Supplementary Material

Linear Regression for Panel with Unknown Number of Factors as Interactive Fixed Effects

Hyungsik Roger Moon* Martin Weidner‡

February 7, 2015

Contents

*Department of Economics and USC Dornsife INET, University of Southern California, Los Angeles, CA 90089-0253. Email: moonr@usc.edu. Department of Economics, Yonsei University, Seoul, Korea.

‡Corresponding author. Department of Economics, University College London, Gower Street, London WC1E 6BT, U.K., and CeMMaP. Email: m.weidner@ucl.ac.uk.

S.1 Proofs for Main Text Results

Proof of Theorem 2.1 (Identifictaion). Let $Q(\beta, \Lambda, F) = \mathbb{E}\left(\|Y - \beta \cdot X - \Lambda F'\|_{HS}^2\right)$. Existence of $Q(\beta, \Lambda, F)$ is guaranteed by Assumption ID(i). The statement of the theorem follows if we can show that $Q(\beta, \Lambda, F)$ is uniquely minimized at $\beta = \beta^0$ and $\Lambda F' = \lambda^0 f^{0\prime}$. We have

$$
\begin{aligned}
Q(\beta, \Lambda, F) &= \mathbb{E} \text{ Tr}\left[(Y - \beta \cdot X - \Lambda F')(Y - \beta \cdot X - \Lambda F')'\right] \\
&= \mathbb{E} \text{ Tr}\left[\left(\lambda^0 f^{0\prime} - \Lambda F' - (\beta - \beta^0) \cdot X + e\right)\left(\lambda^0 f^{0\prime} - \Lambda F' - (\beta - \beta^0) \cdot X + e\right)'\right] \\
&= \underbrace{\mathbb{E} \text{ Tr}\left[\left(\lambda^0 f^{0\prime} - \Lambda F' - (\beta - \beta^0) \cdot X\right)\left(\lambda^0 f^{0\prime} - \Lambda F' - (\beta - \beta^0) \cdot X\right)'\right]}_{\equiv Q^*(\beta, \Lambda, F)} + \mathbb{E} \text{ Tr}\left(ee'\right).
\end{aligned}
\tag{S.1}
$$

Here, we used the model, and in the last step we employed Assumption ID(ii). Next, we derive a lower bound on $Q^*(\beta, \Lambda, F)$. We have

$$
\begin{aligned}
Q^*(\beta, \Lambda, F) &\geq \mathbb{E} \text{ Tr}\left[\left(\lambda^0 f^{0\prime} - \Lambda F' - (\beta - \beta^0) \cdot X\right) M_F \left(\lambda^0 f^{0\prime} - \Lambda F' - (\beta - \beta^0) \cdot X\right)'\right] \\
&= \mathbb{E} \text{ Tr}\left[M_F \left(\lambda^0 f^{0\prime} - \Lambda F' - (\beta - \beta^0) \cdot X\right)' \left(\lambda^0 f^{0\prime} - \Lambda F' - (\beta - \beta^0) \cdot X\right) M_F\right] \\
&\geq \mathbb{E} \text{ Tr}\left[M_F \left(\lambda^0 f^{0\prime} - \Lambda F' - (\beta - \beta^0) \cdot X\right)' M_{\lambda^0} \left(\lambda^0 f^{0\prime} - \Lambda F' - (\beta - \beta^0) \cdot X\right) M_F\right] \\
&= \mathbb{E} \text{ Tr}\left[M_F \left((\beta - \beta^0) \cdot X\right)' M_{\lambda^0} \left((\beta - \beta^0) \cdot X\right)\right] \\
&= (\beta - \beta^0)' \left\{\mathbb{E}[x'(M_F \otimes M_{\lambda^0})x]\right\} (\beta - \beta^0).
\end{aligned}
\tag{S.2}
$$

From this and Assumption ID(iii) we conclude that $Q^*(\beta, \Lambda, F) > 0$ for all $\beta \neq \beta^0$. On the other hand, we have $Q^*(\beta^0, \lambda^0, f^0) = 0$. Thus, every minimum of $Q^*(\beta, \Lambda, F)$ satisfies $\beta = \beta^0$. Furthermore, at $\beta = \beta^0$ we have $Q^*(\beta^0, \Lambda, F) = \|\lambda^0 f^{0\prime} - \Lambda F'\|_{HS}^2$, which is zero if and only if $\Lambda F' = \lambda^0 f^{0\prime}$. The minima of $Q^*(\beta, \Lambda, F)$ therefore satisfy $\beta = \beta^0$ and $\Lambda F' = \lambda^0 f^{0\prime}$. Since $Q^*(\beta, \Lambda, F)$ and $Q(\beta, \Lambda, F)$ only differ by a constant the same result holds for $Q(\beta, \Lambda, F)$. Notice that the result that the optimal Λ and F satisfy $\Lambda F' = \lambda^0 f^{0\prime}$ implies that $\text{rank}(\Lambda F') = R^0$, i.e. the true number of factors R^0 is also identified. ■

Proof of Theorem 3.1 (Main Result). Follows from Theorem A.3 and Lemma A.4. ■

Proof of Theorem 3.2 (Consistency of Bias and Variance Estimators). See Section S.5 below. ■

Proof of Theorem 4.1 (Consistency). We first establish a lower bound on $\mathcal{L}_{NT}^R(\beta)$. Let $\Delta\beta = \beta - \beta^0$. Consider the definition of $\mathcal{L}_{NT}^R(\beta)$ in equation (3.3) and plug in the model

$Y = \beta \cdot X + \lambda^0 f^{0\prime} + e$. We have

$$
\begin{aligned}
\mathcal{L}^R_{NT}(\beta) &= \min_{\{\Lambda \in \mathbb{R}^{N\times R},\, F \in \mathbb{R}^{T\times R}\}} \frac{1}{NT} \mathrm{Tr}\left[\left(\Delta\beta \cdot X + e + \lambda^0 f^{0\prime} - \Lambda F'\right) \left(\Delta\beta \cdot X + e + \lambda^0 f^{0\prime} - \Lambda F'\right)' \right] \\
&\geq \min_{\left\{\tilde\Lambda \in \mathbb{R}^{N\times (R+R^0)},\, \tilde F \in \mathbb{R}^{T\times (R+R^0)}\right\}} \frac{1}{NT} \mathrm{Tr}\left[\left(\Delta\beta \cdot X + e - \tilde\Lambda \tilde F'\right) \left(\Delta\beta \cdot X + e - \tilde\Lambda \tilde F'\right)' \right] \\
&= \frac{1}{NT} \min_{\tilde F \in \mathbb{R}^{T\times (R+R^0)}} \mathrm{Tr}\left[\left(\Delta\beta \cdot X + e\right) M_{\tilde F} \left(\Delta\beta \cdot X + e\right)' \right] \\
&= \frac{1}{NT} \min_{\tilde F \in \mathbb{R}^{T\times (R+R^0)}} \Bigg\{ \mathrm{Tr}\left[\left(\Delta\beta \cdot X\right) M_{\tilde F} \left(\Delta\beta \cdot X\right)' \right] + \mathrm{Tr}\left(ee'\right) - \mathrm{Tr}\left(e P_{\tilde F} e'\right) \\
&\qquad\qquad + 2\mathrm{Tr}\left[\left(\Delta\beta \cdot X\right) e'\right] - 2\mathrm{Tr}\left[\left(\Delta\beta \cdot X\right) P_{\tilde F} e'\right] \Bigg\} \\
&\geq \frac{1}{NT} \Bigg\{ \sum_{r=R+R^0+1}^{T} \mu_r \left[(\Delta\beta \cdot X)'(\Delta\beta \cdot X) \right] + \mathrm{Tr}\left(ee'\right) - 2(R+R^0)\|e\|^2 \\
&\qquad\qquad + 2\mathrm{Tr}\left[\left(\Delta\beta \cdot X\right) e'\right] - 2(R+R^0)\|e\| \|\Delta\beta \cdot X\| \Bigg\} \\
&\geq b \, \|\Delta\beta\|^2 + \frac{1}{NT} \mathrm{Tr}\left(ee'\right) + \mathcal{O}_P\left(\frac{1}{\min(N,T)} \right) + \mathcal{O}_P\left(\frac{\|\Delta\beta\|}{\sqrt{\min(N,T)}} \right). \qquad \text{(S.3)}
\end{aligned}
$$

Here, we applied the inequality $|\mathrm{Tr}(A)| \leq \mathrm{rank}(A)\|A\|$ with $A = \left(\Delta\beta \cdot X\right) P_{\widetilde F} e'$ and also with $A = e P_{\widetilde F} e'$. We also used that $\min_{\tilde F} \mathrm{Tr}\left[\left(\Delta\beta \cdot X\right) M_{\widetilde F} \left(\Delta\beta \cdot X\right)' \right] = \sum_{r=R+R^0+1}^{T} \mu_r \left[(\Delta\beta \cdot X)'(\Delta\beta \cdot X) \right]$, which follows by the same logic as equation (3.3) in the main text. In the last step of (S.3) we applied Assumptions SN, EX and NC.

Next, we establish an upper bound on $\mathcal{L}^R_{NT}(\beta^0)$. Since $R \geq R^0$ we can choose $\Lambda\, F' = \lambda^0 f^{0\prime}$ in the minimization problem in the first line of equation (3.3), and therefore

$$
\begin{aligned}
\mathcal{L}^R_{NT}(\beta^0) \; &= \min_{\{\Lambda \in \mathbb{R}^{N\times R},\, F \in \mathbb{R}^{T\times R}\}} \frac{1}{NT} \; \left\| e + \lambda^0 f^{0\prime} - \Lambda\, F' \right\|^2_{HS} \\
&\leq \frac{1}{NT} \, \|e\|^2_{HS} = \frac{1}{NT} \, \mathrm{Tr}\left(ee'\right). \qquad \text{(S.4)}
\end{aligned}
$$

Since we could choose $\beta = \beta^0$ in the minimization of β, the optimal $\widehat\beta_R$ needs to satisfy $\mathcal{L}^R_{NT}(\widehat\beta_R) \leq \mathcal{L}^R_{NT}(\beta^0)$. Together with (S.3) and (S.4) this gives

$$
b \, \|\widehat\beta_R - \beta^0\|^2 + \mathcal{O}_P\left(\frac{\|\widehat\beta_R - \beta^0\|}{\sqrt{\min(N,T)}} \right) + \mathcal{O}_P\left(\frac{1}{\min(N,T)} \right) \; \leq \; 0\,. \qquad \text{(S.5)}
$$

From this it follows that $\|\widehat\beta_R - \beta^0\| = \mathcal{O}_P\left(\min(N,T)^{-1/2}\right)$, which is what we wanted to show. ■

Proof of Theorem 4.2 (Quadratic Approximation of $\mathcal{L}^0_{NT}(\beta)$). See Section S.2 below. ■

Proof of Corollary 4.3 (Asymptotic Characterization of $\widehat{\beta}_{R^0}$).
Define $\gamma \equiv W^{-1}\left(C^{(1)} + C^{(2)}\right)/\sqrt{NT}$. Applying Theorem 4.2 we obtain

$$\mathcal{L}^0_{NT}\left(\widehat{\beta}_{R^0}\right) = \mathcal{L}^0_{NT}(\beta^0) + \left(\widehat{\beta}_{R^0} - \beta^0 - \gamma\right)' W \left(\widehat{\beta}_{R^0} - \beta^0 - \gamma\right) - \gamma' W \gamma + \mathcal{L}^{0,\mathrm{rem}}_{NT}(\widehat{\beta}_{R^0}),$$
$$\mathcal{L}^0_{NT}\left(\beta^0 + \gamma\right) = \mathcal{L}^0_{NT}(\beta^0) - \gamma' W \gamma + \mathcal{L}^{0,\mathrm{rem}}_{NT}(\beta^0 + \gamma). \tag{S.6}$$

The first equation above is obtained by completing the square and using the definition of γ, while the second equation is just a special case of the first. Applying the above to the inequality $\mathcal{L}^0_{NT}(\widehat{\beta}_{R^0}) \leq \mathcal{L}^0_{NT}\left(\beta^0 + \gamma\right)$ gives

$$\left(\widehat{\beta}_{R^0} - \beta^0 - \gamma\right)' W \left(\widehat{\beta}_{R^0} - \beta^0 - \gamma\right) \leq \mathcal{L}^{0,\mathrm{rem}}_{NT}(\beta^0 + \gamma) - \mathcal{L}^{0,\mathrm{rem}}_{NT}(\widehat{\beta}_{R^0}). \tag{S.7}$$

We have $W \geq \mu_K(W)\mathbb{1}_K$, and using Assumption NC we find for $R = R^0$ that

$$\begin{aligned}
\mu_K(W) &= \min_{\{\alpha\in\mathbb{R}^K, \|\alpha\|=1\}} \alpha' W \alpha \\
&= \min_{\{\alpha\in\mathbb{R}^K, \|\alpha\|=1\}} \frac{1}{NT} \mathrm{Tr}\left(M_{\lambda^0}(\alpha\cdot X) M_{f^0} (\alpha\cdot X)'\right) \\
&= \min_{\{\alpha\in\mathbb{R}^K, \|\alpha\|=1\}} \frac{1}{NT} \mathrm{Tr}\left(M_{f^0}(\alpha\cdot X)' M_{\lambda^0} (\alpha\cdot X) M_{f^0}\right) \\
&\geq \frac{1}{NT} \sum_{r=2R^0+1}^{T} \mu_r\left[(\alpha\cdot X)'(\alpha\cdot X)\right] \geq b\,, \qquad \text{wpa1},
\end{aligned} \tag{S.8}$$

and therefore $W^{-1} \leq \mathbb{1}_K/b$ wpa1. Using Assumption SN we find

$$|C^{(1)}_k| \leq \left|\frac{1}{\sqrt{NT}} \mathrm{Tr}\left(X_k e'\right)\right| + \frac{2R^0}{\sqrt{NT}} \|X_k\| \|e\| = \mathcal{O}_P\left(\sqrt{\max(N,T)}\right),$$
$$|C^{(2)}_k| \leq \frac{9R^0}{2\sqrt{NT}} \|e\|^2 \|X_k\| \left\|\lambda^0 \left(\lambda^{0\prime}\lambda^0\right)^{-1} \left(f^{0\prime} f^0\right)^{-1} f^{0\prime}\right\| = \mathcal{O}_P\left(1\right), \tag{S.9}$$

and therefore we have $\gamma = \mathcal{O}_P[(1+\|C^{(1)}\|)/\sqrt{NT}] = o_P(1)$. We also know $\|\widehat{\beta}_{R^0} - \beta^0\| = o_P(1)$ from Theorem 4.1. Thus, the bound on the remainder in Theorem 4.2 becomes applicable and we have

$$\begin{aligned}
\mathcal{L}^{0,\mathrm{rem}}_{NT}(\beta^0 + \gamma) - \mathcal{L}^{0,\mathrm{rem}}_{NT}(\widehat{\beta}_{R^0}) &\leq o_P\left(\frac{1}{NT}\right)\left[\left(1 + \sqrt{NT}\gamma\right)^2 + \left(1 + \sqrt{NT}\|\widehat{\beta}_{R^0} - \beta^0\|\right)^2\right] \\
&= o_P\left(\frac{1}{NT}\right)\left\{\mathcal{O}_P\left[(1+\|C^{(1)}\|)^2\right] + \left(1 + \sqrt{NT}\|\widehat{\beta}_{R^0} - \beta^0\|\right)^2\right\}.
\end{aligned} \tag{S.10}$$

Applying this, and (S.7), and $W^{-1} \leq 1/b$, and the inequality $\sqrt{(x+y)} \leq \sqrt{x} + \sqrt{y}$, which holds for all non-negative real number x, y, we find that

$$\sqrt{NT}\left\|\widehat{\beta}_{R^0}-\beta^0-\gamma\right\| \le o_P\left(1+\|C^{(1)}\|\right)+o_P\left(1+\sqrt{NT}\|\widehat{\beta}_{R^0}-\beta^0\|\right). \tag{S.11}$$

Since $\gamma=\mathcal{O}_P[(1+\|C^{(1)}\|)/\sqrt{NT}]$ it follows from this that $\sqrt{NT}\|\widehat{\beta}_{R^0}-\beta^0\|=\mathcal{O}_P\left(1+\|C^{(1)}\|\right)$, and therefore

$$\sqrt{NT}\left\|\widehat{\beta}_{R^0}-\beta^0-\gamma\right\| \le o_P\left(1+\|C^{(1)}\|\right), \tag{S.12}$$

which is what we wanted to show. ■

Proof of Example in Section 4.3 (Counter Example for $\sqrt{NT}$ Convergence Rate).

Consider the DGP and asymptotic as described in the example in Section 4.3. Let $\mathcal{L}^1_{NT}(\beta)$ be the profile objective function for $R=1$, defined in (3.3). We want to show that for any sequence $\Delta_{NT}>0$ with $\Delta_{NT}=o(N^{-1/2})$ we have

$$\min_{\beta\in\mathbb{R}}\mathcal{L}^1_{NT}(\beta) < \min_{\beta\in[\beta^0-\Delta_{NT},\beta^0+\Delta_{NT}]}\mathcal{L}^1_{NT}(\beta), \qquad \text{wpa1}. \tag{S.13}$$

This implies that $\left\|\widehat{\beta}_1-\beta^0\right\|$ cannot converge to zero at a faster than $\sqrt{N}$ rate.

What is left to do is to proof (S.13). We decompose $Y-\beta\cdot X=e-(\beta-\beta^0)X=e_1(\beta)+e_2(\beta)$, where

$$\begin{aligned}
e_1(\beta) &= \frac{c}{N}\lambda_x(M_{f_x}u'\lambda_x)'+\frac{c}{T}(M_{\lambda_x}uf_x)f_x'-(\beta-\beta^0)\lambda_x f_x', \\
e_2(\beta) &= \widetilde{u}+\frac{c^2(\lambda_x'uf_x)}{NT}\lambda_x f_x'+\frac{c}{N}\lambda_x\lambda_x' uP_{f_x}+\frac{c}{T}P_{\lambda_x}uf_xf_x',
\end{aligned} \tag{S.14}$$

with $\widetilde{u}=u-a(\beta-\beta^0)\widetilde{X}$. Since $\|\lambda_x\|=\mathcal{O}(\sqrt{N})$, $\|f_x\|=\mathcal{O}(\sqrt{T})$, and $\lambda_x'uf_x=\mathcal{O}_P(\sqrt{NT})$ we have $\|e_2(\beta)-\widetilde{u}\|=o_P(N)$. The matrix $\widetilde{u}$ has iid normal entries with mean zero and variance $1+a^2(\beta-\beta^0)^2$. According to Geman (1980) we thus have $\|\widetilde{u}\|^2=(1+a^2(\beta-\beta^0)^2)(\sqrt{N}+\sqrt{T})^2+o_P(N)$. Thus, as $N,T\to\infty$ at the same rate we have

$$\|e_2(\beta)\|^2 \le (1+a^2(\beta-\beta^0)^2)(\sqrt{N}+\sqrt{T})^2+o_P(N). \tag{S.15}$$

Note that $\operatorname{rank}(e_1(\beta))=2$, which implies that e_1 can be written as $e_1=A\widetilde{e}_1B'$, where A is an $N\times 2$ matrix satisfying $A'A=\mathbb{1}_2$, B is a $T\times 2$ matrix satisfying $B'B=\mathbb{1}_2$, and $\widetilde{e}_1$ is a 2×2 matrix, namely

$$\widetilde{e}_1=\begin{pmatrix} (\beta-\beta^0)\|\lambda_x f_x'\| & \left\|\frac{c}{N}\lambda_x(M_{f_x}u'\lambda_x)'\right\| \\ \left\|\frac{c}{T}(M_{\lambda_x}uf_x)f_x'\right\| & 0\end{pmatrix}. \tag{S.16}$$

Using this characterization of e_1 as well as $\|\lambda_x\|^2=N+o(N)$, $\|f_x\|^2=T+o(T)$, $\|M_{f_x}u'\lambda_x\|^2=$

$NT + o_P(NT)$, and $\|M_{\lambda_x} u f_x\|^2 = NT + o_P(NT)$, we find

$$
\begin{aligned}
& \|e_1(\beta)\|^2 \\
& = \mu_1 \left[e_1(\beta)' e_1(\beta) \right] = \mu_1 \left[\widetilde{e}_1(\beta)' \widetilde{e}_1(\beta) \right] \\
& = \mu_1 \left[\begin{pmatrix} \|f_x\|^2 \left(\frac{c^2 \|M_{\lambda_x} u f_x\|^2}{T^2} + \|\lambda_x\|^2 (\beta - \beta^0)^2 \right) & \frac{c \|\lambda_x\|^2 \|f_x\| \|M_{f_x} u' \lambda_x\| (\beta - \beta^0)}{N} \\ \frac{c \|\lambda_x\|^2 \|f_x\| \|M_{f_x} u' \lambda_x\| (\beta - \beta^0)}{N} & \frac{c^2 \|\lambda_x\|^2 \|M_{f_x} u' \lambda_x\|^2}{N^2} \end{pmatrix} \right] \\
& = \mu_1 \left[\begin{pmatrix} c^2 N + NT(\beta - \beta^0)^2 & cT\sqrt{N}(\beta - \beta^0) \\ cT\sqrt{N}(\beta - \beta^0) & c^2 T \end{pmatrix} \right] + o_P \left[\left(\sqrt{N} + \sqrt{NT} \|\beta - \beta^0)\| \right)^2 \right] \\
& = \frac{1}{2} \left(c^2 N + c^2 T + NT(\beta - \beta^0)^2 + \sqrt{\left[c^2 N + c^2 T + NT(\beta - \beta^0)^2 \right]^2 - 4c^4 NT} \right) \\
& \qquad + o_P \left[\left(\sqrt{N} + \sqrt{NT} \|\beta - \beta^0)\| \right)^2 \right] .
\end{aligned}
\tag{S.17}
$$

The objective function for $R = 1$ reads

$$
\begin{aligned}
\mathcal{L}^1_{NT}(\beta) &= \mathcal{L}^0_{NT}(\beta) - \mu_1 \left[(Y - \beta \cdot X)' (Y - \beta \cdot X) \right] \\
&= \mathrm{Tr} \left[(Y - \beta \cdot X)' (Y - \beta \cdot X) \right] - \mu_1 \left[(Y - \beta \cdot X)' (Y - \beta \cdot X) \right] \\
&= \mathrm{Tr}(e'e) + 2(\beta - \beta^0) \mathrm{Tr}(X'e) + (\beta - \beta^0)^2 \mathrm{Tr}(X'X) \\
& \qquad - \mu_1 \left[(e_1(\beta) + e_2(\beta))' (e_1(\beta) + e_2(\beta)) \right] \\
&= \mathrm{Tr}(e'e) + (\beta - \beta^0)^2 (NT + a^2 NT) + \mathcal{O}_P(\sqrt{NT} \|\beta - \beta^0\|) + o_P(NT \|\beta - \beta^0\|^2) \\
& \qquad - \mu_1 \left[(e_1(\beta) + e_2(\beta))' (e_1(\beta) + e_2(\beta)) \right] .
\end{aligned}
\tag{S.18}
$$

We have

$$
\begin{aligned}
& \left| \mu_1 \left[(e_1(\beta) + e_2(\beta))' (e_1(\beta) + e_2(\beta)) \right] - \mu_1 \left[e_1(\beta)' e_1(\beta) \right] \right| \\
& \leq \left\| (e_1(\beta) + e_2(\beta))' (e_1(\beta) + e_2(\beta)) - e_1(\beta)' e_1(\beta) \right\| \\
& \leq 2 \|e_1(\beta)\| \, \|e_2(\beta)\| + \|e_2(\beta)\|^2,
\end{aligned}
\tag{S.19}
$$

and therefore

$$
\begin{aligned}
& \left| \mathcal{L}^1_{NT}(\beta) - \mathrm{Tr}(e'e) - (\beta - \beta^0)^2 (NT + a^2 NT) + \|e_1(\beta)\|^2 \right| \\
& \qquad \leq 2 \|e_1(\beta)\| \, \|e_2(\beta)\| + \|e_2(\beta)\|^2 + \mathcal{O}_P(\sqrt{NT} \|\beta - \beta^0\|) + o_P(NT \|\beta - \beta^0\|^2).
\end{aligned}
\tag{S.20}
$$

Using this inequality together with the results on $\|e_1(\beta)\|$ and $\|e_2(\beta)\|$ above one can show that (for details see below)

$$
\min_{\beta \in [\beta^0 - \Delta_{NT}, \beta^0 + \Delta_{NT}]} \mathcal{L}^1_{NT}(\beta) \geq \mathrm{Tr}(e'e) - T \underbrace{\left[c \max(1, \kappa) + 1 + \kappa \right]^2}_{\equiv f_1(\kappa, a, c)} + o_P(N),
\tag{S.21}
$$

and for $\widetilde{\beta}_{NT} = \beta^0 + c(aNT)^{-1/4}$ we have (again, for details see below)

$$\begin{aligned} &\mathcal{L}^1_{NT}(\widetilde{\beta}_{NT}) \\ &\leq \mathrm{Tr}(e'e) - \underbrace{\left[c^2 g(a,\kappa) - c^2 a^{-1/2}(1+a^2)\kappa - 2c(1+\kappa)\sqrt{g(a,\kappa)} - (1+\kappa)^2\right]}_{\equiv f_2(\kappa,a,c)} T + o_P(N), \quad \text{(S.22)} \end{aligned}$$

where

$$g(a,\kappa) = \frac{1}{2}\left(1+\kappa^2+\frac{\kappa}{\sqrt{a}} + \sqrt{\left(1+\kappa^2+\frac{\kappa}{\sqrt{a}}\right)^2 - 4\kappa^2}\right). \tag{S.23}$$

For $0 < a < (1/2)^{2/3}\min(\kappa^2,\kappa^{-2})$ and $c \geq \frac{(2+\sqrt{2})(1+\kappa)(1+\sqrt{3}a^{-1/4})}{\min(1,\kappa)[1/2 - a^{3/2}\max(\kappa,\kappa^{-1})]}$ one can show that $f_1(\kappa,a,c) < f_2(\kappa,a,c)$ (for details on this below). Thus, for these values of a and c we can conclude that wpa1

$$\min_{\beta\in[\beta^0-\Delta_{NT},\beta^0+\Delta_{NT}]} \mathcal{L}^1_{NT}(\beta) > \mathcal{L}^1_{NT}(\widetilde{\beta}_{NT}) \geq \min_{\beta\in\mathbb{R}} \mathcal{L}^1_{NT}(\beta). \tag{S.24}$$

This is what we wanted to show. In the following we provide more details regarding how to obtain (S.21) and (S.22) and $f_1(\kappa,a,c) < f_2(\kappa,a,c)$.

Derivation of (S.21): Remember $\Delta_{NT} = o(N^{-1/2})$. Thus, for any $\beta \in [\beta^0 - \Delta_{NT}, \beta^0 + \Delta_{NT}]$ we find from (S.15), (S.17), and (S.20) that

$$\begin{aligned} \|e_1(\beta)\|^2 &= c^2\max(N,T) + o_P(N) = c^2\max(1,\kappa^2)T + o_P(N), \\ \|e_2(\beta)\|^2 &= (\sqrt{N}+\sqrt{T})^2 + o_P(N) = (1+\kappa)^2 T + o_P(N), \\ \mathcal{L}^1_{NT}(\beta) &\geq \mathrm{Tr}(e'e) - \|e_1(\beta)\|^2 - 2\|e_1(\beta)\|\,\|e_2(\beta)\| - \|e_2(\beta)\|^2 + o_P(N) \\ &= \mathrm{Tr}(e'e) - (\|e_1(\beta)\| + \|e_2(\beta)\|)^2 + o_P(N) \\ &= \mathrm{Tr}(e'e) - T\,\left[c\max(1,\kappa) + 1 + \kappa\right]^2 + o_P(N). \end{aligned} \tag{S.25}$$

Derivation of (S.22): We defined $\widetilde{\beta}_{NT} = \beta^0 + c(aNT)^{-1/4}$. From (S.15) we find $\|e_2(\beta)\|^2 = (1+\kappa)^2T + o_P(N)$ as before. Furthermore, we find from (S.17) that $\left\|e_1(\widetilde{\beta}_{NT})\right\|^2 = c^2\,T\,g(a,\kappa) + o_P(N)$. Equation (S.20) thus gives

$$\begin{aligned} &\mathcal{L}^1_{NT}(\widetilde{\beta}_{NT}) \\ &\leq \mathrm{Tr}(e'e) + c^2a^{-1/2}(1+a^2)\kappa T - \left\|e_1(\widetilde{\beta}_{NT})\right\|^2 + 2\left\|e_1(\widetilde{\beta}_{NT})\right\|\left\|e_2(\widetilde{\beta}_{NT})\right\| + \left\|e_2(\widetilde{\beta}_{NT})\right\|^2 + o_P(N) \\ &= \mathrm{Tr}(e'e) + \left[c^2a^{-1/2}(1+a^2)\kappa - c^2 g(a,\kappa) + 2c(1+\kappa)\sqrt{g(a,\kappa)} + (1+\kappa)^2\right] T + o_P(N). \quad \text{(S.26)} \end{aligned}$$

Show that $f_1(\kappa, a, c) < f_2(\kappa, a, c)$: Recall

$$\begin{aligned}
f_1(\kappa, a, c) &= (\max\{1,\kappa\} c + 1 + \kappa)^2 = \max\left\{1,\kappa^2\right\} c^2 + 2\max\{1,\kappa\}(1+\kappa)c + (1+\kappa)^2, \\
f_2(\kappa, a, c) &= \left(g(a,\kappa) - \frac{1+a^2}{\sqrt{a}}\kappa\right) c^2 - 2(1+\kappa)\sqrt{g(a,\kappa)}c - (1+\kappa)^2.
\end{aligned}$$

Note that $f_2(\kappa, a, c) - f_1(\kappa, a, c)$ is a quadratic polynomial in c, namely

$$f_2(\kappa, a, c) - f_1(\kappa, a, c) = h_1(a,\kappa) c^2 - 2h_2(a,\kappa) c - h_3(\kappa), \tag{S.27}$$

where

$$\begin{aligned}
h_1(a,\kappa) &= g(a,\kappa) - \frac{1+a^2}{\sqrt{a}}\kappa - \max\left\{1,\kappa^2\right\}, \\
h_2(a,\kappa) &= (1+\kappa)\sqrt{g(a,\kappa)} + \max\{1,\kappa\}(1+\kappa) > 0, \\
h_3(\kappa) &= 2(1+\kappa)^2 > 0.
\end{aligned}$$

We first want to show that $h_1(a,\kappa) > 0$. By assumption we have $a = \epsilon^2 \min\{\kappa^2, \kappa^{-2}\}$ with $0 < \epsilon \le (1/2)^{1/3}$. Suppose that $\kappa \ge 1$, i.e. $a = \frac{\epsilon^2}{\kappa^2}$. Then, we have

$$\begin{aligned}
& h_1(a,\kappa) \\
&= g(a,\kappa) - \frac{1+a^2}{\sqrt{a}}\kappa - \kappa^2 \\
&= \frac{1}{2}\left(1 + \left(1+\frac{1}{\epsilon}\right)\kappa^2 + \sqrt{\left(1 + \left(1+\frac{1}{\epsilon}\right)\kappa^2\right)^2 - 4\kappa^2}\right) - \frac{1}{\epsilon}\kappa^2 - \frac{\epsilon^3}{\kappa^2} - \kappa^2 \\
&= \frac{1}{2} - \frac{\epsilon^3}{\kappa^2} + \left\{\frac{1}{2}\left(1+\frac{1}{\epsilon}\right) + \frac{1}{2}\sqrt{\left(\frac{1}{\kappa^2} + \left(1+\frac{1}{\epsilon}\right)\right)^2 - \frac{4}{\kappa^2}} - \left(1+\frac{1}{\epsilon}\right)\right\}\kappa^2 \\
&= \frac{1}{2} - \frac{\epsilon^3}{\kappa^2} + \left\{\frac{1}{2}\left(1+\frac{1}{\epsilon}\right) + \frac{1}{2}\sqrt{\left(1+\frac{1}{\epsilon}\right)^2 + \left(\frac{2}{\epsilon} - 2\right)\frac{1}{\kappa^2} + \frac{1}{\kappa^4}} - \left(1+\frac{1}{\epsilon}\right)\right\}\kappa^2 \\
&> \frac{1}{2} - \frac{\epsilon^3}{\kappa^2} \ge \frac{1}{2} - \epsilon^3 \ge 0,
\end{aligned}$$

where the first strict inequality holds since

$$\sqrt{\left(1+\frac{1}{\epsilon}\right)^2 + \left(\frac{2}{\epsilon} - 2\right)\frac{1}{\kappa^2} + \frac{1}{\kappa^4}} > \sqrt{\left(1+\frac{1}{\epsilon}\right)^2} = 1 + \frac{1}{\epsilon}.$$

Analogously one can show that $h_1(a,\kappa) > \kappa^2(1/2 - \epsilon^3\kappa^2) > 0$ for $\kappa < 1$. Since $h_1(a,\kappa) > 0$ and $h_3(\kappa) > 0$, the quadratic equation $h_1(a,\kappa) c^2 - 2h_2(a,\kappa) c - h_3(\kappa) = 0$ has two real roots, the larger of which reads

$$c_{\text{bnd}}\left(a,\kappa\right) \;=\; \frac{h_2\left(a,\kappa\right)+\sqrt{h_2\left(a,\kappa\right)^2+h_1\left(a,\kappa\right)h_3\left(\kappa\right)}}{h_1\left(a,\kappa\right)},$$

and we have $f_2(\kappa,a,c)-f_1(\kappa,a,c)>0$ if $c>c_{\text{bnd}}\left(a,\kappa\right)$. Since $\sqrt{x+y}\leq\sqrt{x}+\sqrt{y}$ for all positive numbers x,y, and $h_1\left(a,\kappa\right)h_3\left(\kappa\right)\leq 2h_2\left(a,\kappa\right)^2$ we have

$$c_{\text{bnd}}\left(a,\kappa\right)\leq\frac{2h_2\left(a,\kappa\right)+\sqrt{h_1\left(a,\kappa\right)h_3\left(\kappa\right)}}{h_1\left(a,\kappa\right)}\leq(2+\sqrt{2})\frac{h_2\left(a,\kappa\right)}{h_1\left(a,\kappa\right)}.$$

Above we have already shown the lower bound $h_1\left(a,\kappa\right)>\min(1,\kappa^2)[1/2-\epsilon^3\min(\kappa^2,\kappa^{-2})]=\min(1,\kappa^2)[1/2-a^{3/2}\max(\kappa,\kappa^{-1})]$. In addition, we have $g(a,\kappa)<3\max(1,\kappa^2)/\sqrt{a}$ and therefore $h_2\left(a,\kappa\right)<\max(1,\kappa)\left(1+\kappa\right)\left(1+\sqrt{3}a^{-1/4}\right)$. Thus,

$$c_{\text{bnd}}\left(a,\kappa\right)<\frac{(2+\sqrt{2})\left(1+\kappa\right)\left(1+\sqrt{3}a^{-1/4}\right)}{\min(1,\kappa)[1/2-a^{3/2}\max(\kappa,\kappa^{-1})]}. \tag{S.28}$$

Our assumptions guarantee that c is larger or equal to the rhs of the last inequality, i.e. also $c>c_{\text{bnd}}\left(a,\kappa\right)$ and $f_2(\kappa,a,c)-f_1(\kappa,a,c)>0$. ■

Proof of Theorem A.2 ($N^{3/4}$ Convergence Rate of $\widehat{\beta}_{R^0}$). The result follows from Theorem S.5 and Lemma S.8 below. ■

Proof of Theorem A.3 (Asymptotic Equivalence of $\widehat{\beta}_{R^0}$ and $\widehat{\beta}_R$, $R>R^0$). The result follows from Corollary S.10 and Lemmas S.8 and S.12 below. ■

Proof of Lemma A.4 (Justification of Main Text High-Level Assumptions). See Section S.4.2 below. ■

Proof of Lemma A.1 (Spectral Norm Bound for Random Matrices). Let Σ, η, Ψ, χ be the $N\times N$ matrices with entries Σ_{ij}, η_{ij}, Ψ_{ij} and χ_{ij}, respectively. Assumption (ii) of the Lemma guarantees that

$$\mathbb{E}\|\eta\|_{HS}^2=\sum_{i,j=1}^{n}\mathbb{E}(\eta_{ij}^2)=\mathcal{O}(N^2), \tag{S.29}$$

from which we conclude that $\|\eta\|_{HS}=\mathcal{O}_P(N)$. Analogously, we find that assumption (iv) of the Lemma implies $\|\chi\|_{HS}=\mathcal{O}_P(N)$. Furthermore, assumption (iii) of the Lemma guarantees that $\|\Psi\|_{HS}=\sqrt{\sum_{i,j=1}^{n}\Psi_{ij}^2}=\mathcal{O}_P(N^{1/2})$. Since $\eta^2=N\Psi+N^{1/2}\chi$ we thus have

$$\|\eta^2\|_{HS}=\|N\Psi+N^{1/2}\chi\|_{HS}\leq N\|\Psi\|_{HS}+N^{1/2}\|\chi\|_{HS}=\mathcal{O}_P(N^{3/2}). \tag{S.30}$$

Since Σ is a symmetric positive definite matrix we have $\|\Sigma\|=\mu_1(\Sigma)$, i.e. by assumption (i) of the Lemma we have $\|\Sigma\|=\mathcal{O}(1)$.

Using the above results on $\|\eta\|_{HS}$, $\|\eta^2\|_{HS}$ and $\|\Sigma\|$, and the fact that $ee' = T\Sigma + T^{1/2}\eta$, we obtain

$$\begin{aligned}
\|e\|^4 &= \|(ee')^2\| \le \|(ee')^2\|_{HS} = \|(T\Sigma + T^{1/2}\eta)^2\|_{HS} \\
&\le T^2\|\Sigma^2\|_{HS} + 2T^{3/2}\|\Sigma\eta\|_{HS} + T\|\eta^2\|_{HS} \\
&\le T^2 N^{1/2}\|\Sigma\|^2 + 2T^{3/2}\|\Sigma\|\|\eta\|_{HS} + T\|\eta^2\|_{HS} \\
&= \mathcal{O}_P\left(T^2N^{1/2} + T^{3/2}N + TN^{3/2}\right) = \mathcal{O}_P(N^{5/2}),
\end{aligned} \tag{S.31}$$

where in the second to last line we applied the general matrix norm inequalities $\|A\|_{HS} \le \mathrm{rank}(A)\|A\|$ and $\|CD\|_{HS} \le \|C\|\|D\|_{HS}$ with $A = \Sigma^2$, $C = \Sigma$ and $D = \eta$. We thus conclude that $\|e\| = \mathcal{O}_P(N^{5/8})$. ■

S.2 Details for Quadratic Approximation of $\mathcal{L}^0_{NT}(\beta)$

The following extends the discussion in Section 4.2 and Appendix A.2 of the main paper. Using the perturbation theory of linear operators we provide an asymptotic expansion of the least squares objective function $\mathcal{L}^0_{NT}(\beta)$ when $R = R^0$. Lemma S.1 is the key result of this section, which is afterwards used to show Theorem 4.2. The proofs for the intermediate results of this section are provided in Section S.6 below.

This section is only concerned with $R = R^0$, in which case we write $\widehat{\lambda} := \widehat{\Lambda}_{R^0}$ and $\widehat{f} = \widehat{F}_{R^0}$. It is also convenient to define $\widehat{\lambda}(\beta)$ and $\widehat{f}(\beta)$ as the minimizers of the LS objective for different values of β. We have $\widehat{\lambda} = \widehat{\lambda}(\widehat{\beta}_{R^0})$ and $\widehat{f} = \widehat{f}(\widehat{\beta}_{R^0})$. Finally, we define $M_{\widehat{\lambda}}(\beta) := M_{\widehat{\lambda}(\beta)}$ and $M_{\widehat{f}}(\beta) := M_{\widehat{f}(\beta)}$, and the residuals $\widehat{e}(\beta) := Y - \beta \cdot X - \widehat{\lambda}(\beta)\widehat{f}'(\beta)$.

S.2.1 General Expansion Result and Proof of Theorem 4.2

Definition 1. *For the $N \times R^0$ matrix λ^0 and the $T \times R^0$ matrix f^0 we define*

$$\begin{aligned}
d_{\max}(\lambda^0, f^0) &= \frac{1}{\sqrt{NT}} \left\|\lambda^0 f^{0\prime}\right\| = \sqrt{\frac{1}{NT}\,\mu_1(\lambda^{0\prime} f^0 f^{0\prime} \lambda^0)}\ , \\
d_{\min}(\lambda^0, f^0) &= \sqrt{\frac{1}{NT}\,\mu_{R^0}(\lambda^{0\prime} f^0 f^{0\prime} \lambda^0)}\ ,
\end{aligned} \tag{S.32}$$

i.e. $d_{\max}(\lambda^0, f^0)$ and $d_{\min}(\lambda^0, f^0)$ are the square roots of the maximal and the minimal eigenvalue of $\lambda^{0\prime} f^0 f^{0\prime}\lambda^0/NT$. Furthermore, the convergence radius $r_0(\lambda^0, f^0)$ is defined by

$$r_0(\lambda^0, f^0) = \left(\frac{4d_{\max}(\lambda^0, f^0)}{d^2_{\min}(\lambda^0, f^0)} + \frac{1}{2d_{\max}(\lambda^0, f^0)}\right)^{-1} . \tag{S.33}$$

Lemma S.1. *If the following condition is satisfies*

$$\sum_{k=1}^{K} \left|\beta_k - \beta_k^0\right| \frac{\|X_k\|}{\sqrt{NT}} + \frac{\|e\|}{\sqrt{NT}} \; < \; r_0(\lambda^0, f^0) \; , \tag{S.34}$$

then

(i) the profile least squares objective function can be written as a power series in the $K+1$ parameters $\epsilon_0 = \|e\|/\sqrt{NT}$ and $\epsilon_k = \beta_k^0 - \beta_k$, namely

$$\mathcal{L}_{NT}^0\left(\beta\right) = \frac{1}{NT} \sum_{g=2}^{\infty} \sum_{k_1=0}^{K} \sum_{k_2=0}^{K} \ldots \sum_{k_g=0}^{K} \epsilon_{k_1}\,\epsilon_{k_2}\,\ldots\,\epsilon_{k_g}\, L^{(g)}\left(\lambda^0,\, f^0,\, X_{k_1},\, X_{k_2}, \ldots, X_{k_g}\right) ,$$

where the expansion coefficients are given by[1]

$$\begin{aligned} L^{(g)}\left(\lambda^0,\, f^0,\, X_{k_1},\, X_{k_2}, \ldots, X_{k_g}\right) &= \widetilde{L}^{(g)}\left(\lambda^0,\, f^0,\, X_{(k_1},\, X_{k_2}, \ldots, X_{k_g)}\right) \\ &= \frac{1}{g!}\left[\widetilde{L}^{(g)}\left(\lambda^0,\, f^0,\, X_{k_1},\, X_{k_2}, \ldots, X_{k_g}\right) + \textit{all permutations of } k_1, \ldots, k_g\right] , \end{aligned}$$

i.e. $L^{(g)}$ is obtained by total symmetrization of the last g arguments of[2]

$$\begin{aligned} &\widetilde{L}^{(g)}\left(\lambda^0,\, f^0,\, X_{k_1},\, X_{k_2}, \ldots, X_{k_g}\right) \\ &\quad = \sum_{p=1}^{g} (-1)^{p+1} \sum_{\substack{\nu_1+\ldots+\nu_P = g \\ m_1+\ldots+m_{p+1} = p-1 \\ 2 \geq \nu_j \geq 1 \; , \; m_j \geq 0}} \mathrm{Tr}\left(S^{(m_1)}\, \mathcal{T}_{k_1\ldots}^{(\nu_1)}\, S^{(m_2)} \ldots S^{(m_P)}\, \mathcal{T}_{\ldots k_g}^{(\nu_P)}\, S^{(m_{p+1})}\right) , \end{aligned}$$

with

$$\begin{aligned} S^{(0)} &= -M_{\lambda^0} \; , & S^{(m)} &= \left[\lambda^0(\lambda^{0\prime}\lambda^0)^{-1}(f^{0\prime}f^0)^{-1}(\lambda^{0\prime}\lambda^0)^{-1}\lambda^{0\prime}\right]^m \; , \; \textit{for } m \geq 1, \\ \mathcal{T}_k^{(1)} &= \lambda^0\, f^{0\prime}\, X_k' + X_k\, f^0\, \lambda^{0\prime} \; , & \mathcal{T}_{k_1 k_2}^{(2)} &= X_{k_1}\, X_{k_2}' \; , \qquad \textit{for } k, k_1, k_2 = 0 \ldots K \; , \\ X_0 &= \frac{\sqrt{NT}}{\|e\|}\, e \; , & X_k &= X_k \; , \qquad \textit{for } k = 1 \ldots K \; . \end{aligned}$$

[1] Here we use the round bracket notation $(k_1, k_2, \ldots, k_g)$ for total symmetrization of these indices, e.g. $\widetilde{L}^{(2)}\left(\lambda^0,\, f^0,\, X_{(k_1},\, X_{k_2)}\right) = \frac{1}{2}\left[\widetilde{L}^{(2)}\left(\lambda^0,\, f^0,\, X_{k_1},\, X_{k_2}\right) + \widetilde{L}^{(2)}\left(\lambda^0,\, f^0,\, X_{k_2},\, X_{k_1}\right)\right]$.

[2] One finds $\widetilde{L}^{(1)}\left(\lambda^0,\, f^0,\, X_{k_1},\, X_{k_2}, \ldots, X_{k_g}\right) = 0$, which is why the sum in the power series of $\mathcal{L}_{NT}^0$ starts from $g = 2$ instead of $g = 1$. For $g = 2$ and $g = 3$ we have

$$\begin{aligned} L^{(2)}\left(\lambda^0,\, f^0,\, X_{k_1},\, X_{k_2}\right) &= \mathrm{Tr}\left(M_{\lambda^0}\, X_{k_1}\, M_{f^0}\, X_{k_2}\right) \; , \\ L^{(3)}\left(\lambda^0,\, f^0,\, X_{k_1},\, X_{k_2},\, X_{k_3}\right) &= -\frac{1}{3}\Big[\mathrm{Tr}\left(M_{\lambda^0}\, X_{k_1}\, M_f\, X_{k_2}'\, \lambda^0\, (\lambda^{0\prime}\lambda^0)^{-1}\, (f^{0\prime}f^0)^{-1}\, f^{0\prime}\, X_{k_3}'\right) \\ &\qquad + 5 \text{ permutations of } k_1 \ldots k_3\Big]. \end{aligned}$$

(ii) the projector $M_{\widehat{\lambda}}(\beta)$ can be written as a power series in the same parameters ϵ_k ($k = 0, \ldots, K$), namely

$$M_{\widehat{\lambda}}(\beta) = \sum_{g=0}^{\infty} \sum_{k_1=0}^{K} \sum_{k_2=0}^{K} \ldots \sum_{k_g=0}^{K} \epsilon_{k_1}\, \epsilon_{k_2}\, \ldots\, \epsilon_{k_g}\, M^{(g)}\left(\lambda^0,\, f^0,\, X_{k_1},\, X_{k_2}, \ldots, X_{k_g}\right) \ ,$$

where the expansion coefficients are given by $M^{(0)}(\lambda^0,\, f^0) = M_{\lambda^0}$, and for $g \geq 1$

$$\begin{aligned} M^{(g)}\left(\lambda^0,\, f^0,\, X_{k_1},\, X_{k_2}, \ldots, X_{k_g}\right) &= \widetilde{M}^{(g)}\left(\lambda^0,\, f^0,\, X_{(k_1},\, X_{k_2}, \ldots, X_{k_g)}\right) \\ &= \frac{1}{g!}\left[\widetilde{M}^{(g)}\left(X_{k_1},\, X_{k_2}, \ldots, X_{k_g}\right) + \textit{all permutations of } k_1, \ldots, k_g\right] \ , \end{aligned}$$

i.e. $M^{(g)}$ is obtained by total symmetrization of the last g arguments of

$$\begin{aligned} &\widetilde{M}^{(g)}\left(\lambda^0,\, f^0,\, X_{k_1},\, X_{k_2}, \ldots, X_{k_g}\right) \\ &\quad = \sum_{p=1}^{g} (-1)^{p+1} \sum_{\substack{\nu_1 + \ldots + \nu_P = g \\ m_1 + \ldots + m_{p+1} = p \\ 2 \geq \nu_j \geq 1\ ,\ m_j \geq 0}} S^{(m_1)}\, \mathcal{T}^{(\nu_1)}_{k_1 \ldots}\, S^{(m_2)} \ldots S^{(m_P)}\, \mathcal{T}^{(\nu_P)}_{\ldots k_g}\, S^{(m_{p+1})} \ , \end{aligned}$$

where $S^{(m)}$, $\mathcal{T}^{(1)}_k$, $\mathcal{T}^{(2)}_{k_1 k_2}$, and X_k are given above.

(iii) For $g \geq 3$ the coefficients $L^{(g)}$ in the series expansion of $\mathcal{L}^0_{NT}(\beta)$ are bounded as follows

$$\begin{aligned} &\frac{1}{NT}\left|L^{(g)}\left(\lambda^0,\, f^0,\, X_{k_1},\, X_{k_2}, \ldots, X_{k_g}\right)\right| \\ &\qquad \leq \frac{R\, g\, d^2_{\min}(\lambda^0, f^0)}{2} \left(\frac{16\, d_{\max}(\lambda^0, f^0)}{d^2_{\min}(\lambda^0, f^0)}\right)^g \frac{\|X_{k_1}\|}{\sqrt{NT}}\, \frac{\|X_{k_2}\|}{\sqrt{NT}} \ldots \frac{\|X_{k_g}\|}{\sqrt{NT}} \,. \end{aligned}$$

Under the stronger condition

$$\sum_{k=1}^{K} \left|\beta^0_k - \beta_k\right| \frac{\|X_k\|}{\sqrt{NT}} + \frac{\|e\|}{\sqrt{NT}} \ < \ \frac{d^2_{\min}(\lambda^0, f^0)}{16\, d_{\max}(\lambda^0, f^0)} \ , \tag{S.35}$$

we therefore have the following bound on the remainder when the series expansion for $\mathcal{L}^0_{NT}(\beta)$ is truncated at order $G \geq 2$:

$$\begin{aligned} &\left|\mathcal{L}^0_{NT}(\beta) - \frac{1}{NT} \sum_{g=2}^{G} \sum_{k_1=0}^{K} \ldots \sum_{k_g=0}^{K} \epsilon_{k_1} \ldots \epsilon_{k_g}\, L^{(g)}\left(\lambda^0,\, f^0,\, X_{k_1},\, X_{k_2}, \ldots, X_{k_g}\right)\right| \\ &\qquad\qquad \leq \frac{R\,(G+1)\,\alpha^{G+1}\, d^2_{\min}(\lambda^0, f^0)}{2(1-\alpha)^2} \ , \end{aligned}$$

where

$$\alpha = \frac{16\, d_{\max}(\lambda^0, f^0)}{d^2_{\min}(\lambda^0, f^0)} \left(\sum_{k=1}^{K} \left|\beta^0_k - \beta_k\right| \frac{\|X_k\|}{\sqrt{NT}} + \frac{\|e\|}{\sqrt{NT}} \right) < 1 \ .$$

(iv) *The operator norm of the coefficient $M^{(g)}$ in the series expansion of $M_{\widehat{\lambda}}(\beta)$ is bounded as follows, for $g \geq 1$*

$$\left\| M^{(g)}\left(\lambda^0, f^0, X_{k_1}, X_{k_2}, \ldots, X_{k_g}\right) \right\| \leq \frac{g}{2} \left(\frac{16\, d_{\max}(\lambda^0, f^0)}{d^2_{\min}(\lambda^0, f^0)} \right)^{g} \frac{\|X_{k_1}\|}{\sqrt{NT}} \frac{\|X_{k_2}\|}{\sqrt{NT}} \ldots \frac{\|X_{k_g}\|}{\sqrt{NT}} \ .$$

Under the condition (S.35) *we therefore have the following bound on operator norm of the remainder of the series expansion of $M_{\widehat{\lambda}}(\beta)$, for $G \geq 0$*

$$\left\| M_{\widehat{\lambda}}(\beta) - \sum_{g=0}^{G} \sum_{k_1=0}^{K} \ldots \sum_{k_g=0}^{K} \epsilon_{k_1} \ldots \epsilon_{k_g} M^{(g)}\left(\lambda^0,\ f^0,\ X_{k_1},\ X_{k_2}, \ldots, X_{k_g}\right) \right\| \leq \frac{(G+1)\,\alpha^{G+1}}{2(1-\alpha)^2} \ .$$

Proof of Theorem 4.2 (Quadratic Approximation of $\mathcal{L}^0_{NT}(\beta)$). The R^0 non-zero eigenvalues of the matrix $\lambda^{0\prime} f^0 f^{0\prime} \lambda^0/NT$ are identical to the eigenvalues of the $R^0 \times R^0$ matrix $(f^0 f^{0\prime}/T)^{-1/2}(\lambda^0\lambda^{0\prime}/N)(f^0 f^{0\prime}/T)^{-1/2}$, and Assumption SF guarantees that these eigenvalues, including $d_{\max}(\lambda^0, f^0)$ and $d_{\min}(\lambda^0, f^0)$ converge to positive constants in probability. Therefore, also $r_0(\lambda^0, f^0)$ converges to a positive constant in probability.

Assumptions SF and SN furthermore imply that in the limit $N, T \to \infty$ with $N/T \to \kappa^2$, $0 < \kappa < \infty$, we have

$$\frac{\|\lambda^0\|}{\sqrt{N}} = \mathcal{O}_P(1) \ , \quad \frac{\|f^0\|}{\sqrt{T}} = \mathcal{O}_P(1) \ , \qquad \left\| \left(\frac{\lambda^{0\prime}\lambda^0}{N} \right)^{-1} \right\| = \mathcal{O}_P(1) \ , \quad \left\| \left(\frac{f^{0\prime} f^0}{T} \right)^{-1} \right\| = \mathcal{O}_P(1) \ ,$$
$$\frac{\|X_k\|}{\sqrt{NT}} = \mathcal{O}_P(1) \ , \quad \frac{\|e\|}{\sqrt{NT}} = \mathcal{O}_P\left(N^{-1/2}\right) \ . \tag{S.36}$$

Thus, for $\|\beta - \beta^0\| \leq c_{NT}$, $c_{NT} = o(1)$, we have $\alpha \to 0$ as $N, T \to \infty$, i.e. the condition (S.35) in part (iii) of Lemma S.1 is asymptotically satisfied, and by applying the Lemma we find

$$\frac{1}{NT}(\epsilon_0)^{g-r} L^{(g)}\left(\lambda^0, f^0, X_{k_1}, \ldots, X_{k_r}, X_0, \ldots, X_0\right) = \mathcal{O}_P\left(\left(\frac{\|e\|}{\sqrt{NT}} \right)^{g-r} \right) = \mathcal{O}_P\left(N^{-\frac{g-r}{2}}\right) , \tag{S.37}$$

where we used $\epsilon_0 X_0 = e$ and the linearity of $L^{(g)}\left(\lambda^0,\ f^0,\ X_{k_1},\ X_{k_2}, \ldots, X_{k_g}\right)$ in the arguments X_k. Truncating the expansion of $\mathcal{L}^0_{NT}(\beta)$ at order $G = 3$ and applying the corresponding result in Lemma S.1(iii) we obtain

$$\begin{aligned}
\mathcal{L}^0_{NT}(\beta) =& \frac{1}{NT} \sum_{k_1,k_2=0}^{K} \epsilon_{k_1}\epsilon_{k_2} L^{(2)}\left(\lambda^0, f^0, X_{k_1}, X_{k_2}\right) \\
&+ \frac{1}{NT} \sum_{k_1,k_2,k_3=0}^{K} \epsilon_{k_1}\epsilon_{k_2}\epsilon_{k_3} L^{(3)}\left(\lambda^0, f^0, X_{k_1}, X_{k_2}, X_{k_3}\right) + \mathcal{O}_P\left(\alpha^4\right) \\
=& \mathcal{L}^0_{NT}(\beta^0) - \frac{2}{\sqrt{NT}}\left(\beta-\beta^0\right)'\left(C^{(1)}+C^{(2)}\right) \\
&+ \left(\beta-\beta^0\right)' W \left(\beta-\beta^0\right) + \mathcal{L}^{0,\mathrm{rem}}_{NT}(\beta) \, ,
\end{aligned} \tag{S.38}$$

where, using (S.37) we find

$$\begin{aligned}
\mathcal{L}^{0,\mathrm{rem}}_{NT}(\beta) =& \frac{3}{NT} \sum_{k_1,k_2=1}^{K} \epsilon_{k_1}\epsilon_{k_2}\epsilon_0 L^{(3)}\left(\lambda^0, f^0, X_{k_1}, X_{k_2}, X_0\right) \\
&+ \frac{1}{NT} \sum_{k_1,k_2,k_3=1}^{K} \epsilon_{k_1}\epsilon_{k_2}\epsilon_{k_3} L^{(3)}\left(\lambda^0, f^0, X_{k_1}, X_{k_2}, X_{k_3}\right) \\
&+ \mathcal{O}_P\left[\left(\sum_{k=1}^{K} \left|\beta^0_k - \beta_k\right| \frac{\|X_k\|}{\sqrt{NT}} + \frac{\|e\|}{\sqrt{NT}}\right)^4\right] - \mathcal{O}_P\left[\left(\frac{\|e\|}{\sqrt{NT}}\right)^4\right] \\
=& \mathcal{O}_P\left(\|\beta-\beta^0\|^2 N^{-1/2}\right) + \mathcal{O}_P\left(\|\beta-\beta^0\|^3\right) + \mathcal{O}_P\left(\|\beta-\beta^0\| N^{-3/2}\right) \\
&+ \mathcal{O}_P\left(\|\beta-\beta^0\|^2 N^{-1}\right) + \mathcal{O}_P\left(\|\beta-\beta^0\|^3 N^{-1/2}\right) + \mathcal{O}_P\left(\|\beta-\beta^0\|^4\right) \, .
\end{aligned} \tag{S.39}$$

Here $\mathcal{O}_P\left[\left(\frac{\|e\|}{\sqrt{NT}}\right)^4\right]$ is not just some term of that order, but exactly the term of that order contained in $\mathcal{O}_P(\alpha^4) = \mathcal{O}_P\left[\left(\sum_{k=1}^{K} |\beta^0_k - \beta_k| \frac{\|X_k\|}{\sqrt{NT}} + \frac{\|e\|}{\sqrt{NT}}\right)^4\right]$. This term is not present in $\mathcal{L}^{0,\mathrm{rem}}_{NT}(\beta)$ since it is already contained in $\mathcal{L}^0_{NT}(\beta^0)$.[3] Equation (S.39) shows that the remainder satisfies the bound stated in the theorem, which concludes the proof. ■

[3] Alternatively, we could have truncated the expansion at order $G = 4$. Then, the term $\mathcal{O}_P\left[\left(\frac{\|e\|}{\sqrt{NT}}\right)^4\right]$ would be more explicit, namely it would equal $\frac{1}{NT}\epsilon_0^4 L^{(4)}\left(\lambda^0, f^0, X_0, X_0, X_0, X_0\right)$, which is clearly contained in $\mathcal{L}^0_{NT}(\beta^0)$.

S.2.2 Expansion of Other Quantities

Lemma S.2. *Define the pseudo-inverses* $(\lambda^0 f^{0\prime})^\dagger \equiv f^0\,(f^{0\prime} f^0)^{-1}\,(\lambda^{0\prime}\lambda^0)^{-1}\lambda^{0\prime}$ *and* $(f^0\lambda^{0\prime})^\dagger \equiv \lambda^0\,(\lambda^{0\prime}\lambda^0)^{-1}\,(f^{0\prime} f^0)^{-1}\,f^{0\prime}$*. Under the assumptions of Theorem 4.2 we have*

$$M_{\widehat{\lambda}}(\beta) = M_{\lambda^0} + M^{(1)}_{\widehat{\lambda},e} + M^{(2)}_{\widehat{\lambda},e} - \sum_{k=1}^{K} \left(\beta_k - \beta^0_k\right) M^{(1)}_{\widehat{\lambda},X,k} + M^{(\text{rem})}_{\widehat{\lambda}}(\beta)\ ,$$

$$M_{\widehat{f}}(\beta) = M_{f^0} + M^{(1)}_{\widehat{f},e} + M^{(2)}_{\widehat{f},e} - \sum_{k=1}^{K} \left(\beta_k - \beta^0_k\right) M^{(1)}_{\widehat{f},X,k} + M^{(\text{rem})}_{\widehat{f}}(\beta)\ ,$$

where the expansion coefficients in the expansion of $M_{\widehat{\lambda}}(\beta)$ *are* $N\times N$ *matrices given by*

$$\begin{aligned}
M^{(1)}_{\widehat{\lambda},e} &= -\,M_{\lambda^0}\,e\,(\lambda^0 f^{0\prime})^\dagger \,-\, (f^0\lambda^{0\prime})^\dagger\,e'\,M_{\lambda^0}\ ,\\
M^{(1)}_{\widehat{\lambda},X,k} &= -\,M_{\lambda^0}\,X_k\,(\lambda^0 f^{0\prime})^\dagger \,-\, (f^0\lambda^{0\prime})^\dagger\,X_k'\,M_{\lambda^0}\ ,\\
M^{(2)}_{\widehat{\lambda},e} &= M_{\lambda^0}\,e\,(\lambda^0 f^{0\prime})^\dagger\,e\,(\lambda^0 f^{0\prime})^\dagger + (f^0\lambda^{0\prime})^\dagger\,e'\,(f^0\lambda^{0\prime})^\dagger\,e'\,M_{\lambda^0}\\
&\quad - M_{\lambda^0}\,e\,M_{f^0}\,e'\,(f^0\lambda^{0\prime})^\dagger\,(\lambda^0 f^{0\prime})^\dagger - (f^0\lambda^{0\prime})^\dagger\,(\lambda^0 f^{0\prime})^\dagger\,e\,M_{f^0}\,e'\,M_{\lambda^0}\\
&\quad - M_{\lambda^0}\,e\,(\lambda^0 f^{0\prime})^\dagger\,(f^0\lambda^{0\prime})^\dagger\,e'\,M_{\lambda^0} + (f^0\lambda^{0\prime})^\dagger\,e'\,M_{\lambda^0}\,e\,(\lambda^0 f^{0\prime})^\dagger\ ,
\end{aligned}$$

and analogously we have $T\times T$ *matrices*

$$\begin{aligned}
M^{(1)}_{\widehat{f},e} &= -\,M_{f^0}\,e'\,(f^0\lambda^{0\prime})^\dagger \,-\, (\lambda^0 f^{0\prime})^\dagger\,e\,M_{f^0}\ ,\\
M^{(1)}_{\widehat{f},X,k} &= -\,M_{f^0}\,X_k'\,(f^0\lambda^{0\prime})^\dagger \,-\, (\lambda^0 f^{0\prime})^\dagger\,X_k\,M_{f^0}\ ,\\
M^{(2)}_{\widehat{f},e} &= M_{f^0}\,e'\,(f^0\lambda^{0\prime})^\dagger\,e'\,(f^0\lambda^{0\prime})^\dagger + (\lambda^0 f^{0\prime})^\dagger\,e\,(\lambda^0 f^{0\prime})^\dagger\,e\,M_{f^0}\\
&\quad - M_{f^0}\,e'\,M_{\lambda^0}\,e\,(\lambda^0 f^{0\prime})^\dagger\,(f^0\lambda^{0\prime})^\dagger - (\lambda^0 f^{0\prime})^\dagger\,(f^0\lambda^{0\prime})^\dagger\,e'\,M_{\lambda^0}\,e\,M_{f^0}\\
&\quad - M_{f^0}\,e'\,(f^0\lambda^{0\prime})^\dagger\,(\lambda^0 f^{0\prime})^\dagger\,e\,M_{f^0} + (\lambda^0 f^{0\prime})^\dagger\,e\,M_{f^0}\,e'\,(f^0\lambda^{0\prime})^\dagger\ .
\end{aligned}$$

Finally, the remainder terms of the expansions satisfy for any sequence $c_{NT}\to 0$

$$\sup_{\{\beta:\|\beta-\beta^0\|\le c_{NT}\}} \frac{\left\| M^{(\text{rem})}_{\widehat{\lambda}}(\beta)\right\|}{\|\beta-\beta^0\|^2 + N^{-1/2}\|\beta-\beta^0\| \ + N^{-3/2}} = \mathcal{O}_P(1)\ ,$$

$$\sup_{\{\beta:\|\beta-\beta^0\|\le c_{NT}\}} \frac{\left\| M^{(\text{rem})}_{\widehat{f}}(\beta)\right\|}{\|\beta-\beta^0\|^2 + N^{-1/2}\,\|\beta-\beta^0\| \ + N^{-3/2}} = \mathcal{O}_P(1)\ .$$

Lemma S.3. *Let* $(\lambda^0 f^{0\prime})^\dagger$ *and* $(f^0\lambda^{0\prime})^\dagger$ *as defined in Lemma S.2 above. Under the assumptions of Theorem 4.2 we have*

$$\widehat{e}(\beta) = M_{\lambda^0}\,e\,M_{f^0} + \widehat{e}^{(1)}_e + \widehat{e}^{(2)}_e - \sum_{k=1}^{K} \left(\beta_k - \beta^0_k\right)\left(\widehat{e}^{(1)}_{X,k} + \widehat{e}^{(2)}_{X,k}\right) + \widehat{e}^{(\text{rem})}(\beta)\ ,$$

where the $N \times T$ matrix valued expansion coefficients read

$$
\begin{aligned}
\widehat{e}^{(1)}_{X,k} &= M_{\lambda^0}\, X_k\, M_{f^0} \ , \\
\widehat{e}^{(2)}_{X,k} &= -M_{\lambda^0} X_k M_{f^0} e' (f^0 \lambda^{0\prime})^{\dagger} - M_{\lambda^0} e M_{f^0} X_k' (f^0 \lambda^{0\prime})^{\dagger} - (f^0 \lambda^{0\prime})^{\dagger} X_k' M_{\lambda^0} e M_{f^0} \\
&\quad - (f^0 \lambda^{0\prime})^{\dagger} e' M_{\lambda^0} X_k M_{f^0} - M_{\lambda^0} X_k (\lambda^0 f^{0\prime})^{\dagger} e M_{f^0} - M_{\lambda^0} e (\lambda^0 f^{0\prime})^{\dagger}\, X_k\, M_{f^0} \ , \\
\widehat{e}^{(1)}_{e} &= -M_{\lambda^0}\, e\, M_{f^0}\, e' \, (f^0 \lambda^{0\prime})^{\dagger} - (f^0 \lambda^{0\prime})^{\dagger}\, e'\, M_{\lambda^0}\, e\, M_{f^0} - M_{\lambda^0}\, e\, (\lambda^0 f^{0\prime})^{\dagger}\, e\, M_{f^0} \ , \\
\widehat{e}^{(2)}_{e} &= M_{\lambda^0} e M_{f^0}\, e' \, (f^0 \lambda^{0\prime})^{\dagger}\, e' \, (f^0 \lambda^{0\prime})^{\dagger} - M_{\lambda^0} e M_{f^0}\, e'\, M_{\lambda^0}\, e \, (\lambda^0 f^{0\prime})^{\dagger}\, (f^0 \lambda^{0\prime})^{\dagger} \\
&\quad - M_{\lambda^0} e M_{f^0}\, e' \, (f^0 \lambda^{0\prime})^{\dagger}\, (\lambda^0 f^{0\prime})^{\dagger}\, e\, M_{f^0} + \, M_{\lambda^0}\, e\, (\lambda^0 f^{0\prime})^{\dagger} e M_{f^0}\, e' \, (f^0 \lambda^{0\prime})^{\dagger} \\
&\quad + \, (f^0 \lambda^{0\prime})^{\dagger}\, e'\, M_{\lambda^0} e M_{f^0}\, e' \, (f^0 \lambda^{0\prime})^{\dagger} + \, M_{\lambda^0}\, e\, (\lambda^0 f^{0\prime})^{\dagger} e (\lambda^0 f^{0\prime})^{\dagger}\, e\, M_{f^0} \\
&\quad + \, (f^0 \lambda^{0\prime})^{\dagger}\, e'\, M_{\lambda^0} e (\lambda^0 f^{0\prime})^{\dagger}\, e\, M_{f^0} + (f^0 \lambda^{0\prime})^{\dagger}\, e' \, (f^0 \lambda^{0\prime})^{\dagger}\, e'\, M_{\lambda^0} e M_{f^0} \\
&\quad - (f^0 \lambda^{0\prime})^{\dagger}\, (\lambda^0 f^{0\prime})^{\dagger}\, e\, M_{f^0}\, e'\, M_{\lambda^0} e M_{f^0} - M_{\lambda^0}\, e\, (\lambda^0 f^{0\prime})^{\dagger}\, (f^0 \lambda^{0\prime})^{\dagger}\, e'\, M_{\lambda^0} e M_{f^0} \ ,
\end{aligned}
$$

and the remainder term satisfies for any sequence $c_{NT} \to 0$

$$
\sup_{\{\beta : \|\beta - \beta^0\| \le c_{NT}\}} \frac{\left\| \widehat{e}^{(\mathrm{rem})}(\beta) \right\|}{N \|\beta - \beta^0\|^2 + \|\beta - \beta^0\| \ + N^{-1}} = \mathcal{O}_P\left(1\right) \ .
$$

S.3 Details for $N^{3/4}$-Convergence Rate of $\widehat{\beta}_R$

This section extends the discussion in Section A.3 of the main paper. We provide the high-level Assumption HL1 under which $N^{3/4}\left(\widehat{\beta}_R - \beta^0\right) = \mathcal{O}_P(1)$ can be shown, see Theorem S.5 below. Lemma S.8 then provides the connection between our main text assumptions and Assumption HL1. The proofs are provided in Section S.6 below. Combining Theorem S.5 and Lemma S.8 yields Theorem A.2 in the main text.

We first note that equation (3.3) implies that

$$
\begin{aligned}
\mathcal{L}^R_{NT}(\beta) &= \mathcal{L}^0_{NT}(\beta) - \frac{1}{NT} \sum_{r=R^0+1}^{R} \mu_r \left[(Y - \beta \cdot X)' \, (Y - \beta \cdot X) \right] \\
&= \mathcal{L}^0_{NT}(\beta) - \frac{1}{NT} \sum_{r=1}^{R-R^0} \mu_r \left[\widehat{e}'(\beta) \widehat{e}(\beta) \right] . \qquad \text{(S.40)}
\end{aligned}
$$

The extra term $\frac{1}{NT} \sum_{r=R^0+1}^{R} \mu_r \left[(Y - \beta \cdot X)' \, (Y - \beta \cdot X) \right]$ is due to overfitting on the extra factors. In the second line of (S.40) we used that $\widehat{e}'(\beta)\widehat{e}(\beta)$ is the residual of $(Y - \beta \cdot X)' \, (Y - \beta \cdot X)$ after subtracting the first R^0 principal components, which implies that the eigenvalues of these two matrices are the same, except from the R^0 largest ones which are missing in $\widehat{e}'(\beta)\widehat{e}(\beta)$. The decomposition in equation (S.40) together with the expansion result for $\widehat{e}(\beta)$ in Lemma S.3 give rise to the following Lemma.

Lemma S.4. *Under Assumption SF and SN and for $R > R^0$ we have*

$$\mathcal{L}_{NT}^{R}(\beta) = \mathcal{L}_{NT}^{0}(\beta) - \frac{1}{NT} \sum_{r=1}^{R-R^0} \mu_r \left[A(\beta)\right] + \mathcal{L}_{NT}^{R,\mathrm{rem},1}(\beta),$$

where $A(\beta) = M_{f^0} \left[e - \Delta\beta \cdot X\right]' M_{\lambda^0} \left[e - \Delta\beta \cdot X\right] M_{f^0}$, with $\Delta\beta = \beta - \beta^0$, and for any constant $c > 0$ we have

$$\sup_{\{\beta: \sqrt{N}\|\beta-\beta^0\| \leq c\}} \frac{\left|\mathcal{L}_{NT}^{R,\mathrm{rem},1}(\beta)\right|}{\sqrt{N} + \sqrt{NT}\ \|\beta - \beta^0\|} = \mathcal{O}_p\left(\frac{1}{NT}\right).$$

The following high-level assumption guarantees that the β-dependence of $\frac{1}{NT}\sum_{r=1}^{R-R^0} \mu_r \left[A(\beta)\right]$ is small, so that apart from a constant the approximate quadratic expansions of $\mathcal{L}_{NT}^{R}(\beta)$ and $\mathcal{L}_{NT}^{0}(\beta)$ around β^0 are identical.

Assumption HL1 (First High-Level Assumption on Matrix Spectra). *Let $\Delta\beta = \beta - \beta^0$ and*

$$\begin{aligned} d(\beta) = \sum_{r=1}^{R-R^0} \Big\{ & \mu_r \left[M_{f^0} \left(e - \Delta\beta \cdot X\right)' M_{\lambda^0} \left(e - \Delta\beta \cdot X\right) M_{f^0}\right] \\ & - \mu_r \left[M_{f^0} e' M_{\lambda^0} e M_{f^0}\right] - \mu_r \left[M_{f^0} \left(\Delta\beta \cdot X\right)' M_{\lambda^0} \left(\Delta\beta \cdot X\right) M_{f^0}\right] \Big\}. \end{aligned}$$

For all constants $c > 0$ we assume that

$$\sup_{\{\beta: \sqrt{N}\|\beta-\beta^0\| \leq c\}} \frac{\max[d(\beta), 0]}{\sqrt{N} + N^{5/4}\|\beta - \beta^0\| + N^2\|\beta - \beta^0\|^2/\log(N)} = \mathcal{O}_P\left(1\right).$$

Combining Lemma S.4 with this high-level assumption yields the following theorem.

Theorem S.5. *Let $R > R^0$, let Assumptions SF, SN, NC, EX and HL1 be satisfied and furthermore assume that $C^{(1)} = \mathcal{O}_P(N^{1/4})$. In the limit $N, T \to \infty$ with $N/T \to \kappa^2$, $0 < \kappa < \infty$, we then have $N^{3/4}\left(\widehat{\beta}_R - \beta^0\right) = \mathcal{O}_P(1)$.*

The theorem follows from the inequality $\mathcal{L}_{NT}^{R}(\widehat{\beta}_R) \leq \mathcal{L}_{NT}^{R}(\beta^0)$ by applying Lemma S.4, Assumption HL1, and our expansion of $\mathcal{L}_{NT}^{0}(\beta)$. The detailed proof is given below.

S.3.1 Justification of Assumption HL1

We first present two technical Lemmas, which are used to show Lemma S.8 below.

Lemma S.6. *Let g be an $N \times Q$ matrix and h be a $T \times Q$ matrix such that $g'g = h'h = \mathbb{1}_Q$.*

Let U be an $N \times T$ matrix and C a $Q \times Q$ matrix. Assume that $\operatorname{rank}[(U'g, h)] = 2Q$. *Let*[4]

$$\Delta_{\max} = \max_{r \in \{1,2,\ldots,\min(R,Q)\}} \left[\mu_r(g'UU'g) - \mu_{r+Q-\min(Q,R)}(g'UU'g) \right].$$

We then have

$$\begin{aligned} &\sum_{r=1}^{R} \mu_r \left[(U + gCh')' (U + gCh') \right] \\ &\qquad \leq \sum_{r=1}^{R} \mu_r \left(U'U + \|g'UU'g\| \, P_{\left(M_{U'g} h\right)} + \Delta_{\max} P_{(U'g)} \right) \\ &\qquad\qquad + \sum_{r=1}^{\min(Q,R)} \mu_r \left(CC' + g'UhC' + Ch'U'g \right). \end{aligned}$$

Lemma S.7. *Let e be an $N \times T$ matrix, whose columns e_t, $t = 1, \ldots T$, are distributed as $e_t \sim iid\mathcal{N}(0, \Sigma)$, where Σ is a symmetric positive definite non-random $N \times N$ matrix with eigenvalues $\mu_1(\Sigma)$, $\ldots$, $\mu_N(\Sigma)$. Let A be a symmetric positive definite non-random $T \times T$ matrix with* $\operatorname{rank}(A) = Q$. *Let n be the number of eigenvalues of Σ that is larger or equal than $\|A\|/T$, i.e. $n \leq N$ is the largest integer such that $\mu_n(\Sigma) \geq \|A\|/T$. Consider an asymptotic where $N, T, n \to \infty$ jointly, while Q and R are constant positive integers. We then have*

$$\sum_{r=1}^{R} \mu_r \left(e'e + A \right) - \sum_{r=1}^{R} \mu_r \left(e'e \right) = \mathcal{O}_P \left(\sqrt{(N+T)T/n} \right).$$

The following Lemma connects Lemma S.5 to the main text.

Lemma S.8. *Let $R > R^0$ and let Assumptions SF hold. Let either Assumption DX-1 or DX-2 be satisfied. Consider $N, T \to \infty$ with $N/T \to \kappa^2$, $0 < \kappa < \infty$. Then Assumptions SN and HL1 are satisfied. If, in addition, Assumption EX holds, then we have $C^{(1)} = \mathcal{O}_P(N^{1/4})$.*

Combining Theorem S.5 and Lemma S.8 we obtain Theorem A.2 in the main text.

S.4 Details for Asymptotic Equivalence of $\widehat{\beta}_{R^0}$ and $\widehat{\beta}_R$

This section extends the discussion of Section A.4 in the main paper. By applying the expansion of $\widehat{e}(\beta)$ in equation (S.40) to the expression for $\mathcal{L}^R_{NT}(\beta)$ one obtains the following.

Lemma S.9. *Under Assumption SF and SN and for $R > R^0$ we have*

$$\mathcal{L}^R_{NT}(\beta) = \mathcal{L}^0_{NT}(\beta) - \frac{1}{NT} \sum_{r=1}^{R-R^0} \mu_r \left[B(\beta) + B'(\beta) \right] + \mathcal{L}^{R,\text{rem}}_{NT}(\beta),$$

[4] Note that $\Delta_{\max} = 0$ if $R \geq Q$, and that $\Delta_{\max} \leq \mu_1(g'UU'g) - \mu_Q(g'UU'g)$ for $R < Q$.

where

$$
\begin{aligned}
B(\beta) = {} & \tfrac{1}{2} M_{f^0} \left[e - (\beta - \beta^0) \cdot X\right]' M_{\lambda^0} \left[e - (\beta - \beta^0) \cdot X\right] M_{f^0} \\
& - M_{f^0} e' M_{\lambda^0} e M_{f^0} e' \lambda^0 (\lambda^{0\prime} \lambda^0)^{-1} (f^{0\prime} f^0)^{-1} f^{0\prime} \\
& + M_{f^0} \left[(\beta - \beta^0) \cdot X - e\right]' M_{\lambda^0} e f^0 (f^{0\prime} f^0)^{-1} (\lambda^{0\prime} \lambda^0)^{-1} \lambda^{0\prime} e M_{f^0} \\
& + M_{f^0} e' M_{\lambda^0} \left[(\beta - \beta^0) \cdot X\right] f^0 (f^{0\prime} f^0)^{-1} (\lambda^{0\prime} \lambda^0)^{-1} \lambda^{0\prime} e M_{f^0} \\
& + M_{f^0} e' M_{\lambda^0} e f^0 (f^{0\prime} f^0)^{-1} (\lambda^{0\prime} \lambda^0)^{-1} \lambda^{0\prime} \left[(\beta - \beta^0) \cdot X\right] M_{f^0} \\
& + B^{(eeee)} + M_{f^0} B^{(\mathrm{rem},1)}(\beta) P_{f^0} + P_{f^0} B^{(\mathrm{rem},2)} P_{f^0},
\end{aligned}
$$

and

$$
\begin{aligned}
B^{(eeee)} = {} & - M_{f^0} e' M_{\lambda^0} e M_{f^0} e' \lambda^0 (\lambda^{0\prime} \lambda^0)^{-1} (f^{0\prime} f^0)^{-1} (\lambda^{0\prime} \lambda^0)^{-1} \lambda^{0\prime} e M_{f^0} \\
& + M_{f^0} e' M_{\lambda^0} e f^0 (f^{0\prime} f^0)^{-1} (\lambda^{0\prime} \lambda^0)^{-1} \lambda^{0\prime} e f^0 (f^{0\prime} f^0)^{-1} (\lambda^{0\prime} \lambda^0)^{-1} \lambda^{0\prime} e M_{f^0} \\
& - \tfrac{1}{2} M_{f^0} e' M_{\lambda^0} e f^0 (f^{0\prime} f^0)^{-1} (\lambda^{0\prime} \lambda^0)^{-1} (f^{0\prime} f^0)^{-1} f^{0\prime} e' M_{\lambda^0} e M_{f^0} \\
& + \tfrac{1}{2} M_{f^0} e' \lambda^0 (\lambda^{0\prime} \lambda^0)^{-1} (f^{0\prime} f^0)^{-1} f^{0\prime} e' M_{\lambda^0} e f^0 (f^{0\prime} f^0)^{-1} (\lambda^{0\prime} \lambda^0)^{-1} \lambda^{0\prime} e M_{f^0} \; .
\end{aligned}
$$

Here, $B^{(\mathrm{rem},1)}(\beta)$ and $B^{(\mathrm{rem},2)}$ are $T \times T$ matrices, $B^{(\mathrm{rem},2)}$ is independent of β and satisfies $\|B^{(\mathrm{rem},2)}\| = \mathcal{O}_P(1)$, and for any constant $c > 0$

$$
\sup_{\{\beta : \sqrt{N} \|\beta - \beta^0\| \le c\}} \frac{\|B^{(\mathrm{rem},1)}(\beta)\|}{1 + \sqrt{NT}\, \|\beta - \beta^0\|} = \mathcal{O}_p\left(1\right),
$$

$$
\sup_{\{\beta : \sqrt{N} \|\beta - \beta^0\| \le c\}} \frac{\left|\mathcal{L}^{R,\mathrm{rem}}_{NT}(\beta)\right|}{(1 + \sqrt{NT}\, \|\beta - \beta^0\|)^2} = o_p\left(\frac{1}{NT}\right).
$$

Here, the remainder term $\mathcal{L}^{R,\mathrm{rem}}_{NT}(\beta)$ stems from terms in $\widehat{e}'(\beta)\widehat{e}(\beta)$ whose spectral norm is smaller than $o_P(1)$ within a $\sqrt{N}$ shrinking neighborhood of β after dividing by $\left(1 + \sqrt{NT}\, \|\beta - \beta^0\|\right)^2$. Using Weyl's inequality those terms can be separated from the eigenvalues $\mu_r\left[\widehat{e}'(\beta)\widehat{e}(\beta)\right]$. The expression for $B(\beta)$ looks complicated, in particular the terms in $B^{(eeee)}$. Note however, that $B^{(eeee)}$ is β-independent and satisfies $\|B^{(eeee)}\| = \mathcal{O}_P(1)$ under our assumptions, so that it is relatively easy to deal with these terms. Note furthermore that the structure of $B(\beta)$ is closely related to the expansion of $\mathcal{L}^0_{NT}(\beta)$, since by definition we have $\mathcal{L}^0_{NT}(\beta) = (NT)^{-1}\mathrm{Tr}(\widehat{e}'(\beta)\widehat{e}(\beta))$, which can be approximated by $(NT)^{-1}\mathrm{Tr}(B(\beta) + B'(\beta))$. Plugging the definition of $B(\beta)$ into $(NT)^{-1}\mathrm{Tr}(B(\beta) + B'(\beta))$ one indeed recovers the terms of the approximated Hessian and score provided by Theorem 4.2, which is a convenient consistency check. We do not give explicit formulas for $B^{(\mathrm{rem},1)}(\beta)$ and $B^{(\mathrm{rem},2)}$, because those terms enter $B(\beta)$ projected by P_{f^0}, which makes them orthogonal to the leading term in $B(\beta) + B'(\beta)$, so that they can only have limited influence on the eigenvalues of $B(\beta) + B'(\beta)$. The bounds on the norms of $B^{(\mathrm{rem},1)}(\beta)$ and $B^{(\mathrm{rem},2)}$ provided in the lemma are sufficient for all conclusions on the properties of $\mu_r\left[B(\beta) + B'(\beta)\right]$ below. The proof of the lemma can be found in the section S.6 below. The lemma motivates the following high-level assumption.

Assumption HL2 (Second High-Level Assumption on Matrix Spectra). *For all con-*

stants $c > 0$

$$\sup_{\{\beta: N^{3/4}\|\beta-\beta^0\|\leq c\}} \frac{\left|\sum_{r=1}^{R-R^0} \left\{\mu_r\left[B(\beta)+B'(\beta)\right] - \mu_r\left[B(\beta^0)+B'(\beta^0)\right]\right\}\right|}{(1+\sqrt{NT}\|\beta-\beta^0\|)^2} = o_P(1),$$

where $B(\beta)$ was defined in Lemma S.9.

Combining Lemma S.4, Assumption HL2, and Theorem 4.2, we find that the profile objective function for $R > R^0$ can be written as

$$\mathcal{L}_{NT}^R(\beta) = \mathcal{L}_{NT}^R(\beta^0) - \frac{2}{\sqrt{NT}}\left(\beta-\beta^0\right)'\left(C^{(1)}+C^{(2)}\right) + \left(\beta-\beta^0\right)' W \left(\beta-\beta^0\right) + \mathcal{L}_{NT}^{R,\mathrm{rem},2}(\beta),$$

with a remainder term that satisfies for all constants $c > 0$

$$\sup_{\{\beta: N^{3/4}\|\beta-\beta^0\|\leq c\}} \frac{\left|\mathcal{L}_{NT}^{R,\mathrm{rem},2}(\beta)\right|}{\left(1+\sqrt{NT}\;\|\beta-\beta^0\|\right)^2} = o_p\left(\frac{1}{NT}\right).$$

This result, together with our $N^{3/4}$-consistency result for $\widehat{\beta}_R$, gives rise to the following corollary.

Corollary S.10. *Let $R > R^0$, let Assumptions SF, SN, NC, EX, HL1 and HL2 be satisfied and furthermore assume that $C^{(1)} = \mathcal{O}_P(1)$. In the limit $N, T \to \infty$ with $N/T \to \kappa^2$, $0 < \kappa < \infty$, we then have*

$$\sqrt{NT}\left(\widehat{\beta}_R - \beta^0\right) = W^{-1}\left(C^{(1)}+C^{(2)}\right) + o_P(1) = \mathcal{O}_P(1).$$

The proof of Corollary S.10 is analogous to that of Corollary 4.3. The combination of both corollaries shows that our main result holds under high-level assumptions, i.e. the limiting distributions of $\widehat{\beta}_R$ and $\widehat{\beta}_{R^0}$ are indeed identical.

S.4.1 Justification of Assumption HL2

The following is a technical lemma, which is crucially used in the proof of Lemma S.12 below.

Lemma S.11. *Let A and B be symmetric $n \times n$ matrices, and let A be positive semi-definite. Let $\mu_1(A) \geq \mu_2(A) \geq \ldots \geq \mu_n(A) \geq 0$ be the sorted eigenvalues of A, and let $\nu_1, \nu_2, \ldots, \nu_n$ be the corresponding eigenvectors that are orthogonal and normalized such that $\|\nu_i\| = 1$ for $i = 1, \ldots, n$. Let $b = \max_{i,j=1,\ldots,n} |\nu_i' B \nu_j|$. Let r and q be positive integers with $r < q \leq n$, and let $\sum_{i=q}^n b\left(\mu_r(A) - \mu_i(A)\right)^{-1} \leq 1$ be satisfied. Then we have*

$$|\mu_r(A+B) - \mu_r(A)| \leq \frac{(q-1)\,b}{1 - \sum_{i=q}^n \frac{b}{\mu_r(A)-\mu_i(A)}}$$

The following Lemma provides conditions under which Assumption HL2 is satisfied. It crucially connects the current section with the main text.

Lemma S.12. *Let Assumptions SF, SN and EV hold, let $R > R^0$ and consider a limit $N, T \to \infty$ with $N/T \to \kappa^2$, $0 < \kappa < \infty$. Then, for all constants $c > 0$ and $r = 1, \ldots, R - R^0$ we have*

$$\sup_{\{\beta: N^{3/4}\|\beta-\beta^0\| \leq c\}} |\mu_r\left(B(\beta) + B'(\beta)\right) - \rho_r| = o_P(1),$$

which implies that Assumption HL2 is satisified.

S.4.2 Sufficiency of Low-Level Assumptions in Main Text

The following Lemma summarizes some properties of the singular value vectors v_r and w_r of $M_{f^0} e M_{\lambda^0}$ for the case where e_{it} is *iid* normally distributed. Those properties are used in the proof of the main text Lemma A.4 below.

Lemma S.13. *Let Assumption LL hold and let v_r and w_r be defined as in Assumption EV. Then the following holds.*

(i) Let $\widetilde{v}$ be an N-vector with iid$\mathcal{N}(0,1)$ entries; let $\widetilde{w}$ be a T-vector, independent of $\widetilde{v}$, also with iid$\mathcal{N}(0,1)$ entries; and let $\widetilde{v}$ and $\widetilde{w}$ be independent of λ^0, f^0, $\overline{X}_k$, and $\widetilde{X}_k^{\text{str}}$ and eP_{f^0}. Then, for all $r, s = 1, \ldots, \min(N,T) - R^0$ we have

$$\begin{pmatrix} v_r \\ w_s \end{pmatrix} \underset{d}{=} \begin{pmatrix} \|M_{\lambda^0}\widetilde{v}\|^{-1} M_{\lambda^0}\widetilde{v} \\ \|M_{f^0}\widetilde{w}\|^{-1} M_{f^0}\widetilde{w} \end{pmatrix},$$

where $=_d$ refers to equally distributed. Furthermore, the squares of $\|M_{\lambda^0}\widetilde{v}\|^{-1}$ and $\|M_{f^0}\widetilde{w}\|^{-1}$ have inverse chi-square distributions with $N - R^0$ and $T - R^0$ degrees of freedom, respectively, which implies that for every $\xi > 0$ there exists a constant $c > 0$ such that we have

$$\mathbb{E}\left(\sqrt{N}\|M_{\lambda^0}\widetilde{v}\|^{-1}\right)^{\xi} < c, \qquad \mathbb{E}\left(\sqrt{T}\|M_{f^0}\widetilde{w}\|^{-1}\right)^{\xi} < c,$$

for all $N > 4\xi + R^0$ and $T > 4\xi + R^0$.

(ii) There exists $\varepsilon \in [0, 1/12)$ such that as N, T become large we have

$$\max_{r,s,\tau}\left|\sum_{t=\tau+1}^{T} w_{r,t} w_{s,t-\tau}\right| = \mathcal{O}_P(T^{-1/2+\varepsilon}),$$

where $r, s = 1, \ldots, \min(N,T) - R^0$ and $\tau = 1, 2, \ldots, T-1$.

(iii) The matrices $P_{\lambda^0} e P_{f^0}$, $P_{\lambda^0} e M_{f^0}$, $M_{\lambda^0} e P_{f^0}$ and $M_{\lambda^0} e M_{f^0}$ are all mutually independent, and their entries have uniformly bounded moments of arbitrary order.

The proof of Lemma S.13 is given in section S.6.3 below.

Proof of Lemma A.4 (Justification of Main Text High-Level Assumptions). # First, we show that Assumptions SN, EX and DX-1 are satisfied, and that $C^{(1)} = \mathcal{O}_P(1)$:

Since e_{it} is iid $\mathcal{N}(0,\sigma^2)$ we have $\|e\| = \mathcal{O}_P(\sqrt{N})$ as N, T grow at the same rate, see e.g. Geman (1980). This also implies that $\|\widetilde{X}_k^{\text{weak}}\| \leq \sum_{\tau=1}^{\infty} |\gamma_\tau| \|e\| = \mathcal{O}_P(\sqrt{N})$. Assumption LL therefore implies that Assumption DX-1 holds with $\widetilde{X}_k = \widetilde{X}_k^{\text{str}} + \widetilde{X}_k^{\text{weak}}$ and $\Sigma = \sigma^2 \mathbb{1}$. Note that for this Σ we have $g'\Sigma g = \sigma^2 \mathbb{1}_Q = \|g'\Sigma g\| \mathbb{1}_Q$ and $\mu_n(\Sigma) = \sigma^2 = \|g'\Sigma g\|$ for all n. Assumption DX-1 also implies that Assumption SN holds, as also noted in Lemma S.8.

Since we assume that $\mathbb{E}\,|X_{k,it}|^2$ is uniformly bounded we have $\mathbb{E}\frac{1}{NT}\text{Tr}(X_k'X_k) = \frac{1}{NT}\sum_{i,t} \mathbb{E}X_{k,it}^2 = \mathcal{O}(1)$ and therefore $\text{Tr}(X_k'X_k) = \mathcal{O}_P(NT)$. We also have $\mathbb{E}\left[\text{Tr}(X_k e')^2 | X_k\right] = \sigma^2 \text{Tr}(X_k'X_k) = \mathcal{O}_P(NT)$, and therefore $\frac{1}{\sqrt{NT}}\text{Tr}(X_k\, e') = \mathcal{O}_P(1)$, i.e. Assumption EX holds. By replacing X_k with $M_{\lambda^0} X_k M_{f^0}$ in the previous argument we also find that $C^{(1)} = \mathcal{O}_P(1)$.

Assumption EV(i) holds for any $c < c_{\max} = \lim_{N,T\to\infty} \left(\sqrt{N} + \sqrt{T}\right)^2 / N$, because from Theorem 1 in Soshnikov (2002) we know that $\rho_{R-R^0}/N - c_{\max} = \mathcal{O}_P(N^{-2/3})$. Some more details are given below.

We now show that Assumption EV(ii) holds with $q_{NT} = \log(N) N^{1/6}$. Without loss of generality, we set $\sigma = 1$ in this part of the proof. We want to show that $q_{NT} = \log(N) N^{1/6}$ also satisfies

$$\frac{1}{q_{NT}(T-R^0)} \sum_{r=q_{NT}}^{Q} \frac{1}{\mu_{R-R^0} - \mu_r} = \mathcal{O}_P(1),$$

where $\mu_r \equiv \rho_r/(T-R^0)$. Note that it is not important whether the sum runs to $Q = N - R^0$ or $Q = T - R^0$, since the contributions of small eigenvalues between $r = N - R^0$ and $r = T - R^0$ are of order $\mathcal{O}_P(1)$ anyways. Without loss of generality let $\lim_{N,T\to\infty} N/T = \kappa^2 \leq 1$ in the rest of this proof (the proof for $\kappa \geq 1$ is analogous, since all arguments are symmetric under interchange of N and T). Let $\mu_{NT} = \left[(N-R^0)^{1/2} + (T-R^0)^{1/2}\right]^2$, $\sigma_{NT} = \left[(N-R^0)^{1/2} + (T-R^0)^{1/2}\right]\left[(N-R^0)^{-1/2} + (T-R^0)^{-1/2}\right]^{1/3}$, $\overline{x} = \lim_{N,T\to\infty} \mu_{NT}/(T-R^0) = (1+\kappa)^2$, and $\underline{x} = (1-\kappa)^2$. From Theorem 1 in Soshnikov (2002) we know that the joint distribution of $\sigma_{NT}^{-1}(\rho_1 - \mu_{NT}, \rho_2 - \mu_{NT}, \ldots, \rho_{R^0+1} - \mu_{NT})$ converges to the Tracy-Widom law, i.e. to the limiting distribution of the first $R^0 + 1$ eigenvalues of the Gaussian Orthogonal Ensemble. Note that σ_{NT} is of order $N^{1/3}$, and that the Tracy-Widom law is a continuous distribution, so that the result of Soshnikov implies that

$$\overline{x} - \mu_{R-R^0} = \mathcal{O}_P(N^{-2/3}) \, , \qquad (\mu_{R-R^0} - \mu_{R-R^0+1})^{-1} = \mathcal{O}_P\left(N^{2/3}\right) \, . \tag{S.41}$$

The empirical distribution of the μ_r is defined as $F_{NT}(x) = Q^{-1}\sum_{r=1}^{Q} 1(\mu_r \leq x)$, where $1(.)$ is the indicator function. This empirical distribution converges to the Marchenko-Pastur limiting spectral distribution $F_{\text{LSD}}(x)$, which has domain $[\underline{x}, \overline{x}]$, and whose density $f_{\text{LSD}}(x) = dF_{\text{LSD}}(x)/dx$ is given by

$$f_{\text{LSD}}(x) = \frac{1}{2\pi\kappa^2 x}\sqrt{(\overline{x}-x)(x-\underline{x})} \, . \tag{S.42}$$

An upper bound for $f_{\text{LSD}}(x)$ is given by $\frac{1}{2\pi\kappa^2 \underline{x}}\sqrt{(\overline{x}-x)(\overline{x}-\underline{x})}$, and by integrating that upper bound we obtain

$$1 - F_{\mathrm{LSD}}(x) \leq a\,(\overline{x} - x)^{3/2} , \qquad\qquad a = \frac{2}{3\pi\kappa^{3/2}\underline{x}} . \tag{S.43}$$

From Theorem 1.2 in Götze and Tikhomirov (2007) we know that

$$\sup_x |F_{NT}(x) - F_{LSD}(x)| = \mathcal{O}_P(N^{-1/2}) . \tag{S.44}$$

Let $c_{1,NT} = \lceil 2N^{1/2+\epsilon} \rceil$ and $c_{2,NT} = \lceil 2N^{3/4} \rceil$, where $\lceil a \rceil$ is the smallest integer larger or equal to a. Plugging in $x = \mu_{c_{1,NT}}$ into the result of Götze and Tikhomirov, and using $F_{NT}(\mu_r) = 1 - (r-1)/N$, we find

$$\begin{aligned} a\left(\overline{x} - \mu_{c_{1,NT}}\right)^{3/2} &\geq 1 - F_{\mathrm{LSD}}(\mu_{c_{1,NT}}) = \frac{c_{1,NT} - 1}{N} + \mathcal{O}_P(N^{-1/2}) \\ &\geq N^{-1/2+\epsilon}, \qquad \text{wpa1.} \end{aligned} \tag{S.45}$$

Using this and (S.41) we obtain $(\mu_{R-R^0} - \mu_{c_1})^{-1} = \mathcal{O}_P\left(N^{1/3-2/3\epsilon}\right)$. Analogously one can show that $(\mu_{R-R^0} - \mu_{c_2})^{-1} = \mathcal{O}_P\left(N^{1/6}\right)$. In the following we just write q, c_1 and c_2 for q_{NT}, $c_{1,NT}$ and $c_{2,NT}$. Combining the above results we find

$$\begin{aligned} \frac{1}{q\,n}\sum_{r=q}^{Q}\frac{1}{\mu_{R-R^0} - \mu_r} &= \frac{1}{q\,n}\sum_{r=q}^{c_1-1}\frac{1}{\mu_{R-R^0} - \mu_r} + \frac{1}{q\,n}\sum_{r=c_1}^{c_2-1}\frac{1}{\mu_{R-R^0} - \mu_r} + \frac{1}{q\,n}\sum_{r=c_2}^{Q}\frac{1}{\mu_{R-R^0} - \mu_r} \\ &\leq \frac{c_1}{qn(\mu_{R-R^0} - \mu_{R-R^0+1})} + \frac{c_2}{qn(\mu_{R-R^0} - \mu_{c_1})} + \frac{Q}{qn(\mu_{R-R^0} - \mu_{c_2})} \\ &= \mathcal{O}_P(1) + \mathcal{O}_P(N^{-1/12-5/3\epsilon}) + \mathcal{O}_P(N^{-\epsilon}) = \mathcal{O}_P(1) . \end{aligned}$$

This is what we wanted to show.

We now show that Assumption EV(iii) holds with $q_{NT} = \log(N)N^{1/6}$. Define

$$\begin{aligned} d_{NT}^{(1)} &= \max_{r,s,k} |v_r' X_k w_s|, & d_{NT}^{(2)} &= \max_r \|v_r' e P_{f^0}\|, & d_{NT}^{(3)} &= \max_r \|w_r' e' P_{\lambda^0}\|, \\ d_{NT}^{(4)} &= N^{-3/4}\max_r \|v_r' X_k P_{f^0}\|, & d_{NT}^{(5)} &= N^{-3/4}\max_r \|w_r' X_k' P_{\lambda^0}\|. \end{aligned} \tag{S.46}$$

Furthermore, define $d_{NT} = \max\left(1,\ d_{NT}^{(1)},\ d_{NT}^{(2)},\ d_{NT}^{(3)},\ d_{NT}^{(4)},\ d_{NT}^{(5)}\right)$. Then, Assumption EV($iii$) can be summarized as $d_{NT}q_{NT} = o_P(N^{1/4})$, i.e. given our choice of q_{NT} we need to show that $d_{NT} = o_P(N^{1/12}/\log(N))$.

We decompose $X_k = X_k^{(A)} + X_k^{(B)}$, where $X_k^{(A)} = \overline{X}_k + \widetilde{X}_k^{(str)} + \sum_{\tau=1}^{t-1}\gamma_\tau[eP_{f^0}]_{i,t-\tau}$ and $X_k^{(B)} = \sum_{\tau=1}^{t-1}\gamma_\tau[eM_{f^0}]_{i,t-\tau}$. Note that $X_k^{(A)}$ contains the strictly exogenous part of the regressor X_k, but also contains the part of $\widetilde{X}_{k,it}^{\text{weak}}$, which is independent of $M_{\lambda^0}eM_{f^0}$, i.e. $X_k^{(A)}$ is independent of ρ_r, v_r and w_r. $X_k^{(B)}$ is the part of the weakly exogenous part $\widetilde{X}_{k,it}^{\text{weak}}$ that is not independent of ρ_r, v_r and w_r. Following the decomposition $X_k = X_k^{(A)} + X_k^{(B)}$ we also introduce the corresponding decomposition of $d_{NT}^{(1)} = d_{NT}^{(A,1)} + d_{NT}^{(B,1)}$, where $d_{NT}^{(A,1)} = \max_{r,s,k}|v_r' X_k^{(A)} w_s|$ and

$d_{NT}^{(B,1)} = \max_{r,s,k} |v_r' \widetilde{X}_k^{(B)} w_s|$, and analogously we define $d_{NT}^{(A,4)}$, $d_{NT}^{(B,4)}$, $d_{NT}^{(A,5)}$, and $d_{NT}^{(B,5)}$. Note that $d_{NT}^{(1)} \leq d_{NT}^{(A,1)} + d_{NT}^{(B,1)}$ and analogously for $d_{NT}^{(4)}$ and $d_{NT}^{(5)}$.

Using Lemma S.13 and Holder's inequality we have for sufficiently large N, T

$$
\begin{aligned}
&\mathbb{E}\left[\left| v_r' X_k^{(A)} w_s \right|^{25} \middle| X_k^{(A)} \right] \\
&= \mathbb{E}\left[\left| \frac{\widetilde{v}' M_{\lambda^0} X_k^{(A)} M_{f^0} \widetilde{w}}{\|M_{\lambda^0}\widetilde{v}\| \|M_{f^0}\widetilde{w}\|} \right|^{25} \middle| X_k^{(A)} \right] \\
&= \mathbb{E}\left[\left(\sqrt{N} \|M_{\lambda^0}\widetilde{v}\|^{-1} \sqrt{T} \|M_{f^0}\widetilde{w}\|^{-1} \left| \frac{1}{\sqrt{NT}} \sum_{i,t} \widetilde{v}_i \widetilde{w}_t [M_{\lambda^0} X_k^{(A)} M_{f^0}]_{it} \right| \right)^{25} \middle| X_k^{(A)} \right] \\
&\leq \left\{ \mathbb{E}\left(\sqrt{N} \|M_{\lambda^0}\widetilde{v}\|^{-1} \right)^{\xi} \right\}^{25/\xi} \left\{ \mathbb{E}\left(\sqrt{T} \|M_{f^0}\widetilde{w}\|^{-1} \right)^{\xi} \right\}^{25/\xi} \\
&\qquad \left\{ \mathbb{E}\left[\left(\frac{1}{\sqrt{NT}} \sum_{i,t} \widetilde{v}_i \widetilde{w}_t [M_{\lambda^0} X_k^{(A)} M_{f^0}]_{it} \right)^{26} \middle| X_k^{(A)} \right] \right\}^{25/26} \\
&\leq C \left[\frac{1}{NT} \sum_{it} \left([M_{\lambda^0} X_k^{(A)} M_{f^0}]_{it} \right)^2 \right]^{13},
\end{aligned}
$$

where ξ satisfied $2/\xi + 1/26 = 1/25$, and C is a global constant. Here, as everywhere else in the paper and supplementary material, we implicitly also condition on λ^0 and f^0. Since we assume that $\mathbb{E}\,|(M_{\lambda^0} X_k M_{f^0})_{it}|^{26}$ and therefore also $\mathbb{E}\left|(M_{\lambda^0} X_k^{(A)} M_{f^0})_{it}\right|^{26}$ is uniformly bounded we thus obtain that $\mathbb{E}\left[\left|v_r' X_k^{(A)} w_s\right|^{25}\right]$ is also uniformly bounded. We thus conclude that

$$
\mathbb{E}\left(\max_{r,s} |v_r' X_k^{(A)} w_s| \right)^{25} = \mathbb{E}\left(\max_{r,s} |v_r' X_k^{(A)} w_s|^{25} \right) \leq \mathbb{E}\left(\sum_{r,s} |v_r' X_k^{(A)} w_s|^{25} \right) = \mathcal{O}\left(N^2\right), \quad \text{(S.47)}
$$

which implies that $d_{NT}^{(A,1)} = \mathcal{O}_P\left(N^{2/25}\right) = o_P(N^{1/12}/\log(N))$.

We have

$$
\begin{aligned}
d_{NT}^{(A,4)} = N^{-3/4} \max_{r,k} \|v_r' X_k^{(A)} P_{f^0}\| &\leq N^{-3/4} \max_{r,k} \|v_r' X_k^{(A)}\| \\
&\leq N^{-3/4} \sqrt{T} \max_{r,t,k} |v_r' X_{k,\cdot t}^{(A)}| \\
&\leq N^{-3/4} \sqrt{T} \max_{r,t,k} \sum_{i=1}^{N} v_{r,i} X_{k,it}^{(A)}, \qquad \text{(S.48)}
\end{aligned}
$$

where $t = 1, \ldots, T$, and we applied the inequality $\|z\| \leq \sqrt{T} \max_t z_t$, which holds for all T-vectors z. The remaining treatment of $d_{NT}^{(4)}$ is analogous to that of $d_{NT}^{(1)}$. Using Lemma S.13

and the assumption that $(M_{\lambda^0} X_k)_{it}$ and thus also $(M_{\lambda^0} X_k^{(A)})_{it}$ has uniformly bounded 8'th moment one can show that $\mathbb{E}\left[\left|\sum_{i=1}^N v_{r,i} X_{k,it}\right|^7\right]$ is also uniformly bounded, which then implies $d_{NT}^{(A,4)} = \mathcal{O}_P\left(N^{-3/4}\sqrt{T}N^{2/7}\right) = \mathcal{O}_P\left(N^{1/28}\right) = o_P(N^{1/12}/\log(N))$. Analogously one obtains $d_{NT}^{(A,5)} = o_P(N^{1/12}/\log(N))$.

Since $[M_{\lambda^0}\widetilde{X}_k^{(B)}]_{it} = \sum_{\tau=1}^{t-1}\gamma_\tau [M_{\lambda^0} e M_{f^0}]_{i,t-\tau} = \sum_{r=1}^{Q}\sqrt{\rho_r}\, v_{r,i}\sum_{\tau=1}^{t-1}\gamma_\tau w'_{r,t-\tau}$ we find

$$
\begin{aligned}
d_{NT}^{(B,1)} &= \max_{r,s,k} |v_r' \widetilde{X}_k^{(B)} w_s| \\
&= \max_{r,s,k} |v_r' M_{\lambda^0} \widetilde{X}_k^{(B)} w_s| \\
&= \max_{r,s,k} \left| \sqrt{\rho_r} \sum_{t=1}^{T} \sum_{\tau=1}^{t-1} \gamma_\tau w'_{r,t-\tau} w_{s,t} \right| \\
&\leq \sqrt{\rho_1} \max_{r,s,k} \left| \sum_{t=1}^{T} \sum_{\tau=1}^{t-1} \gamma_\tau w'_{r,t-\tau} w_{s,t} \right| \\
&\leq \|e\| \max_{r,s,k} \left(\sum_{\tau=1}^{T-1} |\gamma_\tau| \right) \left(\max_{r,s,\tau} \left| \sum_{t=\tau+1}^{T} w_{r,t} w_{s,t-\tau} \right| \right) = \mathcal{O}_P\left(N^{\epsilon}\right) = o_P(N^{1/12}/\log(N)), \quad \text{(S.49)}
\end{aligned}
$$

where we used that $v_r' v_r = 1$ and $v_r' v_s = 0$ for $r \neq s$, and we also employed Lemma S.13 in the last step, which guarantees that $\epsilon < 1/12$. We thus have shown that $d_{NT}^{(1)} = o_P(N^{1/12}/\log(N))$.

We have $\|X_k^{(B)}\| \leq \sum_{\tau=1}^{t-1} |\gamma_\tau| \|e\| = \mathcal{O}_P(\sqrt{N})$ and therefore

$$
d_{NT}^{(B,4)} = N^{-3/4} \max_{r} \|v_r' X_k^{(B)} P_{f^0}\| \leq N^{-3/4} \|X_k^{(B)}\| = \mathcal{O}_P\left(N^{-1/4}\right), \tag{S.50}
$$

and therefore $d_{NT}^{(4)} = o_P(N^{1/12}/\log(N))$. Analogously we obtain $d_{NT}^{(B,5)} = \mathcal{O}_P\left(N^{-1/4}\right)$ and thus $d_{NT}^{(5)} = o_P(N^{1/12}/\log(N))$.

Let $\widetilde{f}$ be an $N \times R^0$ matrix such that $P_{f^0} = P_{\widetilde{f}}$, i.e. the column spaces of f^0 and $\widetilde{f}$ are identical, and $\widetilde{f}'\widetilde{f} = \mathbb{1}_{R^0}$. Then we have $\|v_r' e P_{f^0}\| = \|v_r' e \widetilde{f}'\|$. Note that $e\widetilde{f}'$ is an $N \times R^0$ matrix with iid normal entries, independently distributed of v_r for all $r = 1, \ldots, Q$. Together with the distributional characterization of v_r in Lemma S.13 it is then easy to show that $\max_r \|v_r' e P_{f^0}\| = \mathcal{O}_P(N^{\delta})$ for any $\delta > 0$, and the same is true for $\max_r \|w_r' e' P_{\lambda^0}\|$, i.e. we have $d_{NT}^{(2)} = o_P(N^{1/12}/\log(N))$ and $d_{NT}^{(3)} = o_P(N^{1/12}/\log(N))$. We have thus shown that Assumption EV(iii) holds. ■

S.5 Estimated Factors, Loadings, Variance, and Bias (Proof of Theorem 3.2)

The goal of this section if to prove Theorem 3.2 in the main text. The Lemmas S.14 and S.15 are intermediate results that are used in the proof of the main theorem below.

Lemma S.14. *Let A, B, ν_i, $i = 1, \ldots, n$, b be defined as in Lemma S.11 (but we do not require the assumption of Lemma S.11 here). Assume that for all $r \in \{1, 2, \ldots, R+1\}$ we have $|\mu_r(A+B) - \mu_r(A)| \leq c_1$ and for all $r \in \{1, 2, \ldots, R\}$ we have $\mu_r(A) - \mu_{r+1}(A) \geq c_2$ for positive constants c_1 and c_2.*[5] *Furthermore, let $\widetilde{\nu}_i$, $i = 1, \ldots, n$ be the eigenvector of $A+B$ corresponding to the eigenvalue $\mu_i(A+B)$, normalized such that $\|\widetilde{\nu}_i\| = 1$. Then for $r \in \{1, 2, \ldots, R\}$ we have*

$$\|\widetilde{\nu}_r - \nu_r\|^2 \leq \frac{2(4^r - 1)(b + c_1)}{3c_2}.$$

Proof of Lemma S.14. Since A and B are symmetric $\{\nu_i\}$ and $\{\widetilde{\nu}_i\}$ are orthogonal bases of $\mathbb{R}^n$. With $\omega_{ri} := \nu_r' \widetilde{\nu}_i \in [-1, 1]$ we thus have $\nu_r = \sum_{i=1}^n \omega_{ri} \widetilde{\nu}_i$. Note that $\sum_{i=1}^n \omega_{ri}^2 = 1$, and also $\sum_{r=1}^n \omega_{ri}^2 = 1$. Let $q \in \{1, \ldots, R\}$. We have

$$\left| \sum_{r=1}^q \nu_r'(A+B)\nu_r - \sum_{r=1}^q \mu_r(A) \right| = \left| \sum_{r=1}^q \nu_r' B \nu_r \right| \leq qb,$$

and

$$\sum_{r=1}^q \nu_r'(A+B)\nu_r = \sum_{r=1}^q \sum_{i,j=1}^n \omega_{ri}\omega_{rj} \widetilde{\nu}_i'(A+B)\widetilde{\nu}_j = \sum_{r=1}^q \sum_{i=1}^n \omega_{ri}^2 \mu_i(A+B).$$

Therefore

$$\sum_{r=1}^q \sum_{i=1}^n \omega_{ri}^2 \mu_i(A+B) - \sum_{r=1}^q \mu_r(A) \geq -qb.$$

Using $\mu_{q+1}(A+B) \geq \mu_i(A+B)$ for $q+1 \leq i$ we obtain

$$\sum_{r=1}^q \sum_{i=1}^q \omega_{ri}^2 \mu_i(A+B) + \mu_{q+1}(A+B) \sum_{r=1}^q \sum_{i=q+1}^n \omega_{ri}^2 - \sum_{r=1}^q \mu_r(A) \geq -qb,$$

and using $|\mu_i(A+B) - \mu_i(A)| \leq c_1$ for $i \in \{1, \ldots, q+1\}$ and $\sum_{i=1}^n \omega_{ri}^2 = 1$ we find

$$\sum_{r=1}^q \sum_{i=1}^q \omega_{ri}^2 \mu_i(A) + \mu_{q+1}(A) \sum_{r=1}^q \underbrace{\sum_{i=q+1}^n \omega_{ri}^2}_{=1-\sum_{i=1}^q \omega_{ri}^2} - \underbrace{\sum_{r=1}^q \mu_r(A)}_{=\sum_{i=1}^q \mu_i(A)} \geq -qb - qc_1.$$

The last inequality can be rewritten as

$$\sum_{i=1}^q [\mu_i(A) - \mu_{q+1}(A)] \left[1 - \sum_{r=1}^q \omega_{ri}^2\right] \leq q(b + c_1).$$

[5] The inequality $|\mu_r(A+B) - \mu_r(A)| \leq c_1$ can be justified by applying Lemma S.11.

Using $\mu_i(A) - \mu_{q+1}(A) \geq c_2$ for $i \in \{1, \ldots, q\}$ we obtain

$$q - \sum_{i=1}^{q} \sum_{r=1}^{q} \omega_{ri}^2 \leq \frac{q(b+c_1)}{c_2}. \tag{S.51}$$

For $q = 1$ we find $1 - \omega_{11}^2 \leq \frac{b+c_1}{c_2}$, which also implies that $\omega_{12}^2 \leq 1 - \omega_{11}^2 \leq \frac{b+c_1}{c_2}$ and $\omega_{21}^2 \leq 1 - \omega_{11}^2 \leq \frac{b+c_1}{c_2}$. Using these results and (S.51) for $q = 2$ gives $1 - \omega_{22}^2 \leq \frac{2(b+c_1)}{c_2} + (1 - \omega_{11}^2) + \omega_{12}^2 + \omega_{21}^2 \leq \frac{5(b+c_1)}{c_2}$. By continuing to apply (S.51) recursively for increasing q one obtains for $r \in \{1, \ldots, R\}$ that

$$1 - \omega_{rr}^2 \leq \frac{(4^r - 1)}{3} \frac{(b+c_1)}{c_2},$$

and therefore

$$\|\widetilde{\nu}_r - \nu_r\|^2 = (\widetilde{\nu}_r - \nu_r)'(\widetilde{\nu}_r - \nu_r) = 2\left(1 - \omega_{rr}^2\right) \leq \frac{2(4^r - 1)(b+c_1)}{3c_2}.$$

■

It is convenient to introduce some additional notation. We decompose $\widehat{F}_R = (\widehat{F}_R^0, \widehat{F}_R^{\text{red}})$, where the $T \times R^0$ matrix $\widehat{F}_R^0$ contains the eigenvectors corresponding to the R^0 largest eigenvalues of the $T \times T$ matrx $\left(Y - \widehat{\beta}_R \cdot X\right)' \left(Y - \widehat{\beta}_R \cdot X\right)$, while the $T \times (R - R^0)$ matrix $\widehat{F}_R^{\text{red}}$ contains the eigenvalues corresponding to the next $R - R^0$ largest eigenvalues of this $T \times T$ matrix. Note that $\widehat{F}_R^0 = \widehat{f}(\widehat{\beta}_R)$. By applying Lemma S.2 we find that $P_{\widehat{F}_R^0} = \mathbb{1}_T - M_{\widehat{F}_R^0}$ satisfies $\left\| P_{\widehat{F}_R^0} - P_{f^0} \right\| = \mathcal{O}_P(1/\sqrt{N})$ under our assumptions. See also Bai (2009) and Moon and Weidner (2014) for further details on the estimated factors for $R = R^0$. The new difficulty in this section is to work out the asymptotic behavior of the redundantly estimated factors $\widehat{F}_R^{\text{red}}$. Note that $\widehat{F}_R^{\text{red}}$ are the leading principal components (eigenvalues corresponding to $R - R^0$ largest eigenvalues) of $\widehat{e}'(\widehat{\beta}_R)\widehat{e}(\widehat{\beta}_R)$, with $\widehat{e}(\beta)$ defined at the beginning of Section S.2 (residuals after subtracting only R^0 principal component).

Ananlogous to the decomposition $\widehat{F}_R = (\widehat{F}_R^0, \widehat{F}_R^{\text{red}})$ we also introduce the decomposition $\widehat{\Lambda}_R = (\widehat{\Lambda}_R^0, \widehat{\Lambda}_R^{\text{red}})$ for the estimated factor loadings.

Following Moon and Weidner (2014) we define the truncation kernel function $\Gamma : \mathbb{R} \to \mathbb{R}$ by $\Gamma(x) = 1$ for $|x| \leq 1$, and $\Gamma(x) = 0$ otherwise. Furthermore, for a $T \times T$ matrix B with elements B_{ts} we define the right-sided Kernel truncation of B as the $T \times T$ matrix B^{truncR} with elements $B_{ts}^{\text{truncR}} = \Gamma\left(\frac{s-t}{M}\right) B_{ts}$ for $t < s$, and $B_{ts}^{\text{truncR}} = 0$ otherwise. Note that B^{truncR} depends on the bandwidth parameter M, but this dependence is suppressed in the notation. With this definition we have

$$\widehat{B}_{R,k} = \sum_{t=1}^{T} \sum_{s=t+1}^{t+M} P_{\widehat{F}_R,ts} \left[\frac{1}{N} \sum_{i=1}^{N} \widehat{e}_{R,it} X_{k,is} \right] = \frac{1}{N} \mathrm{Tr}\left[P_{\widehat{F}_R} \left(\widehat{e}'_R X_k \right)^{\text{truncR}} \right]$$

Lemma S.15. *Under the assumptions of Theorem 3.2 we have* *(i)* $N^{-1}\left\|\mathbb{E}(e'X_k) - (\widehat{e}'_R X_k)^{\mathrm{truncR}}\right\| = o_P(1)$,

(ii) $N^{-1}\left\|\mathbb{E}(e'X_k)\right\| = \mathcal{O}(1)$,

(iii) $N^{-1}\mathrm{Tr}\left[P_{f^{\mathrm{red}}}\mathbb{E}(e'X_k)\right] = o_P(1)$.

Proof of Lemma S.15. # Part (i): For $R = R^0$ statement (i) is identical to Lemma S.10.5(i) in Moon and Weidner (2014), and the proof there also applies to $R > R^0$, with only one additional issue left to work out: Namely, we have $\widehat{e}_R = \widehat{e}(\widehat{\beta}_R) - \widehat{\Lambda}^{\mathrm{red}\prime}_R \widehat{F}^{\mathrm{red}}_R$, i.e. we have to account for the fact that $R - R^0$ redundant principal components are subtracted from the residuals $\widehat{e}(\widehat{\beta}_R)$ that were introduced in Section S.2 based on the correct number of factors R^0. The fact that $\widehat{\beta}_R$ instead of $\widehat{\beta}_{R^0}$ is used to define the residuals makes no difference in the proof, since both are $\sqrt{NT}$ consistent under our assumptions. In addition to the proof of Lemma S.10.5 already provided in Moon and Weidner (2014) we therefore also need to show that

$$N^{-1}\left\|\left(\widehat{F}^{\mathrm{red}\prime}_R \widehat{\Lambda}^{\mathrm{red}}_R X_k\right)^{\mathrm{truncR}}\right\| = o_P(1).$$

We have

$$\widehat{\Lambda}^{\mathrm{red}\prime}_R \widehat{F}^{\mathrm{red}}_R = \left[\widehat{e}(\widehat{\beta}_R)\right] P_{\widehat{F}^{\mathrm{red}}_R}, \qquad\qquad \lambda^{\mathrm{red}\prime} f^{\mathrm{red}} = M_{\lambda^0} e M_{f^0} P_{f^{\mathrm{red}}} = M_{\lambda^0} e P_{f^{\mathrm{red}}}.$$

The current lemma is only used for the proof of the last part of Theorem 3.2 (i.e. $\widehat{B}_R = B + o_P(1)$). The proof of Theorem 3.2 below starts by showing $\left\|P_{\widehat{F}^{\mathrm{red}}_R} - P_{f^{\mathrm{red}}}\right\| = \mathcal{O}_P(N^{-1/6}\log N)$, and we will already make use of this result here. Applying Lemma S.3 we furthermore find that $\left\|\widehat{e}(\widehat{\beta}_R) - M_{\lambda^0} e M_{f^0}\right\| = \mathcal{O}_P(1)$. Using this we find that

$$\begin{aligned}
\left\|\widehat{\Lambda}^{\mathrm{red}\prime}_R \widehat{F}^{\mathrm{red}}_R - \lambda^{\mathrm{red}\prime} f^{\mathrm{red}}\right\| &= \left\|\left[\widehat{e}(\widehat{\beta}_R)\right] P_{\widehat{F}^{\mathrm{red}}_R} - M_{\lambda^0} e M_{f^0} P_{f^{\mathrm{red}}}\right\| \\
&= \left\|\left(\left[\widehat{e}(\widehat{\beta}_R)\right] - M_{\lambda^0} e M_{f^0}\right) P_{\widehat{F}^{\mathrm{red}}_R} + M_{\lambda^0} e M_{f^0}\left(P_{\widehat{F}^{\mathrm{red}}_R} - P_{f^{\mathrm{red}}}\right)\right\| \\
&\leq \left\|\widehat{e}(\widehat{\beta}_R) - M_{\lambda^0} e M_{f^0}\right\| + \|e\|\left\|P_{\widehat{F}^{\mathrm{red}}_R} - P_{f^{\mathrm{red}}}\right\| \\
&= \mathcal{O}_P(1) + \mathcal{O}_P(N^{1/2-1/6}\log N) = \mathcal{O}_P(N^{1/3}\log N).
\end{aligned}$$

Let $C = \widehat{\Lambda}^{\mathrm{red}\prime}_R \widehat{F}^{\mathrm{red}}_R - \lambda^{\mathrm{red}\prime} f^{\mathrm{red}}$. We have shown $\|C\| = \mathcal{O}_P(N^{1/3}\log N)$. For $t = 1, \ldots, T$ let C_t and $X_{k,t}$ be the t'th column of the $N \times T$ matrices C and X_k, so that $C'_t X_{k,s}$ is the element at position (t,s) in the $T \times T$ matrix $C'X_k$. Remember also that we assume $\max_t \|X_{k,t}\| = \mathcal{O}_P(\log N\sqrt{N})$. Using Lemma S.8.3 in Moon and Weidner (2014) we have

$$\begin{aligned}
N^{-1}\left\|(C'X_k)^{\mathrm{truncR}}\right\| &\leq \frac{M}{N}\max_{t,s}|C'_t X_{k,s}| \leq \frac{M}{N}\max_{t,s}\|C_t\|\,\|X_{k,s}\| \leq \frac{M}{N}\|C\|\max_t\|X_{k,t}\| \\
&= \mathcal{O}_P\left(\frac{MN^{1/3+1/2}(\log N)^2}{N}\right) = \mathcal{O}_P\left(MN^{-1/6}(\log N)^2\right) = o_P(1).
\end{aligned}$$

We have

$$N^{-1}\left\|\left(\widehat{F}_R^{\mathrm{red}\prime}\widehat{\Lambda}_R^{\mathrm{red}}X_k\right)^{\mathrm{truncR}}\right\| \leq N^{-1}\left\|\left(f^{\mathrm{red}\prime}\lambda^{\mathrm{red}}X_k\right)^{\mathrm{truncR}}\right\| + N^{-1}\left\|\left(C'X_k\right)^{\mathrm{truncR}}\right\|$$
$$= N^{-1}\left\|\left(f^{\mathrm{red}\prime}\lambda^{\mathrm{red}}X_k\right)^{\mathrm{truncR}}\right\| + o_P(1).$$

Thus, what is left to show is that $N^{-1}\left\|\left(f^{\mathrm{red}\prime}\lambda^{\mathrm{red}}X_k\right)^{\mathrm{truncR}}\right\| = o_P(1)$. In the notation of Assumption EV we have $f^{\mathrm{red}\prime}\lambda^{\mathrm{red}} = \sum_{r=1}^{R-R^0}\sqrt{\rho_r}w_r v_r'$. We have $\sqrt{\rho_r} \leq \|e\| = \mathcal{O}_P(\sqrt{N})$. The distribution of the unit vectors w_r and w_r is characterized by Lemma S.13, from which it is easy to show that $\max_t |w_{r,t}| = \mathcal{O}_P(N^{-1/2+1/8})$. Again using Lemma S.8.3 in Moon and Weidner (2014) we thus find

$$N^{-1}\left\|\left(f^{\mathrm{red}\prime}\lambda^{\mathrm{red}}X_k\right)^{\mathrm{truncR}}\right\| = N^{-1}\left\|\left(\sum_{r=1}^{R-R^0}\sqrt{\rho_r}w_r v_r' X_k\right)^{\mathrm{truncR}}\right\| \leq \sum_{r=1}^{R-R^0}\frac{\sqrt{\rho_r}}{N}\left\|\left(w_r v_r' X_k\right)^{\mathrm{truncR}}\right\|$$
$$\leq \sum_{r=1}^{R-R^0}\frac{M\sqrt{\rho_r}}{N}\left(\max_t |w_{r,t}|\right)\left(\max_t |v_r' X_{k,t}|\right)$$
$$\leq \sum_{r=1}^{R-R^0}\frac{M\sqrt{\rho_r}}{N}\left(\max_t |w_{r,t}|\right)\left(\max_t \|X_{k,t}\|\right) = o_P(1).$$

Part (ii): We have $N^{-1}\mathbb{E}(e_t' X_{k,s}) = N^{-1}\mathbb{E}(e_t'\widetilde{X}_{k,s}^{\mathrm{weak}}) = \sigma^2\gamma_{s-t}$ for $s > t$, and $= 0$ otherwise. Therefore $N^{-1}\|\mathbb{E}(e'X_k)\| \leq \sqrt{N^{-1}\|\mathbb{E}(e'X_k)\|_1 N^{-1}\|\mathbb{E}(e'X_k)\|_\infty} \leq \sigma^2\sum_{t=1}^{\infty}|\gamma_t| = \mathcal{O}(1)$.

Part (iii): In the notation of Assumption EV we have $P_{f^{\mathrm{red}}} = P_{(w_1,\ldots,w_{R-R^0})} = \sum_{r=1}^{R-R^0} w_r w_r'$. We thus have

$$N^{-1}\mathrm{Tr}\left[P_{f^{\mathrm{red}}}\mathbb{E}(e'X_k)\right] = \sum_{r=1}^{R-R^0}\sum_{t=1}^{T}\sum_{s=t+1}^{T}\gamma_{s-t}w_{r,s}w_{r,t}.$$

Again, using the distributional characterization of w_r in Lemma S.13 it is easy to show that this term is $o_P(1)$. ■

Proof of Theorem 3.2 (Consistency of Bias and Variance Estimators).
Consistency for factors: We want to apply Lemma S.14 with $A = M_{f^0}e'M_{\lambda^0}eM_{f^0}$ and $A+B = \widehat{e}'(\widehat{\beta}_R)\widehat{e}(\widehat{\beta}_R)$, i.e. in the notation of Lemma S.14 we have $f^{\mathrm{red}} = (\nu_1,\nu_2,\ldots,\nu_{R-R^0})H_1$ and $\widehat{F}_R^{\mathrm{red}} = (\widetilde{\nu}_1,\widetilde{\nu}_2,\ldots,\widetilde{\nu}_{R-R^0})H_2$, for some invertible $(R-R^0)\times(R-R^0)$ matrices H_1 and H_2. Note that $\mu_r(A) = \rho_r$ by definition of ρ_r in the main text, and by applying Lemma S.3 and the definition of $B(\beta)$ (which is different from B here) in Lemma S.9 we find $\mu_r(A+B) = \mu_r(B(\beta)+B'(\beta)) + \mathcal{O}_P(1)$, so that by also applying Lemma S.12 we find $|\mu_r(A+B) - \mu_r(A)| = \mathcal{O}_P(1)$. Note that Lemma S.12 only states this for $r = 1,\ldots,R-R^0$, but it can also be shown for $r = R-R^0+1$ by following the proof in Lemma S.12, which is what we require here, since we want to apply Lemma S.14 with the R in the Lemma equal to $R-R^0$ here. The assumption $|\mu_r(A+B) - \mu_r(A)| \leq c_1$ is therefore satisfied with $c_1 = \mathcal{O}_P(1)$. Following the steps in the

proof of Lemma S.9 and Lemma S.12 we also find that $b = \max_{i,j=1,\ldots,n} |\nu_i' B \nu_j| = \mathcal{O}_P(1)$ here. From Johnstone (2001) and Soshnikov (2002) it is known that $(\mu_1(A), \ldots, \mu_{R-R^0+1}(A))$ properly shifted and rescaled (by $N^{1/3}$) are jointly asymptotically distributed according to the Tracy-Widom law, which is a continuous distribution, from which we can conclude that $\mu_r(A) - \mu_{r+1}(A) \geq c_2$ holds for $c_2 = o_P(N^{1/3})$, e.g. $c_2 = N^{1/3}/\log(N)^2$. By applying Lemma S.14 we thus obtain $\|\widetilde{\nu}_r - \nu_r\|^2 = \mathcal{O}_P(N^{-1/3}(\log N)^2)$ for $r \in \{1, \ldots, R - R^0\}$. We thus also have

$$\left\| P_{\widehat{F}_R^{\rm red}} - P_{f^{\rm red}} \right\| = \left\| P_{(\nu_1,\ldots,\nu_{R-R^0})} - P_{(\widetilde{\nu}_1,\ldots,\widetilde{\nu}_{R-R^0})} \right\| = \mathcal{O}_P(N^{-1/6} \log N).$$

Together with the above result $\left\| P_{\widehat{F}_R^0} - P_{f^0} \right\| = \mathcal{O}_P(1/\sqrt{N})$ we thus have

$$\left\| P_{\widehat{F}_R} - P_{[f^0, f^{\rm red}]} \right\| = \mathcal{O}_P(N^{-1/6} \log N) = o_P(1),$$

which is the first statement of Theorem 3.2. Note that f^0 and $f^{\rm red}$ are orthogonal (i.e. we have $f^{0\prime} f^{\rm red} = 0$), so that $P_{[f^0, f^{\rm red}]} = P_{f^0} + P_{f^{\rm red}}$.

Consistency for factors loadings: The problem is symmetric under exchange of $N \leftrightarrow T$, so analogous to the proof for the factors one also finds $\left\| P_{\widehat{\Lambda}_R} - P_{[\lambda^0, \lambda^{\rm red}]} \right\| = o_P(1)$.

Consistency of $\widehat{\sigma}_R^2$: Using the definition of $\mathcal{L}_{NT}^R(\beta)$, Lemma S.4, Theorem 4.2, the definition of $\mathcal{L}_{NT}^0(\beta^0)$, and the WLLN we obtain

$$\begin{aligned}
\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (\widehat{e}_{R,it})^2 &= \mathcal{L}_{NT}^R(\widehat{\beta}_R) = \mathcal{L}_{NT}^0(\widehat{\beta}_R) + \mathcal{O}_P\left(\frac{\|e\|^2}{NT} + \frac{\sqrt{N}}{NT} \right) = \mathcal{L}_{NT}^0(\beta^0) + \mathcal{O}_P(1/N) \\
&= \frac{1}{NT} \mathrm{Tr}(M_{\widehat{\lambda}} e M_{\widehat{f}} e') + \mathcal{O}_P(1/N) = \frac{1}{NT} \mathrm{Tr}(ee') + \mathcal{O}_P\left(\frac{\|e\|^2}{NT} + 1/N \right) \\
&= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T e_{it}^2 + \mathcal{O}_P(1/N) = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \mathbb{E}(e_{it}^2) + o_P(1) \\
&= \sigma^2 + o_P(1).
\end{aligned}$$

We thus also have $\widehat{\sigma}_R^2 = \sigma^2 + o_P(1)$.

Consistency of $\widehat{W}_R$: Using that $M_{\widehat{F}_R} - M_{[f^0, f^{\rm red}]} = P_{[f^0, f^{\rm red}]} - P_{\widehat{F}_R}$ and $M_{\widehat{\Lambda}_R} - M_{[\lambda^0, \lambda^{\rm red}]} = P_{[\lambda^0, \lambda^{\rm red}]} - P_{\widehat{\Lambda}_R}$ are low rank matrices (rank $\leq 2R$) and satisfy the spectral norm bounds

$\left\| M_{\widehat{F}_R} - M_{[f^0, f^{\rm red}]} \right\| = o_P(1)$ and $\left\| M_{\widehat{\Lambda}_R} - M_{[\lambda^0, \lambda^{\rm red}]} \right\| = o_P(1)$ we obtain

$$\begin{aligned}
\widehat{W}_{R,k_1k_2} &= \frac{1}{NT} \mathrm{Tr}\left(M_{\widehat{\Lambda}_R} X_{k_1} M_{\widehat{F}_R} X'_{k_2} \right) \\
&= \frac{1}{NT} \mathrm{Tr}\left(M_{[\lambda^0, \lambda^{\rm red}]} X_{k_1} M_{[f^0, f^{\rm red}]} X'_{k_2} \right) + o_P\left(\frac{(\|X_{k_1}\| + \|X_{k_2}\|)^2}{NT} \right) \\
&= \frac{1}{NT} \mathrm{Tr}\left(M_{[\lambda^0, \lambda^{\rm red}]} X_{k_1} M_{[f^0, f^{\rm red}]} X'_{k_2} \right) + o_P(1) \\
&= \underbrace{\frac{1}{NT} \mathrm{Tr}\left(M_{\lambda^0} X_{k_1} M_{f^0} X'_{k_2} \right)}_{=W_{k_1k_2}} - \frac{1}{NT} \mathrm{Tr}\left(P_{\lambda^{\rm red}} X_{k_1} M_{f^0} X'_{k_2} \right) \\
&\qquad - \frac{1}{NT} \mathrm{Tr}\left(M_{\lambda^0} X_{k_1} P_{f^{\rm red}} X'_{k_2} \right) + \frac{1}{NT} \mathrm{Tr}\left(P_{\lambda^{\rm red}} X_{k_1} P_{f^{\rm red}} X'_{k_2} \right),
\end{aligned}$$

where in the last step we used that $M_{[\lambda^0, \lambda^{\rm red}]} = M_{\lambda^0} - P_{\lambda^{\rm red}}$ and $M_{[f^0, f^{\rm red}]} = M_{f^0} - P_{f^{\rm red}}$. Remember that $P_{f^{\rm red}} = P_{(\nu_1, \nu_2, \ldots, \nu_{R-R^0})}$ in the notation used here, and $P_{f^{\rm red}} = P_{(w_1, w_2, \ldots, w_{R-R^0})}$ in the notation of Assumption EV. Using the characterization of the distribution of w_s given in Lemma S.13, and the methods used in the proof of Lemma S.12 we obtain $\left\| X_k P_{f^{\rm red}} \right\| = o_P(\sqrt{NT})$, and analogously $\| P_{\lambda^{\rm red}} X_k \| = o_P(\sqrt{NT})$. We thus obtain

$$\left| \frac{1}{NT} \mathrm{Tr}\left(P_{\lambda^{\rm red}} X_{k_1} M_{f^0} X'_{k_2} \right) \right| \leq \frac{R}{NT} \| P_{\lambda^{\rm red}} X_{k_1} \| \, \| X_{k_2} \| = o_P(1),$$

and analogously we find $\frac{1}{NT} \mathrm{Tr}\left(M_{\lambda^0} X_{k_1} P_{f^{\rm red}} X'_{k_2} \right) = o_P(1)$ and $\frac{1}{NT} \mathrm{Tr}\left(P_{\lambda^{\rm red}} X_{k_1} P_{f^{\rm red}} X'_{k_2} \right) = o_P(1)$. Combining this with the above result for $\widehat{W}_{R,k_1k_2}$ gives $\widehat{W}_R = W + o_P(1)$.

\# Consistency $\widehat{B}_R$: Remember that $B_k = \frac{1}{N} \mathrm{Tr}[P_{f^0} \mathbb{E}(e' X_k)]$ and $\widehat{B}_{R,k} = \frac{1}{N} \mathrm{Tr}\left[P_{\widehat{F}_R} \left(\widehat{e}'_R X_k \right)^{\rm truncR} \right]$. Using Lemma S.15 we find

$$\begin{aligned}
\widehat{B}_{R,k} &= \frac{1}{N} \mathrm{Tr}\left[P_{\widehat{F}_R} \mathbb{E}(e' X_k) \right] + o_P(1) && \text{(using part (i) of the lemma)} \\
&= \frac{1}{N} \mathrm{Tr}\left[P_{[f^0, f^{\rm red}]} \mathbb{E}(e' X_k) \right] + o_P(1) && \text{(using part (ii) and } \left\| P_{\widehat{F}_R} - P_{[f^0, f^{\rm red}]} \right\| = o_P(1)\text{)} \\
&= B_k + \frac{1}{N} \mathrm{Tr}\left[P_{f^{\rm red}} \mathbb{E}(e' X_k) \right] + o_P(1) && \text{(using } P_{[f^0, f^{\rm red}]} = P_{f^0} + P_{f^{\rm red}}\text{)} \\
&= B_k + o_P(1), && \text{(using part (iii))}
\end{aligned}$$

which is the desired consistency result for $\widehat{B}_R$. ■

S.6 Proofs for Intermediate Results

S.6.1 Proofs for Expansions of $\mathcal{L}^0_{NT}(\beta)$, $M_{\widehat{\lambda}}(\beta)$, $M_{\widehat{f}}(\beta)$ and $\widehat{e}(\beta)$

Proof of Lemma S.1.

(i,ii) We apply perturbation theory in Kato (1980). The unperturbed operator is $\mathcal{T}^{(0)} = \lambda^0 f^{0\prime} f^0 \lambda^{0\prime}$, the perturbed operator is $\mathcal{T} = \mathcal{T}^{(0)} + \mathcal{T}^{(1)} + \mathcal{T}^{(2)}$ (*i.e.* the parameter κ that appears in Kato is set to 1), where $\mathcal{T}^{(1)} = \sum_{k=0}^K \epsilon_k X_k f^0 \lambda^{0\prime} + \lambda^0 f^{0\prime} \sum_{k=0}^K \epsilon_k X_k'$, and $\mathcal{T}^{(2)} = \sum_{k_1=0}^K \sum_{k_2=0}^K \epsilon_{k_1} \epsilon_{k_2} X_{k_1} X_{k_2}'$. The matrices $\mathcal{T}$ and $\mathcal{T}^0$ are real and symmetric (which implies that they are normal operators), and positive semi-definite. We know that $\mathcal{T}^{(0)}$ has an eigenvalue 0 with multiplicity $N - R^0$, and the separating distance of this eigenvalue is $d = NT d^2_{\min}(\lambda^0, f^0)$. The bound (S.34) guarantees that

$$\|\mathcal{T}^{(1)} + \mathcal{T}^{(2)}\| \leq \frac{NT}{2}\, d^2_{\min}(\lambda^0, f^0) \; . \tag{S.52}$$

By Weyl's inequality we therefore find that the $N - R^0$ smallest eigenvalues of $\mathcal{T}$ (also counting multiplicity) are all smaller than $\frac{NT}{2} d^2_{\min}(\lambda^0, f^0)$, and they "originate" from the zero-eigenvalue of $\mathcal{T}^{(0)}$, with the power series expansion for $\mathcal{L}^0_{NT}(\beta)$ given in (2.22) and (2.18) at p.77/78 of Kato, and the expansion of $M_{\widehat{\lambda}}$ given in (2.3) and (2.12) at p.75,76 of Kato. We still need to justify the convergence radius of this series. Since we set the complex parameter κ in Kato to 1, we need to show that the convergence radius (r_0 in Kato's notation) is at least 1. The condition (3.7) in Kato p.89 reads $\|\mathcal{T}^{(n)}\| \leq a c^{n-1}$, $n = 1, 2, \ldots$, and it is satisfied for $a = 2\sqrt{NT} d_{\max}(\lambda^0, f^0) \sum_{k=0}^K |\epsilon_k| \|X_k\|$ and $c = \sum_{k=0}^K |\epsilon_k| \|X_k\| / \sqrt{NT}/2/d_{\max}(\lambda^0, f^0)$. According to equation (3.51) in Kato p.95, we therefore find that the power series for $\mathcal{L}^0_{NT}(\beta)$ and $M_{\widehat{\lambda}}$ are convergent ($r_0 \geq 1$ in his notation) if $1 \leq \left(\frac{2a}{d} + c\right)^{-1}$, and this becomes exactly our condition (S.34).

When $\mathcal{L}^0_{NT}(\beta)$ is approximated up to order $G \in \mathbb{N}$, Kato's equation (3.6) at p.89 gives the following bound on the remainder

$$\left| \mathcal{L}^0_{NT}(\beta) - \frac{1}{NT} \sum_{g=2}^G \sum_{k_1=0}^K \cdots \sum_{k_g=0}^K \epsilon_{k_1} \ldots \epsilon_{k_g} L^{(g)}\left(\lambda^0, f^0, X_{k_1}, X_{k_2}, \ldots, X_{k_g}\right) \right| \leq \frac{(N - R^0)\gamma^{G+1} d^2_{\min}(\lambda^0, f^0)}{4(1-\gamma)} \, , \tag{S.53}$$

where

$$\gamma = \frac{\sum_{k=1}^K |\beta^0_k - \beta_k| \frac{\|X_k\|}{\sqrt{NT}} + \frac{\|e\|}{\sqrt{NT}}}{r_0(\lambda^0, f^0)} < 1 \; . \tag{S.54}$$

This bound again shows convergence of the series expansion, since $\gamma^{G+1} \to 0$ as $G \to \infty$. Unfortunately, for our purposes this is not a good bound since it still involves the factor $N - R^0$ (in Kato this factor is hidden since his $\widehat{\lambda}(\kappa)$ is the average of the eigenvalues, not the sum), but as we show below this can be avoided.

(iii,iv) We have $\|S^{(m)}\| = (NT d^2_{\min}(\lambda^0, f^0))^{-m}$, $\|\mathcal{T}^{(1)}_k\| \leq 2\sqrt{NT} d_{\max}(\lambda^0, f^0) \|X_k\|$, and $\|\mathcal{T}^{(2)}_{k_1 k_2}\| \leq$

$\|X_{k_1}\|\|X_{k_2}\|$. Therefore

$$\left\| S^{(m_1)}\, \mathcal{T}^{(\nu_1)}_{k_1\ldots}\, S^{(m_2)} \,\ldots\, S^{(m_P)}\, \mathcal{T}^{(\nu_P)}_{\ldots k_g}\, S^{(m_{p+1})} \right\|$$
$$\leq \left(NT d^2_{\min}(\lambda^0, f^0)\right)^{-\sum m_j} \left(2\sqrt{NT} d_{\max}(\lambda^0, f^0)\right)^{2p-\sum \nu_j} \|X_{k_1}\|\|X_{k_2}\| \ldots \|X_{k_g}\| \,. \tag{S.55}$$

We have

$$\sum_{\substack{\nu_1+\ldots+\nu_P=g \\ 2\geq\nu_j\geq 1}} 1 \leq 2^p \,,$$
$$\sum_{\substack{m_1+\ldots+m_{p+1}=p-1 \\ m_j\geq 0}} 1 \leq \sum_{\substack{m_1+\ldots+m_{p+1}=p \\ m_j\geq 0}} 1 = \frac{(2p)!}{(p!)^2} \leq 4^p \,. \tag{S.56}$$

Using this we find[6]

$$\left\| M^{(g)}\left(\lambda^0,\, f^0,\, X_{k_1},\, X_{k_2}, \ldots, X_{k_g}\right)\right\|$$
$$\leq \left(2\sqrt{NT} d_{\max}(\lambda^0, f^0)\right)^{-g} \|X_{k_1}\|\|X_{k_2}\| \ldots \|X_{k_g}\| \sum_{p=\lceil g/2\rceil}^{g} \left(\frac{32\, d^2_{\max}(\lambda^0, f^0)}{d^2_{\min}(\lambda^0, f^0)}\right)^p$$
$$\leq \frac{g}{2} \left(\frac{16\, d_{\max}(\lambda^0, f^0)}{d^2_{\min}(\lambda^0, f^0)}\right)^g \frac{\|X_{k_1}\|}{\sqrt{NT}} \frac{\|X_{k_2}\|}{\sqrt{NT}} \ldots \frac{\|X_{k_g}\|}{\sqrt{NT}} \,. \tag{S.57}$$

For $g \geq 3$ there always appears at least one factor $S^{(m)}$, $m \geq 1$, inside the trace of the terms that contribute to $L^{(g)}$, and we have $\mathrm{rank}(S^{(m)}) = R^0$ for $m \geq 1$. Using $\mathrm{Tr}(A) \leq \mathrm{rank}(A)\|A\|$, and the equations (S.55) and (S.56), we therefore find[7] for $g \geq 3$

$$\frac{1}{NT}\left| L^{(g)}\left(\lambda^0,\, f^0,\, X_{k_1},\, X_{k_2}, \ldots, X_{k_g}\right)\right|$$
$$\leq R^0\, d^2_{\min}(\lambda^0, f^0) \left(2\sqrt{NT} d_{\max}(\lambda^0, f^0)\right)^{-g}$$
$$\|X_{k_1}\|\|X_{k_2}\| \ldots \|X_{k_g}\| \sum_{p=\lceil g/2\rceil}^{g} \left(\frac{32\, d^2_{\max}(\lambda^0, f^0)}{d^2_{\min}(\lambda^0, f^0)}\right)^p$$
$$\leq \frac{R^0\, g\, d^2_{\min}(\lambda^0, f^0)}{2} \left(\frac{16\, d_{\max}(\lambda^0, f^0)}{d^2_{\min}(\lambda^0, f^0)}\right)^g \frac{\|X_{k_1}\|}{\sqrt{NT}} \frac{\|X_{k_2}\|}{\sqrt{NT}} \ldots \frac{\|X_{k_g}\|}{\sqrt{NT}} \,. \tag{S.58}$$

[6]The sum over p only starts from $\lceil g/2\rceil$, the smallest integer larger or equal $g/2$, because $\nu_1 + \ldots + \nu_P = g$ can not be satisfied for smaller p, since $\nu_j \leq 2$.

[7]The calculation for the bound of $L^{(g)}$ is almost identical to the one for $M^{(g)}$. But now there appears an additional factor R^0 from the rank, and since $\sum m_j = p - 1$ (not p as before), there is also an additional factor $NT d^2_{\min}(\lambda^0, f^0)$.

This implies for $g \geq 3$

$$\frac{1}{NT} \left| \sum_{k_1=0}^{K} \sum_{k_2=0}^{K} \cdots \sum_{k_g=0}^{K} \epsilon_{k_1} \epsilon_{k_2} \ldots \epsilon_{k_g} L^{(g)} \left(\lambda^0, f^0, X_{k_1}, X_{k_2}, \ldots, X_{k_g}\right) \right|$$
$$\leq \frac{R^0 \, g \, d_{\min}^2(\lambda^0, f^0)}{2} \left(\frac{16 \, d_{\max}(\lambda^0, f^0)}{d_{\min}^2(\lambda^0, f^0)} \right)^g \left(\sum_{k=0}^{K} \frac{\|\epsilon_k X_k\|}{\sqrt{NT}} \right)^g . \tag{S.59}$$

Therefore for $G \geq 2$ we have

$$\left| \mathcal{L}_{NT}^0 (\beta) - \frac{1}{NT} \sum_{g=2}^{G} \sum_{k_1=0}^{K} \cdots \sum_{k_g=0}^{K} \epsilon_{k_1} \ldots \epsilon_{k_g} L^{(g)} \left(\lambda^0, f^0, X_{k_1}, X_{k_2}, \ldots, X_{k_g}\right) \right|$$
$$= \frac{1}{NT} \sum_{g=G+1}^{\infty} \sum_{k_1=0}^{K} \sum_{k_2=0}^{K} \cdots \sum_{k_g=0}^{K} \epsilon_{k_1} \epsilon_{k_2} \ldots \epsilon_{k_g} L^{(g)} \left(\lambda^0, f^0, X_{k_1}, X_{k_2}, \ldots, X_{k_g}\right)$$
$$\leq \sum_{g=G+1}^{\infty} \frac{R^0 \, g \, \alpha^g \, d_{\min}^2(\lambda^0, f^0)}{2}$$
$$\leq \frac{R^0 \, (G+1) \, \alpha^{G+1} \, d_{\min}^2(\lambda^0, f^0)}{2(1-\alpha)^2} , \tag{S.60}$$

where

$$\alpha = \frac{16 \, d_{\max}(\lambda^0, f^0)}{d_{\min}^2(\lambda^0, f^0)} \sum_{k=0}^{K} \frac{\|\epsilon_k X_k\|}{\sqrt{NT}}$$
$$= \frac{16 \, d_{\max}(\lambda^0, f^0)}{d_{\min}^2(\lambda^0, f^0)} \left(\sum_{k=1}^{K} \left|\beta_k^0 - \beta_k\right| \frac{\|X_k\|}{\sqrt{NT}} + \frac{\|e\|}{\sqrt{NT}} \right) < 1 . \tag{S.61}$$

Using the same argument we can start from equation (S.57) to obtain the boundfor the remainder of the series expansion for $M_{\widehat{\lambda}}(\beta)$.

Note that compared to the bound (S.53) on the remainder, the new bound (S.60) only shows convergence of the power series within the the smaller convergence radius $\frac{d_{\min}^2(\lambda^0, f^0)}{16 \, d_{\max}(\lambda^0, f^0)} < r_0(\lambda^0, f^0)$. However, the factor $N - R^0$ does not appear in this new bound, which is crucial for our approximations.

■

Proof of Lemma S.2. The general expansion of $M_{\widehat{\lambda}}(\beta)$ is given in Lemma S.1. The present Lemma just makes this expansion explicit for the first few orders. The bound on the remainder $M_{\widehat{\lambda}}^{(\mathrm{rem})}(\beta)$ is obtained from the bound (S.57) by the same logic as in the proof of Theorem 4.2. The analogous result for $M_{\widehat{f}}(\beta)$ is obtained by applying the symmetry $N \leftrightarrow T$, $\lambda \leftrightarrow f$, $e \leftrightarrow e'$, $X_k \leftrightarrow X'_k$. ■

Proof of Lemma S.3. The general expansion of $M_{\widehat{\lambda}}(\beta)$ is given in Lemma S.1, and the analogous expansion for $M_{\widehat{f}}(\beta)$ is obtained by applying the symmetry $N \leftrightarrow T$, $\lambda \leftrightarrow f$, $e \leftrightarrow e'$,

$X_k \leftrightarrow X_k'$. Lemma S.2 above provides a more explicit version of these projector expansions. For the residuals $\widehat{e}(\beta)$ we have

$$\widehat{e}(\beta) = M_{\widehat{\lambda}}(\beta)\ (Y - \beta \cdot X)\ M_{\widehat{f}}(\beta) = M_{\widehat{\lambda}}(\beta)\ \left[e - \left(\beta - \beta^0\right) \cdot X + \lambda^0 f^{0\prime}\right]\ M_{\widehat{f}}(\beta)\ , \tag{S.62}$$

and plugging in the expansions of $M_{\widehat{\lambda}}(\beta)$ and $M_{\widehat{f}}(\beta)$ it is straightforward to derive the expansion of $\widehat{e}(\beta)$ from this, including the bound on the remainder. ■

S.6.2 Proofs for $N^{3/4}$ Convergence Rate Result

Proof of Lemma S.4. The result follows from Lemma S.9 by applying Weyl's inequality, because the terms in $B(\beta) + B'(\beta)$ in addition to $A(\beta)$ all have a spectral norm of order $\mathcal{O}_P(\sqrt{N})$ for $\sqrt{N}\|\beta - \beta^0\| \leq c$. ■

Proof of Theorem S.5. From Theorem 4.1 we know that $\sqrt{N}(\widehat{\beta}_R - \beta_0) = \mathcal{O}_P(1)$, so that the bounds in Lemma S.4 and Assumption HL1 are applicable. Since $\widehat{\beta}_R$ minimizes $\mathcal{L}_{NT}^R(\beta)$ it must in particular satisfy $\mathcal{L}_{NT}^R(\widehat{\beta}_R) \leq \mathcal{L}_{NT}^R(\beta^0)$. Applying this, Lemma S.4, and Assumption HL1 we obtain

$$\begin{aligned}
0 &\geq \mathcal{L}_{NT}^R(\widehat{\beta}_R) - \mathcal{L}_{NT}^R(\beta^0) \\
&= \mathcal{L}_{NT}^0(\widehat{\beta}_R) - \mathcal{L}_{NT}^0(\beta^0) - \frac{1}{NT} \sum_{r=1}^{R-R^0} \left[\mu_r\left(A(\widehat{\beta}_R)\right) - \mu_r\left(A(\beta^0)\right)\right] \\
&\quad + \frac{1}{NT}\mathcal{O}_P\left[\sqrt{N} + \sqrt{NT}\|\widehat{\beta}_R - \beta^0\|\right] \\
&\geq \mathcal{L}_{NT}^0(\widehat{\beta}_R) - \mathcal{L}_{NT}^0(\beta^0) - \frac{1}{NT} \sum_{r=1}^{R-R^0} \mu_r\left[M_{f^0}\left(\Delta\beta \cdot X\right)' M_{\lambda^0}\left(\Delta\beta \cdot X\right) M_{f^0}\right] \\
&\quad + \frac{1}{NT}\mathcal{O}_P\left[\sqrt{N} + N^{5/4}\|\widehat{\beta}_R - \beta^0\| + N^2\|\widehat{\beta}_R - \beta^0\|/\log(N)\right].
\end{aligned} \tag{S.63}$$

Applying Theorem 4.2 then gives

$$\begin{aligned}
&\left(\widehat{\beta}_R - \beta^0\right)' W \left(\widehat{\beta}_R - \beta^0\right) - \frac{2}{\sqrt{NT}}\left(\widehat{\beta}_R - \beta^0\right)' \left(C^{(1)} + C^{(2)}\right) \\
&\leq \frac{1}{NT}\Bigg\{ \sum_{r=1}^{R-R^0} \mu_r\left[M_{f^0}\left(\Delta\beta \cdot X\right)' M_{\lambda^0}\left(\Delta\beta \cdot X\right) M_{f^0}\right] \\
&\qquad + \mathcal{O}_P\left[\sqrt{N} + N^{5/4}\|\widehat{\beta}_R - \beta^0\| + N^2\|\widehat{\beta}_R - \beta^0\|/\log(N)\right]\Bigg\}.
\end{aligned} \tag{S.64}$$

Our assumptions guarantee $C^{(2)} = \mathcal{O}_P(1)$, and we explicitly assume $C^{(1)} = \mathcal{O}_P(N^{1/4})$. Furthermore, Assumption NC guarantees that

$$(\Delta\beta)' W (\Delta\beta) - \frac{1}{NT} \sum_{r=1}^{R-R^0} \mu_r \left[M_{f^0} \left(\Delta\beta \cdot X \right)' M_{\lambda^0} \left(\Delta\beta \cdot X \right) M_{f^0} \right] \geq b \|\Delta\beta\|^2, \tag{S.65}$$

which we apply for $\Delta\beta = \widehat{\beta}_R - \beta^0$. Thus, we obtain

$$b \left(N^{3/4} \|\widehat{\beta}_R - \beta^0\| \right)^2 \leq \mathcal{O}_P(1) + \mathcal{O}_P \left(N^{3/4} \|\widehat{\beta}_R - \beta^0\| \right) + o_P \left[\left(N^{3/4} \|\widehat{\beta}_R - \beta^0\| \right)^2 \right], \tag{S.66}$$

from which we can conclude that $N^{3/4}\|\widehat{\beta}_R - \beta^0\| = \mathcal{O}_P(1)$, which proves the lemma. ■

Proof of Lemma S.6. Note that $P_g = gg'$ and $P_h = hh'$. We decompose

$$\left(U + gCh' \right)' \left(U + gCh' \right) = A_1 + A_2(C) \,, \tag{S.67}$$

where

$$\begin{aligned} A_1 &\equiv U'U + \|g'UU'g\| \, P_{\left(M_{U'g} h\right)} + \Delta_{\max} P_{(U'g)} \,, \\ A_2(C) &\equiv \left(U + gCh' \right)' P_g \left(U + gCh' \right) - U' P_g U - \|g'UU'g\| \, P_{\left(M_{U'g} h\right)} - \Delta_{\max} P_{(U'g)}. \end{aligned} \tag{S.68}$$

By Weyl's inequality we then have

$$\sum_{r=1}^{R} \mu_r \left[\left(U + gCh' \right)' \left(U + gCh' \right) \right] \leq \sum_{r=1}^{R} \mu_r(A_1) + \sum_{r=1}^{R} \mu_r \left[A_2(C) \right] . \tag{S.69}$$

We have $A_2(C) = P_{(h,U'g)} A_2(C) P_{(h,U'g)}$, i.e. $A_2(C)$ has $T - 2Q$ zero-eigenvalues and only $2Q$ non-zero eigenvalues. Let $\widetilde{h} = (h, U'g)[(h, U'g)'(h, U'g)]^{-1/2}$, which is a $T \times 2Q$ matrix that satisfies $\widetilde{h}'\widetilde{h} = \mathbb{1}_{2Q}$ and $\widetilde{h}\widetilde{h}' = P_{(h,U'g)}$. We then have

$$\sum_{r=1}^{R} \mu_r \left[A_2(C) \right] = \sum_{r=1}^{\min(R,2Q)} \mu_r \left[\widetilde{h}' A_2(C) \widetilde{h} \right], \tag{S.70}$$

and

$$
\begin{aligned}
&\sum_{r=1}^{\min(R,2Q)} \mu_r \left[\widetilde{h}' A_2(C))\widetilde{h}\right] \\
&\leq \sum_{r=1}^{\min(R,2Q)} \mu_r \left[\widetilde{h}' \left(U + gCh'\right)' P_g \left(U + gCh'\right) \widetilde{h}\right] \\
&\quad + \sum_{r=1}^{\min(R,2Q)} \mu_r \left[\widetilde{h}' \left(-U' P_g U - \|g'UU'g\| P_{\left(M_{U'g} h\right)} - \Delta_{\max} P_{(U'g)}\right) \widetilde{h}\right] \\
&= \sum_{r=1}^{\min(R,Q)} \mu_r \left[g' \left(U + gCh'\right) \left(U + gCh'\right)' g\right] \\
&\quad - \sum_{r=2Q-\min(R,2Q)+1}^{2Q} \mu_r \left[\widetilde{h}' \left(U' P_g U + \|g'UU'g\| P_{\left(M_{U'g} h\right)} + \Delta_{\max} P_{(U'g)}\right) \widetilde{h}\right]. \qquad \text{(S.71)}
\end{aligned}
$$

Here, in the first step we again used Weyl's inequality, and in the second step we used that the Q non-zero eigenvalues of $\widetilde{h}' \left(U + gCh'\right)' gg' \left(U + gCh'\right) \widetilde{h}$ are identical to the eigenvalues of $g' \left(U + gCh'\right) \left(U + gCh'\right)' g$, and that the eigenvalues of a matrix are equal to minus the eigenvalues of the negative of the matrix (but interchanging the ordering of the eigenvalues).

The eigenvalues of $\widetilde{h}' \left(U' P_g U + \|g'UU'g\| P_{\left(M_{U'g} h\right)} + \Delta_{\max} P_{(U'g)}\right) \widetilde{h}$ are given by Q eigenvalues equal to $\|g'UU'g\|$ (stemming from $\|g'UU'g\| P_{\left(M_{U'g} h\right)}$), while the remaining Q eigenvalues are given by $\mu_r(U' P_g U) + \Delta_{\max}$, $r = 1, \ldots, Q$, and satisfy $\mu_{r+R-\min(Q,R)}(U' P_g U) + \Delta_{\max} \geq \mu_r(U' P_g U)$, for $r \in \{1, 2, \ldots, \min(R, Q)\}$ (by the definition of $\Delta_{\max}$). Therefore we have

$$
\sum_{r=2Q-\min(R,2Q)+1}^{2Q} \mu_r \left[\widetilde{h}' \left(U' P_g U + \|g'UU'g\| P_{\left(M_{U'g} h\right)} + \Delta_{\max} P_{(U'g)}\right) \widetilde{h}\right] \geq \sum_{r=1}^{\min(R,Q)} \mu_r(U' P_g U). \qquad \text{(S.72)}
$$

We can thus conclude that

$$
\begin{aligned}
&\sum_{r=1}^{\min(R,2Q)} \mu_r \left[\widetilde{h}' A_2(C))\widetilde{h}\right] \\
&\leq \sum_{r=1}^{\min(R,Q)} \mu_r \left[g' \left(U + gCh'\right) \left(U + gCh'\right)' g\right] - \sum_{r=1}^{\min(R,Q)} \mu_r(g'UU'g) \\
&\leq \sum_{r=1}^{\min(Q,R)} \mu_r \left(CC' + g'UhC' + Ch'U'g\right). \qquad \text{(S.73)}
\end{aligned}
$$

Combining (S.69), (S.70) and (S.73) gives the statement of the lemma. ■

Proof of Lemma S.7. Let h be a $T \times Q$ matrix whose span equals the span of A, i.e. $P_h A = A$, and that satisfies $h'h = \mathbb{1}_Q$, and let $\rho = \|A\|/T$. Then $A \leq T\rho P_h$, which implies $\sum_{r=1}^{R} \mu_r\left(e'e + A\right) \leq \sum_{r=1}^{R} \mu_r\left(e'e + T\rho P_h\right)$.

The distribution of e is invariant under orthogonal transformations $e \mapsto eO$, where O is an arbitrary orthogonal $T \times T$ matrix, i.e. $OO' = \mathbb{1}_T$. The distribution of the eigenvalues of $e'e + T\rho P_h$ therefore does not depend on h at all, but only on ρ and Σ. We can therefore choose h arbitrarily, even as a random matrix (but independent from e). Let u be a $Q \times T$ matrix that is independent of e, and whose columns u_t, $t = 1, \dots T$, are distributed as $u_t \sim iid\mathcal{N}(0, \rho\mathbb{1}_Q)$. We choose h such that the span of h equals the span of u', i.e. $uP_h = P_h$. Since we consider an asymptotic where Q is finite, while $T \to \infty$ it is easy to verify that $\|T\rho P_h - u'u\| = \mathcal{O}_P(\sqrt{T})$, which implies $\sum_{r=1}^{R} \mu_r\left(e'e + T\rho P_h\right) = \sum_{r=1}^{R} \mu_r\left(e'e + u'u\right) + \mathcal{O}_P(\sqrt{T})$.

Let $U = (e', u')'$ and $E = (e', 0_{T\times Q})'$, which are $(N+Q) \times T$ matrices. The non-zero eigenvalues of the $T \times T$ matrices $U'U = e'e + u'u$ and $E'E = e'e$ are equal to the non-zero eigenvalues of the $(N+Q) \times (N+Q)$ matrices UU' and EE', respectively. Let v be the $(N+Q) \times R$ matrix whose columns equal to the normalized eigenvectors that correspond to the R largest eigenvalues of UU'. We then have

$$\begin{aligned}
\sum_{r=1}^{R} \mu_r\left(e'e + u'u\right) &= \sum_{r=1}^{R} \mu_r\left(UU'\right) = \operatorname{Tr}\left(v'UU'v\right), \\
\sum_{r=1}^{R} \mu_r\left(e'e\right) &= \sum_{r=1}^{R} \mu_r\left(EE'\right) \geq \operatorname{Tr}\left(v'EE'v\right),
\end{aligned} \tag{S.74}$$

where the last inequality follows from the maximization property of the eigenvalues of EE'. Decompose $v = (v_1', v_2')'$ into the $N \times R$ matrix v_1 and the $Q \times R$ matrix v_2. We then have

$$\begin{aligned}
\sum_{r=1}^{R} \mu_r\left(e'e + u'u\right) - \sum_{r=1}^{R} \mu_r\left(e'e\right) &\leq \operatorname{Tr}\left(v'UU'v\right) - \operatorname{Tr}\left(v'EE'v\right) \\
&= \operatorname{Tr}\left[v' \begin{pmatrix} 0_{N\times N} & eu' \\ ue' & uu' \end{pmatrix} v\right] \\
&= 2\operatorname{Tr}\left(v_1'eu'v_2\right) + \operatorname{Tr}\left(v_2'uu'v_2\right) \\
&\leq 2R\|v_1'eu'v_2\| + R\|v_2'uu'v_2\| \\
&\leq 2R\|e\|\|u\|\|v_2\| + R\|u\|^2\|v_2\|^2,
\end{aligned} \tag{S.75}$$

where we used that for any square matrix B we have $\operatorname{Tr}(B) \leq \operatorname{rank}(B)\|B\|$, and also that $\|v_1\| \leq 1$. We have $\|e\| = \mathcal{O}_P(\sqrt{\max(N,T)}) = \mathcal{O}_P(\sqrt{N+T})$, $\|u\| = \mathcal{O}_P(\sqrt{T})$ and, as will be shown below, $\|v_2\| = \mathcal{O}_P(1/\sqrt{n})$. Therefore

$$\sum_{r=1}^{R} \mu_r\left(e'e + u'u\right) - \sum_{r=1}^{R} \mu_r\left(e'e\right) = \mathcal{O}_P(\sqrt{(N+T)T/n}). \tag{S.76}$$

Combining the above results we find

$$
\begin{aligned}
\sum_{r=1}^{R} \mu_r \left(e'e + A\right) &\leq \sum_{r=1}^{R} \mu_r \left(e'e + T\rho P_h\right) \\
&\leq \sum_{r=1}^{R} \mu_r \left(e'e + u'u\right) + \mathcal{O}_P(\sqrt{T}) \\
&\leq \sum_{r=1}^{R} \mu_r \left(e'e\right) + \mathcal{O}_P\left(\sqrt{(N+T)T/n}\right) + \mathcal{O}_P(\sqrt{T}) \\
&\leq \sum_{r=1}^{R} \mu_r \left(e'e\right) + \mathcal{O}_P\left(\sqrt{(N+T)T/n}\right), \qquad \text{(S.77)}
\end{aligned}
$$

where in the last step we used that $N/n \geq 1$. The last statement is what we wanted to show. However, we still have to justify that $\|v_2\| = \mathcal{O}_P(1/\sqrt{n})$. For this we first note that increasing the eigenvalues of Σ can only decrease $\|v_2\|$. Without loss of generality we can therefore consider the case where all the n eigenvalues of Σ that are smaller than ρ are increased to be exactly equal to ρ. In that case the distribution of U is symmetric under left-multiplication with orthogonal $O(n+Q)$ matrices, which only act on the the $(n+Q)$-dimensional eigenspace of the $(N+Q)\times(N+Q)$ covariance matrix of U corresponding to eigenvector ρ. Since the distribution of U has this symmetry, the same needs to be true for the distribution of the eigenvectors v of UU'. Since Q is finite, while $n \to \infty$ this implies that $\|v_2\| = \mathcal{O}_P(1/\sqrt{n})$. ■

Proof of Lemma S.8, Part 1. Here, we consider the case where Assumption DX-1 holds, and show that Lemma S.8 holds in that case.

We want to show that $C^{(1)} = \mathcal{O}_P(N^{1/4})$. By definition of $C^{(1)}$ and Assumption EX we have

$$
\begin{aligned}
C_k^{(1)} &= \frac{1}{\sqrt{NT}} \operatorname{Tr}(M_{\lambda^0}\, X_k\, M_{f^0}\, e') \\
&= \frac{1}{\sqrt{NT}} \operatorname{Tr}(X_k e') - \frac{1}{\sqrt{NT}} \operatorname{Tr}(P_{\lambda^0}\, X_k\, e') + \frac{1}{\sqrt{NT}} \operatorname{Tr}(P_{\lambda^0}\, X_k\, P_{f^0}\, e') \\
&= \mathcal{O}_P(1) - \frac{1}{\sqrt{NT}} \operatorname{Tr}(P_{\lambda^0}\, X_k\, e') + \frac{1}{\sqrt{NT}} \operatorname{Tr}(P_{\lambda^0}\, X_k\, P_{f^0}\, e'). \qquad \text{(S.78)}
\end{aligned}
$$

Since $\|\widetilde{X}_k\| = \mathcal{O}_P(N^{3/4})$ we have

$$
\left| \frac{1}{\sqrt{NT}} \operatorname{Tr}(P_{\lambda^0}\, \widetilde{X}_k\, e') \right| \leq \frac{R}{\sqrt{NT}} \|\widetilde{X}_k\| \|e\| = \mathcal{O}_P(N^{1/4}), \qquad \text{(S.79)}
$$

i.e. $\frac{1}{\sqrt{NT}} \operatorname{Tr}(P_{\lambda^0}\, \widetilde{X}_k\, e') = \mathcal{O}_P(N^{1/4})$. Analogously we obtain $\frac{1}{\sqrt{NT}} \operatorname{Tr}(P_{\lambda^0}\, \widetilde{X}_k\, P_{f^0}\, e') = \mathcal{O}_P(N^{1/4})$. Regarding the $\overline{X}_k$ contribution to $C_k^{(1)}$, consider $e = \Sigma^{1/2} u$, i.e. case (a) of Assumption DX-1 (the proof for case (b) is analogous). Using our assumptions on the distribution of e and $\overline{X}_k$ we have $\mathbb{E}\left[\operatorname{Tr}(P_{\lambda^0}\, \overline{X}_k e')^2 | X_k, \lambda^0, \Sigma\right] = \operatorname{Tr}(\overline{X}_k' P_{\lambda^0} \Sigma P_{\lambda^0} \overline{X}_k) \leq \operatorname{rank}(\overline{X}_k) \|\overline{X}_k\|^2 \|\Sigma\| = \mathcal{O}_P(NT)$, and therefore $\frac{1}{\sqrt{NT}} \mathrm{Tr}(P_{\lambda^0}\, \overline{X}_k\, e') = \mathcal{O}_P(1)$. Analogously we find $\frac{1}{\sqrt{NT}} \mathrm{Tr}(P_{\lambda^0}\, \overline{X}_k\, P_{f^0}\, e') = \mathcal{O}_P(1)$. Combining the above results gives $C^{(1)} = \mathcal{O}_P(N^{1/4})$.

We want to show that Assumption SN holds. We have $\|X_k\| \leq \|\overline{X}_k\| + \|\widetilde{X}_k\| = \mathcal{O}_P(\sqrt{NT}) + \mathcal{O}_P(N^{3/4}) = \mathcal{O}_P(\sqrt{NT})$, i.e. Assumption SN($i$) is satisfied. In the following we assume that $e = \Sigma^{1/2} u$, i.e. case (a) of Assumption DX-1. The proof for case (b) follows by symmetry of the problem ($N \leftrightarrow T$). We have $\|e\| = \|\Sigma\|^{1/2}\|u\| = \mathcal{O}_P(1)\|u\|$, since we assume that $\|\Sigma\| = \mathcal{O}_P(1)$. Thus, we are left to show $\|u\| = \mathcal{O}_P(\sqrt{\max(N,T)})$. Lemma S.8 assumes $N/T \to \kappa^2$, but it turns out that this assumption is not necessary to show $\|u\| = \mathcal{O}_P(\sqrt{\max(N,T)})$, i.e. for the moment consider an arbitrary limit $N, T \to \infty$. By assumption, the errors u_{it} are iid $\mathcal{N}(0,1)$. Since an arbitrary limit $N, T \to \infty$ is not considered very often in Random Matrix Theory, we define the $\max(N,T) \times \max(N,T)$ matrix $u^{\rm big}$, which contains u as a submatrix, and whose remaining elements are also iid $\mathcal{N}(0,1)$ and independent of u. We then have $\|u\| \leq \|u^{\rm big}\| = \mathcal{O}_P(\sqrt{\max(N,T)})$, where the last step is due to Geman (1980).

Finally, we show that Assumption HL1 holds. Consider case (a) of Assumption DX-1(ii) in the following. Using the decomposition $X_k = \overline{X}_k + \widetilde{X}_k$ we have

$$
\begin{aligned}
& \sum_{r=1}^{R-R^0} \left\{ \mu_r \left[M_{f^0}\, (e - \Delta\beta \cdot X)'\, M_{\lambda^0}\, (e - \Delta\beta \cdot X)\, M_{f^0} \right] - \mu_r \left[M_{f^0}\, (\Delta\beta \cdot X)'\, M_{\lambda^0}\, (\Delta\beta \cdot X)\, M_{f^0} \right] \right\} \\
&= \sum_{r=1}^{R-R^0} \left\{ \mu_r \left[M_{f^0} \left(e - \Delta\beta \cdot \overline{X} \right)' M_{\lambda^0} \left(e - \Delta\beta \cdot \overline{X} \right) M_{f^0} \right] - \mu_r \left[M_{f^0} \left(\Delta\beta \cdot \overline{X} \right)' M_{\lambda^0} \left(\Delta\beta \cdot \overline{X} \right) M_{f^0} \right] \right\} \\
&\qquad\qquad + \mathcal{O}_P(\|e\| \|\widetilde{X}_k\| \|\Delta\beta\|) + \mathcal{O}_P(\|\widetilde{X}_k\| \|X_k\| \|\Delta\beta\|^2) \\
&= \sum_{r=1}^{R-R^0} \left\{ \mu_r \left[M_{f^0} \left(e - \Delta\beta \cdot \overline{X} \right)' M_{\lambda^0} \left(e - \Delta\beta \cdot \overline{X} \right) M_{f^0} \right] - \mu_r \left[M_{f^0} \left(\Delta\beta \cdot \overline{X} \right)' M_{\lambda^0} \left(\Delta\beta \cdot \overline{X} \right) M_{f^0} \right] \right\} \\
&\qquad\qquad + \mathcal{O}_P(N^{5/4} \|\Delta\beta\|) + \mathcal{O}_P(N^{7/4} \|\Delta\beta\|^2). \qquad \text{(S.80)}
\end{aligned}
$$

We now apply Lemma S.6 with $U = M_{\lambda^0} e M_{f^0}$ and $gCh' = -M_{\lambda^0}(\Delta\beta \cdot \overline{X}) M_{f^0}$, where g and h

are define in Assumption DX-1, and $C = g'(\Delta\beta \cdot \overline{X})h$. We obtain

$$
\begin{aligned}
&\sum_{r=1}^{R-R^0} \Big\{ \mu_r \left[M_{f^0} \left(e - \Delta\beta \cdot \overline{X} \right)' M_{\lambda^0} \left(e - \Delta\beta \cdot \overline{X} \right) M_{f^0} \right] \\
&\quad \le \sum_{r=1}^{R-R^0} \mu_r \left(M_{f^0} e' M_{\lambda^0} e M_{f^0} + \| g' e M_{f^0} e' g \| \, P_{\left(M_{[M_{f^0} e' g]} h \right)} + \Delta_{\max} P_{(M_{f^0} e' g)} \right) \\
&\qquad\quad + \sum_{r=1}^{\min(Q, R-R^0)} \mu_r \left[M_{f^0} \left(\Delta\beta \cdot \overline{X} \right)' M_{\lambda^0} \left(\Delta\beta \cdot \overline{X} \right) M_{f^0} \right] + \mathcal{O}_P \left(\| g' e h \| \| \overline{X} \| \| \Delta\beta \| \right) \\
&\quad \le \sum_{r=1}^{R-R^0} \mu_r \left(M_{f^0} e' M_{\lambda^0} e M_{f^0} + \| g' e M_{f^0} e' g \| \, P_h \right) \\
&\qquad\quad + \sum_{r=1}^{R-R^0} \mu_r \left[M_{f^0} \left(\Delta\beta \cdot \overline{X} \right)' M_{\lambda^0} \left(\Delta\beta \cdot \overline{X} \right) M_{f^0} \right] + \mathcal{O}_P \left(\sqrt{NT} \| \Delta\beta \| \right) + \mathcal{O}_P(\sqrt{N}) \\
&\quad \le \sum_{r=1}^{R-R^0} \mu_r \left(M_{f^0} e' M_{\lambda^0} e M_{f^0} + T \| g' \Sigma g \| \, P_h \right) \\
&\qquad\quad + \sum_{r=1}^{R-R^0} \mu_r \left[M_{f^0} \left(\Delta\beta \cdot \overline{X} \right)' M_{\lambda^0} \left(\Delta\beta \cdot \overline{X} \right) M_{f^0} \right] + \mathcal{O}_P \left(\sqrt{NT} \| \Delta\beta \| \right) + \mathcal{O}_P(\sqrt{N}),
\end{aligned}
\tag{S.81}
$$

where we used that under our assumptions we have

(i) $\| g' e h \| = \mathcal{O}_P(1)$,

(ii) $g' e M_{f^0} e' g = T g' \Sigma g + \mathcal{O}_P(\sqrt{N})$,

(iii) $\Delta_{\max} \equiv \max_{r \in \{1,2,\ldots,\min(R,Q)\}} \left[\mu_r(g' e M_{f^0} e' g) - \mu_{r+Q-\min(Q,R)}(g' e M_{f^0} e' g) \right] = \mathcal{O}_P(\sqrt{N})$,

(iv) $\left\| P_{\left(M_{[M_{f^0} e' g]} h \right)} - P_h \right\| = \mathcal{O}_P(N^{-1/2})$.

Statement (i) above holds, because $g' e h = g' \Sigma^{1/2} u h$ is a projection of u to a $Q \times Q$ submatrix, with $g' \Sigma^{1/2}$ and h independent of u, and $\| g' \Sigma^{1/2} \| = \mathcal{O}_P(1)$ and $\| h \| = 1$.

Statement (ii) holds, because we can calculate the expectation and variance of $g' e M_{f^0} e' g = g' \Sigma^{1/2} u u' \Sigma^{1/2} g$ conditional on $\Sigma^{1/2} g$ to show that $g' \Sigma^{1/2} u u' \Sigma^{1/2} g = g' \Sigma^{1/2} \mathbb{E}(u u') \Sigma^{1/2} g + \mathcal{O}_P(\sqrt{N})$, with $\mathbb{E}(u u') = T \mathbb{1}_N$.

Statement (iii) holds, because assume that either $R \ge Q$, in which case $\Delta_{\max} = 0$, or we assume $g' \Sigma g = \| g' \Sigma g \| \mathbb{1}_Q + \mathcal{O}_P(N^{-1/2})$, so that $g' e M_{f^0} e' g = T \| g' \Sigma g \| \mathbb{1}_Q + \mathcal{O}_P(\sqrt{N})$, where $T \| g' \Sigma g \| \mathbb{1}_Q$ gives no contribution to $\Delta_{\max}$.

We now apply Lemma S.7 with "e" in the Lemma equal to $M_{\lambda^0} e M_{f^0}$, "Σ" in the Lemma equal to $M_{\lambda^0} \Sigma M_{\lambda^0}$, and $A = T \| g' \Sigma g \| \, P_h$. We have $\mu_{n-R^0}(M_{\lambda^0} \Sigma M_{\lambda^0}) \ge \mu_n(\Sigma)$ and $g' \Sigma g =$

$g'M_{\lambda^0}\Sigma M_{\lambda^0}g$. We therefore choose "$n$" in the Lemma equal to $n-R^0$ when applying Lemma S.7, and our assumption $\mu_n(\Sigma) \geq \|g'\Sigma g\|$ with $1/n = \mathcal{O}_P(1/N)$ is now used. When employing Lemma S.7 here we also use that rotational invariance of $e'M_{\lambda^0}e = u'\Sigma^{1/2}M_{\lambda^0}\Sigma^{1/2}u$ allows us to treat $M_{f^0}e'M_{\lambda^0}eM_{f^0}$ as an $(N-R^0)\times(N-R^0)$ matrix, which requires that u is *iid* normally distributed. By Lemma S.7 we then have

$$\begin{aligned}
&\sum_{r=1}^{R-R^0} \mu_r \left(M_{f^0}e'M_{\lambda^0}eM_{f^0} + T\|g'\Sigma g\|\, P_h\right) \\
&= \sum_{r=1}^{R-R^0} \mu_r \left(M_{f^0}e'M_{\lambda^0}eM_{f^0}\right) + \mathcal{O}_P\left(\sqrt{(N+T-2R^0)(T-R^0)/(n-R^0)}\right) \\
&= \sum_{r=1}^{R-R^0} \mu_r \left(M_{f^0}e'M_{\lambda^0}eM_{f^0}\right) + \mathcal{O}_P(\sqrt{N}). \qquad \text{(S.82)}
\end{aligned}$$

Combining this with (S.80) and (S.81) gives Assumption HL1. ■

Proof of Lemma S.8, Part 2. Here, we consider the case where Assumption DX-2 holds, and show that Lemma S.8 holds in that case.

Using the assumption $M_{\lambda^0}\overline{X}_k M_{f^0} = 0$ simplifies the calculation in (S.80), namely

$$\begin{aligned}
&\sum_{r=1}^{R-R^0} \mu_r \left[M_{f^0}\left(e-\Delta\beta\cdot X\right)' M_{\lambda^0}\left(e-\Delta\beta\cdot X\right) M_{f^0}\right] \\
&= \sum_{r=1}^{R-R^0} \mu_r \left[M_{f^0}\left(e-\Delta\beta\cdot \overline{X}\right)' M_{\lambda^0}\left(e-\Delta\beta\cdot \overline{X}\right) M_{f^0}\right] + \mathcal{O}_P(\|e\|\|\widetilde{X}_k\|\|\Delta\beta\|) + \mathcal{O}_P(\|\widetilde{X}_k\|^2\|\Delta\beta\|^2) \\
&= \sum_{r=1}^{R-R^0} \mu_r \left[M_{f^0}e'M_{\lambda^0}eM_{f^0}\right] + \mathcal{O}_P(N^{5/4}\|\Delta\beta\|) + \mathcal{O}_P(N^{3/2}\|\Delta\beta\|^2), \qquad \text{(S.83)}
\end{aligned}$$

and analogously we obtain $\sum_{r=1}^{R-R^0} \mu_r \left[M_{f^0}\left(\Delta\beta\cdot X\right)' M_{\lambda^0}\left(\Delta\beta\cdot X\right) M_{f^0}\right] = \mathcal{O}_P(N^{3/2}\|\Delta\beta\|^2)$. We therefore have

$$d(\beta) = \mathcal{O}_P(N^{5/4}\|\Delta\beta\|) + \mathcal{O}_P(N^{3/2}\|\Delta\beta\|^2), \qquad \text{(S.84)}$$

which implies that Assumption HL1 holds. The result for $C^{(1)}$ follows because with $M_{\lambda^0}\overline{X}_k M_{f^0} = 0$ we find

$$\begin{aligned}
C_k^{(1)} &= \frac{1}{\sqrt{NT}} \operatorname{Tr}(M_{\lambda^0}\, X_k\, M_{f^0}\, e') \\
&= \frac{1}{\sqrt{NT}} \operatorname{Tr}(X_k e') + \mathcal{O}_P(\|e\|\|\widetilde{X}_k\|/\sqrt{NT}) \\
&= \mathcal{O}_P(1) + \mathcal{O}_P(N^{1/4}). \qquad \text{(S.85)}
\end{aligned}$$

Finally, Assumption SN holds obviously under Assumption DX-2. ■

S.6.3 Proofs for Details on Asymptotic Equivalence

Proof of Lemma S.9. Applying the expansion of $\widehat{e}(\beta)$ in Lemma S.3 together with $\|M_{\lambda^0} e M_{f^0}\| = \mathcal{O}_P(\sqrt{N})$, $\|\widehat{e}^{(1)}_e\| = \mathcal{O}_P(1)$, $\|\widehat{e}^{(2)}_e\| = \mathcal{O}_P(N^{-1/2})$, $\|\widehat{e}^{(1)}_k\| = \mathcal{O}_P(N)$ $\|\widehat{e}^{(2)}_k\| = \mathcal{O}_P(\sqrt{N})$ and the bound on $\|\widehat{e}^{(\text{rem})}\|$ given in the Lemma we obtain

$$\widehat{e}'(\beta)\widehat{e}(\beta) = B(\beta) + B'(\beta) + T^{(\text{rem})}(\beta) \ , \tag{S.86}$$

where the terms $B^{(\text{rem},1)}(\beta)$ and $B^{(\text{rem},2)}$ in $B(\beta)$ are given by

$$\begin{aligned} B^{(\text{rem},1)}(\beta) &= M_{f^0}[(\beta - \beta^0 \cdot X)]' M_{\lambda^0} e M_{f^0} e' \lambda^0 (\lambda^{0\prime}\lambda^0)^{-1} (f^{0\prime} f^0)^{-1} f^{0\prime} \\ &\quad + M_{f^0} e' M_{\lambda^0} [(\beta - \beta^0 \cdot X)] M_{f^0} e' \lambda^0 (\lambda^{0\prime}\lambda^0)^{-1} (f^{0\prime} f^0)^{-1} f^{0\prime} \\ &\quad + M_{f^0} e' M_{\lambda^0} e M_{f^0} [(\beta - \beta^0 \cdot X)]' \lambda^0 (\lambda^{0\prime}\lambda^0)^{-1} (f^{0\prime} f^0)^{-1} f^{0\prime} \\ &\quad + M_{f^0} \left(M_{f^0} e' M_{\lambda^0} \widehat{e}^{(2)}_e + \widehat{e}^{(1)\prime}_e \widehat{e}^{(2)}_e + \widehat{e}^{(2)\prime}_e M_{\lambda^0} e' M_{f^0} \right) P_{f^0} \ , \\ B^{(\text{rem},2)} &= \tfrac{1}{2} P_{f^0} \left(M_{f^0} e' M_{\lambda^0} \widehat{e}^{(2)}_e + \widehat{e}^{(1)\prime}_e \widehat{e}^{(2)}_e + \widehat{e}^{(2)\prime}_e M_{\lambda^0} e' M_{f^0} \right) P_{f^0} \\ &= f^0 (f^{0\prime} f^0)^{-1} (\lambda^{0\prime}\lambda^0)^{-1} \lambda^{0\prime} e M_{f^0} e' M_{\lambda^0} e M_{f^0} e' \lambda^0 (\lambda^{0\prime}\lambda^0)^{-1} (f^{0\prime} f^0)^{-1} f^{0\prime}, \end{aligned} \tag{S.87}$$

and for $\sqrt{N}\|\beta - \beta^0\| \leq c$ (which implies $\|\widehat{e}(\beta)\| = \mathcal{O}_P(\sqrt{N})$) we have

$$\|T^{(\text{rem})}(\beta)\| = \mathcal{O}_P(N^{-1/2}) + \|\beta - \beta^0\| \mathcal{O}_P(N^{1/2}) + \|\beta - \beta^0\|^2 \mathcal{O}_P(N^{3/2}) \ . \tag{S.88}$$

which holds uniformly over β. Note also that

$$B^{(eeee)} + B^{(eeee)\prime} = M_{f^0} \left(M_{f^0} e' M_{\lambda^0} \widehat{e}^{(2)}_e + \widehat{e}^{(1)\prime}_e \widehat{e}^{(2)}_e + \widehat{e}^{(2)\prime}_e M_{\lambda^0} e' M_{f^0} \right) M_{f^0}. \tag{S.89}$$

Thus, we have $\|B^{(\text{rem},2)}\| = \mathcal{O}_P(1)$, and for $\sqrt{N}\|\beta - \beta^0\| \leq c$ we have $\|B^{(\text{rem},1)}(\beta)\| = \mathcal{O}_P(1) + \|\beta - \beta^0\| \mathcal{O}_P(N)$, and by Weyl's inequality

$$\mu_r \left[\widehat{e}'(\beta)\widehat{e}(\beta) \right] = \mu_r \left[B(\beta) + B'(\beta) \right] + o_P \left[\left(1 + \|\beta - \beta^0\| \right)^2 \right] \ , \tag{S.90}$$

again uniformly over β. This proves the lemma. ■

Proof of Corollary S.10. From Theorem S.5 we know that $N^{3/4}\|\widehat{\beta}_R - \beta^0\| = \mathcal{O}_P(1)$, so that the bound in Assumption HL2 becomes applicable. Let $\gamma \equiv W^{-1}\left(C^{(1)} + C^{(2)}\right)/\sqrt{NT} = \mathcal{O}_P(1/\sqrt{NT})$, as in the proof proof of Corollary 4.3. Since $\widehat{\beta}_R$ minimizes $\mathcal{L}^R_{NT}(\beta)$ it must in particular satisfy $\mathcal{L}^R_{NT}(\widehat{\beta}_R) \leq \mathcal{L}^R_{NT}(\beta^0 + \gamma)$. Using Lemma S.9 and Assumption HL2 it follows that

$$\mathcal{L}^0_{NT}(\widehat{\beta}_R) \leq \mathcal{L}^0_{NT}\left(\beta^0 + \gamma\right) + \frac{1}{NT}\, o_P \left[\left(1 + \sqrt{NT}\|\widehat{\beta}_R - \beta^0\|^2 \right)^2 \right]. \tag{S.91}$$

The rest of the proof is analogous to the proof of Corrollary 4.3. ■

Proof of Lemma S.11. For the eigenvalues of $A + B$ we have

$$\mu_r(A+B) = \min_{\Gamma} \max_{\{\gamma \colon \|\gamma\|=1,\, P_\Gamma \gamma = 0\}} \gamma'(A+B)\gamma \,, \tag{S.92}$$

where Γ is a $n \times (r-1)$ matrix with full rank $r-1$, and γ is a $n \times 1$ vector. In the following we only consider those γ that lie in the span of the first r eigenvectors A, i.e. $\gamma = \sum_{i=1}^r c_i \nu_i$. The condition $\|\gamma\| = 1$ implies $\sum_{i=1}^r c_i^2 = 1$. The column space of Γ is $(r-1)$-dimensional. Therefore, for a given $\gamma = \sum_{i=1}^r c_i \nu_i$ there always exists a Γ such that the conditions $\|\gamma\| = 1$ and $P_\Gamma \gamma = 0$ *uniquely* determine γ up to the sign. We thus have

$$\begin{aligned}
\mu_r(A+B) &\geq \min_{\Gamma} \max_{\left\{\gamma \colon \gamma = \sum_{i=1}^r c_i \nu_i,\, \|\gamma\|=1,\, P_\Gamma \gamma = 0\right\}} \gamma'(A+B)\gamma \\
&= \min_{\left\{\gamma \colon \gamma = \sum_{i=1}^r c_i \nu_i,\, \|\gamma\|=1\right\}} \gamma'(A+B)\gamma \\
&\geq \min_{\left\{(c_1,\ldots,c_r) \colon \sum_{i=1}^r c_i^2 = 1\right\}} \left[\sum_{i=1}^r c_i^2 \, \mu_i(A) - b \left(\sum_{i=1}^r |c_i| \right)^2 \right] \\
&\geq \mu_r(A) - r\, b \\
&\geq \mu_r(A) - \frac{(q-1)\, b}{1 - \sum_{i=q}^n \frac{b}{\mu_r(A) - \mu_i(A)}} \,,
\end{aligned} \tag{S.93}$$

where we used that $q - 1 \geq r$ and that the additional fraction we multiplied with is larger than one. This is the lower bound for $\mu_r(A+B)$ that we wanted to show. We now want to derive the upper bound. Let $\widetilde{A}$, $\widetilde{B}$ and $\bar{B}$ be $(n-r+1) \times (n-r+1)$ matrices defined by $\widetilde{A}_{ij} = \nu'_{i+r-1} A \nu_{j+r-1}$, $\widetilde{B}_{ij} = \nu'_{i+r-1} B \nu_{j+r-1}$, and $\bar{B}_{ij} = b$, where $i, j = 1, \ldots, n-r+1$. We can choose $\Gamma = (\nu_1, \nu_2, \ldots, \nu_{r-1})$ in the above minimization problem, in which case γ is restricted to the span of ν_r, $\nu_{r+1}, \ldots, \nu_n$. Therefore

$$\begin{aligned}
\mu_r(A+B) &\leq \max_{\{\widetilde{\gamma} \colon \|\widetilde{\gamma}\|=1\}} \widetilde{\gamma}'(\widetilde{A} + \widetilde{B})\widetilde{\gamma} \\
&= \mu_1(\widetilde{A} + \widetilde{B}) \,,
\end{aligned} \tag{S.94}$$

where $\widetilde{\gamma}$ is a $(n-r+1)$-dimensional vector, whose components are denoted $\widetilde{\gamma}_i$, $i = 1, \ldots, n-r+1$, in the following. Note that $\widetilde{A}$ is a diagonal matrix with entries $\mu_{i+r-1}(A)$, $i = 1, \ldots, n-r+1$.

Therefore

$$\mu_r(A+B) \le \max_{\{\widetilde{\gamma}:\, \|\widetilde{\gamma}\|=1\}} \left[\sum_{i=1}^{n+r-1} (\widetilde{\gamma}_i)^2 \mu_{i+r-1}(A) + \sum_{i,j=1}^{n+r-1} \widetilde{\gamma}_i\, \widetilde{\gamma}_j\, \widetilde{B}_{ij} \right]$$
$$\le \max_{\{\widetilde{\gamma}:\, \|\widetilde{\gamma}\|=1\}} \left[\sum_{i=1}^{n+r-1} (\widetilde{\gamma}_i)^2 \mu_{i+r-1}(A) + b \sum_{i,j=1}^{n+r-1} |\widetilde{\gamma}_i|\, |\widetilde{\gamma}_j| \right]$$
$$= \max_{\{\widetilde{\gamma}:\, \|\widetilde{\gamma}\|=1\}} \left[\sum_{i=1}^{n+r-1} (\widetilde{\gamma}_i)^2 \mu_{i+r-1}(A) + \sum_{i,j=1}^{n+r-1} \widetilde{\gamma}_i\, \widetilde{\gamma}_j\, \bar{B}_{ij} \right]$$
$$= \mu_1(\widetilde{A} + \bar{B}) \ . \tag{S.95}$$

In the last maximization problem the maximum is always attained at a point with $\widetilde{\gamma}_i \ge 0$, which is why we could omit the absolute values around $\widetilde{\gamma}_i$.

The eigenvalue $\widetilde{\mu} \equiv \mu_1(\widetilde{A} + \bar{B})$ is a solution of the characteristic polynomial of $\widetilde{A} + \bar{B}$ which can be written as

$$1 = \sum_{i=r}^{n} \frac{b}{\widetilde{\mu} - \mu_i(A)} \ , \tag{S.96}$$

where $\mu_i(A) = \mu_{i-r+1}(\widetilde{A})$ are the eigenvalues of $\widetilde{A}$. In addition we have $\widetilde{\mu} = \mu_1(\widetilde{A} + \bar{B}) > \mu_1(\widetilde{A}) = \mu_r(A)$, because $\bar{B}$ is positive semi-definite (which gives $\ge$) and the eigenvectors of $\widetilde{A}$ do not agree with those of $\bar{B}$ (which gives $\ne$). From the characteristic polynomial we therefore find

$$1 = \sum_{i=r}^{q-1} \frac{b}{\widetilde{\mu} - \mu_i(A)} + \sum_{i=q}^{n} \frac{b}{\widetilde{\mu} - \mu_i(A)}$$
$$\le \frac{b(q-1)}{\widetilde{\mu} - \mu_r(A)} + \sum_{i=q}^{n} \frac{b}{\mu_r(A) - \mu_i(A)}. \tag{S.97}$$

Since we assume $1 \ge \sum_{i=q}^{n} \frac{b}{\mu_r(A) - \mu_i(A)}$, this gives an upper bound on $\widetilde{\mu}$, and since $\mu_r(A+B) \le \widetilde{\mu}$ the same bound holds for $\mu_r(A+B)$, namely

$$\mu_r(A+B) \le \mu_r(A) + \frac{(q-1)\, b}{1 - \sum_{i=q}^{n} \frac{b}{\mu_r(A) - \mu_i(A)}} \ . \tag{S.98}$$

This is what we wanted to show. ■

Proof of Lemma S.12. Consider the case $T \le N$, so that for $Q = \min(N,T) - R^0$ defined in Assumption EV we have $Q = T - R^0$. If $N < T$, then we interchange the role of N and T in the following proof.[8]

[8] We consider limits $N, T \to \infty$ with $N/T \to \kappa^2$. For $\kappa^2 > 1$ we have $T < N$ holding asymptotically, while for $\kappa^2 < 1$ we have $T > N$ holding asymptotically and the role of N and T in the proof needs to be interchanged.

Define

$$
\begin{aligned}
C^{\pm}(\beta) = B(\beta) + B'(\beta) &\pm \left(\sqrt{\frac{4}{aN}} M_{f^0} B^{(\mathrm{rem},1)}(\beta) P_{f^0} \mp \sqrt{\frac{aN}{4}} P_{f^0} \right) \\
&\qquad\qquad \times \left(\sqrt{\frac{4}{aN}} M_{f^0} B^{(\mathrm{rem},1)}(\beta) P_{f^0} \mp \sqrt{\frac{aN}{4}} P_{f^0} \right)' \\
&\pm \left(\sqrt{\frac{4}{aN}} M_{f^0} e' M_{\lambda^0} e M_{f^0} e' \lambda^0 (\lambda^{0\prime}\lambda^0)^{-1} (f^{0\prime} f^0)^{-1} f^{0\prime} \pm \sqrt{\frac{aN}{4}} P_{f^0} \right) \\
&\quad \times \left(\sqrt{\frac{4}{aN}} M_{f^0} e' M_{\lambda^0} e M_{f^0} e' \lambda^0 (\lambda^{0\prime}\lambda^0)^{-1} (f^{0\prime} f^0)^{-1} f^{0\prime} \pm \sqrt{\frac{aN}{4}} P_{f^0} \right)' . \qquad \text{(S.99)}
\end{aligned}
$$

Since $C^{+}(\beta)$ [respectively $C^{-}(\beta)$] is obtained by adding [respectively subtracting] a positive definite matrix from $B(\beta) + B'(\beta)$, we have

$$
\mu_r \left(C^{-}(\beta) \right) \leq \mu_r \left(B(\beta) + B'(\beta) \right) \leq \mu_r \left(C^{+}(\beta) \right) . \qquad \text{(S.100)}
$$

The advantage of considering $C^{\pm}(\beta)$ instead of $B(\beta) + B'(\beta)$ directly is that there are no "mixed terms" in $C^{\pm}(\beta)$, which start with M_{f^0} and end with P_{f^0}, or vice versa, i.e. we can write $C^{\pm}(\beta) = C_1^{\pm}(\beta) + C_2^{\pm}$, where $C_1^{\pm}(\beta) = M_{f^0} C_1^{\pm}(\beta) M_{f^0}$ and $C_2^{\pm} = P_{f^0} C_2^{\pm} P_{f^0}$. Concretely, we have

$$
\begin{aligned}
C_1^{\pm}(\beta) = A(\beta) &\pm \frac{4}{aN} M_{f^0} B^{(\mathrm{rem},1)}(\beta) P_{f^0} B^{(\mathrm{rem},1)\prime}(\beta) M_{f^0} \\
&\pm \frac{4}{aN} M_{f^0} e' M_{\lambda^0} e M_{f^0} e' \lambda^0 (\lambda^{0\prime}\lambda^0)^{-1} (f^{0\prime} f^0)^{-1} (\lambda^{0\prime}\lambda^0)^{-1} \lambda^{0\prime} e M_{f^0} e' M_{\lambda^0} e M_{f^0} \\
&+ M_{f^0} \left[(\beta - \beta^0) \cdot X - e \right]' M_{\lambda^0} e f^0 (f^{0\prime} f^0)^{-1} (\lambda^{0\prime}\lambda^0)^{-1} \lambda^{0\prime} e M_{f^0} \\
&+ M_{f^0} e' M_{\lambda^0} \left[(\beta - \beta^0) \cdot X \right] f^0 (f^{0\prime} f^0)^{-1} (\lambda^{0\prime}\lambda^0)^{-1} \lambda^{0\prime} e M_{f^0} \\
&+ M_{f^0} e' M_{\lambda^0} e f^0 (f^{0\prime} f^0)^{-1} (\lambda^{0\prime}\lambda^0)^{-1} \lambda^{0\prime} \left[(\beta - \beta^0) \cdot X \right] M_{f^0} \\
&+ \text{the last three lines transposed} + B^{(eeee)} + B^{(eeee)\prime}, \\
C_2^{\pm} = P_{f^0} B^{(\mathrm{rem},2)} P_{f^0} &+ P_{f^0} B^{(\mathrm{rem},2)\prime} P_{f^0} \pm \frac{aN}{2} P_{f^0}. \qquad \text{(S.101)}
\end{aligned}
$$

In the rest of the proof we always assume that $N^{3/4} \left\| \beta - \beta^0 \right\| \leq c$. We apply Lemma S.11 to $C_1^{\pm}(\beta)$, with the A in the lemma equal to the leading term $M_{f^0} e' M_{\lambda^0} e M_{f^0}$, the B in the lemma equal to the remainder of $C_1^{\pm}(\beta)$, and $q = q_{NT}$. Assumption EV introduces ρ_r and w_r as the eigenvalues and corresponding eigenvectors of $M_{f^0} e' M_{\lambda^0} e M_{f^0}$, where $r = 1, \ldots, Q$ with

For $\kappa^2 = 1$ there is a subtlety, because neither $T \leq N$ nor $N \leq T$ needs to hold asymptotically (the ordering of N and T can change arbitrarily often while N and T grow). We could rule out this subtlety by only considering asymptotic sequences that satisfy either always $T \leq N$ or always $N \leq T$, which would not diminish the practical implications of our results in any way. The proof can also be adjusted to jointly consider the cases $T \leq N$ and $N \leq T$ in the asymptotic, which is not complicated, but cumbersome.

$Q = \min(N,T) - R^0 = T - R^0$. If we can show that

$$\sum_{r=q_{NT}}^{T-R^0} \frac{b_{NT}}{\rho_{R-R^0} - \rho_r} = o_P(1), \tag{S.102}$$

then Lemma S.11 becomes applicable asymptotically, and for $r = 1, \ldots, R - R^0$ we have wpa1

$$\left|\mu_r\left(C_1^{\pm}(\beta)\right) - \rho_r\right| \le \frac{(q_{NT}-1)\, b_{NT}}{1 - \sum_{s=q_{NT}}^{T-R^0} \frac{b_{NT}}{\rho_r - \rho_s}} \le \frac{q_{NT}\, b_{NT}}{1 - \sum_{s=q_{NT}}^{T-R^0} \frac{b_{NT}}{\rho_{R-R^0} - \rho_s}}\,, \tag{S.103}$$

where

$$b_{NT} = \max_{r,s} \left| w_r' \left(C_1^{\pm}(\beta) - M_{f^0} e' M_{\lambda^0} e M_{f^0}\right) w_s \right|. \tag{S.104}$$

We now check how the different terms in $C_1^{\pm}(\beta) - M_{f^0} e' M_{\lambda^0} e M_{f^0}$ contribute to b_{NT}. Using the definition of d_{NT} in equation (S.46) we have

$$\begin{aligned}
\max_{r,s} \left| w_r' M_{f^0} e' M_{\lambda^0} [(\beta - \beta^0) \cdot X] M_{f^0} w_s \right| &\le K \|e\| \|\beta - \beta^0\| \max_{k,r,s} \|v_r' X_k w_s\| \\
&\le d_{NT} \mathcal{O}_P(N^{-1/4}), \\
\max_{r,s} \left| w_r' M_{f^0} [(\beta - \beta^0) \cdot X]' M_{\lambda^0} [(\beta - \beta^0) \cdot X] M_{f^0} w_s \right| &\le K^2 \left\|\beta - \beta^0\right\|^2 \max_{k,r} \|M_{\lambda^0} X_k w_r\|^2 \\
&\le K^2 N \left\|\beta - \beta^0\right\|^2 \max_{k,r,s} \|v_r' X_k w_s\|^2 \\
&\le d_{NT}^2 \mathcal{O}_P(N^{-1/2}), \\
\max_{r,s} \left| w_r' \frac{4}{aN} M_{f^0} B^{(\text{rem},1)}(\beta) P_{f^0} B^{(\text{rem},1)\prime}(\beta) M_{f^0} w_s \right| &\le \frac{4}{aN} \|B^{(\text{rem},1)}(\beta)\|^2 = \mathcal{O}_P(N^{-1/2}),
\end{aligned}$$

$$
\begin{aligned}
&\max_{r,s} \left| w_r' \frac{4}{aN} M_{f^0} e' M_{\lambda^0} e M_{f^0} e' \lambda^0 (\lambda^{0\prime}\lambda^0)^{-1} (f^{0\prime} f^0)^{-1} (\lambda^{0\prime}\lambda^0)^{-1} \lambda^{0\prime} e M_{f^0} e' M_{\lambda^0} e M_{f^0} w_s \right| \\
&\qquad \leq \frac{4}{aN} \|e\|^4 \left\| \lambda^0 (\lambda^{0\prime}\lambda^0)^{-1} (f^{0\prime} f^0)^{-1} (\lambda^{0\prime}\lambda^0)^{-1} \lambda^{0\prime} \right\| \max_r \|w_r' e' P_{\lambda^0}\|^2 \leq d_{NT} \mathcal{O}_P(N^{-1}), \\
&\max_{r,s} \left| w_r' M_{f^0} e' M_{\lambda^0} e f^0 (f^{0\prime} f^0)^{-1} (\lambda^{0\prime}\lambda^0)^{-1} \lambda^{0\prime} e M_{f^0} w_s \right| \\
&\qquad \leq \|e\| \left\| f^0 (f^{0\prime} f^0)^{-1} (\lambda^{0\prime}\lambda^0)^{-1} \lambda^{0\prime} \right\| \max_s \|v_s' e P_{f^0}\| \max_r \|w_r' e' P_{\lambda^0}\| \leq d_{NT}^2 \mathcal{O}_P(N^{-1/2}), \\
&\max_{r,s} \left| w_r' M_{f^0} \left[(\beta - \beta^0) \cdot X \right]' M_{\lambda^0} e f^0 (f^{0\prime} f^0)^{-1} (\lambda^{0\prime}\lambda^0)^{-1} \lambda^{0\prime} e M_{f^0} w_s \right| \\
&\qquad = \max_{r,s} \left| w_r' \left[(\beta - \beta^0) \cdot X \right]' \left(\sum_q v_q' v_q \right) e f^0 (f^{0\prime} f^0)^{-1} (\lambda^{0\prime}\lambda^0)^{-1} \lambda^{0\prime} e w_s \right| \\
&\qquad \leq K \|\beta - \beta^0\| N \max_{r,s,k} |v_s' X_k w_r| \max_r \|v_r' e P_{f^0}\| \max_r \|w_r' e' P_{\lambda^0}\| \left\| f^0 (f^{0\prime} f^0)^{-1} (\lambda^{0\prime}\lambda^0)^{-1} \lambda^{0\prime} \right\| \\
&\qquad \leq d_{NT}^3 \mathcal{O}_P(N^{-3/4}), \\
&\max_{r,s} \left| w_r' M_{f^0} e' M_{\lambda^0} \left[(\beta - \beta^0) \cdot X \right] f^0 (f^{0\prime} f^0)^{-1} (\lambda^{0\prime}\lambda^0)^{-1} \lambda^{0\prime} e M_{f^0} w_s \right| \\
&\qquad \leq K \|e\| \|\beta - \beta^0\| \left\| f^0 (f^{0\prime} f^0)^{-1} (\lambda^{0\prime}\lambda^0)^{-1} \lambda^{0\prime} \right\| \max_{r,k} \|v_r' X_k P_{f^0}\| \max_r \|w_r' e' P_{\lambda^0}\| \\
&\qquad \leq d_{NT}^2 \mathcal{O}_P(N^{-1/2}), \\
&\max_{r,s} \left| w_r' M_{f^0} e' M_{\lambda^0} e f^0 (f^{0\prime} f^0)^{-1} (\lambda^{0\prime}\lambda^0)^{-1} \lambda^{0\prime} \left[(\beta - \beta^0) \cdot X \right] M_{f^0} w_s \right| \\
&\qquad \leq K \|e\| \|\beta - \beta^0\| \left\| f^0 (f^{0\prime} f^0)^{-1} (\lambda^{0\prime}\lambda^0)^{-1} \lambda^{0\prime} \right\| \max_{r,k} \|v_r' e P_{f^0}\| \max_r \|w_r' X_k' P_{\lambda^0}\| \\
&\qquad \leq d_{NT}^2 \mathcal{O}_P(N^{-1/2}).
\end{aligned}
$$

and analogously one can check that

$$
\max_{r,s} \left| w_r' B^{(eeee)} w_s \right| \leq d_{NT}^2 \mathcal{O}_P(N^{-1}) + d_{NT}^3 \mathcal{O}_P(N^{-3/2}). \tag{S.105}
$$

All in all, we thus have

$$
\begin{aligned}
b_{NT} &\leq \mathcal{O}_P(N^{-1/2}) + d_{NT} \mathcal{O}_P(N^{-1/4}) + d_{NT}^2 \mathcal{O}_P(N^{-1/2}) + d_{NT}^3 \mathcal{O}_P(N^{-3/4}) \\
&\leq d_{NT} \mathcal{O}_P(N^{-1/4}) \ ,
\end{aligned} \tag{S.106}
$$

where in the last step we used that by assumption $d_{NT} \geq 1$ and $d_{NT} = o_P(N^{1/4})$. Therefore

$$
\sum_{r=q_{NT}}^{T-R^0} \frac{b_{NT}}{\rho_{R-R^0} - \rho_r} = q_{NT} d_{NT} \mathcal{O}_P(N^{-1/4}) \frac{1}{q_{NT}} \leq \sum_{r=q_{NT}}^{T-R^0} \frac{1}{\rho_{R-R^0} - \rho_r} = o_P(1), \tag{S.107}
$$

so that Lemma S.11 is indeed applicable asymptotically, and we find

$$
\left| \mu_r \left(C_1^{\pm}(\beta) \right) - \rho_r \right| \ \leq \ \frac{q_{NT} \, b_{NT}}{1 - o_P(1)} \ \leq \ q_{NT} \, d_{NT} \, \mathcal{O}_P(N^{-1/4}) = o_P(1) \ . \tag{S.108}
$$

For $t = 1, \ldots, R - R^0$ we thus have

$$\mu_r\left(C_1^{\pm}(\beta)\right) = \rho_r + o_P(1) \geq \rho_{R-R^0} + o_P(1) \geq \|C_2^{\pm}\|, \quad \text{wpa1}, \tag{S.109}$$

where the last step follows because $\|C_2^{\pm}\| = aN/2 + \mathcal{O}_P(1)$ and we assumed $\rho_{R-R^0} > aN$, wpa1. Since $C^{\pm}(\beta)$ is block-diagonal with blocks $C_1^{\pm}(\beta)$ and $C_2^{\pm}$ (in the basis defined by f^0), and $\mu_r\left(C_1^{\pm}(\beta)\right) \geq \|C_2^{\pm}\|$, it must be the case that wpa1 the largest $R - R^0$ eigenvalues of $C^{\pm}(\beta)$ are those of $C_1^{\pm}(\beta)$. Thus,

$$\left|\mu_r\left(C^{\pm}(\beta)\right) - \rho_r\right| = o_P(1)\ , \tag{S.110}$$

and also

$$\left|\mu_r\left(B(\beta) + B'(\beta)\right) - \rho_r\right| = o_P(1)\ , \tag{S.111}$$

which holds uniformly over all β with $N^{3/4}\,\|\beta - \beta^0\| \leq c$. This concludes the proof. ■

Proof of Lemma S.13. # Part (i). Since e has $iid\mathcal{N}(0, \sigma^2)$ entries, independent of λ^0 and f^0, rotational invariance dictates that the distribution of v_r and w_r is given by the Haar measure on the unit sphere of dimension $N - R^0$ and $T - R^0$, respectively, and the the lemma just provides a concrete representation of this. The bounds on $\mathbb{E}\left(\sqrt{N}\|M_{\lambda^0}\widetilde{v}\|^{-1}\right)^{\xi}$ and $\mathbb{E}\left(\sqrt{T}\|M_{f^0}\widetilde{w}\|^{-1}\right)^{\xi}$ follow, because the inverse chi-square distribution with dof ν possesses all moments smaller than $\nu/2$.

Part (ii). Using part (i) of the lemma we have

$$w_r \underset{d}{=} \|M_{f^0}\widetilde{w}\|^{-1} M_{f^0}\widetilde{w} = \left\|\frac{M_{f^0}\widetilde{w}}{\sqrt{T}}\right\|^{-1}\left(\frac{\widetilde{w}}{\sqrt{T}} - \frac{P_{f^0}\widetilde{w}}{\sqrt{T}}\right), \tag{S.112}$$

where $\widetilde{w}$ be a T-vector with $iid\mathcal{N}(0, 1)$ entries. It is also useful to define the time shift operator $L : \mathbb{R}^T \to \mathbb{R}^T$, which satisfies $(Lw_r)_t = w_{r,t-1}$, and therefore $(L^\tau w_r)_t = w_{r,t-\tau}$. We then have

$$\begin{aligned}\sum_{t=\tau+1}^{T} w_{r,t} w_{r,t-\tau} &= w_r' L^\tau w_r \\ &\underset{d}{=} \left\|\frac{M_{f^0}\widetilde{w}}{\sqrt{T}}\right\|^{-2} \frac{1}{T}\left(\widetilde{w}' L^\tau \widetilde{w} - \widetilde{w}' L^\tau P_{f^0}\widetilde{w} - \widetilde{w}' P_{f^0} L^\tau \widetilde{w} + \widetilde{w}' P_{f^0} L^\tau P_{f^0}\widetilde{w}\right).\end{aligned} \tag{S.113}$$

Given the distribution of $\widetilde{w}$ it is easy to show that $\left|\frac{\widetilde{w}' L^\tau \widetilde{w}}{\sqrt{T}}\right|$ has arbitrary high bounded moments as T becomes large, i.e. we have $\mathbb{E}\left|\frac{\widetilde{w}' L^\tau \widetilde{w}}{\sqrt{T}}\right|^{\xi} = \mathcal{O}(1)$ for any $\tau \geq 1$ and any $\xi > 0$. Furthermore, using that $\|L\| = 1$ we can bound

$$\begin{aligned}|\widetilde{w}' L^\tau P_{f^0}\widetilde{w}| &\leq \|\widetilde{w}\|\|P_{f^0}\widetilde{w}\| \\ |\widetilde{w}' P_{f^0} L^\tau \widetilde{w}| &\leq \|\widetilde{w}\|\|P_{f^0}\widetilde{w}\| \\ |\widetilde{w}' P_{f^0} L^\tau P_{f^0}\widetilde{w}| &\leq \|P_{f^0}\widetilde{w}\|^2,\end{aligned} \tag{S.114}$$

where $\|\widetilde{w}\|^2 \underset{d}{=} \chi^2(T)$ and $\|P_{f^0}\widetilde{w}\|^2 \underset{d}{=} \chi^2(R^0)$. Note that the rhs of the inequalities in the last display do not depend on τ, i.e. the bounds are uniform over τ. The χ-square distribution with R^0 degrees of freedom does not depend on T and has finite moments of all orders. Since $\|\widetilde{w}\|^2$ is $\chi^2(T)$ distributed we find that $\frac{1}{\sqrt{T}}\|\widetilde{w}\|$ has arbitrarily high uniformly bounded moments as T becomes large. Combining these results we obtain that all moments of $\left|\frac{1}{\sqrt{T}}\left(\widetilde{w}'L^\tau\widetilde{w} - \widetilde{w}'L^\tau P_{f^0}\widetilde{w} - \widetilde{w}'P_{f^0}L^\tau\widetilde{w} + \widetilde{w}'P_{f^0}L^\tau P_{f^0}\widetilde{w}\right)\right|$ are uniformly bounded as T becomes large. Part (i) of the lemma shows that the same is true for $\left\|\frac{M_{f^0}\widetilde{w}}{\sqrt{T}}\right\|^{-2}$. Using Holder's inequality we thus find that for all $\xi > 0$ we have

$$\mathbb{E}\left|\sum_{t=\tau+1}^{T} w_{r,t}w_{r,t-\tau}\right|^\xi = \mathbb{E}\left|\|M_{f^0}\widetilde{w}\|^{-2}\widetilde{w}'M_{f^0}L^\tau M_{f^0}\widetilde{w}\right|^\xi = \mathcal{O}(1/\sqrt{T}), \tag{S.115}$$

uniformly over r and τ. From this we obtain $\max_{r,\tau}\left|\sum_{t=\tau+1}^{T} w_{r,t}w_{r,t-\tau}\right| = \mathcal{O}_P(T^{-1/2+\varepsilon})$ for any $\varepsilon > 0$ (namely $\varepsilon = 2/\xi$). This is the statement of the lemma for the special case where $r = s$.

What is left to show is that $\max_{r\neq s}\max_\tau\left|\sum_{t=\tau+1}^{T} w_{r,t}w_{s,t-\tau}\right| = \mathcal{O}_P(T^{-1/2+\varepsilon})$, for $\varepsilon \in [0, 1/12)$. Let $\widetilde{w}^a$ and $\widetilde{w}^b$ be two T-vector with $iid\mathcal{N}(0,1)$ entries, independent of each other, and independent of f^0. Then we have for any $r, s = 1, \ldots, Q$ with $r \neq s$ that

$$\begin{pmatrix} w_r \\ w_s \end{pmatrix} \underset{d}{=} \begin{pmatrix} \|M_{f^0}\widetilde{w}^a\|^{-1}M_{f^0}\widetilde{w}^a \\ \|M_{f^0}M_{w^a}\widetilde{w}^b\|^{-1}M_{f^0}M_{w^a}\widetilde{w}^b \end{pmatrix}. \tag{S.116}$$

Note that this representation of the joint distribution accounts for the constraint $w_r'w_s = 0$, in addition to $\|w_r\| = \|w_s\| = 1$ and the invariance under the orthogonal group $O(T - R^0)$. Using this representation of the joint distribution of w_r and w_s the proof is now analogous to the case $r = s$. The result can be shown for any $\varepsilon > 0$.

\# Part (iii). This again follows since e has $iid\mathcal{N}(0,\sigma^2)$ entries and from the resulting rotational invariance of e wrt to orthogonal $O(N)$ and $O(T)$ rotations from the left and right, respectively. ■

S.7 Additional Monte Carlo Simulations

S.7.1 "Empirical Monte Carlo"

The static model in the empirical illustration reads

$$Y_{it} = \sum_{k=1}^{8} \beta_k X_{k,it} + \alpha_i + \gamma_i\, t + \delta_i\, t^2 + \mu_t + \lambda_i' f_t + e_{it}.$$

As described in the main text, estimates for β, λ_i and f_t are obtained by applying the LS estimation procedure with $R = 4$ to $\widetilde{Y}_{it} = \sum_{k=1}^{8} \beta_k \widetilde{X}_{k,it} + \widetilde{\lambda}_i'\widetilde{f}_t + \widetilde{e}_{it}$, where $\widetilde{Y} = M_{1_N} Y M_{(1_T,\mathbf{t},\mathbf{t^2})}$ and $\widetilde{X}_k = M_{1_N} X_k M_{(1_T,\mathbf{t},\mathbf{t^2})}$ are the outcome variable and regressors after projecting out α_i, γ_i,

δ_i and μ_t. We then construct the bias corrected estimator $\widehat{\beta}_R^{\text{BC}}$, as reported in the $R = 4$ column of Table 3. We afterwards estimate α_i, γ_i, δ_i and μ_t by applying least squares with outcome variable given by the residuals $Y_{it} - \sum_{k=1}^{8} \widehat{\beta}_{R,k}^{\text{BC}} X_{k,it} - \widehat{\Lambda}'_{R,i}\widehat{F}_{R,t}$, obtaining $\widehat{\alpha}_{R,i}$, $\widehat{\gamma}_{R,i}$, $\widehat{\delta}_{R,i}$ and $\widehat{\mu}_{R,t}$.

For the simulation we generate e_{it} according to the $MA(1)$ process

$$e_{it}^s = 0.1(v_{it} + v_{i,t-1}),$$

where $v_{it} \sim iid\, t(5)$, i.e. v_{it} has a Student's t-distribution with 5 degrees of freedom. The factor 0.1 in the formula for e_{it}^s was chosen to reproduce standard deviations for $\widehat{\beta}_k^{\text{BC}}$ in the simulation that are close to the estimated standard errors in the actual application.

We set $R^0 = 4$ and generate the simulated outcome variable as

$$Y_{it}^s = \sum_{k=1}^{8} \widehat{\beta}_{R^0,k}^{\text{BC}} X_{k,it} + \widehat{\alpha}_{R^0,i} + \widehat{\gamma}_{R^0,i}\, t + \widehat{\delta}_{R^0,i}\, t^2 + \widehat{\mu}_{R^0,t} + \widehat{\Lambda}'_{R^0,i}\widehat{F}_{R^0,t} + e_{it}^s.$$

The sample size is $N = 48$ and $T = 33$, as in the real data. We generate 10.000 Monte Carlo samples in this way and for each sample apply the same bias corrected estimator for β that was reported in Table 3 for the real data.

Table 1 reports the finite sample bias and standard deviation of $\widehat{\beta}_R^{\text{BC}}$ for $R \in \{0, \ldots, 9\}$. We also report the empirical size of a size 5% t-test for whether each coefficient is equal to its true value. The standard error estimator used for the t-test allows for heteroscedasticity and serial correlation.

We find that the bias corrected estimates for β_k, $k = 1, \ldots, 8$, are essentially unbiased when $R \geq R^0$ factors are used in the estimation, but for $R < R^0$ the coefficient estimates are often biased. For β_k, $k \geq 3$, there are only small changes in the standard deviation of the estimator between $R = 4$ and $R = 9$, but for β_k, $k = 1, 2$, we observe standard deviation inflation of up to 25% between $R = 4$ and $R = 9$.

For $k \geq 5$ the empirical sizes of the t-test are quite accurate, but for $k \leq 4$ the finite sample t-test overrejects the null even for $R = R^0 = 4$.

Given the relatively small sample size the difference between $R = 9$ and $R^0 = 4$ is relatively large, and some finite sample inefficiency and size distortions are not too surprising.

S.7.2 Dynamic Model

Here, we consider an AR(1) panel model with two factors ($R^0 = 2$) and the following data generating process (DGP):

$$Y_{it} = \beta^0 Y_{i,t-1} + \sum_{r=1}^{2} \lambda_{ir} f_{tr} + e_{it}, \qquad f_{tr} = 0.5\, f_{t-1,r} + \frac{\varepsilon_{tr}}{\sqrt{1 - 0.5^2}}. \tag{S.117}$$

The random variables λ_{ir}, ε_{tr} and e_{it} are mutually independent; with $\lambda_{ir} \sim iid\mathcal{N}(1, 1)$; and ε_{tr} and $e_{it} \sim iid\mathcal{N}(0, 1)$. The AR(1) processes for Y_{it} and f_{tr} are initiated with 100 time periods before the actual estimation sample starts, so that the initial conditions roughly correspond to the long-run static distribution. We choose $\beta^0 \in \{0.2, 0.5, 0.8\}$, and use $10,000$ repetitions in our simulation. The true number of factors is chosen to be $R^0 = 2$. For each draw of Y and X we compute the LS estimator $\widehat{\beta}_R$ according to equation (3.1) for different values of R.

Table 2 reports bias and standard deviation of the estimator $\widehat{\beta}_R$ for $N = 300$ and different combinations of R, T and β^0. Table 3 reports various quantiles of the distribution of $\sqrt{NT}(\widehat{\beta}_R - \beta^0)$ for $N = 300$ and different combinations of R, T and β^0. Table 4 reports the size of a t-test with nominal size equal to 5% for $R \geq R^0$. We use the results in Bai (2009) and Moon and Weidner (2014) to correct for the leading $1/N$ (not actually present in our DGP) and $1/T$ (present in our DGP) biases in $\widehat{\beta}_R$ before calculating the t-test statistics, allowing for predetermined regressors and heteroscedsticity in both panel dimensions when estimating the bias and standard deviation of $\widehat{\beta}_R$.

S.8 Comments Regarding Numerical Calculation of $\widehat{\beta}_R$

Different iteration schemes can be used to implement the LS estimator defined in (3.1) numerically:

(1) Ahn, Lee and Schmidt (2001) use an iteration scheme were the following steps are repeated until convergence: (a) for fixed $\widetilde{\beta}$ find $\widetilde{F}$ and $\widetilde{\Lambda}$ that minimize the LS objective function in (3.1) via principal component analysis (but $\widetilde{\Lambda}$ need not actually be computed); (b) for fixed $\widetilde{F}$ find $\widetilde{\beta}$ and $\widetilde{\Lambda}$ that minimize the LS objective function in (3.1) (but $\widetilde{\Lambda}$ need not actually be computed, because $\widetilde{\beta}$ can be obtained by regressing Y on $X_k M_{\widetilde{F}}$).

(2) Alternatively, Bai (2009) proposes the following iteration steps: (a) for fixed $\widetilde{\beta}$ find $\widetilde{F}$ and $\widetilde{\Lambda}$ that minimize the LS objective function in (3.1) via principal component analysis; (b) for fixed $\widetilde{F}$ and $\widetilde{\Lambda}$ find $\widetilde{\beta}$ that minimizes the LS objective function in (3.1) by running a regression of $(Y - \widetilde{\Lambda}\widetilde{F}')$ on X_k.

(3) Another iteration scheme, which we have used in our implementation, and we have not found discussed previously in the literature, is the following: (a) for fixed $\widetilde{\beta}$ find $\widetilde{F}$ and $\widetilde{\Lambda}$ that minimize the LS objective function in (3.1) via principal component analysis; (b) for fixed $\widetilde{F}$ and $\widetilde{\Lambda}$ find $\widetilde{\beta}$ that minimizes the alternative objective function $\|M_{\widetilde{\Lambda}}(Y - \beta \cdot X) M_{\widetilde{F}}\|^2_{HS}$ by running a regression of Y on $M_{\widetilde{\Lambda}} X_k M_{\widetilde{F}}$.

All three iteration schemes have the same step (a), i.e. differ from each other only in step (b). Each step of the iteration schemes (1) and (2) minimizes the LS objective function, i.e. those schemes guarantee that the sum of squared residuals is non-increasing in each step. In contrast, step (b) in scheme (3) minimizes an alternative objective function, i.e. it is possible that the LS objective function in (3.1) is actually increasing during that step. However, this step can nevertheless be justified, namely one can show that close to any (local) minimum the profile objective function $\mathcal{L}^R_{NT}(\beta)$ is well approximated by the alternative objective function $\frac{1}{NT}\|M_{\widetilde{\Lambda}}(Y - \beta \cdot X) M_{\widetilde{F}}\|^2_{HS}$, i.e. step (b) in scheme (3) is minimizing an approximation of $\mathcal{L}^R_{NT}(\beta)$.[9]

[9]Step (b) in scheme (1) and (2) can be equivalently described as minimizing the objective functions $\frac{1}{NT}\|(Y - \beta \cdot X) M_{\widetilde{F}}\|^2_{HS}$ and $\frac{1}{NT}\|(Y - \beta \cdot X - \widetilde{\Lambda}\widetilde{F}')\|^2_{HS}$, respectively, which are also approximations of $\mathcal{L}^R_{NT}(\beta)$. However,

Bai (2009) points out that the iteration scheme (2) is somewhat more robust towards the choice of starting value for β, which was confirmed in our simulations exercises, both compared to scheme (1) and to scheme (3). However, once close to a (local) minimum of the LS objective function we found the convergence rate of scheme (3) to be significantly faster than the conference rates of scheme (1) and (2). Scheme (1) performed between scheme (2) and (3) in terms of both robustness and speed. Each iteration scheme therefore has its relative advantages and disadvantages. We use scheme (3) for our final implementation, because the LS objective (and the profile objective function $\mathcal{L}^R_{NT}(\beta)$) can have multiple local minima, so that multiple optimization runs with different starting values are usually necessary anyways to achieve confidence that the global minimum was actually found. By using scheme (3) we minimize the time required for each optimization run, which enables us to try out more starting values within the same amount of total CPU time. Combining different iteration schemes (e.g. starting with scheme (2) and switching to scheme (3) once close to a minimum) might also be a good idea, which we have not explored, however.

S.9 Verifying the Assumptions in Bai (2009) for Example in Section 4.3

Throughout this section we only consider the particular DGP in the example of Section 4.3 in the main text. For this DGP it is easy to see that the OLS estimator $\widehat{\beta}_0$ (the LS-estimator with $R = R^0 = 0$) is $\sqrt{NT}$-consistent, while the example shows that $\widehat{\beta}_1$ (the LS-estimator with $R = 1$) is only $\sqrt{N}$-consistent. In the following we show that the regularity conditions imposed in Bai (2009) are also satisfied for this DGP. This is interesting to verify, since then example also shows that we need stronger Assumptions than those imposed in Bai (2009) in order to derive our results for $\widehat{\beta}_R$ for $R > R^0$.

Verifying Assumptions A, B, D, E

- Since $R^0 = 0$ we find that Assumption A in Bai (2009) becomes $\frac{1}{NT}\sum_{i,t} X^2_{it} > 0$, which is satisfied. The assumption would also be satisfied for $R^0 > 0$, since the component $\widetilde{X}$ makes the regressors X a "high-rank regressor".

- Bai's Assumption B is trivially satisfied for $R^0 = 0$.

- Assumption D in Bai (2009) requires strict exogeneity of the regressors in the sense that X and e are independent, which is also satisfied.[10]

those approximations are less precise than the approximation in step (b) of scheme (3). Namely, close to the minimizer $\widehat{\beta}_R$ of $\mathcal{L}^R_{NT}(\beta)$ we have $\mathcal{L}^R_{NT}(\beta) = \frac{1}{NT}\|M_{\widetilde{\Lambda}}(Y - \beta \cdot X)M_{\widetilde{F}}\|^2_{HS} + \mathcal{O}_P(\|\beta - \widehat{\beta}_R\|^3)$, while the other two approximations have remainders of order $\|\beta - \widehat{\beta}_R\|^2$.

[10] One could also consider λ_x and f_x as random, but independent of e and $\widetilde{X}$. In that case X and e are still strictly exogenous in the sense of mean-independence, i.e. we have $\mathbb{E}(e|X) = 0$, but e and X are not fully independent. However, our Corollary 4.3 in the main text (see also Moon and Weidner (2014)) shows that the asymptotic distribution of $\widehat{\beta}_{R^0}$ can be derived under the weaker exogeneity assumption $\mathbb{E}(e_{it}X_{it}) = 0$. Full independence of e and X is therefore only assumed for convenience in Bai (2009), and his results on $\widehat{\beta}_{R^0}$ remain

- Finally, Assumption E in Bai (2009) becomes $\frac{1}{\sqrt{NT}}\sum_{it} X_{it}e_{it} \to_d \mathcal{N}(0, D_Z)$, where $D_Z = \lim_{N,T\to\infty} \mathrm{Var}\left[\frac{1}{\sqrt{NT}}\sum_{it} X_{it}e_{it}\right]$. This is also satisfied, since $X_{it}e_{it}$ is independent across i and over t, and has bounded variance.

Verifying Assumption C

A more difficult task is to verify Assumption C in Bai (2009), which contains regularity conditions for e_{it}.[11] In our notation the assumption reads

(i) $\mathbb{E}(e_{it}) = 0$ and $\mathbb{E}(e_{it}^8) \leq M$,

(ii) Let $\mathbb{E}(e_{it}e_{js}) = \sigma_{ij,ts}$. Then, $\frac{1}{N}\sum_{i,j=1}^{N} \sup_{t,s} |\sigma_{ij,ts}| \leq M$, $\frac{1}{T}\sum_{t,s=1}^{T} \sup_{i,j} |\sigma_{ij,ts}| \leq M$, and $\frac{1}{NT}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T} |\sigma_{ij,ts}| \leq M$. Also, the largest eigenvalue of $\mathbb{E}(e_i e_i')$ is bounded uniformly in i and T.

(iii) For every (t,s), $\mathbb{E}\left|\frac{1}{N^{1/2}}\sum_{i=1}^{N}[e_{is}e_{it} - \mathbb{E}(e_{is}e_{it})]\right|^4 \leq M$.

(iv) Moreover,

$$\begin{aligned} \frac{1}{NT^2}\sum_{t,s,u,v}\sum_{i,j} |\mathrm{Cov}(e_{it}e_{is}, e_{ju}e_{jv})| &\leq M \\ \frac{1}{N^2T}\sum_{t,s}\sum_{i,j,k,l} |\mathrm{Cov}(e_{it}e_{jt}, e_{ks}e_{ls})| &\leq M. \end{aligned}$$

In the following M always denotes some global constant, whose precise value may change from equation to equation. Furthermore, we simply write λ and f instead of λ_x and f_x. We use notation $1\{\cdot\}$ to denote the indicator function.

We also define $v_1 = c\frac{\lambda}{\sqrt{N}}\frac{\lambda' u}{\sqrt{N}}$, $v_2 = c\frac{uf}{\sqrt{T}}\frac{f'}{\sqrt{T}}$, $v_3 = \left(\frac{c^2}{\sqrt{NT}}\lambda' uf\right)\frac{\lambda}{\sqrt{N}}\frac{f'}{\sqrt{T}}$, and $v = v_1 + v_2 + v_3$. We then have

$$e = \left(I + c\frac{\lambda\lambda'}{N}\right) u \left(I + c\frac{ff'}{T}\right) = u + v. \tag{S.118}$$

unchanged when only imposing $\mathbb{E}(e|X) = 0$ instead.

[11]Essentially, Assumption C requires that e_{it} is mean zero and weakly correlated across i and t. Thus, it plays the same role as our high-level Assumption SN(ii), which is easy to check since $\|e\| \leq \left\|\mathbb{1}_N + c\frac{\lambda_x\lambda_x'}{N}\right\| \|u\| \left\|\mathbb{1}_T + c\frac{f_x f_x'}{T}\right\|$ and we have $\left\|\mathbb{1}_N + c\frac{\lambda_x\lambda_x'}{N}\right\| \leq \|\mathbb{1}_N\| + c\left\|\frac{\lambda_x\lambda_x'}{N}\right\| = \mathcal{O}(1)$, and $\left\|\mathbb{1}_T + c\frac{f_x f_x'}{T}\right\| \leq \|\mathbb{1}_T\| + c\left\|\frac{f_x f_x'}{T}\right\| = \mathcal{O}(1)$, and $\|u\| = \mathcal{O}_P(\sqrt{\min(N,T)})$ — see Appendix A.1 in the main text regarding the last statement.

Notice that

$$\begin{aligned} v_{it} &= \frac{\lambda_i}{\sqrt{N}}\left(\frac{c}{\sqrt{N}}\sum_{h=1}^{N}\lambda_h u_{ht}\right)+\left(\frac{c}{\sqrt{T}}\sum_{\tau=1}^{T}u_{i\tau}f_\tau\right)\frac{f_t}{\sqrt{T}}+\left(\frac{c^2}{\sqrt{NT}}\lambda' u f\right)\frac{\lambda_i}{\sqrt{N}}\frac{f_t}{\sqrt{T}} \\ &= \frac{\lambda_i}{\sqrt{N}}\tilde{\lambda}\left(\lambda,u_{\cdot t}\right)+\tilde{f}\left(f,u_{i\cdot}\right)\frac{f_t}{\sqrt{T}}+\tilde{v}\left(\lambda,f,u\right)\frac{\lambda_i}{\sqrt{N}}\frac{f_t}{\sqrt{T}} \\ &= v_{1,it}+v_{2,it}+v_{3,it}, \end{aligned}$$

where we defined $\tilde{\lambda}\left(\lambda,u_{\cdot t}\right)$, $\tilde{f}\left(f,u_{i\cdot}\right)$ and $\tilde{v}\left(\lambda,f,u\right)$ implicitly. We also define $g_{1,ij,ts}=u_{it}u_{js}$, $g_{2,ij,ts}=u_{it}v_{js}$, $g_{3,ij,ts}=v_{it}u_{js}$, and $g_{4,ij,ts}=v_{it}v_{js}$, so that

$$e_{it}e_{js} = g_{1,ij,ts}+g_{2,ij,ts}+g_{3,ij,ts}+g_{4,ij,ts}.$$

In the following we discuss part (i), (ii), (iii) and (iv) of Assumption C separately:

Part (i)

This is straightforward to check.

Part (ii)

Let $\sigma_{k,ij,ts}=\mathbb{E}\left(g_{k,ij,ts}\right).$

1. $\sigma_{1,ij,ts}$: The desired result follows since

$$\sigma_{1,ij,ts}=1\left\{i=j,t=s\right\}.$$

2. $\sigma_{2,ij,ts}$: By definition,

$$\sigma_{2,ij,ts}=\mathbb{E}\left(u_{it}v_{js}\right)=\mathbb{E}\left(u_{it}v_{1,js}\right)+\mathbb{E}\left(u_{it}v_{2,js}\right)+\mathbb{E}\left(u_{it}v_{3,js}\right).$$

Direct calculations show that

$$\begin{aligned}
\mathbb{E}\left(u_{it}v_{1,js}\right) &= \mathbb{E}\left[u_{it}\frac{\lambda_j}{\sqrt{N}}\left(\frac{c}{\sqrt{N}}\sum_{h=1}^{N}\lambda_h u_{hs}\right)\right] = \frac{c}{N}\sum_{h=1}^{N}\lambda_j\lambda_h\mathbb{E}\left(u_{hs}u_{it}\right) \\
&= \frac{c}{N}\sum_{h=1}^{N}\lambda_j\lambda_h 1\left\{i=h, s=t\right\} \\
&= \frac{c}{N}\lambda_i\lambda_j 1\left\{t=s\right\}. \\
\mathbb{E}\left(u_{it}v_{2,js}\right) &= \mathbb{E}\left[u_{it}\left(\frac{c}{\sqrt{T}}\sum_{\tau=1}^{T}u_{j\tau}f_\tau\right)\frac{f_s}{\sqrt{T}}\right] = \frac{c}{T}\sum_{\tau=1}^{T}f_\tau f_s\mathbb{E}\left(u_{it}u_{j\tau}\right) \\
&= \frac{c}{T}\sum_{\tau=1}^{T}f_\tau f_s 1\left\{i=j, t=\tau\right\} \\
&= \frac{c}{T}f_t f_s 1\left\{i=j\right\}. \\
\mathbb{E}\left(u_{it}v_{3,js}\right) &= \mathbb{E}\left[u_{it}\left(\frac{c^2}{\sqrt{NT}}\sum_{h=1}^{N}\sum_{\tau=1}^{T}\lambda_h f_\tau u_{h\tau}\right)\frac{\lambda_j}{\sqrt{N}}\frac{f_s}{\sqrt{T}}\right] = \frac{c^2}{NT}\sum_{h=1}^{N}\sum_{\tau=1}^{T}\lambda_h\lambda_j f_\tau f_s\mathbb{E}\left(u_{it}u_{h\tau}\right) \\
&= \frac{c^2}{NT}\sum_{h=1}^{N}\sum_{\tau=1}^{T}\lambda_h\lambda_j f_\tau f_s 1\left\{i=h, t=\tau\right\} \\
&= \frac{c^2}{NT}\lambda_i\lambda_j f_t f_s.
\end{aligned}$$

Combining these, we have

$$\sigma_{2,ij,ts} = \frac{c}{N}\lambda_i\lambda_j 1\left\{t=s\right\} + \frac{c}{T}f_t f_s 1\left\{i=j\right\} + \frac{c^2}{NT}\lambda_i\lambda_j f_t f_s.$$

Since λ_i and f_t are bounded, we have the desired result,

$$\begin{aligned}
\frac{1}{N}\sum_{i,j=1}^{N}\sup_{t,s}\left|\frac{c}{N}\lambda_i\lambda_j 1\left\{t=s\right\} + \frac{c}{T}f_t f_s 1\left\{i=j\right\} + \frac{c^2}{NT}\lambda_i\lambda_j f_t f_s\right| &\leq M, \\
\frac{1}{T}\sum_{t,s=1}^{T}\sup_{i,j}\left|\frac{c}{N}\lambda_i\lambda_j 1\left\{t=s\right\} + \frac{c}{T}f_t f_s 1\left\{i=j\right\} + \frac{c^2}{NT}\lambda_i\lambda_j f_t f_s\right| &\leq M, \\
\frac{1}{NT}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}\left|\frac{c}{N}\lambda_i\lambda_j 1\left\{t=s\right\} + \frac{c}{T}f_t f_s 1\left\{i=j\right\} + \frac{c^2}{NT}\lambda_i\lambda_j f_t f_s\right| &\leq M.
\end{aligned}$$

3. $\sigma_{3,ij,ts}$: The result for the term $\sigma_{3,ij,ts}$ follows similarly to the case of $\sigma_{2,ij,ts}$.

4. $g_{4,ij,ts}$: By definition, we have

$$
\begin{aligned}
\mathbb{E}\left(g_{4,ij,ts}\right) &= \sum_{k=1}^{3}\sum_{l=1}^{3}\mathbb{E}\left(v_{k,it}v_{l,js}\right) \\
&= \mathbb{E}\left(\begin{array}{c}\left[\begin{array}{c}\frac{\lambda_i}{\sqrt{N}}\left(\frac{c}{\sqrt{N}}\sum_{k=1}^{N}\lambda_k u_{kt}\right)+\left(\frac{c}{\sqrt{T}}\sum_{p=1}^{T}f_p u_{ip}\right)\frac{f_t}{\sqrt{T}} \\ +\frac{\lambda_i}{\sqrt{N}}\left(\frac{c^2}{\sqrt{NT}}\sum_{k=1}^{N}\sum_{p=1}^{T}\lambda_k f_p u_{kp}\right)\frac{f_t}{\sqrt{T}}\end{array}\right] \\ \times\left[\begin{array}{c}\frac{\lambda_j}{\sqrt{N}}\left(\frac{c}{\sqrt{N}}\sum_{l=1}^{N}\lambda_l u_{ls}\right)+\left(\frac{c}{\sqrt{T}}\sum_{q=1}^{T}f_q u_{jq}\right)\frac{f_s}{\sqrt{T}} \\ +\frac{\lambda_j}{\sqrt{N}}\left(\frac{c^2}{\sqrt{NT}}\sum_{l=1}^{N}\sum_{q=1}^{T}\lambda_q f_l u_{lq}\right)\frac{f_s}{\sqrt{T}}\end{array}\right]\end{array}\right).
\end{aligned}
$$

Notice that

$$
\begin{aligned}
\mathbb{E}\left(v_{1,it}v_{1,js}\right) &= \frac{c^2}{N^2}\sum_{k=1}^{N}\sum_{l=1}^{N}\lambda_i\lambda_k\lambda_j\lambda_l\mathbb{E}\left(u_{kt}u_{ls}\right) \\
&= \frac{c^2}{N}\lambda_i\lambda_j\left(\frac{1}{N}\sum_{k=1}^{N}\lambda_k^2\right)1\left\{t=s\right\}.
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left(v_{1,it}v_{2,js}\right) &= \mathbb{E}\left[\frac{\lambda_i}{\sqrt{N}}\left(\frac{c}{\sqrt{N}}\sum_{k=1}^{N}\lambda_k u_{kt}\right)\left(\frac{c}{\sqrt{T}}\sum_{q=1}^{T}f_q u_{jq}\right)\frac{f_s}{\sqrt{T}}\right] \\
&= \frac{c^2}{NT}\left[\sum_{k=1}^{N}\sum_{q=1}^{T}\lambda_i\lambda_k f_q f_s\mathbb{E}\left(u_{kt}u_{jq}\right)\right] \\
&= \frac{c^2}{NT}\lambda_i\lambda_j f_t f_s,
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left(v_{1,it}v_{3,js}\right) &= \mathbb{E}\left[\frac{\lambda_i}{\sqrt{N}}\left(\frac{c}{\sqrt{N}}\sum_{k=1}^{N}\lambda_k u_{kt}\right)\frac{\lambda_j}{\sqrt{N}}\left(\frac{c^2}{\sqrt{NT}}\sum_{l=1}^{N}\sum_{q=1}^{T}\lambda_l f_q u_{lq}\right)\frac{f_s}{\sqrt{T}}\right] \\
&= \frac{c^3}{N^2T}\mathbb{E}\left[\sum_{k=1}^{N}\sum_{l=1}^{N}\sum_{q=1}^{T}\lambda_i\lambda_k\lambda_j\lambda_l f_q f_s\mathbb{E}\left(u_{kt}u_{lq}\right)\right] \\
&= \frac{c^3}{NT}\lambda_i\lambda_j f_t f_s\left(\frac{1}{N}\sum_{k=1}^{N}\lambda_k^2\right),
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left(v_{2,it}v_{2,js}\right) &= \mathbb{E}\left(\left(\frac{c}{\sqrt{T}}\sum_{p=1}^{T} f_p u_{ip}\right)\frac{f_t}{\sqrt{T}}\left(\frac{c^2}{\sqrt{T}}\sum_{q=1}^{T} f_q u_{jq}\right)\frac{f_s}{\sqrt{T}}\right) \\
&= \frac{c^3}{T^2}\sum_{p=1}^{T}\sum_{q=1}^{T} f_p f_t f_q f_s \mathbb{E}\left(u_{ip}u_{jq}\right) \\
&= \frac{c^3}{T}\left(\frac{1}{T}\sum_{p=1}^{T} f_p^2\right) f_t f_s 1\left\{i=j\right\},
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left(v_{2,it}v_{3,js}\right) &= \mathbb{E}\left[\left(\frac{c}{\sqrt{T}}\sum_{p=1}^{T} f_p u_{ip}\right)\frac{f_t}{\sqrt{T}}\frac{\lambda_j}{\sqrt{N}}\left(\frac{c^2}{\sqrt{NT}}\sum_{l=1}^{N}\sum_{q=1}^{T} \lambda_l f_q u_{lq}\right)\frac{f_s}{\sqrt{T}}\right] \\
&= \frac{c^3}{NT^2}\sum_{p=1}^{T}\sum_{l=1}^{N}\sum_{q=1}^{T} f_p f_t \lambda_j \lambda_l f_q f_s \mathbb{E}\left(u_{ip}u_{lq}\right) \\
&= \frac{c^3}{NT}\lambda_i\lambda_j f_t f_s\left(\frac{1}{T}\sum_{p=1}^{T} f_p^2\right),
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{E}\left(v_{3,it}v_{3,js}\right) &= \mathbb{E}\left(\frac{\lambda_i}{\sqrt{N}}\left(\frac{c^2}{\sqrt{NT}}\sum_{k=1}^{N}\sum_{p=1}^{T} \lambda_k f_p u_{kp}\right)\frac{f_t}{\sqrt{T}}\frac{\lambda_j}{\sqrt{N}}\left(\frac{c^2}{\sqrt{NT}}\sum_{l=1}^{N}\sum_{q=1}^{T} \lambda_q f_l u_{lq}\right)\frac{f_s}{\sqrt{T}}\right) \\
&= \frac{c^4}{N^2T^2}\sum_{k=1}^{N}\sum_{l=1}^{N}\sum_{p=1}^{T}\sum_{q=1}^{T} \lambda_i\lambda_k\lambda_j\lambda_q f_p f_t f_l f_s \mathbb{E}\left(u_{kp}u_{lq}\right) \\
&= \frac{c^4}{NT}\lambda_i\lambda_j f_t f_s\left(\frac{1}{N}\sum_{k=1}^{N}\lambda_k f_k\right)\left(\frac{1}{T}\sum_{p=1}^{T}\lambda_p f_p\right).
\end{aligned}
$$

From these and using the boundedness of λ_i and f_t we obtain

$$
\frac{1}{N}\sum_{i,j=1}^{N}\sup_{t,s}\left|\sigma_{4,ij,ts}\right| \leq M,\ \frac{1}{T}\sum_{t,s=1}^{T}\sup_{i,j}\left|\sigma_{4,ij,ts}\right| \leq M,\ \frac{1}{NT}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}\left|\sigma_{4,ij,ts}\right| \leq M.
$$

Combining these, we have the desired result.

What is left to show is the bound on the largest eigenvalue of $\Omega_i = \mathbb{E}\left(e_i e_i'\right)$, which is equivalent to the spectral norm of Ω_i. The spectral norm of a symmetric matrix is bounded by the infinity norm, i.e. we have $\mu_1(\Omega_i) = \|\Omega_i\| \leq \|\Omega_i\|_\infty = \max_t \sum_s |\Omega_{i,ts}|$. For the elements $\Omega_{i,ts}$ of the matrix Ω_i we have $\Omega_{i,ts} = \mathbb{E}\left(e_{it}e_{is}\right) = \sigma_{ii,ts}$. We thus have

$$
\mu_1(\Omega_i) \leq \max_t \sum_s \left|\sigma_{ii,ts}\right| = \mathcal{O}(1),
$$

where the last step follows by the above results on $\sigma_{ij,ts} = \sum_{k=1}^{4}\sigma_{k,ij,ts}$.

<u>**Part** *(iii)*</u>

Write

$$
\begin{aligned}
\mathbb{E}\left[\frac{1}{N^{1/2}}\sum_{i=1}^{N}\left[e_{is}e_{it}-\mathbb{E}\left(e_{is}e_{it}\right)\right]\right]^{4} &= \mathbb{E}\left\{\sum_{k=1}^{4}\left[\frac{1}{N^{1/2}}\sum_{i=1}^{N}\left(g_{k,ii,ts}-\sigma_{k,ii,ts}\right)\right]\right\}^{4} \\
&\leq M\left\{\mathbb{E}\left[\frac{1}{N^{1/2}}\sum_{i=1}^{N}\left(g_{k,ii,ts}-\sigma_{k,ii,ts}\right)\right]^{4}\right\}.
\end{aligned}
$$

1. $g_{1,ii,ts}$: Since $u_{it}\sim iidN\left(0,1\right)$ across i and over t, it is straightforward to see that for all t,s,

$$\mathbb{E}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(g_{1,ii,ts}-\sigma_{1,ii,ts}\right)\right]^{4}\leq M.$$

2. $g_{2,ii,ts}$: Next, notice that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(g_{2,ii,ts}-\sigma_{2,ii,ts}\right)=\sum_{k=1}^{3}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(u_{it}v_{k,is}-\mathbb{E}\left(u_{it}v_{k,is}\right)\right).$$

   Due to the boundedness of λ_i and f_t and iid normality of u_{it}, we have the following. First,

$$
\begin{aligned}
&\mathbb{E}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(u_{it}v_{1,is}-\mathbb{E}\left(u_{it}v_{1,is}\right)\right)\right)^{4} \\
=\ &\mathbb{E}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{u_{it}\frac{\lambda_j}{\sqrt{N}}\left(\frac{c}{\sqrt{N}}\sum_{h=1}^{N}\lambda_h u_{hs}\right)-\frac{c}{N}\lambda_i\lambda_j 1\left\{t=s\right\}\right\}\right]^{4} \\
=\ &\lambda_j^4 c^4\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\left\{u_{it}\left(\frac{1}{\sqrt{N}}\sum_{h=1}^{N}\lambda_h u_{hs}\right)-\frac{1}{\sqrt{N}}\lambda_i 1\left\{t=s\right\}\right\}\right]^{4} \\
\leq\ &M,
\end{aligned}
$$

   where the last equality follows since

$$\left(\frac{1}{N}\sum_{i=1}^{N}a_i\right)^{4}\leq\frac{1}{N}\sum_{i=1}^{N}a_i^4,$$

   and $\sup_{i,t,s}\mathbb{E}\left(\left\{u_{it}\left(\frac{1}{\sqrt{N}}\sum_{h=1}^{N}\lambda_h u_{hs}\right)-\frac{1}{\sqrt{N}}\lambda_i 1\left\{t=s\right\}\right\}^{4}\right)\leq M.$

Second, similarly to the first case, we have

$$
\begin{aligned}
& \mathbb{E}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(u_{it}v_{2,is}-\mathbb{E}\left(u_{it}v_{2,is}\right)\right)\right)^4 \\
= \; & \mathbb{E}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{u_{it}\left(\frac{c}{\sqrt{T}}\sum_{\tau=1}^{T}u_{i\tau}f_\tau\right)\frac{f_s}{\sqrt{T}}-\frac{cf_sf_t}{T}\right\}\right]^4 \\
= \; & f_s^4c^4\mathbb{E}\left[\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left\{u_{it}\left(\frac{1}{\sqrt{T}}\sum_{\tau=1}^{T}u_{i\tau}f_\tau\right)f_s-\frac{f_t}{\sqrt{T}}\right\}\right]^4 \\
\leq \; & M.
\end{aligned}
$$

Third,

$$
\begin{aligned}
& \mathbb{E}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(u_{it}v_{3,is}-\mathbb{E}\left(u_{it}v_{3,is}\right)\right)\right)^4 \\
= \; & \mathbb{E}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{u_{it}\left(\frac{c^2}{\sqrt{NT}}\sum_{h=1}^{N}\sum_{\tau=1}^{T}\lambda_hf_\tau u_{h\tau}\right)\frac{\lambda_i}{\sqrt{N}}\frac{f_s}{\sqrt{T}}-\frac{c^2}{NT}\lambda_i^2f_tf_s\right\}\right)^4 \\
= \; & f_s^4c^8\mathbb{E}\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left\{u_{it}\left(\frac{c^2}{\sqrt{NT}}\sum_{h=1}^{N}\sum_{\tau=1}^{T}\lambda_hf_\tau u_{h\tau}\right)\frac{\lambda_i}{\sqrt{N}}-\frac{c^2}{N\sqrt{T}}\lambda_i^2f_t\right\}\right)^4 \\
\leq \; & M.
\end{aligned}
$$

Combining these, we have

$$
\mathbb{E}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(g_{2,ii,ts}-\sigma_{2,ii,ts}\right)\right)^4 < M.
$$

3. $g_{3,ii,ts}$: Similarly, we find

$$
\mathbb{E}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(g_{3,ii,ts}-\sigma_{3,ii,ts}\right)\right)^4 < M,
$$

   because $g_{3,ii,ts}=g_{2,ii,st}$.

4. $g_{4,ii,ts}$: Finally,

$$
\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(g_{4,ii,ts}-\sigma_{4,ii,ts}\right)=\sum_{k=1}^{3}\sum_{l=1}^{3}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(v_{k,it}v_{l,is}-\mathbb{E}\left(v_{k,it}v_{l,is}\right)\right).
$$

Notice that

$$
\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(v_{1,it}v_{1,is}-\mathbb{E}\left(v_{1,it}v_{1,is}\right)\right)\\
=\;&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{\lambda_i}{\sqrt{N}}\left(\frac{c}{\sqrt{N}}\sum_{k=1}^{N}\lambda_k u_{kt}\right)\frac{\lambda_i}{\sqrt{N}}\left(\frac{c}{\sqrt{N}}\sum_{l=1}^{N}\lambda_l u_{ls}\right)-\frac{c^2}{N}\lambda_i^2\left(\frac{1}{N}\sum_{k=1}^{N}\lambda_k^2\right)1\left\{t=s\right\}\right)\\
=\;&\frac{c^2}{\sqrt{N}}\frac{1}{N}\sum_{i=1}^{N}\left\{\lambda_i^2\left\{\left(\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\lambda_k u_{kt}\right)\left(\frac{1}{\sqrt{N}}\sum_{l=1}^{N}\lambda_l u_{ls}\right)-\left(\frac{1}{N}\sum_{k=1}^{N}\lambda_k^2\right)1\left\{t=s\right\}\right\}\right\}.
\end{aligned}
$$

Notice that $\sup_{i,t,s}\mathbb{E}\left[\lambda_i^2\left\{\left(\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\lambda_k u_{kt}\right)\left(\frac{1}{\sqrt{N}}\sum_{l=1}^{N}\lambda_l u_{ls}\right)-\left(\frac{1}{N}\sum_{k=1}^{N}\lambda_k^2\right)1\left\{t=s\right\}\right\}\right]^4\leq M$. Therefore, we have

$$
\begin{aligned}
&\mathbb{E}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(v_{1,it}v_{1,is}-\mathbb{E}\left(v_{1,it}v_{1,is}\right)\right)\right]^4\\
\leq\;&\left(\frac{c^8}{N^2}\right)\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[\left\{\lambda_i^2\left\{\left(\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\lambda_k u_{kt}\right)\left(\frac{1}{\sqrt{N}}\sum_{l=1}^{N}\lambda_l u_{ls}\right)-\left(\frac{1}{N}\sum_{k=1}^{N}\lambda_k^2\right)1\left\{t=s\right\}\right\}\right\}^4\right]\\
\leq\;&M.
\end{aligned}
$$

Similarly, we can show the rest of the cases.

Part *(iv)*

Without loss of generality, we set $N=T$ here. We show that $\frac{1}{NT^2}\sum_{t,s,u,v}\sum_{i,j}\left|\mathrm{Cov}\left(e_{it}e_{is},e_{ju}e_{jv}\right)\right|\leq M$. The other case follows by the same fashion because the DGP is symmetric between i and t. Notice that

$$
\mathrm{Cov}\left(e_{it}e_{is},e_{ju}e_{jv}\right)=\mathrm{Cov}\left(\sum_{k=1}^{4}g_{k,ii,ts},\sum_{k=1}^{4}g_{k,jj,uv}\right)=\sum_{k=1}^{4}\sum_{l=1}^{4}\mathrm{Cov}\left(g_{k,ii,ts},g_{l,jj,uv}\right).
$$

Among $\{\mathrm{Cov}\left(g_{1,ii,ts},g_{1,jj,uv}\right)\}$ there are six kinds, (a) the term of (u,u) and (u,u) (b) the terms of (u,u) and (u,v) (c) the terms of (u,u) and (v,v), (d) the terms of (u,v) and (u,v), (e) the terms of (u,v) and (v,v), and (f) the term of (v,v) and (v,v).

In what follows we use "two pairs among $\{t_1,t_2,t_3,t_4\}$" to denote the sum of the three terms like $1\{t_1=t_2\}1\{t_3=t_4\}$.

The main step in establishing the required result, $\frac{1}{NT^2}\sum_{t,s,u,v}\sum_{i,j}\left|\mathrm{Cov}\left(g_{k,ii,ts},g_{l,jj,uv}\right)\right|\leq M$, is to find an upper bound of $\left|\mathrm{Cov}\left(g_{k,ii,ts},g_{l,jj,uv}\right)\right|$ in the form of $\frac{1}{N^aT^b}1\{$ some pairs of indices$\}$, so that the power of $NT^2N^aT^b=N^{a+b+3}$, $a+b+3$, is larger than or equal to the number of outstanding summations.

In the following proofs, we use the following fact multiple times:

$$
\begin{aligned}
&\mathbb{E}\left(u_{it}u_{is}u_{ju}u_{jv}\right)-\mathbb{E}\left(u_{it}u_{is}\right)\mathbb{E}\left(u_{ju}u_{jv}\right)\\
&\quad=1\{i\neq j\}1\{t=s\}1\{u=v\}+1\{i=j\}1\{\text{two pairs among }\{t,s,u,v\}\}\\
&\qquad-1\{t=s\}1\{u=v\}\\
&\quad=1\{i=j\}1\{\text{two pairs among }\{t,s,u,v\}\}. \qquad\text{(S.119)}
\end{aligned}
$$

1. $\text{Cov}\left(u_{it}u_{is}, u_{ju}u_{jv}\right)$: Notice that

$$\begin{aligned}\text{Cov}\left(g_{1,ii,ts}, g_{2,jj,uv}\right) &= \text{Cov}\left(u_{it}u_{is}, u_{ju}u_{jv}\right) = \mathbb{E}\left(u_{it}u_{is}u_{ju}u_{jv}\right) - \mathbb{E}\left(u_{it}u_{is}\right)\mathbb{E}\left(u_{ju}u_{jv}\right) \\ &= 1\left\{i=j\right\}1\left\{\text{two pairs among } \left\{t,s,u,v\right\}\right\}.\end{aligned}$$

   This implies that out of the six summations over indices (t,s,u,v,i,j), only three summations matter. Therefore, we have

$$\frac{1}{NT^2}\sum_{t,s,u,v}\sum_{i,j}\left|\text{Cov}\left(g_{1,ii,ts}, g_{1,jj,uv}\right)\right| \leq M.$$

2. $\text{Cov}\left(u_{it}u_{is}, u_{ju}v_{jv}\right)$: Notice that

$$\text{Cov}\left(u_{it}u_{is}, u_{ju}v_{jv}\right) = \sum_{k=1}^{3}\text{Cov}\left(u_{it}u_{is}, u_{ju}v_{k,jv}\right).$$

   (a) Notice that

$$\begin{aligned}\text{Cov}\left(u_{it}u_{is}, u_{ju}v_{k,jv}\right) &= \mathbb{E}\left(u_{it}u_{is}u_{ju}v_{1,jv}\right) - \mathbb{E}\left(u_{it}u_{is}\right)\mathbb{E}\left(u_{ju}v_{1,jv}\right) \\ &= \frac{c}{N}\left(\sum_{h=1}^{N}\lambda_j\lambda_h\left\{\mathbb{E}\left(u_{it}u_{is}u_{ju}u_{hv}\right) - \mathbb{E}\left(u_{it}u_{is}\right)\mathbb{E}\left(u_{ju}u_{hv}\right)\right\}\right) \\ &= \frac{c}{N}\lambda_j^2\left\{\mathbb{E}\left(u_{it}u_{is}u_{ju}u_{jv}\right) - \mathbb{E}\left(u_{it}u_{is}\right)\mathbb{E}\left(u_{ju}u_{jv}\right)\right\} \\ &\leq \frac{c}{N}\lambda_j^2 1\left\{i=j\right\}1\left\{\text{two pairs among } \left\{t,s,u,v\right\}\right\}.\end{aligned}$$

   Therefore,

$$\begin{aligned}&\frac{1}{NT^2}\sum_{t,s,u,v}\sum_{i,j}\left|\mathbb{E}\left(u_{it}u_{is}u_{ju}v_{1,jv}\right)\right| \\ &\quad\leq \frac{M}{N^2T^2}\sum_{t,s,u,v}\sum_{i,j}1\left\{i=j\right\}1\left\{\text{two pairs among } \left\{t,s,u,v\right\}\right\} \leq M.\end{aligned}$$

   (b) Similarly, we have

$$\begin{aligned}\text{Cov}\left(u_{it}u_{is}, u_{ju}v_{2,jv}\right) &= \mathbb{E}\left(u_{it}u_{is}u_{ju}v_{2,jv}\right) - \mathbb{E}\left(u_{it}u_{is}\right)\mathbb{E}\left(u_{ju}v_{2,jv}\right) \\ &= \frac{c}{T}\left(\sum_{\tau=1}^{T}\left\{\mathbb{E}\left(u_{it}u_{is}u_{ju}u_{j\tau}\right) - \mathbb{E}\left(u_{it}u_{is}\right)\mathbb{E}\left(u_{ju}u_{j\tau}\right)\right\}f_\tau f_v\right) \\ &= \frac{c}{T}\sum_{\tau=1}^{T}f_\tau f_v 1\left\{i=j\right\}1\left\{\text{two pairs among } \left\{t,s,u,\tau\right\}\right\},\end{aligned}$$

   which leads the desired result

$$\frac{1}{NT^2}\sum_{t,s,u,v}\sum_{i,j}\left|\text{Cov}\left(u_{it}u_{is}, u_{ju}v_{2,jv}\right)\right| \leq M.$$

(c) Notice that

$$
\begin{aligned}
\mathrm{Cov}\left(u_{it}u_{is}u_{ju}v_{3,jv}\right) &= \mathbb{E}\left(u_{it}u_{is}u_{ju}v_{3,jv}\right) - \mathbb{E}\left(u_{it}u_{is}\right)\mathbb{E}\left(u_{ju}v_{3,jv}\right) \\
&= \frac{c^2}{NT}\left(\sum_{h=1}^{N}\sum_{\tau=1}^{T}\lambda_j\lambda_h f_\tau f_v\left[\mathbb{E}\left(u_{it}u_{is}u_{ju}u_{h\tau}\right) - \mathbb{E}\left(u_{it}u_{is}\right)\mathbb{E}\left(u_{ju}u_{h\tau}\right)\right]\right) \\
&= \frac{c^2}{NT}\left(\sum_{\tau=1}^{T}\lambda_j^2 f_\tau f_v\left[\mathbb{E}\left(u_{it}u_{is}u_{ju}u_{j\tau}\right) - \mathbb{E}\left(u_{it}u_{is}\right)\mathbb{E}\left(u_{ju}u_{j\tau}\right)\right]\right) \\
&= \frac{c^2}{NT}\left(\sum_{\tau=1}^{T}\lambda_j^2 f_\tau f_v 1\left\{i=j\right\}1\left\{\text{two pairs among }\ \{t,s,u,\tau\}\right\}\right).
\end{aligned}
$$

Therefore,

$$
\frac{1}{NT^2}\sum_{t,s,u,v}\sum_{i,j}\left|\mathrm{Cov}\left(u_{it}u_{is}u_{ju}v_{1,jv}\right)\right| \le M.
$$

Combining these, we have the desired result.

3. $\mathrm{Cov}\left(u_{it}u_{is}, v_{ju}v_{jv}\right)$: Notice that

$$
\begin{aligned}
&\mathrm{Cov}\left(u_{it}u_{is}, v_{ju}v_{jv}\right) \\
&= \mathbb{E}\left(u_{it}u_{is}v_{ju}v_{jv}\right) - \mathbb{E}\left(u_{it}u_{is}\right)\mathbb{E}\left(v_{ju}v_{jv}\right) \\
&= \mathbb{E}\left(\begin{array}{c} u_{it}u_{is}\left(\frac{\lambda_j}{\sqrt{N}}\left(\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\lambda_k u_{ku}\right) + \left(\frac{1}{\sqrt{T}}\sum_{p=1}^{T}f_p u_{jp}\right)\frac{f_u}{\sqrt{T}} + \tilde{v}\left(\lambda,f,u\right)\frac{\lambda_j}{\sqrt{N}}\frac{f_u}{\sqrt{T}}\right) \\ \times\left(\frac{\lambda_j}{\sqrt{N}}\left(\frac{1}{\sqrt{N}}\sum_{l=1}^{N}\lambda_l u_{lv}\right) + \left(\frac{1}{\sqrt{T}}\sum_{q=1}^{T}f_q u_{jq}\right)\frac{f_v}{\sqrt{T}} + \tilde{v}\left(\lambda,f,u\right)\frac{\lambda_j}{\sqrt{N}}\frac{f_v}{\sqrt{T}}\right)\end{array}\right) \\
&\quad - 1\left\{t=s\right\}\mathbb{E}\left(\begin{array}{c} \left(\frac{\lambda_j}{\sqrt{N}}\left(\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\lambda_k u_{ku}\right) + \left(\frac{1}{\sqrt{T}}\sum_{p=1}^{T}f_p u_{jp}\right)\frac{f_u}{\sqrt{T}} + \tilde{v}\left(\lambda,f,u\right)\frac{\lambda_j}{\sqrt{N}}\frac{f_u}{\sqrt{T}}\right) \\ \times\left(\frac{\lambda_j}{\sqrt{N}}\left(\frac{1}{\sqrt{N}}\sum_{l=1}^{N}\lambda_l u_{lv}\right) + \left(\frac{1}{\sqrt{T}}\sum_{q=1}^{T}f_q u_{jq}\right)\frac{f_v}{\sqrt{T}} + \tilde{v}\left(\lambda,f,u\right)\frac{\lambda_j}{\sqrt{N}}\frac{f_v}{\sqrt{T}}\right)\end{array}\right).
\end{aligned}
$$

Here there are 9 terms in the product.

(a) Notice that

$$
\begin{aligned}
\mathrm{Cov}\left(u_{it}u_{is}, v_{1,ju}v_{1,jv}\right) &= \mathbb{E}\left(u_{it}u_{is}\frac{\lambda_j}{\sqrt{N}}\left(\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\lambda_k u_{ku}\right)\frac{\lambda_j}{\sqrt{N}}\left(\frac{1}{\sqrt{N}}\sum_{l=1}^{N}\lambda_l u_{lv}\right)\right) \\
&\quad - \mathbb{E}\left(u_{it}u_{is}\right)\mathbb{E}\left(\frac{\lambda_j}{\sqrt{N}}\left(\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\lambda_k u_{ku}\right)\frac{\lambda_j}{\sqrt{N}}\left(\frac{1}{\sqrt{N}}\sum_{l=1}^{N}\lambda_l u_{lv}\right)\right) \\
&= \frac{1}{N^2}\sum_{k=1}^{N}\sum_{l=1}^{N}\lambda_j^2\lambda_k\lambda_l\left(\mathbb{E}\left(u_{it}u_{is}u_{ku}u_{lv}\right) - \mathbb{E}\left(u_{it}u_{is}\right)\mathbb{E}\left(u_{ku}u_{lv}\right)\right) \\
&= \frac{1}{N^2}\sum_{k=1}^{N}\lambda_j^2\lambda_k^2\left(\mathbb{E}\left(u_{it}u_{is}u_{ku}u_{kv}\right) - \mathbb{E}\left(u_{it}u_{is}\right)\mathbb{E}\left(u_{ku}u_{kv}\right)\right) \\
&= \frac{1}{N^2}\sum_{k=1}^{N}\lambda_j^2\lambda_k^2\left\{1\left\{i=k\right\}1\left\{\text{two pairs among }\ \{t,s,u,v\}\right\}\right\}.
\end{aligned}
$$

Therefore,

$$\frac{1}{NT^2}\sum_{t,s,u,v}\sum_{i,j}\left|\mathrm{Cov}\left(u_{it}u_{is},v_{1,ju}v_{1,jv}\right)\right|\leq M.$$

(b) Notice that

$$\begin{aligned}
\mathrm{Cov}\left(u_{it}u_{is},v_{2,ju}v_{2,jv}\right) &= \mathbb{E}\left(u_{it}u_{is}\left(\frac{1}{\sqrt{T}}\sum_{p=1}^{T}f_pu_{jp}\right)\frac{f_u}{\sqrt{T}}\left(\frac{1}{\sqrt{T}}\sum_{q=1}^{T}f_qu_{jq}\right)\frac{f_v}{\sqrt{T}}\right)\\
&\quad-\mathbb{E}\left(u_{it}u_{is}\right)\mathbb{E}\left(\left(\frac{1}{\sqrt{T}}\sum_{p=1}^{T}f_pu_{jp}\right)\frac{f_u}{\sqrt{T}}\left(\frac{1}{\sqrt{T}}\sum_{q=1}^{T}f_qu_{jq}\right)\frac{f_v}{\sqrt{T}}\right)\\
&=\frac{1}{T^2}\sum_{p=1}^{T}\sum_{q=1}^{T}f_pf_qf_uf_v\left[\mathbb{E}\left(u_{it}u_{is}u_{jp}u_{jq}\right)-\mathbb{E}\left(u_{it}u_{is}\right)\mathbb{E}\left(u_{jp}u_{jq}\right)\right]\\
&=\frac{1}{T^2}\sum_{p=1}^{T}\sum_{q=1}^{T}f_pf_qf_uf_v\left[1\left\{i=j\right\}1\left\{\text{two pairs among }\ \{t,s,p,q\}\right\}\right].
\end{aligned}$$

Therefore,

$$\frac{1}{NT^2}\sum_{t,s,u,v}\sum_{i,j}\left|\mathrm{Cov}\left(u_{it}u_{is},v_{2,ju}v_{2,jv}\right)\right|\leq M.$$

(c) Notice that

$$\begin{aligned}
&\mathrm{Cov}\left(u_{it}u_{is},v_{3,ju}v_{3,jv}\right)\\
&=\mathbb{E}\left(u_{it}u_{is}\left(\frac{1}{\sqrt{NT}}\sum_{k=1}^{N}\sum_{p=1}^{T}\lambda_kf_pu_{kp}\right)\frac{\lambda_jf_u}{\sqrt{NT}}\left(\frac{1}{\sqrt{NT}}\sum_{l=1}^{N}\sum_{q=1}^{T}\lambda_lf_qu_{lq}\right)\frac{f_v\lambda_j}{\sqrt{NT}}\right)\\
&\quad-\mathbb{E}\left(u_{it}u_{is}\right)\mathbb{E}\left(\left(\frac{1}{\sqrt{NT}}\sum_{k=1}^{N}\sum_{p=1}^{T}\lambda_kf_pu_{kp}\right)\frac{\lambda_jf_u}{\sqrt{NT}}\left(\frac{1}{\sqrt{NT}}\sum_{l=1}^{N}\sum_{q=1}^{T}\lambda_lf_qu_{lq}\right)\frac{f_v\lambda_j}{\sqrt{NT}}\right)\\
&=\frac{1}{N^2T^2}\sum_{k=1}^{N}\sum_{p=1}^{T}\sum_{l=1}^{N}\sum_{q=1}^{T}\lambda_k\lambda_j^2\lambda_lf_pf_uf_qf_v\left[\mathbb{E}\left(u_{it}u_{is}u_{kp}u_{lq}\right)-\mathbb{E}\left(u_{it}u_{is}\right)\mathbb{E}\left(u_{kp}u_{lq}\right)\right]\\
&=\frac{1}{N^2T^2}\sum_{k=1}^{N}\sum_{p=1}^{T}\sum_{q=1}^{T}\lambda_k^2\lambda_j^2\mathbb{E}f_pf_uf_qf_v\left[\mathbb{E}\left(u_{it}u_{is}u_{kp}u_{kq}\right)-\mathbb{E}\left(u_{it}u_{is}\right)\mathbb{E}\left(u_{kp}u_{kq}\right)\right]\\
&=\frac{1}{N^2T^2}\sum_{k=1}^{N}\sum_{p=1}^{T}\sum_{q=1}^{T}\lambda_k^2\lambda_j^2\mathbb{E}f_pf_uf_qf_v\left[1\left\{i=k\right\}1\left\{\text{two pairs among }\ \{t,s,p,q\}\right\}\right].
\end{aligned}$$

Then,

$$\frac{1}{NT^2}\sum_{t,s,u,v}\sum_{i,j}\left|\mathrm{Cov}\left(u_{it}u_{is},v_{2,ju}v_{2,jv}\right)\right|\leq M.$$

(d) The desired result follows similarly since

$$
\begin{aligned}
\operatorname{Cov}\left(u_{it}u_{is}, v_{1,ju}v_{2,jv}\right) &= \mathbb{E}\left(u_{it}u_{is}\frac{\lambda_j}{\sqrt{N}}\left(\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\lambda_k u_{ku}\right)\left(\frac{1}{\sqrt{T}}\sum_{q=1}^{T}f_q u_{jq}\right)\frac{f_v}{\sqrt{T}}\right) \\
&\quad - \mathbb{E}\left(u_{it}u_{is}\right)\mathbb{E}\left(\frac{\lambda_j}{\sqrt{N}}\left(\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\lambda_k u_{ku}\right)\left(\frac{1}{\sqrt{T}}\sum_{q=1}^{T}f_q u_{jq}\right)\frac{f_v}{\sqrt{T}}\right) \\
&= \frac{1}{NT}\sum_{k=1}^{N}\sum_{q=1}^{T}\lambda_j\lambda_k f_q f_v\left\{\mathbb{E}\left(u_{it}u_{is}u_{ku}u_{jq}\right) - \mathbb{E}\left(u_{it}u_{is}\right)\mathbb{E}\left(u_{ku}u_{jq}\right)\right\} \\
&= \frac{1}{NT}\sum_{q=1}^{T}\lambda_j^2 f_q f_v\left\{\mathbb{E}\left(u_{it}u_{is}u_{ju}u_{jq}\right) - \mathbb{E}\left(u_{it}u_{is}\right)\mathbb{E}\left(u_{ju}u_{jq}\right)\right\} \\
&= \frac{1}{NT}\sum_{q=1}^{T}\lambda_j^2 f_q f_v\left[1\left\{i=j\right\}1\left\{\text{two pairs among }\ \{t,s,u,q\}\right\}\right].
\end{aligned}
$$

(e) The desired result follows since

$$
\begin{aligned}
&\operatorname{Cov}\left(u_{it}u_{is}, v_{1,ju}v_{3,jv}\right) \\
&= \mathbb{E}\left(u_{it}u_{is}\frac{\lambda_j}{\sqrt{N}}\left(\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\lambda_k u_{ku}\right)\left(\frac{1}{\sqrt{NT}}\sum_{l=1}^{N}\sum_{q=1}^{T}\lambda_l f_q u_{lq}\right)\frac{\lambda_j f_v}{\sqrt{NT}}\right) \\
&\quad - \mathbb{E}\left(u_{it}u_{is}\right)\mathbb{E}\left(\frac{\lambda_j}{\sqrt{N}}\left(\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\lambda_k u_{ku}\right)\left(\frac{1}{\sqrt{NT}}\sum_{l=1}^{N}\sum_{q=1}^{T}\lambda_l f_q u_{lq}\right)\frac{\lambda_j f_v}{\sqrt{NT}}\right) \\
&= \frac{1}{N^2T}\sum_{k=1}^{N}\sum_{l=1}^{N}\sum_{q=1}^{T}\mathbb{E}\left(\lambda_j^2\lambda_k\lambda_l\right)\mathbb{E}\left(f_q f_v\right)\left\{\mathbb{E}\left(u_{it}u_{is}u_{ku}u_{lq}\right) - \mathbb{E}\left(u_{it}u_{is}\right)\mathbb{E}\left(u_{ku}u_{lq}\right)\right\} \\
&= \text{it should be [q=v and k=l]} \\
&= \frac{1}{N^2T}\sum_{k=1}^{N}\mathbb{E}\left(\lambda_j^2\lambda_k^2\right)\mathbb{E}\left(f_v^2\right)\left\{\mathbb{E}\left(u_{it}u_{is}u_{ku}u_{kv}\right) - \mathbb{E}\left(u_{it}u_{is}\right)\mathbb{E}\left(u_{ku}u_{kv}\right)\right\} \\
&= \text{it should be that } i=k \\
&= \frac{1}{N^2T}\mathbb{E}\left(\lambda_j^2\lambda_i^2\right)\mathbb{E}\left(f_v^2\right)\left\{\mathbb{E}\left(u_{it}u_{is}u_{iu}u_{iv}\right) - \mathbb{E}\left(u_{it}u_{is}\right)\mathbb{E}\left(u_{iu}u_{iv}\right)\right\} \\
&\leq \frac{1}{N^2T}\mathbb{E}\left(\lambda_j^2\lambda_i^2\right)1\left\{\text{two pairs among }\ \{t,s,u,v\}\right\}.
\end{aligned}
$$

(f) Similarly, the desired result follows since

$$
\begin{aligned}
&\operatorname{Cov}\left(u_{it}u_{is}, v_{2,ju}v_{3,jv}\right)\\
&= \mathbb{E}\left(u_{it}u_{is}\left(\frac{1}{\sqrt{T}}\sum_{p=1}^{T} f_p u_{jp}\right)\frac{f_u}{\sqrt{T}}\left(\frac{1}{\sqrt{NT}}\sum_{l=1}^{N}\sum_{q=1}^{T}\lambda_l f_q u_{lq}\right)\frac{\lambda_j f_v}{\sqrt{NT}}\right)\\
&\quad - \mathbb{E}\left(u_{it}u_{is}\right)\mathbb{E}\left(\left(\frac{1}{\sqrt{T}}\sum_{p=1}^{T} f_p u_{jp}\right)\frac{f_u}{\sqrt{T}}\left(\frac{1}{\sqrt{NT}}\sum_{l=1}^{N}\sum_{q=1}^{T}\lambda_l f_q u_{lq}\right)\frac{\lambda_j f_v}{\sqrt{NT}}\right)\\
&= \frac{1}{NT^2}\sum_{l=1}^{N}\sum_{p=1}^{T}\sum_{q=1}^{T}\lambda_j\lambda_l f_p f_q f_u f_v\left\{\mathbb{E}\left(u_{it}u_{is}u_{jp}u_{lq}\right)-\mathbb{E}\left(u_{it}u_{is}\right)\mathbb{E}\left(u_{jp}u_{lq}\right)\right\}\\
&= \frac{1}{NT^2}\sum_{p=1}^{T}\sum_{q=1}^{T}\lambda_j^2 f_p f_q f_u f_v\left\{\mathbb{E}\left(u_{it}u_{is}u_{jp}u_{jq}\right)-\mathbb{E}\left(u_{it}u_{is}\right)\mathbb{E}\left(u_{jp}u_{jq}\right)\right\}\\
&= \frac{1}{NT^2}\sum_{p=1}^{T}\sum_{q=1}^{T}\lambda_j^2 f_p f_q f_u f_v\left[1\left\{i=j\right\}1\left\{\text{two pairs among }\left\{t,s,u,q\right\}\right\}\right].
\end{aligned}
$$

4. $\operatorname{Cov}\left(u_{it}v_{is}, u_{ju}v_{jv}\right)$: Notice that

$$
\begin{aligned}
&\operatorname{Cov}\left(u_{it}v_{is}, u_{ju}v_{jv}\right)\\
&= \mathbb{E}\left(u_{it}v_{is}u_{ju}v_{jv}\right)-\mathbb{E}\left(u_{it}v_{is}\right)\mathbb{E}\left(u_{ju}v_{jv}\right)\\
&= \mathbb{E}\left(\begin{array}{c} u_{it}\left(\frac{\lambda_i}{\sqrt{N}}\left(\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\lambda_k u_{ks}\right)+\left(\frac{1}{\sqrt{T}}\sum_{p=1}^{T} f_p u_{ip}\right)\frac{f_s}{\sqrt{T}}+\tilde{v}\left(\lambda,f,u\right)\frac{\lambda_i}{\sqrt{N}}\frac{f_s}{\sqrt{T}}\right)\\ \times u_{ju}\left(\frac{\lambda_j}{\sqrt{N}}\left(\frac{1}{\sqrt{N}}\sum_{l=1}^{N}\lambda_l u_{lv}\right)+\left(\frac{1}{\sqrt{T}}\sum_{q=1}^{T} f_q u_{jq}\right)\frac{f_v}{\sqrt{T}}+\tilde{v}\left(\lambda,f,u\right)\frac{\lambda_j}{\sqrt{N}}\frac{f_v}{\sqrt{T}}\right)\end{array}\right)\\
&\quad - \mathbb{E}\left(u_{it}v_{is}\right)\mathbb{E}\left(u_{ju}v_{jv}\right).
\end{aligned}
$$

(a) The desired result follows since

$$
\begin{aligned}
&\operatorname{Cov}\left(u_{it}v_{1,is}, u_{ju}v_{1,jv}\right)\\
&= \mathbb{E}\left(u_{it}v_{1,is}u_{ju}v_{1,jv}\right)-\mathbb{E}\left(u_{it}v_{1,is}\right)\mathbb{E}\left(u_{ju}v_{1,jv}\right)\\
&= \mathbb{E}\left(u_{it}u_{ju}\frac{\lambda_i}{\sqrt{N}}\left(\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\lambda_k u_{ks}\right)\frac{\lambda_j}{\sqrt{N}}\left(\frac{1}{\sqrt{N}}\sum_{l=1}^{N}\lambda_l u_{lv}\right)\right)\\
&\quad - \mathbb{E}\left(u_{it}\frac{\lambda_i}{\sqrt{N}}\left(\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\lambda_k u_{ks}\right)\right)\mathbb{E}\left(u_{ju}\frac{\lambda_j}{\sqrt{N}}\left(\frac{1}{\sqrt{N}}\sum_{l=1}^{N}\lambda_l u_{lv}\right)\right)\\
&= \frac{1}{N^2}\sum_{k=1}^{N}\sum_{l=1}^{N}\lambda_i\lambda_k\lambda_j\lambda_l\left\{\mathbb{E}\left(u_{it}u_{ju}u_{ks}u_{lv}\right)-\mathbb{E}\left(u_{it}u_{ks}\right)\mathbb{E}\left(u_{ju}u_{lv}\right)\right\}.
\end{aligned}
$$

So,

$$
\begin{aligned}
&\frac{1}{NT^2}\sum_{t,s,u,v}\sum_{i,j}\left|\operatorname{Cov}\left(u_{it}v_{1,is},u_{ju}v_{1,jv}\right)\right| \\
&=\frac{1}{NT^2}\sum_{t,s,u,v}\sum_{i,j}\frac{1}{N^2}\sum_{k=1}^{N}\sum_{l=1}^{N}\left[\lambda_i\lambda_k\lambda_j\lambda_l\left\{\begin{array}{c}\mathbb{E}\left(u_{it}u_{ju}u_{ks}u_{lv}\right)\\ -\mathbb{E}\left(u_{it}u_{ks}\right)\mathbb{E}\left(u_{ju}u_{lv}\right)\end{array}\right\}\right] \\
&\leq\frac{M}{N^3T^2}\sum_{t,s,u,v}\sum_{i,j,k,l}1\left\{\text{two pairs among }\ \{i,j,k,l\}\right\}1\left\{\text{two pairs among }\ \{t,s,u,v\}\right\} \\
&\leq M.
\end{aligned}
$$

(b) Also, we have

$$
\begin{aligned}
&\operatorname{Cov}\left(u_{it}v_{2,is},u_{ju}v_{2,jv}\right) \\
&=\mathbb{E}\left(u_{it}v_{2,is}u_{ju}v_{2,jv}\right)-\mathbb{E}\left(u_{it}v_{2,is}\right)\mathbb{E}\left(u_{ju}v_{2,jv}\right) \\
&=\mathbb{E}\left(u_{it}u_{ju}\left(\frac{1}{\sqrt{T}}\sum_{p=1}^{T}f_pu_{ip}\right)\frac{f_s}{\sqrt{T}}\left(\frac{1}{\sqrt{T}}\sum_{q=1}^{T}f_qu_{jq}\right)\frac{f_v}{\sqrt{T}}\right) \\
&\quad-\mathbb{E}\left(u_{it}\left(\frac{1}{\sqrt{T}}\sum_{p=1}^{T}f_pu_{ip}\right)\frac{f_s}{\sqrt{T}}\right)\mathbb{E}\left(u_{ju}\left(\frac{1}{\sqrt{T}}\sum_{q=1}^{T}f_qu_{jq}\right)\frac{f_v}{\sqrt{T}}\right) \\
&=\frac{1}{T^2}\sum_{p=1}^{T}\sum_{q=1}^{T}f_pf_sf_qf_v\left\{\mathbb{E}\left(u_{it}u_{ip}u_{jq}u_{ju}\right)-\mathbb{E}\left(u_{it}u_{ip}\right)\mathbb{E}\left(u_{ju}u_{jq}\right)\right\} \\
&=\frac{1}{T^2}\sum_{p=1}^{T}\sum_{q=1}^{T}f_pf_sf_qf_v\left\{1\left\{i=j\right\}1\left\{\text{two pairs among }\ \{t,p,q,u\}\right\}\right\}.
\end{aligned}
$$

So,

$$
\frac{1}{NT^2}\sum_{t,s,u,v}\sum_{i,j}\left|\operatorname{Cov}\left(u_{it}v_{2,is},u_{ju}v_{2,jv}\right)\right|\leq M.
$$

(c) We can show the rest of the cases analogously.

5. $\operatorname{Cov}\left(u_{it}v_{is},v_{ju}v_{jv}\right)$: There are 4 kinds, (i) # of $v_{3,\cdot\cdot}=0$, (ii) # of $v_{3,\cdot\cdot}=1$, (iii) # of $v_{3,\cdot\cdot}=2$, and (iv) # of $v_{3,\cdot\cdot}=4$.

   (a) When # of $v_{3,\cdot\cdot}=0$: For example, $\operatorname{Cov}\left(u_{it}v_{1,is},v_{1,ju}v_{1,jv}\right)$. The desired result follows since

$$
\begin{aligned}
&\frac{1}{NT^2}\sum_{t,s,u,v}\sum_{i,j}\left|\mathbb{E}\left(u_{it}v_{1,is}v_{1,ju}v_{2,jv}\right)-\mathbb{E}\left(u_{it}v_{1,is}\right)\mathbb{E}\left(v_{1,ju}v_{2,jv}\right)\right| \\
&\leq\frac{M}{NT^2}\sum_{t,s,u,v}\sum_{i,j}\frac{1}{N^2T}\sum_{i^*,j^*}\sum_{v^*}\left(\mathbb{E}\left(u_{it}u_{i^*s}u_{j^*u}u_{jv^*}\right)-\mathbb{E}\left(u_{it}u_{i^*s}\right)\mathbb{E}\left(u_{j^*u}u_{jv^*}\right)\right) \\
&\leq\frac{M}{N^3T^3}\sum_{t,s,u,v,v^*}\sum_{i,j,i^*,j^*}1\left\{\text{two pairs among }\ \{i,i^*,j,j^*\}\right\}1\left\{\text{two pairs among }\ \{t,s,u,v^*\}\right\} \\
&\leq M.
\end{aligned}
$$

(b) When # of $v_{3,\cdot\cdot} = 1$: For example, $\mathrm{Cov}\left(u_{it}v_{3,is}, v_{1,ju}v_{2,jv}\right)$. The desired result follows since

$$\begin{aligned}
&\frac{1}{NT^2}\sum_{t,s,u,v}\sum_{i,j}\left|\mathrm{Cov}\left(u_{it}v_{3,is}, v_{1,ju}v_{2,jv}\right)\right| \\
&= \frac{1}{NT^2}\sum_{t,s,u,v}\sum_{i,j}\left|\mathbb{E}\left(u_{it}v_{3,is}v_{1,ju}v_{2,jv}\right) - \mathbb{E}\left(u_{it}v_{3,is}\right)\mathbb{E}\left(v_{1,ju}v_{2,jv}\right)\right| \\
&\leq \frac{M}{NT^2}\frac{1}{N^2T^2}\sum_{t,s,u,v}\sum_{i,j}\sum_{i^*,j^*}\sum_{s^*,v^*}\left\{\mathbb{E}\left(u_{it}u_{i^*s^*}u_{j^*u}u_{jv^*}\right) - \mathbb{E}\left(u_{it}u_{i^*s^*}\right)\mathbb{E}\left(u_{j^*u}u_{jv^*}\right)\right\} \\
&= \frac{M}{N^3T^4}\sum_{t,s,u,v}\sum_{i,j}\sum_{i^*,j^*}\sum_{s^*,v^*} 1\left\{\text{two pairs among } \left\{i,i^*,j,j^*\right\}\right\} 1\left\{\text{two pairs among } \left\{t,s^*,u,v^*\right\}\right\} \\
&\leq M.
\end{aligned}$$

(c) When # of $v_{3,\cdot\cdot} = 2$: For example, $\mathrm{Cov}\left(u_{it}v_{3,is}, v_{3,ju}v_{2,jv}\right)$. The desired result follows since

$$\begin{aligned}
&\frac{1}{NT^2}\sum_{t,s,u,v}\sum_{i,j}\left|\mathrm{Cov}\left(u_{it}v_{3,is}, v_{3,ju}v_{2,jv}\right)\right| \\
&= \frac{1}{NT^2}\sum_{t,s,u,v}\sum_{i,j}\left|\mathbb{E}\left(u_{it}v_{3,is}v_{1,ju}v_{2,jv}\right) - \mathbb{E}\left(u_{it}v_{3,is}\right)\mathbb{E}\left(v_{3,ju}v_{2,jv}\right)\right| \\
&\leq \frac{M}{NT^2}\sum_{t,s,u,v}\sum_{i,j}\frac{1}{N^2T^3}\sum_{i^*,j^*}\sum_{u^*,s^*,v^*}\left(\mathbb{E}\left(u_{it}u_{i^*s^*}u_{j^*u^*}u_{jv^*}\right) - \mathbb{E}\left(u_{it}u_{i^*s^*}\right)\mathbb{E}\left(u_{j^*u^*}u_{jv^*}\right)\right) \\
&= \frac{M}{N^3T^5}\sum_{t,s,u,v}\sum_{i,j}\sum_{i^*,j^*}\sum_{u^*,s^*,v^*} 1\left\{\text{two pairs among } \left\{i,i^*,j,j^*\right\}\right\} 1\left\{\text{two pairs among } \left\{t,s^*,u^*,v^*\right\}\right\} \\
&\leq M.
\end{aligned}$$

(d) For the other cases, notice that an additional $v_{3,\cdot\cdot}$ term adds an extra summation. However, it also increases the order of the denominator by one. Therefore, the required result follows analogously.

6. $\mathrm{Cov}\left(v_{it}v_{is}, v_{ju}v_{jv}\right)$: Follows analogously to the previous cases.

References

Ahn, S. C., Lee, Y. H., and Schmidt, P. (2001). GMM estimation of linear panel data models with time-varying individual effects. *Journal of Econometrics*, 101(2):219–255.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77(4):1229–1279.

Geman, S. (1980). A limit theorem for the norm of random matrices. *Annals of Probability*, 8(2):252–261.

Johnstone, I. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Annals of Statistics*, 29(2):295–327.

Kato, T. (1980). *Perturbation Theory for Linear Operators*. Springer-Verlag.

Moon, H. and Weidner, M. (2014). Dynamic Linear Panel Regression Models with Interactive Fixed Effects. *CeMMAP working paper series*.

Soshnikov, A. (2002). A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. *Journal of Statistical Physics*, 108(5):1033–1056.

	k = 1			k = 2			k = 3			k = 4		
	BIAS	SD	SIZE	BIAS	SD	SIZE	BIAS	SD	SIZE	BIAS	SD	SIZE
R = 0	-0.032	0.042	0.003	-0.176	0.055	0.604	-0.214	0.067	0.561	-0.161	0.081	0.175
R = 1	-0.022	0.043	0.013	-0.078	0.056	0.144	-0.100	0.068	0.157	-0.098	0.084	0.088
R = 2	0.001	0.050	0.011	-0.055	0.074	0.105	-0.091	0.094	0.147	-0.125	0.115	0.158
R = 3	0.044	0.056	0.151	0.035	0.086	0.121	0.037	0.111	0.101	0.045	0.139	0.095
R = 4	0.003	0.054	0.100	0.002	0.079	0.107	0.004	0.098	0.084	0.003	0.120	0.073
R = 5	0.002	0.058	0.121	0.002	0.084	0.127	0.004	0.103	0.099	0.004	0.122	0.081
R = 6	0.001	0.061	0.138	0.001	0.089	0.143	0.003	0.107	0.112	0.003	0.125	0.091
R = 7	0.001	0.064	0.152	0.000	0.092	0.156	0.002	0.110	0.119	0.003	0.127	0.092
R = 8	0.001	0.067	0.165	-0.001	0.096	0.162	0.002	0.112	0.128	0.002	0.126	0.094
R = 9	0.000	0.070	0.180	-0.001	0.098	0.167	0.001	0.114	0.131	0.001	0.126	0.092

	k = 5			k = 6			k = 7			k = 8		
	BIAS	SD	SIZE	BIAS	SD	SIZE	BIAS	SD	SIZE	BIAS	SD	SIZE
R = 0	-0.124	0.098	0.039	-0.115	0.117	0.031	-0.054	0.141	0.008	0.004	0.168	0.004
R = 1	-0.018	0.102	0.015	0.049	0.122	0.019	0.177	0.147	0.104	0.291	0.175	0.192
R = 2	-0.100	0.132	0.079	-0.089	0.148	0.055	-0.024	0.168	0.026	0.023	0.191	0.021
R = 3	0.056	0.164	0.094	0.057	0.179	0.081	0.083	0.200	0.094	0.092	0.220	0.083
R = 4	0.002	0.137	0.061	-0.001	0.152	0.052	0.001	0.170	0.050	-0.001	0.191	0.041
R = 5	0.003	0.137	0.063	0.000	0.151	0.053	0.001	0.170	0.051	-0.001	0.190	0.044
R = 6	0.002	0.139	0.068	-0.001	0.152	0.056	0.000	0.169	0.052	-0.002	0.189	0.045
R = 7	0.002	0.139	0.071	-0.001	0.151	0.059	0.000	0.168	0.051	-0.002	0.186	0.043
R = 8	0.002	0.137	0.071	0.000	0.148	0.056	-0.001	0.164	0.052	-0.003	0.181	0.041
R = 9	0.001	0.136	0.067	0.000	0.146	0.052	-0.001	0.160	0.047	-0.002	0.175	0.035

Table 1: Results of the Empirical Monte Carlo Simulation. Bias and standard deviation (SD) of $\widehat{\beta}^{\mathrm{BC}}_{R,k}$ are reported for the regressor $k = 1, \ldots, 8$ using $R = 0, \ldots, 9$ factors in the estimation procedure. We also report the empirical size of a 5% nominal size t-test of the hypothesis $H_0 : \beta_k = \beta_k^0$. Results are based on $10,000$ repetitions.

beta = 0.2

N=300	T=30		T=100		T=300		T=1000	
R	Bias	SD	Bias	SD	Bias	SD	Bias	SD
0	0.3443	0.0960	0.3569	0.0532	0.3605	0.0304	0.3617	0.0171
1	0.2044	0.1054	0.2295	0.0574	0.2362	0.0339	0.2384	0.0198
2	-0.0376	0.0157	-0.0109	0.0061	-0.0036	0.0033	-0.0011	0.0018
3	-0.0589	0.0250	-0.0119	0.0066	-0.0037	0.0034	-0.0011	0.0018
4	-0.0996	0.0354	-0.0131	0.0073	-0.0037	0.0035	-0.0011	0.0018
5	-0.1631	0.0356	-0.0143	0.0081	-0.0039	0.0036	-0.0011	0.0018

beta = 0.5

N=300	T=30		T=100		T=300		T=1000	
R	Bias	SD	Bias	SD	Bias	SD	Bias	SD
0	0.2439	0.0708	0.2554	0.0380	0.2591	0.0216	0.2604	0.0120
1	0.1551	0.0863	0.1780	0.0457	0.1841	0.0265	0.1862	0.0152
2	-0.0378	0.0178	-0.0104	0.0057	-0.0034	0.0030	-0.0010	0.0016
3	-0.0757	0.0351	-0.0115	0.0063	-0.0035	0.0031	-0.0010	0.0016
4	-0.1632	0.0415	-0.0127	0.0071	-0.0035	0.0031	-0.0010	0.0016
5	-0.2644	0.0275	-0.0140	0.0079	-0.0036	0.0032	-0.0010	0.0016

beta = 0.8

N=300	T=30		T=100		T=300		T=1000	
R	Bias	SD	Bias	SD	Bias	SD	Bias	SD
0	0.1030	0.0439	0.1115	0.0215	0.1146	0.0119	0.1157	0.0064
1	0.0656	0.0552	0.0826	0.0272	0.0873	0.0154	0.0892	0.0086
2	-0.0277	0.0209	-0.0065	0.0044	-0.0021	0.0021	-0.0006	0.0011
3	-0.1281	0.0482	-0.0073	0.0051	-0.0021	0.0022	-0.0006	0.0011
4	-0.2727	0.0409	-0.0083	0.0060	-0.0022	0.0022	-0.0006	0.0011
5	-0.4075	0.0368	-0.0097	0.0076	-0.0023	0.0023	-0.0006	0.0011

Table 2: For $N = 300$ and different combinations of T and true parameter β^0 we report the bias and standard deviation of the estimator $\widehat{\beta}_R$, for $R = 0, 1, \ldots, 5$, based on simulations with $10,000$ repetition of design (S.117), where the true number of factors is $R^0 = 2$.

N=300, T=300, beta = 0.2

R	0.03	0.05	0.10	0.25	0.50	0.75	0.90	0.95	0.98
2	-3.04	-2.72	-2.35	-1.74	-1.07	-0.38	0.21	0.57	0.87
3	-3.08	-2.78	-2.40	-1.79	-1.10	-0.40	0.20	0.60	0.90
4	-3.16	-2.86	-2.48	-1.84	-1.12	-0.40	0.22	0.60	0.94
5	-3.25	-2.95	-2.55	-1.89	-1.16	-0.42	0.22	0.59	0.95

N=300, T=1000, beta = 0.2

R	0.03	0.05	0.10	0.25	0.50	0.75	0.90	0.95	0.98
2	-2.49	-2.19	-1.84	-1.25	-0.59	0.06	0.67	1.01	1.32
3	-2.53	-2.21	-1.84	-1.26	-0.59	0.08	0.69	1.03	1.35
4	-2.54	-2.24	-1.87	-1.27	-0.60	0.09	0.69	1.04	1.35
5	-2.58	-2.25	-1.90	-1.28	-0.61	0.10	0.70	1.08	1.37

N=300, T=300, beta = 0.5

R	0.03	0.05	0.10	0.25	0.50	0.75	0.90	0.95	0.98
2	-2.75	-2.48	-2.16	-1.62	-1.02	-0.40	0.15	0.48	0.74
3	-2.82	-2.54	-2.22	-1.65	-1.04	-0.42	0.15	0.48	0.77
4	-2.91	-2.61	-2.28	-1.71	-1.06	-0.42	0.16	0.49	0.77
5	-2.97	-2.68	-2.34	-1.76	-1.09	-0.44	0.14	0.50	0.79

N=300, T=1000, beta = 0.5

R	0.03	0.05	0.10	0.25	0.50	0.75	0.90	0.95	0.98
2	-2.23	-1.98	-1.67	-1.13	-0.56	0.03	0.56	0.87	1.13
3	-2.28	-2.00	-1.67	-1.15	-0.55	0.04	0.57	0.88	1.15
4	-2.27	-2.02	-1.70	-1.16	-0.56	0.04	0.58	0.89	1.16
5	-2.29	-2.02	-1.71	-1.17	-0.57	0.05	0.59	0.90	1.17

N=300, T=300, beta = 0.8

R	0.03	0.05	0.10	0.25	0.50	0.75	0.90	0.95	0.98
2	-1.85	-1.66	-1.44	-1.05	-0.61	-0.19	0.18	0.43	0.64
3	-1.92	-1.72	-1.47	-1.08	-0.64	-0.19	0.18	0.43	0.64
4	-1.98	-1.76	-1.52	-1.10	-0.65	-0.20	0.19	0.42	0.66
5	-2.02	-1.80	-1.56	-1.14	-0.67	-0.21	0.19	0.44	0.66

N=300, T=1000, beta = 0.8

R	0.03	0.05	0.10	0.25	0.50	0.75	0.90	0.95	0.98
2	-1.51	-1.33	-1.10	-0.74	-0.34	0.07	0.45	0.67	0.86
3	-1.52	-1.35	-1.12	-0.75	-0.34	0.07	0.46	0.69	0.87
4	-1.53	-1.36	-1.13	-0.76	-0.34	0.07	0.47	0.69	0.88
5	-1.54	-1.36	-1.15	-0.77	-0.35	0.07	0.47	0.69	0.89

Table 3: For $N = 300$ and different combinations of T and true parameter β^0 we report certain quantiles of the distribution of $\sqrt{NT}(\widehat{\beta}_R - \beta^0)$, for $R = 2, 3, 4, 5$, based on simulations with $10,000$ repetition of design (S.117), where the true number of factors is $R^0 = 2$.

R	beta = 0.2, N=300 T=30	T=100	T=300	T=1000	beta = 0.5, N=300 T=30	T=100	T=300	T=1000	beta = 0.8, N=300 T=30	T=100	T=300	T=1000
2	0.114	0.058	0.051	0.048	0.299	0.096	0.061	0.051	0.452	0.156	0.076	0.054
3	0.346	0.066	0.053	0.050	0.707	0.124	0.065	0.052	0.953	0.210	0.084	0.056
4	0.721	0.084	0.057	0.049	0.966	0.160	0.069	0.052	1.000	0.270	0.091	0.059
5	0.963	0.102	0.059	0.051	1.000	0.205	0.077	0.053	1.000	0.335	0.098	0.058

Table 4: The empirical size of a t-test with 5% nominal size is reported for $N = 300$ and different combinations of T, R and true parameter β^0, based on $10,000$ repetition of design (S.117). A bias corrected estimator for β is used to calculate the test statistics, and we allow for predetermined regressors and heteroscedastic errors when estimating bias and standard deviation. Results for $R = 0, 1$ are not reported since those have size=1 due to misspecification.