Supplementary Appendix - Estimating Latent Asset-Pricing Factors

Martin Lettau
Haas School of Business, University of California at Berkeley, Berkeley, CA 94720,
lettau@berkeley.edu,NBER,CEPR

Markus Pelger
Department of Management Science & Engineering, Stanford University, Stanford, CA 94305,
mpelger@stanford.edu

Abstract
This draft: December 7, 2018
This supplementary appendix provides the proofs for the strong factor model and additional simulation results.
1. Proofs for the Strong Factor Model

The proofs for the strong factor model follow from the results in Bai (2003) after replacing the factors and asset space by their projected counterpart. We replace the factors and asset space by their projected counterpart $W_F$ and $W_X$ in Bai’s (2003) proofs with $W^2 = \left(I_T + \gamma^T ax^3\right)^T$. First, we state the relevant results from Bai (2003) using our notation and then show how to map our framework into his model. Note that the Assumptions in Bai (2003) do not require to demean the data. Assumption 3 is identical to Assumptions A-G in Bai (2003):

**Assumption 3: Bai (2003)**

**A:** Factors: $E[\|F_t\|^4] \leq M < \infty$ and $\frac{1}{T} \sum_{t=1}^T F_t F_t^T \rightarrow P$ for some $K \times K$ positive definite matrix $\Sigma_F$

**B:** Factor loadings: $\|\Lambda_i\| < \infty$, and $\|\Lambda^T \Lambda / N - \Sigma_\Lambda\| \rightarrow 0$ for some $K \times K$ positive definite matrix $\Sigma_\Lambda$.

**C:** Time and cross-section dependence and heteroskedasticity: There exists a positive constant $M < \infty$ such that for all $N$ and $T$:

1. $E[e_{t,i}] = 0$, $E[|e_{t,i}|^8] \leq M$.
2. $E[N^{-1} \sum_{i=1}^N e_{s,i} e_{t,i}] = \xi(s, t), |\xi(s, s)| \leq M$ for all $s$ and for every $t \leq T$ it holds $\sum_{s=1}^T |\xi(s, t)| \leq M$
3. $E[e_{t,i} e_{t,j}] = \tau_{ij,t}$ with $|\tau_{ij,t}| \leq |\tau_{ij}|$ for some $\tau_{ij}$ and for all $t$ and for every $i \leq N$ it holds $\sum_{i=1}^N |\tau_{ij}| \leq M$.
4. $E[e_{t,i} e_{s,j}] = \tau_{ij,ts}$ and $(NT)^{-1} \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^T \sum_{s=1}^T |\tau_{ij,ts}| \leq M$.
5. For every $(t, s), E\left[|N^{-1/2} \sum_{i=1}^N (e_{s,i} e_{t,i}) - E[e_{s,t} e_{t,i}]|^4\right] \leq M$.

**D:** Weak dependence between factors and idiosyncratic errors:

1. $E\left[\frac{1}{N} \sum_{i=1}^N \|\frac{1}{\sqrt{T}} \sum_{t=1}^T F_t e_{t,i}\|^2\right] \leq M$.

**E:** Moments and Central Limit Theorem: There exists an $M < \infty$ such that for all $N$ and $T$:

1. For each $t, E\left[\|\frac{1}{\sqrt{NT}} \sum_{s=1}^T \sum_{k=1}^N F_s(e_{s,k} e_{t,k} - E[e_{s,k} e_{t,k}])\|^2\right] \leq M$.
2. The $K \times K$ matrices satisfy $E\left[\|\frac{1}{\sqrt{NT}} \sum_{t=1}^T \sum_{i=1}^N F_t \Lambda_i^T e_{t,i}\|^2\right] \leq M$.
3. For each $t$ as $N \rightarrow \infty$:

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \Lambda_i e_{t,i} \stackrel{d}{\rightarrow} N(0, \Gamma_t),$$

where $\Gamma_t = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N \Lambda_i \Lambda_j^T E[e_{t,i} e_{t,j}]$

4. For each $i$ as $T \rightarrow \infty$:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T F_t e_{t,i} \stackrel{d}{\rightarrow} N(0, \Omega_{11,i})$$
where $\Omega_{11,i} = p \lim_{T \to \infty} \frac{1}{T} \sum_{i=1}^{T} \sum_{t=1}^{T} E \left[ F_t F_s^\top e_{s,i} e_{t,i} \right]$. 

**F:** The eigenvalues of the $K \times K$ matrix $\Sigma \hat{\Lambda} \tilde{\Sigma}_F$ are distinct.

The following theorem combines Theorem 1, 2 and 3 in Bai (2003):

**Theorem 3: Asymptotic distribution in Bai (2003)**

Assume Assumption 1 holds. Define the estimator $\hat{\Lambda}$ as the eigenvectors of the $K$ largest eigenvalues of $\frac{1}{N^2} X^\top X$ multiplied by $\sqrt{N}$ and $\hat{F} = X \hat{\Lambda} (\hat{\Lambda}^\top \hat{\Lambda})^{-1}$. The following expansions hold with $H = \left( \frac{1}{T} F^\top F \right) \left( \frac{1}{N} \Lambda \hat{\Lambda} \right) V_{TN}^{-1}$ and $V_{TN}$ is a diagonal matrix of the largest $K$ eigenvalues of $\frac{1}{N^2} X^\top X$:

1. $\sqrt{T} \left( H^\top \hat{\Lambda}_i - \Lambda_i \right) \xrightarrow{D} N (0, \tilde{\Sigma}_F \Omega_{11,i} \tilde{\Sigma}_F^{-1})$

2. $\sqrt{N} \left( H^{-1} \hat{F}_t - F_t \right) \xrightarrow{D} N (0, \Lambda \hat{\Lambda}^{-1})$

3. $\sqrt{\delta} \left( \hat{C}_t,i - C_t,i \right) = \sqrt{\frac{N}{T}} F^\top \left( \frac{1}{T} F^\top F \right)^{-1} \frac{1}{\sqrt{T}} F^\top e_i + \frac{\sqrt{\delta}}{\sqrt{N}} \Lambda^\top \left( \frac{1}{N} \Lambda \hat{\Lambda} \right)^{-1} \frac{1}{\sqrt{N}} \Lambda^\top e_t + o_p (1)$

with $\delta = \min (N, T)$.

Furthermore

1. If $\frac{T}{N} \to 0$, then the asymptotic distribution of the loadings estimator is given by

   $\sqrt{T} \left( H^\top \hat{\Lambda}_i - \Lambda_i \right) \xrightarrow{D} N (0, \tilde{\Sigma}_F \Omega_{11,i} \tilde{\Sigma}_F^{-1})$

2. If $\frac{N}{T} \to 0$, then the asymptotic distribution of the factors is given by

   $\sqrt{N} \left( H^{-1} \hat{F}_t - F_t \right) \xrightarrow{D} N (0, \Lambda \hat{\Lambda}^{-1})$

The assumptions for RP-PCA in the strong factor model are slightly stronger than Assumption 3. More specifically we add assumption D.2, E.1 and E.2 to Bai (2003)'s assumptions and slightly modify A, C.3 and F. We denote by $\hat{F}_t = \frac{1}{T} \sum_{t=1}^{T} F_t$ and $\hat{e}_i = \frac{1}{T} \sum_{t=1}^{T} e_{t,i}$ the time-series mean of the factors and idiosyncratic component. The matrix $M_1 = I_T - \frac{1}{T} 1 1^\top$ demeans a time-series vector. The constant $M$ is finite and can change from line to line.

**Assumption 1: Strong Factor Model**

**A:** Factors: $E [ F_t^4 ] \leq M < \infty$ and $\frac{1}{T} \sum_{t=1}^{T} F_t \xrightarrow{p} \mu_F$ and $\frac{1}{T} \sum_{t=1}^{T} F_t W^2 F_t^\top \xrightarrow{p} \Sigma_F + (1 + \gamma) \mu_F \mu_F^\top$ which is a $K \times K$ positive definite matrix.

**B:** Factor loadings: $||\Lambda_i|| \leq \hat{\Lambda} < \infty$, and $|| \Lambda^\top \Lambda / N - \Sigma_\Lambda || \to 0$ for some $K \times K$ positive definite matrix $\Sigma_\Lambda$.

**C:** Time and cross-section dependence and heteroskedasticity: There exists a positive constant $M < \infty$ such that for all $N$ and $T$:
1. \( E[e_{t,i}] = 0, E[|e_{t,i}|^8] \leq M \).
2. \( E[N^{-1} \sum_{i=1}^{N} e_{s,i} e_{t,i}] = \xi(s,t), |\xi(s,s)| \leq M \) for all \( s \) and for every \( t \leq T \) it holds \( \sum_{s=1}^{T} |\xi(s,t)| \leq M \).
3. \( E[e_{t,i} e_{s,j}] = \tau_{ij,ts} \) with \( |\tau_{ij,ts}| \leq |\tau_{ij}| \) for some \( \tau_{ij} \) and for all \( s, t \) and for every \( i \leq N \) it holds \( \sum_{i=1}^{N} |\tau_{ij}| \leq M \).
4. \( E[e_{t,i} e_{s,j}] = \tau_{ij,ts} \) and \( (NT)^{-1} \sum_{t=1}^{T} \sum_{s=1}^{N} \sum_{i=1}^{N} \sum_{j=1}^{N} |\tau_{ij,ts}| \leq M \).
5. For every \((t,s)\), \( E\left[ \left| N^{-1/2} \sum_{i=1}^{N} (e_{s,i} e_{t,i}) - E[e_{s,i} e_{t,i}] \right|^4 \right] \leq M \).

D: Weak dependence between factors and idiosyncratic errors:
1. \( E\left[ \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} F_t e_{t,i} \right\|^2 \right] \leq M \).
2. \( E\left[ \frac{1}{N} \sum_{i=1}^{N} \left\| \sqrt{T} \hat{F} \hat{e}_i \right\|^2 \right] \leq M \).

E: Moments and Central Limit Theorem: There exists an \( M < \infty \) such that for all \( N \) and \( T \):
1. For each \( t \), \( E\left[ \left\| \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} F_s (e_{s,k} e_{t,k} - E[e_{s,k} e_{t,k}]) \right\|^2 \right] \leq M \) and \( E\left[ \left\| \frac{\sqrt{T}}{\sqrt{N}} \sum_{k=1}^{N} \hat{F} (\hat{e}_k e_{t,k} - E[\hat{e}_k e_{t,k}]) \right\|^2 \right] \leq M \).
2. The \( K \times K \) matrices satisfy \( E\left[ \left\| \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \sum_{i=1}^{N} F_t \Lambda^T_i e_{t,i} \right\|^2 \right] \leq M \) and \( E\left[ \left\| \frac{\sqrt{T}}{\sqrt{N}} \sum_{i=1}^{N} \hat{F} \Lambda^T_i \hat{e}_i \right\|^2 \right] \leq M \).
3. For each \( t \) as \( N \to \infty \):
   \[
   \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Lambda_i e_{t,i} \overset{d}{\to} N(0, \Gamma_t),
   \]
   where \( \Gamma_t = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \Lambda_i \Lambda_j^T E[e_{t,i} e_{t,j}] \).
4. For each \( i \) as \( T \to \infty \):
   \[
   \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} F_t e_{t,i}, \frac{1}{\sqrt{T}} \sum_{t=1}^{T} e_{t,i} \right) \overset{d}{\to} N(0, \Omega_i),
   \]
   where \( \Omega_i = p \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} E\left[ \begin{pmatrix} F_t F_s^T e_{s,j} e_{t,i} & F_t e_{s,j} e_{t,i} \\ F_s^T e_{s,j} e_{t,i} & e_{s,j} e_{t,i} \end{pmatrix} \right] \).

F: The eigenvalues of the \( K \times K \) matrix \( \Sigma (\Sigma_F + (\gamma + 1) \mu_F \mu_F^T) \) are distinct.

Theorem 1 is a consequence of Theorem 3:

**Theorem 1:** Asymptotic distribution of RP-PCA in strong factor model
Assume Assumption 1 holds. Define the estimator \( \hat{\Lambda} \) as the eigenvectors of the \( K \) largest eigenvalues of \( \frac{1}{NT} X^T W^2 X \) multiplied by \( \sqrt{N} \) and \( \hat{F} = X \hat{\Lambda} (\hat{\Lambda}^T \hat{\Lambda})^{-1} \) with \( W^2 = (I_T + \gamma \frac{11^T}{T}) \). Then:
1. If \( \min(N,T) \to \infty \), then for any \( \gamma \in [-1, \infty) \) the factors and loadings can be estimated consis-
tently pointwise.

2. If $\sqrt{T/N} \to 0$, then the asymptotic distribution of the loadings estimator is given by

$$
\sqrt{T} \left( H^\top \hat{\Lambda}_t - \Lambda_t \right) \stackrel{D}{\to} N(0, \Phi_t)
$$

$$
\Phi_t = (\Sigma + (y + 1)\mu_F \mu_F^\top)^{-1} \left( \Omega_{11,i} + y\mu_F \Omega_{21,i} + y^2 \mu_F \Omega_{22,i} \mu_F^\top \right) \left( \Sigma + (y + 1)\mu_F \mu_F^\top \right)^{-1}
$$

$$
H = \left( \frac{1}{T} F^\top W^2 F \right) \left( \frac{1}{N}\Lambda \Lambda \right) V_{TN}^{-1}
$$

and $V_{TN}$ is a diagonal matrix of the largest $K$ eigenvalues of $\frac{1}{NT} X^\top W^2 X$.

3. If $\sqrt{N/T} \to 0$, then the asymptotic distribution of the factors is not affected by the choice of $\gamma$.

4. For any choice of $\gamma \in [-1,\infty)$ the common components can be estimated consistently if $\min(N,T) \to \infty$. The asymptotic distribution of the common component depends on $\gamma$ if and only if $N/T$ does not go to zero. For $T/N \to 0$

$$
\sqrt{T} \left( \hat{C}_{t,i} - C_{t,i} \right) \stackrel{D}{\to} N \left( 0, F_t^\top \Phi_t F_t \right).
$$

Proof. We replace the returns, factors and idiosyncratic terms by their projected counterpart:

$$
\tilde{X} = XW \quad \tilde{F} = FW \quad \tilde{e} = eW \quad W = I_T + \frac{\sqrt{N}}{T} \mathbb{1} \mathbb{1}^\top \quad \tilde{y} = \sqrt{y + 1} - 1.
$$

Note that $WW = I_T + \frac{\gamma}{T} \mathbb{1} \mathbb{1}^\top$ and $\tilde{F}_t = F_t + \tilde{y} \tilde{F}$ and $\tilde{e}_{t,i} = e_{t,i} + \tilde{y} \tilde{e}_i$. We will show that under Assumption 1, the projected model with $\tilde{X}, \tilde{F}, \tilde{e}$ satisfies Assumption 3. Thus, the following expansions hold:

1. $\sqrt{T} \left( H^\top \tilde{\Lambda}_i - \Lambda_i \right) = \left( \frac{1}{T} F^\top W^2 F \right)^{-1} \frac{1}{\sqrt{T}} F^\top W^2 e_i + O_p \left( \frac{\sqrt{T}}{N} \right) + o_p(1)$

2. $\sqrt{N} \left( \tilde{\Lambda}_t - \Lambda_t \right) = \left( \frac{1}{N} \Lambda^\top \Lambda \right)^{-1} \frac{1}{\sqrt{N}} \Lambda^\top e_t + O_p \left( \frac{\sqrt{N}}{T} \right) + o_p(1)$

3. $\sqrt{\delta} \left( \hat{C}_{t,i} - C_{t,i} \right) = \frac{\sqrt{\delta}}{\sqrt{T}} F_t^\top \left( \frac{1}{T} F^\top W^2 F \right)^{-1} \frac{1}{\sqrt{T}} F^\top W^2 e_i + \frac{\sqrt{\delta}}{\sqrt{N}} \Lambda_i^\top \left( \frac{1}{N} \Lambda^\top \Lambda \right)^{-1} \frac{1}{\sqrt{N}} \Lambda^\top e_t + o_p(1)$

with $\delta = \min(N,T)$.

The distribution results are then an immediate consequence. In particular for $\sqrt{T/N} \to 0$ the asymptotic distribution of the loadings is described by the limit of $\left( \frac{1}{T} F^\top W^2 F \right)^{-1} \frac{1}{\sqrt{T}} F^\top W^2 e_i$. Note that

$$
\sqrt{T} F^\top W^2 e_i = \left( I_K, yF \right) \left( \frac{1}{T} \sum_{t=1}^T F_t e_{t,i} \right) \stackrel{D}{\to} N \left( 0, \left( I_K, y\mu_F \right) \left( \Omega_{11,i}, \Omega_{12,i}, \Omega_{21,i}, \Omega_{22,i} \right) \left( I_K, y\mu_F \right)^\top \right)
$$

and

$$
\frac{1}{T} F^\top W^2 F \to \Sigma_F + (y + 1)\mu_F \mu_F^\top.
$$

Now we show step by step that Assumptions A to F in Assumption 3 are satisfied for the projected
data:

A: \( E[\|\tilde{F}_t\|^4] \leq ME[\|F_t\|^4] \leq M < \infty \) and \( \frac{1}{T} \sum_{t=1}^{T} \tilde{F}_t \tilde{F}_t^\top \xrightarrow{P} \Sigma_F + \gamma \mu_F \mu_F^\top \) positive definite.

B: Factor loadings are not affected by the projection.

C: Time and cross-section dependence and heteroskedasticity: There exists a positive constant \( M < \infty \) (which can be different from line to line) such that for all \( N \) and \( T \):

1. \( E[\tilde{e}_{t,i}] = 0, E[|\tilde{e}_{t,i}|^8] \leq M \).

2. \( E[N^{-1} \sum_{i=1}^{N} \tilde{e}_{s,i} \tilde{e}_{t,i}] = E[N^{-1} \sum_{i=1}^{N} e_{s,i} e_{t,i}] + \tilde{y} \frac{1}{T} \sum_{s=1}^{N} E[N^{-1} \sum_{i=1}^{N} e_{s,i} e_{t,i}] + \tilde{y}^2 \frac{1}{T^2} \sum_{s=1}^{N} \sum_{i=1}^{N} E[N^{-1} \sum_{i=1}^{N} e_{s,i} e_{t,i}] \leq (1 + \tilde{y})^2 \xi(s,t) \) and hence \( \sum_{s=1}^{T} E[N^{-1} \sum_{i=1}^{N} \tilde{e}_{s,i} \tilde{e}_{t,i}] \leq M \).

3. \( E[\tilde{e}_{t,i} \tilde{e}_{t,j}] = E[e_{t,i} e_{t,j}] + \tilde{y} \frac{1}{T} \sum_{s=1}^{N} E[e_{s,i} e_{s,j}] + \tilde{y}^2 \frac{1}{T^2} \sum_{s=1}^{N} \sum_{i=1}^{N} E[e_{s,i} e_{s,j}] \leq M \tau_{ij} \) which yields \( \sum_{s=1}^{N} E[e_{t,i} e_{t,j}] \leq M \).

4. \( E[\tilde{e}_{t,i} \tilde{e}_{s,j}] = E[e_{t,i} e_{s,j}] + \tilde{y} \frac{1}{T} \sum_{s=1}^{N} E[e_{s,i} e_{t,j}] + \tilde{y}^2 \frac{1}{T^2} \sum_{s=1}^{N} \sum_{i=1}^{N} E[e_{s,i} e_{s,j}] + \tilde{y}^2 \frac{1}{T^2} \sum_{s=1}^{N} \sum_{i=1}^{N} E[e_{s,i} e_{h,j}] \) and thus

\[
(NT)^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} |E[\tilde{e}_{t,i} \tilde{e}_{s,j}]| \leq (NT)^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} |(1 + 2\tilde{y} + \tilde{y}^2) \tau_{ij,at}| \leq M.
\]

5. Define \( v_{s,t} = N^{-1/2} \sum_{i=1}^{N} (e_{s,i} e_{t,i}) - E[e_{s,i} e_{t,i}] \). We have

\[
E[|N^{-1/2} \sum_{i=1}^{N} (\tilde{e}_{s,i} \tilde{e}_{t,i}) - E[\tilde{e}_{s,t} \tilde{e}_{t,i}]|^4] \]

\[
\leq E[|v_{s,t}|^4] + \tilde{y}^4 \frac{1}{T^2} \sum_{t_1}^{T} \sum_{t_2}^{T} \sum_{t_3}^{T} E[|v_{s,t_1} v_{s,t_2} v_{s,t_3} v_{s,t_4}|] \]

\[
+ \tilde{y}^8 \frac{1}{T^4} \sum_{s_1}^{T} \sum_{s_2}^{T} \sum_{s_3}^{T} \sum_{s_4}^{T} E[|v_{s_1,t_1} v_{s_2,t_2} v_{s_3,t_3} v_{s_4,t_4}|] \leq M
\]

as \( E[|v_{s_1,t_1} v_{s_2,t_2} v_{s_3,t_3} v_{s_4,t_4}|] \leq \max_{s,t} E[|v_{s,t}|^4] \leq M \).

D: Weak dependence between factors and idiosyncratic errors: Note that

\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tilde{F}_t \tilde{e}_{t,i} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} F_t e_{t,i} + \tilde{y} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} F_t \tilde{e}_i + \tilde{y} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tilde{F}_t \tilde{e}_i + \tilde{y}^2 \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tilde{F}_t \tilde{e}_i \]

\[
= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} F_t e_{t,i} + (2\tilde{y} + \tilde{y}^2) \sqrt{T} \tilde{e}_i.
\]

Thus D.1 and D.2 imply \( E\left[ \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tilde{F}_t \tilde{e}_{t,i} \right]^2 \leq M \).
E: Moments and Central Limit Theorem:

1. We have

\[
\frac{1}{\sqrt{T}} \sum_{s=1}^{T} \tilde{F}_s(\tilde{e}_{s,k}\tilde{e}_{t,k} - E[\tilde{e}_{s,k}\tilde{e}_{t,k}])
\]

\[
= \frac{1}{\sqrt{T}} \sum_{s=1}^{T} F_s(e_{s,k}e_{t,k} - E[e_{s,k}e_{t,k}]) + \tilde{\gamma} \frac{1}{\sqrt{T}} \sum_{s=1}^{T} \tilde{F}(e_{s,k}\tilde{e}_{t,k} - E[e_{s,k}\tilde{e}_{t,k}])
\]

\[
+ \tilde{\gamma}^2 \frac{1}{\sqrt{T}} \sum_{s=1}^{T} F_s(\tilde{e}_{s,k}e_{t,k} - E[\tilde{e}_{s,k}e_{t,k}]) + \tilde{\gamma}^2 \frac{1}{\sqrt{T}} \sum_{s=1}^{T} \tilde{F}(e_{s,k}\tilde{e}_{t,k} - E[e_{s,k}\tilde{e}_{t,k}])
\]

\[
+ \tilde{\gamma}^2 \frac{1}{\sqrt{T}} \sum_{s=1}^{T} F_s(\tilde{e}_{s,k}e_{t,k} - E[\tilde{e}_{s,k}e_{t,k}])
\]

\[
\leq \frac{1}{\sqrt{T}} \sum_{s=1}^{T} F_s(e_{s,k}e_{t,k} - E[e_{s,k}e_{t,k}]) + \frac{1}{\sqrt{T}} \sum_{s=1}^{T} F_s(e_{s,k}\tilde{e}_{t,k} - E[e_{s,k}\tilde{e}_{t,k}])
\]

\[
+ M\sqrt{T}\tilde{F}(e_{s,k}e_{t,k} - E[e_{s,k}e_{t,k}]) + \frac{1}{\sqrt{T}} \sum_{s=1}^{T} F_s(e_{s,k}\tilde{e}_{t,k} - E[e_{s,k}\tilde{e}_{t,k}]).
\]

The two assumptions in E.1 therefore imply

\[
E \left[ \left\| \frac{1}{\sqrt{NT}} \sum_{s=1}^{T} \sum_{k=1}^{N} \tilde{F}_s(\tilde{e}_{s,k}\tilde{e}_{t,k} - E[\tilde{e}_{s,k}\tilde{e}_{t,k}]) \right\|^2 \right] \leq M.
\]

2. The relevant term can be decomposed into

\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tilde{F}_t \Lambda_i^\top \tilde{e}_{t,i} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} F_t \Lambda_i^\top e_{t,i} + \tilde{\gamma} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} F_t \Lambda_i^\top \tilde{e}_{t,i} + \frac{1}{\sqrt{T}} \sum_{t=1}^{T} F \Lambda_i^\top \tilde{e}_{t,i}
\]

\[
+ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} F_t \Lambda_i^\top e_{t,i} + (2\tilde{\gamma} + \tilde{\gamma}^2) \sqrt{T} \tilde{F} \Lambda_i^\top \tilde{e}_{t,i}.
\]

Hence, under the two assumptions of E.2 we have

\[
E \left[ \left\| \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \sum_{i=1}^{N} \tilde{F}_t \Lambda_i^\top \tilde{e}_{t,i} \right\|^2 \right] \leq M.
\]

3. The cross-sectional central limit theorem is not affected by the projection as

\[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Lambda_i \tilde{e}_{t,i} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Lambda_i e_{t,i} + \tilde{\gamma} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Lambda_i \tilde{e}_{t,i} + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Lambda_i \tilde{e}_{t,i}
\]

\[
\text{and } \frac{1}{T\sqrt{N}} \sum_{i=1}^{N} \sum_{t=1}^{T} \Lambda_i e_{t,i} = O_p \left( \frac{1}{\sqrt{T}} \right) \text{ by Assumption C.4.}
\]

4. As discussed before the time-series central limit theorem takes the form

\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tilde{F}_t \tilde{e}_{t,i} \overset{D}{\rightarrow} N \left( 0, \left( \Omega_{11,i} + \gamma \Omega_{21,i} + \gamma \Omega_{12,i} + \gamma^2 \Omega_{22,i} \right) \right).
\]

F: $K \times K$ matrix $\Sigma_F + (y + 1)\mu_F^\top \mu_F$ is the corresponding limiting matrix for the projected data
which has distinct eigenvalues by Assumption F.

In order to get a better intuition we consider an example with i.i.d. residuals over time. This simplified model is more comparable to the weak factor model.

Example 1: Simplified Strong Factor Model

1. **Rate:** Assume that $\frac{N}{T} \to c$ with $0 < c < \infty$.

2. **Factors:** The factors $F$ are uncorrelated among each other and are independent of $e$ and $\Lambda$ and have bounded first four moments.

   \[
   \hat{\mu}_F := \frac{1}{T} \sum_{t=1}^{T} F_t \to \mu_F \quad \hat{\Sigma}_F := \frac{1}{T} \sum_{t=1}^{T} F_t F_t^\top - \hat{\mu}_F \hat{\mu}_F^\top \to \Sigma_F = \begin{pmatrix} \sigma^2_{F1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma^2_{FK} \end{pmatrix}.
   \]

3. **Loadings:** $\Lambda^\top \Lambda / N \overset{p}{\to} I_K$ and all loadings are bounded. The loadings are independent of the factors and residuals.

4. **Residuals:** Residual matrix can be represented as $e = \epsilon \Sigma$ with $\epsilon_{t,i} \overset{i.i.d.}{\sim} N(0, 1)$. All elements and all row sums of absolute values of $\Sigma$ are bounded.

Corollary 1: Simplified Strong Factor Model:

The assumptions of example 1 hold. The factors and loadings can be estimated consistently. The asymptotic distribution of the factors is not affected by $\gamma$. The asymptotic distribution of the loadings is given by

\[
\sqrt{T} \left( H^\top \hat{\Lambda}_i - \Lambda_i \right) \overset{D}{\to} N(0, \Omega_i),
\]

where $E[e_{t,i}^2] = \sigma^2_{e_i}$ and

\[
\Phi_i = \sigma^2_{e_i} \left( \Sigma_F + (1 + \gamma) \mu_F \mu_F^\top \right)^{-1} \left( \Sigma_F + (1 + \gamma)^2 \mu_F \mu_F^\top \right) \left( \Sigma_F + (1 + \gamma) \mu_F \mu_F^\top \right)^{-1}.
\]

The optimal choice for the weight minimizing the asymptotic variance is $\gamma = 0$. Choosing $\gamma = -1$, i.e. the covariance matrix for factor estimation, is not efficient.

Proof. We show that the assumptions in Example 1 satisfy the conditions of Assumption 1. Conditions A, B, C.1 and C.2 are trivially satisfied. Condition C.3 is the boundedness of the row sums of absolute values of $\Sigma$. C.4 follows from C.3 and the independence of the residuals over time. C.5 requires more work. As the residuals are independent over time, it is sufficient to deal with $E\left[ \frac{1}{\sqrt{N}} \sum_{t=1}^{N} \left( e_{t,i}^2 - \Sigma_{ii} \right) \right]$. Because $\frac{1}{\sqrt{N}} \sum_{t=1}^{N} (e_{t,i}^2 - \Sigma_{ii})$ is normally distributed it is sufficient to show
that its variance is finite in the limit. Note that the variance equals \( \frac{1}{N} \Sigma \otimes \Sigma \). As \( \sum_{i=1}^{N} |\Sigma_{i,j}| \) is bounded we also have that \( \sum_{i=1}^{N} |\Sigma_{i,j}|^2 \) is bounded. If the number of non-zero elements in \( \Sigma \) is growing with \( N \) then \( |\Sigma_{i,j}| < 1 \) except for finitely many elements. Therefore sum of squared elements has to be bounded. D.1 and D.2 follow from the independence of the factors with the residuals and the time-series independence of the residuals. Note that \( E[\frac{1}{T} \sum_{t=1}^{T} e_{t,i}^2] \) is bounded in the limit. E.1 follows from similar arguments as C.5. E.2 uses again the independence of factors and residuals and \( \frac{1}{N} \Lambda^T \Sigma \Lambda \leq M \). We don’t need to show the central limit theorem in E.3 but it is sufficient to note that \( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Lambda_{i} e_{t,i} \) does not depend on \( \gamma \). E.4 is a martingale difference sequence central limit theorem. Note that

\[
\Omega_i = \begin{pmatrix} (\Sigma_F + \mu_F \mu_T^T) \sigma_{\epsilon_i}^2 & \mu_F \sigma_{\epsilon_i}^2 \\ \mu_F^T \sigma_{\epsilon_i}^2 & \sigma_{\epsilon_i}^2 \end{pmatrix}
\]

which immediately yields

\[
\Phi_i = \sigma_{\epsilon_i}^2 (\Sigma_F + (1 + \gamma) \mu_F \mu_T^T)^{-1} \left( \Sigma_F + (1 + \gamma)^2 \mu_F \mu_T^T \right) (\Sigma_F + (1 + \gamma) \mu_F \mu_T^T)^{-1}.
\]

Taking the derivative of \( \Phi_i(\gamma) \) and setting it to zero yields the first-order condition

\[
- \left( \Sigma_F + (1 + \gamma)^2 \mu_F \mu_T^T \right) \mu_F \mu_T^T \left( \Sigma_F + (1 + \gamma) \mu_F \mu_T^T \right) + \left( \Sigma_F + (1 + \gamma) \mu_F \mu_T^T \right) (2(1 + \gamma) \mu_F \mu_T^T) \left( \Sigma_F + (1 + \gamma) \mu_F \mu_T^T \right) - \left( \Sigma_F + (1 + \gamma) \mu_F \mu_T^T \right) \mu_F \mu_T^T \left( \Sigma_F + (1 + \gamma)^2 \mu_F \mu_T^T \right) = 0
\]

For \( \mu \neq 0 \) this has the unique solution \( \gamma = 0 \) with optimal covariance matrix \( \Phi_i = \sigma_{\epsilon_i}^2 (\Sigma_F + \mu_F \mu_T^T)^{-1} \) which is the smallest asymptotic variance. \( \square \)
2. Simulation

2.1. Multi-Factor Model

Figure 1: $N = 74, T = 650$: Correlation of estimated rotated factors with true factors in-sample and out-of-sample for different variances and Sharpe-ratios of the fourth factor and for different RP-weights $\gamma$. 
Figure 2: $N = 74, T = 650$: Sharpe ratios of estimated rotated factors in-sample and out-of-sample for different variances and Sharpe-ratios of the fourth factor and for different RP-weights $\gamma$. 

1. Factor SR (IS) for $\sigma_F^2=0.03$    1. Factor SR (OOS) for $\sigma_F^2=0.03$    1. Factor SR (IS) for $\sigma_F^2=0.1$    1. Factor SR (OOS) for $\sigma_F^2=0.1$

2. Factor SR (IS) for $\sigma_F^2=0.03$    2. Factor SR (OOS) for $\sigma_F^2=0.03$    2. Factor SR (IS) for $\sigma_F^2=0.1$    2. Factor SR (OOS) for $\sigma_F^2=0.1$

3. Factor SR (IS) for $\sigma_F^2=0.03$    3. Factor SR (OOS) for $\sigma_F^2=0.03$    3. Factor SR (IS) for $\sigma_F^2=0.1$    3. Factor SR (OOS) for $\sigma_F^2=0.1$

4. Factor SR (IS) for $\sigma_F^2=0.03$    4. Factor SR (OOS) for $\sigma_F^2=0.03$    4. Factor SR (IS) for $\sigma_F^2=0.1$    4. Factor SR (OOS) for $\sigma_F^2=0.1$
2.2. Single-Factor Model with $N = 370$ and $T = 650$

Figure 3: $N = 370$, $T = 650$: Correlations and Sharpe-ratios as a function of the RP-weight $\gamma$ for different variances and Sharpe-ratios. The residuals have cross-sectional correlation defined by the band-diagonal matrix.
Figure 4: $N = 370$, $T = 650$: Correlations and Sharpe-ratios as a function of the RP-weight $\gamma$ for different variances and Sharpe-ratios. The residuals have the empirical residual correlation matrix.
2.3. Single-Factor Model with $N = 74$ and $T = 650$

![Graphs showing correlations and Sharpe-ratios as a function of the RP-weight $\gamma$ for different variances and Sharpe-ratios.]

Figure 5: $N = 74, T = 650$: Correlations and Sharpe-ratios as a function of the RP-weight $\gamma$ for different variances and Sharpe-ratios.
2.4. Single-Factor Model with $N = 25$ and $T = 240$

![Graph showing correlations and Sharpe-ratios as a function of the RP-weight $\gamma$ for different variances and Sharpe-ratios.]

Figure 6: $N = 25$, $T = 240$: Correlations and Sharpe-ratios as a function of the RP-weight $\gamma$ for different variances and Sharpe-ratios.
2.5. Pricing Errors for Single-Factor Model

Figure 7: $N = 370, T = 650$: Root-mean-squared pricing errors as a function of the RP-weight $\gamma$ for different variances and Sharpe-ratios.

Figure 8: $N = 74, T = 650$: Root-mean-squared pricing errors as a function of the RP-weight $\gamma$ for different variances and Sharpe-ratios.
Figure 9: $N = 25, T = 240$: Root-mean-squared pricing errors as a function of the RP-weight $\gamma$ for different variances and Sharpe-ratios.