Dynamic Linear Panel Regression Models with Interactive Fixed Effects

Supplementary Material (Not for Publication)

Hyungsik Roger Moon* Martin Weidner*

July 15, 2010

S.1 Proofs for Appendix A (Examples of Error Distributions)

Here we give the proof of the result $\|e\| = \mathcal{O}_p(\sqrt{\max(N,T)})$ for the different examples of error distributions presented in the main text.

Proof of example (i). Latala (2005) showed that for a $N \times T$ matrix e with independent entries we have

$$\mathbb{E}\|e\| \,\leq\, c \left(\max_i \sqrt{\sum_t \mathbb{E}e_{it}^2} + \max_j \sqrt{\sum_i \mathbb{E}e_{it}^2} + \sqrt[4]{\sum_{i,t} \mathbb{E}e_{it}^4} \right) \,, \tag{S.1.1}$$

where c is some universal constant (independent of N and T, and of the distribution of e). Since we assumed uniformly bounded 4'th moments for e_{it} we thus have $\mathbb{E}\|e\| = \mathcal{O}(\sqrt{T}) + \mathcal{O}(\sqrt{N}) + \mathcal{O}((TN)^{1/4})$, which implies $\mathbb{E}\|e\| = \mathcal{O}(\sqrt{\max(N,T)})$. Therefore $\|e\| = \mathcal{O}_p(\sqrt{\max(N,T)})$. ∎

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

$$\begin{aligned} e &= \sum_{j=0}^{\infty} \mathrm{diag}(\psi_j)\, U_{-j} \\ &= \sum_{j=0}^{T} \mathrm{diag}(\psi_j)\, U_{-j} + r_{NT}, \,, \end{aligned} \tag{S.1.2}$$

where we cut the sum at T, which results in the remainder $r_{NT} = \sum_{j=T+1}^{\infty} \mathrm{diag}(\psi_j)\, U_{-j}$. When

*Department of Economics, University of Southern California, 3620 South Vermont Avenue, Kaprielian Hall 300, Los Angeles, CA 90089-0253. Email: moonr@usc.edu, mweidner@usc.edu.

approximating an MA(∞) by a finite MA(T) process we have for the remainder

$$
\begin{aligned}
\mathbb{E}\left(\|r_{NT}\|_F\right)^2 &= \sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}\left(r_{NT}\right)_{ij}^2 \leq \sigma_u^2 \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=T+1}^{\infty}\psi_{ij}^2 \\
&\leq \sigma_u^2\, N\, T \sum_{j=T+1}^{\infty} \max_i \left(\psi_{ij}^2\right) \\
&\leq \sigma_u^2\, N \sum_{j=T+1}^{\infty} j \max_i \left(\psi_{ij}^2\right) \ , \qquad \text{(S.1.3)}
\end{aligned}
$$

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

$$
\mathbb{E}\left(\frac{\left(\|r_{NT}\|_F\right)^2}{N}\right) \longrightarrow 0 \ , \qquad \text{(S.1.4)}
$$

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 $\|\operatorname{diag}(\psi_j)\| \leq \max_i \left(|\psi_{ij}|\right)$.

Combining these results we find

$$
\begin{aligned}
\|e\| &\leq \sum_{j=0}^{T} \|\operatorname{diag}(\psi_j)\|\, \|U_{-j}\| + \|r_{NT}\| \\
&\leq \sum_{j=0}^{T} \max_i \left(|\psi_{ij}|\right) \|V\| + o_p(\sqrt{N}) \\
&\leq \left[\sum_{j=0}^{\infty} \max_i \left(|\psi_{ij}|\right)\right] \mathcal{O}_p(\sqrt{\max(N, 2T)}) + o_p(\sqrt{N}) \\
&\leq \mathcal{O}_p(\sqrt{\max(N, T)}) \ . \qquad \text{(S.1.5)}
\end{aligned}
$$

This is what we wanted to show. ∎

Proof of 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 a $N \times T$ 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

$$
\|\sigma\| \leq \|\sigma\|_1 \equiv \max_{j=1\ldots N} \sum_{i=1}^{N} |\sigma_{ij}| < L \ , \qquad \text{(S.1.6)}
$$

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.2 Comments on assumption 4 on the regressors

Consistency of the QMLE $\hat{\beta}$ requires that the regressors not only satisfy the standard no-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 $\text{plim}_{N,T\to\infty} W_{NT} > 0$ 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 likelihood expansion in theorem 3.1, *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 obtains consistency of the QMLE. 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 $W_{NT} = (NT)^{-1} \operatorname{Tr}(M_{f^0}\, vw'\, M_{\lambda^0}\, wv')$, and therefore $\text{plim}_{N,T\to\infty} W_{NT} = \text{plim}_{N,T\to\infty} (NT)^{-1} \operatorname{Tr}(M_{f^0}\, vw'\, M_{\lambda^0}\, wv') > 0$. 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 QMLE is not consistent (both β^0 and $\beta^0 + 1$ could be the true parameter, but the QMLE can not 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 can't 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 QMLE 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 can not be replaced by the weaker condition $\text{plim}_{N,T\to\infty} W_{NT} > 0$.

S.3 Some Matrix Algebra

The following statements are true for *real* matrices – throughout this paper no complex numbers appear.

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

$$\|A\|_1 = \max_{j=1\ldots m} \sum_{i=1}^{n} |A_{ij}| \ , \quad \|A\|_\infty = \max_{i=1\ldots n} \sum_{j=1}^{m} |A_{ij}| \ , \quad \|A\|_{\max} = \max_{i=1\ldots n} \max_{j=1\ldots m} |A_{ij}| \ . \tag{S.3.1}$$

Theorem S.3.1 (Some useful Inequalities)**.** *Let A be a $n \times m$ matrix, B be a $m \times p$ matrix, and C and D be $n \times n$ matrices. Then we have:*

$$\begin{array}{rl}
\text{(i)} & \|A\| \leq \|A\|_F \leq \|A\| \operatorname{rank}(A)^{1/2} \ , \\
\text{(ii)} & \|AB\| \leq \|A\| \, \|B\| \ , \\
\text{(iii)} & \|AB\|_F \leq \|A\|_F \, \|B\| \leq \|A\|_F \, \|B\|_F \ , \\
\text{(iv)} & |\mathrm{Tr}(AB)| \leq \|A\|_F \, \|B\|_F \ , \qquad \textit{for } n = p, \\
\text{(v)} & |\mathrm{Tr}\,(C)| \leq \|C\| \operatorname{rank}(C) \ , \\
\text{(vi)} & \|C\| \leq \mathrm{Tr}\,(C) \ , \qquad \textit{for } C \textit{ symmetric and } C \geq 0, \\
\text{(vii)} & \|A\|^2 \leq \|A\|_1 \, \|A\|_\infty \ , \\
\text{(viii)} & \|A\|_{\max} \leq \|A\| \leq \sqrt{nm} \, \|A\|_{\max} \ , \\
\text{(ix)} & \|A'CA\| \leq \|A'DA\| \ , \qquad \textit{for } C \textit{ symmetric and } C \leq D. \\
 & \textit{For } C,\ D \textit{ symmetric, and } i = 1, \ldots, n \textit{ we have:} \\
\text{(x)} & \mu_i(C) + \mu_n(D) \leq \mu_i(C+D) \leq \mu_i(C) + \mu_1(D) \ , \\
\text{(xi)} & \mu_i(C) \leq \mu_i(C+D) \ , \qquad \textit{for } D \geq 0, \\
\text{(xii)} & \mu_i(C) - \|D\| \leq \mu_i(C+D) \leq \mu_i(C) + \|D\| \ .
\end{array}$$

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

$$\begin{aligned}
\|AB\|_F^2 &= \mathrm{Tr}(ABB'A') \\
&= \mathrm{Tr}[\|B\|^2 AA' - A(\|B\|^2 \mathbb{I} - BB')A'] \\
&\leq \|B\|^2 \mathrm{Tr}(AA') = \|B\|^2 \ \|A\|_F^2 \ ,
\end{aligned} \tag{S.3.2}$$

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 $n \times \operatorname{rank}(C)$ that satisfy $U'U = O'O = \mathbb{I}$ and D is a $\operatorname{rank}(C) \times \operatorname{rank}(C)$

diagonal matrix with entries $s_i(C)$. We then have $\|O\| = \|U\| = 1$ and $\|D\| = \|C\|$ and therefore

$$\begin{aligned} |\mathrm{Tr}(C)| &= |\mathrm{Tr}(UDO')| = |\mathrm{Tr}(DO'U)| \\ &= \left| \sum_{i=1}^{\mathrm{rank}(C)} \eta_i' DO'U\eta_i \right| \\ &\le \sum_{i=1}^{\mathrm{rank}(C)} \|D\|\|O'\|\|U\| = \mathrm{rank}(C)\|C\| \, . \end{aligned} \tag{S.3.3}$$

For (vi) let e_1 be a vector that satisfied $\|e_1\| = 1$ and $\|C\| = e_1' C e_1$. 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 $\mathrm{Tr}\,(C) = \sum_i e_i' C e_i \ge e_1 C e_1 = \|C\|$, which is what we wanted to show. For (vii) we refer to Golub, van Loan (1996), p.15. For (viii) let e be the vector the vector that satisfies $\|e\| = 1$ and $\|A'CA\| = e'A'CAe$. Since $A'CA$ is symmetric such an e has to exist. Since $C \le D$ we then have $\|C\| = (e'A')C(Ae) \le (e'A')D(Ae) \le \|A'DA\|$. 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 $\|A'CA\| = e_1'A'DAe_1 - e_1'A'(D-C)Ae_1 \le e_1'A'DAe_1 \le \|A'DA\|$. 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 $\mu_n(D) \ge 0$ for $D \ge 0$, and since $-\|D\| \le \mu_i(D) \ge \|D\|$ for $i = 1, \ldots, n$. ∎

Definition S.3.2. *Let A be an $n \times r_1$ matrix and B be an $n \times r_2$ matrix with* $\mathrm{rank}(A) = r_1$ *and* $\mathrm{rank}(B) = r_2$. *The smallest principal angle $\theta_{A,B} \in [0, \pi/2]$ between the linear subspaces* $\mathrm{span}(A) = \{Aa |\, a \in \mathbb{R}^{r_1}\}$ *and* $\mathrm{span}(B) = \{Bb |\, b \in \mathbb{B}^{r_2}\}$ *of $\mathbb{R}^n$ is defined by*

$$\cos(\theta_{A,B}) = \max_{0 \ne a \in \mathbb{R}^{r_1}} \max_{0 \ne b \in \mathbb{R}^{r_2}} \frac{a'A'Bb}{\|Aa\|\|Bb\|} \, . \tag{S.3.4}$$

Theorem S.3.3. *Let A be an $n \times r_1$ matrix and B be an $n \times r_2$ matrix with* $\mathrm{rank}(A) = r_1$ *and* $\mathrm{rank}(B) = r_2$. *Then we have the following alternative characterizations of the smallest principal angle between* $\mathrm{span}(A)$ *and* $\mathrm{span}(B)$

$$\begin{aligned} \sin(\theta_{A,B}) &= \min_{0 \ne a \in \mathbb{R}^{r_1}} \frac{\|M_B\, A\, a\|}{\|A\, a\|} \\ &= \min_{0 \ne b \in \mathbb{R}^{r_2}} \frac{\|M_A\, B\, b\|}{\|B\, b\|} \, . \end{aligned} \tag{S.3.5}$$

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

$$\cos(\theta_{A,B}) = \min_{0 \ne a \in \mathbb{R}^{r_1}} \frac{\|P_B\, A\, a\|}{\|A\, a\|} = \min_{0 \ne b \in \mathbb{R}^{r_2}} \frac{\|P_A\, B\, b\|}{\|A\, b\|} \, . \tag{S.3.6}$$

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

Proof of Theorem B.1. Let

$$
\begin{aligned}
S_1(Z) &= \min_{f,\lambda} \operatorname{Tr}\left[\left(Z-\lambda f'\right)\left(Z'-f\lambda'\right)\right] \ , \\
S_2(Z) &= \min_{f} \operatorname{Tr}(Z\, M_f\, Z') \ , \\
S_3(Z) &= \min_{\lambda} \operatorname{Tr}(Z'\, M_\lambda\, Z) \ , \\
S_4(Z) &= \min_{\tilde\lambda,\tilde f} \operatorname{Tr}(M_{\tilde\lambda}\, Z\, M_{\tilde f}\, Z') \ , \\
S_5(Z) &= \sum_{i=R+1}^{T} \mu_i(Z'Z) \ , \\
S_6(Z) &= \sum_{i=R+1}^{N} \mu_i(ZZ') \ .
\end{aligned}
\tag{S.3.7}
$$

The theorem claims

$$
S_1(Z) \ = \ S_2(Z) \ = \ S_3(Z) \ = \ S_4(Z) \ = \ S_5(Z) \ = \ S_6(Z) \ . \tag{S.3.8}
$$

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

$$
\lambda'\, Z = \lambda'\, \lambda\, f' \ . \tag{S.3.9}
$$

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

$$
\begin{aligned}
\operatorname{Tr}\left[\left(Z-\lambda f'\right)'\left(Z-\lambda f'\right)\right] &= \operatorname{Tr}\left(Z'Z - Z'\lambda f'\right) \\
&= \operatorname{Tr}\left(Z'Z - Z'\lambda(\lambda'\lambda)^{-1}(\lambda'\lambda) f'\right) \\
&= \operatorname{Tr}\left(Z'Z - Z'\lambda(\lambda'\lambda)^{-1}(\lambda'\lambda)\lambda'\, Z\right) \\
&= \operatorname{Tr}\left(Z'\, M_\lambda\, Z\right) \ .
\end{aligned}
\tag{S.3.10}
$$

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

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

$$
\begin{aligned}
0 &\le \operatorname{Tr}\left\{\left(P_{\hat\lambda}[ZZ' - \omega_R\mathbb{I}_N]P_{\hat\lambda} - M_{\hat\lambda}[ZZ' - \omega_R\mathbb{I}_N]M_{\hat\lambda}\right)\left(M_\lambda - M_{\hat\lambda}\right)^2\right\} \\
&= \operatorname{Tr}\left\{\left(P_{\hat\lambda}[ZZ' - \omega_R\mathbb{I}_N]P_{\hat\lambda} + M_{\hat\lambda}[ZZ' - \omega_R\mathbb{I}_N]M_{\hat\lambda}\right)\left(M_\lambda - M_{\hat\lambda}\right)\right\} \\
&= \operatorname{Tr}\left[Z'\, M_\lambda\, Z\right] - \operatorname{Tr}\left[Z'\, M_{\hat\lambda}\, Z\right] + \omega_R\,\left[\operatorname{rank}(M_\lambda) - \operatorname{rank}(M_{\hat\lambda})\right] \ ,
\end{aligned}
\tag{S.3.11}
$$

[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.

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

$$\mathrm{Tr}\left[Z' M_{\hat{\lambda}} Z\right] \leq \mathrm{Tr}\left[Z' M_{\lambda} Z\right] . \tag{S.3.12}$$

This shows that $M_{\hat{\lambda}}$ is the optimal choice in the minimization problem of $S_3(Z)$, *i.e.* the optimal $\lambda = \hat{\lambda}$ is chosen such that the span of the N-dimensional vectors $\hat{\lambda}_r$ $(r = 1 \ldots R)$ 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 $\tilde{\lambda}$ such that the span of the N-dimensional vectors $\tilde{\lambda}_r$ $(r = 1 \ldots R_1)$ equals to the span of the R_1 eigenvectors that correspond to the R_1 largest eigenvalues of ZZ'. In addition, we can choose $\tilde{f}$ such that the span of the T-dimensional vectors $\tilde{f}_r$ $(r = 1 \ldots R_2)$ 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 $\tilde{\lambda}$ and $\tilde{f}$ 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 $M_{\tilde{\lambda}} \, Z \, M_{\tilde{f}} = Z - \tilde{Z}$, where

$$\tilde{Z} = P_{\tilde{\lambda}} \, Z \, M_{\tilde{f}} + Z \, P_{\tilde{f}} . \tag{S.3.13}$$

Since $\mathrm{rank}(Z) \leq \mathrm{rank}(P_{\tilde{\lambda}} \, Z \, M_{\tilde{f}}) + \mathrm{rank}(Z \, P_{\tilde{f}}) = R_1 + R_2 = R$, we can always write $\tilde{Z} = \lambda f'$ for some appropriate $N \times R$ and $T \times R$ matrices λ and f. This shows that

$$\begin{aligned} S_4(Z) &= \min_{\tilde{\lambda}, \tilde{f}} \mathrm{Tr}(M_{\tilde{\lambda}} \, Z \, M_{\tilde{f}} \, Z') \\ &\geq \min_{\{\tilde{Z} \,:\, \mathrm{rank}(\tilde{Z}) \leq R\}} \mathrm{Tr}((Z - \tilde{Z})(Z - \tilde{Z})') \\ &= \min_{f, \lambda} \mathrm{Tr}\left[\left(Z - \lambda f'\right)\left(Z' - f \lambda'\right)\right] = S_1(Z) . \end{aligned} \tag{S.3.14}$$

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.4 Supplement to the Consistency Proof (Appendix B)

Lemma S.4.1. *Under assumption 1 and 4 there exists a constant $B_0 > 0$ such that for the matrices w and v introduced in assumption 4 we have*

$$\begin{aligned} w' \, M_{\lambda^0} \, w \; - \; B_0 \, w' \, v \geq 0 \, , & \qquad \textit{wpa1,} \\ v' \, M_{f^0} \, v \; - \; B_0 \, v' \, v \geq 0 \, , & \qquad \textit{wpa1.} \end{aligned} \tag{S.4.1}$$

Proof. We can decompose $w = \widetilde{w} \, \bar{w}$, where $\widetilde{w}$ is a $N \times \mathrm{rank}(w)$ 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

$$\frac{\|M_v \, \lambda^0 \, a\|^2}{\|\lambda^0 \, a\|^2} \; = \; \frac{a' \, \lambda^{0\prime} \, M_v \, \lambda^0 \, a}{a' \, \lambda^{0\prime} \, \lambda^0 \, a} > \frac{B}{B_1} \, , \qquad \text{wpa1.} \tag{S.4.2}$$

Applying theorem S.3.3 we find

$$\min_{0\neq b\in\mathbb{R}^{\mathrm{rank}(w)}} \frac{b'\,\widetilde{w}'\,M_{\lambda^0}\,\widetilde{w}\,b}{b'\,\widetilde{w}'\,\widetilde{w}\,b} = \min_{0\neq a\in\mathbb{R}^R} \frac{a'\,\lambda^{0\prime}\,M_w\,\lambda^0\,a}{a'\,\lambda^{0\prime}\,\lambda^0\,a} > \frac{B}{B_1}\,, \qquad \text{wpa1.} \tag{S.4.3}$$

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.4.2. *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*

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

$$\min_{\{\alpha\in\mathbb{R}^{K_1},\|\alpha\|=1\}} \frac{\left\|M_{\lambda^0}\,X_{\mathrm{low},\alpha}\,M_{f^0}\,X'_{\mathrm{low},\alpha}\,M_{\lambda^0}\right\|}{NT} > B\,, \qquad \textit{wpa1.} \tag{S.4.4}$$

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.4.1 and part (v), (vi) and (ix) of theorem S.3.1 we find

$$\begin{aligned}
\left\|M_{\lambda^0}\,X_{\mathrm{low},\alpha}\,M_{f^0}\,X'_{\mathrm{low},\alpha}\,M_{\lambda^0}\right\| &= \left\|M_{\lambda^0}\,w\,\mathrm{diag}(\alpha')\,v'\,M_{f^0}\,v\,\mathrm{diag}(\alpha')\,w'M_{\lambda^0}\right\| \\
&\geq B_0\,\left\|M_{\lambda^0}\,w\,\mathrm{diag}(\alpha')\,v'\;v\,\mathrm{diag}(\alpha')\,w'M_{\lambda^0}\right\| \\
&\geq \frac{B_0}{K_1}\,\mathrm{Tr}\left[M_{\lambda^0}\,w\,\mathrm{diag}(\alpha')\,v'\;v\,\mathrm{diag}(\alpha')\,w'M_{\lambda^0}\right] \\
&= \frac{B_0}{K_1}\,\mathrm{Tr}\left[v\,\mathrm{diag}(\alpha')\,w'M_{\lambda^0}w\,\mathrm{diag}(\alpha')\,v'\right] \\
&\geq \frac{B_0}{K_1}\,\left\|v\,\mathrm{diag}(\alpha')\,w'M_{\lambda^0}w\,\mathrm{diag}(\alpha')\,v'\right\| \\
&\geq \frac{B_0^2}{K_1}\,\left\|v\,\mathrm{diag}(\alpha')\,w'w\,\mathrm{diag}(\alpha')\,v'\right\| \\
&\geq \frac{B_0^2}{K_1^2}\,\mathrm{Tr}\left[v\,\mathrm{diag}(\alpha')\,w'w\,\mathrm{diag}(\alpha')\,v'\right] \\
&= \frac{B_0^2}{K_1^2}\mathrm{Tr}\left[X_{\mathrm{low},\alpha}X'_{\mathrm{low},\alpha}\right]\,.
\end{aligned} \tag{S.4.5}$$

Thus we have $\left\|M_{\lambda^0}\,X_{\mathrm{low},\alpha}\,M_{f^0}\,X'_{\mathrm{low},\alpha}\,M_{\lambda^0}\right\|/(NT) \geq (B_0/K_1)^2\,\alpha'\,W^{\mathrm{low}}_{NT}\,\alpha$, where the $K_1\times K_1$ matrix W^{low}_{NT} is defined by $W^{\mathrm{low}}_{NT,l_1l_2} = (NT)^{-1}\mathrm{Tr}\left(X_{l_1}X'_{l_2}\right)$, *i.e.* it is a submatrix of W_{NT}. Since by assumption W_{NT} and thus W^{low}_{NT} 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 $\tilde{S}^{(2)}_{NT}(\beta, f)$ that was used in the consistency proof. Remember that

$$\tilde{S}^{(2)}_{NT}(\beta, f) = \frac{1}{NT} \operatorname{Tr} \left[\left(\lambda^0 f^{0\prime} + \sum_{k=1}^{K} (\beta^0_k - \beta_k) X_k \right) M_f \left(\lambda^0 f^{0\prime} + \sum_{k=1}^{K} (\beta^0_k - \beta_k) X_k \right)' P_{(\lambda_0, w)} \right] . \tag{S.4.6}$$

We want to show that under the assumptions of theorem 2.1 there exist finite positive constants a_0, a_1, a_2, a_3 and a_4 such that

$$\begin{aligned} \tilde{S}^{(2)}_{NT}(\beta, f) \geq & \frac{a_0 \left\| \beta^{\text{low}} - \beta^{\text{low}}_0 \right\|^2}{\left\| \beta^{\text{low}} - \beta^{\text{low}}_0 \right\|^2 + a_1 \left\| \beta^{\text{low}} - \beta^{\text{low}}_0 \right\| + a_2} \\ & - a_3 \left\| \beta^{\text{high}} - \beta^{\text{high}}_0 \right\| - a_4 \left\| \beta^{\text{high}} - \beta^{\text{high}}_0 \right\| \left\| \beta^{\text{low}} - \beta^{\text{low}}_0 \right\| , \qquad \text{wpa1.} \end{aligned} \tag{S.4.7}$$

Proof of the lower bound on $\tilde{S}^{(2)}_{NT}(\beta, f)$. Applying theorem B.1 and part (xi) of theorem S.3.1 we find that

$$\begin{aligned} \tilde{S}^{(2)}_{NT}(\beta, f) \geq & \frac{1}{NT} \mu_{R+1} \left[\left(\lambda^0 f^{0\prime} + \sum_{k=1}^{K} (\beta^0_k - \beta_k) X_k \right)' P_{(\lambda_0, w)} \left(\lambda^0 f^{0\prime} + \sum_{k=1}^{K} (\beta^0_k - \beta_k) X_k \right) \right] \\
= & \frac{1}{NT} \mu_{R+1} \Bigg[\left(\lambda^0 f^{0\prime} + \sum_{l=1}^{K_1} (\beta^0_l - \beta_l) w_l v_l' \right)' \left(\lambda^0 f^{0\prime} + \sum_{l=1}^{K_1} (\beta^0_l - \beta_l) w_l v_l' \right) \\
& + \left(\lambda^0 f^{0\prime} + \sum_{l=1}^{K_1} (\beta^0_l - \beta_l) w_l v_l' \right)' P_{(\lambda_0, w)} \sum_{m=K_1}^{K} (\beta^0_m - \beta_m) X_m \\
& + \sum_{m=K_1}^{K} (\beta^0_m - \beta_m) X_m' P_{(\lambda_0, w)} \left(\lambda^0 f^{0\prime} + \sum_{l=1}^{K_1} (\beta^0_l - \beta_l) w_l v_l' \right) \\
& + \sum_{m=K_1}^{K} (\beta^0_m - \beta_m) X_m' P_{(\lambda_0, w)} \sum_{m=K_1}^{K} (\beta^0_m - \beta_m) X_m \Bigg] \\
\geq & \frac{1}{NT} \mu_{R+1} \Bigg[\left(\lambda^0 f^{0\prime} + \sum_{l=1}^{K_1} (\beta^0_l - \beta_l) w_l v_l' \right)' \left(\lambda^0 f^{0\prime} + \sum_{l=1}^{K_1} (\beta^0_l - \beta_l) w_l v_l' \right) \\
& + \left(\lambda^0 f^{0\prime} + \sum_{l=1}^{K_1} (\beta^0_l - \beta_l) w_l v_l' \right)' P_{(\lambda_0, w)} \sum_{m=K_1}^{K} (\beta^0_m - \beta_m) X_m \\
& + \sum_{m=K_1}^{K} (\beta^0_m - \beta_m) X_m' P_{(\lambda_0, w)} \left(\lambda^0 f^{0\prime} + \sum_{l=1}^{K_1} (\beta^0_l - \beta_l) w_l v_l' \right) \Bigg] \\
\geq & \frac{1}{NT} \mu_{R+1} \left[\left(\lambda^0 f^{0\prime} + \sum_{l=1}^{K_1} (\beta^0_l - \beta_l) w_l v_l' \right)' \left(\lambda^0 f^{0\prime} + \sum_{l=1}^{K_1} (\beta^0_l - \beta_l) w_l v_l' \right) \right] \\
& - a_3 \left\| \beta^{\text{high}} - \beta^{\text{high}}_0 \right\| - a_4 \left\| \beta^{\text{high}} - \beta^{\text{high}}_0 \right\| \left\| \beta^{\text{low}} - \beta^{\text{low}}_0 \right\| , \qquad \text{wpa1,} \end{aligned} \tag{S.4.8}$$

where $a_3 > 0$ and $a_4 > 0$ are appropriate constants. For the last step we used part (xii) of theorem S.3.1 and the fact that

$$\frac{1}{NT}\left\|\sum_{m=K_1}^{K}(\beta_m^0-\beta_m)X_m' P_{(\lambda_0,w)}\left(\lambda^0 f^{0\prime}+\sum_{l=1}^{K_1}(\beta_l^0-\beta_l)w_l v_l'\right)\right\|$$
$$\leq K\left\|\beta^{\text{high}}-\beta_0^{\text{high}}\right\|\max_m\left\|\frac{X_m}{\sqrt{NT}}\right\|\left(\left\|\frac{\lambda^0 f^{0\prime}}{\sqrt{NT}}\right\|+K\left\|\beta^{\text{low}}-\beta_0^{\text{low}}\right\|\max_l\left\|\frac{w_l v_l'}{\sqrt{NT}}\right\|\right). \tag{S.4.9}$$

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 $\tilde{S}^{(2)}_{NT}(\beta,f)\geq \mu_{R+1}(A'A)/(NT)+$ terms containing β^{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

$$\begin{aligned}A'A\;&\geq\;A'A-(a^{1/2}A_3'+a^{-1/2}A_2')(a^{1/2}A_3+a^{-1/2}A_2)\\&=\left[A_1'A_1-(a-1)\,A_3'A_3\right]\;+\;(1-a^{-1})A_2'A_2\,,\end{aligned}\tag{S.4.10}$$

where $\geq$ for matrices refers to the difference being positive definite, and a is a positive number, namely $a=1+\mu_R(A_1'A_1)/(2\,\|A_3\|^2)$. 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

$$\begin{aligned}\frac{1}{NT}\,\mu_{R+1}\left(A'A\right)&\geq\frac{1}{NT}\,\mu_{R+1}\left[\left(A_1'A_1-(a-1)\,A_3'A_3\right)\;+\;(1-a^{-1})A_2'A_2\right]\\&\geq\frac{1}{NT}\,\min\left\{(1-a^{-1})\|A_2\|^2\,,\;\;\mu_R\left[A_1'A_1-(a-1)\,A_3'A_3\right]\right\}\;.\end{aligned}\tag{S.4.11}$$

Using theorem S.3.1(xii) and our particular choice of a we find

$$\begin{aligned}\mu_R\;\left[A_1'A_1-(a-1)\,A_3'A_3\right]&\geq\;\mu_R(A_1'A_1)-\left\|(a-1)A_3'A_3\right\|\\&=\frac{1}{2}\,\mu_R(A_1'A_1)\;.\end{aligned}\tag{S.4.12}$$

Therefore

$$\begin{aligned}\frac{1}{NT}\,\mu_{R+1}(A'A)&\geq\frac{1}{2\,NT}\,\mu_R(A_1'A_1)\,\min\left\{1\,,\;\;\frac{2\,\|A_2\|^2}{2\,\|A_3\|^2+\mu_R(A_1'A_1)}\right\}\\&\geq\frac{1}{NT}\,\frac{\|A_2\|^2\,\mu_R(A_1'A_1)}{2\,\|A\|^2+\mu_R(A_1'A_1)}\,,\end{aligned}\tag{S.4.13}$$

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

$$
\begin{aligned}
\frac{\|A\|}{\sqrt{NT}} &\leq \frac{\|\lambda^0 f^{0\prime}\|}{\sqrt{NT}} + \sum_{l=1}^{K_1} |\beta_l^0 - \beta_l| \frac{\|w_l v_l'\|}{\sqrt{NT}} \leq c_0 + c_1 \left\| \beta^{\rm low} - \beta_0^{\rm low} \right\| , \quad \text{wpa1} , \\
\frac{\mu_R(A_1' A_1)}{NT} &= \frac{\mu_R\left(f^0 \lambda^{0\prime} M_w \lambda^0 f^{0\prime}\right)}{NT} \geq c_2 , \quad \text{wpa1} , \\
\frac{\|A_2\|^2}{NT} &= \mu_1 \left[\sum_{l_1=1}^{K_1} (\beta_{l_1}^0 - \beta_{l_1})\, w_{l_1} v_{l_1}' \, M_{f^0} \sum_{l_2=1}^{K_1} (\beta_{l_2}^0 - \beta_{l_2})\, v_{l_2} w_{l_2}' \right] \\
&\geq c_3 \left\| \beta^{\rm low} - \beta_0^{\rm low} \right\|^2 , \quad \text{wpa1} ,
\end{aligned}
\tag{S.4.14}
$$

were for the last inequality we used lemma S.4.2.

We thus have

$$
\frac{1}{NT}\, \mu_{R+1}\left(A'A\right) \geq \frac{c_3 \left\| \beta^{\rm low} - \beta_0^{\rm low} \right\|^2}{1 + \frac{2}{c_2}\left(c_0 + c_1 \left\| \beta^{\rm low} - \beta_0^{\rm low} \right\|\right)^2} , \quad \text{wpa1} . \tag{S.4.15}
$$

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

$$
\frac{1}{NT}\, \mu_{R+1}\left(A'A\right) \geq \frac{a_0 \left\| \beta^{\rm low} - \beta_0^{\rm low} \right\|^2}{\left\| \beta^{\rm low} - \beta_0^{\rm low} \right\|^2 + a_1 \left\| \beta^{\rm low} - \beta_0^{\rm low} \right\| + a_2} , \quad \text{wpa1} , \tag{S.4.16}
$$

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

S.5 Proof of Corollary 3.2

All that is left to show for Corollary 3.2 is that the matrix $W_{NT} = W_{NT}(\lambda^0, f^0, X_k)$ does not become singular as $N, T \to \infty$.

Proof. Remember that

$$
W_{NT} = \frac{1}{NT} \mathrm{Tr}(M_{f^0} X_{k_1}' M_{\lambda^0} X_{k_2}) . \tag{S.5.1}
$$

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

$$
\begin{aligned}
\mu_K\left(W_{NT}\right) &= \min_{\{a \in \mathbb{R}^K,\, a \neq 0\}} \frac{a' W_{NT}\, a}{\|a\|^2} \\
&= \min_{\{a \in \mathbb{R}^K,\, a \neq 0\}} \frac{1}{NT\, \|a\|^2} \mathrm{Tr}\left[M_{f^0} \left(\sum_{k_1=1}^{K} a_{k_1} X_{k_1}' \right) M_{\lambda^0} \left(\sum_{k_2=1}^{K} a_{k_2} X_{k_2} \right) \right] \\
&= \min_{\substack{\{\alpha \in \mathbb{R}^{K_1},\ \varphi \in \mathbb{R}^{K_2} \\ \alpha \neq 0,\ \varphi \neq 0\}}} \frac{\mathrm{Tr}\left[M_{f^0} \left(X_{{\rm low},\varphi}' + X_{{\rm high},\alpha}' \right) M_{\lambda^0} \left(X_{{\rm low},\varphi} + X_{{\rm high},\alpha} \right) \right]}{NT\, \left(\|\alpha\|^2 + \|\varphi\|^2\right)} ,
\end{aligned}
\tag{S.5.2}
$$

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]

$$X_{\text{low},\varphi} = \sum_{l=1}^{K_1} \varphi_l \, X_l \, , \qquad\qquad X_{\text{high},\alpha} = \sum_{m=K_1+1}^{K} \alpha_m \, X_m \, . \tag{S.5.3}$$

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 $N \times (R + K_1)$ matrix, while $M_{\lambda^0} w$ is also a $N \times K_1$ matrix. Using this we obtain

$$\begin{aligned}
&\mu_K\left(W_{NT}\right) \\
&= \min_{\substack{\{\varphi \in \mathbb{R}^{K_1},\ \alpha \in \mathbb{R}^{K_2} \\ \varphi \neq 0,\ \alpha \neq 0\}}} \frac{1}{NT\ (\|\varphi\|^2 + \|\alpha\|^2)} \Big\{ \mathrm{Tr}\left[M_{f^0} \ \left(X'_{\text{low},\varphi} + X'_{\text{high},\alpha}\right) \ M_{(\lambda^0,w)} \ \left(X_{\text{low},\varphi} + X_{\text{high},\alpha}\right)\right] \\
&\qquad\qquad + \mathrm{Tr}\left[M_{f^0} \ \left(X'_{\text{low},\varphi} + X'_{\text{high},\alpha}\right) \ P_{(M_{\lambda^0} w)} \ \left(X_{\text{low},\varphi} + X_{\text{high},\alpha}\right)\right] \Big\} \\
&= \min_{\substack{\{\varphi \in \mathbb{R}^{K_1},\ \alpha \in \mathbb{R}^{K_2} \\ \varphi \neq 0,\ \alpha \neq 0\}}} \frac{1}{NT\ (\|\varphi\|^2 + \|\alpha\|^2)} \Big\{ \mathrm{Tr}\left[M_{f^0} \, X'_{\text{high},\alpha} \, M_{(\lambda^0,w)} \, X_{\text{high},\alpha}\right] \\
&\qquad\qquad + \mathrm{Tr}\left[M_{f^0} \ \left(X'_{\text{low},\varphi} + X'_{\text{high},\alpha}\right) \ P_{(M_{\lambda^0} w)} \ \left(X_{\text{low},\varphi} + X_{\text{high},\alpha}\right)\right] \Big\} \, .
\end{aligned} \tag{S.5.4}$$

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

$$\begin{aligned}
\frac{1}{NT} \mathrm{Tr}\left[M_{f^0} \, X'_{\text{high},\alpha} \, M_{(\lambda^0,w)} \, X_{\text{high},\alpha}\right] &\geq \ c_1 \|\alpha\|^2 \ , \quad \text{wpa1}, \\
\frac{1}{NT} \mathrm{Tr}\left[M_{f^0} \ \left(X'_{\text{low},\varphi} + X'_{\text{high},\alpha}\right) \ P_{(M_{\lambda^0} w)} \ \left(X_{\text{low},\varphi} + X_{\text{high},\alpha}\right)\right] &\geq 0 \ , \\
\frac{1}{NT} \mathrm{Tr}\left[M_{f^0} \, X'_{\text{low},\varphi} \, P_{(M_{\lambda^0} w)} \, X_{\text{low},\varphi}\right] &\geq \ c_2 \, \|\varphi\|^2 \ , \quad \text{wpa1}, \\
\frac{1}{NT} \mathrm{Tr}\left[M_{f^0} \, X'_{\text{low},\varphi} \, P_{(M_{\lambda^0} w)} \, X_{\text{high},\alpha}\right] &\geq -\frac{c_3}{2} \, \|\varphi\| \|\alpha\| \ , \quad \text{wpa1}, \\
\frac{1}{NT} \mathrm{Tr}\left[M_{f^0} \, X'_{\text{high},\alpha} \, P_{(M_{\lambda^0} w)} \, X_{\text{high},\alpha}\right] &\geq 0 \ ,
\end{aligned} \tag{S.5.5}$$

and we want to justify these inequalities now. The second and the last equation in (S.5.5) are true because *e.g.* $\mathrm{Tr}\left[M_{f^0} \, X'_{\text{high},\alpha} \, P_{(M_{\lambda^0} w)} \, X_{\text{high},\alpha}\right] = \mathrm{Tr}\left[M_{f^0} \, X'_{\text{high},\alpha} \, P_{(M_{\lambda^0} w)} \, X_{\text{high},\alpha} \, M_{f^0}\right]$, and the trace of a symmetric positive semi-definite matrix is non-negative. The first inequality in (S.5.5) is true because $\mathrm{rank}(f^0) + \mathrm{rank}(\lambda^0, w) = 2R + K_1$ and using theorem B.1 and assumption 4 we have

$$\frac{1}{NT\|\alpha\|^2} \mathrm{Tr}\left[M_{f^0} \, X'_{\text{high},\alpha} \, M_{(\lambda^0,w)} \, X_{\text{high},\alpha}\right] \geq \frac{1}{NT\|\alpha\|^2} \mu_{2R+K_1+1}\left[X_{\text{high},\alpha} \, X'_{\text{high},\alpha}\right] > b \ , \quad \text{wpa1}, \tag{S.5.6}$$

i.e. we can set $c_1 = b$. The third inequality in (S.5.5) is true because according theorem S.3.1(v)

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

we have

$$
\begin{aligned}
\frac{1}{NT}\mathrm{Tr}\left[M_{f^0}\,X'_{\mathrm{low},\varphi}\,P_{(M_{\lambda^0}w)}\,X_{\mathrm{high},\alpha}\right] &\geq -\frac{K_1}{NT}\,\|X_{\mathrm{low},\varphi}\|\,\|X_{\mathrm{high},\alpha}\| \\
&\geq -\frac{K_1}{NT}\,\|X_{\mathrm{low},\varphi}\|_F\,\|X_{\mathrm{high},\alpha}\|_F \\
&\geq -K_1\,K_1\,K_2\,\|\varphi\|\,\|\alpha\|\,\max_{k_1=1\ldots K_1}\left\|\frac{X_{k_1}}{\sqrt{NT}}\right\|_F\,\max_{k_2=K_1+1\ldots K}\left\|\frac{X_{k_2}}{\sqrt{NT}}\right\|_F \\
&\geq -\frac{c_3}{2}\,\|\varphi\|\,\|\alpha\|\,, \qquad\qquad (\mathrm{S.5.7})
\end{aligned}
$$

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.5.5) holds. Note that $X'_{\mathrm{low},\varphi}\,P_{(M_{\lambda^0}w)}\,X_{\mathrm{low},\varphi} = X'_{\mathrm{low},\varphi}\,M_{\lambda^0}\,X_{\mathrm{low},\varphi}$, *i.e.* we need to show that

$$
\frac{1}{NT}\mathrm{Tr}\left[M_{f^0}\,X'_{\mathrm{low},\varphi}\,M_{\lambda^0}\,X_{\mathrm{low},\varphi}\right] \geq\; c_2\,\|\varphi\|^2\,. \qquad (\mathrm{S.5.8})
$$

Using part (vi) or theorem S.3.1 we find

$$
\begin{aligned}
\frac{1}{NT}\mathrm{Tr}\left[M_{f^0}\,X'_{\mathrm{low},\varphi}\,M_{\lambda^0}\,X_{\mathrm{low},\varphi}\right] &= \frac{1}{NT}\mathrm{Tr}\left[M_{\lambda^0}\,X_{\mathrm{low},\varphi}\,M_{f^0}\,X'_{\mathrm{low},\varphi}\,M_{\lambda^0}\right] \\
&\geq \frac{1}{NT}\left\|M_{\lambda^0}\,X_{\mathrm{low},\varphi}\,M_{f^0}\,X'_{\mathrm{low},\varphi}\,M_{\lambda^0}\right\|\,, \qquad (\mathrm{S.5.9})
\end{aligned}
$$

and according to lemma S.4.2 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.5.5) in equation (S.5.4) we obtain

$$
\begin{aligned}
\mu_K\left(W_{NT}\right) &\geq \min_{\substack{\{\varphi\in\mathbb{R}^{K_1},\ \alpha\in\mathbb{R}^{K_2}\\ \varphi\neq 0,\ \alpha\neq 0\}}}\frac{1}{\|\varphi\|^2+\|\alpha\|^2}\left\{c_1\|\alpha\|^2+\max\left[0,\,c_2\|\varphi\|^2-c_3\|\varphi\|\|\alpha\|\right]\right\} \\
&\geq \min\left(\frac{c_2}{2},\,\frac{c_1c_2^2}{c_2^2+c_3^2}\right)\,,\quad \text{wpa1}. \qquad (\mathrm{S.5.10})
\end{aligned}
$$

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

S.6 Supplement to the Proof of Theorem 3.3

Proof of lemma C.1. # We start with the proof for the operator norm of the weakly exogenous part of the regressors. We can rewrite equation (3.2) as

$$
X_k^{\mathrm{weak}} = \sum_{\tau=1}^{T-1} C_{k,\tau}\,e_{(\tau)}\,, \qquad (\mathrm{S.6.1})
$$

where the $C_{k,\tau}$ are diagonal $N\times N$ matrices with diagonal entries $C_{k,\tau,ii} = c_{k,i\tau}$, and the $e_{(\tau)}$ are $N\times T$ matrix obtained by shifting all entries of e downwards by p rows and filling the upper p rows with zeros, *i.e.* $e_{(\tau),it} = e_{i,t-\tau}$ for $t > p$, and $e_{(\tau),it} = 0$ otherwise. We then have $\|e_{(\tau)}\| \leq \|e\|$ and $\|C_{k,\tau}\| = \max_i |c_{k,i\tau}| \leq \alpha^\tau$. Therefore

$$\begin{aligned}
\left\| X_k^{\text{weak}} \right\| &\leq \sum_{\tau=1}^{T-1} \|C_{k,\tau}\| \, \|e_{(\tau)}\| \\
&\leq \sum_{\tau=1}^{T-1} \alpha^\tau \, \|e\| \\
&\leq \frac{\alpha}{1-\alpha} \, \mathcal{O}_p(N^{1/2}) = \mathcal{O}_p(N^{1/2}) \, . \qquad \text{(S.6.2)}
\end{aligned}$$

This is what we wanted to show.

To prove the result on the operator norm of $P_{\lambda^0} e P_{f^0}$, we first find that $\frac{1}{\sqrt{NT}} \left\| f^{0\prime} e \lambda^0 \right\|_F = \mathcal{O}_p(1)$, because

$$\begin{aligned}
\mathbb{E} \left(\frac{\left\| f^{0\prime} e \lambda^0 \right\|_F}{\sqrt{NT}} \right)^2 &= \frac{1}{NT} E \left(\sum_{i=1}^N \sum_{t=1}^T e_{it} f_t^{0\prime} \lambda_i^0 \right)^2 \\
&= \frac{1}{NT} \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^T \sum_{s=1}^T E\left(e_{it} e_{js}\right) f_t^{0\prime} \lambda_i^0 \lambda_j^{0\prime} f_s^0 \\
&= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \mathbb{E}\left(e_{it}^2\right) f_t^{0\prime} \lambda_i^0 \lambda_i^{0\prime} f_t^0 \\
&= \mathcal{O}(1) \, . \qquad \text{(S.6.3)}
\end{aligned}$$

Using this we obtain

$$\begin{aligned}
\|P_{\lambda^0} e P_{f^0}\| &= \|\lambda^0 (\lambda^{0\prime}\lambda^0)^{-1} \lambda^{0\prime} e f^0 (f^{0\prime} f^0)^{-1} f^{0\prime}\| \\
&\leq \|\lambda^0\| \left\| (\lambda^{0\prime}\lambda^0)^{-1} \right\| \|\lambda^{0\prime} e f^0\| \left\| (f^{0\prime} f^0)^{-1} \right\| \|f^{0\prime}\| \\
&\leq \mathcal{O}_p(N^{1/2}) \mathcal{O}_p(N^{-1}) \|\lambda^{0\prime} e f^0\|_F \mathcal{O}_p(T^{-1}) \mathcal{O}_p(T^{1/2}) = \mathcal{O}_p(1) \, , \qquad \text{(S.6.4)}
\end{aligned}$$

where we used part (i) and (ii) of theorem S.3.1.

To show the next statement of the lemma we first note that $N^{-1}T^{-1/2}\|\lambda^{0\prime} e X_k^{\text{str}\prime}\|_F = \mathcal{O}_p(1)$, which is true because

$$\begin{aligned}
\mathbb{E}\left(N^{-1}T^{-1/2}\|\lambda^{0\prime} e X_k^{\text{str}\prime}\|_F\right)^2 &= N^{-2}T^{-1} \mathbb{E}\left[\sum_{r=1}^R \sum_{j=1}^N \left(\sum_{i=1}^N \sum_{t=1}^T \lambda_{ir}^0 e_{it} X_{k,jt}^{\text{str}} \right)^2 \right] \\
&= N^{-2}T^{-1} \sum_{r=1}^R \sum_{j=1}^N \sum_{i=1}^N \sum_{t=1}^T \mathbb{E}\left(e_{it}^2\right) (\lambda_{ir}^0)^2 \mathbb{E}\left[(X_{k,jt}^{\text{str}})^2\right] = \mathcal{O}(1) \, .
\end{aligned}$$

(S.6.5)

Therefore we find

$$\begin{aligned}
\|P_{\lambda^0} e X_k^{\text{str}\prime}\| &= \|\lambda^0 (\lambda^{0\prime}\lambda^0)^{-1} \lambda^{0\prime} e X_k^{\text{str}\prime}\| \\
&\leq \|\lambda^0 (\lambda^{0\prime}\lambda^0)^{-1}\| \|\lambda^{0\prime} e X_k^{\text{str}\prime}\| \\
&\leq \|\lambda^0 (\lambda^{0\prime}\lambda^0)^{-1}\| \|\lambda^{0\prime} e X_k^{\text{str}\prime}\|_F = \mathcal{O}_p(\sqrt{NT}) \, , \qquad \text{(S.6.6)}
\end{aligned}$$

where we used part (i) and (ii) of theorem S.3.1. This is what we wanted to show.

The proof for $\|P_{f^0} e' X_k^{\rm str}\| = \mathcal{O}_p(\sqrt{NT})$ is analogous. ∎

Lemma S.6.1. *Suppose that A and B are a $T\times T$ and an $N\times N$ matrices that are independent of e such that $\mathbb{E}\,\|A\|_F^2 = \mathcal{O}\,(NT)$ and $\mathbb{E}\,\|B\|_F^2 = \mathcal{O}\,(NT)$, and let assumption 5 be satisfied. Then there exists a finite constant c_0, independent of N,T, such that*

$$
\begin{aligned}
&(a) \qquad && \mathbb{E}\left\{\mathrm{Tr}\left[\left(e'e - \mathbb{E}\left(e'e\right)\right) A\right]\right\}^2 \leq c_0\, N\, \mathbb{E}\left(\|A\|_F^2\right) \ , \\
&(b) \qquad && \mathbb{E}\left\{\mathrm{Tr}\left[\left(ee' - \mathbb{E}\left(ee'\right)\right) B\right]\right\}^2 \leq c_0\, T\, \mathbb{E}\left(\|B\|_F^2\right) \ . \qquad \text{(S.6.7)}
\end{aligned}
$$

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

$$
\begin{aligned}
\mathrm{Tr}\left\{\left(e'e - \mathbb{E}\left(e'e\right)\right) A\right\} &= \sum_{t=1}^{T}\sum_{s=1}^{T} \left(e'e - \mathbb{E}\left(e'e\right)\right)_{ts} A_{ts} \\
&= \sum_{t=1}^{T}\sum_{s=1}^{T} \left(\sum_{i=1}^{N} \left(e_{it}e_{is} - \mathbb{E}\left(e_{it}e_{is}\right)\right)\right) A_{ts}. \qquad \text{(S.6.8)}
\end{aligned}
$$

To compute its variance, we write

$$
\begin{aligned}
&\mathbb{E}\left(\mathrm{Tr}\left\{\left(e'e - \mathbb{E}\left(e'e\right)\right) A\right\}\right)^2 \\
&\quad = \sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{p=1}^{T}\sum_{q=1}^{T} \mathbb{E}\left\{\left(\sum_{i=1}^{N} \left(e_{it}e_{is} - \mathbb{E}\left(e_{it}e_{is}\right)\right)\right)\left(\sum_{j=1}^{N} \left(e_{jp}e_{jq} - \mathbb{E}\left(e_{jp}e_{jq}\right)\right)\right)\right\} A_{ts}A_{pq}. \qquad \text{(S.6.9)}
\end{aligned}
$$

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

$$
\begin{aligned}
&\mathbb{E}\left\{\left(\sum_{i=1}^{N} \left(e_{it}e_{is} - \mathbb{E}\left(e_{it}e_{is}\right)\right)\right)\left(\sum_{j=1}^{N} \left(e_{jp}e_{jq} - \mathbb{E}\left(e_{jp}e_{jq}\right)\right)\right)\right\} \\
&\qquad = \sum_{i=1}^{N}\sum_{j=1}^{N} \left\{\mathbb{E}\left(e_{it}e_{is}e_{jp}e_{jq}\right) - \mathbb{E}\left(e_{it}e_{is}\right)\mathbb{E}\left(e_{jp}e_{jq}\right)\right\} \\
&\qquad = \begin{cases} \Sigma_{it}\Sigma_{is} & \text{if } \ (t=p)\neq(s=q) \ \text{ and } \ (i=j) \\ \Sigma_{it}\Sigma_{is} & \text{if } \ (t=q)\neq(s=p) \ \text{ and } \ (i=j) \\ \mathbb{E}\left(e_{it}^4\right) - \Sigma_{it}^2 & \text{if } \ (t=s=p=q) \ \text{ and } \ (i=j) \\ 0 & \text{otherwise.} \end{cases} \qquad \text{(S.6.10)}
\end{aligned}
$$

Therefore,

$$
\mathbb{E}\left(\mathrm{Tr}\left\{\left(e'e - \mathbb{E}\left(e'e\right)\right) A\right\}\right)^2 \leq \sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N} \Sigma_{it}\Sigma_{is}\left(\mathbb{E}\left(A_{ts}^2\right) + \mathbb{E}\left(A_{ts}A_{st}\right)\right) + \sum_{t=1}^{T}\sum_{i=1}^{N} \left(\mathbb{E}\left(e_{it}^4\right) - \Sigma_{it}^2\right)\mathbb{E}A_{tt}^2. \qquad \text{(S.6.11)}
$$

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

$$\begin{aligned}
\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\Sigma_{it}\Sigma_{is}\left(\mathbb{E}A_{ts}^2\right) &= \mathbb{E}\left(\sum_{i=1}^{N}\operatorname{Tr}\left(A'\Sigma^i A\Sigma^i\right)\right) \\
&\leq \sum_{i=1}^{N}\mathbb{E}\left\|A\Sigma^i\right\|_F^2 \leq \sum_{i=1}^{N}\left\|\Sigma^i\right\|^2 \mathbb{E}\left\|A\right\|_F^2 \\
&\leq N\left(\sup_{it}\Sigma_{it}^2\right)\mathbb{E}\left\|A\right\|_F^2 . \qquad \text{(S.6.12)}
\end{aligned}$$

Also,

$$\begin{aligned}
\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\Sigma_{it}\Sigma_{is}\mathbb{E}\left(A_{ts}A_{st}\right) &= \mathbb{E}\left[\sum_{i=1}^{N}\operatorname{Tr}\left(\Sigma^i A A\Sigma^i\right)\right] \\
&\leq \sum_{i=1}^{N}\mathbb{E}\left\|\Sigma^i A\right\|_F \left\|A\Sigma^i\right\|_F \leq \sum_{i=1}^{N}\left\|\Sigma^i\right\|^2 \mathbb{E}\left\|A\right\|_F^2 \\
&\leq N\left(\sup_{it}\Sigma_{it}^2\right)\mathbb{E}\left\|A\right\|_F^2 \ . \qquad \text{(S.6.13)}
\end{aligned}$$

Finally,

$$\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\mathbb{E}\left(e_{it}^4\right) - \Sigma_{it}^2\right)\mathbb{E}A_{tt}^2 \ \leq \ N\left(\sup_{it}\mathbb{E}\left(e_{it}^4\right)\right)\mathbb{E}\left\|A\right\|_F^2 . \qquad \text{(S.6.14)}$$

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

We can now give the proof of the two lemmas in appendix C of the main text.

Proof of lemma C.2.
First, we want to prove the statements (a), (b), (c), (d) and (e). We have $\|\lambda^0\,(\lambda^{0\prime}\lambda^0)^{-1}\,(f^{0\prime}f^0)^{-1}\,f^{0\prime}\| = \mathcal{O}_p((NT)^{-1/2})$, $\|e\| = \mathcal{O}_p(N^{1/2})$, and according to lemma C.1 also $\|X_k^{\text{weak}}\| = \mathcal{O}_p(N^{1/2})$. Using this and part (v) of theorem S.3.1 we find, *e.g.*

$$\frac{1}{NT}\operatorname{Tr}(M_{f^0}\,X_{k_1}^{\text{weak}\prime}\,P_{\lambda^0}\,X_{k_2}^{\text{weak}}) \leq \frac{R}{NT}\left\|X_{k_1}^{\text{weak}\prime}\right\|\left\|X_{k_2}^{\text{weak}}\right\| = \mathcal{O}_p(N^{-1}) = o_p(1) \ , \qquad \text{(S.6.15)}$$

and

$$\begin{aligned}
&\frac{1}{\sqrt{NT}}\operatorname{Tr}\left(e\,M_{f^0}\,e'\,M_{\lambda^0}\,X_k^{\text{weak}}\,f^0\,(f^{0\prime}f^0)^{-1}\,(\lambda^{0\prime}\lambda^0)^{-1}\,\lambda^{0\prime}\right) \\
&\qquad \leq \frac{R}{\sqrt{NT}}\|e\|^2\left\|X_k^{\text{weak}}\right\|\left\|f^0\,(f^{0\prime}f^0)^{-1}\,(\lambda^{0\prime}\lambda^0)^{-1}\,\lambda^{0\prime}\right\| = \mathcal{O}_p(N^{-1/2}) = o_p(1) \ , \qquad \text{(S.6.16)}
\end{aligned}$$

and analogously for the three statements (b), (d) and (e).

The proofs for part (f), (g) and (h) are similar, but in addition we use $\|X_k^{\rm str}\| = \mathcal{O}_p(\sqrt{NT})$ and $\|P_{\lambda^0} e P_{f^0}\| = \mathcal{O}_p(1)$, which was shown in lemma C.1. We then obtain *e.g.*

$$\begin{aligned}
&\frac{1}{\sqrt{NT}} \mathrm{Tr}\left(e P_{f^0}\, e'\, M_{\lambda^0}\, X_k^{\rm str}\, f^0\, (f^{0\prime} f^0)^{-1}\, (\lambda^{0\prime}\lambda^0)^{-1}\, \lambda^{0\prime}\right) \\
&\qquad = \frac{1}{\sqrt{NT}} \mathrm{Tr}\left(P_{\lambda^0} e P_{f^0}\, e'\, M_{\lambda^0}\, X_k^{\rm str}\, f^0\, (f^{0\prime} f^0)^{-1}\, (\lambda^{0\prime}\lambda^0)^{-1}\, \lambda^{0\prime}\right) \\
&\qquad \leq \frac{R}{\sqrt{NT}} \left\|P_{\lambda^0} e P_{f^0}\right\| \|e\| \|X_k^{\rm str}\| \left\|f^0\, (f^{0\prime} f^0)^{-1}\, (\lambda^{0\prime}\lambda^0)^{-1}\, \lambda^{0\prime}\right\| = \mathcal{O}_p(N^{-1/2}) = o_p(1)\ ,
\end{aligned} \tag{S.6.17}$$

and analogously for part (f) and (h).

To prove statement (i) we need to use additionally $\|P_{f^0} e' X_k^{\rm str}\| = \mathcal{O}_p(\sqrt{NT})$, which was also shown in lemma C.1. We find

$$\begin{aligned}
&\frac{1}{\sqrt{NT}} \mathrm{Tr}\left(e' M_{\lambda^0}\, X_k^{\rm str}\, M_{f^0}\, e'\, \lambda^0\, (\lambda^{0\prime}\lambda^0)^{-1}\, (f^{0\prime} f^0)^{-1}\, f^{0\prime}\right) \\
&\qquad = \frac{1}{\sqrt{NT}} \mathrm{Tr}\left(P_{f^0}\, e'\, X_k^{\rm str}\, M_{f^0}\, e'\, \lambda^0\, (\lambda^{0\prime}\lambda^0)^{-1}\, (f^{0\prime} f^0)^{-1}\, f^{0\prime}\right) \\
&\qquad\qquad - \frac{1}{\sqrt{NT}} \mathrm{Tr}\left(P_{f^0}\, e' P_{\lambda^0}\, X_k^{\rm str}\, M_{f^0}\, e'\, \lambda^0\, (\lambda^{0\prime}\lambda^0)^{-1}\, (f^{0\prime} f^0)^{-1}\, f^{0\prime}\right) \\
&\qquad \leq \frac{R}{\sqrt{NT}} \|P_{f^0}\, e'\, X_k^{\rm str}\| \|e\|\, \|\lambda^0\, (\lambda^{0\prime}\lambda^0)^{-1}\, (f^{0\prime} f^0)^{-1}\, f^{0\prime}\| \\
&\qquad\qquad - \frac{R}{\sqrt{NT}} \|P_{f^0}\, e' P_{\lambda^0}\| \|X_k^{\rm str}\| \|e\| \|\lambda^0\, (\lambda^{0\prime}\lambda^0)^{-1}\, (f^{0\prime} f^0)^{-1}\, f^{0\prime}\| \\
&\qquad = \mathcal{O}_p(T^{-1/2}) = o_p(1)\ .
\end{aligned} \tag{S.6.18}$$

Statement (j) follows if we can show that the second moment of $e' P_{\lambda^0} X_k^{\rm weak}/\sqrt{NT}$ is $o(1)$. We find

$$\begin{aligned}
\mathbb{E}\left(\frac{1}{\sqrt{NT}}\, \mathrm{Tr}\left(e' P_{\lambda^0} X_k^{\rm weak}\right)\right)^2 &= \frac{1}{NT} \mathbb{E}\left(\sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^T e_{it} X_{k,jt}^{\rm weak} \lambda_i^{0\prime} \left(\lambda^{0\prime}\lambda^0\right)^{-1} \lambda_j^0\right)^2 \\
&= \frac{1}{NT} \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^N \sum_{l=1}^N \sum_{t=1}^T \sum_{s=1}^T E\left(e_{it} X_{k,jt}^{\rm weak} e_{ks} X_{k,ls}^{\rm weak}\right) \lambda_i^{0\prime} \left(\lambda^{0\prime}\lambda^0\right)^{-1} \lambda_j^0 \lambda_k^{0\prime} \left(\lambda^{0\prime}\lambda^0\right)^{-1} \lambda_l^0.
\end{aligned} \tag{S.6.19}$$

Notice that $\mathbb{E}\left(e_{it} X_{k,jt}^{\rm weak} e_{ks} X_{k,ls}^{\rm weak}\right) = 0$ unless $(t = s)$ and $(i = k)$ and $(j = l)$. Therefore

$$\begin{aligned}
\mathbb{E}\left(\frac{1}{\sqrt{NT}}\, \mathrm{Tr}\left(e' P_{\lambda^0} X_k^{\rm weak}\right)\right)^2 &= \frac{1}{NT} \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^T \mathbb{E}\left(e_{it}^2\right) \mathbb{E}\left[\left(X_{jt}^{\rm weak}\right)^2\right] \left(\lambda_i^{0\prime} \left(\lambda^{0\prime}\lambda^0\right)^{-1} \lambda_j^0\right)^2 \\
&\leq \frac{\bar{B}\, \|P_{\lambda^0}\|_F^2}{N} = \mathcal{O}\left(N^{-1}\right) = o(1)\ ,
\end{aligned} \tag{S.6.20}$$

where we used $\sup_{it} \mathbb{E}\left(e_{it}^2\right) \leq \bar{B}$ and $\sup_{it} \mathbb{E}\left[\left(X_{it}^{\rm weak}\right)^2\right] \leq \bar{B}$. This is what we wanted to show.

Now we want to prove part (k) and (l) of the present lemma. For part (k) we define (we fix the regressor index k here)

$$B = M_{\lambda^0}\, X_k^{\rm str}\, f^0\, (f^{0\prime} f^0)^{-1}\, (\lambda^{0\prime}\lambda^0)^{-1}\, \lambda^{0\prime}\, . \tag{S.6.21}$$

Using part (i) and (ii) of theorem S.3.1 we find

$$\begin{aligned} \|B\|_F &\leq R^{1/2} \|B\| \\ &\leq R^{1/2} \|X_k^{\rm str}\| \, \left\| f^0\, (f^{0\prime} f^0)^{-1}\, (\lambda^{0\prime}\lambda^0)^{-1}\, \lambda^{0\prime} \right\| \\ &\leq R^{1/2} \|X_k^{\rm str}\|_F \, \left\| f^0\, (f^{0\prime} f^0)^{-1}\, (\lambda^{0\prime}\lambda^0)^{-1}\, \lambda^{0\prime} \right\| \ . \end{aligned} \tag{S.6.22}$$

and therefore

$$\begin{aligned} \mathbb{E}\left(\|B\|_F^2\right) &\leq R \left\| f^0\, (f^{0\prime} f^0)^{-1}\, (\lambda^{0\prime}\lambda^0)^{-1}\, \lambda^{0\prime} \right\|^2 \mathbb{E} \|X_k^{\rm str}\|_F^2 \\ &= \mathcal{O}(1)\, , \end{aligned} \tag{S.6.23}$$

where we used $\mathbb{E}\|X_k^{\rm str}\|_F^2 = \mathcal{O}(NT)$, which is true since we assumed uniformly bounded second (actually even higher) moments of $X_{k,it}^{\rm str}$. Applying lemma S.6.1 we therefore find

$$\mathbb{E}\left(\frac{1}{\sqrt{NT}} {\rm Tr}\left\{ \left[ee' - \mathbb{E}\left(ee'\right)\right] B \right\} \right)^2 \leq c_0\, \frac{T}{NT}\, \mathbb{E}\left(\|B\|_F^2\right) = o(1)\, , \tag{S.6.24}$$

and thus

$$\frac{1}{\sqrt{NT}} {\rm Tr}\left\{ \left[ee' - \mathbb{E}\left(ee'\right)\right] B \right\} = o_p(1)\, , \tag{S.6.25}$$

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

Finally, we want to show part (m) of the lemma. Write $\tilde f_{ts} = f_t' \, (f'f)^{-1}\, f_s$. The required result follows by showing

$$\begin{aligned} & \mathbb{E}\left[\frac{1}{N} \sum_{t=1}^T \sum_{s=1}^T \sum_{i=1}^N \left(e_{it} \tilde f_{ts} X_{k,is}^{\rm weak} - \mathbb{E}\left(e_{it} \tilde f_{ts} X_{k,is}^{\rm weak} \right)\right) \right]^2 \\ = \ & \frac{1}{N^2} \sum_{t=1}^T \sum_{s=1}^T \sum_{p=1}^T \sum_{q=1}^T \sum_{i=1}^N \sum_{j=1}^N \mathbb{E}\left(e_{it} e_{jp} X_{k,is}^{\rm weak} X_{k,jq}^{\rm weak} \tilde f_{ts} \tilde f_{pq} \right) - \left[\frac{1}{N} \sum_{t=1}^T \sum_{s=1}^T \sum_{i=1}^N \mathbb{E}\left(e_{it} \tilde f_{ts} X_{k,is}^{\rm weak} \right) \right]^2 \\ = \ & \frac{1}{N^2} \sum_{t=1}^T \sum_{s=1}^T \sum_{p=1}^T \sum_{q=1}^T \sum_{i=1}^N \mathbb{E}\left(e_{it} e_{ip} X_{k,is}^{\rm weak} X_{k,iq}^{\rm weak} \tilde f_{ts} \tilde f_{pq} \right) - \left[\frac{1}{N} \sum_{t=1}^T \sum_{s=1}^T \sum_{i=1}^N \mathbb{E}\left(e_{it} \tilde f_{ts} X_{k,is}^{\rm weak} \right) \right]^2 \\ \leq \ & \frac{1}{N^2} \sum_{i=1}^N \mathbb{E} \left[\begin{array}{c} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T e_{it} f_t' \left(\frac{f'f}{T}\right)^{-1/2} \right) \left(\frac{1}{\sqrt{T}} \sum_{s=1}^T X_{k,is}^{\rm weak} \left(\frac{f'f}{T}\right)^{-1/2} f_s \right) \\ \times \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T e_{it} f_t' \left(\frac{f'f}{T}\right)^{-1/2} \right) \left(\frac{1}{\sqrt{T}} \sum_{s=1}^T X_{k,is}^{\rm weak} \left(\frac{f'f}{T}\right)^{-1/2} f_s \right) \end{array} \right] \\ \leq \ & \frac{1}{N^2} \sum_{i=1}^N \mathbb{E} \left[\left\| \frac{1}{\sqrt{T}} \sum_{t=1}^T e_{it} f_t' \left(\frac{f'f}{T}\right)^{-1/2} \right\|^2 \right] \left[\left\| \frac{1}{\sqrt{T}} \sum_{t=1}^T X_{k,is}^{\rm weak} \left(\frac{f'f}{T}\right)^{-1/2} f_t \right\|^2 \right] \\ \leq \ & \frac{1}{N^2} \sum_{i=1}^N \left[\mathbb{E} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^T e_{it} f_t' \left(\frac{f'f}{T}\right)^{-1/2} \right\|^4 \right] \left[\mathbb{E} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^T X_{k,is}^{\rm weak} \left(\frac{f'f}{T}\right)^{-1/2} f_t \right\|^4 \right] \\ = \ & O\left(\frac{1}{N}\right), \end{aligned} \tag{S.6.26}$$

where the last line holds because

$$\sup_{i,T} \mathbb{E} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} e_{it} f_t' \left(\frac{f'f}{T} \right)^{-1/2} \right\|^4 \quad < \quad c_0 \ ,$$

$$\sup_{i,T} \mathbb{E} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} X_{k,is}^{\text{weak}} \left(\frac{f'f}{T} \right)^{-1/2} f_t \right\|^4 \quad < \quad c_0 \ , \tag{S.6.27}$$

for some c_0 independent of N, T. ∎

Proof of lemma C.3. Let c be a K-vector such that $\|c\| = 1$. The required result follows by the Cramer-Wold device, if we show that

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} e_{it} \mathfrak{X}_{it}' c \Rightarrow \mathcal{N}\left(0, c'\Omega c\right) \ . \tag{S.6.28}$$

Denote $Z_{iT} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} e_{it} \mathfrak{X}_{it}' c$. Notice that conditioning on X^{str}, the $\{Z_{iT}\}_{i=1,\ldots,N}$ are independent with $\mathbb{E}\left(Z_{iT} \middle| X^{\text{str}}\right) = 0$ and

$$\mathbb{E}\left(Z_{iT}^2 \middle| X^{\text{str}}\right) = c' \left(\frac{1}{T} \sum_{t=1}^{T} \Sigma_{it} \mathbb{E}\left(\mathfrak{X}_{it}\mathfrak{X}_{it}'\right) \middle| X^{\text{str}} \right) c. \tag{S.6.29}$$

Using our assumptions we thus find that

$$\frac{1}{N} \sum_{i=1}^{N} \mathbb{E}\left(Z_{iT}^2 \middle| X^{\text{str}}\right) \underset{p}{\rightarrow} c'\Omega c > 0. \tag{S.6.30}$$

Notice also that $\sup_{i,T} \mathbb{E}\left(Z_{iT}^4 \middle| X^{\text{str}}\right) = \mathcal{O}_p(1)$. Define $s_{NT}^2 = \sum_{i=1}^{N} \mathbb{E}\left(Z_{iT}^2 \middle| X^{\text{str}}\right)$. Then, the Lindeberg-Feller condition is satisfied since for any $\epsilon > 0$,

$$\begin{aligned}
\sum_{i=1}^{N} \mathbb{E}\left[\left| \frac{Z_{iT}}{s_{NT}} \right|^2 1\left\{ \left| \frac{Z_{iT}}{s_{NT}} \right| \geq \epsilon \right\} \middle| X^{\text{str}} \right] &\leq \frac{1}{\epsilon^2} \frac{1}{s_{NT}^4} \sum_{i=1}^{N} \mathbb{E}\left(|Z_{iT}|^4 \middle| X^{\text{str}} \right) \\
&\leq \frac{1}{\epsilon^2} \frac{1}{N} \frac{1}{\left(s_{NT}^2/N\right)^2} \sup_{i,T} \mathbb{E}\left(|Z_{iT}|^4 \middle| X^{\text{str}} \right) \underset{p}{\rightarrow} 0.
\end{aligned} \tag{S.6.31}$$

By the conditional version of theorem 2 of Phillips and Moon (1999), and since the conditional limiting distribution does not depend on X^{str}, we obtain the required result. ∎

S.7 Supplement to the Proof of Theorem 3.6

The following lemma gives a useful bound on the maximum of (correlated) random variables

Lemma S.7.1. *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 $\max_i \mathbb{E}|Z_i|^\gamma \leq B$, i.e. the γ'th moment of the Z_i are finite and uniformly bounded. For $n \to \infty$ we then have*

$$\max_i |Z_i| = \mathcal{O}_p\left(n^{1/\gamma}\right) \ . \tag{S.7.1}$$

Proof. Using Jensen's inequality one obtains $\mathbb{E}\max_i |Z_i| \leq (\mathbb{E}\max_i |Z_i|^\gamma)^{1/\gamma} \leq (\mathbb{E}\sum_{i=1}^n |Z_i|^\gamma)^{1/\gamma} \leq (n \max_i \mathbb{E}|Z_i|^\gamma)^{1/\gamma} \leq n^{1/\gamma} B^{1/\gamma}$. Markov's inequality then gives equation (S.7.1). ∎

Lemma S.7.2. *Let*

$$
\begin{aligned}
\bar{Z}^{(1)}_{k,t\tau} &= N^{-1/2} \sum_{i=1}^N \left[e_{it} X_{k,i\tau} - \mathbb{E}\left(e_{it} X_{k,i\tau} \right) \right] \ , \\
\bar{Z}^{(2)}_{t} &= N^{-1/2} \sum_{i=1}^N \left[e_{it}^2 - \mathbb{E}\left(e_{it}^2 \right) \right] \ , \\
\bar{Z}^{(3)}_{i} &= T^{-1/2} \sum_{t=1}^T \left[e_{it}^2 - \mathbb{E}\left(e_{it}^2 \right) \right] \ .
\end{aligned} \tag{S.7.2}
$$

Under assumption 5 we have

$$
\begin{aligned}
\mathbb{E}\left| \bar{Z}^{(1)}_{k,t\tau} \right|^4 &\leq B \ , \\
\mathbb{E}\left| \bar{Z}^{(2)}_{t\tau} \right|^4 &\leq B \ , \\
\mathbb{E}\left| \bar{Z}^{(3)}_{i} \right|^4 &\leq B \ ,
\end{aligned} \tag{S.7.3}
$$

for some $B > 0$, i.e. the 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 $Z^{(1)}_{k,t\tau,i} = e_{it} X_{k,i\tau} - \mathbb{E}\left(e_{it} X_{k,i\tau} \right)$. 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 $\mathbb{E}(Z_i) = 0$ and $\mathbb{E}(Z_i Z_j Z_k Z_l) = 0$ if $i \notin \{j, k, l\}$ (and the same holds for permutations of i, j, k, l) —- the latter follows from the fact that conditional on the strictly exogenous part of the regressors $X_k^{\rm str}$ we find Z_i to have mean zero and to be independent from $Z_j Z_k Z_l$ if $i \notin \{j, k, l\}$. Using this we compute

$$
\begin{aligned}
\mathbb{E}\left(\sum_{i=1}^N Z_i \right)^4 &= \sum_{i,j,k,l=1}^N \mathbb{E}\left(Z_i Z_j Z_k Z_l \right) \\
&= 3 \sum_{i \neq j} \mathbb{E}\left(Z_i^2 \, Z_j^2 \right) + \sum_i \mathbb{E}\left(Z_i^4 \right) \\
&= 3 \sum_{i,j=1}^N \mathbb{E}\left(Z_i^2 \right) \, \mathbb{E}\left(Z_j^2 \right) + \sum_{i=1}^N \left\{ \mathbb{E}\left(Z_i^4 \right) - 3 \left[\mathbb{E}\left(Z_i^2 \right) \right]^2 \right\} \ ,
\end{aligned} \tag{S.7.4}
$$

Since we argued that $\mathbb{E}\left(Z_i^4 \right)$ 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.7.3. *For a $T \times T$ matrix A we have*[3]

$$\left\|A^{\rm truncR}\right\| \le M \left\|A^{\rm truncR}\right\|_{\max} \equiv M \max_t \max_{t<\tau\le t+M} |A_{t\tau}| ,$$
$$\left\|A^{\rm truncL}\right\| \le M \left\|A^{\rm truncL}\right\|_{\max} \equiv M \max_t \max_{t-M\le\tau<t} |A_{t\tau}| . \tag{S.7.5}$$

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

$$\begin{aligned}\left\|A^{\rm truncR}\right\|_1 &= \max_{t=1\ldots T} \sum_{\tau=t+1}^{t+M} |A_{t\tau}| \\ &\le M \max_{t,\tau=1\ldots T} \max_{t<\tau\le t+M} |A_{t\tau}| ,\end{aligned} \tag{S.7.6}$$

and analogously we find the same bound for the ∞-norm $\left\|A^{\rm truncR}\right\|_\infty$. Applying part (vii) of theorem S.3.1 we therefore also get this bound for the operator norm $\|A^{\rm truncR}\|$. The proof for $A^{\rm truncL}$ is analogous. ∎

Proof of lemma E.3. # First, we are going to show $A_1 = o_p(1)$. Let $B_{1,it} = \hat{\mathfrak{X}}_{it} - \mathfrak{X}_{it}$, $B_{2,it} = e_{it}^2 \mathfrak{X}_{it}$, and $B_{3,it} = e_{it}^2 \hat{\mathfrak{X}}_{it}$. 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

$$A_{1,k_1k_2} = \frac{1}{NT}\operatorname{Tr}\left(B_{1,k_1}B'_{3,k_2} + B_{2,k_1}B'_{1,k_2}\right) . \tag{S.7.7}$$

We find that

$$B_{1,k} = -P_{f^0}X_k^{\rm weak}M_{\lambda^0} - X_k^{\rm weak}P_{\lambda^0} + (M_{\hat\lambda} - M_{\lambda^0})X_kM_{f^0} + M_{\hat\lambda}X_k(M_{\hat f} - M_{f^0}) , \tag{S.7.8}$$

which satisfies $\|B_{1,k}\| = \mathcal{O}_p(N^{1/2})$ since $\|M_{\hat\lambda} - M_{\lambda^0}\| = \mathcal{O}_p(N^{-1/2})$, $\|M_{\hat f} - M_{f^0}\| = \mathcal{O}_p(N^{-1/2})$, $\|X_k\| = \mathcal{O}_p(\sqrt{NT}) = \mathcal{O}_p(N)$, and $\|X_k^{\rm weak}\| = \mathcal{O}_p(N^{1/2})$. In addition we have $\operatorname{rank}(B_{1,k}) \le 6R$.

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

$$\begin{aligned}\|B_{2,k}\|^4 &\le \|B_{2,k}\|_F^4 \\ &= \left(\sum_{i=1}^N\sum_{t=1}^T e_{it}^4 \mathfrak{X}_{k,it}^2\right)^2 \\ &\le \left(\sum_{i=1}^N\sum_{t=1}^T e_{it}^8\right)\left(\sum_{i=1}^N\sum_{t=1}^T \mathfrak{X}_{k,it}^4\right) = \mathcal{O}_p(NT)\,\mathcal{O}_p(NT) ,\end{aligned} \tag{S.7.9}$$

i.e. $\|B_{2,k}\| = \mathcal{O}_p(\sqrt{NT})$, and analogously $\|B_{3,k}\| = \mathcal{O}_p(\sqrt{NT})$. Therefore

$$\begin{aligned}|A_{1,k_1k_2}| &\le \frac{6R}{NT}\left(\|B_{1,k_1}\|\|B_{3,k_2}\| + \|B_{2,k_1}\|\|B_{1,k_2}\|\right) \\ &= \frac{6R}{NT}\left(\mathcal{O}_p(N^{1/2})\mathcal{O}_p(\sqrt{NT}) + \mathcal{O}_p(\sqrt{NT})\mathcal{O}_p(N^{1/2})\right) = o_p(1) .\end{aligned} \tag{S.7.10}$$

[3] 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 $1 \le \tau \le T$. Since this would complicate notation, we prefer the convention that $A_{t\tau} = 0$ for $t < 1$ or $\tau < 1$ of $t > T$ or $\tau > T$.

This is what we wanted to show.

Next, we want to show $A_2 = o_p(1)$. According to theorem D.1 we have $e - \hat{e} = C_1 + C_2$, where we defined $C_1 = -\sum_{k=1}^K \left(\hat{\beta}_k - \beta_k^0\right) X_k$, and $C_2 = \sum_{k=1}^K \left(\hat{\beta}_k - \beta_k^0\right) \left(P_{\lambda^0}\, X_k\, M_{f^0} + X_k\, P_{f^0}\right) + P_{\lambda^0}\, e\, M_{f^0} + e\, P_{f^0} - \hat{e}_e^{(1)} - \hat{e}^{(\mathrm{rem})}$, 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 D.1). Using this notation we have

$$A_2 = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (e_{it} + \hat{e}_{it})(C_{1,it} + C_{2,it}) \hat{\mathfrak{X}}_{it} \hat{\mathfrak{X}}'_{it} \ , \tag{S.7.11}$$

which can also be written as

$$A_{2,k_1k_2} = -\sum_{k_3=1}^K \left(\hat{\beta}_{k_3} - \beta_{k_3}^0\right) (C_{5,k_1k_2k_3} + C_{6,k_1k_2k_3}) + \frac{1}{NT} \mathrm{Tr}\left(C_2\, C_{3,k_1k_2}\right) + \frac{1}{NT} \mathrm{Tr}\left(C_2\, C_{4,k_1k_2}\right) \ , \tag{S.7.12}$$

where we defined

$$\begin{aligned}
C_{3,k_1k_2,it} &= e_{it} \hat{\mathfrak{X}}_{k_1,it} \hat{\mathfrak{X}}_{k_2,it} \ , \\
C_{4,k_1k_2,it} &= \hat{e}_{it} \hat{\mathfrak{X}}_{k_1,it} \hat{\mathfrak{X}}_{k_2,it} \ , \\
C_{5,k_1k_2k_3} &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \ e_{it} \hat{\mathfrak{X}}_{k_1,it} \hat{\mathfrak{X}}_{k_2,it} X_{k_3,it} \ , \\
C_{6,k_1k_2k_3} &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \ \hat{e}_{it} \hat{\mathfrak{X}}_{k_1,it} \hat{\mathfrak{X}}_{k_2,it} X_{k_3,it} \ .
\end{aligned} \tag{S.7.13}$$

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

$$\begin{aligned}
\|C_{3,k_1k_2}\|^4 &\leq \|C_{3,k_1k_2}\|_F^4 \\
&= \left(\sum_{i=1}^N \sum_{t=1}^T e_{it}^2 \hat{\mathfrak{X}}_{k_1,it}^2 \hat{\mathfrak{X}}_{k_2,it}^2 \right)^2 \\
&\leq \left(\sum_{i=1}^N \sum_{t=1}^T e_{it}^4 \right) \left(\sum_{i=1}^N \sum_{t=1}^T \hat{\mathfrak{X}}_{k_1,it}^4 \hat{\mathfrak{X}}_{k_2,it}^4 \right) \\
&= \mathcal{O}_p(N^2 T^2) \ ,
\end{aligned} \tag{S.7.14}$$

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

$$\begin{aligned}
\|C_{4,k_1k_2}\|^2 &\leq \|C_{3,k_1k_2}\|_F^2 \\
&= \sum_{i=1}^N \sum_{t=1}^T \hat{e}_{it}^2 \hat{\mathfrak{X}}_{k_1,it}^2 \hat{\mathfrak{X}}_{k_2,it}^2 \\
&\leq \left(\sum_{i=1}^N \sum_{t=1}^T \hat{e}_{it}^2 \right) \max_{i=1\ldots N} \max_{t=1\ldots T} \left(\hat{\mathfrak{X}}_{k_1,it}^2 \hat{\mathfrak{X}}_{k_2,it}^2 \right) \\
&\leq \left(\sum_{i=1}^N \sum_{t=1}^T e_{it}^2 \right) \max_{i=1\ldots N} \max_{t=1\ldots T} \left(\hat{\mathfrak{X}}_{k_1,it}^2 \hat{\mathfrak{X}}_{k_2,it}^2 \right) \\
&= \mathcal{O}_p(NT) \mathcal{O}_p((NT)^{(4/(8+\epsilon))}) = o_p((NT)^{(3/4)}) \ .
\end{aligned} \tag{S.7.15}$$

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

For C_5 we find

$$\begin{aligned} C_{5,k_1k_2k_3}^2 &\leq \left(\frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T e_{it}^2\right)\left(\frac{1}{NT}\hat{\mathfrak{X}}_{k_1,it}^2 \hat{\mathfrak{X}}_{k_2,it}^2 X_{k_3,it}^2\right) \\ &= \mathcal{O}_p(1) \ , \end{aligned} \tag{S.7.16}$$

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 $\sum_{i=1}^N \sum_{t=1}^T \hat{e}_{it}^2 \leq \sum_{i=1}^N \sum_{t=1}^T e_{it}^2$.

Using these results we obtain

$$\begin{aligned} |A_{2,k_1k_2}| &\leq -\sum_{k_3=1}^K \left\| \hat{\beta}_{k_3} - \beta_{k_3}^0 \right\| |C_{5,k_1k_2k_3} + C_{6,k_1k_2k_3}| + \frac{11R}{NT}\|C_2\|\|C_{3,k_1k_2}\| + \frac{11R}{NT}\|C_2\|\|C_{4,k_1k_2}\| \\ &= \mathcal{O}_p((NT)^{-1/2})\mathcal{O}_p(1) + \frac{11R}{NT}\mathcal{O}_p(N^{1/2})\mathcal{O}_p(\sqrt{NT}) + \frac{11R}{NT}\mathcal{O}_p(N^{1/2})o_p((NT)^{3/4}) = o_p(1) \ . \end{aligned} \tag{S.7.17}$$

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 E.4. By lemma E.2 we know that asymptotically $P_{\hat{f}}$ is close to P_{f^0} and therefore $\text{rank}(P_{\hat{f}} P_{f^0}) = \text{rank}(P_{f^0} P_{f^0}) = R$, *i.e.* $\text{rank}(P_{\hat{f}} f^0) = R$ asymptotically. We can therefore write $\hat{f} = P_{\hat{f}} f^0 H$, 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 $\hat{f}$ and f^0 we have $H = (\hat{f}' P_{\hat{f}} f^0/T)^{-1} = (\hat{f}' f^0/T)^{-1}$, and therefore $\|H^{-1}\| \leq \|\hat{f}\|\|f^0\|/T = \mathcal{O}_p(1)$. We also have $\hat{f} = f^0 H + (P_{\hat{f}} - P_{f^0}) f^0 H$, and thus $H = f^{0\prime}\hat{f}/T - f^{0\prime}(P_{\hat{f}} - P_{f^0}) f^0 H/T$, *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 $\hat{f}$ in terms of $P_{\hat{f}}$ as above is that the result $\left\| P_{\hat{f}} - P_{f^0} \right\| = \mathcal{O}_p\left(T^{-1/2}\right)$ of lemma E.2 immediately implies

$$\left\| \hat{f} - f^0\, H \right\| = \mathcal{O}_p\left(1\right) \ . \tag{S.7.18}$$

The FOC wrt λ in the minimization of the first line in equation (2.5) reads

$$\hat{\lambda}\, \hat{f}' \hat{f} = \left(Y - \sum_{k=1}^K \hat{\beta}_k X_k\right) \hat{f} \ , \tag{S.7.19}$$

which yields

$$
\begin{aligned}
\hat{\lambda} &= \left[\lambda^0 f^{0\prime} - \sum_{k=1}^{K} \left(\hat{\beta}_k - \beta_k^0\right) X_k\right] \hat{f} \left(\hat{f}'\hat{f}\right)^{-1} \\
&= \left[\lambda^0 f^{0\prime} + \sum_{k=1}^{K} \left(\beta_k^0 - \hat{\beta}_k\right) X_k + e\right] P_{\hat{f}} f^0 \left(f^{0\prime} P_{\hat{f}} f^0\right)^{-1} \left(H'\right)^{-1} \\
&= \lambda^0 \left(H'\right)^{-1} + \lambda^0 f^{0\prime} \left(P_{\hat{f}} - P_{f^0}\right) f^0 \left(f^{0\prime} P_{\hat{f}} f^0\right)^{-1} \left(H'\right)^{-1} \\
&\qquad + \lambda^0 f^{0\prime} f^0 \left[\left(f^{0\prime} P_{\hat{f}} f^0\right)^{-1} - \left(f^{0\prime} f^0\right)^{-1}\right] \left(H'\right)^{-1} \\
&\qquad + \left[\sum_{k=1}^{K} \left(\beta_k^0 - \hat{\beta}_k\right) X_k + e\right] P_{\hat{f}} f^0 \left(f^{0\prime} P_{\hat{f}} f^0\right)^{-1} \left(H'\right)^{-1} \ .
\end{aligned} \tag{S.7.20}
$$

We have $\left(f^{0\prime} P_{\hat{f}} f^0 \big/ T\right)^{-1} - \left(f^{0\prime} f^0 / T\right)^{-1} = \mathcal{O}_p(T^{-1/2})$, because $\left\|P_{\hat{f}} - P_{f^0}\right\| = \mathcal{O}_p\left(T^{-1/2}\right)$ 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 E.1 also $\|\hat{\beta} - \beta^0\| = \mathcal{O}_p(1/\sqrt{NT})$. Therefore

$$
\left\|\hat{\lambda} - \lambda^0 \left(H'\right)^{-1}\right\| = \mathcal{O}_p\left(1\right) \ , \tag{S.7.21}
$$

which is what we wanted to prove.

Next, we want to show

$$
\begin{aligned}
\left\|\left(\frac{\hat{\lambda}'\hat{\lambda}}{N}\right)^{-1} - \left(\frac{(H)^{-1} \lambda^{0\prime} \lambda^0 \left(H'\right)^{-1}}{N}\right)^{-1}\right\| &= \mathcal{O}_p\left(N^{-1/2}\right) \ , \\
\left\|\left(\frac{\hat{f}'\hat{f}}{T}\right)^{-1} - \left(\frac{H' f^{0\prime} f^0 H}{T}\right)^{-1}\right\| &= \mathcal{O}_p\left(T^{-1/2}\right) \ .
\end{aligned} \tag{S.7.22}
$$

Let $A = N^{-1} \hat{\lambda}' \hat{\lambda}$ and $B = N^{-1} (H)^{-1} \lambda^{0\prime} \lambda^0 (H')^{-1}$. Using (S.7.21) we find

$$
\begin{aligned}
\|A - B\| &= \frac{1}{2N} \left\|\left[\hat{\lambda}' + (H)^{-1} \lambda^{0\prime}\right] \left[\hat{\lambda} - \lambda^0 \left(H'\right)^{-1}\right] + \left[\hat{\lambda}' - (H)^{-1} \lambda^{0\prime}\right] \left[\hat{\lambda} + \lambda^0 \left(H'\right)^{-1}\right]\right\| \\
&= N^{-1} \mathcal{O}_p(N^{1/2}) \mathcal{O}_p(1) = \mathcal{O}_p\left(N^{-1/2}\right) \ .
\end{aligned} \tag{S.7.23}
$$

By assumption 1 we know that

$$
\left\|\left(\frac{\lambda^{0\prime}\lambda^0}{N}\right)^{-1}\right\| = \mathcal{O}_p(1) \ . \tag{S.7.24}
$$

and thus also $\left\|B^{-1}\right\| = \mathcal{O}_p(1)$, and therefore $\left\|A^{-1}\right\| = \mathcal{O}_p(1)$ (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

$$
\begin{aligned}
\left\|A^{-1} - B^{-1}\right\| &\leq \left\|A^{-1}\right\| \left\|B^{-1}\right\| \|A - B\| \\
&= \mathcal{O}_p\left(N^{-1/2}\right) \ .
\end{aligned} \tag{S.7.25}
$$

Thus, we have shown the first statement of (S.7.22), and analogously one can show the second one. Combining (S.7.21), (S.7.19) and (S.7.22) we obtain

$$\begin{aligned}
&\left\| \frac{\hat\lambda}{\sqrt{N}} \left(\frac{\hat\lambda'\hat\lambda}{N}\right)^{-1} \left(\frac{\hat f'\hat f}{T}\right)^{-1} \frac{\hat f'}{\sqrt{T}} - \frac{\lambda^0}{\sqrt{N}} \left(\frac{\lambda^{0\prime}\lambda^0}{N}\right)^{-1} \left(\frac{f^{0\prime}f^0}{T}\right)^{-1} \frac{f^{0\prime}}{\sqrt{T}} \right\| \\
&= \left\| \frac{\hat\lambda}{\sqrt{N}} \left(\frac{\hat\lambda'\hat\lambda}{N}\right)^{-1} \left(\frac{\hat f'\hat f}{T}\right)^{-1} \frac{\hat f'}{\sqrt{T}} - \frac{\lambda^0\,(H')^{-1}}{\sqrt{N}} \left(\frac{(H)^{-1}\,\lambda^{0\prime}\lambda^0\,(H')^{-1}}{N}\right)^{-1} \left(\frac{H' f^{0\prime} f^0 H}{T}\right)^{-1} \frac{H' f^{0\prime}}{\sqrt{T}} \right\| \\
&\qquad = \mathcal{O}_p\left(N^{-1/2}\right) , \qquad\qquad \text{(S.7.26)}
\end{aligned}$$

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

Proof of lemma E.5. # Part A of the proof: We start by showing

$$N^{-1} \left\| \mathbb{E}\left[e' M_{\lambda^0} X_k - \left(e' M_{\lambda^0} X_k\right)^{\text{truncR}} \right] \right\| = o_p(1) . \qquad \text{(S.7.27)}$$

This statement is trivial for the strictly exogenous part of the regressor, since then the expectation is just 0, *i.e.* what we actually have to show is

$$N^{-1} \left\| \mathbb{E}\left[e' M_{\lambda^0} X_k^{\text{weak}} - \left(e' M_{\lambda^0} X_k^{\text{weak}}\right)^{\text{truncR}} \right] \right\| = o_p(1) . \qquad \text{(S.7.28)}$$

Let $A = e' M_{\lambda^0} X_k$ 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 $\mathbb{E}(A_{t\tau}) = 0$ for $t \geq \tau$. For $t < \tau$ we define the $N \times N$ matrices $C_{k,t\tau} \equiv \mathbb{E}(e_t' X_{k,\tau})$, which by assumption are diagonal and satisfy $|C_{k,t\tau,ii}| < V_e\,\alpha^{\tau-t}$ for all i, j, where V_e is the uniform bound of $\mathbb{E}e_{it}^2$. We then have $\text{Tr}(C_{k,t\tau}) < N\,V_e\,\alpha^{\tau-t}$ and $\|C_{k,t\tau}\| = \|C_{k,t\tau}\|_{\max} < V_e\,\alpha^{\tau-t}$. Therefore $|\mathbb{E}(A_{t\tau})| = \left|\mathbb{E}\sum_{i,j=1}^N e_{it} M_{\lambda^0,ij} X_{k,j\tau}\right| = |\text{Tr}(C_{k,t\tau} M_{\lambda^0})| \leq |\text{Tr}(C_{k,t\tau})| + |\text{Tr}(C_{k,t\tau} P_{\lambda^0})| \leq |\text{Tr}(C_{k,t\tau})| + R\|C_{k,t\tau}\|$, *i.e.* we obtain $|\mathbb{E}(A_{t\tau})| \leq (N+R)\,V_e\,\alpha^{\tau-t}$ for $t < \tau$. Using this we have $\mathbb{E}(B_{t\tau}) = 0$ for $\tau \leq t + M$, and $|\mathbb{E}B_{t\tau}| < (N+R)\,\alpha^{\tau-t}$ for $\tau > t + M$. Therefore

$$\begin{aligned}
\|\mathbb{E}(B)\|_1 &= \max_{t=1\ldots T} \sum_{\tau=1}^T |\mathbb{E}(B_{t\tau})| \\
&< (N+R) \max_{t=1\ldots T} \sum_{\tau=t+M+1}^T \alpha^{\tau-t} < \frac{(N+R)\,\alpha^M}{1-\alpha} , \qquad \text{(S.7.29)}
\end{aligned}$$

and analogously we can show $\|\mathbb{E}(B)\|_\infty < (N+R)\,\alpha^M/(1-\alpha)$. Using part (vii) of theorem S.3.1 we therefore find $\|\mathbb{E}(B)\| < (N+R)\,\alpha^M/(1-\alpha) = o_p(N)$, which is equivalent to equation (S.7.27) that we wanted to show in this part of the proof. Here we used $M \to \infty$. The proof of

$$\begin{aligned}
N^{-1} \left\| \mathbb{E}\left[e' M_{\lambda^0} e) - \left(e' M_{\lambda^0} e\right)^{\text{truncD}} \right] \right\| &= o_p(1) , \\
T^{-1} \left\| \mathbb{E}\left[e M_{f^0} e') - \left(e M_{f^0} e'\right)^{\text{truncD}} \right] \right\| &= o_p(1) , \qquad \text{(S.7.30)}
\end{aligned}$$

is analogous.

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

$$N^{-1} \left\| \left[e' M_{\lambda^0} X_k \, - \, \mathbb{E}\left(e' M_{\lambda^0} X_k \right) \right]^{\text{truncR}} \right\| = o_p(1) \, . \tag{S.7.31}$$

Using lemma S.7.3 we have

$$\begin{aligned}
N^{-1} \left\| \left[e' M_{\lambda^0} X_k \, - \, \mathbb{E}\left(e' M_{\lambda^0} X_k \right) \right]^{\text{truncR}} \right\| &\le M \, \max_t \, \max_{t < \tau \le t+M} \, N^{-1} \left| e_t' M_{\lambda^0} X_{k,\tau} - \mathbb{E}\left(e_t' M_{\lambda^0} X_{k,\tau} \right) \right| \\
&\le M \, \max_t \, \max_{t < \tau \le t+M} \, N^{-1} \left| \sum_{i=1}^N \left[e_{it} X_{k,i\tau} - \mathbb{E}\left(e_{it} X_{k,i\tau} \right) \right] \right| \\
&\quad + M \, \max_t \, \max_{t < \tau \le t+M} \, N^{-1} \left| e_t' P_{\lambda^0} X_{k,\tau} - \mathbb{E}\left(e_t' P_{\lambda^0} X_{k,\tau} \right) \right| \\
&\le M \, N^{-1/2} \, \max_t \, \max_{t < \tau \le t+M} \left| \bar{Z}^{(1)}_{k,t\tau} \right| \\
&\quad + M \, \max_t \, \max_{t < \tau \le t+M} \, N^{-1} \left| e_t' P_{\lambda^0} X_{k,\tau} \right| \\
&\quad + M \, \max_t \, \max_{t < \tau \le t+M} \, N^{-1} \left| \mathbb{E}\left(e_t' P_{\lambda^0} X_{k,\tau} \right) \right| \, .
\end{aligned} \tag{S.7.32}$$

According to lemma S.7.2 we know that $\mathbb{E} \left| \bar{Z}^{(1)}_{k,t\tau} \right|^4$ is bounded uniformly across t and τ. Applying lemma S.7.1 we therefore find $\max_t \, \max_{t<\tau\le t+M} \bar{Z}^{(1)}_{t\tau} = \mathcal{O}_p((MT)^{1/4})$. Thus we have

$$M \, N^{-1/2} \, \max_t \, \max_{t < \tau \le t+M} \left| \bar{Z}^{(1)}_{t\tau} \right| = \mathcal{O}_p\left(M \, N^{-1/2} \, (MT)^{1/4} \right) \, = \, o_p(1) \, . \tag{S.7.33}$$

Here we used $M^5/T \to 0$.

What is left to show is that the two terms involving $e_t' P_{\lambda^0} X_{k,\tau}$ are also $o_p(1)$. We have

$$\begin{aligned}
\max_t \, \max_{t < \tau \le t+M} \, M \, N^{-1} \left| e_t' P_{\lambda^0} X_{k,\tau} \right| &\le \frac{M}{\sqrt{N}} \left| \left(\max_t \frac{\sum_{i=1}^N e_{it} \lambda_i^0}{\sqrt{N}} \right) \left(\frac{\lambda^{0\prime} \lambda^0}{N} \right)^{-1} \left(\max_\tau \frac{\sum_{i=1}^N \, X_{k,i\tau} \lambda_i^{0\prime}}{N} \right) \right| \\
&\le \frac{M}{\sqrt{N}} \, \mathcal{O}_p(T^{1/8}) \, \mathcal{O}_p(1) \, \mathcal{O}_p(T^{1/8}) = o_p(1) \, .
\end{aligned} \tag{S.7.34}$$

Here we applied lemma S.7.1 and used the fact that both $N^{-1/2} \sum_{i=1}^N e_{it} \lambda_i^0$ and $N^{-1} \sum_{i=1}^N \, X_{k,it} \lambda_i^0$ have bounded 8'th moments uniformly across t. We also used $M^4/T \to 0$. Using the above decomposition of $e_t' P_{\lambda^0} X_{k,\tau}$ we find that $\mathbb{E}\left(N^{-1/2} e_t' P_{\lambda^0} X_{k,\tau} \right)$ is uniformly bounded across t, τ, and therefore

$$\max_t \, \max_{t < \tau \le t+M} \, M \, N^{-1} \left| \mathbb{E}\left(e_t' P_{\lambda^0} X_{k,\tau} \right) \right| \le M \, N^{-1/2} \, \mathcal{O}_p(1) = o_p(1) \, . \tag{S.7.35}$$

The proof of

$$\begin{aligned}
N^{-1} \left\| \left[e' M_{\lambda^0} e \, - \, \mathbb{E}\left(e' M_{\lambda^0} e \right) \right]^{\text{truncD}} \right\| &= o_p(1) \, , \\
T^{-1} \left\| \left[e M_{f^0} e' \, - \, \mathbb{E}\left(e M_{f^0} e' \right) \right]^{\text{truncD}} \right\| &= o_p(1) \, ,
\end{aligned} \tag{S.7.36}$$

is analogous.

Part C of the proof: Finally, we have to show

$$N^{-1} \left\| \left[e' M_{\lambda^0} X_k \, - \, \hat{e}' \, X_k \right]^{\text{truncR}} \right\| = o_p(1) \; . \tag{S.7.37}$$

According to theorem D.1 we have $\hat{e} = M_{\lambda^0} e M_{f^0} + e_{\text{rem}}$, where $e_{\text{rem}} = \hat{e}^{(1)}_e - \sum_{k=1}^K \left(\hat{\beta}_k - \beta_k^0 \right) \hat{e}^{(1)}_k + \hat{e}^{(\text{rem})}$, and using corollary E.1 we find that the remainder term satisfies $\|e_{\text{rem}}\| = \mathcal{O}_p(1)$. We have to show

$$\begin{aligned} N^{-1} \left\| \left[P_{\lambda_0} e' \, X_k \right]^{\text{truncR}} \right\| &= o_p(1) \; , \\ N^{-1} \left\| \left[\hat{e}'_{\text{rem}} \, X_k \right]^{\text{truncR}} \right\| &= o_p(1) \; . \end{aligned} \tag{S.7.38}$$

Using lemma S.7.3 we find

$$\begin{aligned} N^{-1} \left\| \left[\hat{e}'_{\text{rem}} \, X_k \right]^{\text{truncR}} \right\| &= \frac{M}{N} \max_{t,\tau} \hat{e}'_{\text{rem},t} \, X_{k,\tau} \\ &\leq \frac{M}{N} \max_{t,\tau} \|\hat{e}_{\text{rem},t}\| \, \|X_{k,\tau}\| \\ &\leq \frac{M}{N} \|\hat{e}_{\text{rem}}\| \max_{\tau} \|X_{k,\tau}\| \\ &\leq \frac{M}{N} \mathcal{O}_p(1) \mathcal{O}_p(N^{1/2} T^{1/8}) = o_p(1) \; , \end{aligned} \tag{S.7.39}$$

where we used the fact that the norm of each column $\hat{e}_{\text{rem},t}$ is smaller than the operator norm of the whole matrix $\hat{e}_{\text{rem}}$. In addition we used lemma S.7.1 and the fact that $N^{-1/2} \|X_{k,\tau}\| = \sqrt{N^{-1} \sum_{i=1}^N X^2_{k,i\tau}}$ has finite 8'th moment in order to show $\max_\tau \|X_{k,\tau}\| = \mathcal{O}_p(N^{1/2} T^{1/8})$.

Using again lemma S.7.3 we find

$$\begin{aligned} N^{-1} \left\| \left[P_{f^0} e' \, X_k \right]^{\text{truncR}} \right\| &\leq N^{-1} \, M \max_{t,\tau = 1 \ldots T} \left| f^0_t \, (f^{0\prime} \, f^0)^{-1} \, f^{0\prime} \, e' \, X_{k,\tau} \right| \\ &\leq N^{-1} \, M \, \|e\| \, \|f^0\| \, \left\| (f^{0\prime} \, f^0)^{-1} \right\| \max_t \|f^0_t\| \max_\tau \|X_{k,\tau}\| \\ &= N^{-1} \, M \, \mathcal{O}_p(N^{1/2}) \, \mathcal{O}_p(T^{1/2}) \, \mathcal{O}_p(T^{-1}) \, \mathcal{O}_p(N^{1/2} T^{1/8}) = o_p(1) \; . \end{aligned} \tag{S.7.40}$$

Thus, we proved equation (S.7.37).

The proof of

$$\begin{aligned} N^{-1} \left\| \left[e' M_{\lambda^0} e \, - \, \hat{e}' \, \hat{e} \right]^{\text{truncD}} \right\| &= o_p(1) \; , \\ T^{-1} \left\| \left[e M_{f^0} e' \, - \, \hat{e} \, \hat{e}' \right]^{\text{truncD}} \right\| &= o_p(1) \; , \end{aligned} \tag{S.7.41}$$

is analogous.

The combination of part A, B and C proves the theorem. ∎

Proof of lemma E.6. Using theorem D.1 and E.1 we find $\|\hat{e}\| = \mathcal{O}_p(N^{1/2})$. Applying lemma

S.7.3 we therefore find

$$
\begin{aligned}
N^{-1} \left\| \left(\hat{e}' X_k\right)^{\mathrm{truncR}} \right\| &\leq \frac{M}{N} \max_{t,\tau} \hat{e}_t' X_{k,\tau} \\
&\leq \frac{M}{N} \max_{t,\tau} \|\hat{e}_t\| \, \|X_{k,\tau}\| \\
&\leq \frac{M}{N} \|\hat{e}\| \max_{\tau} \|X_{k,\tau}\| \\
&\leq \frac{M}{N} \mathcal{O}_p(N^{1/2}) \mathcal{O}_p(N^{1/2} T^{1/8}) = \mathcal{O}_p(M T^{1/8}) \, , \qquad \text{(S.7.42)}
\end{aligned}
$$

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. ∎

References

Andrews, D. W. K. (1999). Estimation when a parameter is on a boundary. *Econometrica*, 67(6):1341–1384.

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

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

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.

Phillips, P. C. B. and Moon, H. R. (1999). Linear regression limit theory for nonstationary panel data. *Econometrica*, 67(5):1057–1111.