

Likelihood Expansion for Panel Regression Models with Factors

Supplementary Material (Not for Publication)

Hyungsik Roger Moon* Martin Weidner*

May 2009

Abstract

This is a supplementary material that contains all the omitted proofs and derivations of the results in Moon and Weidner (2009). This note is not for publication.

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

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

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 $\psi_j = (\psi_{1j}, \dots, \psi_{Nj})$ 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} , $i = 1 \dots N$, $t = 1 - j, \dots, T - j$. We can then write equation (A.1) in matrix notation as

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

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

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

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

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

which implies $(\|r_{NT}\|_F)^2 = \mathcal{O}_p(N)$, 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} , $i = 1 \dots N$, $t = 1 - T, \dots, T$. For $j = 0 \dots T$ 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 $\|\text{diag}(\psi_j)\| \leq \max_i(|\psi_{ij}|)$.

Combining these results we find

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

This is what we wanted to show. ■

Proof of example (iii). Since σ and Σ are positive definite, there exists 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

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

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 for the regressors not only to satisfy the standard no-collinearity condition in assumption 4(a), but also the additional conditions on high- and low-rank regressors in

assumption 4(b). 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 $\text{plim}_{N,T \rightarrow \infty} W_{NT} > 0$ on the regressors (instead of assumption 4), where W_{NT} is the $K \times K$ matrix defined by $W_{NT, k_1 k_2} = (NT)^{-1} \text{Tr}(M_{f^0} X'_{k_1} M_{\lambda^0} X_{k_2})$. This matrix appears as the approximate Hessian in the likelihood expansion in theorem 3.1, *i.e.* the condition $\text{plim}_{N,T \rightarrow \infty} W_{NT} > 0$ 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 $w = (w_1, \dots, w_N)'$, and the $T \times 1$ vector $v = (v_1, \dots, v_T)'$ are such that the $N \times 2$ matrix $\Lambda = (\lambda^0, w)$ and the $T \times 2$ matrix $F = (f^0, v)$ satisfy $\text{plim}_{N,T \rightarrow \infty} (\Lambda' \Lambda / N) > 0$, $\text{plim}_{N,T \rightarrow \infty} (F' F / T) > 0$. In this case, we have $W_{NT} = (NT)^{-1} \text{Tr}(M_{f^0} v v' M_{\lambda^0} w w')$, and therefore $\text{plim}_{N,T \rightarrow \infty} W_{NT} = \text{plim}_{N,T \rightarrow \infty} (NT)^{-1} \text{Tr}(M_{f^0} v v' M_{\lambda^0} w w') > 0$. However, β is not identified because $\beta^0 X + \lambda^0 f^{0'} = (\beta^0 + 1)X - w v'$, *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 $\text{plim}_{N,T \rightarrow \infty} (\Lambda' \Lambda / N) > 0$. Let the $T \times 1$ vectors f^0, v_1 and v_2 be such that $F = (f^0, v_1, v_2)$ satisfies $\text{plim}_{N,T \rightarrow \infty} (F' F / T) > 0$. 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 $\text{plim}_{N,T \rightarrow \infty} W_{NT} > 0$. However, again β_k is not identified, because $\sum_{k=1}^4 \beta_k^0 X_k + \lambda^0 f^{0'} = \sum_{k=1}^4 (\beta_k^0 + 1) X_k - (\lambda^0 + w_1 + w_2)(f^{0'} + v_1 + v_2)'$, *i.e.* we can't distinguish between the true parameters and $(\beta, \lambda, f) = (\beta^0 + 1, -\lambda^0 - w_1 - w_2, f^{0'} + v_1 + v_2)$. 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 $\text{rank}(X) = 1$, 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 $\text{plim}_{N,T \rightarrow \infty} W_{NT} > 0$.

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

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

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:*

- (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) $|\operatorname{Tr}(AB)| \leq \|A\|_F \|B\|_F$, for $n = p$,
- (v) $|\operatorname{Tr}(C)| \leq \|C\| \operatorname{rank}(C)$,
- (vi) $\|C\| \leq \operatorname{Tr}(C)$, for C symmetric and $C \geq 0$,
- (vii) $\|A\|^2 \leq \|A\|_1 \|A\|_\infty$,
- (viii) $\|A\|_{\max} \leq \|A\| \leq \sqrt{nm} \|A\|_{\max}$,
- (ix) $\|A'CA\| \leq \|A'DA\|$, for C symmetric and $C \leq D$.

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

- (x) $\mu_i(C) + \mu_n(D) \leq \mu_i(C + D) \leq \mu_i(C) + \mu_1(D)$,
- (xi) $\mu_i(C) \leq \mu_i(C + D)$, for $D \geq 0$,
- (xii) $\mu_i(C) - \|D\| \leq \mu_i(C + D) \leq \mu_i(C) + \|D\|$.

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

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

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

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

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 $e_i, i = 2, \dots, n$, such that $e_i, i = 1, \dots, n$, becomes a orthonormal basis of the vector space of $n \times 1$ vectors. Since C is positive semi definite we then have $\operatorname{Tr}(C) = \sum_i e_i' C e_i \geq e_1' C e_1 = \|C\|$, 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 $\|A'CA\| = e' A' C A e$. Since $A'CA$ is symmetric such an e has to exist. Since $C \leq D$ we then have $\|C\| = (e' A') C (A e) \leq (e' A') D (A e) \leq \|A' D A\|$. This is what we wanted to show. For inequality (ix) let e_1 be a vector that satisfied $\|e_1\| = 1$ and $\|A'CA\| = e_1' A' C A e_1$. Then we have $\|A'CA\| = e_1' A' D A e_1 - e_1' A' (D - C) A e_1 \leq e_1' A' D A e_1 \leq \|A' D A\|$. 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 $i = 1, \dots, n$. ■

Definition S.3.2. 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 $\text{span}(A) = \{Aa \mid a \in \mathbb{R}^{r_1}\}$ and $\text{span}(B) = \{Bb \mid b \in \mathbb{R}^{r_2}\}$ of \mathbb{R}^n is defined by

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

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

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

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

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

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

Proof. Proof of Theorem B.1 Let

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

The theorem claims

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

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

$$\lambda' Z = \lambda' \lambda f' . \quad (\text{S.3.9})$$

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

$$\begin{aligned} \text{Tr} [(Z - \lambda f')' (Z - \lambda f')] &= \text{Tr} (Z' Z - Z' \lambda f') \\ &= \text{Tr} (Z' Z - Z' \lambda (\lambda' \lambda)^{-1} (\lambda' \lambda) f') \\ &= \text{Tr} (Z' Z - Z' \lambda (\lambda' \lambda)^{-1} (\lambda' \lambda) \lambda' Z) \\ &= \text{Tr} (Z' M_\lambda Z) . \end{aligned} \quad (\text{S.3.10})$$

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 eigenvalues¹ 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

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

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

$$\text{Tr} [Z' M_{\hat{\lambda}} Z] \leq \text{Tr} [Z' M_{\lambda} Z] . \quad (\text{S.3.12})$$

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 $\hat{\lambda}_r$ ($r = 1 \dots R$) 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 $\tilde{\lambda}_r$ ($r = 1 \dots R_1$) 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 ($r = 1 \dots R_2$) 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

$$\tilde{Z} = P_{\tilde{\lambda}} Z M_{\tilde{f}} + Z P_{\tilde{f}} . \quad (\text{S.3.13})$$

Since $\text{rank}(Z) \leq \text{rank}(P_{\tilde{\lambda}} Z M_{\tilde{f}}) + \text{rank}(Z P_{\tilde{f}}) = R_1 + R_2 = R$, 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

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

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*

$$\begin{aligned} w' M_{\lambda^0} w - B_0 w' v &\geq 0 , & \text{wpa1,} \\ v' M_{f^0} v - B_0 v' v &\geq 0 , & \text{wpa1.} \end{aligned} \quad (\text{S.4.1})$$

¹If 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.

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

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

$$\frac{\|M_v \lambda^0 a\|^2}{\|\lambda^0 a\|^2} = \frac{a' \lambda^{0'} M_v \lambda^0 a}{a' \lambda^{0'} \lambda^0 a} > \frac{B}{B_1}, \quad \text{wpa1.} \quad (\text{S.4.2})$$

Applying theorem S.3.3 we find

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

Therefore we find for every $\text{rank}(w) \times 1$ vector b that $b' (\tilde{w}' M_{\lambda^0} \tilde{w} - (B/B_1) \tilde{w}' \tilde{w}) b > 0$, wpa1. Thus $\tilde{w}' M_{\lambda^0} \tilde{w} - (B/B_1) \tilde{w}' \tilde{w} > 0$, 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 to show. Analogously we can show the statement for v . ■

As a consequence of 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 $X_{\text{low},\alpha} = \sum_{l=1}^{K_1} \alpha_l X_l$ be a linear combination of the low-rank regressors. Then there exists some constant $B > 0$ such that*

$$\begin{aligned} \min_{\{\alpha \in \mathbb{R}^{K_1}, \|\alpha\|=1\}} \frac{\|X_{\text{low},\alpha} M_{f^0} X'_{\text{low},\alpha}\|}{NT} &> B, \quad \text{wpa1,} \\ \min_{\{\alpha \in \mathbb{R}^{K_1}, \|\alpha\|=1\}} \frac{\|M_{\lambda^0} X_{\text{low},\alpha} M_{f^0} X'_{\text{low},\alpha} M_{\lambda^0}\|}{NT} &> B, \quad \text{wpa1.} \end{aligned} \quad (\text{S.4.4})$$

Proof. Note that $\|M_{\lambda^0} X_{\text{low},\alpha} M_{f^0} X'_{\text{low},\alpha} M_{\lambda^0}\| \leq \|X_{\text{low},\alpha} M_{f^0} X'_{\text{low},\alpha}\|$, 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 $X_{\text{low},\alpha} = w \text{diag}(\alpha') v'$. Using lemma S.4.1 and part (v), (vi) and (ix) of theorem S.3.1 we find

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

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

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

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

$$\tilde{S}_{NT}^{(2)}(\beta, f) \geq \frac{a_0 \left\| \beta^{\text{low}} - \beta_0^{\text{low}} \right\|^2}{\left\| \beta^{\text{low}} - \beta_0^{\text{low}} \right\|^2 + a_1 \left\| \beta^{\text{low}} - \beta_0^{\text{low}} \right\| + a_2 - a_3 \left\| \beta^{\text{high}} - \beta_0^{\text{high}} \right\| - a_4 \left\| \beta^{\text{high}} - \beta_0^{\text{high}} \right\| \left\| \beta^{\text{low}} - \beta_0^{\text{low}} \right\|}, \quad \text{wpa1}. \quad (\text{S.4.7})$$

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

$$\begin{aligned} \tilde{S}_{NT}^{(2)}(\beta, f) &\geq \frac{1}{NT} \mu_{R+1} \left[\left(\lambda^0 f^{0r} + \sum_{k=1}^K (\beta_k^0 - \beta_k) X_k \right)' P_{(\lambda_0, w)} \left(\lambda^0 f^{0r} + \sum_{k=1}^K (\beta_k^0 - \beta_k) X_k \right) \right] \\ &= \frac{1}{NT} \mu_{R+1} \left[\left(\lambda^0 f^{0r} + \sum_{l=1}^{K_1} (\beta_l^0 - \beta_l) w_l v_l' \right)' \left(\lambda^0 f^{0r} + \sum_{l=1}^{K_1} (\beta_l^0 - \beta_l) w_l v_l' \right) \right. \\ &\quad + \left(\lambda^0 f^{0r} + \sum_{l=1}^{K_1} (\beta_l^0 - \beta_l) w_l v_l' \right)' P_{(\lambda_0, w)} \sum_{m=K_1}^K (\beta_m^0 - \beta_m) X_m \\ &\quad + \sum_{m=K_1}^K (\beta_m^0 - \beta_m) X_m' P_{(\lambda_0, w)} \left(\lambda^0 f^{0r} + \sum_{l=1}^{K_1} (\beta_l^0 - \beta_l) w_l v_l' \right) \\ &\quad \left. + \sum_{m=K_1}^K (\beta_m^0 - \beta_m) X_m' P_{(\lambda_0, w)} \sum_{m=K_1}^K (\beta_m^0 - \beta_m) X_m \right] \\ &\geq \frac{1}{NT} \mu_{R+1} \left[\left(\lambda^0 f^{0r} + \sum_{l=1}^{K_1} (\beta_l^0 - \beta_l) w_l v_l' \right)' \left(\lambda^0 f^{0r} + \sum_{l=1}^{K_1} (\beta_l^0 - \beta_l) w_l v_l' \right) \right. \\ &\quad + \left(\lambda^0 f^{0r} + \sum_{l=1}^{K_1} (\beta_l^0 - \beta_l) w_l v_l' \right)' P_{(\lambda_0, w)} \sum_{m=K_1}^K (\beta_m^0 - \beta_m) X_m \\ &\quad \left. + \sum_{m=K_1}^K (\beta_m^0 - \beta_m) X_m' P_{(\lambda_0, w)} \left(\lambda^0 f^{0r} + \sum_{l=1}^{K_1} (\beta_l^0 - \beta_l) w_l v_l' \right) \right] \\ &\geq \frac{1}{NT} \mu_{R+1} \left[\left(\lambda^0 f^{0r} + \sum_{l=1}^{K_1} (\beta_l^0 - \beta_l) w_l v_l' \right)' \left(\lambda^0 f^{0r} + \sum_{l=1}^{K_1} (\beta_l^0 - \beta_l) w_l v_l' \right) \right. \\ &\quad \left. - a_3 \left\| \beta^{\text{high}} - \beta_0^{\text{high}} \right\| - a_4 \left\| \beta^{\text{high}} - \beta_0^{\text{high}} \right\| \left\| \beta^{\text{low}} - \beta_0^{\text{low}} \right\| \right], \quad \text{wpa1}, \quad (\text{S.4.8}) \end{aligned}$$

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

$$\begin{aligned} & \frac{1}{NT} \left\| \sum_{m=K_1}^K (\beta_m^0 - \beta_m) X'_m P_{(\lambda_0, w)} \left(\lambda^0 f^{0'} + \sum_{l=1}^{K_1} (\beta_l^0 - \beta_l) w_l v'_l \right) \right\| \\ & \leq K \left\| \beta^{\text{high}} - \beta_0^{\text{high}} \right\| \max_m \left\| \frac{X_m}{\sqrt{NT}} \right\| \left(\left\| \frac{\lambda^0 f^{0'}}{\sqrt{NT}} \right\| + K \left\| \beta^{\text{low}} - \beta_0^{\text{low}} \right\| \max_l \left\| \frac{w_l v'_l}{\sqrt{NT}} \right\| \right). \end{aligned} \quad (\text{S.4.9})$$

Our assumptions guarantee that the operator norms of $\lambda^0 f^{0'}/\sqrt{NT}$ and X_m/\sqrt{NT} are bounded from above as $N, T \rightarrow \infty$, which results in finite constants a_3 and a_4 .

We write the above result as $\tilde{S}_{NT}^{(2)}(\beta, f) \geq \mu_{R+1}(A'A)/(NT) + \text{terms containing } \beta^{\text{high}}$, where we defined $A = \lambda^0 f^{0'} + \sum_{l=1}^{K_1} (\beta_l^0 - \beta_l) w_l v'_l$. Also write $A = A_1 + A_2 + A_3$, with $A_1 = M_w A P_{f^0} = M_w \lambda^0 f^{0'}$, $A_2 = P_w A M_{f^0} = \sum_{l=1}^{K_1} (\beta_l^0 - \beta_l) w_l v'_l M_{f^0}$, $A_3 = P_w A P_{f^0} = P_w \lambda^0 f^{0'} + \sum_{l=1}^{K_1} (\beta_l^0 - \beta_l) w_l v'_l P_{f^0}$. We then find $A'A = A'_1 A_1 + (A'_2 + A'_3)(A_2 + A_3)$ and

$$\begin{aligned} A'A & \geq A'A - (a^{1/2} A'_3 + a^{-1/2} A'_2)(a^{1/2} A_3 + a^{-1/2} A_2) \\ & = [A'_1 A_1 - (a-1) A'_3 A_3] + (1-a^{-1}) A'_2 A_2, \end{aligned} \quad (\text{S.4.10})$$

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 by the operator norm $(1-a^{-1})\|A_2\|^2$. We therefore find

$$\begin{aligned} \frac{1}{NT} \mu_{R+1}(A'A) & \geq \frac{1}{NT} \mu_{R+1} \left[(A'_1 A_1 - (a-1) A'_3 A_3) + (1-a^{-1}) A'_2 A_2 \right] \\ & \geq \frac{1}{NT} \min \left\{ (1-a^{-1})\|A_2\|^2, \mu_R [A'_1 A_1 - (a-1) A'_3 A_3] \right\}. \end{aligned} \quad (\text{S.4.11})$$

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

$$\begin{aligned} \mu_R [A'_1 A_1 - (a-1) A'_3 A_3] & \geq \mu_R(A'_1 A_1) - \|(a-1) A'_3 A_3\| \\ & = \frac{1}{2} \mu_R(A'_1 A_1). \end{aligned} \quad (\text{S.4.12})$$

Therefore

$$\begin{aligned} \frac{1}{NT} \mu_{R+1}(A'A) & \geq \frac{1}{2NT} \mu_R(A'_1 A_1) \min \left\{ 1, \frac{2\|A_2\|^2}{2\|A_3\|^2 + \mu_R(A'_1 A_1)} \right\} \\ & \geq \frac{1}{NT} \frac{\|A_2\|^2 \mu_R(A'_1 A_1)}{2\|A\|^2 + \mu_R(A'_1 A_1)}, \end{aligned} \quad (\text{S.4.13})$$

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

$$\begin{aligned} \frac{\|A\|}{\sqrt{NT}} & \leq \frac{\|\lambda^0 f^{0'}\|}{\sqrt{NT}} + \sum_{l=1}^{K_1} |\beta_l^0 - \beta_l| \frac{\|w_l v'_l\|}{\sqrt{NT}} \leq c_0 + c_1 \left\| \beta^{\text{low}} - \beta_0^{\text{low}} \right\|, \quad \text{wpa1}, \\ \frac{\mu_R(A'_1 A_1)}{NT} & = \frac{\mu_R(f^0 \lambda^{0'} M_w \lambda^0 f^{0'})}{NT} \geq c_2, \quad \text{wpa1}, \\ \frac{\|A_2\|^2}{NT} & = \mu_1 \left[\sum_{l_1=1}^{K_1} (\beta_{l_1}^0 - \beta_{l_1}) w_{l_1} v'_{l_1} M_{f^0} \sum_{l_2=1}^{K_1} (\beta_{l_2}^0 - \beta_{l_2}) v_{l_2} w'_{l_2} \right] \\ & \geq c_3 \left\| \beta^{\text{low}} - \beta_0^{\text{low}} \right\|^2, \quad \text{wpa1}, \end{aligned} \quad (\text{S.4.14})$$

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

We thus have

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

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

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

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

S.5 Proof of Corollary 3.4

As discussed in the main text, all that is left to prove corollary 3.4 is that the matrix $W_{NT} = W_{NT}(\lambda^0, f^0, X_k)$ does not become singular as $N, T \rightarrow \infty$.

Proof. Remember that

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

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

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

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 regressors²

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

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(ii), *i.e.*

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

(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

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

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

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

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

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

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

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

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

$$\frac{1}{NT} \text{Tr} [M_{f^0} X'_{\text{low},\varphi} M_{\lambda^0} X_{\text{low},\varphi}] \geq c_2 \|\varphi\|^2. \tag{S.5.8}$$

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

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

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

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

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 4.1

Proof of lemma D.1. # We start with the proof for the operator norm of the weakly exogenous part of the regressors. We can rewrite equation (4.1) as

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

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

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

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 $\frac{1}{\sqrt{NT}} \|f^{0'} e \lambda^0\|_F = \mathcal{O}_p(1)$, because

$$\begin{aligned} \mathbb{E} \left(\frac{\|f^{0'} e \lambda^0\|_F}{\sqrt{NT}} \right)^2 &= \frac{1}{NT} E \left(\sum_{i=1}^N \sum_{t=1}^T e_{it} f_t^{0'} \lambda_i^0 \right)^2 \\ &= \frac{1}{NT} \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^T \sum_{s=1}^T E(e_{it} e_{js}) f_t^{0'} \lambda_i^0 \lambda_j^{0'} f_s^0 \\ &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \mathbb{E}(e_{it}^2) f_t^{0'} \lambda_i^0 \lambda_i^{0'} f_t^0 \\ &= \mathcal{O}(1). \end{aligned} \quad (\text{S.6.3})$$

Using this we obtain

$$\begin{aligned}
\|P_{\lambda^0} e P_{f^0}\| &= \|\lambda^0 (\lambda^0 \lambda^0)^{-1} \lambda^{0'} e f^0 (f^0 f^0)^{-1} f^{0'}\| \\
&\leq \|\lambda^0\| \|(\lambda^0 \lambda^0)^{-1}\| \|\lambda^{0'} e f^0\| \|(f^0 f^0)^{-1}\| \|f^{0'}\| \\
&\leq \mathcal{O}_p(N^{1/2}) \mathcal{O}_p(N^{-1}) \|\lambda^{0'} e f^0\|_F \mathcal{O}_p(T^{-1}) \mathcal{O}_p(T^{1/2}) = \mathcal{O}_p(1), \tag{S.6.4}
\end{aligned}$$

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

To show the next statement of the lemma we first note that $N^{-1} T^{-1/2} \|\lambda^{0'} e X_k^{\text{str}'}\|_F = \mathcal{O}_p(1)$, which is true because

$$\begin{aligned}
\mathbb{E} \left(N^{-1} T^{-1/2} \|\lambda^{0'} e X_k^{\text{str}'}\|_F \right)^2 &= N^{-2} T^{-1} \mathbb{E} \left[\sum_{r=1}^R \sum_{j=1}^N \left(\sum_{i=1}^N \sum_{t=1}^T \lambda_{ir}^0 e_{it} X_{k,jt}^{\text{str}} \right)^2 \right] \\
&= N^{-2} T^{-1} \sum_{r=1}^R \sum_{j=1}^N \sum_{i=1}^N \sum_{t=1}^T \mathbb{E} (e_{it}^2) (\lambda_{ir}^0)^2 \mathbb{E} [(X_{k,jt}^{\text{str}})^2] = \mathcal{O}(1). \tag{S.6.5}
\end{aligned}$$

Therefore we find

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

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

The proof for $\|P_{f^0} e' X_k^{\text{str}}\| = \mathcal{O}_p(\sqrt{NT})$ 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*

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

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

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

To compute its variance, we write

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

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

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

Therefore,

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

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

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

Also,

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

Finally,

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

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

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

Proof of lemma D.2.

First, we want to prove the statements (a), (b), (c), (d) and (e). We have $\|\lambda^0 (\lambda^{0'} \lambda^0)^{-1} (f^{0'} f^0)^{-1} f^{0'}\| = \mathcal{O}_p((NT)^{-1/2})$, $\|e\| = \mathcal{O}_p(N^{1/2})$, and according to lemma D.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.*

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

and

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

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 $\|X_k^{\text{str}}\| = \mathcal{O}_p(\sqrt{NT})$ and $\|P_{\lambda^0} e P_{f^0}\| = \mathcal{O}_p(1)$, which was shown in lemma D.1. We then obtain *e.g.*

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

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

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

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

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

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

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

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

where we used $\sup_{it} \mathbb{E}(e_{it}^2) \leq \bar{B}$ and $\sup_{it} \mathbb{E} \left[(X_{it}^{\text{weak}})^2 \right] \leq \bar{B}$. 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)

$$B = M_{\lambda^0} X_k^{\text{str}} f^0 (f^{0'} f^0)^{-1} (\lambda^{0'} \lambda^0)^{-1} \lambda^{0'}. \quad (\text{S.6.21})$$

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

$$\begin{aligned} \|B\|_F &\leq R^{1/2} \|B\| \\ &\leq R^{1/2} \|X_k^{\text{str}}\| \|f^0 (f^{0'} f^0)^{-1} (\lambda^{0'} \lambda^0)^{-1} \lambda^{0'}\| \\ &\leq R^{1/2} \|X_k^{\text{str}}\|_F \|f^0 (f^{0'} f^0)^{-1} (\lambda^{0'} \lambda^0)^{-1} \lambda^{0'}\|. \end{aligned} \quad (\text{S.6.22})$$

and therefore

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

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

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

and thus

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

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}^{\text{weak}} - \mathbb{E} \left(e_{it} \tilde{f}_{ts} X_{k,is}^{\text{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}^{\text{weak}} X_{k,jq}^{\text{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}^{\text{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}^{\text{weak}} X_{k,iq}^{\text{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}^{\text{weak}} \right) \right]^2 \\ &\leq \frac{1}{N^2} \sum_{i=1}^N \mathbb{E} \left[\begin{aligned} &\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}^{\text{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}^{\text{weak}} \left(\frac{f' f}{T} \right)^{-1/2} f_s \right) \end{aligned} \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}^{\text{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}^{\text{weak}} \left(\frac{f' f}{T} \right)^{-1/2} f_t \right\|^4 \right] \\ &= \mathcal{O} \left(\frac{1}{N} \right), \end{aligned} \quad (\text{S.6.26})$$

where the last line holds because

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

for some c_0 independent of N, T . ■

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

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

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

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

Using our assumptions we thus find that

$$\frac{1}{N} \sum_{i=1}^N \mathbb{E}(Z_{iT}^2 | X^{\text{str}}) \xrightarrow{p} c' \Omega c > 0. \quad (\text{S.6.30})$$

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

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

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 4.4

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

Lemma S.7.1. *Let Z_i , $i = 1, 2, \dots, n$, 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 $n \rightarrow \infty$ we then have*

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

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*

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

Under assumption 5 we have

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

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

Proof. # We start with the proof for $\bar{Z}_{k,t\tau}^{(1)}$. Define $Z_{k,t\tau,i}^{(1)} = e_{it} X_{k,i\tau} - \mathbb{E}(e_{it} X_{k,i\tau})$. 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 $Z_{k,t\tau,i}^{(1)}$ uniformly across k, i, t, τ . For ease of notation we now fix k, t, τ and write $Z_i = Z_{k,t\tau,i}^{(1)}$. 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^{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

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

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

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

Lemma S.7.3. *For a $T \times T$ matrix A we have³*

$$\begin{aligned}\|A^{\text{truncR}}\| &\leq M \|A^{\text{truncR}}\|_{\max} \equiv M \max_t \max_{t < \tau \leq t+M} |A_{t\tau}| , \\ \|A^{\text{truncL}}\| &\leq M \|A^{\text{truncL}}\|_{\max} \equiv M \max_t \max_{t-M \leq \tau < t} |A_{t\tau}| .\end{aligned}\tag{S.7.5}$$

³For 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 $\tau < 1$ or $t > T$ or $\tau > T$.

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

$$\begin{aligned} \|A^{\text{truncR}}\|_1 &= \max_{t=1\dots T} \sum_{\tau=t+1}^{t+M} |A_{t\tau}| \\ &\leq M \max_{t,\tau=1\dots T} \max_{t < \tau \leq t+M} |A_{t\tau}|, \end{aligned} \quad (\text{S.7.6})$$

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

Proof of lemma E.3. # First, we are going to show $A_1 = o_p(1)$. Let $B_{1,it} = \hat{\mathbf{x}}_{it} - \mathbf{x}_{it}$, $B_{2,it} = e_{it}^2 \mathbf{x}_{it}$, and $B_{3,it} = e_{it}^2 \hat{\mathbf{x}}_{it}$. 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 $k = 1, \dots, K$. We have $A_1 = (NT)^{-1} \sum_i \sum_t (B_{1,it} B'_{2,it} + B_{3,it} B'_{1,it})$, or equivalently

$$A_{1,k_1 k_2} = \frac{1}{NT} \text{Tr} (B_{1,k_1} B'_{3,k_2} + B_{2,k_1} B'_{1,k_2}). \quad (\text{S.7.7})$$

We find that

$$B_{1,k} = -P_{f_0} X_k^{\text{weak}} M_{\lambda^0} - X_k^{\text{weak}} P_{\lambda^0} + (M_{\hat{\lambda}} - M_{\lambda^0}) X_k M_{f_0} + M_{\hat{\lambda}} X_k (M_{\hat{f}} - M_{f_0}), \quad (\text{S.7.8})$$

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 $\|X_k^{\text{weak}}\| = \mathcal{O}_p(N^{1/2})$. In addition we have $\text{rank}(B_{1,k}) \leq 6R$.

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

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

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

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

This is what we wanted to show.

Next, we want to show $A_2 = o_p(1)$. According to theorem 3.3 we have $e - \hat{e} = C_1 + C_2$, where we defined $C_1 = -\sum_{k=1}^K (\hat{\beta}_k - \beta_k^0) X_k$, and $C_2 = \sum_{k=1}^K (\hat{\beta}_k - \beta_k^0) (P_{\lambda^0} X_k M_{f_0} + X_k P_{f_0}) + P_{\lambda^0} e M_{f_0} + e P_{f_0} - \hat{e}_e^{(1)} - \hat{e}^{(\text{rem})}$, which satisfies $\|C_2\| = \mathcal{O}_p(N^{1/2})$, and $\text{rank}(C_2) \leq 11R$ (actually, one can easily prove $\leq 5R$, but this does not follow from theorem 3.3). Using this notation we have

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

which can also be written as

$$A_{2,k_1 k_2} = - \sum_{k_3=1}^K \left(\hat{\beta}_{k_3} - \beta_{k_3}^0 \right) (C_{5,k_1 k_2 k_3} + C_{6,k_1 k_2 k_3}) + \frac{1}{NT} \text{Tr} (C_2 C_{3,k_1 k_2}) + \frac{1}{NT} \text{Tr} (C_2 C_{4,k_1 k_2}), \quad (\text{S.7.12})$$

where we defined

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

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

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

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

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

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

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

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 $\Gamma(\cdot)$ 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 $\hat{f} = P_{\hat{f}}f^0H$, where $H = H_{NT}$ is a non-singular $R \times R$ matrix.

We now want to show $\|H\| = \mathcal{O}_p(1)$ and $\|H^{-1}\| = \mathcal{O}_p(1)$. 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 $\hat{f} = f^0H + (P_{\hat{f}} - P_{f^0})f^0H$, and thus $H = f^{0'}\hat{f}/T - f^{0'}(P_{\hat{f}} - P_{f^0})f^0H/T$, i.e. $\|H\| \leq \mathcal{O}_p(1) + \|H\|\mathcal{O}_p(T^{-1/2})$ 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 $\|P_{\hat{f}} - P_{f^0}\| = \mathcal{O}_p(T^{-1/2})$ of lemma E.2 immediately implies

$$\|\hat{f} - f^0H\| = \mathcal{O}_p(1) . \quad (\text{S.7.18})$$

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

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

which yields

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

We have $(f^{0'} P_{\hat{f}} f^0 / T)^{-1} - (f^{0'} f^0 / T)^{-1} = \mathcal{O}_p(T^{-1/2})$, because $\|P_{\hat{f}} - P_{f^0}\| = \mathcal{O}_p(T^{-1/2})$ and $f^{0'} f^0 / T$ 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

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

which is what we wanted to prove.

Next, we want to show

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

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

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

By assumption 1 we know that

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

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

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

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

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

which is equivalent to (E.5). 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

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

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

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

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 $\Gamma(\cdot)$) 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

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

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 $M \rightarrow \infty$. The proof of

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

is analogous.

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

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

Using lemma S.7.3 we have

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

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

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

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

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

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

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 $N^{-1} \sum_{i=1}^N X_{k,i\tau} \lambda_i^{0'}$ have bounded 8'th moments uniformly across t . We also used $M^4/T \rightarrow 0$. Using the above decomposition of $e'_t P_{\lambda^0} X_{k,\tau}$ we find that $\mathbb{E}(N^{-1/2} e'_t P_{\lambda^0} X_{k,\tau})$ is uniformly bounded across t, τ , and therefore

$$\max_t \max_{t < \tau \leq t+M} M N^{-1} |\mathbb{E}(e'_t P_{\lambda^0} X_{k,\tau})| \leq M N^{-1/2} \mathcal{O}_p(1) = o_p(1). \quad (\text{S.7.35})$$

The proof of

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

is analogous.

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

$$N^{-1} \left\| [e' M_{\lambda_0} X_k - \hat{e}' X_k]^{\text{truncR}} \right\| = o_p(1). \quad (\text{S.7.37})$$

According to theorem 3.3 we have $\hat{e} = M_{\lambda_0} e M_{f^0} + e_{\text{rem}}$, where $e_{\text{rem}} = \hat{e}_e^{(1)} - \sum_{k=1}^K (\hat{\beta}_k - \beta_k^0) \hat{e}_k^{(1)} + \hat{e}^{(\text{rem})}$, and using corollary E.1 we find that the remainder term satisfies $\|e_{\text{rem}}\| = \mathcal{O}_p(1)$. We have to show

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

Using lemma S.7.3 we find

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

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}_{rem} . In addition we used lemma S.7.1 and the fact that $N^{-1/2} \|X_{k,\tau}\| = \sqrt{N^{-1} \sum_{i=1}^N X_{k,i\tau}^2}$ 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

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

Thus, we proved equation (S.7.37).

The proof of

$$\begin{aligned} N^{-1} \left\| [e' M_{\lambda_0} e - \hat{e}' \hat{e}]^{\text{truncD}} \right\| &= o_p(1), \\ T^{-1} \left\| [e M_{f^0} e' - \hat{e} \hat{e}']^{\text{truncD}} \right\| &= o_p(1), \end{aligned} \quad (\text{S.7.41})$$

is analogous.

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

Proof of lemma E.6. Using theorem 3.3 and E.1 we find $\|\hat{e}\| = \mathcal{O}_p(N^{1/2})$. Applying lemma S.7.3 we therefore find

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

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.8 Algebraic Results for Higher Order Approximations

Applying the formula (C.5) for $L^{(g)}$ we find for $g \leq 5$

$$\begin{aligned}
L^{(1)}(\lambda^0, f^0, X_{\kappa_1}) &= 0, \\
L^{(2)}(\lambda^0, f^0, X_{\kappa_1}, X_{\kappa_2}) &= \text{Tr}(M_{\lambda^0} X_{\kappa_1} M_{f^0} X_{\kappa_2}), \\
L^{(3)}(\lambda^0, f^0, X_{\kappa_1}, X_{\kappa_2}, X_{\kappa_3}) &= -\frac{1}{6} \left[\text{Tr}(2 M_{\lambda^0} X_{\kappa_1} M_f X'_{\kappa_2} \lambda^0 (\lambda^{0'} \lambda^0)^{-1} (f^{0'} f^0)^{-1} f^{0'} X'_{\kappa_3}) \right. \\
&\quad \left. + 5 \text{ permutations of } \kappa_1 \dots \kappa_3 \right], \\
L^{(4)}(\lambda^0, f^0, X_{\kappa_1}, X_{\kappa_2}, X_{\kappa_3}, X_{\kappa_4}) \\
&= \frac{1}{24} \text{Tr} \left[-X_{\kappa_1} f^0 (f^{0'} f^0)^{-1} (\lambda^{0'} \lambda^0)^{-1} (f^{0'} f^0)^{-1} f^{0'} X'_{\kappa_2} M_{\lambda^0} X_{\kappa_3} M_{f^0} X'_{\kappa_4} M_{\lambda^0} \right. \\
&\quad - X_{\kappa_1} M_{f^0} X'_{\kappa_2} M_{\lambda^0} X_{\kappa_3} M_{f^0} X'_{\kappa_4} \lambda^0 (\lambda^{0'} \lambda^0)^{-1} (f^{0'} f^0)^{-1} (\lambda^{0'} \lambda^0)^{-1} \lambda^{0'} \\
&\quad + 2X_{\kappa_1} f^0 (f^{0'} f^0)^{-1} (\lambda^{0'} \lambda^0)^{-1} \lambda^{0'} X_{\kappa_2} f^0 (f^{0'} f^0)^{-1} (\lambda^{0'} \lambda^0)^{-1} \lambda^{0'} X_{\kappa_3} M_{f^0} X'_{\kappa_4} M_{\lambda^0} \\
&\quad + X_{\kappa_1} f^0 (f^{0'} f^0)^{-1} (\lambda^{0'} \lambda^0)^{-1} \lambda^{0'} X_{\kappa_2} M_{f^0} X'_{\kappa_3} \lambda^0 (\lambda^{0'} \lambda^0)^{-1} (f^{0'} f^0)^{-1} f^{0'} X'_{\kappa_4} M_{\lambda^0} \\
&\quad \left. + 23 \text{ permutations of } \kappa_1 \dots \kappa_4 \text{ for all four terms} \right], \\
L^{(5)}(\lambda^0, f^0, X_{\kappa_1}, X_{\kappa_2}, X_{\kappa_3}, X_{\kappa_4}, X_{\kappa_5}) &= \frac{1}{120} \text{Tr} \left[\right. \\
&\quad - 2X_{\kappa_1} f^0 (f^{0'} f^0)^{-1} (\lambda^{0'} \lambda^0)^{-1} (f^{0'} f^0)^{-1} (\lambda^{0'} \lambda^0)^{-1} \lambda^{0'} X_{\kappa_2} M_{f^0} X'_{\kappa_3} M_{\lambda^0} X_{\kappa_4} M_{f^0} X'_{\kappa_5} M_{\lambda^0} \\
&\quad + 2X_{\kappa_1} f^0 (f^{0'} f^0)^{-1} (\lambda^{0'} \lambda^0)^{-1} (f^{0'} f^0)^{-1} f^{0'} X'_{\kappa_2} M_{\lambda^0} X_{\kappa_3} f^0 (f^{0'} f^0)^{-1} (\lambda^{0'} \lambda^0)^{-1} \lambda^{0'} X_{\kappa_4} M_{f^0} X'_{\kappa_5} M_{\lambda^0} \\
&\quad + 2X_{\kappa_1} f^0 (f^{0'} f^0)^{-1} (\lambda^{0'} \lambda^0)^{-1} (f^{0'} f^0)^{-1} f^{0'} X'_{\kappa_2} M_{\lambda^0} X_{\kappa_3} M_{f^0} X'_{\kappa_4} M_{\lambda^0} X_{\kappa_5} f^0 (f^{0'} f^0)^{-1} (\lambda^{0'} \lambda^0)^{-1} \lambda^{0'} \\
&\quad + 2X_{\kappa_1} f^0 (f^{0'} f^0)^{-1} (\lambda^{0'} \lambda^0)^{-1} \lambda^{0'} X_{\kappa_2} M_{f^0} X'_{\kappa_3} M_{\lambda^0} X_{\kappa_4} M_{f^0} X'_{\kappa_5} \lambda^0 (\lambda^{0'} \lambda^0)^{-1} (f^{0'} f^0)^{-1} (\lambda^{0'} \lambda^0)^{-1} \lambda^{0'} \\
&\quad + 2X_{\kappa_1} f^0 (f^{0'} f^0)^{-1} (\lambda^{0'} \lambda^0)^{-1} \lambda^{0'} X_{\kappa_2} M_{f^0} X'_{\kappa_3} \lambda^0 (\lambda^{0'} \lambda^0)^{-1} (f^{0'} f^0)^{-1} (\lambda^{0'} \lambda^0)^{-1} \lambda^{0'} X_{\kappa_4} M_{f^0} X'_{\kappa_5} M_{\lambda^0} \\
&\quad - 2X_{\kappa_1} f^0 (f^{0'} f^0)^{-1} (\lambda^{0'} \lambda^0)^{-1} \lambda^{0'} X_{\kappa_2} f^0 (f^{0'} f^0)^{-1} (\lambda^{0'} \lambda^0)^{-1} \lambda^{0'} X_{\kappa_3} f^0 \\
&\quad \quad \quad (f^{0'} f^0)^{-1} (\lambda^{0'} \lambda^0)^{-1} \lambda^{0'} X_{\kappa_4} M_{f^0} X'_{\kappa_5} M_{\lambda^0} \\
&\quad - 2X_{\kappa_1} f^0 (f^{0'} f^0)^{-1} (\lambda^{0'} \lambda^0)^{-1} \lambda^{0'} X_{\kappa_2} f^0 (f^{0'} f^0)^{-1} (\lambda^{0'} \lambda^0)^{-1} \lambda^{0'} X_{\kappa_3} M_{f^0} \\
&\quad \quad \quad X'_{\kappa_4} \lambda^0 (\lambda^{0'} \lambda^0)^{-1} (f^{0'} f^0)^{-1} f^{0'} X'_{\kappa_5} M_{\lambda^0} \\
&\quad \left. + 119 \text{ permutations of } \kappa_1 \dots \kappa_5 \text{ for all seven terms} \right]. \tag{S.8.1}
\end{aligned}$$

These results can be helpful for future applications of the likelihood-expansion, *e.g.* to discuss higher orders of the asymptotic theory of the QMLE.

References

- Bai, J. (2009). Panel data models with interactive fixed effects. *Forthcoming in Econometrica*.
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.