Supplementary Material

Dynamic Linear Panel Regression Models with Interactive Fixed Effects

Hyungsik Roger Moon[‡] Martin Weidner[§]

December 22, 2014

S.1 Proof of Identification (Theorem 2.1)

Proof of Theorem 2.1. Let $Q(\beta, \lambda, f) \equiv \mathbb{E}\left(\| Y - \beta \cdot X - \lambda f' \|_F^2 \,\middle|\, \lambda^0, f^0, w \right)$, where $\beta \in \mathbb{R}^K$, $\lambda \in \mathbb{R}^{N \times R}$ and $f \in \mathbb{R}^{T \times R}$. We have

$$
\begin{aligned}
& Q(\beta, \lambda, f) \\
&= \mathbb{E}\left\{ \mathrm{Tr}\left[(Y - \beta \cdot X - \lambda f')' (Y - \beta \cdot X - \lambda f') \right] \,\middle|\, \lambda^0, f^0, w \right\} \\
&= \mathbb{E}\left\{ \mathrm{Tr}\left[\left(\lambda^0 f^{0\prime} - \lambda f' - (\beta - \beta^0) \cdot X + e \right)' \left(\lambda^0 f^{0\prime} - \lambda f' - (\beta - \beta^0) \cdot X + e \right) \right] \,\middle|\, \lambda^0, f^0, w \right\} \\
&= \mathbb{E}\left[\mathrm{Tr}\,(e'e) \,\middle|\, \lambda^0, f^0, w \right] \\
&\qquad + \underbrace{\mathbb{E}\left\{ \mathrm{Tr}\left[\left(\lambda^0 f^{0\prime} - \lambda f' - (\beta - \beta^0) \cdot X \right)' \left(\lambda^0 f^{0\prime} - \lambda f' - (\beta - \beta^0) \cdot X \right) \right] \,\middle|\, \lambda^0, f^0, w \right\}}_{\equiv Q^*(\beta, \lambda, f)} .
\end{aligned}
$$

In the last step we used Assumption ID(ii). Since $\mathbb{E}\left[\mathrm{Tr}\,(e'e) \,\middle|\, \lambda^0, f^0, w \right]$ is independent of β, λ, f, we find that minimizing $Q(\beta, \lambda, f)$ is equivalent to minimizing $Q^*(\beta, \lambda, f)$. We decompose

[‡]Department of Economics, University of Southern California, Los Angeles, CA 90089-0253. Email: moonr@usc.edu. Department of Economics, Yonsei University, Seoul, Korea.

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

$Q^*(\beta,\lambda,f)$ as follows

$$
\begin{aligned}
&Q^*(\beta,\lambda,f)\\
&=\mathbb{E}\left\{\mathrm{Tr}\left[\left(\lambda^0 f^{0\prime}-\lambda f'-(\beta-\beta^0)\cdot X\right)'\left(\lambda^0 f^{0\prime}-\lambda f'-(\beta-\beta^0)\cdot X\right)\right]\Big|\lambda^0,f^0,w\right\}\\
&=\mathbb{E}\left\{\mathrm{Tr}\left[\left(\lambda^0 f^{0\prime}-\lambda f'-(\beta-\beta^0)\cdot X\right)' M_{(\lambda,\lambda^0,w)}\left(\lambda^0 f^{0\prime}-\lambda f'-(\beta-\beta^0)\cdot X\right)\right]\Big|\lambda^0,f^0,w\right\}\\
&\quad+\mathbb{E}\left\{\mathrm{Tr}\left[\left(\lambda^0 f^{0\prime}-\lambda f'-(\beta-\beta^0)\cdot X\right)' P_{(\lambda,\lambda^0,w)}\left(\lambda^0 f^{0\prime}-\lambda f'-(\beta-\beta^0)\cdot X\right)\right]\Big|\lambda^0,f^0,w\right\}\\
&=\underbrace{\mathbb{E}\left\{\mathrm{Tr}\left[\left((\beta^{\mathrm{high}}-\beta^{0,\mathrm{high}})\cdot X_{\mathrm{high}}\right)' M_{(\lambda,\lambda^0,w)}\left((\beta^{\mathrm{high}}-\beta^{0,\mathrm{high}})\cdot X_{\mathrm{high}}\right)\right]\Big|\lambda^0,f^0,w\right\}}_{\equiv Q^{\mathrm{high}}(\beta^{\mathrm{high}},\lambda)}\\
&\quad+\underbrace{\mathbb{E}\left\{\mathrm{Tr}\left[\left(\lambda^0 f^{0\prime}-\lambda f'-(\beta-\beta^0)\cdot X\right)' P_{(\lambda,\lambda^0,w)}\left(\lambda^0 f^{0\prime}-\lambda f'-(\beta-\beta^0)\cdot X\right)\right]\Big|\lambda^0,f^0,w\right\}}_{\equiv Q^{\mathrm{low}}(\beta,\lambda,f)},
\end{aligned}
$$

where $(\beta^{\mathrm{high}}-\beta^{0,\mathrm{high}})\cdot X_{\mathrm{high}}=\sum_{m=K_1+1}^{K}(\beta_m-\beta_m^0)X_m$. A lower bound on $Q^{\mathrm{high}}(\beta^{\mathrm{high}},\lambda)$ is given by

$$
\begin{aligned}
&Q^{\mathrm{high}}(\beta^{\mathrm{high}},\lambda)\\
&\geq \min_{\widetilde{\lambda}\in\mathbb{R}^{N\times(R+R+\mathrm{rank}(w))}}\mathbb{E}\left\{\mathrm{Tr}\left[\left((\beta^{\mathrm{high}}-\beta^{0,\mathrm{high}})\cdot X_{\mathrm{high}}\right)' M_{(\widetilde{\lambda},\lambda,w)}\left((\beta^{\mathrm{high}}-\beta^{0,\mathrm{high}})\cdot X_{\mathrm{high}}\right)\right]\Big|\lambda^0,f^0,w\right\}\\
&=\sum_{r=R+R+\mathrm{rank}(w)}^{\min(N,T)}\mu_r\left\{\mathbb{E}\left[\left((\beta^{\mathrm{high}}-\beta^{0,\mathrm{high}})\cdot X_{\mathrm{high}}\right)\left((\beta^{\mathrm{high}}-\beta^{0,\mathrm{high}})\cdot X_{\mathrm{high}}\right)'\Big|\lambda^0,f^0,w\right]\right\}.
\end{aligned}
\tag{S.1.1}
$$

Since $Q^*(\beta,\lambda,f)$, $Q^{\mathrm{high}}(\beta^{\mathrm{high}},\lambda)$, and $Q^{\mathrm{low}}(\beta,\lambda,f)$, are expectations of traces of positive semi-definite matrices we have $Q^*(\beta,\lambda,f)\geq 0$, $Q^{\mathrm{high}}(\beta^{\mathrm{high}},\lambda)\geq 0$, and $Q^{\mathrm{low}}(\beta,\lambda,f)\geq 0$ for all β, λ, f. Let $\bar\beta$, $\bar\lambda$ and $\bar f$ be the parameter values that minimize $Q(\beta,\lambda,f)$, and thus also $Q^*(\beta,\lambda,f)$. Since $Q^*(\beta^0,\lambda^0,f^0)=0$ we have $Q^*(\bar\beta,\bar\lambda,\bar f)=\min_{\beta,\lambda,f}Q^*(\beta,\lambda,f)=0$. This implies $Q^{\mathrm{high}}(\bar\beta^{\mathrm{high}},\bar\lambda)=0$ and $Q^{\mathrm{low}}(\bar\beta,\bar\lambda,\bar f)=0$. Assumption ID($v$), the lower bound (S.1.1), and $Q^{\mathrm{high}}(\bar\beta^{\mathrm{high}},\bar\lambda)=0$ imply that $\bar\beta^{\mathrm{high}}=\beta^{0,\mathrm{high}}$. Using this we find that

$$
\begin{aligned}
&Q^{\mathrm{low}}(\bar\beta,\bar\lambda,\bar f)\\
&=\mathbb{E}\left\{\mathrm{Tr}\left[\left(\lambda^0 f^{0\prime}-\bar\lambda\bar f'-(\bar\beta^{\mathrm{low}}-\beta^{0,\mathrm{low}})\cdot X_{\mathrm{low}}\right)'\left(\lambda^0 f^{0\prime}-\bar\lambda\bar f'-(\bar\beta^{\mathrm{low}}-\beta^{0,\mathrm{low}})\cdot X_{\mathrm{low}}\right)\right]\Big|\lambda^0,f^0,w\right\},\\
&\geq\min_f\mathbb{E}\left\{\mathrm{Tr}\left[\left(\lambda^0 f^{0\prime}-\bar\lambda f'-(\bar\beta^{\mathrm{low}}-\beta^{0,\mathrm{low}})\cdot X_{\mathrm{low}}\right)'\left(\lambda^0 f^{0\prime}-\bar\lambda f'-(\bar\beta^{\mathrm{low}}-\beta^{0,\mathrm{low}})\cdot X_{\mathrm{low}}\right)\right]\Big|\lambda^0,f^0,w\right\}\\
&=\mathbb{E}\left\{\mathrm{Tr}\left[\left(\lambda^0 f^{0\prime}-(\bar\beta^{\mathrm{low}}-\beta^{0,\mathrm{low}})\cdot X_{\mathrm{low}}\right)' M_{\bar\lambda}\left(\lambda^0 f^{0\prime}-(\bar\beta^{\mathrm{low}}-\beta^{0,\mathrm{low}})\cdot X_{\mathrm{low}}\right)\right]\Big|\lambda^0,f^0,w\right\},
\end{aligned}
\tag{S.1.2}
$$

where $(\bar\beta^{\rm low} - \beta^{0,{\rm low}}) \cdot X_{\rm low} = \sum_{l=1}^{K_1} (\bar\beta_l - \beta_l^0) X_l$. Since $Q^{\rm low}(\bar\beta, \bar\lambda, \bar f) = 0$ and the last expression in (S.1.2) is non-negative we must have

$$\mathbb{E}\left\{ \mathrm{Tr}\left[\left(\lambda^0 f^{0\prime} - (\bar\beta^{\rm low} - \beta^{0,{\rm low}}) \cdot X_{\rm low} \right)' M_{\bar\lambda} \left(\lambda^0 f^{0\prime} - (\bar\beta^{\rm low} - \beta^{0,{\rm low}}) \cdot X_{\rm low} \right) \right] \Big| \lambda^0, f^0, w \right\} = 0.$$

Using $M_{\bar\lambda} = M_{\bar\lambda} M_{\bar\lambda}$ and the cyclicality of the trace we obtain from the last equality that

$$\mathrm{Tr}\left\{ M_{\bar\lambda} A M_{\bar\lambda} \right\} = 0,$$

where $A = \mathbb{E}\left[\left(\lambda^0 f^{0\prime} - (\bar\beta^{\rm low} - \beta^{0,{\rm low}}) \cdot X_{\rm low} \right) \left(\lambda^0 f^{0\prime} - (\bar\beta^{\rm low} - \beta^{0,{\rm low}}) \cdot X_{\rm low} \right)' \Big| \lambda^0, f^0, w \right]$. The trace of a positive semi-definite matrix is only equal to zero if the matrix itself is equal to zero, so we find

$$M_{\bar\lambda} A M_{\bar\lambda} = 0,$$

This together with the fact that A itself is positive semi definite implies that (note that A positive semi-definite implies $A = CC'$ for some matrix C, and $M_{\bar\lambda} A M_{\bar\lambda} = 0$ then implies $M_{\bar\lambda} C = 0$, i.e. $C = P_{\bar\lambda} C$)

$$A = P_{\bar\lambda} A P_{\bar\lambda},$$

and therefore $\mathrm{rank}(A) \le \mathrm{rank}(P_{\bar\lambda}) \le R$. We have thus shown that

$$\mathrm{rank}\left\{ \mathbb{E}\left[\left(\lambda^0 f^{0\prime} - (\bar\beta^{\rm low} - \beta^{0,{\rm low}}) \cdot X_{\rm low} \right) \left(\lambda^0 f^{0\prime} - (\bar\beta^{\rm low} - \beta^{0,{\rm low}}) \cdot X_{\rm low} \right)' \Big| \lambda^0, f^0, w \right] \right\} \le R.$$

We furthermore find

$$\begin{aligned}
R &\ge \mathrm{rank}\left\{ \mathbb{E}\left[\left(\lambda^0 f^{0\prime} - (\bar\beta^{\rm low} - \beta^{0,{\rm low}}) \cdot X_{\rm low} \right) \left(\lambda^0 f^{0\prime} - (\bar\beta^{\rm low} - \beta^{0,{\rm low}}) \cdot X_{\rm low} \right)' \Big| \lambda^0, f^0, w \right] \right\} \\
&\ge \mathrm{rank}\left\{ M_w \mathbb{E}\left[\left(\lambda^0 f^{0\prime} - (\bar\beta^{\rm low} - \beta^{0,{\rm low}}) \cdot X_{\rm low} \right) P_{f^0} \left(\lambda^0 f^{0\prime} - (\bar\beta^{\rm low} - \beta^{0,{\rm low}}) \cdot X_{\rm low} \right)' M_w \Big| \lambda^0, f^0, w \right] \right\} \\
&\quad + \mathrm{rank}\left\{ P_w \mathbb{E}\left[\left(\lambda^0 f^{0\prime} - (\bar\beta^{\rm low} - \beta^{0,{\rm low}}) \cdot X_{\rm low} \right) M_{f^0} \left(\lambda^0 f^{0\prime} - (\bar\beta^{\rm low} - \beta^{0,{\rm low}}) \cdot X_{\rm low} \right)' P_w \Big| \lambda^0, f^0, w \right] \right\} \\
&\ge \mathrm{rank}\left[M_w \lambda^0 f^{0\prime} f^0 \lambda^{0\prime} M_w \right] \\
&\quad + \mathrm{rank}\left\{ \mathbb{E}\left[\left((\bar\beta^{\rm low} - \beta^{0,{\rm low}}) \cdot X_{\rm low} \right) M_{f^0} \left((\bar\beta^{\rm low} - \beta^{0,{\rm low}}) \cdot X_{\rm low} \right)' \Big| \lambda^0, f^0, w \right] \right\}.
\end{aligned}$$

Assumption ID(iv) guarantees that $\mathrm{rank}\left(M_w \lambda^0 f^{0\prime} f^0 \lambda^{0\prime} M_w \right) = \mathrm{rank}\left(\lambda^0 f^{0\prime} f^0 \lambda^{0\prime} \right) = R$, i.e. we must have

$$\mathbb{E}\left[\left((\bar\beta^{\rm low} - \beta^{0,{\rm low}}) \cdot X_{\rm low} \right) M_{f^0} \left((\bar\beta^{\rm low} - \beta^{0,{\rm low}}) \cdot X_{\rm low} \right)' \Big| \lambda^0, f^0, w \right] = 0.$$

According to Assumption ID(iii) this implies $\bar\beta^{\rm low} = \beta^{0,{\rm low}}$, i.e. we have $\bar\beta = \beta^0$. This also implies $Q^*(\bar\beta, \bar\lambda, \bar f) = \|\lambda^0 f^{0\prime} - \bar\lambda \bar f'\|_F^2 = 0$, and thereofere $\bar\lambda \bar f' = \lambda^0 f^{0\prime}$. ∎

S.2 Examples of Error Distributions

The following Lemma provides examples of error distributions that satisfy $\|e\| = \mathcal{O}_p(\sqrt{\max(N,T)})$ as $N, T \to \infty$. Example (i) is particularly relevant for us, because those assumptions on e_{it} are imposed in Assumption 5 in the main text, i.e. under those main text assumptions we indeed have $\|e\| = \mathcal{O}_p(\sqrt{\max(N,T)})$.

Lemma S.2.1. *For each of the following distributional assumptions on the errors e_{it}, $i = 1, \ldots, N$, $t = 1, \ldots, T$, we have $\|e\| = \mathcal{O}_p(\sqrt{\max(N,T)})$.*

(i) The e_{it} are independent across i and t, conditional on $\mathcal{C}$, and satisfy $\mathbb{E}(e_{it}|\mathcal{C}) = 0$, and $\mathbb{E}(e_{it}^4|\mathcal{C})$ is bounded uniformly by a non-random constant, uniformly over i, t and N, T. Here $\mathcal{C}$ can be any conditioning sigma-field, including the empty one (corresponding to unconditional expectations).

(ii) The e_{it} follow different MA(∞) *process for each i, namely*

$$e_{it} = \sum_{\tau=0}^{\infty} \psi_{i\tau}\, u_{i,t-\tau} \ , \qquad \textit{for } i = 1 \ldots N,\ t = 1 \ldots T \ , \tag{S.2.1}$$

where the u_{it}, $i = 1 \ldots N$, $t = -\infty \ldots T$ are independent random variables with $\mathbb{E}u_{it} = 0$ and $\mathbb{E}u_{it}^4$ uniformly bounded across i, t and N, T. The coefficients $\psi_{i\tau}$ satisfy

$$\sum_{\tau=0}^{\infty} \tau \max_{i=1\ldots N} \psi_{i\tau}^2 < B \ , \qquad\qquad \sum_{\tau=0}^{\infty} \max_{i=1\ldots N} |\psi_{i\tau}| < B \ , \tag{S.2.2}$$

for a finite constant B which is independent of N and T.

(iii) The error matrix e is generated as $e = \sigma^{1/2}\, u\, \Sigma^{1/2}$, where u is an $N \times T$ matrix with independently distributed entries u_{it} and $\mathbb{E}u_{it} = 0$, $\mathbb{E}u_{it}^2 = 1$, and $\mathbb{E}u_{it}^4$ is bounded uniformly across i, t and N, T. Here σ is the $N \times N$ cross-sectional covariance matrix, and Σ is the $T \times T$ time-serial covariance matrix, and they satisfy

$$\max_{j=1\ldots N} \sum_{i=1}^{N} |\sigma_{ij}| < B \ , \qquad\qquad \max_{\tau=1\ldots T} \sum_{t=1}^{T} |\Sigma_{t\tau}| < B \ , \tag{S.2.3}$$

for some finite constant B which is independent of N and T. In this example we have $\mathbb{E}e_{it}e_{j\tau} = \sigma_{ij}\Sigma_{t\tau}$.

Proof of Lemma S.2.1, Example (i). Latala (2005) showed that for a $N \times T$ matrix e with independent entries, conditional on $\mathcal{C}$, we have

$$\mathbb{E}\left(\|e\| \,|\mathcal{C}\right) \,\leq\, c \left\{ \max_i \left[\sum_t \mathbb{E}\left(e_{it}^2 \,|\mathcal{C}\right) \right]^{1/2} + \max_j \left[\sum_i \mathbb{E}\left(e_{it}^2 \,|\mathcal{C}\right) \right]^{1/2} + \left[\sum_{i,t} \mathbb{E}\left(e_{it}^4 \,|\mathcal{C}\right) \right]^{1/4} \right\} ,$$

where c is some universal constant. Since we assumed uniformly bounded 4'th conditional moments for e_{it} we thus have $\|e\| = \mathcal{O}_P(\sqrt{T}) + \mathcal{O}_P(\sqrt{N}) + \mathcal{O}_P((TN)^{1/4}) = \mathcal{O}_p(\sqrt{\max(N,T)})$. ∎

Example (ii). Let $\psi_j = (\psi_{1j}, \ldots, \psi_{Nj})$ be an $N \times 1$ vector for each $j \geq 0$. Let U_{-j} be an $N \times T$ sub-matrix of (u_{it}) consisting of u_{it}, $i = 1 \ldots N$, $t = 1-j, \ldots, T-j$. We can then write equation (S.2.1) in matrix notation as

$$\begin{aligned} e &= \sum_{j=0}^{\infty} \operatorname{diag}(\psi_j)\, U_{-j} \\ &= \sum_{j=0}^{T} \operatorname{diag}(\psi_j)\, U_{-j} + r_{NT}, \end{aligned}$$

where we cut the sum at T, which results in the remainder $r_{NT} = \sum_{j=T+1}^{\infty} \operatorname{diag}(\psi_j)\, U_{-j}$. When approximating an MA(∞) by a finite MA(T) process we have for the remainder

$$\begin{aligned} \mathbb{E}\left(\|r_{NT}\|_F\right)^2 = \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbb{E}\left(r_{NT}\right)_{ij}^2 &\leq \sigma_u^2 \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{j=T+1}^{\infty} \psi_{ij}^2 \\ &\leq \sigma_u^2\, N\, T \sum_{j=T+1}^{\infty} \max_i \left(\psi_{ij}^2\right) \\ &\leq \sigma_u^2\, N \sum_{j=T+1}^{\infty} j \max_i \left(\psi_{ij}^2\right) , \end{aligned}$$

where σ_u^2 is the variance of u_{it}. Therefore, for $T \to \infty$ we have

$$\mathbb{E}\left(\frac{\left(\|r_{NT}\|_F\right)^2}{N} \right) \longrightarrow 0 ,$$

which implies $\left(\|r_{NT}\|_F\right)^2 = \mathcal{O}_p(N)$, and therefore $\|r_{NT}\| \leq \|r_{NT}\|_F = \mathcal{O}_p(\sqrt{N})$.

Let V be the $N \times 2T$ matrix consisting of u_{it}, $i = 1 \ldots N$, $t = 1-T, \ldots, T$. For $j = 0 \ldots T$ the matrices U_{-j} are sub-matrices of V, and therefore $\|U_{-j}\| \leq \|V\|$. From example (i) we know that $\|V\| = \mathcal{O}_p(\sqrt{\max(N, 2T)})$. Furthermore, we know that $\|\operatorname{diag}(\psi_j)\| \leq \max_i \left(\left|\psi_{ij}\right|\right)$.

Combining these results we find

$$\begin{aligned}
\|e\| &\leq \sum_{j=0}^{T} \|\operatorname{diag}(\psi_j)\| \, \|U_{-j}\| + \|r_{NT}\| \\
&\leq \sum_{j=0}^{T} \max_i \left(|\psi_{ij}|\right) \|V\| + o_p(\sqrt{N}) \\
&\leq \left[\sum_{j=0}^{\infty} \max_i \left(|\psi_{ij}|\right)\right] \mathcal{O}_p(\sqrt{\max(N, 2T)}) + o_p(\sqrt{N}) \\
&\leq \mathcal{O}_p(\sqrt{\max(N, T)}) \ .
\end{aligned}$$

This is what we wanted to show. ∎

Example (iii). Since σ and Σ are positive definite, there exits a symmetric $N \times N$ matrix ϕ and a symmetric $T \times T$ matrix ψ such that $\sigma = \phi^2$ and $\Sigma = \psi^2$. The error term can then be generated as $e = \phi u \psi$, where u is an $N \times T$ matrix with iid entries u_{it} such that $\mathbb{E}(u_{it}) = 0$ and $\mathbb{E}(u_{it}^4) < \infty$. Given this definition of e we immediately have $\mathbb{E} e_{it} = 0$ and $\mathbb{E} e_{it} e_{j\tau} = \sigma_{ij} \Sigma_{t\tau}$. What is left to show is that $\|e\| = \mathcal{O}_p(\sqrt{\max(N, T)})$. From example (i) we know that $\|u\| = \mathcal{O}_p(\sqrt{\max(N, T)})$. Using the inequality $\|\sigma\| \leq \sqrt{\|\sigma\|_1 \|\sigma\|_\infty} = \|\sigma\|_1$, where $\|\sigma\|_1 = \|\sigma\|_\infty$ because σ is symmetric we find

$$\|\sigma\| \leq \|\sigma\|_1 \equiv \max_{j=1\ldots N} \sum_{i=1}^{N} |\sigma_{ij}| < L \ ,$$

and analogously $\|\Sigma\| < L$. Since $\|\sigma\| = \|\phi\|^2$ and $\|\Sigma\| = \|\psi\|^2$, we thus find $\|e\| \leq \|\phi\| \|u\| \|\psi\| \leq L \mathcal{O}_p(\sqrt{\max(N, T)})$, *i.e.* $\|e\| = \mathcal{O}_p(\sqrt{\max(N, T)})$. ∎

S.3 Comments on assumption 4 on the regressors

Consistency of the LS estimator $\widehat{\beta}$ requires that the regressors not only satisfy the standard non-collinearity condition in assumption 4(i), but also the additional conditions on high- and low-rank regressors in assumption 4(ii). Bai (2009) considers the special cases of only high-rank and only low-rank regressors. As low-rank regressors he considers only cross-sectional invariant and time-invariant regressors, and he shows that if only these two types of regressors are present, one can show consistency under the assumption $\operatorname{plim}_{N,T\to\infty} W_{NT} > 0$ on the regressors (instead of assumption 4), where W_{NT} is the $K \times K$ matrix defined by $W_{NT,k_1k_2} = (NT)^{-1} \operatorname{Tr}(M_{f^0} X'_{k_1} M_{\lambda^0} X_{k_2})$. This matrix appears as the approximate Hessian in the profile objective expansion in theorem 4.1, *i.e.* the condition $\text{plim}_{N,T\to\infty} W_{NT} > 0$ is very natural in the context of the interactive fixed effect models, and one may wonder whether also for the general case one can replace assumption 4 with this weaker condition and still obtain consistency of the LS estimator. Unfortunately, this is not the case, and below we present two simple counter examples that show this.

(i) Let there only be one factor ($R = 1$) f_t^0 with corresponding factor loadings λ_i^0. Let there only be one regressor ($K = 1$) of the form $X_{it} = w_i v_t + \lambda_i^0 f_t^0$. Assume that the $N \times 1$ vector $w = (w_1, \ldots, w_N)'$, and the $T \times 1$ vector $v = (v_1, \ldots, v_N)'$ are such that the $N \times 2$ matrix $\Lambda = (\lambda^0, w)$ and and the $T \times 2$ matrix $F = (f^0, v)$ satisfy $\text{plim}_{N,T\to\infty} (\Lambda'\Lambda/N) > 0$, $\text{plim}_{N,T\to\infty} (F'F/T) > 0$. In this case, we have $W_{NT} = (NT)^{-1}\,\text{Tr}(M_{f^0}\, vw'\, M_{\lambda^0}\, wv')$, and therefore $\text{plim}_{N,T\to\infty} W_{NT} = \text{plim}_{N,T\to\infty} (NT)^{-1}\,\text{Tr}(M_{f^0}\, vw'\, M_{\lambda^0}\, wv') > 0$. However, β is not identified because $\beta^0 X + \lambda^0 f^{0\prime} = (\beta^0 + 1) X - wv'$, *i.e.* it is not possible to distinguish $(\beta, \lambda, f) = (\beta^0, \lambda^0, f^0)$ and $(\beta, \lambda, f) = (\beta^0 + 1, -w, v)$. This implies that the LS estimator is not consistent (both β^0 and $\beta^0 + 1$ could be the true parameter, but the LS estimator cannot be consistent for both).

(ii) Let there only be one factor ($R = 1$) f_t^0 with corresponding factor loadings λ_i^0. Let the $N \times 1$ vectors λ^0, w_1 and w_2 be such that $\Lambda = (\lambda^0, w_1, w_2)$ satisfies $\text{plim}_{N,T\to\infty} (\Lambda'\Lambda/N) > 0$. Let the $T \times 1$ vectors f^0, v_1 and v_2 be such that $F = (f^0, v_1, v_2)$ satisfies $\text{plim}_{N,T\to\infty} (F'F/T) > 0$. Let there be four regressors ($K = 4$) defined by $X_1 = w_1 v_1'$, $X_2 = w_2 v_2'$, $X_3 = (w_1 + \lambda^0)(v_2 + f^0)'$, $X_4 = (w_2 + \lambda^0)(v_1 + f^0)'$. In this case, one can easily check that $\text{plim}_{N,T\to\infty} W_{NT} > 0$. However, again β_k is not identified, because $\sum_{k=1}^4 \beta_k^0 X_k + \lambda^0 f^{0\prime} = \sum_{k=1}^4 (\beta_k^0 + 1) X_k - (\lambda^0 + w_1 + w_2)(f^{0\prime} + v_1 + v_2)'$, *i.e.* we cannot distinguish between the true parameters and $(\beta, \lambda, f) = (\beta^0 + 1,\ -\lambda^0 - w_1 - w_2,\ f^{0\prime} + v_1 + v_2)$. Again, as a consequence the LS estimator is not consistent in this case.

In example (ii), there are only low-rank regressors with $\text{rank}(X_l) = 1$. One can easily check that assumption 4 is not satisfied for this example. In example (i) the regressor is a low-rank regressor with $\text{rank}(X) = 2$. In our present version of assumption 4 we only consider low-rank regressors with rank $\text{rank}(X) = 1$, but (as already noted in a footnote in the main paper) it is straightforward to extend the assumption and the consistency proof to low-rank regressors with rank larger than one. Independent of whether we extend the assumption or not, the regressor X of example (i) fails to satisfy assumption 4. This justifies our formulation of assumption 4, because it shows that in general the assumption cannot be replaced by the weaker condition $\text{plim}_{N,T\to\infty} W_{NT} > 0$.

S.4 Some Matrix Algebra (including Proof of Lemma A.1)

The following statements are true for real matrices (throughout the whole paper and supplementary material we never use complex numbers anywhere). Let A be an arbitrary $n \times m$ matrix. In addition to the operator (or spectral) norm $\|A\|$ and to the Frobenius (or Hilbert-Schmidt) norm $\|A\|_F$, it is also convenient to define the 1-norm, the ∞-norm, and the max-norm by

$$\|A\|_1 = \max_{j=1\ldots m} \sum_{i=1}^{n} |A_{ij}| \; , \quad \|A\|_\infty = \max_{i=1\ldots n} \sum_{j=1}^{m} |A_{ij}| \; , \quad \|A\|_{\max} = \max_{i=1\ldots n} \max_{j=1\ldots m} |A_{ij}| \; .$$

Lemma S.4.1 (Some useful Inequalities). *Let A be an $n \times m$ matrix, B be an $m \times p$ matrix, and C and D be $n \times n$ matrices. Then we have:*

$$
\begin{aligned}
&(i) && \|A\| \leq \|A\|_F \leq \|A\| \operatorname{rank}(A)^{1/2} \; , \\
&(ii) && \|AB\| \leq \|A\| \, \|B\| \; , \\
&(iii) && \|AB\|_F \leq \|A\|_F \, \|B\| \leq \|A\|_F \, \|B\|_F \; , \\
&(iv) && |\operatorname{Tr}(AB)| \leq \|A\|_F \, \|B\|_F \; , \qquad \text{for } n = p, \\
&(v) && |\operatorname{Tr}(C)| \leq \|C\| \operatorname{rank}(C) \; , \\
&(vi) && \|C\| \leq \operatorname{Tr}(C) \; , \qquad \text{for } C \text{ symmetric and } C \geq 0, \\
&(vii) && \|A\|^2 \leq \|A\|_1 \, \|A\|_\infty \; , \\
&(viii) && \|A\|_{\max} \leq \|A\| \leq \sqrt{nm} \, \|A\|_{\max} \; , \\
&(ix) && \|A'CA\| \leq \|A'DA\| \; , \qquad \text{for } C \text{ symmetric and } C \leq D.
\end{aligned}
$$

For C, D symmetric, and $i = 1, \ldots, n$ we have:

$$
\begin{aligned}
&(x) && \mu_i(C) + \mu_n(D) \leq \mu_i(C + D) \leq \mu_i(C) + \mu_1(D) \; , \\
&(xi) && \mu_i(C) \leq \mu_i(C + D) \; , \qquad \text{for } D \geq 0, \\
&(xii) && \mu_i(C) - \|D\| \leq \mu_i(C + D) \leq \mu_i(C) + \|D\| \; .
\end{aligned}
$$

Proof. Here we use notation $s_i(A)$ for the i'th largest singular value of a matrix A.
(i) We have $\|A\| = s_1(A)$, and $\|A\|_F^2 = \sum_{i=1}^{\operatorname{rank}(A)} (s_i(A))^2$. The inequalities follow directly from this representation. (ii) This inequality is true for all unitarily invariant norms, see *e.g.* Bhatia (1997). (iii) can be shown as follows

$$
\begin{aligned}
\|AB\|_F^2 &= \operatorname{Tr}(ABB'A') \\
&= \operatorname{Tr}[\|B\|^2 AA' - A(\|B\|^2 \mathbb{I} - BB')A'] \\
&\leq \|B\|^2 \operatorname{Tr}(AA') = \|B\|^2 \, \|A\|_F^2 \; ,
\end{aligned}
$$

where we used that $A(\|B\|^2\mathbb{I} - BB')A'$ is positive definite. Relation (iv) is just the Cauchy Schwarz inequality. To show (v) we decompose $C = UDO'$ (singular value decomposition), where U and O are $n \times \text{rank}(C)$ that satisfy $U'U = O'O = \mathbb{I}$ and D is a $\text{rank}(C) \times \text{rank}(C)$ diagonal matrix with entries $s_i(C)$. We then have $\|O\| = \|U\| = 1$ and $\|D\| = \|C\|$ and therefore

$$\begin{aligned}
|\text{Tr}(C)| &= |\text{Tr}(UDO')| = |\text{Tr}(DO'U)| \\
&= \left| \sum_{i=1}^{\text{rank}(C)} \eta_i' DO'U\eta_i \right| \\
&\leq \sum_{i=1}^{\text{rank}(C)} \|D\|\|O'\|\|U\| = \text{rank}(C)\|C\| \,.
\end{aligned}$$

For (vi) let e_1 be a vector that satisfied $\|e_1\| = 1$ and $\|C\| = e_1'Ce_1$. Since C is symmetric such an e_1 has to exist. Now choose e_i, $i = 2, \ldots, n$, such that e_i, $i = 1, \ldots, n$, becomes a orthonormal basis of the vector space of $n \times 1$ vectors. Since C is positive semi definite we then have $\text{Tr}\,(C) = \sum_i e_i'Ce_i \geq e_1Ce_1 = \|C\|$, which is what we wanted to show. For (vii) we refer to Golub and van Loan (1996), p.15. For (viii) let e be the vector that satisfies $\|e\| = 1$ and $\|A'CA\| = e'A'CAe$. Since $A'CA$ is symmetric such an e has to exist. Since $C \leq D$ we then have $\|C\| = (e'A')C(Ae) \leq (e'A')D(Ae) \leq \|A'DA\|$. This is what we wanted to show. For inequality (ix) let e_1 be a vector that satisfied $\|e_1\| = 1$ and $\|A'CA\| = e_1'A'CAe_1$. Then we have $\|A'CA\| = e_1'A'DAe_1 - e_1'A'(D - C)Ae_1 \leq e_1'A'DAe_1 \leq \|A'DA\|$. Statement (x) is a special case of Weyl's inequality, see *e.g.* Bhatia (1997). The Inequalities (xi) and (xii) follow directly from (ix) since $\mu_n(D) \geq 0$ for $D \geq 0$, and since $-\|D\| \leq \mu_i(D) \leq \|D\|$ for $i = 1, \ldots, n$. ∎

Definition S.4.2. *Let A be an $n \times r_1$ matrix and B be an $n \times r_2$ matrix with* $\text{rank}(A) = r_1$ *and* $\text{rank}(B) = r_2$. *The smallest principal angle $\theta_{A,B} \in [0, \pi/2]$ between the linear subspaces* $\text{span}(A) = \{Aa|\, a \in \mathbb{R}^{r_1}\}$ *and* $\text{span}(B) = \{Bb|\, b \in \mathbb{B}^{r_2}\}$ *of $\mathbb{R}^n$ is defined by*

$$\cos(\theta_{A,B}) = \max_{0 \neq a \in \mathbb{R}^{r_1}} \max_{0 \neq b \in \mathbb{R}^{r_2}} \frac{a'A'Bb}{\|Aa\|\|Bb\|} \,.$$

Lemma S.4.3. *Let A be an $n \times r_1$ matrix and B be an $n \times r_2$ matrix with* $\text{rank}(A) = r_1$ *and* $\text{rank}(B) = r_2$. *Then we have the following alternative characterizations of the smallest principal angle between* $\text{span}(A)$ *and* $\text{span}(B)$

$$\begin{aligned}
\sin(\theta_{A,B}) &= \min_{0 \neq a \in \mathbb{R}^{r_1}} \frac{\|M_B\, A\, a\|}{\|A\, a\|} \\
&= \min_{0 \neq b \in \mathbb{R}^{r_2}} \frac{\|M_A\, B\, b\|}{\|B\, b\|} \,.
\end{aligned}$$

Proof. Since $\|M_B\, A\, a\|^2 + \|P_B\, A\, a\|^2 = \|A\, a\|^2$ and $\sin(\theta_{A,B})^2 + \cos(\theta_{A,B})^2 = 1$, we find that proving the theorem is equivalent to proving

$$\cos(\theta_{A,B}) = \min_{0 \neq a \in \mathbb{R}^{r_1}} \frac{\|P_B\, A\, a\|}{\|A\, a\|} = \min_{0 \neq b \in \mathbb{R}^{r_2}} \frac{\|P_A\, B\, b\|}{\|A\, b\|} \; .$$

This result is theorem 8 in Galantai, Hegedus (2006), and the proof can be found there. ∎

Proof of Lemma A.1. Let

$$\begin{aligned}
S_1(Z) &= \min_{f,\lambda} \operatorname{Tr}\left[(Z - \lambda f')\,(Z' - f\,\lambda')\right] \; , \\
S_2(Z) &= \min_{f} \operatorname{Tr}(Z\, M_f\, Z') \; , \\
S_3(Z) &= \min_{\lambda} \operatorname{Tr}(Z'\, M_\lambda\, Z) \; , \\
S_4(Z) &= \min_{\tilde{\lambda},\tilde{f}} \operatorname{Tr}(M_{\tilde{\lambda}}\, Z\, M_{\tilde{f}}\, Z') \; , \\
S_5(Z) &= \sum_{i=R+1}^{T} \mu_i(Z'Z) \; , \\
S_6(Z) &= \sum_{i=R+1}^{N} \mu_i(ZZ') \; .
\end{aligned}$$

The theorem claims

$$S_1(Z) \;=\; S_2(Z) \;=\; S_3(Z) \;=\; S_4(Z) \;=\; S_5(Z) \;=\; S_6(Z) \; .$$

We find:

(i) The non-zero eigenvalues of $Z'Z$ and ZZ' are identical, so in the sums in $S_5(Z)$ and in $S_6(Z)$ we are summing over identical values, which shows $S_5(Z) = S_6(Z)$.

(ii) Starting with $S_1(Z)$ and minimizing with respect to f we obtain the first order condition

$$\lambda'\, Z = \lambda'\, \lambda\, f' \; .$$

Putting this into the objective function we can integrate out f, namely

$$\begin{aligned}
\operatorname{Tr}\left[(Z - \lambda f')'\,(Z - \lambda f')\right] &= \operatorname{Tr}\left(Z'Z - Z'\lambda f'\right) \\
&= \operatorname{Tr}\left(Z'Z - Z'\lambda(\lambda'\lambda)^{-1}(\lambda'\lambda) f'\right) \\
&= \operatorname{Tr}\left(Z'Z - Z'\lambda(\lambda'\lambda)^{-1}(\lambda'\lambda)\lambda'\, Z\right) \\
&= \operatorname{Tr}\left(Z'\, M_\lambda\, Z\right) \; .
\end{aligned}$$

This shows $S_1(Z) = S_3(Z)$. Analogously, we can integrate out λ to obtain $S_1(Z) = S_2(Z)$.

(iii) Let $M_{\widehat{\lambda}}$ be the projector on the $N-R$ eigenspaces corresponding to the $N-R$ smallest eigenvalues[1] of ZZ', let $P_{\widehat{\lambda}} = \mathbb{I}_N - M_{\widehat{\lambda}}$, and let ω_R be the R'th largest eigenvalue of ZZ'. We then know that the matrix $P_{\widehat{\lambda}}[ZZ' - \omega_R \mathbb{I}_N] P_{\widehat{\lambda}} - M_{\widehat{\lambda}}[ZZ' - \omega_R \mathbb{I}_N] M_{\widehat{\lambda}}$ is positive semi-definite. Thus, for an arbitrary $N \times R$ matrix λ with corresponding projector M_λ we have

$$\begin{aligned}
0 &\leq \mathrm{Tr}\left\{ \left(P_{\widehat{\lambda}}[ZZ' - \omega_R \mathbb{I}_N] P_{\widehat{\lambda}} - M_{\widehat{\lambda}}[ZZ' - \omega_R \mathbb{I}_N] M_{\widehat{\lambda}} \right) \left(M_\lambda - M_{\widehat{\lambda}} \right)^2 \right\} \\
&= \mathrm{Tr}\left\{ \left(P_{\widehat{\lambda}}[ZZ' - \omega_R \mathbb{I}_N] P_{\widehat{\lambda}} + M_{\widehat{\lambda}}[ZZ' - \omega_R \mathbb{I}_N] M_{\widehat{\lambda}} \right) \left(M_\lambda - M_{\widehat{\lambda}} \right) \right\} \\
&= \mathrm{Tr}\left[Z' \, M_\lambda \, Z \right] - \mathrm{Tr}\left[Z' \, M_{\widehat{\lambda}} \, Z \right] + \omega_R \, \left[\mathrm{rank}(M_\lambda) - \mathrm{rank}(M_{\widehat{\lambda}}) \right] \ ,
\end{aligned}$$

and since $\mathrm{rank}(M_{\widehat{\lambda}}) = N - R$ and $\mathrm{rank}(M_\lambda) \leq N - R$ we have

$$\mathrm{Tr}\left[Z' \, M_{\widehat{\lambda}} \, Z \right] \leq \mathrm{Tr}\left[Z' \, M_\lambda \, Z \right] \ .$$

This shows that $M_{\widehat{\lambda}}$ is the optimal choice in the minimization problem of $S_3(Z)$, *i.e.* the optimal $\lambda = \widehat{\lambda}$ is chosen such that the span of the N-dimensional vectors $\widehat{\lambda}_r$ $(r = 1 \ldots R)$ equals to the span of the R eigenvectors that correspond to the R largest eigenvalues of ZZ'. This shows that $S_3(Z) = S_6(Z)$. Analogously one can show that $S_2(Z) = S_5(Z)$.

(iv) In the minimization problem in $S_4(Z)$ we can choose $\widetilde{\lambda}$ such that the span of the N-dimensional vectors $\widetilde{\lambda}_r$ $(r = 1 \ldots R_1)$ equals to the span of the R_1 eigenvectors that correspond to the R_1 largest eigenvalues of ZZ'. In addition, we can choose $\widetilde{f}$ such that the span of the T-dimensional vectors $\widetilde{f}_r$ $(r = 1 \ldots R_2)$ equals to the span of the R_2 eigenvectors that correspond to the (R_1+1)-largest up to the R-largest eigenvalue of $Z'Z$. With this choice of $\widetilde{\lambda}$ and $\widetilde{f}$ we actually project out all the R largest eigenvalues of $Z'Z$ and ZZ'. This shows that $S_4(Z) \leq S_5(Z)$. (This result is actually best understood by using the singular value decomposition of Z.)

We can write $M_{\widetilde{\lambda}} \, Z \, M_{\widetilde{f}} = Z - \widetilde{Z}$, where

$$\widetilde{Z} = P_{\widetilde{\lambda}} \, Z \, M_{\widetilde{f}} + Z \, P_{\widetilde{f}} \ .$$

Since $\mathrm{rank}(Z) \leq \mathrm{rank}(P_{\widetilde{\lambda}} \, Z \, M_{\widetilde{f}}) + \mathrm{rank}(Z \, P_{\widetilde{f}}) = R_1 + R_2 = R$, we can always write $\widetilde{Z} = \lambda f'$

[1]If an eigenvalue has multiplicity m, we count it m times when finding the $N-R$ smallest eigenvalues. In this terminology we always have exactly N eigenvalues of ZZ', but some may appear multiple times.

for some appropriate $N \times R$ and $T \times R$ matrices λ and f. This shows that

$$\begin{aligned} S_4(Z) &= \min_{\bar\lambda, \bar f} \mathrm{Tr}(M_{\widetilde\lambda}\, Z\, M_{\widetilde f}\, Z') \\ &\geq \min_{\{\widetilde Z\,:\,\mathrm{rank}(\widetilde Z)\leq R\}} \mathrm{Tr}((Z-\widetilde Z)(Z-\widetilde Z)') \\ &= \min_{f,\lambda} \mathrm{Tr}\left[(Z-\lambda f')\,(Z'-f\lambda')\right] = S_1(Z)\ . \end{aligned}$$

Thus we have shown here $S_1(Z) \leq S_4(Z) \leq S_5(Z)$, and actually this holds with equality since $S_1(Z) = S_5(Z)$ was already shown above.

∎

S.5 Supplement to the Consistency Proof (Appendix A)

Lemma S.5.1. *Under assumptions 1 and 4 there exists a constant $B_0 > 0$ such that for the matrices w and v introduced in assumption 4 we have*

$$\begin{aligned} w'\, M_{\lambda^0}\, w \,-\, B_0\, w'\, w &\geq 0\ , \qquad \textit{wpa1,} \\ v'\, M_{f^0}\, v \,-\, B_0\, v'\, v &\geq 0\ , \qquad \textit{wpa1.} \end{aligned}$$

Proof. We can decompose $w = \widetilde w\, \bar w$, where $\widetilde w$ is an $N \times \mathrm{rank}(w)$ matrix and $\bar w$ is a $\mathrm{rank}(w) \times K_1$ matrix. Note that $\widetilde w$ has full rank, and $M_w = M_{\widetilde w}$.

By assumption 1(i) we know that $\lambda^{0\prime}\lambda^0/N$ has a probability limit, i.e there exists some $B_1 > 0$ such that $\lambda^{0\prime}\lambda^0/N < B_1 \mathbb{I}_R$ wpa1. Using this and assumption 4 we find that for any $R \times 1$ vector $a \neq 0$ we have

$$\frac{\|M_v\, \lambda^0\, a\|^2}{\|\lambda^0\, a\|^2} = \frac{a'\, \lambda^{0\prime}\, M_v\, \lambda^0\, a}{a'\, \lambda^{0\prime}\, \lambda^0\, a} > \frac{B}{B_1}\ , \qquad \text{wpa1.}$$

Applying Lemma S.4.3 we find

$$\min_{0\neq b\in\mathbb{R}^{\mathrm{rank}(w)}} \frac{b'\, \widetilde w'\, M_{\lambda^0}\, \widetilde w\, b}{b'\, \widetilde w'\, \widetilde w\, b} = \min_{0\neq a\in\mathbb{R}^R} \frac{a'\, \lambda^{0\prime}\, M_w\, \lambda^0\, a}{a'\, \lambda^{0\prime}\, \lambda^0\, a} > \frac{B}{B_1}\ , \qquad \text{wpa1.}$$

Therefore we find for every $\mathrm{rank}(w) \times 1$ vector b that $b'\, (\widetilde w'\, M_{\lambda^0}\, \widetilde w - (B/B_1)\widetilde w' \widetilde w\,)\, b > 0$, wpa1. Thus $\widetilde w'\, M_{\lambda^0}\, \widetilde w - (B/B_1)\, \widetilde w'\, \widetilde w > 0$, wpa1. Multiplying from the left with $\bar w'$ and from the right with $\bar w$ we obtain $w'\, M_{\lambda^0}\, w - (B/B_1)\, w'\, w \geq 0$, wpa1. This is what we wanted do show. Analogously we can show the statement for v. ∎

As a consequence of the this lemma we obtain some properties of the low-rank regressors summarized in the following lemma.

Lemma S.5.2. *Let the assumptions 1 and 4 be satisfied and let* $X_{\mathrm{low},\alpha} = \sum_{l=1}^{K_1} \alpha_l X_l$ *be a linear combination of the low-rank regressors. Then there exists some constant* $B > 0$ *such that*

$$\min_{\{\alpha \in \mathbb{R}^{K_1}, \|\alpha\|=1\}} \frac{\left\| X_{\mathrm{low},\alpha}\, M_{f^0}\, X'_{\mathrm{low},\alpha} \right\|}{NT} > B \ , \qquad wpa1,$$

$$\min_{\{\alpha \in \mathbb{R}^{K_1}, \|\alpha\|=1\}} \frac{\left\| M_{\lambda^0}\, X_{\mathrm{low},\alpha}\, M_{f^0}\, X'_{\mathrm{low},\alpha}\, M_{\lambda^0} \right\|}{NT} > B \ , \qquad wpa1.$$

Proof. Note that $\left\| M_{\lambda^0}\, X_{\mathrm{low},\alpha}\, M_{f^0}\, X'_{\mathrm{low},\alpha}\, M_{\lambda^0} \right\| \leq \left\| X_{\mathrm{low},\alpha}\, M_{f^0}\, X'_{\mathrm{low},\alpha} \right\|$, because $\|M_{\lambda^0}\| = 1$, *i.e.* if we can show the second inequality of the lemma we have also shown the first inequality.

We can write $X_{\mathrm{low},\alpha} = w \, \mathrm{diag}(\alpha')\, v'$. Using Lemma S.5.1 and part (v), (vi) and (ix) of Lemma S.4.1 we find

$$\begin{aligned}
\left\| M_{\lambda^0}\, X_{\mathrm{low},\alpha}\, M_{f^0}\, X'_{\mathrm{low},\alpha}\, M_{\lambda^0} \right\| &= \| M_{\lambda^0}\, w\, \mathrm{diag}(\alpha')\, v'\, M_{f^0}\, v\, \mathrm{diag}(\alpha')\, w' M_{\lambda^0} \| \\
&\geq B_0 \ \| M_{\lambda^0}\, w\, \mathrm{diag}(\alpha')\, v' \ v\, \mathrm{diag}(\alpha')\, w' M_{\lambda^0} \| \\
&\geq \frac{B_0}{K_1} \, \mathrm{Tr} \left[M_{\lambda^0}\, w\, \mathrm{diag}(\alpha')\, v' \ v\, \mathrm{diag}(\alpha')\, w' M_{\lambda^0} \right] \\
&= \frac{B_0}{K_1} \, \mathrm{Tr} \left[v\, \mathrm{diag}(\alpha')\, w' M_{\lambda^0} w\, \mathrm{diag}(\alpha')\, v' \right] \\
&\geq \frac{B_0}{K_1} \ \| v\, \mathrm{diag}(\alpha')\, w' M_{\lambda^0} w\, \mathrm{diag}(\alpha')\, v' \| \\
&\geq \frac{B_0^2}{K_1} \ \| v\, \mathrm{diag}(\alpha')\, w' w\, \mathrm{diag}(\alpha')\, v' \| \\
&\geq \frac{B_0^2}{K_1^2} \, \mathrm{Tr} \left[v\, \mathrm{diag}(\alpha')\, w' w\, \mathrm{diag}(\alpha')\, v' \right] \\
&= \frac{B_0^2}{K_1^2} \mathrm{Tr} \left[X_{\mathrm{low},\alpha} X'_{\mathrm{low},\alpha} \right] \ .
\end{aligned}$$

Thus we have $\left\| M_{\lambda^0}\, X_{\mathrm{low},\alpha}\, M_{f^0}\, X'_{\mathrm{low},\alpha}\, M_{\lambda^0} \right\| / (NT) \geq (B_0/K_1)^2 \, \alpha' \, W_{NT}^{\mathrm{low}} \, \alpha$, where the $K_1 \times K_1$ matrix W_{NT}^{low} is defined by $W_{NT, l_1 l_2}^{\mathrm{low}} = (NT)^{-1} \mathrm{Tr} \left(X_{l_1} X'_{l_2} \right)$, *i.e.* it is a submatrix of W_{NT}. Since W_{NT} and thus W_{NT}^{low} converges to a positive definite matrix the lemma is proven by the inequality above. ∎

Using the above lemmas we can now prove the lower bound on $\widetilde{S}_{NT}^{(2)}(\beta, f)$ that was used in the consistency proof. Remember that

$$\widetilde{S}_{NT}^{(2)}(\beta, f) = \frac{1}{NT} \, \mathrm{Tr} \left[\left(\lambda^0 \, f^{0\prime} + \sum_{k=1}^{K} (\beta_k^0 - \beta_k) X_k \right) \, M_f \, \left(\lambda^0 \, f^{0\prime} + \sum_{k=1}^{K} (\beta_k^0 - \beta_k) X_k \right)' \, P_{(\lambda^0, w)} \right] \ .$$

We want to show that under the assumptions of theorem 3.1 there exist finite positive constants a_0, a_1, a_2, a_3 and a_4 such that

$$\widetilde{S}^{(2)}_{NT}(\beta,f) \geq \frac{a_0 \left\|\beta^{\rm low} - \beta^{0,\rm low}\right\|^2}{\left\|\beta^{\rm low} - \beta^{0,\rm low}\right\|^2 + a_1 \left\|\beta^{\rm low} - \beta^{0,\rm low}\right\| + a_2} \\ - a_3 \left\|\beta^{\rm high} - \beta^{0,\rm high}\right\| - a_4 \left\|\beta^{\rm high} - \beta^{0,\rm high}\right\| \left\|\beta^{\rm low} - \beta^{0,\rm low}\right\| , \qquad \text{wpa1.}$$

Proof of the lower bound on $\widetilde{S}^{(2)}_{NT}(\beta,f)$. Applying Lemma A.1 and part (xi) of Lemma S.4.1 we find that

$$\begin{aligned}
\widetilde{S}^{(2)}_{NT}(\beta,f) &\geq \frac{1}{NT}\,\mu_{R+1}\left[\left(\lambda^0 f^{0\prime} + \sum_{k=1}^{K}(\beta^0_k - \beta_k)X_k\right)' P_{(\lambda^0,w)} \left(\lambda^0 f^{0\prime} + \sum_{k=1}^{K}(\beta^0_k - \beta_k)X_k\right)\right] \\
&= \frac{1}{NT}\,\mu_{R+1}\Bigg[\left(\lambda^0 f^{0\prime} + \sum_{l=1}^{K_1}(\beta^0_l - \beta_l)w_l\, v_l'\right)' \left(\lambda^0 f^{0\prime} + \sum_{l=1}^{K_1}(\beta^0_l - \beta_l)w_l\, v_l'\right) \\
&\qquad + \left(\lambda^0 f^{0\prime} + \sum_{l=1}^{K_1}(\beta^0_l - \beta_l)w_l\, v_l'\right)' P_{(\lambda^0,w)} \sum_{m=K_1}^{K}(\beta^0_m - \beta_m)X_m \\
&\qquad + \sum_{m=K_1}^{K}(\beta^0_m - \beta_m)X_m' P_{(\lambda^0,w)} \left(\lambda^0 f^{0\prime} + \sum_{l=1}^{K_1}(\beta^0_l - \beta_l)w_l\, v_l'\right) \\
&\qquad + \sum_{m=K_1}^{K}(\beta^0_m - \beta_m)X_m' P_{(\lambda^0,w)} \sum_{m=K_1}^{K}(\beta^0_m - \beta_m)X_m\Bigg] \\
&\geq \frac{1}{NT}\,\mu_{R+1}\Bigg[\left(\lambda^0 f^{0\prime} + \sum_{l=1}^{K_1}(\beta^0_l - \beta_l)w_l\, v_l'\right)' \left(\lambda^0 f^{0\prime} + \sum_{l=1}^{K_1}(\beta^0_l - \beta_l)w_l\, v_l'\right) \\
&\qquad + \left(\lambda^0 f^{0\prime} + \sum_{l=1}^{K_1}(\beta^0_l - \beta_l)w_l\, v_l'\right)' P_{(\lambda^0,w)} \sum_{m=K_1}^{K}(\beta^0_m - \beta_m)X_m \\
&\qquad + \sum_{m=K_1}^{K}(\beta^0_m - \beta_m)X_m' P_{(\lambda^0,w)} \left(\lambda^0 f^{0\prime} + \sum_{l=1}^{K_1}(\beta^0_l - \beta_l)w_l\, v_l'\right)\Bigg] \\
&\geq \frac{1}{NT}\,\mu_{R+1}\left[\left(\lambda^0 f^{0\prime} + \sum_{l=1}^{K_1}(\beta^0_l - \beta_l)w_l\, v_l'\right)' \left(\lambda^0 f^{0\prime} + \sum_{l=1}^{K_1}(\beta^0_l - \beta_l)w_l\, v_l'\right)\right] \\
&\qquad - a_3 \left\|\beta^{\rm high} - \beta^{0,\rm high}\right\| - a_4 \left\|\beta^{\rm high} - \beta^{0,\rm high}\right\| \left\|\beta^{\rm low} - \beta^{0,\rm low}\right\| , \qquad \text{wpa1,}
\end{aligned}$$

where $a_3 > 0$ and $a_4 > 0$ are appropriate constants. For the last step we used part (xii) of

Lemma S.4.1 and the fact that

$$\frac{1}{NT}\left\|\sum_{m=K_1}^{K}(\beta_m^0-\beta_m)X_m' P_{(\lambda^0,w)}\left(\lambda^0 f^{0\prime}+\sum_{l=1}^{K_1}(\beta_l^0-\beta_l)w_l\, v_l'\right)\right\|$$
$$\leq K\,\left\|\beta^{\rm high}-\beta^{0,\rm high}\right\|\max_m\left\|\frac{X_m}{\sqrt{NT}}\right\|\left(\left\|\frac{\lambda^0 f^{0\prime}}{\sqrt{NT}}\right\|+K\,\left\|\beta^{\rm low}-\beta^{0,\rm low}\right\|\max_l\left\|\frac{w_l v_l'}{\sqrt{NT}}\right\|\right)\,.$$

Our assumptions guarantee that the operator norms of $\lambda^0 f^{0\prime}/\sqrt{NT}$ and $X_m/\sqrt{NT}$ are bounded from above as $N,T\to\infty$, which results in finite constants a_3 and a_4.

We write the above result as $\widetilde{S}^{(2)}_{NT}(\beta,f)\geq \mu_{R+1}(A'A)/(NT)+\text{terms containing }\beta^{\rm high}$, where we defined $A=\lambda^0 f^{0\prime}+\sum_{l=1}^{K_1}(\beta_l^0-\beta_l)\,w_l\,v_l'$. Also write $A=A_1+A_2+A_3$, with $A_1=M_w\,A\,P_{f^0}=M_w\,\lambda^0\,f^{0\prime}$, $A_2=P_w\,A\,M_{f^0}=\sum_{l=1}^{K_1}(\beta_l^0-\beta_l)\,w_l\,v_l'\,M_{f^0}$, $A_3=P_w\,A\,P_{f^0}=P_w\,\lambda^0\,f^{0\prime}+\sum_{l=1}^{K_1}(\beta_l^0-\beta_l)\,w_l\,v_l'\,P_f$. We then find $A'A=A_1'A_1+(A_2'+A_3')(A_2+A_3)$ and

$$\begin{aligned}A'A\;&\geq\;A'A-(a^{1/2}A_3'+a^{-1/2}A_2')(a^{1/2}A_3+a^{-1/2}A_2)\\&=[A_1'A_1-(a-1)\,A_3'A_3]\;+\;(1-a^{-1})A_2'A_2\,,\end{aligned}$$

where $\geq$ for matrices refers to the difference being positive definite, and a is a positive number. We choose $a=1+\mu_R(A_1'A_1)/(2\,\|A_3\|^2)$. The reason for this choice becomes clear below.

Note that $[A_1'A_1-(a-1)\,A_3'A_3]$ has at most rank R (asymptotically it has exactly rank R). The non-zero eigenvalues of $A'A$ are therefore given by the (at most) R non-zero eigenvalues of $[A_1'A_1-(a-1)\,A_3'A_3]$ and the non-zero eigenvalues of $(1-a^{-1})A_2'A_2$, the largest one of the latter being given given by the operator norm $(1-a^{-1})\|A_2\|^2$. We therefore find

$$\begin{aligned}\frac{1}{NT}\,\mu_{R+1}\,(A'A)&\geq\frac{1}{NT}\,\mu_{R+1}\left[(A_1'A_1-(a-1)\,A_3'A_3)\;+\;(1-a^{-1})A_2'A_2\right]\\&\geq\frac{1}{NT}\,\min\left\{(1-a^{-1})\|A_2\|^2\,,\;\;\mu_R\,[A_1'A_1-(a-1)\,A_3'A_3]\right\}\,.\end{aligned}$$

Using Lemma S.4.1(xii) and our particular choice of a we find

$$\begin{aligned}\mu_R\,[A_1'A_1-(a-1)\,A_3'A_3]\geq\;&\mu_R(A_1'A_1)-\|(a-1)A_3'A_3\|\\=\;&\frac{1}{2}\,\mu_R(A_1'A_1)\,.\end{aligned}$$

Therefore

$$\begin{aligned}\frac{1}{NT}\,\mu_{R+1}(A'A)&\geq\frac{1}{2\,NT}\,\mu_R(A_1'A_1)\,\min\left\{1\,,\;\;\frac{2\,\|A_2\|^2}{2\,\|A_3\|^2+\mu_R(A_1'A_1)}\right\}\\&\geq\frac{1}{NT}\,\frac{\|A_2\|^2\,\mu_R(A_1'A_1)}{2\,\|A\|^2+\mu_R(A_1'A_1)}\,,\end{aligned}$$

where we used $\|A\| \geq \|A_3\|$ and $\|A\| \geq \|A_2\|$.

Our assumptions guarantee that there exist positive constants c_0, c_1, c_2 and c_3 such that

$$\begin{aligned}
\frac{\|A\|}{\sqrt{NT}} &\leq \frac{\|\lambda^0 f^{0\prime}\|}{\sqrt{NT}} + \sum_{l=1}^{K_1} |\beta_l^0 - \beta_l| \frac{\|w_l\, v_l'\|}{\sqrt{NT}} \leq c_0 + c_1 \left\|\beta^{\rm low} - \beta^{0,\rm low}\right\| , \quad \text{wpa1} , \\
\frac{\mu_R(A_1' A_1)}{NT} &= \frac{\mu_R\left(f^0\, \lambda^{0\prime}\, M_w\, \lambda^0\, f^{0\prime}\right)}{NT} \geq c_2 , \quad \text{wpa1} , \\
\frac{\|A_2\|^2}{NT} &= \mu_1 \left[\sum_{l_1=1}^{K_1} (\beta_{l_1}^0 - \beta_{l_1})\, w_{l_1}\, v_{l_1}'\, M_{f^0} \sum_{l_2=1}^{K_1} (\beta_{l_2}^0 - \beta_{l_2})\, v_{l_2}\, w_{l_2}' \right] \\
&\geq c_3 \left\|\beta^{\rm low} - \beta^{0,\rm low}\right\|^2 , \quad \text{wpa1} ,
\end{aligned}$$

were for the last inequality we used Lemma S.5.2.

We thus have

$$\frac{1}{NT}\, \mu_{R+1}\left(A'A\right) \geq \frac{c_3 \left\|\beta^{\rm low} - \beta^{0,\rm low}\right\|^2}{1 + \frac{2}{c_2}\left(c_0 + c_1 \left\|\beta^{\rm low} - \beta^{0,\rm low}\right\|\right)^2} , \quad \text{wpa1} .$$

Defining $a_0 = \frac{c_2 c_3}{2 c_1^2}$, $a_1 = \frac{2 c_0}{c_1}$ and $a_2 = \frac{c_2}{2 c_1^2}$ we thus obtain

$$\frac{1}{NT}\, \mu_{R+1}\left(A'A\right) \geq \frac{a_0 \left\|\beta^{\rm low} - \beta^{0,\rm low}\right\|^2}{\left\|\beta^{\rm low} - \beta^{0,\rm low}\right\|^2 + a_1 \left\|\beta^{\rm low} - \beta^{0,\rm low}\right\| + a_2} , \quad \text{wpa1} ,$$

i.e. we have shown the desired bound on $\widetilde{S}_{NT}^{(2)}(\beta, f)$. ∎

S.6 Regarding the Proof of Corollary 4.2

As discussed in the main text, the proof of Corollary 4.2 is provided in Moon and Weidner (2013). All that is left to show here is that the matrix $W_{NT} = W_{NT}(\lambda^0,\, f^0,\, X_k)$ does not become singular as $N, T \to \infty$ under our assumptions.

Proof. Remember that

$$W_{NT} = \frac{1}{NT} \mathrm{Tr}(M_{f^0}\, X_{k_1}'\, M_{\lambda^0}\, X_{k_2}) .$$

The smallest eigenvalue of the symmetric matrix $W(\lambda^0,\ f^0,\ X_k)$ is given my

$$
\begin{aligned}
\mu_K\left(W_{NT}\right) &= \min_{\{a\in\mathbb{R}^K,\ a\neq 0\}} \frac{a'\,W_{NT}\,a}{\|a\|^2} \\
&= \min_{\{a\in\mathbb{R}^K,\ a\neq 0\}} \frac{1}{NT\,\|a\|^2} \mathrm{Tr}\left[M_{f^0} \left(\sum_{k_1=1}^K a_{k_1} X'_{k_1}\right) M_{\lambda^0} \left(\sum_{k_2=1}^K a_{k_2} X_{k_2}\right)\right] \\
&= \min_{\substack{\{\alpha\in\mathbb{R}^{K_1},\ \varphi\in\mathbb{R}^{K_2} \\ \alpha\neq 0,\ \varphi\neq 0\}}} \frac{\mathrm{Tr}\left[M_{f^0}\ \left(X'_{\mathrm{low},\varphi}+X'_{\mathrm{high},\alpha}\right)\ M_{\lambda^0}\ \left(X_{\mathrm{low},\varphi}+X_{\mathrm{high},\alpha}\right)\right]}{NT\ \left(\|\alpha\|^2+\|\varphi\|^2\right)} \ ,
\end{aligned}
$$

where we decomposed $a=(\varphi',\alpha')'$, with φ and α being vectors of length K_1 and K_2, respectively, and we defined linear combinations of high- and low-rank regressors[2]

$$
X_{\mathrm{low},\varphi} = \sum_{l=1}^{K_1} \varphi_l\, X_l \ , \qquad\qquad X_{\mathrm{high},\alpha} = \sum_{m=K_1+1}^{K} \alpha_m\, X_m \ .
$$

We have $M_{\lambda^0} = M_{(\lambda^0,w)} + P_{(M_{\lambda^0}w)}$, where w is the $N\times K_1$ matrix defined in assumption 4, *i.e.* (λ^0,w) is an $N\times(R+K_1)$ matrix, while $M_{\lambda^0}w$ is also an $N\times K_1$ matrix. Using this we obtain

$$
\begin{aligned}
&\mu_K\left(W_{NT}\right) \\
&= \min_{\substack{\{\varphi\in\mathbb{R}^{K_1},\ \alpha\in\mathbb{R}^{K_2} \\ \varphi\neq 0,\ \alpha\neq 0\}}} \frac{1}{NT\ \left(\|\varphi\|^2+\|\alpha\|^2\right)} \Bigg\{ \mathrm{Tr}\left[M_{f^0}\ \left(X'_{\mathrm{low},\varphi}+X'_{\mathrm{high},\alpha}\right)\ M_{(\lambda^0,w)}\ \left(X_{\mathrm{low},\varphi}+X_{\mathrm{high},\alpha}\right)\right] \\
&\qquad\qquad + \mathrm{Tr}\left[M_{f^0}\ \left(X'_{\mathrm{low},\varphi}+X'_{\mathrm{high},\alpha}\right)\ P_{(M_{\lambda^0}w)}\ \left(X_{\mathrm{low},\varphi}+X_{\mathrm{high},\alpha}\right)\right] \Bigg\} \\
&= \min_{\substack{\{\varphi\in\mathbb{R}^{K_1},\ \alpha\in\mathbb{R}^{K_2} \\ \varphi\neq 0,\ \alpha\neq 0\}}} \frac{1}{NT\ \left(\|\varphi\|^2+\|\alpha\|^2\right)} \Bigg\{ \mathrm{Tr}\left[M_{f^0}\, X'_{\mathrm{high},\alpha}\, M_{(\lambda^0,w)}\, X_{\mathrm{high},\alpha}\right] \\
&\qquad\qquad + \mathrm{Tr}\left[M_{f^0}\ \left(X'_{\mathrm{low},\varphi}+X'_{\mathrm{high},\alpha}\right)\ P_{(M_{\lambda^0}w)}\ \left(X_{\mathrm{low},\varphi}+X_{\mathrm{high},\alpha}\right)\right] \Bigg\} \ .
\end{aligned}
\tag{S.6.1}
$$

We note that there exists finite positive constants c_1, c_2, c_3 such that

$$
\begin{aligned}
\frac{1}{NT}\mathrm{Tr}\left[M_{f^0}\, X'_{\mathrm{high},\alpha}\, M_{(\lambda^0,w)}\, X_{\mathrm{high},\alpha}\right] &\geq\ c_1\|\alpha\|^2 \ , \quad \text{wpa1}, \\
\frac{1}{NT}\mathrm{Tr}\left[M_{f^0}\ \left(X'_{\mathrm{low},\varphi}+X'_{\mathrm{high},\alpha}\right)\ P_{(M_{\lambda^0}w)}\ \left(X_{\mathrm{low},\varphi}+X_{\mathrm{high},\alpha}\right)\right] &\geq 0 \ , \\
\frac{1}{NT}\mathrm{Tr}\left[M_{f^0}\, X'_{\mathrm{low},\varphi}\, P_{(M_{\lambda^0}w)}\, X_{\mathrm{low},\varphi}\right] &\geq\ c_2\,\|\varphi\|^2 \ , \quad \text{wpa1}, \\
\frac{1}{NT}\mathrm{Tr}\left[M_{f^0}\, X'_{\mathrm{low},\varphi}\, P_{(M_{\lambda^0}w)}\, X_{\mathrm{high},\alpha}\right] &\geq -\frac{c_3}{2}\,\|\varphi\|\|\alpha\| \ , \quad \text{wpa1}, \\
\frac{1}{NT}\mathrm{Tr}\left[M_{f^0}\, X'_{\mathrm{high},\alpha}\, P_{(M_{\lambda^0}w)}\, X_{\mathrm{high},\alpha}\right] &\geq 0 \ ,
\end{aligned}
\tag{S.6.2}
$$

[2] As in assumption 4 the components of α are denoted $\alpha_{K_1+1},\ldots,\alpha_K$ to simplify notation.

and we want to justify these inequalities now. The second and the last equation in (S.6.2) are true because *e.g.* $\text{Tr}\left[M_{f^0}\, X'_{\text{high},\alpha}\, P_{(M_{\lambda^0} w)}\, X_{\text{high},\alpha}\right] = \text{Tr}\left[M_{f^0}\, X'_{\text{high},\alpha}\, P_{(M_{\lambda^0} w)}\, X_{\text{high},\alpha}\, M_{f^0}\right]$, and the trace of a symmetric positive semi-definite matrix is non-negative. The first inequality in (S.6.2) is true because $\text{rank}(f^0) + \text{rank}(\lambda^0, w) = 2R + K_1$ and using Lemma A.1 and assumption 4 we have

$$\frac{1}{NT\|\alpha\|^2}\text{Tr}\left[M_{f^0}\, X'_{\text{high},\alpha}\, M_{(\lambda^0,w)}\, X_{\text{high},\alpha}\right] \geq \frac{1}{NT\|\alpha\|^2}\mu_{2R+K_1+1}\left[X_{\text{high},\alpha}\, X'_{\text{high},\alpha}\right] > b\,, \quad \text{wpa1},$$

i.e. we can set $c_1 = b$. The third inequality in (S.6.2) is true because according Lemma S.4.1(v) we have

$$\begin{aligned}
\frac{1}{NT}\text{Tr}\left[M_{f^0}\, X'_{\text{low},\varphi}\, P_{(M_{\lambda^0} w)}\, X_{\text{high},\alpha}\right] &\geq -\frac{K_1}{NT}\, \|X_{\text{low},\varphi}\|\, \|X_{\text{high},\alpha}\| \\
&\geq -\frac{K_1}{NT}\, \|X_{\text{low},\varphi}\|_F\, \|X_{\text{high},\alpha}\|_F \\
&\geq -K_1\, K_1\, K_2\, \|\varphi\|\, \|\alpha\| \max_{k_1=1\ldots K_1}\left\|\frac{X_{k_1}}{\sqrt{NT}}\right\|_F \max_{k_2=K_1+1\ldots K}\left\|\frac{X_{k_2}}{\sqrt{NT}}\right\|_F \\
&\geq -\frac{c_3}{2}\, \|\varphi\|\, \|\alpha\|\,,
\end{aligned}$$

where we used that assumption 4 implies that $\left\|X_k/\sqrt{NT}\right\|_F < C$ holds wpa1 for some constant C as, and we set $c_3 = K_1\, K_1\, K_2\, C^2$. Finally, we have to argue that the third inequality in (S.6.2) holds. Note that $X'_{\text{low},\varphi}\, P_{(M_{\lambda^0} w)}\, X_{\text{low},\varphi} = X'_{\text{low},\varphi}\, M_{\lambda^0}\, X_{\text{low},\varphi}$, *i.e.* we need to show that

$$\frac{1}{NT}\text{Tr}\left[M_{f^0}\, X'_{\text{low},\varphi}\, M_{\lambda^0}\, X_{\text{low},\varphi}\right] \geq\; c_2\, \|\varphi\|^2\,.$$

Using part (vi) or Lemma S.4.1 we find

$$\begin{aligned}
\frac{1}{NT}\text{Tr}\left[M_{f^0}\, X'_{\text{low},\varphi}\, M_{\lambda^0}\, X_{\text{low},\varphi}\right] &= \frac{1}{NT}\text{Tr}\left[M_{\lambda^0}\, X_{\text{low},\varphi}\, M_{f^0}\, X'_{\text{low},\varphi}\, M_{\lambda^0}\right] \\
&\geq \frac{1}{NT}\left\|M_{\lambda^0}\, X_{\text{low},\varphi}\, M_{f^0}\, X'_{\text{low},\varphi}\, M_{\lambda^0}\right\|\,,
\end{aligned}$$

and according to Lemma S.5.2 this expression is bounded by some positive constant times $\|\varphi\|^2$ (in the lemma we have $\|\varphi\| = 1$, but all expressions are homogeneous in $\|\varphi\|$).

Using the inequalities (S.6.2) in equation (S.6.1) we obtain

$$\begin{aligned}
\mu_K\left(W_{NT}\right) \geq & \min_{\substack{\{\varphi\, \in\, \mathbb{R}^{K_1},\ \alpha\, \in\, \mathbb{R}^{K_2} \\ \varphi \neq 0,\ \alpha \neq 0\}}} \frac{1}{\|\varphi\|^2 + \|\alpha\|^2}\left\{c_1\|\alpha\|^2 + \max\left[0,\, c_2\|\varphi\|^2 - c_3\|\varphi\|\|\alpha\|\right]\right\} \\
&\geq \min\left(\frac{c_2}{2},\, \frac{c_1 c_2^2}{c_2^2 + c_3^2}\right)\,, \quad \text{wpa1}.
\end{aligned}$$

Thus, the smallest eigenvalue of W_{NT} is bounded from below by a positive constant as $N, T \to \infty$, *i.e.* W_{NT} is non-degenerate and invertible. ∎

S.7 Proof of Examples for Assumption 5

Proof of Example 1. We want to show that the conditions of Assumption 5 are satisfied. Conditions (i)-(iii) immediately follow by the assumptions of the example.

For condition (iv), notice that $\text{Cov}\left(X_{it}, X_{is}|\mathcal{C}\right) = \mathbb{E}\left(U_{it}U_{is}\right)$. Since $|\beta^0| < 1$ and $\sup_{it}\mathbb{E}(e_{it}^2) < \infty$, it follows that

$$\begin{aligned}
\frac{1}{NT}\sum_{i=1}^{N}\sum_{t,s=1}^{T}\left|\text{Cov}\left(X_{it}, X_{is}|\mathcal{C}\right)\right| &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t,s=1}^{T}\left|\mathbb{E}\left(U_{it}U_{is}\right)\right| \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t,s=1}^{T}\sum_{p,q=0}^{\infty}\left|(\beta^0)^{p+q}\mathbb{E}\left(e_{it-p}e_{is-q}\right)\right| < \infty.
\end{aligned}$$

For condition (v), notice by the independence between the sigma field $\mathcal{C}$ and the error terms $\{e_{it}\}$ that we have for some finite constant M,

$$\begin{aligned}
& \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t,s,u,v=1}^{T}\left|\text{Cov}\left(e_{it}\widetilde{X}_{is}, e_{iu}\widetilde{X}_{iv}|\mathcal{C}\right)\right| \\
= & \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t,s,u,v=1}^{T}\left|\text{Cov}\left(e_{it}U_{is}, e_{iu}U_{iv}\right)\right| \\
= & \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t,s,u,v=1}^{T}\sum_{p,q=0}^{\infty}\left|\left(\beta^0\right)^{p+q}\mathbb{E}\left(e_{it}e_{is-p}e_{iu}e_{iv-q}\right) - \left(\beta^0\right)^{p}\mathbb{E}\left(e_{it}e_{is-p}\right)\left(\beta^0\right)^{q}\mathbb{E}\left(e_{iu}e_{iv-q}\right)\right| \\
\leq & \frac{M}{T^2}\sum_{t,s,u,v=1}^{T}\sum_{p,q=0}^{\infty}\left|\beta^0\right|^{p+q}\left[\mathbb{I}\left\{t=u\right\}\mathbb{I}\left\{s-p=v-q\right\} + \mathbb{I}\left\{t=v-q\right\}\mathbb{I}\left\{s-p=u\right\}\right] \\
= & \frac{M}{T^2}\sum_{t,u,s,v=1}^{T}\sum_{k=-\infty}^{s}\sum_{l=-\infty}^{v}\left|\beta^0\right|^{s-k+v-l}\mathbb{I}\left\{t=u\right\}\mathbb{I}\left\{k=l\right\} + M\left(\frac{1}{T}\sum_{\substack{s,u=1\\ s-u\geq 0}}^{T}\left|\beta^0\right|^{s-u}\right)\left(\frac{1}{T}\sum_{\substack{v,t=1\\ v-t\geq 0}}^{T}\left|\beta^0\right|^{v-t}\right) \\
= & \frac{M}{T}\sum_{s,v=1}^{T}\sum_{k=-\infty}^{\min\{s,v\}}\left|\beta^0\right|^{s+v-2k} + M\left(\frac{1}{T}\sum_{\substack{s,u=1\\ s-u\geq 0}}^{T}\left|\beta^0\right|^{s-u}\right)\left(\frac{1}{T}\sum_{\substack{v,t=1\\ v-t\geq 0}}^{T}\left|\beta^0\right|^{v-t}\right).
\end{aligned}$$

Notice that

$$
\begin{aligned}
& \frac{1}{T}\sum_{s,v=1}^{T}\sum_{k=-\infty}^{\min\{s,v\}}\left|\beta^0\right|^{s+v-2k} \\
= & \frac{2}{T}\sum_{s=2}^{T}\sum_{v=1}^{s}\sum_{k=-\infty}^{v}\left|\beta^0\right|^{s-v+2(v-k)}+\frac{2}{T}\sum_{s=1}^{T}\sum_{k=-\infty}^{s}\left|\beta^0\right|^{2(s-k)} \\
= & \frac{2}{T}\sum_{s=2}^{T}\sum_{v=1}^{s}\left|\beta^0\right|^{s-v}\sum_{l=0}^{\infty}\left|\beta^0\right|^{2l}+\frac{2}{T}\sum_{s=1}^{T}\sum_{l=0}^{\infty}\left|\beta^0\right|^{2l} \\
= & \frac{2}{1-\left|\beta^0\right|^2}\frac{1}{T}\sum_{s=2}^{T}\sum_{v=1}^{s}\left|\beta^0\right|^{s-v}+\frac{2}{1-\left|\beta^0\right|^2} \\
= & \left(\frac{2}{1-\left|\beta^0\right|^2}\right)\sum_{l=1}^{T-1}\left|\beta^0\right|^{l}\left(1-\frac{l}{T}\right)+\frac{2}{1-\left|\beta^0\right|^2} \\
= & O(1),
\end{aligned}
$$

and

$$
\frac{1}{T}\sum_{\substack{s,u=1\\ s-u\geq 0}}^{T}\left|\beta^0\right|^{s-u}=\frac{1}{T}\sum_{s=1}^{T}\sum_{u=1}^{s}\left|\beta^0\right|^{s-u}=\sum_{l=0}^{T-1}\left|\beta^0\right|^{l}\left(1-\frac{l}{T}\right)=O(1).
$$

Therefore, we have the desired result that

$$
\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t,s,u,v=1}^{T}\left|\mathrm{Cov}\left(e_{it}\widetilde{X}_{is},e_{iu}\widetilde{X}_{iv}|\mathcal{C}\right)\right|=\mathcal{O}_p(1).
$$

∎

Preliminaries for Proof of Example 2

- Although we observe X_{it} for $1\leq t\leq T$, here we treat that $Z_{it}=(e_{it},X_{it})$ has infinite past and future over time. Define

$$
\mathcal{G}_{\tau}^{t}(i)=\mathcal{C}\vee\sigma\left(\{X_{is}:\tau\leq s\leq t\}\right)\text{ and }\mathcal{H}_{\tau}^{t}(i)=\mathcal{C}\vee\sigma\left(\{Z_{it}:\tau\leq s\leq t\}\right).
$$

  Then, by definition, we have $\mathcal{G}_{\tau}^{t}(i),\mathcal{H}_{\tau}^{t}(i)\subset\mathcal{F}_{\tau}^{t}(i)$ for all τ,t,i. By Assumption (iv) of Example 2, the time series of $\{X_{it}:-\infty<t<\infty\}$ and $\{Z_{it}:-\infty<t<\infty\}$ are conditional $\alpha-$ mixing conditioning on $\mathcal{C}$ uniformly in i.

- Mixing inequality: The following inequality is a conditional version of the α-mixing inequality of Hall and Heyde (1980, p. 278). Suppose that X_{it} is a $\mathcal{F}_t$-measurable random variable with $\mathbb{E}\left(|X_{it}|^{\max\{p,q\}}|\mathcal{C}\right)<\infty$, where $p,q>1$ with $1/p+1/q<1$. Denote

$\|X_{it}\|_{\mathcal{C},p} = \left(\mathbb{E}\left(|X_{it}|^p |\mathcal{C}\right)\right)^{1/p}$. Then, for each i, we have

$$\left|\mathrm{Cov}\left(X_{it}, X_{it+m}|\mathcal{C}\right)\right| \le 8 \|X_{it}\|_{\mathcal{C},p} \|X_{it+m}\|_{\mathcal{C},q} \alpha_m^{1-\frac{1}{p}-\frac{1}{q}}(i). \qquad \text{(S.7.1)}$$

Proof of Example 2. Again, we want to show that the conditions of Assumption 5 are satisfied. Conditions (i)-(iii) immediately follow by the assumptions of the example.

For condition (iv), we apply the mixing inequality $(S.7.1)$ with $p = q > 4$. Then, we have

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t,s=1}^{T}\left|\mathrm{Cov}\left(X_{it}, X_{is}|\mathcal{C}\right)\right| \\
\le\ & \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{m=0}^{T-t}\left|\mathrm{Cov}\left(X_{it}, X_{it+m}|\mathcal{C}\right)\right| = \frac{2}{NT}\sum_{i=1}^{N}\sum_{m=0}^{T-1}\sum_{t=1}^{T-m}\left|\mathrm{Cov}\left(X_{it}, X_{it+m}|\mathcal{C}\right)\right| \\
=\ & \frac{16}{NT}\sum_{i=1}^{N}\sum_{m=0}^{T-1}\sum_{t=1}^{T-m}\|X_{it}\|_{\mathcal{C},p}\|X_{it+m}\|_{\mathcal{C},p}\,\alpha_m(i)^{\frac{p-2}{P}} \\
\le\ & 16\left(\sup_{i,t}\|X_{it}\|_{\mathcal{C},p}^2\right)\sum_{m=0}^{\infty}\alpha_m^{\frac{p-2}{P}} \\
\le\ & \mathcal{O}_p(1),
\end{aligned}$$

where the last line holds since $\sup_{i,t}\|X_{it}\|_{\mathcal{C},p}^2 = \mathcal{O}_p(1)$ for some $p > 4$ as assumed in the example (2), and $\sum_{m=0}^{\infty}\alpha_m^{\frac{p-2}{P}} = \sum_{m=0}^{\infty} m^{-\zeta\frac{p-2}{P}} = \mathcal{O}(1)$ due to $\zeta > 3\frac{4p}{4p-1}$ and $p > 4$.

For condition (v), we need to show

$$\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t,s,u,v=1}^{T}\left|\mathrm{Cov}\left(e_{it}\widetilde{X}_{is}, e_{iu}\widetilde{X}_{iv}|\mathcal{C}\right)\right| = \mathcal{O}_p(1).$$

Notice that

$$\begin{aligned}
&\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t,s,u,v=1}^{T}\left|\mathrm{Cov}\left(e_{it}\widetilde{X}_{is}, e_{iu}\widetilde{X}_{iv}|\mathcal{C}\right)\right| \\
=\ & \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t,s,u,v=1}^{T}\left|\mathbb{E}\left(e_{it}\widetilde{X}_{is}e_{iu}\widetilde{X}_{iv}|\mathcal{C}\right) - \mathbb{E}\left(e_{it}\widetilde{X}_{is}|\mathcal{C}\right)\mathbb{E}\left(e_{iu}\widetilde{X}_{iv}|\mathcal{C}\right)\right| \\
\le\ & \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t,s,u,v=1}^{T}\left|\mathbb{E}\left(e_{it}\widetilde{X}_{is}e_{iu}\widetilde{X}_{iv}|\mathcal{C}\right)\right| + \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t,s=1}^{T}\mathbb{E}\left(e_{it}\widetilde{X}_{is}|\mathcal{C}\right)\right)^2 \\
=\ & I + II, \text{ say.}
\end{aligned}$$

First, for term I, there are finite number of different orderings among the indices t, s, u, v. We consider the case $t \le s \le u \le v$ and establish the desired result. The rest of the cases are the same. Note that

$$
\begin{aligned}
& \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{k=0}^{T-t}\sum_{l=0}^{T-k}\sum_{m=0}^{T-l}\left|\mathbb{E}\left(e_{it}\widetilde{X}_{it+k}e_{it+k+l}\widetilde{X}_{it+k+l+m}|\mathcal{C}\right)\right| \\
\leq\ & \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{\substack{0\leq l,m\leq k\\ 0\leq k+l+m\leq T-t}}\left|\mathbb{E}\left(e_{it}\left(\widetilde{X}_{it+k}e_{it+k+l}\widetilde{X}_{it+k+l+m}\right)|\mathcal{C}\right)\right| \\
& +\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{\substack{0\leq k,m\leq l\\ 0\leq k+l+m\leq T-t}}\left|\mathbb{E}\left[\left(e_{it}\widetilde{X}_{it+k}\right)\left(e_{it+k+l}\widetilde{X}_{it+k+l+m}\right)|\mathcal{C}\right]\right. \\
& \qquad\qquad\qquad\qquad\qquad\left.-\mathbb{E}\left(e_{it}\widetilde{X}_{it+k}|\mathcal{C}\right)\mathbb{E}\left(e_{it+k+l}\widetilde{X}_{it+k+l+m}|\mathcal{C}\right)\right| \\
& +\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{\substack{0\leq k,m\leq l\\ 0\leq k+l+m\leq T-t}}\mathbb{E}\left(e_{it}\widetilde{X}_{it+k}|\mathcal{C}\right)\mathbb{E}\left(e_{it+k+l}\widetilde{X}_{it+k+l+m}|\mathcal{C}\right) \\
& +\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{\substack{0\leq p,l\leq m\\ 0\leq k+l+m\leq T-t}}\left|\mathbb{E}\left[\left(e_{it}\widetilde{X}_{it+k}e_{it+k+l}\right)\widetilde{X}_{it+k+l+m}|\mathcal{C}\right]\right| \\
=\ & I_1+I_2+I_3+I_4,\ \text{say.}
\end{aligned}
$$

By applying the mixing inequality $(S.7.1)$ to $\left|\mathbb{E}\left(e_{it}\left(\widetilde{X}_{it+k}e_{it+k+l}\widetilde{X}_{it+k+l+m}\right)|\mathcal{C}\right)\right|$ with e_{it} and $\widetilde{X}_{it+k}e_{it+k+l}\widetilde{X}_{it+k+l+m}$, we have

$$
\begin{aligned}
& \left|\mathbb{E}\left(e_{it}\left(\widetilde{X}_{it+k}e_{it+k+l}\widetilde{X}_{it+k+l+m}\right)|\mathcal{C}\right)\right| \\
\leq\ & 8\left\|e_{it}\right\|_{\mathcal{C},p}\left\|\widetilde{X}_{it+k}e_{it+k+l}\widetilde{X}_{it+k+l+m}\right\|_{\mathcal{C},q}\alpha_k^{1-\frac{1}{p}-\frac{1}{q}}(i) \\
\leq\ & 8\left\|e_{it}\right\|_{\mathcal{C},p}\left\|\widetilde{X}_{it+k}\right\|_{\mathcal{C},3q}\left\|e_{it+k+l}\right\|_{\mathcal{C},3q}\left\|\widetilde{X}_{it+k+l+m}\right\|_{\mathcal{C},3q}\alpha_k^{1-\frac{1}{p}-\frac{1}{q}}(i),
\end{aligned}
$$

where the last inequality follows by the generalized Holder's inequality. Choose $p=3q>4$. Then,

$$
\begin{aligned}
I_1 \leq\ & \frac{8}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{\substack{0\leq l,m\leq k\\ 0\leq k+l+m\leq T-t}}\left\|e_{it}\right\|_{\mathcal{C},p}\left\|\widetilde{X}_{it+k}\right\|_{\mathcal{C},p}\left\|e_{it+k+l}\right\|_{\mathcal{C},p}\left\|\widetilde{X}_{it+k+l+m}\right\|_{\mathcal{C},p}\alpha_k^{1-\frac{1}{4p}}(i) \\
\leq\ & 8\left(\sup_{i,t}\left\|e_{it}\right\|_{\mathcal{C},p}^2\right)\left(\sup_{i,t}\left\|\widetilde{X}_{it+k}\right\|_{\mathcal{C},p}^2\right)\frac{1}{T^2}\sum_{t=1}^{T}\sum_{\substack{0\leq l,m\leq k\\ 0\leq k+l+m\leq T-t}}\alpha_k^{1-\frac{1}{4p}} \\
\leq\ & 8\left(\sup_{i,t}\left\|e_{it}\right\|_{\mathcal{C},p}^2\right)\left(\sup_{i,t}\left\|\widetilde{X}_{it+k}\right\|_{\mathcal{C},p}^2\right)\sum_{k=0}^{\infty}k^2\alpha_k^{1-\frac{1}{4p}} \\
\leq\ & \mathcal{O}_p(1),
\end{aligned}
$$

where the last line holds since we assume in the example (2) that $\left(\sup_{i,t}\|e_{it}\|_{\mathcal{C},p}^2\right)\left(\sup_{i,t}\left\|\widetilde{X}_{it+k}\right\|_{\mathcal{C},p}^2\right) = \mathcal{O}_p(1)$ for some $p > 4$,, and $\sum_{m=0}^{\infty} m^2 \alpha_m^{1-\frac{1}{4p}} = \sum_{m=0}^{\infty} m^{2-\zeta\frac{4p-1}{4p}} = O(1)$ due to $\zeta > 3\frac{4p}{4p-1}$ and $p > 4$.

By applying similar argument, we can also show that

$$I_2, I_3, I_4 = \mathcal{O}_p(1) .$$

∎

S.8 Supplement to the Proof of Theorem 4.3

Notation $\mathbb{E}_{\mathcal{C}}$ and $\mathrm{Var}_{\mathcal{C}}$ and $\mathrm{Cov}_{\mathcal{C}}$: In the remainder of this supplementary file we write $\mathbb{E}_{\mathcal{C}}$, $\mathrm{Var}_{\mathcal{C}}$ and $\mathrm{Cov}_{\mathcal{C}}$ for the expectation, variance and covariance operators conditional on $\mathcal{C}$, i.e. $\mathbb{E}_{\mathcal{C}}(A) = \mathbb{E}(A|\mathcal{C})$, $\mathrm{Var}_{\mathcal{C}}(A) = \mathrm{Var}(A|\mathcal{C})$ and $\mathrm{Cov}_{\mathcal{C}}(A, B) = \mathrm{Cov}(A, B|\mathcal{C})$.

What is left to show to complete the proof of Theorem 4.3 is that Lemma B.1 and Lemma B.2 in the main text appendix hold. Before showing this, we first present two further intermediate lemmas.

Lemma S.8.1. *Under the assumptions of Theorem 4.3 we have for $k = 1, \ldots, K$ that*

$$\begin{aligned}
&(a) && \|P_{\lambda^0}\widetilde{X}_k\| = o_p(\sqrt{NT}) \ , \\
&(b) && \|\widetilde{X}_k P_{f^0}\| = o_p(\sqrt{NT}) \ , \\
&(c) && \|P_{\lambda^0} e X_k'\| = o_p(N^{3/2}) \, , \\
&(d) && \|P_{\lambda^0} e P_{f^0}\| = \mathcal{O}_p(1) \ .
\end{aligned}$$

Proof of Lemma S.8.1. # Part (a): We have

$$\begin{aligned}
\|P_{\lambda^0}\widetilde{X}_k\| &= \|\lambda^0(\lambda^{0\prime}\lambda^0)^{-1}\lambda^{0\prime}\widetilde{X}_k\| \\
&\le \|\lambda^0(\lambda^{0\prime}\lambda^0)^{-1}\|\|\lambda^{0\prime}\widetilde{X}_k\| \\
&\le \|\lambda^0\|\|(\lambda^{0\prime}\lambda^0)^{-1}\|\|\lambda^{0\prime}\widetilde{X}_k\|_F = \mathcal{O}_p(N^{-1/2})\|\lambda^{0\prime}\widetilde{X}_k\|_F \ ,
\end{aligned}$$

where we used part (i) and (ii) of Lemma S.4.1 and Assumption 1. We have

$$\begin{aligned}
\mathbb{E}\left\{\mathbb{E}_{\mathcal{C}}\left[\|\lambda^{0\prime}\widetilde{X}_k\|_F^2\right]\right\} &= \mathbb{E}\left\{\sum_{r=1}^{R}\sum_{t=1}^{T}\mathbb{E}_{\mathcal{C}}\left[\left(\sum_{i=1}^{N}\lambda_{ir}^0\widetilde{X}_{k,it}\right)^2\right]\right\} \\
&= \mathbb{E}\left\{\sum_{r=1}^{R}\sum_{t=1}^{T}\sum_{i=1}^{N}(\lambda_{ir}^0)^2\mathbb{E}_{\mathcal{C}}\left(\widetilde{X}_{k,it}^2\right)\right\} \\
&= \sum_{r=1}^{R}\sum_{t=1}^{T}\sum_{i=1}^{N}\mathbb{E}\left[(\lambda_{ir}^0)^2\mathrm{Var}_{\mathcal{C}}\left(X_{k,it}\right)\right] \\
&= \mathcal{O}_p(NT),
\end{aligned}$$

where we used that $\widetilde{X}_{k,it}$ is mean zero and independent across i, conditional on $\mathcal{C}$, and our bounds on the moments of λ_{ir}^0 and $X_{k,it}$. We therefore have $\|\lambda^{0\prime}\widetilde{X}_k\|_F = \mathcal{O}_p(\sqrt{NT})$ and the above inequality thus gives $\|P_{\lambda^0}\widetilde{X}_k\| = \mathcal{O}_p(\sqrt{T}) = o_p(\sqrt{NT})$.

The proof for part (b) is similar. As above we first obtain $\|\widetilde{X}_k P_{f^0}\| = \|P_{f^0}\widetilde{X}_k'\| \leq \mathcal{O}_p(T^{-1/2})\|f^{0\prime}\widetilde{X}_k'\|_F$. Next, we have

$$\begin{aligned}
\mathbb{E}_{\mathcal{C}}\left[\|f^{0\prime}\widetilde{X}_k'\|_F^2\right] &= \sum_{r=1}^{R}\sum_{i=1}^{N}\mathbb{E}_{\mathcal{C}}\left[\left(\sum_{t=1}^{T}f_{tr}^0\widetilde{X}_{k,it}\right)^2\right] \\
&= \sum_{r=1}^{R}\sum_{i=1}^{N}\sum_{t,s=1}^{T}f_{tr}^0 f_{sr}^0\mathbb{E}_{\mathcal{C}}\left(\widetilde{X}_{k,it}\widetilde{X}_{k,is}\right) \\
&\leq \left[\sum_{r=1}^{R}\left(\max_t|f_{tr}^0|\right)^2\right]\sum_{i=1}^{N}\sum_{t,s=1}^{T}\left|\mathrm{Cov}_{\mathcal{C}}\left(X_{k,it},X_{k,is}\right)\right| \\
&= \mathcal{O}_p(T^{2/(4+\epsilon)})\,\mathcal{O}_p(NT) = o_p(NT^2),
\end{aligned}$$

where we used that uniformly bounded $\mathbb{E}\|f_t^0\|^{4+\epsilon}$ implies that $\max_t|f_{tr}^0| = \mathcal{O}_p(T^{1/(4+\epsilon)})$. We thus have $\|f^{0\prime}\widetilde{X}_k'\|_F^2 = o_p(T\sqrt{N})$ and therefore $\|\widetilde{X}_k P_{f^0}\| = o_p(\sqrt{NT})$.

Next, we show part (c). First, we have

$$\begin{aligned}
\mathbb{E}\left\{\mathbb{E}_{\mathcal{C}}\left[\left(\|\lambda^{0\prime}eX_k'\|_F\right)^2\right]\right\} &= \mathbb{E}\left\{\mathbb{E}_{\mathcal{C}}\left[\sum_{r=1}^{R}\sum_{j=1}^{N}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_{ir}^0 e_{it}X_{k,jt}\right)^2\right]\right\} \\
&= \mathbb{E}\left\{\sum_{r=1}^{R}\sum_{i,j,l=1}^{N}\sum_{t,s=1}^{T}\lambda_{ir}^0\lambda_{lr}^0\mathbb{E}_{\mathcal{C}}\left(e_{it}e_{ls}X_{k,jt}X_{k,js}\right)\right\} \\
&= \sum_{r=1}^{R}\sum_{i,j=1}^{N}\sum_{t=1}^{T}\mathbb{E}\left[(\lambda_{ir}^0)^2\mathbb{E}_{\mathcal{C}}\left(e_{it}^2X_{k,jt}^2\right)\right] = \mathcal{O}(N^2T)\ ,
\end{aligned}$$

where we used that $\mathbb{E}_{\mathcal{C}}\left(e_{it}e_{ls}X_{k,jt}X_{k,js}\right)$ is only non-zero if $i = l$ (because of cross-sectional independence conditional on $\mathcal{C}$) and $t = s$ (because regressors are pre-determined). We can thus conclude that $\|\lambda^{0\prime} e X_k'\|_F = \mathcal{O}_p(N\sqrt{T})$. Using this we find

$$
\begin{aligned}
\|P_{\lambda^0} e X_k'\| &= \|\lambda^0(\lambda^{0\prime}\lambda^0)^{-1}\lambda^{0\prime} e X_k'\| \\
&\leq \|\lambda^0(\lambda^{0\prime}\lambda^0)^{-1}\| \|\lambda^{0\prime} e X_k'\| \\
&\leq \|\lambda^0\| \|(\lambda^{0\prime}\lambda^0)^{-1}\| \|\lambda^{0\prime} e X_k'\|_F = \mathcal{O}_p(N^{-1/2})\mathcal{O}_p(N\sqrt{T}) = \mathcal{O}_p(\sqrt{NT}) \,.
\end{aligned}
$$

This is what we wanted to show.

For part (d), we first find that $\frac{1}{\sqrt{NT}}\left\|f^{0\prime} e \lambda^0\right\|_F = \mathcal{O}_p\left(1\right)$, because

$$
\begin{aligned}
\mathbb{E}\left\{\mathbb{E}_{\mathcal{C}}\left[\left(\frac{\left\|f^{0\prime} e \lambda^0\right\|_F}{\sqrt{NT}}\right)^2\right]\right\} &= \mathbb{E}\left\{\frac{1}{NT}\mathbb{E}_{\mathcal{C}}\left[\left(\sum_{i=1}^N\sum_{t=1}^T e_{it} f_t^{0\prime}\lambda_i^0\right)^2\right]\right\} \\
&= \mathbb{E}\left\{\frac{1}{NT}\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^T\sum_{s=1}^T \mathbb{E}_{\mathcal{C}}\left(e_{it}e_{js}\right) f_t^{0\prime}\lambda_i^0\lambda_j^{0\prime} f_s^0\right\} \\
&= \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \mathbb{E}\left[\mathbb{E}_{\mathcal{C}}\left(e_{it}^2\right) f_t^{0\prime}\lambda_i^0\lambda_i^{0\prime} f_t^0\right] \\
&= \mathcal{O}\left(1\right),
\end{aligned}
$$

where we used that e_{it} is independent across i and over t, conditional on $\mathcal{C}$. Thus we obtain

$$
\begin{aligned}
\|P_{\lambda^0} e P_{f^0}\| &= \|\lambda^0(\lambda^{0\prime}\lambda^0)^{-1}\lambda^{0\prime} e f^0 (f^{0\prime} f^0)^{-1} f^{0\prime}\| \\
&\leq \|\lambda^0\| \left\|(\lambda^{0\prime}\lambda^0)^{-1}\right\| \|\lambda^{0\prime} e f^0\| \left\|(f^{0\prime} f^0)^{-1}\right\| \|f^{0\prime}\| \\
&\leq \mathcal{O}_p(N^{1/2})\mathcal{O}_p(N^{-1})\|\lambda^{0\prime} e f^0\|_F \mathcal{O}_p(T^{-1})\mathcal{O}_p(T^{1/2}) = \mathcal{O}_p(1) \,,
\end{aligned}
$$

where we used part (i) and (ii) of Lemma S.4.1. ∎

Lemma S.8.2. *Suppose that A and B are a $T \times T$ and an $N \times N$ matrices that are independent of e, conditional on $\mathcal{C}$, such that $\mathbb{E}_{\mathcal{C}}\left(\|A\|_F^2\right) = \mathcal{O}_p\left(NT\right)$ and $\mathbb{E}_{\mathcal{C}}\left(\|B\|_F^2\right) = \mathcal{O}_p\left(NT\right)$, and let Assumption 5 be satisfied. Then there exists a finite non-random constant c_0 such that*

$$
\begin{aligned}
&(a) \qquad \mathbb{E}_{\mathcal{C}}\left(\left\{\mathrm{Tr}\left[\left(e'e - \mathbb{E}_{\mathcal{C}}\left(e'e\right)\right)A\right]\right\}^2\right) \leq c_0\, N\, \mathbb{E}_{\mathcal{C}}\left(\|A\|_F^2\right) \,, \\
&(b) \qquad \mathbb{E}_{\mathcal{C}}\left(\left\{\mathrm{Tr}\left[\left(ee' - \mathbb{E}_{\mathcal{C}}\left(ee'\right)\right)B\right]\right\}^2\right) \leq c_0\, T\, \mathbb{E}_{\mathcal{C}}\left(\|B\|_F^2\right) \,.
\end{aligned}
$$

Proof. # Part (a): Denote A_{ts} to be the $(t,s)^{th}$ element of A. We have

$$\begin{aligned}\mathrm{Tr}\left\{\left(e'e-\mathbb{E}_{\mathcal{C}}\left(e'e\right)\right)A\right\} &= \sum_{t=1}^{T}\sum_{s=1}^{T}\left(e'e-\mathbb{E}_{\mathcal{C}}\left(e'e\right)\right)_{ts}A_{ts}\\ &= \sum_{t=1}^{T}\sum_{s=1}^{T}\left(\sum_{i=1}^{N}\left(e_{it}e_{is}-\mathbb{E}_{\mathcal{C}}\left(e_{it}e_{is}\right)\right)\right)A_{ts}.\end{aligned}$$

Therefore

$$\begin{aligned}&\mathbb{E}_{\mathcal{C}}\left(\mathrm{Tr}\left\{\left(e'e-\mathbb{E}_{\mathcal{C}}\left(e'e\right)\right)A\right\}\right)^2\\ &\quad=\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{p=1}^{T}\sum_{q=1}^{T}\mathbb{E}_{\mathcal{C}}\left[\left(\sum_{i=1}^{N}\left(e_{it}e_{is}-\mathbb{E}_{\mathcal{C}}\left(e_{it}e_{is}\right)\right)\right)\left(\sum_{j=1}^{N}\left(e_{jp}e_{jq}-\mathbb{E}_{\mathcal{C}}\left(e_{jp}e_{jq}\right)\right)\right)\right]\mathbb{E}_{\mathcal{C}}\left(A_{ts}A_{pq}\right).\end{aligned}$$

Let $\Sigma_{it}=\mathbb{E}_{\mathcal{C}}(e_{it}^2)$. Then we find

$$\begin{aligned}&\mathbb{E}_{\mathcal{C}}\left\{\left(\sum_{i=1}^{N}\left(e_{it}e_{is}-\mathbb{E}_{\mathcal{C}}\left(e_{it}e_{is}\right)\right)\right)\left(\sum_{j=1}^{N}\left(e_{jp}e_{jq}-\mathbb{E}_{\mathcal{C}}\left(e_{jp}e_{jq}\right)\right)\right)\right\}\\ &\qquad=\sum_{i=1}^{N}\sum_{j=1}^{N}\left\{\mathbb{E}_{\mathcal{C}}\left(e_{it}e_{is}e_{jp}e_{jq}\right)-\mathbb{E}_{\mathcal{C}}\left(e_{it}e_{is}\right)\mathbb{E}_{\mathcal{C}}\left(e_{jp}e_{jq}\right)\right\}\\ &\qquad=\begin{cases}\Sigma_{it}\Sigma_{is} & \text{if } (t=p)\neq(s=q) \text{ and } (i=j)\\ \Sigma_{it}\Sigma_{is} & \text{if } (t=q)\neq(s=p) \text{ and } (i=j)\\ \mathbb{E}_{\mathcal{C}}\left(e_{it}^4\right)-\Sigma_{it}^2 & \text{if } (t=s=p=q) \text{ and } (i=j)\\ 0 & \text{otherwise.}\end{cases}\end{aligned}$$

Therefore,

$$\mathbb{E}_{\mathcal{C}}\left(\mathrm{Tr}\left\{\left(e'e-\mathbb{E}_{\mathcal{C}}\left(e'e\right)\right)A\right\}\right)^2\leq\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\Sigma_{it}\Sigma_{is}\left(\mathbb{E}_{\mathcal{C}}\left(A_{ts}^2\right)+\mathbb{E}_{\mathcal{C}}\left(A_{ts}A_{st}\right)\right)+\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\mathbb{E}_{\mathcal{C}}\left(e_{it}^4\right)-\Sigma_{it}^2\right)\mathbb{E}_{\mathcal{C}}A_{tt}^2.$$

Define $\Sigma^i=\mathrm{diag}\left(\Sigma_{i1},...,\Sigma_{iT}\right)$. Then, we have

$$\begin{aligned}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\Sigma_{it}\Sigma_{is}\left(\mathbb{E}_{\mathcal{C}}A_{ts}^2\right) &= \mathbb{E}_{\mathcal{C}}\left(\sum_{i=1}^{N}\mathrm{Tr}\left(A'\Sigma^iA\Sigma^i\right)\right)\\ &\leq \sum_{i=1}^{N}\mathbb{E}_{\mathcal{C}}\left\|A\Sigma^i\right\|_F^2\leq\sum_{i=1}^{N}\left\|\Sigma^i\right\|^2\mathbb{E}_{\mathcal{C}}\left\|A\right\|_F^2\\ &\leq N\left(\sup_{it}\Sigma_{it}^2\right)\mathbb{E}_{\mathcal{C}}\left\|A\right\|_F^2. \qquad \text{(S.8.1)}\end{aligned}$$

Also,

$$\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\Sigma_{it}\Sigma_{is}\mathbb{E}_{\mathcal{C}}\left(A_{ts}A_{st}\right) = \mathbb{E}_{\mathcal{C}}\left[\sum_{i=1}^{N}\operatorname{Tr}\left(\Sigma^{i}AA\Sigma^{i}\right)\right]$$
$$\leq \sum_{i=1}^{N}\mathbb{E}_{\mathcal{C}}\left\|\Sigma^{i}A\right\|_{F}\left\|A\Sigma^{i}\right\|_{F} \leq \sum_{i=1}^{N}\left\|\Sigma^{i}\right\|^{2}\mathbb{E}_{\mathcal{C}}\left\|A\right\|_{F}^{2}$$
$$\leq N\left(\sup_{it}\Sigma_{it}^{2}\right)\mathbb{E}_{\mathcal{C}}\left\|A\right\|_{F}^{2}\ . \tag{S.8.2}$$

Finally,

$$\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\mathbb{E}_{\mathcal{C}}\left(e_{it}^{4}\right)-\Sigma_{it}^{2}\right)\mathbb{E}_{\mathcal{C}}A_{tt}^{2} \leq N\left(\sup_{it}\mathbb{E}_{\mathcal{C}}\left(e_{it}^{4}\right)\right)\mathbb{E}_{\mathcal{C}}\left\|A\right\|_{F}^{2}, \tag{S.8.3}$$

and $\sup_{it}\mathbb{E}_{\mathcal{C}}\left(e_{it}^{4}\right)$ is assumed bounded by Assumption 5(vi).

Part (b): The proof is analogous to that of part (a). ∎

Proof of Lemma B.1. # For part (a) we have

$$\left|\frac{1}{\sqrt{NT}}\operatorname{Tr}\left(P_{f^0}\,e'\,P_{\lambda^0}\,\widetilde{X}_k\right)\right| = \left|\frac{1}{\sqrt{NT}}\operatorname{Tr}\left(P_{f^0}\,e'\,P_{\lambda^0}P_{\lambda^0}\widetilde{X}_kP_{f^0}\right)\right|$$
$$\leq \frac{R}{\sqrt{NT}}\left\|P_{\lambda^0}\,e\,P_{f^0}\right\|\left\|P_{\lambda^0}\widetilde{X}_k\right\|\left\|P_{f^0}\right\|$$
$$= \frac{1}{\sqrt{NT}}\,\mathcal{O}_p(1)\,o_p(\sqrt{NT})\,\mathcal{O}_p(1)$$
$$= o_p(1),$$

where the the second last equality follows by Lemma S.8.1 (a) and (d).

To show statement (b) we define $\zeta_{k,ijt}=e_{it}\widetilde{X}_{k,jt}$. We then have

$$\frac{1}{\sqrt{NT}}\operatorname{Tr}\left(P_{\lambda^0}\,e\,\widetilde{X}_k'\right) = \sum_{r,q=1}^{R}\left[\left(\frac{\lambda^{0\prime}\lambda^{0}}{N}\right)^{-1}\right]_{rq}\underbrace{\frac{1}{N\sqrt{NT}}\sum_{t=1}^{T}\sum_{i,j=1}^{N}\lambda_{ir}^{0}\lambda_{jq}^{0}\zeta_{k,ijt}}_{\equiv A_{k,rq}}.$$

We only have $\mathbb{E}_{\mathcal{C}}\left(\zeta_{k,ijt}\zeta_{k,lms}\right)\neq 0$ if $t=s$ (because regressors are pre-determined) and $i=l$ and $j=m$ (because of cross-sectional independence). Therefore

$$\mathbb{E}\left\{\mathbb{E}_{\mathcal{C}}\left(A_{k,rq}^{2}\right)\right\} = \mathbb{E}\left\{\frac{1}{N^3T}\sum_{t,s=1}^{T}\sum_{i,j,l,m=1}^{N}\lambda_{ir}\lambda_{jq}\lambda_{lr}\lambda_{mq}\,\mathbb{E}_{\mathcal{C}}\left(\zeta_{k,ijt}\zeta_{k,lms}\right)\right\}$$
$$= \frac{1}{N^3T}\sum_{t=1}^{T}\sum_{i,j=1}^{N}\mathbb{E}\left[\lambda_{ir}^{2}\lambda_{jq}^{2}\,\mathbb{E}_{\mathcal{C}}\left(\zeta_{k,ijt}^{2}\right)\right] = \mathcal{O}(1/N) = o_p(1).$$

We thus have $A_{k,rq} = o_p(1)$ and therefore also $\frac{1}{\sqrt{NT}} \mathrm{Tr}\left(P_{\lambda^0}\, e\, \widetilde{X}_k'\right) = o_p(1)$.

The proof for statement (c) is similar to that of statement (b). Define $\xi_{k,its} = e_{it}\widetilde{X}_{k,is} - \mathbb{E}_{\mathcal{C}}\left(e_{it}\widetilde{X}_{k,is}\right)$. We then have

$$\frac{1}{\sqrt{NT}} \mathrm{Tr}\left\{P_{f^0}\, \left[e'\, \widetilde{X}_k - \mathbb{E}_{\mathcal{C}}\left(e'\, \widetilde{X}_k\right)\right]\right\} = \sum_{r,q=1}^{R} \left[\left(\frac{f'f}{T}\right)^{-1}\right]_{rq} \underbrace{\frac{1}{T\sqrt{NT}} \sum_{i=1}^{N} \sum_{t,s=1}^{T} f_{tr} f_{sq} \xi_{k,its}}_{\equiv B_{k,rq}} .$$

Therefore

$$\begin{aligned}
\mathbb{E}_{\mathcal{C}}\left(B_{k,rq}^2\right) &= \frac{1}{T^3 N} \sum_{i,j=1}^{N} \sum_{t,s,u,v=1}^{T} f_{tr} f_{sq} f_{ur} f_{vq} \mathbb{E}_{\mathcal{C}}\left(\xi_{k,its}\xi_{k,juv}\right) \\
&\leq \left(\max_{t,\widetilde{r}} |f_{t\widetilde{r}}|\right)^4 \frac{1}{T^3 N} \sum_{i,j=1}^{N} \sum_{t,s,u,v=1}^{T} \left|\mathrm{Cov}_{\mathcal{C}}\left(e_{it}\widetilde{X}_{k,is}, e_{ju}\widetilde{X}_{k,jv}\right)\right| \\
&= \left(\max_{t,\widetilde{r}} |f_{t\widetilde{r}}|\right)^4 \frac{1}{T^3 N} \sum_{i=1}^{N} \sum_{t,s,u,v=1}^{T} \left|\mathrm{Cov}_{\mathcal{C}}\left(e_{it}\widetilde{X}_{k,is}, e_{iu}\widetilde{X}_{k,iv}\right)\right| \\
&= \mathcal{O}_p(T^{4/(4+\epsilon)}) \mathcal{O}_p(1/T) \\
&= o_p(1),
\end{aligned}$$

where we used that that uniformly bounded $\mathbb{E}\|f_t^0\|^{4+\epsilon}$ implies that $\max_t |f_{tr}^0| = \mathcal{O}_p(T^{1/(4+\epsilon)})$.

Part (d) and (e): We have $\|\lambda^0\,(\lambda^{0\prime}\lambda^0)^{-1}\,(f^{0\prime}f^0)^{-1}\,f^{0\prime}\| = \mathcal{O}_p((NT)^{-1/2})$, $\|e\| = \mathcal{O}_p(N^{1/2})$, $\|X_k\| = \mathcal{O}_p(\sqrt{NT})$ and $\|P_{\lambda^0} e P_{f^0}\| = \mathcal{O}_p(1)$, which was shown in Lemma S.8.1. Therefore:

$$\begin{aligned}
&\frac{1}{\sqrt{NT}} \mathrm{Tr}\left(e P_{f^0}\, e'\, M_{\lambda^0}\, X_k\, f^0\,(f^{0\prime}f^0)^{-1}\,(\lambda^{0\prime}\lambda^0)^{-1}\,\lambda^{0\prime}\right) \\
&\qquad = \frac{1}{\sqrt{NT}} \mathrm{Tr}\left(P_{\lambda^0} e P_{f^0}\, e'\, M_{\lambda^0}\, X_k\, f^0\,(f^{0\prime}f^0)^{-1}\,(\lambda^{0\prime}\lambda^0)^{-1}\,\lambda^{0\prime}\right) \\
&\qquad \leq \frac{R}{\sqrt{NT}} \|P_{\lambda^0} e P_{f^0}\|\, \|e\| \|X_k\|\, \left\|f^0\,(f^{0\prime}f^0)^{-1}\,(\lambda^{0\prime}\lambda^0)^{-1}\,\lambda^{0\prime}\right\| = \mathcal{O}_p(N^{-1/2}) = o_p(1) .
\end{aligned}$$

which shows statement (d). The proof for part (e) is analogous.

To prove statement (f) we need to use additionally $\|P_{\lambda^0}\, e\, X_k'\| = o_p(N^{3/2})$, which was also

shown in Lemma S.8.1. We find

$$
\begin{aligned}
&\frac{1}{\sqrt{NT}} \mathrm{Tr}\left(e' M_{\lambda^0}\, X_k\, M_{f^0}\, e'\, \lambda^0\, (\lambda^{0\prime}\lambda^0)^{-1}\, (f^{0\prime} f^0)^{-1}\, f^{0\prime}\right) \\
&\qquad = \frac{1}{\sqrt{NT}} \mathrm{Tr}\left(e' M_{\lambda^0}\, X_k\, e'\, P_{\lambda^0}\, \lambda^0\, (\lambda^{0\prime}\lambda^0)^{-1}\, (f^{0\prime} f^0)^{-1}\, f^{0\prime}\right) \\
&\qquad\quad - \frac{1}{\sqrt{NT}} \mathrm{Tr}\left(e' M_{\lambda^0}\, X_k\, P_{f^0}\, e'\, P_{\lambda^0}\, \lambda^0\, (\lambda^{0\prime}\lambda^0)^{-1}\, (f^{0\prime} f^0)^{-1}\, f^{0\prime}\right) \\
&\qquad \leq \frac{R}{\sqrt{NT}} \|e\| \|P_{\lambda^0}\, e\, X_k'\|\, \|\lambda^0\, (\lambda^{0\prime}\lambda^0)^{-1}\, (f^{0\prime} f^0)^{-1}\, f^{0\prime}\| \\
&\qquad\quad - \frac{R}{\sqrt{NT}} \|e\| \|X_k\| \|P_{\lambda^0}\, e\, P_{f^0}\| \|\lambda^0\, (\lambda^{0\prime}\lambda^0)^{-1}\, (f^{0\prime} f^0)^{-1}\, f^{0\prime}\| \\
&\qquad = o_p(1)\ .
\end{aligned}
$$

\# Now we want to prove part (g) and (h) of the present lemma. For part (g) we have

$$
\begin{aligned}
&\frac{1}{\sqrt{NT}} \mathrm{Tr}\left\{[ee' - \mathbb{E}_{\mathcal{C}}(ee')]\ M_{\lambda^0}\, X_k\, f^0\, (f^{0\prime} f^0)^{-1}\, (\lambda^{0\prime}\lambda^0)^{-1}\, \lambda^{0\prime}\right\} \\
&= \frac{1}{\sqrt{NT}} \mathrm{Tr}\left\{[ee' - \mathbb{E}_{\mathcal{C}}(ee')]\ M_{\lambda^0}\, \overline{X}_k\, f^0\, (f^{0\prime} f^0)^{-1}\, (\lambda^{0\prime}\lambda^0)^{-1}\, \lambda^{0\prime}\right\} \\
&\quad + \frac{1}{\sqrt{NT}} \mathrm{Tr}\left\{[ee' - \mathbb{E}_{\mathcal{C}}(ee')]\ M_{\lambda^0}\, \widetilde{X}_k P_{f^0}\, f^0\, (f^{0\prime} f^0)^{-1}\, (\lambda^{0\prime}\lambda^0)^{-1}\, \lambda^{0\prime}\right\} \\
&= \frac{1}{\sqrt{NT}} \mathrm{Tr}\left\{[ee' - \mathbb{E}_{\mathcal{C}}(ee')]\ M_{\lambda^0}\, \overline{X}_k\, f^0\, (f^{0\prime} f^0)^{-1}\, (\lambda^{0\prime}\lambda^0)^{-1}\, \lambda^{0\prime}\right\} \\
&\quad + \frac{1}{\sqrt{NT}}\, \|ee' - \mathbb{E}_{\mathcal{C}}(ee')\| \left\|\widetilde{X}_k P_{f^0}\right\| \left\|f^0\, (f^{0\prime} f^0)^{-1}\, (\lambda^{0\prime}\lambda^0)^{-1}\, \lambda^{0\prime}\right\| \\
&= \frac{1}{\sqrt{NT}} \mathrm{Tr}\left\{[ee' - \mathbb{E}_{\mathcal{C}}(ee')]\ M_{\lambda^0}\, \overline{X}_k\, f^0\, (f^{0\prime} f^0)^{-1}\, (\lambda^{0\prime}\lambda^0)^{-1}\, \lambda^{0\prime}\right\} + o_p(1)
\end{aligned}
$$

Thus, what is left to prove is that $\frac{1}{\sqrt{NT}} \mathrm{Tr}\left\{[ee' - \mathbb{E}_{\mathcal{C}}(ee')]\ M_{\lambda^0}\, \overline{X}_k\, f^0\, (f^{0\prime} f^0)^{-1}\, (\lambda^{0\prime}\lambda^0)^{-1}\, \lambda^{0\prime}\right\} = o_p(1)$. For this we define

$$
B_k = M_{\lambda^0}\, \overline{X}_k\, f^0\, (f^{0\prime} f^0)^{-1}\, (\lambda^{0\prime}\lambda^0)^{-1}\, \lambda^{0\prime}\ .
$$

Using part (i) and (ii) of Lemma S.4.1 we find

$$
\begin{aligned}
\|B_k\|_F &\leq R^{1/2} \|B_k\| \\
&\leq R^{1/2} \|\overline{X}_k\| \left\|f^0\, (f^{0\prime} f^0)^{-1}\, (\lambda^{0\prime}\lambda^0)^{-1}\, \lambda^{0\prime}\right\| \\
&\leq R^{1/2} \|\overline{X}_k\|_F \left\|f^0\, (f^{0\prime} f^0)^{-1}\, (\lambda^{0\prime}\lambda^0)^{-1}\, \lambda^{0\prime}\right\|\ .
\end{aligned}
$$

and therefore

$$
\begin{aligned}
\mathbb{E}_{\mathcal{C}}\left(\|B_k\|_F^2\right) &\leq R \left\|f^0\, (f^{0\prime} f^0)^{-1}\, (\lambda^{0\prime}\lambda^0)^{-1}\, \lambda^{0\prime}\right\|^2 \mathbb{E}_{\mathcal{C}}\left(\|\overline{X}_k\|_F^2\right) \\
&= \mathcal{O}(1)\ ,
\end{aligned}
$$

where we used $\mathbb{E}_{\mathcal{C}}\left(\|\overline{X}_k\|_F^2\right) = \mathcal{O}(NT)$, which is true since we assumed uniformly bounded moments of $\overline{X}_{k,it}$. Applying Lemma S.8.2 we therefore find

$$\mathbb{E}_{\mathcal{C}}\left(\frac{1}{\sqrt{NT}}\mathrm{Tr}\left\{\left[ee' - \mathbb{E}_{\mathcal{C}}\left(ee'\right)\right] B_k\right\}\right)^2 \le c_0\,\frac{T}{NT}\,\mathbb{E}_{\mathcal{C}}\left(\|B_k\|_F^2\right) = o(1)\ ,$$

and thus

$$\frac{1}{\sqrt{NT}}\mathrm{Tr}\left\{\left[ee' - \mathbb{E}_{\mathcal{C}}\left(ee'\right)\right] B_k\right\} = o_p(1)\ ,$$

which is what we wanted to show. The proof for part (h) is analogous.

Part (i): Conditional on $\mathcal{C}$ the expression $e_{it}^2 \mathfrak{X}_{it}\,\mathfrak{X}_{it}' - \mathbb{E}_{\mathcal{C}}\left(e_{it}^2\,\mathfrak{X}_{it}\,\mathfrak{X}_{it}'\right)$ is mean zero, and it is also uncorrelated across i. This together with the bounded moments that we assume implies that

$$\mathrm{Var}_{\mathcal{C}}\left\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[e_{it}^2\,\mathfrak{X}_{it}\,\mathfrak{X}_{it}' - \mathbb{E}_{\mathcal{C}}\left(e_{it}^2\,\mathfrak{X}_{it}\,\mathfrak{X}_{it}'\right)\right]\right\} = \mathcal{O}_p(1/N) = o_p(1),$$

which shows the required result.

Part (j): Define the $K \times K$ matrix $A = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}^2\left(\mathfrak{X}_{it} + \mathcal{X}_{it}\right)\left(\mathfrak{X}_{it} - \mathcal{X}_{it}\right)'$. Then we have

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}^2\left(\mathfrak{X}_{it}\,\mathfrak{X}_{it}' - \mathcal{X}_{it}\,\mathcal{X}_{it}'\right) = \frac{1}{2}\left(A + A'\right).$$

Let B_k be the $N \times T$ matrix with elements $B_{k,it} = e_{it}^2\left(\mathfrak{X}_{k,it} + \mathcal{X}_{k,it}\right)$. We have $\|B_k\| \le \|B_k\|_F = \mathcal{O}_p(\sqrt{NT})$, because the moments of $B_{k,it}$ are uniformly bounded. The components of A can be written as $A_{lk} = \frac{1}{NT}\mathrm{Tr}[B_l(\mathfrak{X}_k - \mathcal{X}_k)']$. We therefore have

$$|A_{lk}| \le \frac{1}{NT}\mathrm{rank}(\mathfrak{X}_k - \mathcal{X}_k)\|B_l\|\,\|\mathfrak{X}_k - \mathcal{X}_k\|\,.$$

We have $\mathfrak{X}_k - \mathcal{X}_k = \widetilde{X}_k\,P_{f^0} + P_{\lambda^0}\,\widetilde{X}_k\,M_{f^0}$. Therefore $\mathrm{rank}(\mathfrak{X}_k - \mathcal{X}_k) \le 2R$ and

$$\begin{aligned}|A_{lk}| &\le \frac{2R}{NT}\|B_l\|\left(\left\|\widetilde{X}_k\,P_{f^0}\right| + \left\|P_{\lambda^0}\,\widetilde{X}_k\,M_{f^0}\right\|\right)\\ &\le \frac{2R}{NT}\|B_l\|\left(\left\|\widetilde{X}_k\,P_{f^0}\right| + \left\|P_{\lambda^0}\,\widetilde{X}_k\right\|\right) = \frac{2R}{NT}\mathcal{O}_p(\sqrt{NT})o_p(\sqrt{NT}) = o_p(1).\end{aligned}$$

where we used Lemma S.8.1. This shows the desired result. ∎

Proof of Lemma B.2. Let c be a K-vector such that $\|c\| = 1$. The required result follows by the Cramer-Wold device, if we show that

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}\mathfrak{X}_{it}'c\ \Rightarrow\ \mathcal{N}\left(0, c'\Omega c\right)\ .$$

For this, define $\xi_{it} = e_{it} \mathfrak{X}'_{it} c$. Furthermore define $\xi_m = \xi_{M,m} = \xi_{NT,it}$, with $M = NT$ and $m = T(i-1) + t \in \{1, \ldots, M\}$. We then have the following:

(i) Under Assumption 5(i), (ii), (iii) the sequence $\{\xi_m, \, m = 1, \ldots, M\}$ is a martingale difference sequence under the filtration $\mathcal{F}_m = \mathcal{C} \vee \sigma(\{\xi_n : n < m\})$.

(ii) $\mathbb{E}(\xi_{it}^4)$ is uniformly bounded, since by Assumption 5(vi) $\mathbb{E}_{\mathcal{C}} e_{it}^8$ and $\mathbb{E}_{\mathcal{C}} \left(\|X_{it}\|^{8+\epsilon} \right)$ are uniformly bounded by a non-random constant (applying Cauchy-Schwarz and the law of iterated expectations).

(iii) $\frac{1}{M} \sum_{m=1}^M \xi_m^2 = c' \Omega c + o_p(1)$.
This is true, because firstly under our assumptions we have $\mathbb{E}_{\mathcal{C}} \left\{ \left[\frac{1}{M} \sum_{m=1}^M \left(\xi_m^2 - \mathbb{E}_{\mathcal{C}}(\xi_m^2) \right) \right]^2 \right\} = \mathbb{E}_{\mathcal{C}} \left\{ \frac{1}{M^2} \sum_{m=1}^M \left(\xi_m^2 - \mathbb{E}_{\mathcal{C}}(\xi_m^2) \right)^2 \right\} = \mathcal{O}_P(1/M) = o_P(1)$, implying that we have $\frac{1}{M} \sum_{m=1}^M \xi_m^2 = \frac{1}{M} \sum_{m=1}^M \mathbb{E}_{\mathcal{C}}(\xi_m^2) + o_p(1)$. We furthermore have $\frac{1}{M} \sum_{m=1}^M \mathbb{E}_{\mathcal{C}}(\xi_m^2) = \mathrm{Var}_{\mathcal{C}}(M^{-1/2} \sum_{m=1}^M \xi_m)$, and using the result in equation (14) of the main text we find $\mathrm{Var}_{\mathcal{C}}(M^{-1/2} \sum_{m=1}^M \xi_m) = \mathrm{Var}_{\mathcal{C}}((NT)^{-1/2} \sum_{i=1}^N \sum_{t=1}^T \xi_{it}) = c' \Omega c + o_p(1)$.

These three properties of $\{\xi_m, \, m = 1, \ldots, M\}$ allow us to apply Corollary 5.26 in White (2001), which is based on Theorem 2.3 in Mcleish (1974), to obtain that $\frac{1}{\sqrt{M}} \sum_{m=1}^M \xi_m \to_d \mathcal{N}(0, c' \Omega c)$. This concludes the proof, because $\frac{1}{\sqrt{M}} \sum_{m=1}^M \xi_m = \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T e_{it} \mathfrak{X}'_{it} c$. ∎

S.9 Expansions of Projectors and Residuals

The incidental parameter estimators $\widehat{f}$ and $\widehat{\lambda}$ as well as the residuals $\widehat{e}$ enter into the asymptotic bias and variance estimators for the LS estimator $\widehat{\beta}$. To describe the properties of $\widehat{f}$, $\widehat{\lambda}$ and $\widehat{e}$, it is convenient to have asymptotic expansions of the projectors $M_{\widehat{\lambda}}(\beta)$ and $M_{\widehat{f}}(\beta)$ that correspond to the minimizing parameters $\widehat{\lambda}(\beta)$ and $\widehat{f}(\beta)$ in equation (4). Note that the minimizing $\widehat{\lambda}(\beta)$ and $\widehat{f}(\beta)$ can be defined for all values of β, not only for the optimal value $\beta = \widehat{\beta}$. The corresponding residuals are $\widehat{e}(\beta) = Y \, - \, \beta \cdot X \, - \, \widehat{\lambda}(\beta) \, \widehat{f}'(\beta)$.

Theorem S.9.1. *Under Assumptions 1, 3, and 4(i) we have the following expansions*

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

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

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

where the spectral norms of the remainders satisfy for any series $\eta_{NT} \to 0$

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

$$\sup_{\{\beta: \left\|\beta-\beta^0\right\| \le \eta_{NT}\}} \frac{\left\| M^{(\mathrm{rem})}_{\widehat{f}}(\beta) \right\|}{\|\beta - \beta^0\|^2 + (NT)^{-1/2}\, \|e\|\, \|\beta - \beta^0\| \ + (NT)^{-3/2}\, \|e\|^3} = \mathcal{O}_p\left(1\right) \ ,$$

$$\sup_{\{\beta: \left\|\beta-\beta^0\right\| \le \eta_{NT}\}} \frac{\left\| \widehat{e}^{(\mathrm{rem})}(\beta) \right\|}{(NT)^{1/2}\|\beta - \beta^0\|^2 + \|e\|\, \|\beta - \beta^0\| + (NT)^{-1}\|e\|^3} = \mathcal{O}_p\left(1\right) \ ,$$

and we have $\mathrm{rank}(\widehat{e}^{(\mathrm{rem})}(\beta)) \le 7R$*, and the expansion coefficients are given by*

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

analogously

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

and finally

$$
\begin{aligned}
\widehat{e}^{(1)}_k &= M_{\lambda^0}\, X_k\, M_{f^0} \;, \\
\widehat{e}^{(1)}_e &= -M_{\lambda^0}\, e\, M_{f^0}\, e'\, \lambda^0\, (\lambda^{0\prime}\lambda^0)^{-1}\, (f^{0\prime} f^0)^{-1}\, f^{0\prime} \\
&\quad - \lambda^0\, (\lambda^{0\prime}\lambda^0)^{-1}\, (f^{0\prime} f^0)^{-1}\, f^{0\prime}\, e'\, M_{\lambda^0}\, e\, M_{f^0} \\
&\quad - M_{\lambda^0}\, e\, f^0\, (f^{0\prime} f^0)^{-1}\, (\lambda^{0\prime}\lambda^0)^{-1}\, \lambda^{0\prime}\, e\, M_{f^0} \;.
\end{aligned}
$$

Proof. The general expansion of $M_{\widehat{\lambda}}(\beta)$ is given Moon and Weidner (2013), and in the theorem we just make this expansion explicit up to a particular order. The result for $M_{\widehat{f}}(\beta)$ is just obtained by symmetry ($N \leftrightarrow T$, $\lambda \leftrightarrow f$, $e \leftrightarrow e'$, $X_k \leftrightarrow X_k'$). For the residuals $\widehat{e}$ we have

$$
\widehat{e} = M_{\widehat{\lambda}} \left(Y - \sum_{k=1} \widehat{\beta}_k\, X_k \right) = M_{\widehat{\lambda}} \left[e - \left(\widehat{\beta} - \beta^0 \right) \cdot X + \lambda^0 f^{0\prime} \right] \;,
$$

and plugging in the expansion of $M_{\widehat{\lambda}}$ gives the expansion of $\widehat{e}$. We have $\widehat{e}(\beta) = A_0 + \lambda^0 f^{0\prime} - \widehat{\lambda}(\beta)\widehat{f}'(\beta)$, where $A_0 = e - \sum_k(\beta_k - \beta_k^0)X_k$. Therefore $\widehat{e}^{(\text{rem})}(\beta) = A_1 + A_2 + A_3$ with $A_1 = A_0 - M_{\lambda^0}\, A_0\, M_{f^0}$, $A_2 = \lambda^0 f^{0\prime} - \widehat{\lambda}(\beta)\widehat{f}'(\beta)$, and $A_3 = -\widehat{e}^{(1)}_e$. We find $\text{rank}(A_1) \leq 2R$, $\text{rank}(A_2) \leq 2R$, $\text{rank}(A_3) \leq 3R$, and thus $\text{rank}(\widehat{e}^{(\text{rem})}(\beta)) \leq 7R$, as stated in the theorem. ∎

Having expansions for $M_{\widehat{\lambda}}(\beta)$ and $M_{\widehat{f}}(\beta)$ we also have expansions for $P_{\widehat{\lambda}}(\beta) = \mathbb{I}_N - M_{\widehat{\lambda}}(\beta)$ and $P_{\widehat{f}}(\beta) = \mathbb{I}_T - M_{\widehat{f}}(\beta)$. The reason why we give expansions of the projectors and not expansions of $\widehat{\lambda}(\beta)$ and $\widehat{f}(\beta)$ directly is that for the latter we would need to specify a normalization, while the projectors are independent of any normalization choice. An expansion for $\widehat{\lambda}(\beta)$ can for example be defined by $\widehat{\lambda}(\beta) = P_{\widehat{\lambda}}(\beta)\lambda^0$, in which case the normalization of $\widehat{\lambda}(\beta)$ is implicitly defined by the normalization of λ^0.

S.10 Consistency Proof for Bias and Variance Estimators (Proof of Theorem 4.4)

It is convenient to introduce some alternative notation for the Definition 1 in Section 4.3 of the main text.

Definition *Let $\Gamma : \mathbb{R} \to \mathbb{R}$ be the truncation kernel defined by $\Gamma(x) = 1$ for $|x| \leq 1$, and $\Gamma(x) = 0$ otherwise. Let M be a bandwidth parameter that depends on N and T. For an $N \times N$ matrix A with elements A_{ij} and a $T \times T$ matrix B with elements B_{ts} we define*

(i) the diagonal truncations $A^{\text{truncD}} = \text{diag}[(A_{ii})_{i=1,\ldots,N}]$ and $B^{\text{truncD}} = \text{diag}[(B_{tt})_{t=1,\ldots,T}]$.

(ii) the right-sided Kernel truncation of B, which is a $T \times T$ matrix B^{truncR} with elements $B^{\text{truncR}}_{ts} = \Gamma\left(\frac{s-t}{M}\right) B_{ts}$ for $t < s$, and $B^{\text{truncR}}_{ts} = 0$ otherwise.

Here, we suppress the dependence of B^{truncR} on the bandwidth parameter M. Using this notation we can represent the estimators for the bias in Definition 1 as follows:

$$
\begin{aligned}
\widehat{B}_{1,k} &= \frac{1}{N} \text{Tr}\left[P_{\widehat{f}} \, (\widehat{e}' X_k)^{\text{truncR}} \right] , \\
\widehat{B}_{2,k} &= \frac{1}{T} \text{Tr}\left[(\widehat{e}\,\widehat{e}')^{\text{truncD}} \; M_{\widehat{\lambda}} \, X_k \, \widehat{f} \, (\widehat{f}'\widehat{f})^{-1} \, (\widehat{\lambda}'\widehat{\lambda})^{-1} \, \widehat{\lambda}' \right] , \\
\widehat{B}_{3,k} &= \frac{1}{N} \text{Tr}\left[(\widehat{e}'\,\widehat{e})^{\text{truncD}} \; M_{\widehat{f}} \, X_k' \, \widehat{\lambda} \, (\widehat{\lambda}'\widehat{\lambda})^{-1} \, (\widehat{f}'\widehat{f})^{-1} \, \widehat{f}' \right] .
\end{aligned}
$$

Before proving Theorem 4.4 we establish some preliminary results.

Corollary S.10.1. *Under the Assumptions of Theorem 4.3 we have $\sqrt{NT}\left(\widehat{\beta} - \beta^0\right) = \mathcal{O}_p(1)$.*

This corollary directly follows from Theorem 4.3.

Corollary S.10.2. *Under the Assumptions of Theorem 4.4 we have*

$$
\begin{aligned}
\left\| P_{\widehat{\lambda}} - P_{\lambda^0} \right\| &= \left\| M_{\widehat{\lambda}} - M_{\lambda^0} \right\| = \mathcal{O}_p(N^{-1/2}) , \\
\left\| P_{\widehat{f}} - P_{f^0} \right\| &= \left\| M_{\widehat{f}} - M_{f^0} \right\| = \mathcal{O}_p(T^{-1/2}) .
\end{aligned}
$$

Proof. Using $\|e\| = \mathcal{O}_p(N^{1/2})$ and $\|X_k\| = \mathcal{O}_p(N)$ we find that the expansion terms in Theorem S.9.1 satisfy

$$
\left\| M^{(1)}_{\widehat{\lambda},e} \right\| = \mathcal{O}_p(N^{-1/2}) , \qquad \left\| M^{(2)}_{\widehat{\lambda},e} \right\| = \mathcal{O}_p(N^{-1}) , \qquad \left\| M^{(1)}_{\widehat{\lambda},k} \right\| = \mathcal{O}_p(1) .
$$

Together with corollary S.10.1 the result for $\left\| M_{\widehat{\lambda}} - M_{\lambda^0} \right\|$ immediately follows. In addition we have $P_{\widehat{\lambda}} - P_{\lambda^0} = -M_{\widehat{\lambda}} + M_{\lambda^0}$. The proof for $M_{\widehat{f}}$ and $P_{\widehat{f}}$ is analogous. ∎

Lemma S.10.3. *Under the Assumptions of Theorem 4.4 we have*

$$A_1 \equiv \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} e_{it}^2 \left(\mathcal{X}_{it} \mathcal{X}_{it}' - \widehat{\mathcal{X}}_{it} \widehat{\mathcal{X}}_{it}' \right) = o_p(1) \ ,$$

$$A_2 \equiv \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(e_{it}^2 - \widehat{e}_{it}^2 \right) \widehat{\mathcal{X}}_{it} \widehat{\mathcal{X}}_{it}' = o_p(1) \ .$$

Lemma S.10.4. *Let $\widehat{f}$ and f^0 be normalized as $\widehat{f}'\widehat{f}/T = \mathbb{I}_R$ and $f^{0\prime} f^0/T = \mathbb{I}_R$. Then, under the assumptions of Theorem 4.4, there exists an $R \times R$ matrices $H = H_{NT}$ such that*[3]

$$\left\| \widehat{f} - f^0 \, H \right\| = O_p\left(1\right) \ , \qquad \left\| \widehat{\lambda} - \lambda^0 \, \left(H'\right)^{-1} \right\| = O_p\left(1\right) \ .$$

Furthermore

$$\left\| \widehat{\lambda} \, (\widehat{\lambda}' \widehat{\lambda})^{-1} \, (\widehat{f}' \widehat{f})^{-1} \, \widehat{f}' - \lambda^0 \, (\lambda^{0\prime} \lambda^0)^{-1} \, (f^{0\prime} f^0)^{-1} \, f^{0\prime} \right\| = O_p\left(N^{-3/2}\right) \ .$$

Lemma S.10.5. *Under the Assumptions of Theorem 4.4 we have*

$$\text{(i)} \qquad N^{-1} \left\| \mathbb{E}_{\mathcal{C}}(e' X_k) - (\widehat{e}' \, X_k)^{\text{truncR}} \right\| = o_p(1) \ ,$$

$$\text{(ii)} \qquad N^{-1} \left\| \mathbb{E}_{\mathcal{C}}(e' e) - (\widehat{e}' \, \widehat{e})^{\text{truncD}} \right\| = o_p(1) \ ,$$

$$\text{(iii)} \qquad T^{-1} \left\| \mathbb{E}_{\mathcal{C}}(e e') - (\widehat{e} \, \widehat{e}')^{\text{truncD}} \right\| = o_p(1) \ .$$

Lemma S.10.6. *Under the Assumptions of Theorem 4.4 we have*

$$\text{(i)} \qquad N^{-1} \left\| (\widehat{e}' \, X_k)^{\text{truncR}} \right\| = \mathcal{O}_p(M T^{1/8}) \ ,$$

$$\text{(ii)} \qquad N^{-1} \left\| (\widehat{e}' \, \widehat{e})^{\text{truncD}} \right\| = \mathcal{O}_p(1) \ ,$$

$$\text{(iii)} \qquad T^{-1} \left\| (\widehat{e} \, \widehat{e}')^{\text{truncD}} \right\| = \mathcal{O}_p(1) \ .$$

The proof of the above lemmas is given in the supplementary material. Using these lemmas we can now prove Theorem 4.4.

Proof of Theorem 4.4, Part I: show $\widehat{W} = W + o_p(1)$.

[3] We consider a limit $N, T \to \infty$ and for different N, T different H-matrices can be chosen, but we write H instead of H_{NT} to keep notation simple.

Using $|\mathrm{Tr}\,(C)| \leq \|C\|\,\mathrm{rank}\,(C)$ and corollary S.10.2 we find

$$
\begin{aligned}
&\left|\widehat{W}_{k_1k_2} - W_{NT,k_1k_2}\right| \\
&\quad = \left|(NT)^{-1}\mathrm{Tr}\left[\left(M_{\widehat{\lambda}} - M_{\lambda^0}\right)\, X_{k_1}\, M_{\widehat{f}}\, X'_{k_2}\right] + (NT)^{-1}\mathrm{Tr}\left[M_{\lambda^0}\, X_{k_1}\, \left(M_{\widehat{f}} - M_{f^0}\right)\, X'_{k_2}\right]\right. \\
&\quad \leq \frac{2R}{NT}\left\|M_{\widehat{\lambda}} - M_{\lambda^0}\right\| \|X_{k_1}\| \|X_{k_2}\| \frac{2R}{NT}\left\|M_{\widehat{f}} - M_{f^0}\right\| \|X_{k_1}\| \|X_{k_2}\| \\
&\quad = \frac{2R}{NT}\mathcal{O}_p(N^{-1})\mathcal{O}_p(NT) + \frac{2R}{NT}\mathcal{O}_p(T^{-1})\mathcal{O}_p(NT) \\
&\quad = o_p(1)\ .
\end{aligned}
$$

Thus we have $\widehat{W} = W_{NT} + o_p(1) = W + o_p(1)$. ∎

Proof of Theorem 4.4, Part II: show $\widehat{\Omega} = \Omega + o_p(1)$.
Let $\Omega_{NT} \equiv \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}^2\, \mathcal{X}_{it}\, \mathcal{X}'_{it}$. We have $\Omega = \Omega_{NT} + o_P(1) = \widehat{\Omega} + A_1 + A_2 + o_p(1) = \widehat{\Omega} + o_P(1)$, where A_1 and A_2 are defined in Lemma S.10.3, and the lemma states that A_1 and A_2 are $o_p(1)$. ∎

Proof of Theorem 4.4, Part III: show $\widehat{B}_1 = B_1 + o_p(1)$.
Let $B_{1,k,NT} = N^{-1}\,\mathrm{Tr}\,[P_{f^0}\,\mathbb{E}_{\mathcal{C}}\,(e'\,X_k)]$. According to Assumption 6 we have $B_{1,k} = B_{1,k,NT}+o_p(1)$. What is left to show is that $B_{1,k,NT} = \widehat{B}_{1,k} + o_p(1)$. Using $|\mathrm{Tr}\,(C)| \leq \|C\|\,\mathrm{rank}\,(C)$ we find

$$
\begin{aligned}
\left|B_{1,k,NT} - \widehat{B}_1\right| &= \left|\mathbb{E}_{\mathcal{C}}\left[\frac{1}{N}\mathrm{Tr}(P_{f^0}\, e'\ X_k)\right] - \frac{1}{N}\mathrm{Tr}\left[P_{\widehat{f}}\,(\widehat{e}' X_k)^{\mathrm{truncR}}\right]\right| \\
&\leq \left|\frac{1}{N}\mathrm{Tr}\left[\left(P_{f^0} - P_{\widehat{f}}\right)(\widehat{e}' X_k)^{\mathrm{truncR}}\right]\right| \\
&\quad + \left|\frac{1}{N}\mathrm{Tr}\left\{P_{f^0}\left[\mathbb{E}_{\mathcal{C}}\,(e'\ X_k) - (\widehat{e}' X_k)^{\mathrm{truncR}}\right]\right\}\right| \\
&\leq \frac{2R}{N}\left\|P_{f^0} - P_{\widehat{f}}\right\| \left\|(\widehat{e}' X_k)^{\mathrm{truncR}}\right\| \\
&\quad + \frac{R}{N}\|P_{f^0}\| \left\|\mathbb{E}_{\mathcal{C}}\,(e'\ X_k) - (\widehat{e}' X_k)^{\mathrm{truncR}}\right\|\ .
\end{aligned}
$$

We have $\|P_{f^0}\| = 1$. We now apply Lemmas S.10.5, S.10.2 and S.10.6 to find

$$
\left|B_{1,k,NT} - \widehat{B}_1\right| = N^{-1}\left(\mathcal{O}_p(N^{-1/2})\mathcal{O}_p(MNT^{1/8}) + o_p(N)\right) = o_p(1)\ .
$$

This is what we wanted to show. ∎

Proof of Theorem 4.4, final part: show $\widehat{B}_2 = B_2 + o_p(1)$ and $\widehat{B}_3 = B_3 + o_p(1)$.
Define

$$
B_{2,k,NT} = \frac{1}{T}\mathrm{Tr}\left[\mathbb{E}_{\mathcal{C}}\,(ee')\ M_{\lambda^0}\, X_k\, f^0\,(f^{0\prime}f^0)^{-1}\,(\lambda^{0\prime}\lambda^0)^{-1}\,\lambda^{0\prime}\right]\ .
$$

According to Assumption 6 we have $B_{2,k} = B_{2,k,NT} + o_p(1)$. What is left to show is that $B_{2,k,NT} = \widehat{B}_{2,k} + o_p(1)$. We have

$$\begin{aligned}
B_{2,k} - \widehat{B}_{2,k} =& \frac{1}{T} \mathrm{Tr}\left[\mathbb{E}_{\mathcal{C}}\left(ee'\right) M_{\lambda^0} X_k f^0 \left(f^{0\prime} f^0\right)^{-1} \left(\lambda^{0\prime}\lambda^0\right)^{-1} \lambda^{0\prime}\right] \\
& - \frac{1}{T} \mathrm{Tr}\left[\left(\widehat{e}\,\widehat{e}'\right)^{\mathrm{truncD}} M_{\widehat{\lambda}} X_k \widehat{f} \left(\widehat{f}'\widehat{f}\right)^{-1} \left(\widehat{\lambda}'\widehat{\lambda}\right)^{-1} \widehat{\lambda}'\right] \\
=& \frac{1}{T} \mathrm{Tr}\left[\left(\widehat{e}\,\widehat{e}'\right)^{\mathrm{truncD}} M_{\widehat{\lambda}} X_k \left(f^0 \left(f^{0\prime} f^0\right)^{-1} \left(\lambda^{0\prime}\lambda^0\right)^{-1} \lambda^{0\prime} - \widehat{f} \left(\widehat{f}'\widehat{f}\right)^{-1} \left(\widehat{\lambda}'\widehat{\lambda}\right)^{-1} \widehat{\lambda}'\right)\right] \\
& + \frac{1}{T} \mathrm{Tr}\left[\left(\widehat{e}\,\widehat{e}'\right)^{\mathrm{truncD}} \left(M_{\lambda^0} - M_{\widehat{\lambda}}\right) X_k f^0 \left(f^{0\prime} f^0\right)^{-1} \left(\lambda^{0\prime}\lambda^0\right)^{-1} \lambda^{0\prime}\right] \\
& + \frac{1}{T} \mathrm{Tr}\left\{\left[\mathbb{E}_{\mathcal{C}}\left(ee'\right) - \left(\widehat{e}\,\widehat{e}'\right)^{\mathrm{truncD}}\right] M_{\lambda^0} X_k f^0 \left(f^{0\prime} f^0\right)^{-1} \left(\lambda^{0\prime}\lambda^0\right)^{-1} \lambda^{0\prime}\right\} .
\end{aligned}$$

Using $|\mathrm{Tr}\,(C)| \leq \|C\|\,\mathrm{rank}\,(C)$ (which is true for every square matrix C, see the supplementary material) we find

$$\begin{aligned}
\left|B_{2,k} - \widehat{B}_{2,k}\right| \leq& \frac{R}{T} \left\|\left(\widehat{e}\,\widehat{e}'\right)^{\mathrm{truncD}}\right\| \|X_k\| \left\|f^0 \left(f^{0\prime} f^0\right)^{-1} \left(\lambda^{0\prime}\lambda^0\right)^{-1} \lambda^{0\prime} - \widehat{f} \left(\widehat{f}'\widehat{f}\right)^{-1} \left(\widehat{\lambda}'\widehat{\lambda}\right)^{-1} \widehat{\lambda}'\right\| \\
& + \frac{R}{T} \left\|\left(\widehat{e}\,\widehat{e}'\right)^{\mathrm{truncD}}\right\| \left\|M_{\lambda^0} - M_{\widehat{\lambda}}\right\| \|X_k\| \left\|f^0 \left(f^{0\prime} f^0\right)^{-1} \left(\lambda^{0\prime}\lambda^0\right)^{-1} \lambda^{0\prime}\right\| \\
& + \frac{R}{T} \left\|\mathbb{E}_{\mathcal{C}}\left(ee'\right) - \left(\widehat{e}\,\widehat{e}'\right)^{\mathrm{truncD}}\right\| \|X_k\| \left\|f^0 \left(f^{0\prime} f^0\right)^{-1} \left(\lambda^{0\prime}\lambda^0\right)^{-1} \lambda^{0\prime}\right\| .
\end{aligned}$$

Here we used $\|M_{f^0}\| = \left\|M_{\widehat{f}}\right\| = 1$. Using $\|X_k\| = \mathcal{O}_p(\sqrt{NT})$, and applying Lemmas S.10.2, S.10.4, S.10.5 and S.10.6, we now find

$$\begin{aligned}
\left|B_{2,k} - \widehat{B}_{2,k}\right| = T^{-1} \Big[& \mathcal{O}_p(T)\, \mathcal{O}_p((NT)^{1/2})\, \mathcal{O}_p(N^{-3/2}) \\
& + \mathcal{O}_p(T)\, \mathcal{O}_p(N^{-1/2})\, \mathcal{O}_p((NT)^{1/2})\, \mathcal{O}_p((NT)^{-1/2}) \\
& + o_p(T)\, \mathcal{O}_p((NT)^{1/2})\, \mathcal{O}_p((NT)^{-1/2}) \Big] = o_p(1) .
\end{aligned}$$

This is what we wanted to show. The proof of $\widehat{B}_3 = B_3 + o_p(1)$ is analogous. ∎

S.11 Proof of Intermediate Lemma

Here we provide the proof of some intermediate lemmas that were stated and used in Section S.10, but not proved yet, in order to keep that section more readable. The following lemma gives a useful bound on the maximum of (correlated) random variables

Lemma S.11.1. *Let Z_i, $i = 1, 2, \ldots, n$, be n real valued random variables, and let $\gamma \geq 1$ and $B > 0$ be finite constants (independent of n). Assume* $\max_i \mathbb{E}_{\mathcal{C}}|Z_i|^\gamma \leq B$, *i.e. the γ'th moment* *of the Z_i are finite and uniformly bounded. For $n \to \infty$ we then have*

$$\max_i |Z_i| = \mathcal{O}_p\left(n^{1/\gamma}\right) \ . \tag{S.11.1}$$

Proof. Using Jensen's inequality one obtains $\mathbb{E}_{\mathcal{C}} \max_i |Z_i| \leq (\mathbb{E}_{\mathcal{C}} \max_i |Z_i|^{\gamma})^{1/\gamma} \leq (\mathbb{E}_{\mathcal{C}} \sum_{i=1}^n |Z_i|^{\gamma})^{1/\gamma} \leq (n \, \max_i \mathbb{E}_{\mathcal{C}} |Z_i|^{\gamma})^{1/\gamma} \leq n^{1/\gamma} B^{1/\gamma}$. Markov's inequality then gives equation (S.11.1). ∎

Lemma S.11.2. *Let*

$$\begin{aligned}
\bar{Z}^{(1)}_{k,t\tau} &= N^{-1/2} \sum_{i=1}^{N} \left[e_{it} X_{k,i\tau} - \mathbb{E}_{\mathcal{C}}\left(e_{it} X_{k,i\tau}\right)\right] \ , \\
\bar{Z}^{(2)}_{t} &= N^{-1/2} \sum_{i=1}^{N} \left[e_{it}^2 - \mathbb{E}_{\mathcal{C}}\left(e_{it}^2\right)\right] \ , \\
\bar{Z}^{(3)}_{i} &= T^{-1/2} \sum_{t=1}^{T} \left[e_{it}^2 - \mathbb{E}_{\mathcal{C}}\left(e_{it}^2\right)\right] \ .
\end{aligned}$$

Under assumption 5 we have

$$\begin{aligned}
\mathbb{E}_{\mathcal{C}} \left| \bar{Z}^{(1)}_{k,t\tau} \right|^4 &\leq B \ , \\
\mathbb{E}_{\mathcal{C}} \left| \bar{Z}^{(2)}_{t\tau} \right|^4 &\leq B \ , \\
\mathbb{E}_{\mathcal{C}} \left| \bar{Z}^{(3)}_{i} \right|^4 &\leq B \ ,
\end{aligned}$$

for some $B > 0$, i.e. the conditional expectations $\bar{Z}^{(1)}_{k,t\tau}$, $\bar{Z}^{(2)}_{t\tau}$, and $\bar{Z}^{(3)}_{i}$ are uniformly bounded over t, τ, or i, respectively.

Proof. # We start with the proof for $\bar{Z}^{(1)}_{k,t\tau}$. Define $Z^{(1)}_{k,t\tau,i} = e_{it} X_{k,i\tau} - \mathbb{E}_{\mathcal{C}}\left(e_{it} X_{k,i\tau}\right)$. By assumption we have finite 8'th moments for e_{it} and $X_{k,i\tau}$ uniformly across k, i, t, τ, and thus (using Cauchy Schwarz inequality) we have finite 4th moment of $Z^{(1)}_{k,t\tau,i}$ uniformly across k, i, t, τ. For ease of notation we now fix k, t, τ and write $Z_i = Z^{(1)}_{k,t\tau,i}$. We have $\mathbb{E}_{\mathcal{C}}(Z_i) = 0$ and $\mathbb{E}_{\mathcal{C}}(Z_i Z_j Z_k Z_l) = 0$ if $i \notin \{j, k, l\}$ (and the same holds for permutations of i, j, k, l). Using this we compute

$$\begin{aligned}
\mathbb{E}_{\mathcal{C}} \left(\sum_{i=1}^{N} Z_i \right)^4 &= \sum_{i,j,k,l=1}^{N} \mathbb{E}_{\mathcal{C}}\left(Z_i Z_j Z_k Z_l\right) \\
&= 3 \sum_{i \neq j} \mathbb{E}_{\mathcal{C}}\left(Z_i^2 \, Z_j^2\right) + \sum_{i} \mathbb{E}_{\mathcal{C}}\left(Z_i^4\right) \\
&= 3 \sum_{i,j=1}^{N} \mathbb{E}_{\mathcal{C}}\left(Z_i^2\right) \, \mathbb{E}_{\mathcal{C}}\left(Z_j^2\right) + \sum_{i=1}^{N} \left\{ \mathbb{E}_{\mathcal{C}}\left(Z_i^4\right) - 3 \left[\mathbb{E}_{\mathcal{C}}\left(Z_i^2\right)\right]^2 \right\} \ ,
\end{aligned}$$

Since we argued that $\mathbb{E}_{\mathcal{C}}\left(Z_i^4\right)$ is bounded uniformly, the last equation shows that $\bar{Z}^{(1)}_{k,t\tau} = N^{-1/2}\sum_{i=1}^N Z^{(1)}_{k,t\tau,i}$ is bounded uniformly across k, t, τ. This is what we wanted to show.

The proofs for $\bar{Z}^{(2)}_t$ and $\bar{Z}^{(3)}_i$ are analogous. ∎

Lemma S.11.3. *For a $T \times T$ matrix A we have*[4]

$$\left\|A^{\rm truncR}\right\| \;\leq\; M \left\|A^{\rm truncR}\right\|_{\max} \;\equiv\; M \max_t \max_{t<\tau\leq t+M} |A_{t\tau}| \,,$$

Proof. For the 1-norm of $A^{\rm truncR}$ we find

$$\begin{aligned}\left\|A^{\rm truncR}\right\|_1 &= \max_{t=1\ldots T} \sum_{\tau=t+1}^{t+M} |A_{t\tau}| \\ &\leq M \max_{t<\tau\leq t+M} |A_{t\tau}| = M \left\|A^{\rm truncR}\right\|_{\max} \,,\end{aligned}$$

and analogously we find the same bound for the ∞-norm $\left\|A^{\rm truncR}\right\|_\infty$. Applying part (vii) of Lemma S.4.1 we therefore also get this bound for the operator norm $\|A^{\rm truncR}\|$. ∎

Proof of Lemma S.10.3. # We first show $A_1 \equiv (NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T e_{it}^2\left(\mathcal{X}_{it}\mathcal{X}'_{it} - \widehat{\mathcal{X}}_{it}\widehat{\mathcal{X}}'_{it}\right) = o_p(1)$. Let $B_{1,it} = \mathcal{X}_{it} - \widehat{\mathcal{X}}_{it}$, $B_{2,it} = e_{it}^2\mathcal{X}_{it}$, and $B_{3,it} = e_{it}^2\widehat{\mathcal{X}}_{it}$. Note that B_1, B_2, and B_3 can either be viewed as K-vectors for each it, or equivalently as $N\times T$ matrices $B_{1,k}$, $B_{2,k}$, and $B_{3,k}$ for each $k = 1,\ldots,K$. We have $A_1 = (NT)^{-1}\sum_i\sum_t\left(B_{1,it}B'_{2,it} + B_{3,it}B'_{1,it}\right)$, or equivalently

$$A_{1,k_1k_2} = \frac{1}{NT}{\rm Tr}\left(B_{1,k_1}B'_{3,k_2} + B_{2,k_1}B'_{1,k_2}\right) \,.$$

Using $\|M_{\widehat{\lambda}} - M_{\lambda^0}\| = \mathcal{O}_p(N^{-1/2})$, $\|M_{\widehat{f}} - M_{f^0}\| = \mathcal{O}_p(N^{-1/2})$, $\|X_k\| = \mathcal{O}_p(\sqrt{NT}) = \mathcal{O}_p(N)$, we find for $B_{1,k} = (M_{\lambda^0} - M_{\widehat{\lambda}})X_kM_{f^0} + M_{\widehat{\lambda}}X_k(M_{f^0} - M_{\widehat{f}})$ that $\|B_{1,k}\| = \mathcal{O}_p(N^{1/2})$. In addition we have ${\rm rank}(B_{1,k}) \leq 4R$. We also have

$$\begin{aligned}\|B_{2,k}\|^4 &\leq \|B_{2,k}\|_F^4 \\ &= \left(\sum_{i=1}^N\sum_{t=1}^T e_{it}^4\mathcal{X}_{k,it}^2\right)^2 \\ &\leq \left(\sum_{i=1}^N\sum_{t=1}^T e_{it}^8\right)\left(\sum_{i=1}^N\sum_{t=1}^T \mathcal{X}_{k,it}^4\right) = \mathcal{O}_p(NT)\,\mathcal{O}_p(NT)\,,\end{aligned}$$

[4]For the boundaries of τ we could write $\max(1, t-M)$ instead of $t-M$, and $\min(T, t+M)$ instead of $t+M$, to guarantee $1 \leq \tau \leq T$. Since this would complicate notation, we prefer the convention that $A_{t\tau} = 0$ for $t < 1$ or $\tau < 1$ of $t > T$ or $\tau > T$.

which implies $\|B_{2,k}\| = \mathcal{O}_p(\sqrt{NT})$, and analogously we find $\|B_{3,k}\| = \mathcal{O}_p(\sqrt{NT})$. Therefore

$$
\begin{aligned}
|A_{1,k_1k_2}| &\leq \frac{4R}{NT} \left(\|B_{1,k_1}\| \|B_{3,k_2}\| + \|B_{2,k_1}\| \|B_{1,k_2}\| \right) \\
&= \frac{4R}{NT} \left(\mathcal{O}_p(N^{1/2}) \mathcal{O}_p(\sqrt{NT}) + \mathcal{O}_p(\sqrt{NT}) \mathcal{O}_p(N^{1/2}) \right) = o_p(1) \ .
\end{aligned}
$$

This is what we wanted to show.

Finally, we want to show $A_2 \equiv (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T \left(e_{it}^2 - \widehat{e}_{it}^2 \right) \widehat{\mathcal{X}}_{it} \widehat{\mathcal{X}}_{it}' = o_p(1)$. According to theorem S.9.1 we have $e - \widehat{e} = C_1 + C_2$, where we defined $C_1 = - \sum_{k=1}^K \left(\widehat{\beta}_k - \beta_k^0 \right) X_k$, and $C_2 = \sum_{k=1}^K \left(\widehat{\beta}_k - \beta_k^0 \right) \left(P_{\lambda^0} \, X_k \, M_{f^0} + X_k \, P_{f^0} \right) + P_{\lambda^0} \, e \, M_{f^0} + e \, P_{f^0} - \widehat{e}_e^{(1)} - \widehat{e}^{(\mathrm{rem})}$, which satisfies $\|C_2\| = \mathcal{O}_p(N^{1/2})$, and $\mathrm{rank}(C_2) \leq 11R$ (actually, one can easily prove $\leq 5R$, but this does not follow from theorem S.9.1). Using this notation we have

$$
A_2 = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (e_{it} + \widehat{e}_{it})(C_{1,it} + C_{2,it}) \widehat{\mathcal{X}}_{it} \widehat{\mathcal{X}}_{it}' \ ,
$$

which can also be written as

$$
A_{2,k_1k_2} = - \sum_{k_3=1}^K \left(\widehat{\beta}_{k_3} - \beta_{k_3}^0 \right) \left(C_{5,k_1k_2k_3} + C_{6,k_1k_2k_3} \right) + \frac{1}{NT} \mathrm{Tr} \left(C_2 \, C_{3,k_1k_2} \right) + \frac{1}{NT} \mathrm{Tr} \left(C_2 \, C_{4,k_1k_2} \right) \ ,
$$

where we defined

$$
\begin{aligned}
C_{3,k_1k_2,it} &= e_{it} \widehat{\mathcal{X}}_{k_1,it} \widehat{\mathcal{X}}_{k_2,it} \ , \\
C_{4,k_1k_2,it} &= \widehat{e}_{it} \widehat{\mathcal{X}}_{k_1,it} \widehat{\mathcal{X}}_{k_2,it} \ , \\
C_{5,k_1k_2k_3} &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \ e_{it} \widehat{\mathcal{X}}_{k_1,it} \widehat{\mathcal{X}}_{k_2,it} X_{k_3,it} \ , \\
C_{6,k_1k_2k_3} &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \ \widehat{e}_{it} \widehat{\mathcal{X}}_{k_1,it} \widehat{\mathcal{X}}_{k_2,it} X_{k_3,it} \ .
\end{aligned}
$$

Again, since we have uniformly bounded 8'th moments for e_{it} and $X_{k,it}$, we find

$$
\begin{aligned}
\|C_{3,k_1k_2}\|^4 &\leq \|C_{3,k_1k_2}\|_F^4 \\
&= \left(\sum_{i=1}^N \sum_{t=1}^T e_{it}^2 \widehat{\mathcal{X}}_{k_1,it}^2 \widehat{\mathcal{X}}_{k_2,it}^2 \right)^2 \\
&\leq \left(\sum_{i=1}^N \sum_{t=1}^T e_{it}^4 \right) \left(\sum_{i=1}^N \sum_{t=1}^T \widehat{\mathcal{X}}_{k_1,it}^4 \widehat{\mathcal{X}}_{k_2,it}^4 \right) \\
&= \mathcal{O}_p(N^2 T^2) \ ,
\end{aligned}
$$

i.e. $\|C_{3,k_1k_2}\| = \mathcal{O}_p(\sqrt{NT})$. Furthermore

$$\begin{aligned}
\|C_{4,k_1k_2}\|^2 &\leq \|C_{3,k_1k_2}\|_F^2 \\
&= \sum_{i=1}^{N}\sum_{t=1}^{T} \widehat{e}_{it}^2 \widehat{\mathcal{X}}_{k_1,it}^2 \widehat{\mathcal{X}}_{k_2,it}^2 \\
&\leq \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \widehat{e}_{it}^2\right) \max_{i=1\ldots N} \max_{t=1\ldots T} \left(\widehat{\mathcal{X}}_{k_1,it}^2 \widehat{\mathcal{X}}_{k_2,it}^2\right) \\
&\leq \left(\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}^2\right) \max_{i=1\ldots N} \max_{t=1\ldots T} \left(\widehat{\mathcal{X}}_{k_1,it}^2 \widehat{\mathcal{X}}_{k_2,it}^2\right) \\
&= \mathcal{O}_p(NT)\mathcal{O}_p((NT)^{(4/(8+\epsilon))}) = o_p((NT)^{(3/4)}) \ .
\end{aligned}$$

Here we used the assumption that X_k has uniformly bounded moments of order $8+\epsilon$ for some $\epsilon > 0$. We also used $\sum_{i=1}^N \sum_{t=1}^T \widehat{e}_{it}^2 \leq \sum_{i=1}^N \sum_{t=1}^T e_{it}^2$.

For C_5 we find

$$\begin{aligned}
C_{5,k_1k_2k_3}^2 &\leq \left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}^2\right)\left(\frac{1}{NT}\widehat{\mathcal{X}}_{k_1,it}^2 \widehat{\mathcal{X}}_{k_2,it}^2 X_{k_3,it}^2\right) \\
&= \mathcal{O}_p(1) \ ,
\end{aligned}$$

i.e. $C_{5,k_1k_2k_3} = \mathcal{O}_p(1)$, and analogously $C_{6,k_1k_2k_3} = \mathcal{O}_p(1)$, since $\sum_{i=1}^N \sum_{t=1}^T \widehat{e}_{it}^2 \leq \sum_{i=1}^N \sum_{t=1}^T e_{it}^2$.

Using these results we obtain

$$\begin{aligned}
|A_{2,k_1k_2}| &\leq -\sum_{k_3=1}^{K} \left\|\widehat{\beta}_{k_3} - \beta_{k_3}^0\right\| |C_{5,k_1k_2k_3} + C_{6,k_1k_2k_3}| + \frac{11R}{NT}\|C_2\|\|C_{3,k_1k_2}\| + \frac{11R}{NT}\|C_2\|\|C_{4,k_1k_2}\| \\
&= \mathcal{O}_p((NT)^{-1/2})\mathcal{O}_p(1) + \frac{11R}{NT}\mathcal{O}_p(N^{1/2})\mathcal{O}_p(\sqrt{NT}) + \frac{11R}{NT}\mathcal{O}_p(N^{1/2})o_p((NT)^{3/4}) = o_p(1) \ .
\end{aligned}$$

This is what we wanted to show. ∎

Remember that the truncation Kernel $\Gamma(.)$ is defined by $\Gamma(x) = 1$ for $|x| \leq 1$ and $\Gamma(x) = 0$ otherwise. Without loss of generality we assume in the following that the bandwidth parameter M is a positive integer (without this assumption, one needs to replace M everywhere below by the largest integer contained in M, but nothing else changes).

Proof of Lemma S.10.4. By Lemma S.10.2 we know that asymptotically $P_{\widehat{f}}$ is close to P_{f^0} and therefore $\text{rank}(P_{\widehat{f}} P_{f^0}) = \text{rank}(P_{f^0} P_{f^0}) = R$, *i.e.* $\text{rank}(P_{\widehat{f}} f^0) = R$ asymptotically. We can therefore write $\widehat{f} = P_{\widehat{f}} f^0 H$, where $H = H_{NT}$ is a non-singular $R \times R$ matrix.

We now want to show $\|H\| = O_p(1)$ and $\|H^{-1}\| = O_p(1)$. Due to our normalization of $\widehat{f}$ and f^0 we have $H = (\widehat{f}' P_{\widehat{f}} f^0/T)^{-1} = (\widehat{f}' f^0/T)^{-1}$, and therefore $\|H^{-1}\| \leq \|\widehat{f}\| \|f^0\|/T = \mathcal{O}_p(1)$. We also have $\widehat{f} = f^0 H + (P_{\widehat{f}} - P_{f^0}) f^0 H$, and thus $H = f^{0\prime} \widehat{f}/T - f^{0\prime}(P_{\widehat{f}} - P_{f^0}) f^0 H/T$, *i.e.* $\|H\| \leq \mathcal{O}_p(1) + \|H\| \mathcal{O}_p\left(T^{-1/2}\right)$ which shows $\|H\| = \mathcal{O}_p(1)$. Note that all the following results only require $\|H\| = \mathcal{O}_p(1)$ and $\|H^{-1}\| = \mathcal{O}_p(1)$, but apart from that are independent of the choice of normalization.

The advantage of expressing $\widehat{f}$ in terms of $P_{\widehat{f}}$ as above is that the result $\left\| P_{\widehat{f}} - P_{f^0} \right\| = \mathcal{O}_p\left(T^{-1/2}\right)$ of Lemma S.10.2 immediately implies

$$\left\| \widehat{f} - f^0\, H \right\| = \mathcal{O}_p\left(1\right) \ .$$

The FOC wrt λ in the minimization of the first line in equation (4) reads

$$\widehat{\lambda}\, \widehat{f}' \widehat{f} = \left(Y - \sum_{k=1}^{K} \widehat{\beta}_k X_k \right) \widehat{f} \ , \tag{S.11.2}$$

which yields

$$\begin{aligned}
\widehat{\lambda} &= \left[\lambda^0 f^{0\prime} - \sum_{k=1}^{K} \left(\widehat{\beta}_k - \beta_k^0 \right) X_k \right] \widehat{f} \left(\widehat{f}' \widehat{f} \right)^{-1} \\
&= \left[\lambda^0 f^{0\prime} + \sum_{k=1}^{K} \left(\beta_k^0 - \widehat{\beta}_k \right) X_k + e \right] P_{\widehat{f}} f^0 \left(f^{0\prime} P_{\widehat{f}} f^0 \right)^{-1} \left(H' \right)^{-1} \\
&= \lambda^0 \left(H' \right)^{-1} + \lambda^0 f^{0\prime} \left(P_{\widehat{f}} - P_{f^0} \right) f^0 \left(f^{0\prime} P_{\widehat{f}} f^0 \right)^{-1} \left(H' \right)^{-1} \\
&\qquad + \lambda^0 f^{0\prime} f^0 \left[\left(f^{0\prime} P_{\widehat{f}} f^0 \right)^{-1} - \left(f^{0\prime} f^0 \right)^{-1} \right] \left(H' \right)^{-1} \\
&\qquad + \left[\sum_{k=1}^{K} \left(\beta_k^0 - \widehat{\beta}_k \right) X_k + e \right] P_{\widehat{f}} f^0 \left(f^{0\prime} P_{\widehat{f}} f^0 \right)^{-1} \left(H' \right)^{-1} \ .
\end{aligned}$$

We have $\left(f^{0\prime} P_{\widehat{f}} f^0 \big/ T \right)^{-1} - \left(f^{0\prime} f^0/T \right)^{-1} = \mathcal{O}_p(T^{-1/2})$, because $\left\| P_{\widehat{f}} - P_{f^0} \right\| = \mathcal{O}_p\left(T^{-1/2}\right)$ and $f^{0\prime} f^0/T$ by assumption is converging to a positive definite matrix (or given our particular choice of normalization is just the identity matrix $\mathbb{I}_R$) In addition, we have $\|e\| = \mathcal{O}_p(\sqrt{T})$, $\|X_k\| = \mathcal{O}_p(\sqrt{NT})$ and by corollary S.10.1 also $\|\widehat{\beta} - \beta^0\| = \mathcal{O}_p(1/\sqrt{NT})$. Therefore

$$\left\| \widehat{\lambda} - \lambda^0 \left(H' \right)^{-1} \right\| = \mathcal{O}_p\left(1\right) \ , \tag{S.11.3}$$

which is what we wanted to prove.

Next, we want to show

$$\left\| \left(\frac{\widehat{\lambda}' \widehat{\lambda}}{N} \right)^{-1} - \left(\frac{(H)^{-1}\, \lambda^{0\prime} \lambda^0\, (H')^{-1}}{N} \right)^{-1} \right\| = \mathcal{O}_p\left(N^{-1/2}\right) \ ,$$

$$\left\| \left(\frac{\widehat{f}' \widehat{f}}{T} \right)^{-1} - \left(\frac{H' f^{0\prime} f^0 H}{T} \right)^{-1} \right\| = \mathcal{O}_p\left(T^{-1/2}\right) \ . \tag{S.11.4}$$

Let $A = N^{-1} \widehat{\lambda}' \widehat{\lambda}$ and $B = N^{-1}\, (H)^{-1}\, \lambda^{0\prime} \lambda^0\, (H')^{-1}$. Using (S.11.3) we find

$$\begin{aligned} \|A - B\| &= \frac{1}{2N} \left\| \left[\widehat{\lambda}' + (H)^{-1}\, \lambda^{0\prime} \right] \left[\widehat{\lambda} - \lambda^0\, (H')^{-1} \right] + \left[\widehat{\lambda}' - (H)^{-1}\, \lambda^{0\prime} \right] \left[\widehat{\lambda} + \lambda^0\, (H')^{-1} \right] \right\| \\ &= N^{-1}\, \mathcal{O}_p(N^{1/2})\, \mathcal{O}_p(1) = \mathcal{O}_p\left(N^{-1/2}\right) \ . \end{aligned}$$

By assumption 1 we know that

$$\left\| \left(\frac{\lambda^{0\prime} \lambda^0}{N} \right)^{-1} \right\| = \mathcal{O}_p(1) \ .$$

and thus also $\|B^{-1}\| = \mathcal{O}_p(1)$, and therefore $\|A^{-1}\| = \mathcal{O}_p(1)$ (using $\|A - B\| = o_p(1)$ and applying Weyl's inequality to the smallest eigenvalue of B). Since $A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}$ we find

$$\begin{aligned} \left\|A^{-1} - B^{-1}\right\| &\leq \left\|A^{-1}\right\| \, \left\|B^{-1}\right\| \, \|A - B\| \\ &= \mathcal{O}_p\left(N^{-1/2}\right) \ . \end{aligned}$$

Thus, we have shown the first statement of (S.11.4), and analogously one can show the second one. Combining (S.11.3), (S.11.2) and (S.11.4) we obtain

$$\begin{aligned} &\left\| \frac{\widehat{\lambda}}{\sqrt{N}} \left(\frac{\widehat{\lambda}' \widehat{\lambda}}{N} \right)^{-1} \left(\frac{\widehat{f}' \widehat{f}}{T} \right)^{-1} \frac{\widehat{f}'}{\sqrt{T}} - \frac{\lambda^0}{\sqrt{N}} \left(\frac{\lambda^{0\prime} \lambda^0}{N} \right)^{-1} \left(\frac{f^{0\prime} f^0}{T} \right)^{-1} \frac{f^{0\prime}}{\sqrt{T}} \right\| \\ &= \left\| \frac{\widehat{\lambda}}{\sqrt{N}} \left(\frac{\widehat{\lambda}' \widehat{\lambda}}{N} \right)^{-1} \left(\frac{\widehat{f}' \widehat{f}}{T} \right)^{-1} \frac{\widehat{f}'}{\sqrt{T}} - \frac{\lambda^0\, (H')^{-1}}{\sqrt{N}} \left(\frac{(H)^{-1}\, \lambda^{0\prime} \lambda^0\, (H')^{-1}}{N} \right)^{-1} \left(\frac{H' f^{0\prime} f^0 H}{T} \right)^{-1} \frac{H' f^{0\prime}}{\sqrt{T}} \right\| \\ &\qquad = \mathcal{O}_p\left(N^{-1/2}\right) \ , \end{aligned}$$

which is equivalent to the statement in lemma. Note also that $\widehat{\lambda}\, (\widehat{\lambda}' \widehat{\lambda})^{-1}\, (\widehat{f}' \widehat{f})^{-1}\, \widehat{f}'$ is independent of H, *i.e.* independent of the choice of normalization. ∎

Proof of Lemma S.10.5. # Part A of the proof: We start by showing that

$$N^{-1} \left\| \mathbb{E}_{\mathcal{C}} \left[e' X_k - (e' X_k)^{\text{truncR}} \right] \right\| = o_p(1) \; . \tag{S.11.5}$$

Let $A = e'X_k$ and $B = A - A^{\text{truncR}}$. By definition of the left-sided truncation (using the equal weight kernel $\Gamma(.)$) we have $B_{t\tau} = 0$ for $t < \tau \leq t+M$ and $B_{t\tau} = A_{t\tau}$ otherwise. By assumption 5 we have $\mathbb{E}_{\mathcal{C}}(A_{t\tau}) = 0$ for $t \geq \tau$. For $t < \tau$ we have $\mathbb{E}_{\mathcal{C}}(A_{t\tau}) = \sum_{i=1}^N \mathbb{E}_{\mathcal{C}}(e_{it} X_{k,i\tau})$. We thus have $\mathbb{E}_{\mathcal{C}}(B_{t\tau}) = 0$ for $\tau \leq t + M$, and $\mathbb{E}_{\mathcal{C}} B_{t\tau} = \sum_{i=1}^N \mathbb{E}_{\mathcal{C}}(e_{it} X_{k,i\tau})$ for $\tau > t + M$. Therefore

$$\begin{aligned}
\left\| \mathbb{E}_{\mathcal{C}}(B) \right\|_1 &= \max_{t=1\ldots T} \sum_{\tau=1}^{T} \left| \mathbb{E}_{\mathcal{C}}(B_{t\tau}) \right| \\
&\leq \max_{t=1\ldots T} \sum_{\tau=t+M+1}^{T} \left| \sum_{i=1}^{N} \mathbb{E}_{\mathcal{C}}(e_{it} X_{k,i\tau}) \right| \leq N \max_{t=1\ldots T} \sum_{\tau=t+M+1}^{T} c\,(\tau - t)^{-(1+\epsilon)} = o_p(N) \; ,
\end{aligned}$$

where we used $M \to \infty$. Analogously we can show $\left\| \mathbb{E}_{\mathcal{C}}(B) \right\|_\infty = o_p(N)$. Using part (vii) of Lemma S.4.1 we therefore also find $\left\| \mathbb{E}_{\mathcal{C}}(B) \right\| = o_p(N)$, which is equivalent to equation (S.11.5) that we wanted to show in this part of the proof. Analogously we can show that

$$\begin{aligned}
N^{-1} \left\| \mathbb{E}_{\mathcal{C}} \left[e'e - (e'e)^{\text{truncD}} \right] \right\| &= o_p(1) \; , \\
T^{-1} \left\| \mathbb{E}_{\mathcal{C}} \left[ee' - (ee')^{\text{truncD}} \right] \right\| &= o_p(1) \; .
\end{aligned}$$

Part B of the proof: Next, we want to show that

$$N^{-1} \left\| \left[e' X_k \, - \, \mathbb{E}_{\mathcal{C}} \left(e' X_k \right) \right]^{\text{truncR}} \right\| = o_p(1) \; . \tag{S.11.6}$$

Using Lemma S.11.3 we have

$$\begin{aligned}
N^{-1} \left\| \left[e' X_k \, - \, \mathbb{E}_{\mathcal{C}} \left(e' X_k \right) \right]^{\text{truncR}} \right\| &\leq M \max_{t} \max_{t < \tau \leq t+M} N^{-1} \left| e_t' X_{k,\tau} - \mathbb{E}_{\mathcal{C}} \left(e_t' X_{k,\tau} \right) \right| \\
&\leq M \max_{t} \max_{t < \tau \leq t+M} N^{-1} \left| \sum_{i=1}^{N} \left[e_{it} X_{k,i\tau} - \mathbb{E}_{\mathcal{C}} \left(e_{it} X_{k,i\tau} \right) \right] \right| \\
&\leq M \, N^{-1/2} \max_{t} \max_{t < \tau \leq t+M} \left| \bar{Z}^{(1)}_{k,t\tau} \right| .
\end{aligned}$$

According to Lemma S.11.2 we know that $\mathbb{E}_{\mathcal{C}} \left| \bar{Z}^{(1)}_{k,t\tau} \right|^4$ is bounded uniformly across t and τ. Applying Lemma S.11.1 we therefore find $\max_t \max_{t<\tau\leq t+M} \bar{Z}^{(1)}_{t\tau} = \mathcal{O}_p((MT)^{1/4})$. Thus we have

$$M \, N^{-1/2} \max_{t} \max_{t < \tau \leq t+M} \left| \bar{Z}^{(1)}_{t\tau} \right| = \mathcal{O}_p \left(M \, N^{-1/2} \, (MT)^{1/4} \right) \, = \, o_p(1) \; .$$

Here we used $M^5/T \to 0$. Analogously we can show that

$$N^{-1} \left\| \left[e'e \, - \, \mathbb{E}_{\mathcal{C}} \left(e'e \right) \right]^{\rm truncD} \right\| = o_p(1) \; ,$$
$$T^{-1} \left\| \left[ee' \, - \, \mathbb{E}_{\mathcal{C}} \left(ee' \right) \right]^{\rm truncD} \right\| = o_p(1) \; .$$

Part C of the proof: Finally, we want to show that

$$N^{-1} \left\| \left[e' X_k \, - \, \widehat{e}' \, X_k \right]^{\rm truncR} \right\| = o_p(1) \; . \tag{S.11.7}$$

According to theorem S.9.1 we have $\widehat{e} = M_{\lambda^0} e M_{f^0} + e_{\rm rem}$, where $e_{\rm rem} \equiv \widehat{e}^{(1)}_e - \sum_{k=1}^K \left(\widehat{\beta}_k - \beta_k^0 \right) \widehat{e}^{(1)}_k + \widehat{e}^{(\rm rem)}$. We then have

$$\begin{aligned} & N^{-1} \left\| \left[e' X_k \, - \, \widehat{e}' \, X_k \right]^{\rm truncR} \right\| \\ & \leq N^{-1} \left\| \left[e'_{\rm rem} X_k \right]^{\rm truncR} \right\| + N^{-1} \left\| \left[P_{f^0} e' M_{\lambda^0} X_k \right]^{\rm truncR} \right\| + N^{-1} \left\| \left[e' P_{\lambda^0} X_k \right]^{\rm truncR} \right\| . \end{aligned}$$

Using corollary S.10.1 we find that the remainder term satisfies $\|e_{\rm rem}\| = \mathcal{O}_p(1)$. Using Lemma S.11.3 we find

$$\begin{aligned} N^{-1} \left\| \left[e'_{\rm rem} \, X_k \right]^{\rm truncR} \right\| &= \frac{M}{N} \, \max_{t,\tau} \, \widehat{e}'_{{\rm rem},t} \, X_{k,\tau} \\ &\leq \frac{M}{N} \, \max_{t,\tau} \, \|e_{{\rm rem},t}\| \, \|X_{k,\tau}\| \\ &\leq \frac{M}{N} \, \|e_{\rm rem}\| \, \max_{\tau} \|X_{k,\tau}\| \\ &\leq \frac{M}{N} \mathcal{O}_p(1) \mathcal{O}_p(N^{1/2} T^{1/8}) = o_p(1) \; , \end{aligned}$$

where we used the fact that the norm of each column $e_{{\rm rem},t}$ is smaller than the operator norm of the whole matrix $e_{\rm rem}$. In addition we used Lemma S.11.1 and the fact that $N^{-1/2} \left\| X_{k,\tau} \right\| = \sqrt{N^{-1} \sum_{i=1}^N X_{k,i\tau}^2}$ has finite 8'th moment in order to show $\max_\tau \left\| X_{k,\tau} \right\| = \mathcal{O}_p(N^{1/2} T^{1/8})$. Using again Lemma S.11.3 we find

$$\begin{aligned} N^{-1} \left\| \left[P_{f^0} e' M_{\lambda^0} X_k \right]^{\rm truncR} \right\| &\leq N^{-1} \, M \, \max_{t,\tau=1\ldots T} \left| f_t^0 \, (f^{0\prime} \, f^0)^{-1} \, f^{0\prime} \, e' M_{\lambda^0} X_{k,\tau} \right| \\ &\leq N^{-1} \, M \, \|e\| \, \|f^0\| \, \left\| (f^{0\prime} \, f^0)^{-1} \right\| \, \max_t \, \|f_t^0\| \, \max_\tau \|X_{k,\tau}\| \\ &= N^{-1} \, M \, \mathcal{O}_p(N^{1/2}) \, \mathcal{O}_p(T^{1/2}) \, \mathcal{O}_p(T^{-1}) \, \mathcal{O}_p(N^{1/2} T^{1/8}) = o_p(1) \; , \end{aligned}$$

and

$$\begin{aligned} \left\| \left[e' P_{\lambda^0} X_k \right]^{\rm truncR} \right\| &\leq N^{-1/2} M \max_{t=1\ldots T} \left(N^{-1/2} \sum_i e_{it} \lambda_i^0 \right) (N^{-1} \lambda^{0\prime} \, \lambda^0)^{-1} \max_{\tau=1\ldots T} \left(N^{-1} \sum_j \lambda_j^{0\prime} X_{jt} \right) \\ &= N^{-1/2} M \mathcal{O}_p(T^{1/8}) \mathcal{O}_p(1) \mathcal{O}_p(T^{1/8}) = o_p(1). \end{aligned}$$

Thus, we proved equation (S.11.7). Analogously we obtain

$$N^{-1}\left\|\left[e'e - \widehat{e}'\,\widehat{e}\right]^{\text{truncD}}\right\| = o_p(1)\;,$$
$$T^{-1}\left\|\left[ee' - \widehat{e}\,\widehat{e}'\right]^{\text{truncD}}\right\| = o_p(1)\;.$$

Combining (S.11.5), (S.11.6) and (S.11.7) shows that $N^{-1}\left\|\mathbb{E}_{\mathcal{C}}(e'X_k) - (\widehat{e}'\,X_k)^{\text{truncR}}\right\| = o_p(1)$. The proof of the other two statements of the lemma is analogous. ∎

Proof of Lemma S.10.6. Using theorem S.9.1 and S.10.1 we find $\|\widehat{e}\| = \mathcal{O}_p(N^{1/2})$. Applying Lemma S.11.3 we therefore find

$$\begin{aligned}
N^{-1}\left\|(\widehat{e}'\,X_k)^{\text{truncR}}\right\| &\le \frac{M}{N}\max_{t,\tau}|\widehat{e}_t'\,X_{k,\tau}| \\
&\le \frac{M}{N}\max_{t,\tau}\|\widehat{e}_t\|\,\|X_{k,\tau}\| \\
&\le \frac{M}{N}\|\widehat{e}\|\,\max_{\tau}\|X_{k,\tau}\| \\
&\le \frac{M}{N}\mathcal{O}_p(N^{1/2})\mathcal{O}_p(N^{1/2}T^{1/8}) = \mathcal{O}_p(MT^{1/8})\;,
\end{aligned}$$

where we used the result $\max_\tau\|X_{k,\tau}\| = \mathcal{O}_p(N^{1/2}T^{1/8})$ that was already obtained in the proof of the last theorem.

The proof for the statement (ii) and (iii) is analogous. ∎

S.12 Proofs for Section 5 (Testing)

Proof of Theorem 5.1. Using the expansion for $L_{NT}(\beta)$ in Lemma S.1 in the supplementary material of Moon and Weidner (2013) we find for the derivative (the sign convention $\epsilon_k = \beta_k^0 - \beta_k$ results in the minus sign below)

$$\begin{aligned}
\frac{\partial L_{NT}}{\partial \beta_k} &= -\frac{1}{NT}\sum_{g=2}^{\infty} g\sum_{\kappa_1=0}^{K}\sum_{\kappa_2=0}^{K}\ldots\sum_{\kappa_{g-1}=0}^{K}\epsilon_{\kappa_1}\,\epsilon_{\kappa_2}\,\ldots\,\epsilon_{\kappa_{g-1}}\,L^{(g)}\left(\lambda^0,\,f^0,\,X_k,\,X_{\kappa_1},\ldots,X_{\kappa_{g-1}}\right) \\
&= \left[2W_{NT}(\beta-\beta^0)\right]_k - \frac{2}{\sqrt{NT}}C_{NT,k} + \frac{1}{NT}\nabla R_{1,NT,k} + \frac{1}{NT}\nabla R_{2,NT,k}\;,
\end{aligned}$$

where

$$
\begin{aligned}
W_{NT,k_1k_2} &= \frac{1}{NT}\, L^{(2)}\left(\lambda^0,\, f^0,\, X_{k_1},\, X_{k_2}\right)\ , \\
C_{NT,k} &= \frac{1}{2\sqrt{NT}} \sum_{g=2}^{G_e} g\,(\epsilon_0)^{g-1}\, L^{(g)}\left(\lambda^0,\, f^0,\, X_k,\, X_0, \ldots, X_0\right) \\
&= \sum_{g=2}^{G_e} \frac{g}{2\sqrt{NT}}\ L^{(g)}\left(\lambda^0,\, f^0,\, X_k,\, e, \ldots, e\right)\ ,
\end{aligned}
$$

and

$$
\begin{aligned}
\nabla R_{1,NT,k} &= -\sum_{g=G_e+1}^{\infty} g\,(\epsilon_0)^{g-1}\, L^{(g)}\left(\lambda^0,\, f^0,\, X_k,\, X_0, \ldots, X_0\right)\ , \\
&= -\sum_{g=G_e+1}^{\infty} g\ L^{(g)}\left(\lambda^0,\, f^0,\, X_k,\, e, \ldots, e\right)\ , \\
\nabla R_{2,NT,k} &= -\sum_{g=3}^{\infty} g \sum_{r=1}^{g-1} \binom{g-1}{r} \sum_{k_1=1}^{K} \ldots \sum_{k_r=1}^{K} \epsilon_{k_1}\, \ldots\, \epsilon_{k_r}\, (\epsilon_0)^{g-r-1} \\
&\qquad\qquad L^{(g)}\left(\lambda^0,\, f^0,\, X_k,\, X_{k_1}, \ldots, X_{k_r}, X_0, \ldots, X_0\right)\ . \\
&= -\sum_{g=3}^{\infty} g \sum_{r=1}^{g-1} \binom{g-1}{r} \sum_{k_1=1}^{K} \ldots \sum_{k_r=1}^{K} (\beta^0_{k_1} - \beta_{k_1})\, \ldots\, (\beta^0_{k_r} - \beta_{k_r}) \\
&\qquad\qquad L^{(g)}\left(\lambda^0,\, f^0,\, X_k,\, X_{k_1}, \ldots, X_{k_r}, e, \ldots, e\right)\ .
\end{aligned}
$$

The above expressions for W_{NT} and C_{NT} are equivalent to their definitions given in theorem 4.1. Using the bound on $L^{(g)}$ we find[5]

$$
\begin{aligned}
|\nabla R_{1,NT,k}| &\leq c_0\, NT\, \frac{\|X_k\|}{\sqrt{NT}} \sum_{g=G_e+1}^{\infty} g^2 \left(\frac{c_1\|e\|}{\sqrt{NT}}\right)^{g-1} \\
&\leq 2\, c_0\, (1+G_e)^2\, NT\, \frac{\|X_k\|}{\sqrt{NT}} \left(\frac{c_1\|e\|}{\sqrt{NT}}\right)^{G_e} \left[1 - \left(\frac{c_1\|e\|}{\sqrt{NT}}\right)\right]^{-3} = o_p(\sqrt{NT}) \,, \\
|\nabla R_{2,NT,k}| &\leq c_0\, NT\, \frac{\|X_k\|}{\sqrt{NT}} \sum_{g=3}^{\infty} g^2 \sum_{r=1}^{g-1} \binom{g-1}{r} c_1^{g-1} \left(\sum_{\widetilde{k}=1}^{K} |\beta_{\widetilde{k}} - \beta^0_{\widetilde{k}}| \frac{\|X_{\widetilde{k}}\|}{\sqrt{NT}}\right) \\
&\qquad \times \left(\sum_{\widetilde{k}=1}^{K} |\beta_{\widetilde{k}} - \beta^0_{\widetilde{k}}| \frac{\|X_{\widetilde{k}}\|}{\sqrt{NT}} + \frac{\|e\|}{\sqrt{NT}}\right)^{g-2} \\
&\leq c_0\, NT\, \frac{\|X_k\|}{\sqrt{NT}} \sum_{g=3}^{\infty} g^3\, (4c_1)^{g-1} \left(\sum_{\widetilde{k}=1}^{K} |\beta_{\widetilde{k}} - \beta^0_{\widetilde{k}}| \frac{\|X_{\widetilde{k}}\|}{\sqrt{NT}}\right) \left(\sum_{\widetilde{k}=1}^{K} |\beta_{\widetilde{k}} - \beta^0_{\widetilde{k}}| \frac{\|X_{\widetilde{k}}\|}{\sqrt{NT}} + \frac{\|e\|}{\sqrt{NT}}\right)^{g-2} \\
&\leq c_2\, NT\, \frac{\|X_k\|}{\sqrt{NT}} \left(\sum_{\widetilde{k}=1}^{K} |\beta_{\widetilde{k}} - \beta^0_{\widetilde{k}}| \frac{\|X_{\widetilde{k}}\|}{\sqrt{NT}}\right) \left(\sum_{\widetilde{k}=1}^{K} |\beta_{\widetilde{k}} - \beta^0_{\widetilde{k}}| \frac{\|X_{\widetilde{k}}\|}{\sqrt{NT}} + \frac{\|e\|}{\sqrt{NT}}\right) ,
\end{aligned}
$$

where $c_0 = 8Rd_{\max}(\lambda^0, f^0)/2$ and $c_1 = 16 d_{\max}(\lambda^0, f^0)/d^2_{\min}(\lambda^0, f^0)$ both converge to a constants as $N, T \to \infty$, and the very last inequality is only true if $4c_1 \left(\sum_{\widetilde{k}=1}^{K} |\beta_{\widetilde{k}} - \beta^0_{\widetilde{k}}| \frac{\|X_{\widetilde{k}}\|}{\sqrt{NT}} + \frac{\|e\|}{\sqrt{NT}}\right) < 1$, and $c_2 > 0$ is an appropriate positive constant. To show $\nabla R_{1,NT,k} = o_p(NT)$ we used Assumption 3*. From the above inequalities we find for $\eta_{NT} \to \infty$

$$
\begin{aligned}
\sup_{\{\beta: \|\beta-\beta^0\| \leq \eta_{NT}\}} \frac{\|\nabla R_{1,NT}(\beta)\|}{\sqrt{NT}} &= o_p(1) \,, \\
\sup_{\{\beta: \|\beta-\beta^0\| \leq \eta_{NT}\}} \frac{\|\nabla R_{2,NT}(\beta)\|}{NT\, \|\beta-\beta^0\|} &= o_p(1) \,.
\end{aligned}
$$

Thus $R_{NT}(\beta) = R_{1,NT}(\beta) + R_{2,NT}(\beta)$ satisfies the bound in the theorem. ∎

Proof of Theorem 5.2. Using Theorem 4.3 it is straightforward to show that WD^*_{NT} has limiting distribution χ^2_r.

For the LR test we have to show that the estimator $\widehat{c} = (NT)^{-1}\mathrm{Tr}(\widehat{e}(\widehat{\beta})\, \widehat{e}'(\widehat{\beta}))$ is consistent for $c = \mathbb{E}_{\mathcal{C}} e^2_{it}$. As already noted in the main text we have $\widehat{c} = L_{NT}\left(\widehat{\beta}\right)$, and using our expansion and $\sqrt{NT}$-consistency of $\widehat{\beta}$ we immediately obtain

$$
\widehat{c} = \frac{1}{NT}\, \mathrm{Tr}(M_{\lambda^0} e M_{f^0} e') + o_p(1) \,.
$$

[5] Here we use $\binom{n}{k} \leq 4^n$.

Alternatively, one could use the expansion of $\widehat{e}$ in Theorem S.9.1 to show this. From the above result we find

$$\left|\widehat{c} - \frac{1}{NT}\mathrm{Tr}(ee')\right| = \frac{1}{NT}\left|\mathrm{Tr}(P_{\lambda^0} e M_{f^0} e') + \mathrm{Tr}(e P_{f^0} e')\right| + o_p(1)$$
$$\leq \frac{2R}{NT}\|e\|^2 + o_p(1) = o_p(1)\;.$$

By the weak law of large numbers we thus have

$$\widehat{c} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}^2 + o_p(1) = c + o_p(1)\;,$$

i.e. $\widehat{c}$ is indeed consistent for c. Having this one immediately obtains the result for the limiting distribution of LR^*_{NT}.

For the LM test we first want to show that equation (9) holds. Using the expansion of $\widehat{e}$ in Theorem S.9.1 one obtains

$$\begin{aligned}
\sqrt{NT}(\widetilde{\nabla}\mathcal{L}_{NT})_k &= -\frac{2}{\sqrt{NT}}\,\mathrm{Tr}\left(X_k'\widetilde{e}\right) \\
&= \left[2\sqrt{NT}\,W_{NT}\left(\widetilde{\beta}-\beta^0\right)\right]_k + \frac{2}{NT}C^{(1)}(\lambda^0, f^0, X_k, e) + \frac{2}{NT}C^{(2)}(\lambda^0, f^0, X_k, e) \\
&\qquad - \frac{2}{\sqrt{NT}}\,\mathrm{Tr}\left(X_k'\widetilde{e}^{(\mathrm{rem})}\right) \\
&= \left[2\sqrt{NT}\,W_{NT}\left(\widetilde{\beta}-\beta^0\right) + \frac{2}{NT}C_{NT}\right]_k + o_p(1) \\
&= \sqrt{NT}\left[\nabla L_{NT}(\widetilde{\beta})\right]_k + o_p(1)\,,
\end{aligned}$$

which is what we wanted to show. Here we used that $|\mathrm{Tr}\left(X_k'\widetilde{e}^{(\mathrm{rem})}\right)| \leq 7R\|X_k\|\|\widetilde{e}^{(\mathrm{rem})}\| = \mathcal{O}_p(N^{3/2})$. Note that $\|X_k\| = \mathcal{O}_p(N)$, and Theorem S.9.1 and $\sqrt{NT}$-consistency of $\widetilde{\beta}$ imply $\|\widetilde{e}^{(\mathrm{rem})}\| = \mathcal{O}_p(\sqrt{N})$. We also used the expression for $\nabla L_{NT}(\widetilde{\beta})$ given in Theorem 5.1, and the bound on $\nabla R_{NT}(\beta)$ given there.

We now use equation (10) and $\widetilde{W} = W + o_p(1)$, $\widetilde{\Omega} = \Omega + o_p(1)$, and $\widetilde{B} = B + o_p(1)$ to obtain

$$LM^*_{NT} \underset{d}{\longrightarrow} (C-B)'W^{-1}H'(HW^{-1}\Omega W^{-1}H')^{-1}HW^{-1}(C-B)\;.$$

Under H_0 we thus find $LM^*_{NT} \to_d \chi^2_r$. ∎

S.13 Additional Monte Carlo Results

We consider an AR(1) model with R factors

$$Y_{it} = \rho^0\,Y_{i,t-1} + \sum_{r=1}^{R}\lambda^0_{ir}\,f^0_{tr} + e_{it}\;.$$

We draw the e_{it} independently and identically distributed from a t-distribution with five degrees of freedom. The λ^0_{ir} are independently distributed as $\mathcal{N}(1,1)$, and we generate the factors from an AR(1) specification, namely $f^0_{tr} = \rho_f\, f^0_{t-1,r} + u_{tr}$, for each $r = 1, \ldots, R$, where $u_{tr} \sim$ iid$\mathcal{N}(0, (1-\rho_f^2)\sigma_f^2)$. For all simulations we generate 1000 initial time periods for f^0_t and Y_{it} that are not used for estimation. This guarantees that the simulated data used for estimation is distributed according to the stationary distribution of the model.

For $R = 1$ this is exactly the simulation design used in the main text Monte Carlo section, but DGP's with $R > 1$ were not considered in the main text. Table S.1 reports results, where $R = 1$ is used both in the DGP and for the LS estimation. Table S.2 reports results, where $R = 1$ is used in the DGP, but $R = 2$ is used for the LS estimation. Table S.3 reports results, where $R = 2$ is used both in the DGP and for the LS estimation. The results in Table S.1 and S.2 are identical to those reported in the main text Table 1 and 2, except that we also report results for the CCE estimator. The results in Table S.3 are not contained in the main text.

The CCE estimator is obtained by using $\widehat{f}^{\text{proxy}}_t = N^{-1} \sum_i (Y_{it},\, Y_{i,t-1})'$ as a proxy for the factors an then estimating the parameters ρ, λ_{i1}, λ_{i2}, $i = 1, \ldots, N$, via OLS in the linear regression model $Y_{it} = \rho Y_{i,t-1} + \lambda_{i1} \widehat{f}^{\text{proxy}}_{t1} + \lambda_{i2} \widehat{f}^{\text{proxy}}_{t2} + e_{it}$.

The performance of the CCE estimator in Table S.1 and S.2 are identical (up to random MC noise), because the number of factors need not be specified for the CCE estimator, and the DGP's in Table S.1 and S.2 are identical. These tables show that for $R = 1$ in the DGP the CCE estimator performs very well. From Chudik and Pesaran (2013) we expect the CCE estimator to have a bias of order $1/T$ in a dynamic model, which is confirmed in the simulations: the bias of the CCE estimator shrinks roughly in inverse proportion to T, as T becomes larger. The $1/T$ bias of the CCE estimator could be corrected for, and we would expect the bias corrected CCE estimator to perform similarly to the bias corrected LS estimator, although this is not included in the simulations.

However, if there are $R = 2$ factors in the true DGP, then it turns out that the proxies $\widehat{f}^{\text{proxy}}_t$ do not pick those up correctly. Table S.3 shows that for some parameter values and sample sizes (e.g. $\rho^0 = 0.3$ and $T = 10$, or $\rho^0 = 0.9$ and $T = 40$) the CCE estimator is almost unbiased, but for other values, including $T = 80$, the CCE estimator is heavily biased if $R = 2$. In particular, the bias of the CCE estimator does not seem to converge to zero as T becomes large in this case. In contrast, the correctly specified LS estimators (i.e. correctly using $R = 2$ factors in the estimation) performs very well according to Table S.3. However, an incorrectly specified LS estimator, which would underestimate the number of factors (e.g. using $R = 1$ factors in estimation instead of the correct number $R = 2$) would probably perform similarly to the CCE estimator, since not all factors would be corrected for. Overestimating the number of factors (i.e. using $R = 3$ factors in estimation instead of the correct number $R = 2$) should, however, not pose a problem for the LS estimator, according to Moon and Weidner (2013).

References

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

Bhatia, R. (1997). *Matrix Analysis*. Springer-Verlag, New York.

Chudik, A. and Pesaran, H. (2013). Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. *CESIFO WORKING PAPER NO. 4232*.

Galantai, A. and Hegedus, C. J. (2006). Jordan's principal angles in complex vector spaces. *Num. Lin. Alg. Appl.*, 13:589–598.

Golub, G. H. and Van Loan, C. F. (1996). *Matrix Computations (Johns Hopkins Studies in Mathematical Sciences), Third Edition*. The Johns Hopkins University Press.

Latala, R. (2005). Some estimates of norms of random matrices. *Proc. Amer. Math. Soc.*, 133:1273–1282.

McLeish, D. (1974). Dependent central limit theorems and invariance principles. *the Annals of Probability*, pages 620–628.

Moon, H. and Weidner, M. (2013). Linear Regression for Panel with Unknown Number of Factors as Interactive Fixed Effects. *CeMMAP working paper series*.

White, H. (2001). *Asymptotic theory for econometricians*. Academic press New York.

Tables with Simulation Results

		$\rho^0 = 0.3$				$\rho^0 = 0.9$			
		OLS	FLS	BC-FLS	CCE	OLS	FLS	BC-FLS	CCE
$T = 5,\ M = 2$	bias	0.1232	-0.1419	-0.0713	-0.1755	0.0200	-0.3686	-0.2330	-0.3298
	std	0.1444	0.1480	0.0982	0.1681	0.0723	0.1718	0.1301	0.2203
	rmse	0.1898	0.2050	0.1213	0.2430	0.0750	0.4067	0.2669	0.3966
$T = 10,\ M = 3$	bias	0.1339	-0.0542	-0.0201	-0.0819	0.0218	-0.1019	-0.0623	-0.1436
	std	0.1148	0.0596	0.0423	0.0593	0.0513	0.1094	0.0747	0.0972
	rmse	0.1764	0.0806	0.0469	0.1011	0.0557	0.1495	0.0973	0.1734
$T = 20,\ M = 4$	bias	0.1441	-0.0264	-0.0070	-0.0405	0.0254	-0.0173	-0.0085	-0.0617
	std	0.0879	0.0284	0.0240	0.0277	0.0353	0.0299	0.0219	0.0406
	rmse	0.1687	0.0388	0.0250	0.0491	0.0434	0.0345	0.0235	0.0739
$T = 40,\ M = 5$	bias	0.1517	-0.0130	-0.0021	-0.0200	0.0294	-0.0057	-0.0019	-0.0281
	std	0.0657	0.0170	0.0160	0.0166	0.0250	0.0105	0.0089	0.0162
	rmse	0.1654	0.0214	0.0161	0.0260	0.0386	0.0119	0.0091	0.0324
$T = 80,\ M = 6$	bias	0.1552	-0.0066	-0.0007	-0.0100	0.0326	-0.0026	-0.0006	-0.0136
	std	0.0487	0.0112	0.0109	0.0111	0.0179	0.0056	0.0053	0.0073
	rmse	0.1627	0.0130	0.0109	0.0149	0.0372	0.0062	0.0053	0.0154

Table S.1: Same as Table 1 in main paper, but also reporting pooled CCE estimator of Pesaran (2006).

		$\rho^0 = 0.3$				$\rho^0 = 0.9$			
		OLS	FLS	BC-FLS	CCE	OLS	FLS	BC-FLS	CCE
$T = 5,\ M = 2$	bias	0.1239	-0.5467	-0.3721	-0.1767	0.0218	-0.9716	-0.7490	-0.3289
	std	0.1454	0.1528	0.1299	0.1678	0.0731	0.1216	0.1341	0.2203
	rmse	0.1910	0.5676	0.3942	0.2437	0.0763	0.9792	0.7609	0.3958
$T = 10,\ M = 3$	bias	0.1343	-0.1874	-0.1001	-0.0816	0.0210	-0.4923	-0.3271	-0.1414
	std	0.1145	0.1159	0.0758	0.0592	0.0518	0.1159	0.0970	0.0971
	rmse	0.1765	0.2203	0.1256	0.1008	0.0559	0.5058	0.3412	0.1715
$T = 20,\ M = 4$	bias	0.1451	-0.0448	-0.0168	-0.0407	0.0255	-0.1822	-0.1085	-0.0618
	std	0.0879	0.0469	0.0320	0.0277	0.0354	0.0820	0.0528	0.0404
	rmse	0.1696	0.0648	0.0362	0.0492	0.0436	0.1999	0.1207	0.0739
$T = 40,\ M = 5$	bias	0.1511	-0.0161	-0.0038	-0.0199	0.0300	-0.0227	-0.0128	-0.0282
	std	0.0663	0.0209	0.0177	0.0167	0.0250	0.0342	0.0225	0.0164
	rmse	0.1650	0.0264	0.0181	0.0260	0.0390	0.0410	0.0258	0.0326
$T = 80,\ M = 6$	bias	0.1550	-0.0072	-0.0011	-0.0100	0.0325	-0.0030	-0.0010	-0.0136
	std	0.0488	0.0123	0.0115	0.0111	0.0182	0.0064	0.0057	0.0074
	rmse	0.1625	0.0143	0.0116	0.0149	0.0372	0.0071	0.0058	0.0155

Table S.2: Same as Table 2 in main paper, but also reporting pooled CCE estimator of Pesaran (2006).

		$\rho^0 = 0.3$				$\rho^0 = 0.9$			
		OLS	FLS	BC-FLS	CCE	OLS	FLS	BC-FLS	CCE
$T = 5,\ M = 2$	bias	0.1861	-0.4968	-0.3323	-0.1002	0.0309	-0.9305	-0.7057	-0.2750
	std	0.1562	0.1910	0.1580	0.2063	0.0801	0.1644	0.1754	0.2302
	rmse	0.2429	0.5322	0.3680	0.2294	0.0859	0.9449	0.7272	0.3586
$T = 10,\ M = 3$	bias	0.1989	-0.1569	-0.0758	0.0036	0.0326	-0.4209	-0.2732	-0.1040
	std	0.1185	0.1018	0.0700	0.1074	0.0543	0.1607	0.1235	0.1070
	rmse	0.2315	0.1870	0.1031	0.1074	0.0633	0.4505	0.2998	0.1492
$T = 20,\ M = 4$	bias	0.2096	-0.0592	-0.0185	0.0520	0.0366	-0.0741	-0.0406	-0.0310
	std	0.0884	0.0377	0.0287	0.0711	0.0356	0.0859	0.0552	0.0512
	rmse	0.2274	0.0702	0.0341	0.0881	0.0511	0.1134	0.0686	0.0599
$T = 40,\ M = 5$	bias	0.2174	-0.0275	-0.0054	0.0759	0.0404	-0.0134	-0.0047	-0.0012
	std	0.0649	0.0192	0.0170	0.0500	0.0239	0.0166	0.0122	0.0281
	rmse	0.2269	0.0335	0.0179	0.0908	0.0469	0.0214	0.0131	0.0281
$T = 80,\ M = 6$	bias	0.2232	-0.0134	-0.0016	0.0873	0.0433	-0.0052	-0.0012	0.0125
	std	0.0472	0.0118	0.0113	0.0364	0.0164	0.0066	0.0058	0.0176
	rmse	0.2281	0.0179	0.0114	0.0946	0.0463	0.0084	0.0059	0.0216

Table S.3: Analogous to Table 2 in main paper, but with $R = 2$ correctly specified, and also reporting pooled CCE estimator of Pesaran (2006).