin{pmatrix} I_k & -I_k \\ O & I_k \end{pmatrix} \begin{pmatrix} \hat{\beta} - \beta \\ \tilde{\beta} - \beta \end{pmatrix} \xrightarrow{d} N \left[\mathbf{0}, \begin{pmatrix} \Sigma & A_{12} - A_2 \\ A_{21} - A_2 & A_2 \end{pmatrix} \right],$$

which tells us that $A_{21} - A_2$ is the asymptotic covariance of $\tilde{\beta}$ and $\hat{q}$. The result we derived above thus tells us that asymptotically, the asymptotically efficient estimator and the difference $\hat{q}$ of the estimators must be uncorrelated.

The Test Statistic and its Asymptotic Distribution

So far, we have seen that

$$a_N \cdot (\hat{\beta} - \tilde{\beta}) \xrightarrow{d} N \left[\mathbf{0}, \underbrace{A_1 - A_2}_{\Sigma} \right].$$

We now construct and derive the asymptotic distribution of the Hausman test statistic.

Because Σ is symmetric positive semidefinite, it has an eigendecomposition

$$\Sigma = PDP',$$

where P is a $k \times k$ orthogonal matrix and D a diagonal matrix with diagonal entries equal to the eigenvalues $\lambda_1 \geq \cdots \lambda_k \geq 0$ of Σ. Suppose that the rank of Σ is $0 < r \leq k$, so that Σ has exactly r non-zero eigenvalues $\lambda_1 \geq \cdots \geq \lambda_r > 0$. Defining

$$\bar{D} = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_r \end{pmatrix} \in \mathbb{R}^{r \times r},$$

and letting $\bar{P} \in \mathbb{R}^{k \times r}$ collect the first r columns of P, we can easily see that

$$\Sigma = \bar{P} \bar{D} \bar{P}',$$

where $\bar{P}' \bar{P} = I_r$. Since the diagonal elements of $\bar{D}$ are all positive, its matrix square root $\bar{D}^{\frac{1}{2}}$

exists, and we can define

$$\Sigma^{\frac{1}{2}} = \bar{P}\bar{D}^{\frac{1}{2}} \in \mathbb{R}^{k\times r},$$

with its generalized inverse

$$\Sigma^{\frac{1}{2}\dagger} = \left(\Sigma^{\frac{1}{2}\prime}\Sigma^{\frac{1}{2}}\right)^{-1}\Sigma^{\frac{1}{2}\prime} = \bar{D}^{-\frac{1}{2}}\bar{P}' \in \mathbb{R}^{r\times k}.$$

Define $\Sigma^{\dagger}$ as

$$\Sigma^{\dagger} = \Sigma^{\frac{1}{2}\dagger\prime}\Sigma^{\frac{1}{2}\dagger} = \bar{P}\bar{D}^{-1}\bar{P}',$$

and let $X \sim N[\mathbf{0},\Sigma]$. Since

$$Z = \Sigma^{\frac{1}{2}\dagger}X \sim N(\mathbf{0}, I_r),$$

we can obtain the following distribution:

$$X'\Sigma^{\dagger}X = X'\Sigma^{\frac{1}{2}\dagger\prime}\Sigma^{\frac{1}{2}\dagger}X = Z'Z \sim \chi^2_r.$$

Under our regularity assumptions, provided that the vector of signs s is fixed, we can uniquely recover the r orthonormal eigenvectors of $\hat{V}(\hat{\beta}) - \hat{V}(\tilde{\beta})$ corresponding to its r non-zero ordered eigenvalues; they are given by the $k\times r$ random matrix

$$\hat{P}_N = eigvec^s_{k,r}\left(\hat{V}(\hat{\beta}) - \hat{V}(\tilde{\beta})\right).$$

Letting $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_r > 0$ be the r non-zero ordered eigenvalues of $\hat{V}(\hat{\beta}) - \hat{V}(\tilde{\beta})$, by the continuity of ordered eigenvalues and eigenvecctors, and the fact that

$$\hat{V}(\hat{\beta}) - \hat{V}(\tilde{\beta}) \xrightarrow{p} A_1 - A_2 = \Sigma,$$

the continuous mapping theorem tells us that

$$\hat{\lambda}_i \xrightarrow{p} \lambda_i \quad \text{for any } 1 \leq i \leq r,$$
$$\hat{P}_N \xrightarrow{p} eigvec^s_{k,r}(\Sigma) = \bar{P},$$

where the last equality follows because the distribution of $X'\Sigma^{\dagger}X$ does not depend on the sign of $\bar{P}$.

As such, defining

$$\hat{D}_N = \begin{pmatrix} \hat{\lambda}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \hat{\lambda}_r \end{pmatrix}$$

and

$$\left(\hat{V}(\hat{\beta}) - \hat{V}(\tilde{\beta})\right)^{\dagger} = \hat{P}_N \hat{D}_N^{-1} \hat{P}_N,$$

by the continuous mapping theorem again we have

$$\left(\hat{V}(\hat{\beta}) - \hat{V}(\tilde{\beta})\right)^{\dagger} \xrightarrow{p} \bar{P}\bar{D}^{-1}\bar{P} = \Sigma^{\dagger}.$$

The Hausman test statistic is then defined as

$$\hat{H}_N = a_N^2 \cdot (\hat{\beta} - \tilde{\beta})' \left(\hat{V}(\hat{\beta}) - \hat{V}(\tilde{\beta})\right)^{\dagger} (\hat{\beta} - \tilde{\beta}),$$

and we can see that

$$\hat{H}_N \xrightarrow{d} X'\Sigma^{\dagger}X \sim \chi_r^2.$$

Note that, if Σ has full rank, then $\left(\hat{V}(\hat{\beta}) - \hat{V}(\tilde{\beta})\right)^{\dagger}$ is simply the inverse of $\hat{V}(\hat{\beta}) - \hat{V}(\tilde{\beta})$, which shows us that the above case is a generalization of the usual Wald-type statistics.

The Consistency of the Hausman Test

Now that we have constructed the Hausman test statistic and derived its asymptotic distribution under the null, it remains to verify whether this statistic defines a consistent test, that is, a test whose power goes to 1 under the alternative.

Suppose that $\mathcal{M}_2$, not $\mathcal{M}_1$, is the true model.
We first provide a heuristic explanation as to why the test must be consistent. Because $\hat{\beta}$ consistently estimates the true parameter β but $\tilde{\beta}$ does not under model $\mathcal{M}_2$, the difference between the two estimators will be large, meaning that the Hausman test statistic, which is defined as a quadratic form involving the difference between the two estimators, will also tend to infinity as the sample size increases. This indicates that the test will be consistent.

Formally, we assumed that

$$\hat{\beta} - \tilde{\beta} \xrightarrow{p} \beta - \gamma$$

under $\mathcal{M}_2$, and that

$$\hat{V}(\hat{\beta}) - \hat{V}(\tilde{\beta}) \xrightarrow{p} W,$$

where W is a positive definite $k \times k$ matrix. By implication,

$$(\hat{\beta}-\tilde{\beta})'\left(\hat{V}(\hat{\beta})-\hat{V}(\tilde{\beta})\right)^{-1}(\hat{\beta}-\tilde{\beta}) \xrightarrow{p} (\beta-\gamma)'W(\beta-\gamma) > 0,$$

and as such,

$$\hat{H}_N = a_N^2 \cdot (\hat{\beta}-\tilde{\beta})'\left(\hat{V}(\hat{\beta})-\hat{V}(\tilde{\beta})\right)^{-1}(\hat{\beta}-\tilde{\beta}) \xrightarrow{p} +\infty$$

as $N \to \infty$. This implies that

$$\lim_{N\to\infty} \mathbb{P}\left(\hat{H}_N < c\right) = 0$$

for any $c \in \mathbb{R}$, so that the probability of rejecting goes to 1 as $N \to \infty$.

Note that we made the very strong assumption that $\hat{V}(\hat{\beta})-\hat{V}(\tilde{\beta})$ converges to a positive definite matrix under $\mathcal{M}_2$ to prove consistency. This requires $\hat{V}(\tilde{\beta})$ to always be "smaller" than $\hat{V}(\hat{\beta})$