Supplement to 'Individual and Time Effects in Nonlinear Panel Models with Large N, T'

Iván Fernández-Val[‡] Martin Weidner[§]

March 31, 2015

Abstract

This supplemental material contains five appendices. Appendix S.1 presents the results of an empirical application and a Monte Carlo simulation calibrated to the application. Following Aghion *et al.* (2005), we use a panel of U.K. industries to estimate Poisson models with industry and time effects for the relationship between innovation and competition. Appendix S.2 gives the proofs of Theorems 4.3 and 4.4. Appendices S.3, S.4, and S.5 contain the proofs of Appendices B, C, and D, respectively. Appendix S.6 collects some useful intermediate results that are used in the proofs of the main results.

S.1 Relationship between Innovation and Competition

S.1.1 Empirical Example

To illustrate the bias corrections with real data, we revisit the empirical application of Aghion, Bloom, Blundell, Griffith and Howitt (2005) (ABBGH) that estimated a count data model to analyze the relationship between innovation and competition. They used an unbalanced panel of seventeen U.K. industries followed over the 22 years between 1973 and 1994.[1] The dependent variable, Y_{it}, is innovation as measured by a citation-weighted number of patents, and the explanatory variable of interest, Z_{it}, is competition as measured by one minus the Lerner index in the industry-year.

Following ABBGH we consider a quadratic static Poisson model with industry and year effects where

$$Y_{it} \mid Z_i^T, \alpha_i, \gamma_t \sim \mathcal{P}(\exp[\beta_1 Z_{it} + \beta_2 Z_{it}^2 + \alpha_i + \gamma_t]),$$

for $(i = 1, ..., 17; t = 1973, ..., 1994)$, and extend the analysis to a dynamic Poisson model with industry and year effects where

$$Y_{it} \mid Y_i^{t-1}, Z_i^t, \alpha_i, \gamma^t \sim \mathcal{P}(\exp[\beta_Y \log(1 + Y_{i,t-1}) + \beta_1 Z_{it} + \beta_2 Z_{it}^2 + \alpha_i + \gamma_t]),$$

[‡]Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215-1403, USA. Email: `ivanf@bu.edu`

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

[1]We assume that the observations are missing at random conditional on the explanatory variables and unobserved effects and apply the corrections without change since the level of attrition is low in this application.

for ($i = 1, ..., 17; t = 1974, ..., 1994$). In the dynamic model we use the year 1973 as the initial condition for Y_{it}.

Table S1 reports the results of the analysis. Columns (2) and (3) for the static model replicate the empirical results of Table I in ABBGH (p. 708), adding estimates of the APEs. Columns (4) and (5) report estimates of the analytical corrections that do not assume that competition is strictly exogenous with $L = 1$ and $L = 2$, and column (6) reports estimates of the jackknife bias corrections described in equation (3.4) of the paper. Note that we do not need to report separate standard errors for the corrected estimators, because the standard errors of the uncorrected estimators are consistent for the corrected estimators under the asymptotic approximation that we consider.[2] Overall, the corrected estimates, while numerically different from the uncorrected estimates in column (3), agree with the inverted-U pattern in the relationship between innovation and competition found by ABBGH. The close similarity between the uncorrected and bias corrected estimates gives some evidence in favor of the strict exogeneity of competition with respect to the innovation process.

The results for the dynamic model show substantial positive state dependence in the innovation process that is not explained by industry heterogeneity. Uncorrected fixed effects underestimates the coefficient and APE of lag patents relative to the bias corrections, specially relative to the jackknife. The pattern of the differences between the estimates is consistent with the biases that we find in the numerical example in Table S4. Accounting for state dependence does not change the inverted-U pattern, but flattens the relationship between innovation and competition.

Table S.2 implements Chow-type homogeneity tests for the validity of the jackknife corrections. These tests compare the uncorrected fixed effects estimators of the common parameters within the elements of the cross section and time series partitions of the panel. Under time homogeneity, the probability limit of these estimators is the same, so that a standard Wald test can be applied based on the difference of the estimators in the sub panels within the partition. For the static model, the test is rejected at the 1% level in both the cross section and time series partitions. Since the cross sectional partition is arbitrary, these rejection might be a signal of model misspecification. For the dynamic model, the test is rejected at the 1% level in the time series partition, but it cannot be rejected at conventional levels in the cross section partition. The rejection of the time homogeneity might explain the difference between the jackknife and analytical corrections in the dynamic model.

S.1.2 Calibrated Monte Carlo Simulations

We conduct a simulation that mimics the empirical example. The designs correspond to static and dynamic Poisson models with additive individual and time effects. We calibrate all the parameters and exogenous variables using the dataset from ABBGH.

[2] In numerical examples, we find very little gains in terms of the ratio SE/SD and coverage probabilities when we reestimate the standard errors using bias corrected estimates.

S.1.2.1 Static Poisson model

The data generating process is

$$Y_{it} \mid Z_i^T, \alpha, \gamma \sim \mathcal{P}(\exp[Z_{it}\beta_1 + Z_{it}^2\beta_2 + \alpha_i + \gamma_t]), \quad (i = 1, ..., N;\ t = 1, ..., T),$$

where $\mathcal{P}$ denotes the Poisson distribution. The variable Z_{it} is fixed to the values of the competition variable in the dataset and all the parameters are set to the fixed effect estimates of the model. We generate unbalanced panel data sets with $T = 22$ years and three different numbers of industries N: 17, 34, and 51. In the second (third) case, we double (triple) the cross-sectional size by merging two (three) independent realizations of the panel.

Table S3 reports the simulation results for the coefficients β_1 and β_2, and the APE of Z_{it}. We compute the APE using the expression (2.5) with $H(Z_{it}) = Z_{it}^2$. Throughout the table, MLE corresponds to the pooled Poisson maximum likelihood estimator (without individual and time effects), MLE-TE corresponds to the Poisson estimator with only time effects, MLE-FETE corresponds to the Poisson maximum likelihood estimator with individual and time fixed effects, Analytical (L=l) is the bias corrected estimator that uses the analytical correction with $L = l$, and Jackknife is the bias corrected estimator that uses SPJ in both the individual and time dimensions. The analytical corrections are different from the uncorrected estimator because they do not use that the regressor Z_{it} is strictly exogenous. The cross-sectional division in the jackknife follows the order of the observations. The choice of these estimators is motivated by the empirical analysis of ABBGH. All the results in the table are reported in percentage of the true parameter value.

The results of the table agree with the no asymptotic bias result for the Poisson model with exogenous regressors. Thus, the bias of MLE-FETE for the coefficients and APE is negligible relative to the standard deviation and the coverage probabilities get close to the nominal level as N grows. The analytical corrections preserve the performance of the estimators and have very little sensitivity to the trimming parameter. The jackknife correction increases dispersion and rmse, specially for the small cross-sectional size of the application. The estimators that do not control for individual effects are clearly biased.

S.1.2.2 Dynamic Poisson model

The data generating process is

$$Y_{it} \mid Y_i^{t-1}, Z_i^t, \alpha, \gamma \sim \mathcal{P}(\exp[\beta_Y \log(1 + Y_{i,t-1}) + Z_{it}\beta_1 + Z_{it}^2\beta_2 + \alpha_i + \gamma_t]), \quad (i = 1, ..., N; t = 1, ..., T).$$

The competition variable Z_{it} and the initial condition for the number of patents Y_{i0} are fixed to the values in the dataset and all the parameters are set to the fixed effect estimates of the model. To generate panels, we first impute values to the missing observations of Z_{it} using forward and backward predictions from a panel AR(1) linear model with individual and time effects. We then draw panel data sets with $T = 21$ years and three different numbers of industries N: 17, 34, and 51. As in the static model, we double (triple) the cross-sectional size by merging two (three) independent realizations of the panel. We make the generated panels unbalanced by dropping the values corresponding to the missing observations in the original dataset.

Table S4 reports the simulation results for the coefficient β_Y^0 and the APE of $Y_{i,t-1}$. The estimators considered are the same as for the static Poisson model above. We compute the partial effect of $Y_{i,t-1}$ using (2.5) with $Z_{it} = Y_{i,t-1}$, $H(Z_{it}) = \log(1 + Z_{it})$, and dropping the linear term. Table S5 reports the simulation results for the coefficients β_1^0 and β_2^0, and the APE of Z_{it}. We compute the partial effect using (2.5) with $H(Z_{it}) = Z_{it}^2$. Again, all the results in the tables are reported in percentage of the true parameter value.

The results in table S4 show biases of the same order of magnitude as the standard deviation for the fixed effects estimators of the coefficient and APE of $Y_{i,t-1}$, which cause severe undercoverage of confidence intervals. Note that in this case the rate of convergence for the estimator of the APE is $r_{NT} = \sqrt{NT}$, because the individual and time effects are hold fixed across the simulations. The analytical corrections reduce bias by more than half without increasing dispersion, substantially reducing rmse and bringing coverage probabilities closer to their nominal levels. The jackknife corrections reduce bias and increase dispersion leading to lower improvements in rmse and coverage probability than the analytical corrections. The results for the coefficient of Z_{it} in table 8 are similar to the static model. The results for the APE of Z_{it} are imprecise, because the true value of the effect is close to zero.

S.2 Proofs of Theorems 4.3 and 4.4

We start with a lemma that shows the consistency of the fixed effects estimators of averages of the data and parameters. We will use this result to show the validity of the analytical bias corrections and the consistency of the variance estimators.

Lemma S.1. *Let* $G(\beta, \phi) := [N(T-j)]^{-1} \sum_{i,t \geq j+1} g(X_{it}, X_{i,t-j}, \beta, \alpha_i + \gamma_t, \alpha_i + \gamma_{t-j})$ *for* $0 \leq j < T$, *and* $\mathcal{B}_\varepsilon^0$ *be a subset of* $\mathbb{R}^{\dim \beta + 2}$ *that contains an* ε*-neighborhood of* $(\beta, \pi_{it}^0, \pi_{i,t-j}^0)$ *for all* i, t, j, N, T, *and for some* $\varepsilon > 0$. *Assume that* $(\beta, \pi_1, \pi_2) \mapsto g_{itj}(\beta, \pi_1, \pi_2) := g(X_{it}, X_{i,t-j}, \beta, \pi_1, \pi_2)$ *is Lipschitz continuous over* $\mathcal{B}_\varepsilon^0$ *a.s, i.e.* $|g_{itj}(\beta_1, \pi_{11}, \pi_{21}) - g_{itj}(\beta_0, \pi_{10}, \pi_{20})| \leq M_{itj} \|(\beta_1, \pi_{11}, \pi_{21}) - (\beta, \pi_{10}, \pi_{20})\|$ *for all* $(\beta_0, \pi_{10}, \pi_{20}) \in \mathcal{B}_\varepsilon^0$, $(\beta_1, \pi_{11}, \pi_{21}) \in \mathcal{B}_\varepsilon^0$, *and some* $M_{itj} = \mathcal{O}_P(1)$ *for all* i, t, j, N, T. *Let* $(\widehat{\beta}, \widehat{\phi})$ *be an estimator of* (β, ϕ) *such that* $\|\widehat{\beta} - \beta^0\| \to_P 0$ *and* $\|\widehat{\phi} - \phi^0\|_\infty \to_P 0$. *Then,*

$$G(\widehat{\beta}, \widehat{\phi}) \to_P \overline{\mathbb{E}}[G(\beta^0, \phi^0)],$$

provided that the limit exists.

Proof of Lemma S.1. By the triangle inequality

$$|G(\widehat{\beta}, \widehat{\phi}) - \overline{\mathbb{E}}[G(\beta^0, \phi^0)]| \leq |G(\widehat{\beta}, \widehat{\phi}) - G(\beta^0, \phi^0)| + o_P(1),$$

because $|G(\beta^0, \phi^0) - \overline{\mathbb{E}}[G(\beta^0, \phi^0)]| = o_P(1)$. By the local Lipschitz continuity of g_{itj} and the consistency of $(\widehat{\beta}, \widehat{\phi})$,

$$\begin{aligned} |G(\widehat{\beta}, \widehat{\phi}) - G(\beta^0, \phi^0)| &\leq \frac{1}{N(T-j)} \sum_{i,t \geq j+1} M_{itj} \|(\widehat{\beta}, \widehat{\alpha}_i + \widehat{\gamma}_t, \widehat{\alpha}_i + \widehat{\gamma}_{t-j}) - (\beta^0, \alpha_i^0 + \gamma_t^0, \alpha_i^0 + \gamma_{t-j}^0)\| \\ &\leq \frac{1}{N(T-j)} \sum_{i,t \geq j+1} M_{itj} (\|\widehat{\beta} - \beta^0\| + 4\|\widehat{\phi} - \phi^0\|_\infty) \end{aligned}$$

wpa1. The result then follows because $[N(T-j)]^{-1}\sum_{i,\tau\geq t} M_{it\tau} = \mathcal{O}_P(1)$ and $(\|\widehat{\beta}-\beta^0\|+4\|\widehat{\phi}-\phi^0\|_\infty) = o_P(1)$ by assumption. ∎

Proof of Theorem 4.3. We separate the proof in three parts corresponding to the three statements of the theorem.

Part I: Proof of $\widehat{W} \to_P \overline{W}_\infty$. The asymptotic variance and its fixed effects estimators can be expressed as $\overline{W}_\infty = \overline{\mathbb{E}}[W(\beta^0,\phi^0)]$ and $\widehat{W} = W(\widehat{\beta},\widehat{\phi})$, where $W(\beta,\phi)$ has a first order representation as a continuously differentiable transformation of terms that have the form of $G(\beta,\phi)$ in Lemma S.1. The result then follows by the continuous mapping theorem noting that $\|\widehat{\beta}-\beta^0\| \to_P 0$ and $\|\widehat{\phi}-\phi^0\|_\infty \leq \|\widehat{\phi}-\phi^0\|_q \to_P 0$ by Theorem C.1.

Part II: Proof of $\sqrt{NT}(\widetilde{\beta}^A - \beta^0) \to_d \mathcal{N}(0,\overline{W}_\infty^{-1})$. By the argument given after equation (3.3) in the text, we only need to show that $\widehat{B} \to_P \overline{B}_\infty$ and $\widehat{D} \to_P \overline{D}_\infty$. These asymptotic biases and their fixed effects estimators are either time-series averages of fractions of cross-sectional averages, or vice versa. The nesting of the averages makes the analysis a bit more cumbersome than the analysis of $\widehat{W}$, but the result follows by similar standard arguments, also using that $L \to \infty$ and $L/T \to 0$ guarantee that the trimmed estimator in $\widehat{B}$ is also consistent for the spectral expectations; see Lemma 6 in Hahn and Kuersteiner (2011).

Part III: Proof of $\sqrt{NT}(\widetilde{\beta}^J - \beta^0) \to_d \mathcal{N}(0,\overline{W}_\infty^{-1})$. For $\mathcal{T}_1 = \{1,\ldots,\lfloor (T+1)/2\rfloor\}$, $\mathcal{T}_2 = \{\lfloor T/2\rfloor + 1,\ldots,T\}$, $\mathcal{T}_0 = \mathcal{T}_1 \cup \mathcal{T}_2$, $\mathcal{N}_1 = \{1,\ldots,\lfloor (N+1)/2\rfloor\}$, $\mathcal{N}_2 = \{\lfloor N/2\rfloor + 1,\ldots,N\}$, and $\mathcal{N}_0 = \mathcal{N}_1 \cup \mathcal{N}_2$, let $\widehat{\beta}^{(jk)}$ be the fixed effect estimator of β in the subpanel defined by $i \in \mathcal{N}_j$ and $t \in \mathcal{T}_k$.[3] In this notation,

$$\widetilde{\beta}^J = 3\widehat{\beta}^{(00)} - \widehat{\beta}^{(10)}/2 - \widehat{\beta}^{(20)}/2 - \widehat{\beta}^{(01)}/2 - \widehat{\beta}^{(02)}/2.$$

We derive the asymptotic distribution of $\sqrt{NT}(\widetilde{\beta}^J - \beta^0)$ from the joint asymptotic distribution of the vector $\widehat{\mathbb{B}} = \sqrt{NT}(\widehat{\beta}^{(00)} - \beta^0, \widehat{\beta}^{(10)} - \beta^0, \widehat{\beta}^{(20)} - \beta^0, \widehat{\beta}^{(01)} - \beta^0, \widehat{\beta}^{(02)} - \beta^0)$ with dimension $5 \times \dim \beta$. By Theorem C.1,

$$\sqrt{NT}(\widehat{\beta}^{(jk)} - \beta^0) = \frac{2^{1(j>0)}2^{1(k>0)}}{\sqrt{NT}} \sum_{i\in N_j, t\in T_k} [\psi_{it} + b_{it} + d_{it}] + o_P(1),$$

for $\psi_{it} = \overline{W}_\infty^{-1} D_\beta \ell_{it}$, $b_{it} = \overline{W}_\infty^{-1}[U_{it}^{(1a,1)} + U_{it}^{(1b,1,1)}]$, and $d_{it} = \overline{W}_\infty^{-1}[U_{it}^{(1a,4)} + U_{it}^{(1b,4,4)}]$, where the $U_{it}^{(\cdot)}$ is implicitly defined by $U^{(\cdot)} = (NT)^{-1/2}\sum_{i,t} U_{it}^{(\cdot)}$. Here, none of the terms carries a superscript (jk) by Assumption 4.3. The influence function ψ_{it} has zero mean and determines the asymptotic variance $\overline{W}_\infty^{-1}$, whereas b_{it} and d_{it} determine the asymptotic biases $\overline{B}_\infty$ and $\overline{D}_\infty$, but do not affect the asymptotic variance. By this representation,

$$\widehat{\mathbb{B}} \to_d \mathcal{N}\left(\kappa \begin{bmatrix} 1 \\ 1 \\ 1 \\ 2 \\ 2 \end{bmatrix} \otimes \overline{B}_\infty + \kappa^{-1} \begin{bmatrix} 1 \\ 2 \\ 2 \\ 1 \\ 1 \end{bmatrix} \otimes \overline{D}_\infty, \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 0 & 1 & 1 \\ 1 & 0 & 2 & 1 & 1 \\ 1 & 1 & 1 & 2 & 0 \\ 1 & 1 & 1 & 0 & 2 \end{bmatrix} \otimes \overline{W}_\infty^{-1} \right),$$

[3] Note that this definition of the subpanels covers all the cases regardless of whether N and T are even or odd.

where we use that $\{\psi_{it} : 1 \leq i \leq N, 1 \leq t \leq T\}$ is independent across i and martingale difference across t and Assumption 4.3.

The result follows by writing $\sqrt{NT}(\widetilde{\beta}^J - \beta^0) = (3, -1/2, -1/2, -1/2, -1/2)\widehat{\mathbb{B}}$ and using the properties of the multivariate normal distribution. ■

Proof of Theorem 4.4. We separate the proof in three parts corresponding to the three statements of the theorem.

Part I: $\widehat{V}^\delta \to_P \overline{V}^\delta_\infty$. $\overline{V}^\delta_\infty$ and $\widehat{V}^\delta$ have a similar structure to $\overline{W}_\infty$ and $\widehat{W}$ in part I of the proof of Theorem 4.3, so that the consistency follows by an analogous argument.

Part II: $\sqrt{NT}(\widetilde{\delta}^A - \delta^0_{NT}) \to_d \mathcal{N}(0, \overline{V}^\delta_\infty)$. As in the proof of Theorem 4.2, we decompose

$$r_{NT}(\widetilde{\delta}^A - \delta^0_{NT}) = r_{NT}(\delta - \delta^0_{NT}) + \frac{r_{NT}}{\sqrt{NT}}\sqrt{NT}(\widetilde{\delta}^A - \delta).$$

Then, by Mann-Wald theorem,

$$\sqrt{NT}(\widetilde{\delta}^A - \delta) = \sqrt{NT}(\widehat{\delta} - \widehat{B}^\delta/T - \widehat{D}^\delta/N - \delta) \to_d \mathcal{N}(0, \overline{V}^{\delta(1)}_\infty),$$

provided that $\widehat{B}^\delta \to_P \overline{B}^\delta_\infty$ and $\widehat{D}^\delta \to_P \overline{D}^\delta_\infty$, and $r_{NT}(\delta - \delta^0_{NT}) \to_d \mathcal{N}(0, \overline{V}^{\delta(2)}_\infty)$, where $\overline{V}^{\delta(1)}_\infty$ and $\overline{V}^{\delta(2)}_\infty$ are defined as in the proof of Theorem 4.2. The statement thus follows by using a similar argument to part II of the proof of Theorem 4.3 to show the consistency of $\widehat{B}^\delta$ and $\widehat{D}^\delta$, and because $(\delta - \delta^0_{NT})$ and $(\widetilde{\delta}^A - \delta)$ are asymptotically independent, and $\overline{V}^\delta_\infty = \overline{V}^{\delta(2)} + \overline{V}^{\delta(1)} \lim_{N,T\to\infty}(r_{NT}/\sqrt{NT})^2$.

Part III: $\sqrt{NT}(\widetilde{\delta}^J - \delta^0_{NT}) \to_d \mathcal{N}(0, \overline{V}^\delta_\infty)$. As in part II, we decompose

$$r_{NT}(\widetilde{\delta}^J - \delta^0_{NT}) = r_{NT}(\delta - \delta^0_{NT}) + \frac{r_{NT}}{\sqrt{NT}}\sqrt{NT}(\widetilde{\delta}^J - \delta).$$

Then, by an argument similar to part III of the proof of Theorem 4.3,

$$\sqrt{NT}(\widetilde{\delta}^J - \delta) \to_d \mathcal{N}(0, \overline{V}^{\delta(1)}_\infty),$$

and $r_{NT}(\delta - \delta^0_{NT}) \to_d \mathcal{N}(0, \overline{V}^{\delta(2)}_\infty)$, where $\overline{V}^{\delta(1)}_\infty$ and $\overline{V}^{\delta(2)}_\infty$ are defined as in the proof of Theorem 4.2. The statement follows because $(\delta - \delta^0_{NT})$ and $(\widetilde{\delta}^J - \delta)$ are asymptotically independent, and $\overline{V}^\delta_\infty = \overline{V}^{\delta(2)} + \overline{V}^{\delta(1)} \lim_{N,T\to\infty}(r_{NT}/\sqrt{NT})^2$. ■

S.3 Proofs of Appendix B (Asymptotic Expansions)

The following Lemma contains some statements that are not explicitly assumed in Assumptions B.1, but that are implied by it.

Lemma S.1. *Let Assumptions B.1 be satisfied. Then* *(i)* $\mathcal{H}(\beta, \phi) > 0$ *for all* $\beta \in \mathcal{B}(r_\beta, \beta^0)$ *and* $\phi \in \mathcal{B}_q(r_\phi, \phi^0)$ *wpa1,*

$$\sup_{\beta \in \mathcal{B}(r_\beta,\beta^0)} \sup_{\phi \in \mathcal{B}_q(r_\phi,\phi^0)} \|\partial_{\beta\beta'}\mathcal{L}(\beta,\, \phi)\| = \mathcal{O}_P\left(\sqrt{NT}\right),$$

$$\sup_{\beta \in \mathcal{B}(r_\beta,\beta^0)} \sup_{\phi \in \mathcal{B}_q(r_\phi,\phi^0)} \|\partial_{\beta\phi'}\mathcal{L}(\beta,\, \phi)\|_q = \mathcal{O}_P\left((NT)^{1/(2q)}\right),$$

$$\sup_{\beta \in \mathcal{B}(r_\beta,\beta^0)} \sup_{\phi \in \mathcal{B}_q(r_\phi,\phi^0)} \|\partial_{\phi\phi\phi}\mathcal{L}(\beta,\, \phi)\|_q = \mathcal{O}_P\left((NT)^{\epsilon}\right),$$

$$\sup_{\beta \in \mathcal{B}(r_\beta,\beta^0)} \sup_{\phi \in \mathcal{B}_q(r_\phi,\phi^0)} \|\partial_{\beta\phi\phi}\mathcal{L}(\beta,\, \phi)\|_q = \mathcal{O}_P((NT)^{\epsilon}),$$

$$\sup_{\beta \in \mathcal{B}(r_\beta,\beta^0)} \sup_{\phi \in \mathcal{B}_q(r_\phi,\phi^0)} \left\|\mathcal{H}^{-1}(\beta, \phi)\right\|_q = \mathcal{O}_P(1).$$

(ii) Moreover, $\|\mathcal{S}\| = \mathcal{O}_P(1)$, $\left\|\mathcal{H}^{-1}\right\| = \mathcal{O}_P(1)$, $\left\|\overline{\mathcal{H}}^{-1}\right\| = \mathcal{O}_P(1)$, $\left\|\mathcal{H}^{-1} - \overline{\mathcal{H}}^{-1}\right\| = o_P\left((NT)^{-1/8}\right)$, $\left\|\mathcal{H}^{-1} - \left(\overline{\mathcal{H}}^{-1} - \overline{\mathcal{H}}^{-1}\widetilde{\mathcal{H}}\overline{\mathcal{H}}^{-1}\right)\right\| = o_P\left((NT)^{-1/4}\right)$, $\|\partial_{\beta\phi'}\mathcal{L}\| = \mathcal{O}_P\left((NT)^{1/4}\right)$, $\|\partial_{\beta\phi\phi}\mathcal{L}\| = \mathcal{O}_P\left((NT)^{\epsilon}\right)$, $\left\|\sum_g \partial_{\phi\phi'\phi_g}\mathcal{L}\,[\mathcal{H}^{-1}\mathcal{S}]_g\right\| = \mathcal{O}_P\left((NT)^{-1/4+1/(2q)+\epsilon}\right)$, *and* $\left\|\sum_g \partial_{\phi\phi'\phi_g}\mathcal{L}\,[\overline{\mathcal{H}}^{-1}\mathcal{S}]_g\right\| = \mathcal{O}_P\left((NT)^{-1/4+1/(2q)+\epsilon}\right)$.

Proof of Lemma S.1. # Part (i): Let $v \in \mathbb{R}^{\dim\beta}$ and $w, u \in \mathbb{R}^{\dim\phi}$. By a Taylor expansion of $\partial_{\beta\phi'\phi_g}\mathcal{L}(\beta,\, \phi)$ around (β^0, ϕ^0)

$$\begin{aligned}
&\sum_g u_g v' \left[\partial_{\beta\phi'\phi_g}\mathcal{L}(\beta,\, \phi)\right] w \\
&= \sum_g u_g v' \left[\partial_{\beta\phi'\phi_g}\mathcal{L} + \sum_k (\beta_k - \beta_k^0)\partial_{\beta_k\beta\phi'\phi_g}\mathcal{L}(\tilde\beta, \tilde\phi) - \sum_h (\phi_h - \phi_h^0)\partial_{\beta\phi'\phi_g\phi_h}\mathcal{L}(\tilde\beta, \tilde\phi)\right] w,
\end{aligned}$$

with $(\tilde\beta, \tilde\phi)$ between (β^0, ϕ^0) and (β, ϕ). Thus

$$\begin{aligned}
\|\partial_{\beta\phi\phi}\mathcal{L}(\beta,\, \phi)\|_q &= \sup_{\|v\|=1} \sup_{\|u\|_q=1} \sup_{\|w\|_{q/(q-1)}=1} \sum_g u_g v' \left[\partial_{\beta\phi'\phi_g}\mathcal{L}(\beta,\, \phi)\right] w \\
&\le \|\partial_{\beta\phi\phi}\mathcal{L}\|_q + \|\beta - \beta^0\| \sup_{(\tilde\beta,\tilde\phi)} \left\|\partial_{\beta\beta\phi\phi}\mathcal{L}(\tilde\beta, \tilde\phi)\right\|_q + \|\phi - \phi^0\|_q \sup_{(\tilde\beta,\tilde\phi)} \left\|\partial_{\beta\phi\phi\phi}\mathcal{L}(\tilde\beta, \tilde\phi)\right\|_q,
\end{aligned}$$

where the supremum over $(\tilde\beta, \tilde\phi)$ is necessary, because those parameters depend on v, w, u. By Assumption B.1, for large enough N and T,

$$\begin{aligned}
\sup_{\beta \in \mathcal{B}(r_\beta,\beta^0)} \sup_{\phi \in \mathcal{B}_q(r_\phi,\phi^0)} \|\partial_{\beta\phi\phi}\mathcal{L}(\beta,\, \phi)\|_q &\le \|\partial_{\beta\phi\phi}\mathcal{L}\| + r_\beta \sup_{\beta \in \mathcal{B}(r_\beta,\beta^0)} \sup_{\phi \in \mathcal{B}_q(r_\phi,\phi^0)} \|\partial_{\beta\beta\phi\phi}\mathcal{L}(\beta, \phi)\|_q \\
&\quad + r_\phi \sup_{\beta \in \mathcal{B}(r_\beta,\beta^0)} \sup_{\phi \in \mathcal{B}_q(r_\phi,\phi^0)} \|\partial_{\beta\phi\phi\phi}\mathcal{L}(\beta, \phi)\|_q \\
&= \mathcal{O}_P\left[(NT)^{\epsilon} + r_\beta(NT)^{\epsilon} + r_\phi(NT)^{\epsilon}\right] = \mathcal{O}_P\left((NT)^{\epsilon}\right).
\end{aligned}$$

The proofs for the bounds on $\|\partial_{\beta\beta'}\mathcal{L}(\beta,\, \phi)\|$, $\|\partial_{\beta\phi'}\mathcal{L}(\beta,\, \phi)\|_q$ and $\|\partial_{\phi\phi\phi}\mathcal{L}(\beta,\, \phi)\|_q$ are analogous.

Next, we show that $\mathcal{H}(\beta, \phi)$ is non-singular for all $\beta \in \mathcal{B}(r_\beta, \beta^0)$ and $\phi \in \mathcal{B}_q(r_\phi, \phi^0)$ wpa1. By a Taylor expansion and Assumption B.1, for large enough N and T,

$$\begin{aligned}
\sup_{\beta \in \mathcal{B}(r_\beta,\beta^0)} \sup_{\phi \in \mathcal{B}_q(r_\phi,\phi^0)} \|\mathcal{H}(\beta, \phi) - \mathcal{H}\|_q &\le r_\beta \sup_{\beta \in \mathcal{B}(r_\beta,\beta^0)} \sup_{\phi \in \mathcal{B}_q(r_\phi,\phi^0)} \|\partial_{\beta\phi\phi}\mathcal{L}(\beta, \phi)\|_q \\
&\quad + r_\phi \sup_{\beta \in \mathcal{B}(r_\beta,\beta^0)} \sup_{\phi \in \mathcal{B}_q(r_\phi,\phi^0)} \|\partial_{\phi\phi\phi}\mathcal{L}(\beta, \phi)\|_q = o_P(1). \qquad \text{(S.1)}
\end{aligned}$$

Define $\Delta\mathcal{H}(\beta,\phi) = \overline{\mathcal{H}} - \mathcal{H}(\beta,\phi)$. Then $\|\Delta\mathcal{H}(\beta,\phi)\|_q \leq \|\mathcal{H}(\beta,\phi) - \mathcal{H}\|_q + \left\|\widetilde{\mathcal{H}}\right\|_q$, and therefore

$$\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)} \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)} \|\Delta\mathcal{H}(\beta,\phi)\|_q = o_P(1),$$

by Assumption B.1 and equation (S.1).

For any square matrix with $\|A\|_q < 1$, $\left\|(\mathbb{1} - A)^{-1}\right\|_q \leq (1 - \|A\|_q)^{-1}$, see e.g. p.301 in Horn and Johnson (1985). Then

$$\begin{aligned}
\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)} \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)} \left\|\mathcal{H}^{-1}(\beta,\phi)\right\|_q &= \sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)} \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)} \left\|\left(\overline{\mathcal{H}} - \Delta\mathcal{H}(\beta,\phi)\right)^{-1}\right\|_q \\
&= \sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)} \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)} \left\|\overline{\mathcal{H}}^{-1}\left(\mathbb{1} - \Delta\mathcal{H}(\beta,\phi)\overline{\mathcal{H}}^{-1}\right)^{-1}\right\|_q \\
&\leq \left\|\overline{\mathcal{H}}^{-1}\right\|_q \sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)} \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)} \left\|\left(\mathbb{1} - \Delta\mathcal{H}(\beta,\phi)\overline{\mathcal{H}}^{-1}\right)^{-1}\right\|_q \\
&\leq \left\|\overline{\mathcal{H}}^{-1}\right\|_q \sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)} \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)} \left(1 - \left\|\Delta\mathcal{H}(\beta,\phi)\overline{\mathcal{H}}^{-1}\right\|_q\right)^{-1} \\
&\leq \left\|\overline{\mathcal{H}}^{-1}\right\|_q (1 - o_P(1))^{-1} = \mathcal{O}_P(1).
\end{aligned}$$

#Part (ii): By the properties of the ℓ_q-norm and Assumption B.1(v),

$$\|\mathcal{S}\| = \|\mathcal{S}\|_2 \leq (\dim\phi)^{1/2-1/q}\|\mathcal{S}\|_q = \mathcal{O}_p(1).$$

Analogously,

$$\|\partial_{\beta\phi'}\mathcal{L}\| \leq (\dim\phi)^{1/2-1/q}\,\|\partial_{\beta\phi'}\mathcal{L}\|_q = \mathcal{O}_P\left((NT)^{1/4}\right).$$

By Lemma S.4, $\|\overline{\mathcal{H}}^{-1}\|_{q/(q-1)} = \|\overline{\mathcal{H}}^{-1}\|_q$ because $\overline{\mathcal{H}}^{-1}$ is symmetric, and

$$\left\|\overline{\mathcal{H}}^{-1}\right\| = \left\|\overline{\mathcal{H}}^{-1}\right\|_2 \leq \sqrt{\|\overline{\mathcal{H}}^{-1}\|_{q/(q-1)}\|\overline{\mathcal{H}}^{-1}\|_q} = \|\overline{\mathcal{H}}^{-1}\|_q = \mathcal{O}_P(1). \tag{S.2}$$

Analogously,

$$\begin{aligned}
\|\partial_{\beta\phi\phi}\mathcal{L}\| &\leq \|\partial_{\beta\phi\phi}\mathcal{L}\|_q = \mathcal{O}_P\left((NT)^{\epsilon}\right), \\
\left\|\sum_g \partial_{\phi\phi'\phi_g}\mathcal{L}\,[\mathcal{H}^{-1}\mathcal{S}]_g\right\| &\leq \left\|\sum_g \partial_{\phi\phi'\phi_g}\mathcal{L}\,[\mathcal{H}^{-1}\mathcal{S}]_g\right\|_q \\
&\leq \|\partial_{\phi\phi\phi}\mathcal{L}\|_q \left\|\mathcal{H}^{-1}\right\|_q \|\mathcal{S}\|_q = \mathcal{O}_P\left((NT)^{-1/4+1/(2q)+\epsilon}\right), \\
\left\|\sum_g \partial_{\phi\phi'\phi_g}\mathcal{L}\,[\overline{\mathcal{H}}^{-1}\mathcal{S}]_g\right\| &\leq \left\|\sum_g \partial_{\phi\phi'\phi_g}\mathcal{L}\,[\overline{\mathcal{H}}^{-1}\mathcal{S}]_g\right\|_q \\
&\leq \|\partial_{\phi\phi\phi}\mathcal{L}\|_q \left\|\overline{\mathcal{H}}^{-1}\right\|_q \|\mathcal{S}\|_q = \mathcal{O}_P\left((NT)^{-1/4+1/(2q)+\epsilon}\right).
\end{aligned}$$

Assumption B.1 guarantees that $\left\|\overline{\mathcal{H}}^{-1}\right\| \left\|\widetilde{\mathcal{H}}\right\| < 1$ wpa1. Therefore,

$$\mathcal{H}^{-1} = \overline{\mathcal{H}}^{-1}\left(\mathbb{1} + \widetilde{\mathcal{H}}\overline{\mathcal{H}}^{-1}\right)^{-1} = \overline{\mathcal{H}}^{-1}\sum_{s=0}^{\infty}(-\widetilde{\mathcal{H}}\overline{\mathcal{H}}^{-1})^s = \overline{\mathcal{H}}^{-1} - \overline{\mathcal{H}}^{-1}\widetilde{\mathcal{H}}\overline{\mathcal{H}}^{-1} + \overline{\mathcal{H}}^{-1}\sum_{s=2}^{\infty}(-\widetilde{\mathcal{H}}\overline{\mathcal{H}}^{-1})^s.$$

Note that $\left\|\overline{\mathcal{H}}^{-1}\sum_{s=2}^{\infty}(-\widetilde{\mathcal{H}}\overline{\mathcal{H}}^{-1})^s\right\| \leq \left\|\overline{\mathcal{H}}^{-1}\right\|\sum_{s=2}^{\infty}\left(\left\|\overline{\mathcal{H}}^{-1}\right\|\left\|\widetilde{\mathcal{H}}\right\|\right)^s$, and therefore

$$\left\|\mathcal{H}^{-1} - \left(\overline{\mathcal{H}}^{-1} - \overline{\mathcal{H}}^{-1}\widetilde{\mathcal{H}}\overline{\mathcal{H}}^{-1}\right)\right\| \leq \frac{\left\|\overline{\mathcal{H}}^{-1}\right\|^3\left\|\widetilde{\mathcal{H}}\right\|^2}{1-\left\|\overline{\mathcal{H}}^{-1}\right\|\left\|\widetilde{\mathcal{H}}\right\|} = o_P\left((NT)^{-1/4}\right),$$

by Assumption B.1(vi) and equation (S.2).

The results for $\left\|\mathcal{H}^{-1}\right\|$ and $\left\|\mathcal{H}^{-1} - \overline{\mathcal{H}}^{-1}\right\|$ follow immediately. ■

S.3.1 Legendre Transformed Objective Function

We consider the shrinking neighborhood $\mathcal{B}(r_\beta,\beta^0)\times\mathcal{B}_q(r_\phi,\phi^0)$ of the true parameters (β^0,ϕ^0). Statement (i) of Lemma S.1 implies that the objective function $\mathcal{L}(\beta,\,\phi)$ is strictly concave in ϕ in this shrinking neighborhood wpa1. We define

$$\mathcal{L}^*(\beta,\,S) = \max_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)}\left[\mathcal{L}(\beta,\,\phi)-\phi' S\right], \qquad \Phi(\beta,\,S) = \operatorname*{argmax}_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)}\left[\mathcal{L}(\beta,\,\phi)-\phi' S\right], \tag{S.3}$$

where $\beta\in\mathcal{B}(r_\beta,\beta^0)$ and $S\in\mathbb{R}^{\dim\phi}$. The function $\mathcal{L}^*(\beta,\,S)$ is the Legendre transformation of the objective function $\mathcal{L}(\beta,\,\phi)$ in the incidental parameter ϕ. We denote the parameter S as the dual parameter to ϕ, and $\mathcal{L}^*(\beta,\,S)$ as the dual function to $\mathcal{L}(\beta,\,\phi)$. We only consider $\mathcal{L}^*(\beta,\,S)$ and $\Phi(\beta,\,S)$ for parameters $\beta\in\mathcal{B}(r_\beta,\beta^0)$ and $S\in\mathcal{S}(\beta,\mathcal{B}_q(r_\phi,\phi^0))$, where the optimal ϕ is defined by the first order conditions, i.e. is not a boundary solution. We define the corresponding set of pairs $(\beta,\mathcal{S})$ that is dual to $\mathcal{B}(r_\beta,\beta^0)\times\mathcal{B}_q(r_\phi,\phi^0)$ by

$$\mathcal{SB}_r(\beta^0,\phi^0) = \left\{(\beta,\mathcal{S})\in\mathbb{R}^{\dim\beta+\dim\phi}\;:\;(\beta,\Phi(\beta,\,S))\in\mathcal{B}(r_\beta,\beta^0)\times\mathcal{B}_q(r_\phi,\phi^0)\right\}.$$

Assumption B.1 guarantees that for $\beta\in\mathcal{B}(r_\beta,\beta^0)$ the domain $\mathcal{S}(\beta,\mathcal{B}_q(r_\phi,\phi^0))$ includes $S=0$, the origin of $\mathbb{R}^{\dim\phi}$, as an interior point, wpa1, and that $\mathcal{L}^*(\beta,\,S)$ is four times differentiable in a neighborhood of $S=0$ (see Lemma S.2 below). The optimal $\phi=\Phi(\beta,S)$ in equation (S.3) satisfies the first order condition $S=\mathcal{S}(\beta,\phi)$. Thus, for given β, the functions $\Phi(\beta,S)$ and $\mathcal{S}(\beta,\phi)$ are inverse to each other, and the relationship between ϕ and its dual S is one-to-one. This is a consequence of strict concavity of $\mathcal{L}(\beta,\,\phi)$ in the neighborhood of the true parameter value that we consider here.[4] One can show that

$$\Phi(\beta,S) = -\frac{\partial\mathcal{L}^*(\beta,\,S)}{\partial S},$$

which shows the dual nature of the functions $\mathcal{L}(\beta,\,\phi)$ and $\mathcal{L}^*(\beta,\,S)$. For $S=0$ the optimization in (S.3) is just over the objective function $\mathcal{L}(\beta,\phi)$, so that $\Phi(\beta,0)=\widehat{\phi}(\beta)$ and $\mathcal{L}^*(\beta,0)=\mathcal{L}(\beta,\widehat{\phi}(\beta))$, the profile objective function. We already introduced $\mathcal{S}=\mathcal{S}(\beta^0,\phi^0)$, i.e. at $\beta=\beta^0$ the dual of ϕ^0 is $\mathcal{S}$, and vica

[4] Another consequence of strict concavity of $\mathcal{L}(\beta,\,\phi)$ is that the dual function $\mathcal{L}^*(\beta,\,S)$ is strictly convex in S. The original $\mathcal{L}(\beta,\,\phi)$ can be recovered from $\mathcal{L}^*(\beta,\,S)$ by again performing a Legendre transformation, namely

$$\mathcal{L}(\beta,\,\phi) = \min_{S\in\mathbb{R}^{\dim\phi}}\left[\mathcal{L}^*(\beta,\,S)+\phi' S\right].$$

versa. We can write the profile objective function $\mathcal{L}(\beta, \widehat{\phi}(\beta)) = \mathcal{L}^*(\beta, 0)$ as a Taylor series expansion of $\mathcal{L}^*(\beta, S)$ around $(\beta, S) = (\beta^0, \mathcal{S})$, namely

$$\mathcal{L}(\beta, \widehat{\phi}(\beta)) = \mathcal{L}^*(\beta^0, \mathcal{S}) + (\partial_{\beta'}\mathcal{L}^*)\Delta\beta - \Delta\beta'(\partial_{\beta S'}\mathcal{L}^*)\mathcal{S} + \frac{1}{2}\Delta\beta'(\partial_{\beta\beta'}\mathcal{L}^*)\Delta\beta + \dots ,$$

where $\Delta\beta = \beta - \beta^0$, and here and in the following we omit the arguments of $\mathcal{L}^*(\beta, S)$ and of its partial derivatives when they are evaluated at $(\beta^0, \mathcal{S})$. Analogously, we can obtain Taylor expansions for the profile score $\partial_\beta \mathcal{L}(\beta, \widehat{\phi}(\beta)) = \partial_\beta \mathcal{L}^*(\beta, 0)$ and the estimated nuisance parameter $\widehat{\phi}(\beta) = -\partial_S \mathcal{L}^*(\beta, 0)$ in $\Delta\beta$ and $\mathcal{S}$, see the proof of Theorem B.1 below. Apart from combinatorial factors those expansions feature the same coefficients as the expansion of $\mathcal{L}(\beta, \widehat{\phi}(\beta))$ itself. They are standard Taylor expansions that can be truncated at a certain order, and the remainder term can be bounded by applying the mean value theorem.

The functions $\mathcal{L}(\beta, \phi)$ and its dual $\mathcal{L}^*(\beta, S)$ are closely related. In particular, for given β their first derivatives with respect to the second argument $\mathcal{S}(\beta, \phi)$ and $\Phi(\beta, S)$ are inverse functions of each other. We can therefore express partial derivatives of $\mathcal{L}^*(\beta, S)$ in terms of partial derivatives of $\mathcal{L}(\beta, \phi)$. This is done in Lemma S.2. The norms $\|\partial_{\beta SSS}\mathcal{L}^*(\beta, S)\|_q$, $\|\partial_{SSSS}\mathcal{L}^*(\beta, S)\|_q$, etc., are defined as in equation (A.1) and (A.2).

Lemma S.2. *Let assumption B.1 be satisfied.*

(i) The function $\mathcal{L}^*(\beta, S)$ is well-defined and is four times continuously differentiable in $\mathcal{SB}_r(\beta^0, \phi^0)$, wpa1.

(ii) For $\mathcal{L}^* = \mathcal{L}^*(\beta^0, \mathcal{S})$,

$$\partial_S \mathcal{L}^* = -\phi^0, \quad \partial_\beta \mathcal{L}^* = \partial_\beta \mathcal{L}, \quad \partial_{SS'}\mathcal{L}^* = -(\partial_{\phi\phi'}\mathcal{L})^{-1} = \mathcal{H}^{-1}, \quad \partial_{\beta S'}\mathcal{L}^* = -(\partial_{\beta\phi'}\mathcal{L})\mathcal{H}^{-1},$$

$$\partial_{\beta\beta'}\mathcal{L}^* = \partial_{\beta\beta'}\mathcal{L} + (\partial_{\beta\phi'}\mathcal{L})\mathcal{H}^{-1}(\partial_{\phi'\beta}\mathcal{L}), \quad \partial_{SS'S_g}\mathcal{L}^* = -\sum_h \mathcal{H}^{-1}(\partial_{\phi\phi'\phi_h}\mathcal{L})\mathcal{H}^{-1}(\mathcal{H}^{-1})_{gh},$$

$$\partial_{\beta_k SS'}\mathcal{L}^* = \mathcal{H}^{-1}(\partial_{\beta_k\phi'\phi}\mathcal{L})\mathcal{H}^{-1} + \sum_g \mathcal{H}^{-1}(\partial_{\phi_g\phi'\phi}\mathcal{L})\mathcal{H}^{-1}[\mathcal{H}^{-1}\partial_{\beta_k\phi}\mathcal{L}]_g,$$

$$\begin{aligned}\partial_{\beta_k\beta_l S'}\mathcal{L}^* = &-(\partial_{\beta_k\beta_l\phi'}\mathcal{L})\mathcal{H}^{-1} - (\partial_{\beta_l\phi'}\mathcal{L})\mathcal{H}^{-1}(\partial_{\beta_k\phi\phi'}\mathcal{L})\mathcal{H}^{-1} - (\partial_{\beta_k\phi'}\mathcal{L})\mathcal{H}^{-1}(\partial_{\beta_l\phi'\phi}\mathcal{L})\mathcal{H}^{-1} \\ &- \sum_g (\partial_{\beta_k\phi'}\mathcal{L})\mathcal{H}^{-1}(\partial_{\phi_g\phi'\phi}\mathcal{L})\mathcal{H}^{-1}[\mathcal{H}^{-1}\partial_{\beta_l\phi}\mathcal{L}]_g,\end{aligned}$$

$$\begin{aligned}\partial_{\beta_k\beta_l\beta_m}\mathcal{L}^* = &\ \partial_{\beta_k\beta_l\beta_m}\mathcal{L} + \sum_g (\partial_{\beta_k\phi'}\mathcal{L})\mathcal{H}^{-1}(\partial_{\phi_g\phi'\phi}\mathcal{L})\mathcal{H}^{-1}(\partial_{\beta_l\phi}\mathcal{L})[\mathcal{H}^{-1}\partial_{\phi\beta_m}\mathcal{L}]_g \\ &+ (\partial_{\beta_k\phi'}\mathcal{L})\mathcal{H}^{-1}(\partial_{\beta_l\phi'\phi}\mathcal{L})\mathcal{H}^{-1}\partial_{\phi\beta_m}\mathcal{L} + (\partial_{\beta_m\phi'}\mathcal{L})\mathcal{H}^{-1}(\partial_{\beta_k\phi'\phi}\mathcal{L})\mathcal{H}^{-1}\partial_{\phi\beta_l}\mathcal{L} \\ &+ (\partial_{\beta_l\phi'}\mathcal{L})\mathcal{H}^{-1}(\partial_{\beta_m\phi'\phi}\mathcal{L})\mathcal{H}^{-1}\partial_{\phi\beta_k}\mathcal{L} \\ &+ (\partial_{\beta_k\beta_l\phi'}\mathcal{L})\mathcal{H}^{-1}(\partial_{\phi'\beta_m}\mathcal{L}) + (\partial_{\beta_k\beta_m\phi'}\mathcal{L})\mathcal{H}^{-1}(\partial_{\phi'\beta_l}\mathcal{L}) + (\partial_{\beta_l\beta_m\phi'}\mathcal{L})\mathcal{H}^{-1}(\partial_{\phi'\beta_k}\mathcal{L}),\end{aligned}$$

and

$$
\begin{aligned}
\partial_{SS'S_gS_h}\mathcal{L}^* &= \sum_{f,e}\mathcal{H}^{-1}(\partial_{\phi\phi'\phi_f\phi_e}\mathcal{L})\mathcal{H}^{-1}(\mathcal{H}^{-1})_{gf}(\mathcal{H}^{-1})_{he} \\
&\quad + 3\sum_{f,e}\mathcal{H}^{-1}(\partial_{\phi\phi'\phi_e}\mathcal{L})\mathcal{H}^{-1}(\partial_{\phi\phi'\phi_f}\mathcal{L})\mathcal{H}^{-1}(\mathcal{H}^{-1})_{gf}(\mathcal{H}^{-1})_{he}, \\
\partial_{\beta_k SS'S_g}\mathcal{L}^* &= -\sum_{h}\mathcal{H}^{-1}(\partial_{\beta_k\phi'\phi}\mathcal{L})\mathcal{H}^{-1}(\partial_{\phi\phi'\phi_h}\mathcal{L})\mathcal{H}^{-1}[\mathcal{H}^{-1}]_{gh} \\
&\quad - \sum_{h}\mathcal{H}^{-1}(\partial_{\phi\phi'\phi_h}\mathcal{L})\mathcal{H}^{-1}(\partial_{\beta_k\phi'\phi}\mathcal{L})\mathcal{H}^{-1}[\mathcal{H}^{-1}]_{gh} \\
&\quad - \sum_{h}\mathcal{H}^{-1}(\partial_{\phi\phi'\phi_h}\mathcal{L})\mathcal{H}^{-1}[\mathcal{H}^{-1}(\partial_{\beta_k\phi'\phi}\mathcal{L})\mathcal{H}^{-1}]_{gh} \\
&\quad - \sum_{h,f}\mathcal{H}^{-1}(\partial_{\phi_f\phi'\phi}\mathcal{L})\mathcal{H}^{-1}(\partial_{\phi\phi'\phi_h}\mathcal{L})\mathcal{H}^{-1}[\mathcal{H}^{-1}]_{gh}[\mathcal{H}^{-1}\partial_{\beta_k\phi}\mathcal{L}]_f \\
&\quad - \sum_{h,f}\mathcal{H}^{-1}(\partial_{\phi\phi'\phi_h}\mathcal{L})\mathcal{H}^{-1}(\partial_{\phi_f\phi'\phi}\mathcal{L})\mathcal{H}^{-1}[\mathcal{H}^{-1}]_{gh}[\mathcal{H}^{-1}\partial_{\beta_k\phi}\mathcal{L}]_f \\
&\quad - \sum_{h,f}\mathcal{H}^{-1}(\partial_{\phi\phi'\phi_h}\mathcal{L})\mathcal{H}^{-1}[\mathcal{H}^{-1}(\partial_{\phi_f\phi'\phi}\mathcal{L})\mathcal{H}^{-1}]_{gh}[\mathcal{H}^{-1}\partial_{\beta_k\phi}\mathcal{L}]_f \\
&\quad - \sum_{h}\mathcal{H}^{-1}(\partial_{\beta_k\phi\phi'\phi_h}\mathcal{L})\mathcal{H}^{-1}[\mathcal{H}^{-1}]_{gh} \\
&\quad - \sum_{h,f}\mathcal{H}^{-1}(\partial_{\phi\phi'\phi_h\phi_f}\mathcal{L})\mathcal{H}^{-1}[\mathcal{H}^{-1}]_{gh}[\mathcal{H}^{-1}(\partial_{\beta_k\phi}\mathcal{L})]_f.
\end{aligned}
$$

(iii) Moreover,

$$
\begin{aligned}
\sup_{(\beta,S)\in\mathcal{SB}_r(\beta^0,\phi^0)}\|\partial_{\beta\beta\beta}\mathcal{L}^*(\beta,S)\| &= \mathcal{O}_P\left((NT)^{1/2+1/(2q)+\epsilon}\right), \\
\sup_{(\beta,S)\in\mathcal{SB}_r(\beta^0,\phi^0)}\|\partial_{\beta\beta S}\mathcal{L}^*(\beta,S)\|_q &= \mathcal{O}_P\left((NT)^{1/q+\epsilon}\right), \\
\sup_{(\beta,S)\in\mathcal{SB}_r(\beta^0,\phi^0)}\|\partial_{\beta SS}\mathcal{L}^*(\beta,S)\|_q &= \mathcal{O}_P\left((NT)^{1/(2q)+\epsilon}\right), \\
\sup_{(\beta,S)\in\mathcal{SB}_r(\beta^0,\phi^0)}\|\partial_{\beta SSS}\mathcal{L}^*(\beta,S)\|_q &= \mathcal{O}_P\left((NT)^{1/(2q)+2\epsilon}\right), \\
\sup_{(\beta,S)\in\mathcal{SB}_r(\beta^0,\phi^0)}\|\partial_{SSSS}\mathcal{L}^*(\beta,S)\|_q &= \mathcal{O}_P\left((NT)^{2\epsilon}\right).
\end{aligned}
$$

Proof of Lemma S.2. #Part (i): According to the definition (S.3), $\mathcal{L}^*(\beta,S) = \mathcal{L}(\beta,\Phi(\beta,S)) - \Phi(\beta,S)'S$, where $\Phi(\beta,S)$ solves the FOC, $\mathcal{S}(\beta,\Phi(\beta,S)) = S$, i.e. $\mathcal{S}(\beta,.)$ and $\Phi(\beta,.)$ are inverse functions for every β. Taking the derivative of $\mathcal{S}(\beta,\Phi(\beta,S)) = S$ wrt to both S and β yields

$$
\begin{aligned}
[\partial_S\Phi(\beta,S)'][\partial_\phi\mathcal{S}(\beta,\Phi(\beta,S))'] &= \mathbb{1}, \\
[\partial_\beta\mathcal{S}(\beta,\Phi(\beta,S))'] + [\partial_\beta\Phi(\beta,S)'][\partial_\phi\mathcal{S}(\beta,\Phi(\beta,S))'] &= 0.
\end{aligned}
\tag{S.4}
$$

By definition, $\mathcal{S} = \mathcal{S}(\beta^0,\phi^0)$. Therefore, $\Phi(\beta,S)$ is the unique function that satisfies the boundary condition $\Phi(\beta^0,\mathcal{S}) = \phi^0$ and the system of partial differential equations (PDE) in (S.4). Those PDE's can equivalently be written as

$$\partial_S \Phi(\beta, S)' = -[\mathcal{H}(\beta, \Phi(\beta, S))]^{-1},$$
$$\partial_\beta \Phi(\beta, S)' = [\partial_{\beta\phi'} \mathcal{L}(\beta, \Phi(\beta, S))][\mathcal{H}(\beta, \Phi(\beta, S))]^{-1}. \tag{S.5}$$

This shows that $\Phi(\beta, S)$ (and thus $\mathcal{L}^*(\beta, S)$) are well-defined in any neighborhood of $(\beta, S) = (\beta^0, \mathcal{S})$ in which $\mathcal{H}(\beta, \Phi(\beta, S))$ is invertible (inverse function theorem). Lemma S.1 shows that $\mathcal{H}(\beta, \phi)$ is invertible in $\mathcal{B}(r_\beta, \beta^0) \times \mathcal{B}_q(r_\phi, \phi^0)$, wpa1. The inverse function theorem thus guarantee that $\Phi(\beta, S)$ and $\mathcal{L}^*(\beta, S)$ are well-defined in $\mathcal{SB}_r(\beta^0, \phi^0)$. The partial derivatives of $\mathcal{L}^*(\beta, S)$ of up to fourth order can be expressed as continuous transformations of the partial derivatives of $\mathcal{L}(\beta, \phi)$ up to fourth order (see e.g. proof of part (*ii*) of the lemma). Hence, $\mathcal{L}^*(\beta, S)$ is four times continuously differentiable because $\mathcal{L}(\beta, \phi)$ is four times continuously differentiable.

#Part (*ii*): Differentiating $\mathcal{L}^*(\beta, S) = \mathcal{L}(\beta, \Phi(\beta, S)) - \Phi(\beta, S)'S$ wrt β and S and using the FOC of the maximization over ϕ in the definition of $\mathcal{L}^*(\beta, S)$ gives $\partial_\beta \mathcal{L}^*(\beta, S) = \partial_\beta \mathcal{L}(\beta, \Phi(\beta, S))$ and $\partial_S \mathcal{L}^*(\beta, S) = -\Phi(\beta, S)$, respectively. Evaluating this expression at $(\beta, S) = (\beta^0, \mathcal{S})$ gives the first two statements of part (*ii*).

Using $\partial_S \mathcal{L}^*(\beta, S) = -\Phi(\beta, S)$, the PDE (S.5) can be written as

$$\partial_{SS'} \mathcal{L}^*(\beta, S) = \mathcal{H}^{-1}(\beta, \Phi(\beta, S)),$$
$$\partial_{\beta S'} \mathcal{L}^*(\beta, S) = -[\partial_{\beta\phi'} \mathcal{L}(\beta, \Phi(\beta, S))]\mathcal{H}^{-1}(\beta, \Phi(\beta, S)).$$

Evaluating this expression at $(\beta, S) = (\beta^0, \mathcal{S})$ gives the next two statements of part (*ii*).

Taking the derivative of $\partial_\beta \mathcal{L}^*(\beta, S) = \partial_\beta \mathcal{L}(\beta, \Phi(\beta, S))$ wrt to β and using the second equation of (S.5) gives the next statement when evaluated at $(\beta, S) = (\beta^0, \mathcal{S})$.

Taking the derivative of $\partial_{SS'} \mathcal{L}^*(\beta, S) = -[\partial_{\phi\phi'} \mathcal{L}(\beta, \Phi(\beta, S))]^{-1}$ wrt to S_g and using the first equation of (S.5) gives the next statement when evaluated at $(\beta, S) = (\beta^0, \mathcal{S})$.

Taking the derivative of $\partial_{SS'} \mathcal{L}^*(\beta, S) = -[\partial_{\phi\phi'} \mathcal{L}(\beta, \Phi(\beta, S))]^{-1}$ wrt to β_k and using the second equation of (S.5) gives

$$\begin{aligned}\partial_{\beta_k SS'} \mathcal{L}^*(\beta, S) &= \mathcal{H}^{-1}(\beta, \phi)[\partial_{\beta_k \phi' \phi} \mathcal{L}(\beta, \phi)]\mathcal{H}^{-1}(\beta, \phi) \\ &\quad + \sum_g \mathcal{H}^{-1}(\beta, \phi)[\partial_{\phi_g \phi' \phi} \mathcal{L}(\beta, \phi)]\mathcal{H}^{-1}(\beta, \phi)\{\mathcal{H}^{-1}(\beta, \phi)[\partial_{\beta_k \phi} \mathcal{L}(\beta, \phi)]\}_g,\end{aligned} \tag{S.6}$$

where $\phi = \Phi(\beta, S)$. This becomes the next statement when evaluated at $(\beta, S) = (\beta^0, \mathcal{S})$.

We omit the proofs for $\partial_{\beta_k \beta_l S'} \mathcal{L}^*$, $\partial_{\beta_k \beta_l S} \mathcal{L}^*$, $\partial_{SS' S_g S_h} \mathcal{L}^*$ and $\partial_{\beta_k SS' S_g} \mathcal{L}^*$ because they are analogous.

#Part (*iii*): We only show the result for $\|\partial_{\beta SS} \mathcal{L}^*(\beta, S)\|_q$, the proof of the other statements is analogous. By equation (S.6)

$$\|\partial_{\beta SS} \mathcal{L}^*(\beta, S)\|_q \leq \left\|\mathcal{H}^{-1}(\beta, \phi)\right\|_q^2 \|\partial_{\beta\phi\phi} \mathcal{L}(\beta, \phi)\|_q + \left\|\mathcal{H}^{-1}(\beta, \phi)\right\|_q^3 \|\partial_{\phi\phi\phi} \mathcal{L}(\beta, \phi)\|_q \|\partial_{\beta\phi'} \mathcal{L}(\beta, \phi)\|_q ,$$

where $\phi = \Phi(\beta, S)$. Then, by Lemma S.1

$$\begin{aligned}\sup_{(\beta,S)\in\mathcal{SB}_r(\beta^0,\phi^0)} \|\partial_{\beta SS} \mathcal{L}^*(\beta, S)\|_q &\leq \sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)} \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)} \Big[\left\|\mathcal{H}^{-1}(\beta, \phi)\right\|_q^2 \|\partial_{\beta\phi\phi} \mathcal{L}(\beta, \phi)\|_q \\ &\quad + \left\|\mathcal{H}^{-1}(\beta, \phi)\right\|_q^3 \|\partial_{\phi\phi\phi} \mathcal{L}(\beta, \phi)\|_q \|\partial_{\beta\phi'} \mathcal{L}(\beta, \phi)\|_q \Big] = \mathcal{O}\left((NT)^{1/(2q)+\epsilon}\right).\end{aligned}$$

To derive the rest of the bounds we can use that the expressions from part (ii) hold not only for $(\beta^0, \mathcal{S})$, but also for other values (β, S), provided that $(\beta, \Phi(\beta, S)$ is used as the argument on the rhs expressions. ∎

S.3.2 Proofs of Theorem B.1, Corollary B.2, and Theorem B.3

Proof of Theorem B.1, Part 1: Expansion of $\widehat{\phi}(\beta)$. Let $\beta = \beta_{NT} \in \mathcal{B}(\beta^0, r_\beta)$. A Taylor expansion of $\partial_S \mathcal{L}^*(\beta, 0)$ around $(\beta^0, \mathcal{S})$ gives

$$\widehat{\phi}(\beta) = -\partial_S \mathcal{L}^*(\beta, 0) = -\partial_S \mathcal{L}^* - (\partial_{S\beta'} \mathcal{L}^*)\Delta\beta + (\partial_{SS'} \mathcal{L}^*)\mathcal{S} - \frac{1}{2}\sum_g (\partial_{SS'S_g} \mathcal{L}^*)\mathcal{S}\mathcal{S}_g + R^\phi(\beta),$$

where we first expand in β holding $S = \mathcal{S}$ fixed, and then expand in S. For any $v \in \mathbb{R}^{\dim \phi}$ the remainder term satisfies

$$\begin{aligned} v' R^\phi(\beta) = v' \bigg\{ & -\frac{1}{2}\sum_k [\partial_{S\beta'\beta_k} \mathcal{L}^*(\tilde{\beta}, \mathcal{S})](\Delta\beta)(\Delta\beta_k) + \sum_k [\partial_{SS'\beta_k} \mathcal{L}^*(\beta^0, \tilde{S})]\mathcal{S}(\Delta\beta_k) \\ & + \frac{1}{6}\sum_{g,h} [\partial_{SS'S_gS_h} \mathcal{L}^*(\beta^0, \bar{S})]\mathcal{S}\mathcal{S}_g\mathcal{S}_h \bigg\}, \end{aligned}$$

where $\tilde{\beta}$ is between β^0 and β, and $\tilde{S}$ and $\bar{S}$ are between 0 and $\mathcal{S}$. By part (ii) of Lemma S.2,

$$\widehat{\phi}(\beta) - \phi^0 = \mathcal{H}^{-1}(\partial_{\phi\beta'} \mathcal{L})\Delta\beta + \mathcal{H}^{-1}\mathcal{S} + \tfrac{1}{2}\mathcal{H}^{-1}\sum_g (\partial_{\phi\phi'\phi_g} \mathcal{L})\mathcal{H}^{-1}\mathcal{S}(\mathcal{H}^{-1}\mathcal{S})_g + R^\phi(\beta).$$

Using that the vector norm $\|.\|_{q/(q-1)}$ is the dual to the vector norm $\|.\|_q$, Assumption B.1, and Lemmas S.1 and S.2 yields

$$\begin{aligned} \left\| R^\phi(\beta) \right\|_q &= \sup_{\|v\|_{q/(q-1)}=1} v' R^\phi(\beta) \\ &\le \frac{1}{2}\left\| \partial_{S\beta\beta} \mathcal{L}^*(\tilde{\beta}, \mathcal{S}) \right\|_q \|\Delta\beta\|^2 + \left\| \partial_{SS\beta} \mathcal{L}^*(\beta^0, \tilde{S}) \right\|_q \|\mathcal{S}\|_q \|\Delta\beta\| + \frac{1}{6}\left\| \partial_{SSSS} \mathcal{L}^*(\beta^0, \bar{S}) \right\|_q \|\mathcal{S}\|_q^3 \\ &= \mathcal{O}_P\left[(NT)^{1/q+\epsilon} r_\beta \|\Delta\beta\| + (NT)^{-1/4+1/q+\epsilon}\|\Delta\beta\| + (NT)^{-3/4+3/(2q)+2\epsilon} \right] \\ &= o_P\left((NT)^{-1/2+1/(2q)} \right) + o_P\left((NT)^{1/(2q)}\|\beta - \beta^0\| \right), \end{aligned}$$

uniformly over $\beta \in \mathcal{B}(\beta^0, r_\beta)$ by Lemma S.2. ∎

Proof of Theorem B.1, Part 2: Expansion of profile score. Let $\beta = \beta_{NT} \in \mathcal{B}(\beta^0, r_\beta)$. A Taylor expansion of $\partial_\beta \mathcal{L}^*(\beta, 0)$ around $(\beta^0, \mathcal{S})$ gives

$$\partial_\beta \mathcal{L}(\beta, \widehat{\phi}(\beta)) = \partial_\beta \mathcal{L}^*(\beta, 0) = \partial_\beta \mathcal{L}^* + (\partial_{\beta\beta'} \mathcal{L}^*)\Delta\beta - (\partial_{\beta S'} \mathcal{L}^*)\mathcal{S} + \frac{1}{2}\sum_g (\partial_{\beta S'S_g} \mathcal{L}^*)\mathcal{S}\mathcal{S}_g + R_1(\beta),$$

where we first expand in β for fixed $S = \mathcal{S}$, and then expand in S. For any $v \in \mathbb{R}^{\dim \beta}$ the remainder term satisfies

$$\begin{aligned} v' R_1(\beta) = v' \bigg\{ & \frac{1}{2}\sum_k [\partial_{\beta\beta'\beta_k} \mathcal{L}^*(\tilde{\beta}, \mathcal{S})](\Delta\beta)(\Delta\beta_k) - \sum_k [\partial_{\beta\beta_k S'} \mathcal{L}^*(\beta^0, \tilde{S})]\mathcal{S}(\Delta\beta_k) \\ & - \frac{1}{6}\sum_{g,h} [\partial_{\beta S'S_gS_h} \mathcal{L}^*(\beta^0, \bar{S})]\mathcal{S}\mathcal{S}_g\mathcal{S}_h \bigg\}, \end{aligned}$$

where $\tilde{\beta}$ is between β^0 and β, and $\tilde{S}$ and $\bar{S}$ are between 0 and $\mathcal{S}$. By Lemma S.2,

$$\begin{aligned}\partial_\beta \mathcal{L}(\beta, \widehat{\phi}(\beta)) &= \partial_\beta \mathcal{L} + \left[\partial_{\beta\beta'}\mathcal{L} + (\partial_{\beta\phi'}\mathcal{L})\mathcal{H}^{-1}(\partial_{\phi'\beta}\mathcal{L})\right](\beta - \beta^0) + (\partial_{\beta\phi'}\mathcal{L})\mathcal{H}^{-1}\mathcal{S} \\ &\quad + \frac{1}{2}\sum_g \left(\partial_{\beta\phi'\phi_g}\mathcal{L} + [\partial_{\beta\phi'}\mathcal{L}]\,\mathcal{H}^{-1}[\partial_{\phi\phi'\phi_g}\mathcal{L}]\right)[\mathcal{H}^{-1}\mathcal{S}]_g\mathcal{H}^{-1}\mathcal{S} + R_1(\beta),\end{aligned}$$

where for any $v \in \mathbb{R}^{\dim\beta}$,

$$\begin{aligned}\|R_1(\beta)\| &= \sup_{\|v\|=1} v' R_1(\beta) \\ &\leq \frac{1}{2}\left\|\partial_{\beta\beta\beta}\mathcal{L}^*(\tilde{\beta}, \mathcal{S})\right\|\|\Delta\beta\|^2 + (NT)^{1/2-1/q}\left\|\partial_{\beta\beta S}\mathcal{L}^*(\beta^0, \tilde{S})\right\|_q \|\mathcal{S}\|_q\|\Delta\beta\| \\ &\quad + \frac{1}{6}(NT)^{1/2-1/q}\left\|\partial_{\beta SSS}\mathcal{L}^*(\beta^0, \bar{S})\right\|_q\|\mathcal{S}\|_q^3 \\ &= \mathcal{O}_P\left[(NT)^{1/2+1/(2q)+\epsilon} r_\beta\|\Delta\beta\| + (NT)^{1/4+1/(2q)+\epsilon}\|\Delta\beta\| + (NT)^{-1/4+1/q+2\epsilon}\right] \\ &= o_P(1) + o_P(\sqrt{NT}\|\beta - \beta^0\|),\end{aligned}$$

uniformly over $\beta \in \mathcal{B}(\beta^0, r_\beta)$ by Lemma S.2. We can also write

$$\begin{aligned}d_\beta \mathcal{L}(\beta, \widehat{\phi}(\beta)) &= \partial_\beta \mathcal{L} - \sqrt{NT}\,\overline{W}(\Delta\beta) + (\partial_{\beta\phi'}\overline{\mathcal{L}})\overline{\mathcal{H}}^{-1}\mathcal{S} + (\partial_{\beta\phi'}\widetilde{\mathcal{L}})\overline{\mathcal{H}}^{-1}\mathcal{S} - (\partial_{\beta\phi'}\overline{\mathcal{L}})\overline{\mathcal{H}}^{-1}\widetilde{\mathcal{H}}\overline{\mathcal{H}}^{-1}\mathcal{S} \\ &\quad + \frac{1}{2}\sum_g \left(\partial_{\beta\phi'\phi_g}\overline{\mathcal{L}} + [\partial_{\beta\phi'}\overline{\mathcal{L}}]\,\overline{\mathcal{H}}^{-1}[\partial_{\phi\phi'\phi_g}\overline{\mathcal{L}}]\right)[\overline{\mathcal{H}}^{-1}\mathcal{S}]_g\overline{\mathcal{H}}^{-1}\mathcal{S} + R(\beta), \\ &= U - \sqrt{NT}\,\overline{W}(\Delta\beta) + R(\beta),\end{aligned}$$

where we decompose the term linear in $\mathcal{S}$ into multiple terms by using that

$$-(\partial_{\beta S'}\mathcal{L}^*) = (\partial_{\beta\phi'}\mathcal{L})\mathcal{H}^{-1} = \left[(\partial_{\beta\phi'}\overline{\mathcal{L}}) + (\partial_{\beta\phi'}\widetilde{\mathcal{L}})\right]\left[\overline{\mathcal{H}}^{-1} - \overline{\mathcal{H}}^{-1}\widetilde{\mathcal{H}}\overline{\mathcal{H}}^{-1} + \ldots\right].$$

The new remainder term is

$$\begin{aligned}R(\beta) &= R_1(\beta) + (\partial_{\beta\beta'}\widetilde{\mathcal{L}})\Delta\beta + \left[(\partial_{\beta\phi'}\mathcal{L})\mathcal{H}^{-1}(\partial_{\phi'\beta}\mathcal{L}) - (\partial_{\beta\phi'}\overline{\mathcal{L}})\overline{\mathcal{H}}^{-1}(\partial_{\phi'\beta}\overline{\mathcal{L}})\right]\Delta\beta \\ &\quad + (\partial_{\beta\phi'}\mathcal{L})\left[\mathcal{H}^{-1} - \left(\overline{\mathcal{H}}^{-1} - \overline{\mathcal{H}}^{-1}\widetilde{\mathcal{H}}\overline{\mathcal{H}}^{-1}\right)\right]\mathcal{S} - (\partial_{\beta\phi'}\widetilde{\mathcal{L}})\overline{\mathcal{H}}^{-1}\widetilde{\mathcal{H}}\overline{\mathcal{H}}^{-1}\mathcal{S} \\ &\quad + \frac{1}{2}\left[\sum_g \partial_{\beta\phi'\phi_g}\mathcal{L}[\mathcal{H}^{-1}\mathcal{S}]_g\mathcal{H}^{-1}\mathcal{S} - \sum_g \partial_{\beta\phi'\phi_g}\overline{\mathcal{L}}[\overline{\mathcal{H}}^{-1}\mathcal{S}]_g\overline{\mathcal{H}}^{-1}\mathcal{S}\right] \\ &\quad + \frac{1}{2}\left[\sum_g [\partial_{\beta\phi'}\mathcal{L}]\,\mathcal{H}^{-1}[\partial_{\phi\phi'\phi_g}\mathcal{L}][\mathcal{H}^{-1}\mathcal{S}]_g\mathcal{H}^{-1}\mathcal{S} - \sum_g [\partial_{\beta\phi'}\overline{\mathcal{L}}]\,\overline{\mathcal{H}}^{-1}[\partial_{\phi\phi'\phi_g}\overline{\mathcal{L}}][\overline{\mathcal{H}}^{-1}\mathcal{S}]_g\overline{\mathcal{H}}^{-1}\mathcal{S}\right].\end{aligned}$$

By Assumption B.1 and Lemma S.1,

$$
\begin{aligned}
\|R(\beta)\| &\leq \|R_1(\beta)\| + \left\|\partial_{\beta\beta'}\widetilde{\mathcal{L}}\right\| \|\Delta\beta\| + \|\partial_{\beta\phi'}\mathcal{L}\| \left\|\mathcal{H}^{-1} - \overline{\mathcal{H}}^{-1}\right\| \|\partial_{\phi'\beta}\mathcal{L}\| \|\Delta\beta\| \\
&\quad + \left\|\partial_{\beta\phi'}\widetilde{\mathcal{L}}\right\| \left\|\overline{\mathcal{H}}^{-1}\right\| \left(\|\partial_{\phi'\beta}\mathcal{L}\| + \|\partial_{\phi'\beta}\overline{\mathcal{L}}\|\right) \|\Delta\beta\| \\
&\quad + \|\partial_{\beta\phi'}\mathcal{L}\| \left\|\mathcal{H}^{-1} - \left(\overline{\mathcal{H}}^{-1} - \overline{\mathcal{H}}^{-1}\widetilde{\mathcal{H}}\overline{\mathcal{H}}^{-1}\right)\right\| \|\mathcal{S}\| + \left\|\overline{\mathcal{H}}^{-1}\right\|^2 \left\|\partial_{\beta\phi'}\widetilde{\mathcal{L}}\right\| \left\|\widetilde{\mathcal{H}}\right\| \|\mathcal{S}\| \\
&\quad + \frac{1}{2}\|\partial_{\beta\phi\phi}\mathcal{L}\| \left(\|\mathcal{H}^{-1}\| + \left\|\overline{\mathcal{H}}^{-1}\right\|\right) \left\|\mathcal{H}^{-1} - \overline{\mathcal{H}}^{-1}\right\| \|\mathcal{S}\|^2 \\
&\quad + \frac{1}{2}\left\|\overline{\mathcal{H}}^{-1}\right\|^2 \left\|\partial_{\beta\phi\phi}\widetilde{\mathcal{L}}\right\| \|\mathcal{S}\|^2 \\
&\quad + \frac{1}{2}\left\| \sum_g [\partial_{\beta\phi'}\mathcal{L}]\, \mathcal{H}^{-1} [\partial_{\phi\phi'\phi_g}\mathcal{L}][\mathcal{H}^{-1}\mathcal{S}]_g \mathcal{H}^{-1}\mathcal{S} - \sum_g [\partial_{\beta\phi'}\overline{\mathcal{L}}]\, \overline{\mathcal{H}}^{-1} [\partial_{\phi\phi'\phi_g}\overline{\mathcal{L}}][\overline{\mathcal{H}}^{-1}\mathcal{S}]_g \overline{\mathcal{H}}^{-1}\mathcal{S} \right\| \\
&= \|R_1(\beta)\| + o_P(1) + o_P(\sqrt{NT}\|\beta - \beta^0\|) + \mathcal{O}_P\left[(NT)^{-1/8+\epsilon+1/(2q)}\right] \\
&= o_P(1) + o_P(\sqrt{NT}\|\beta - \beta^0\|),
\end{aligned}
$$

uniformly over $\beta \in \mathcal{B}(\beta^0, r_\beta)$. Here we use that

$$
\begin{aligned}
&\left\| \sum_g [\partial_{\beta\phi'}\mathcal{L}]\, \mathcal{H}^{-1} [\partial_{\phi\phi'\phi_g}\mathcal{L}][\mathcal{H}^{-1}\mathcal{S}]_g \mathcal{H}^{-1}\mathcal{S} - \sum_g [\partial_{\beta\phi'}\overline{\mathcal{L}}]\, \overline{\mathcal{H}}^{-1} [\partial_{\phi\phi'\phi_g}\overline{\mathcal{L}}][\overline{\mathcal{H}}^{-1}\mathcal{S}]_g \overline{\mathcal{H}}^{-1}\mathcal{S} \right\| \\
&\leq \|\partial_{\beta\phi'}\mathcal{L}\| \left\|\mathcal{H}^{-1} - \overline{\mathcal{H}}^{-1}\right\| \left(\|\mathcal{H}^{-1}\| + \left\|\overline{\mathcal{H}}^{-1}\right\|\right) \|\mathcal{S}\| \left\| \sum_g \partial_{\phi\phi'\phi_g}\mathcal{L}\, [\mathcal{H}^{-1}\mathcal{S}]_g \right\| \\
&\quad + \|\partial_{\beta\phi'}\mathcal{L}\| \left\|\mathcal{H}^{-1} - \overline{\mathcal{H}}^{-1}\right\| \left\|\overline{\mathcal{H}}^{-1}\right\| \|\mathcal{S}\| \left\| \sum_g \partial_{\phi\phi'\phi_g}\mathcal{L}\, [\overline{\mathcal{H}}^{-1}\mathcal{S}]_g \right\| \\
&\quad + \left\|\partial_{\beta\phi'}\widetilde{\mathcal{L}}\right\| \left\|\overline{\mathcal{H}}^{-1}\right\|^2 \|\mathcal{S}\| \left\| \sum_g \partial_{\phi\phi'\phi_g}\mathcal{L}\, [\overline{\mathcal{H}}^{-1}\mathcal{S}]_g \right\| \\
&\quad + \|\partial_{\beta\phi'}\overline{\mathcal{L}}\| \left\|\overline{\mathcal{H}}^{-1}\right\| \left\| \sum_{g,h} \partial_{\phi\phi_g\phi_h}\widetilde{\mathcal{L}}\, [\overline{\mathcal{H}}^{-1}\mathcal{S}]_g [\overline{\mathcal{H}}^{-1}\mathcal{S}]_h \right\|.
\end{aligned}
$$

■

Proof of Corollary B.2. $\widehat{\beta}$ solves the FOC

$$\partial_\beta \mathcal{L}(\widehat{\beta}, \widehat{\phi}(\widehat{\beta})) = 0.$$

By $\left\|\widehat{\beta} - \beta^0\right\| = o_P(r_\beta)$ and Theorem B.1,

$$0 = \partial_\beta \mathcal{L}(\widehat{\beta}, \widehat{\phi}(\widehat{\beta})) = U - \overline{W}\,\sqrt{NT}(\widehat{\beta} - \beta^0) + o_P(1) + o_P(\sqrt{NT}\|\widehat{\beta} - \beta^0\|).$$

Thus, $\sqrt{NT}(\widehat{\beta} - \beta^0) = \overline{W}^{-1}U + o_P(1) + o_P(\sqrt{NT}\|\widehat{\beta} - \beta^0\|) = \overline{W}_\infty^{-1}U + o_P(1) + o_P(\sqrt{NT}\|\widehat{\beta} - \beta^0\|)$, where we use that $\overline{W} = \overline{W}_\infty + o_P(1)$ is invertible wpa1 and that $\overline{W}^{-1} = \overline{W}_\infty^{-1} + o_P(1)$. We conclude that $\sqrt{NT}(\widehat{\beta} - \beta^0) = \mathcal{O}_P(1)$ because $U = \mathcal{O}_P(1)$, and therefore $\sqrt{NT}(\widehat{\beta} - \beta^0) = \overline{W}_\infty^{-1}U + o_P(1)$. ■

Proof of Theorem B.3. # Consistency of $\widehat{\phi}(\beta)$: Let $\eta = \eta_{NT} > 0$ be such that $\eta = o_P(r_\phi)$, $(NT)^{-1/4+1/(2q)} = o_P(\eta)$, and $(NT)^{1/(2q)} r_\beta = o_P(\eta)$. For $\beta \in \mathcal{B}(r_\beta, \beta^0)$, define

$$\widehat{\phi}^*(\beta) := \underset{\{\phi:\ \|\phi - \phi^0\|_q \leq \eta\}}{\operatorname{argmin}} \|\mathcal{S}(\beta, \phi)\|_q. \tag{S.7}$$

Then, $\|\mathcal{S}(\beta, \widehat{\phi}^*(\beta))\|_q \leq \|\mathcal{S}(\beta, \phi^0)\|_q$, and therefore by a Taylor expansion of $\mathcal{S}(\beta, \phi^0)$ around $\beta = \beta^0$,

$$\begin{aligned}\|\mathcal{S}(\beta, \widehat{\phi}^*(\beta)) - \mathcal{S}(\beta, \phi^0)\|_q &\leq \|\mathcal{S}(\beta, \widehat{\phi}^*(\beta))\|_q + \|\mathcal{S}(\beta, \phi^0)\|_q \leq 2\|\mathcal{S}(\beta, \phi^0)\|_q \\ &\leq 2\|\mathcal{S}\|_q + 2\left\|\partial_{\phi\beta'}\mathcal{L}(\tilde{\beta}, \phi^0)\right\|_q \|\beta - \beta^0\| \\ &= \mathcal{O}_P\left[(NT)^{-1/4+1/(2q)} + (NT)^{1/(2q)}\|\beta - \beta^0\|\right],\end{aligned}$$

uniformly over $\beta \in \mathcal{B}(r_\beta, \beta^0)$, where $\tilde{\beta}$ is between β^0 and β, and we use Assumption B.1(v) and Lemma S.1. Thus,

$$\sup_{\beta \in \mathcal{B}(r_\beta, \beta^0)} \|\mathcal{S}(\beta, \widehat{\phi}^*(\beta)) - \mathcal{S}(\beta, \phi^0)\|_q = \mathcal{O}_P\left[(NT)^{-1/4+1/(2q)} + (NT)^{1/(2q)} r_\beta\right].$$

By a Taylor expansion of $\Phi(\beta, S)$ around $S = \mathcal{S}(\beta, \phi^0)$,

$$\begin{aligned}\left\|\widehat{\phi}^*(\beta) - \phi^0\right\|_q &= \left\|\Phi(\beta, \mathcal{S}(\beta, \widehat{\phi}^*(\beta))) - \Phi(\beta, \mathcal{S}(\beta, \phi^0))\right\|_q \leq \left\|\partial_S \Phi(\beta, \tilde{S})'\right\|_q \left\|\mathcal{S}(\beta, \widehat{\phi}^*(\beta)) - \mathcal{S}(\beta, \phi^0)\right\|_q \\ &= \left\|\mathcal{H}^{-1}(\beta, \Phi(\beta, \tilde{S}))\right\|_q \left\|\mathcal{S}(\beta, \widehat{\phi}^*(\beta)) - \mathcal{S}(\beta, \phi^0)\right\|_q = \mathcal{O}_P(1) \left\|\mathcal{S}(\beta, \widehat{\phi}^*(\beta)) - \mathcal{S}(\beta, \phi^0)\right\|_q,\end{aligned}$$

where $\tilde{S}$ is between $\mathcal{S}(\beta, \widehat{\phi}^*(\beta))$ and $\mathcal{S}(\beta, \phi^0)$ and we use Lemma S.1(i). Thus,

$$\sup_{\beta \in \mathcal{B}(r_\beta, \beta^0)} \left\|\widehat{\phi}^*(\beta) - \phi^0\right\|_q = \mathcal{O}_P\left[(NT)^{-1/4+1/(2q)} + (NT)^{1/(2q)} r_\beta\right] = o_P(\eta).$$

This shows that $\widehat{\phi}^*(\beta)$ is an interior solution of the minimization problem (S.7), wpa1. Thus, $\mathcal{S}(\beta, \widehat{\phi}^*(\beta)) = 0$, because the objective function $\mathcal{L}(\beta, \phi)$ is strictly concave and differentiable, and therefore $\widehat{\phi}^*(\beta) = \widehat{\phi}(\beta)$. We conclude that $\sup_{\beta \in \mathcal{B}(r_\beta, \beta^0)} \left\|\widehat{\phi}(\beta) - \phi^0\right\|_q = \mathcal{O}_P(\eta) = o_P(r_\phi)$.

Consistency of $\widehat{\beta}$: We have already shown that Assumption B.1(ii) is satisfied, in addition to the remaining parts of Assumption B.1, which we assume. The bounds on the spectral norm in Assumption B.1(vi) and in part (ii) of Lemma S.1 can be used to show that $U = \mathcal{O}_P((NT)^{1/4})$.

First, we consider the case $\dim(\beta) = 1$ first. The extension to $\dim(\beta) > 1$ is discussed below. Let $\eta = 2(NT)^{-1/2}\overline{W}^{-1}|U|$. Our goal is to show that $\widehat{\beta} \in [\beta^0 - \eta, \beta^0 + \eta]$. By Theorem B.1,

$$\begin{aligned}\partial_\beta \mathcal{L}(\beta^0 + \eta, \widehat{\phi}(\beta^0 + \eta)) &= U - \overline{W}\sqrt{NT}\eta + o_P(1) + o_P(\sqrt{NT}\eta) = o_P(\sqrt{NT}\eta) - \overline{W}\sqrt{NT}\eta, \\ \partial_\beta \mathcal{L}(\beta^0 - \eta, \widehat{\phi}(\beta^0 - \eta)) &= U + \overline{W}\sqrt{NT}\eta + o_P(1) + o_P(\sqrt{NT}\eta) = o_P(\sqrt{NT}\eta) + \overline{W}\sqrt{NT}\eta,\end{aligned}$$

and therefore for sufficiently large N, T

$$\partial_\beta \mathcal{L}(\beta^0 + \eta, \widehat{\phi}(\beta^0 + \eta)) \leq 0 \leq \partial_\beta \mathcal{L}(\beta^0 - \eta, \widehat{\phi}(\beta^0 - \eta)).$$

Thus, since $\partial_\beta \mathcal{L}(\widehat{\beta}, \widehat{\phi}(\widehat{\beta})) = 0$, for sufficiently large N, T,

$$\partial_\beta \mathcal{L}(\beta^0 + \eta, \widehat{\phi}(\beta^0 + \eta)) \leq \partial_\beta \mathcal{L}(\widehat{\beta}, \widehat{\phi}(\widehat{\beta})) \leq \partial_\beta \mathcal{L}(\beta^0 - \eta, \widehat{\phi}(\beta^0 - \eta)).$$

The profile objective $\mathcal{L}(\beta, \widehat{\phi}(\beta))$ is strictly concave in β because $\mathcal{L}(\beta, \phi)$ is strictly concave in (β, ϕ). Thus, $\partial_\beta \mathcal{L}(\beta, \widehat{\phi}(\beta))$ is strictly decreasing. The previous set of inequalities implies that for sufficiently large N, T

$$\beta^0 + \eta \geq \widehat{\beta} \geq \beta^0 - \eta.$$

We conclude that $\|\widehat{\beta}-\beta^0\| \leq \eta = \mathcal{O}_P((NT)^{-1/4})$. This concludes the proof for $\dim(\beta)=1$.

To generalize the proof to $\dim(\beta)>1$ we define $\beta_\pm = \beta^0 \pm \eta \frac{\widehat{\beta}-\beta^0}{\|\widehat{\beta}-\beta^0\|}$. Let $\langle \beta_-, \beta_+ \rangle = \{r\beta_- + (1-r)\beta_+ \mid r \in [0,1]\}$ be the line segment between β_- and β_+. By restricting attention to values $\beta \in \langle \beta_-, \beta_+ \rangle$ we can repeat the above argument for the case $\dim(\beta)=1$ and thus show that $\widehat{\beta} \in \langle \beta_-, \beta_+ \rangle$, which implies $\|\widehat{\beta}-\beta^0\| \leq \eta = \mathcal{O}_P((NT)^{-1/4})$. ■

S.3.3 Proof of Theorem B.4

Proof of Theorem B.4. A Taylor expansion of $\Delta(\beta,\phi)$ around (β^0,ϕ^0) yields

$$\Delta(\beta,\phi) = \Delta + [\partial_{\beta'}\Delta](\beta-\beta^0) + [\partial_{\phi'}\Delta](\phi-\phi^0) + \tfrac{1}{2}(\phi-\phi^0)'[\partial_{\phi\phi'}\Delta](\phi-\phi^0) + R_1^\Delta(\beta,\phi),$$

with remainder term

$$\begin{aligned}R_1^\Delta(\beta,\phi) &= \tfrac{1}{2}(\beta-\beta^0)'[\partial_{\beta\beta'}\Delta(\bar{\beta},\phi)](\beta-\beta^0) + (\beta-\beta^0)'[\partial_{\beta\phi'}\Delta(\beta^0,\tilde{\phi})](\phi-\phi^0)\\ &\quad + \tfrac{1}{6}\sum_g (\phi-\phi^0)'[\partial_{\phi\phi'\phi_g}\Delta(\beta^0,\bar{\phi})](\phi-\phi^0)[\phi-\phi^0]_g,\end{aligned}$$

where $\bar{\beta}$ is between β and β^0, and $\tilde{\phi}$ and $\bar{\phi}$ are between ϕ and ϕ^0.

By assumption, $\|\widehat{\beta}-\beta^0\| = o_P((NT)^{-1/4})$, and by the expansion of $\widehat{\phi} = \widehat{\phi}(\widehat{\beta})$ in Theorem B.1,

$$\begin{aligned}\|\widehat{\phi}-\phi^0\|_q &\leq \left\|\mathcal{H}^{-1}\right\|_q \|\mathcal{S}\|_q + \left\|\mathcal{H}^{-1}\right\|_q \|\partial_{\phi\beta'}\mathcal{L}\|_q \left\|\widehat{\beta}-\beta^0\right\|_q + \tfrac{1}{2}\left\|\mathcal{H}^{-1}\right\|_q^3 \|\partial_{\phi\phi\phi}\mathcal{L}\|_q \|\mathcal{S}\|_q^2 + \left\|R^\phi(\widehat{\beta})\right\|_q\\ &= \mathcal{O}_P((NT)^{-1/4+1/(2q)}).\end{aligned}$$

Thus, for $\widehat{R}_1^\Delta := R_1^\Delta(\widehat{\beta},\widehat{\phi})$,

$$\begin{aligned}\left|\widehat{R}_1^\Delta\right| &\leq \tfrac{1}{2}\|\widehat{\beta}-\beta^0\|^2 \sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)} \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)} \|\partial_{\beta\beta'}\Delta(\beta,\phi)\|\\ &\quad + (NT)^{1/2-1/q}\|\widehat{\beta}-\beta^0\|\|\widehat{\phi}-\phi^0\|_q \sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)} \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)} \|\partial_{\beta\phi'}\Delta(\beta,\phi)\|_q\\ &\quad + \tfrac{1}{6}(NT)^{1/2-1/q}\|\widehat{\phi}-\phi^0\|_q^3 \sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)} \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)} \|\partial_{\phi\phi\phi}\Delta(\beta,\phi)\|_q\\ &= o_P(1/\sqrt{NT}).\end{aligned}$$

Again by the expansion of $\widehat{\phi} = \widehat{\phi}(\widehat{\beta})$ from Theorem B.1,

$$\begin{aligned}\widehat{\delta}-\delta &= \Delta(\widehat{\beta},\widehat{\phi}) - \Delta = \left(\partial_{\beta'}\Delta + [\partial_\phi\Delta]'\mathcal{H}^{-1}[\partial_{\phi\beta'}\mathcal{L}]\right)(\widehat{\beta}-\beta^0)\\ &\quad + [\partial_\phi\Delta]'\mathcal{H}^{-1}\left(\mathcal{S} + \tfrac{1}{2}\sum_{g=1}^{\dim\phi}[\partial_{\phi\phi'\phi_g}\mathcal{L}]\mathcal{H}^{-1}\mathcal{S}[\mathcal{H}^{-1}\mathcal{S}]_g\right) + \tfrac{1}{2}\,\mathcal{S}'\mathcal{H}^{-1}[\partial_{\phi\phi'}\Delta]\mathcal{H}^{-1}\mathcal{S} + R_2^\Delta,\end{aligned} \tag{S.8}$$

where

$$\begin{aligned}\left|R_2^\Delta\right| &= \left|R_1^\Delta + [\partial_\phi\Delta]'R^\phi(\widehat{\beta}) + \tfrac{1}{2}(\widehat{\phi}-\phi^0+\mathcal{H}^{-1}\mathcal{S})'[\partial_{\phi\phi'}\Delta](\widehat{\phi}-\phi^0-\mathcal{H}^{-1}\mathcal{S})\right|\\ &\leq \left|R_1^\Delta\right| + (NT)^{1/2-1/q}\|\partial_\phi\Delta\|_q \left\|R^\phi(\widehat{\beta})\right\|_q\\ &\quad + \tfrac{1}{2}(NT)^{1/2-1/q}\left\|\widehat{\phi}-\phi^0+\mathcal{H}^{-1}\mathcal{S}\right\|_q \|\partial_{\phi\phi'}\Delta\|_q \left\|\widehat{\phi}-\phi^0-\mathcal{H}^{-1}\mathcal{S}\right\|_q\\ &= o_P(1/\sqrt{NT}),\end{aligned}$$

that uses $\left\|\widehat{\phi}-\phi^0-\mathcal{H}^{-1}\mathcal{S}\right\|_q=\mathcal{O}_P\left((NT)^{-1/2+1/q+\epsilon}\right)$. From equation (S.8), the terms of the expansion for $\widehat{\delta}-\delta$ are analogous to the terms of the expansion for the score in Theorem B.1, with $\Delta(\beta,\phi)$ taking the role of $\frac{1}{\sqrt{NT}}\partial_{\beta_k}\mathcal{L}(\beta,\phi)$. ■

S.4 Proofs of Appendix C (Theorem C.1)

Proof of Theorem C.1, Part (i)**.** Assumption B.1(i) is satisfied because $\lim_{N,T\to\infty}\frac{\dim\phi}{\sqrt{NT}}=\lim_{N,T\to\infty}\frac{N+T}{\sqrt{NT}}=\kappa+\kappa^{-1}$.

Assumption B.1(ii) is satisfied because $\ell_{it}(\beta,\pi)$ and $(v'\phi)^2$ are four times continuously differentiable and the same is true for $\mathcal{L}(\beta,\phi)$.

Let $\overline{\mathcal{D}}=\operatorname{diag}\left(\overline{\mathcal{H}}^*_{(\alpha\alpha)},\overline{\mathcal{H}}^*_{(\gamma\gamma)}\right)$. Then, $\left\|\overline{\mathcal{D}}^{-1}\right\|_\infty=\mathcal{O}_P(1)$ by Assumption 4.1(v). By the properties of the matrix norms and Lemma D.1, $\left\|\overline{\mathcal{H}}^{-1}-\overline{\mathcal{D}}^{-1}\right\|_\infty\le(N+T)\left\|\overline{\mathcal{H}}^{-1}-\overline{\mathcal{D}}^{-1}\right\|_{\max}=\mathcal{O}_P(1)$. Thus, $\left\|\overline{\mathcal{H}}^{-1}\right\|_q\le\left\|\overline{\mathcal{H}}^{-1}\right\|_\infty\le\left\|\overline{\mathcal{D}}^{-1}\right\|_\infty+\left\|\overline{\mathcal{H}}^{-1}-\overline{\mathcal{D}}^{-1}\right\|_\infty=\mathcal{O}_P(1)$ by Lemma S.4 and the triangle inequality. We conclude that Assumption B.1(iv) holds.

We now show that the assumptions of Lemma S.7 are satisfied:

(i) By Lemma S.2, $\chi_i=\frac{1}{\sqrt{T}}\sum_t\partial_{\beta_k}\ell_{it}$ satisfies $\mathbb{E}_\phi(\chi_i^2)\le B$. Thus, by independence across i

$$\mathbb{E}_\phi\left[\left(\frac{1}{\sqrt{NT}}\sum_{i,t}\partial_{\beta_k}\ell_{it}\right)^2\right]=\mathbb{E}_\phi\left[\left(\frac{1}{\sqrt{N}}\sum_i\chi_i\right)^2\right]=\frac{1}{N}\sum_i\mathbb{E}_\phi\chi_i^2\le B,$$

and therefore $\frac{1}{\sqrt{NT}}\sum_{i,t}\partial_{\beta_k}\ell_{it}=\mathcal{O}_P(1)$. Analogously, $\frac{1}{NT}\sum_{i,t}\left\{\partial_{\beta_k\beta_l}\ell_{it}-\mathbb{E}_\phi\left[\partial_{\beta_k\beta_l}\ell_{it}\right]\right\}=\mathcal{O}_P(1/\sqrt{NT})=o_P(1)$. Next,

$$\begin{aligned}&\mathbb{E}_\phi\left(\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)}\sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)}\frac{1}{NT}\sum_{i,t}\partial_{\beta_k\beta_l\beta_m}\ell_{it}(\beta,\pi_{it})\right)^2\\&\le\mathbb{E}_\phi\left(\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)}\sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)}\frac{1}{NT}\sum_{i,t}\left|\partial_{\beta_k\beta_l\beta_m}\ell_{it}(\beta,\pi_{it})\right|\right)^2\le\mathbb{E}_\phi\left(\frac{1}{NT}\sum_{i,t}M(Z_{it})\right)^2\\&\le\mathbb{E}_\phi\frac{1}{NT}\sum_{i,t}M(Z_{it})^2=\frac{1}{NT}\sum_{i,t}\mathbb{E}_\phi M(Z_{it})^2=\mathcal{O}_P(1),\end{aligned}$$

and therefore $\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)}\sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)}\frac{1}{NT}\sum_{i,t}\partial_{\beta_k\beta_l\beta_m}\ell_{it}(\beta,\pi_{it})=\mathcal{O}_P(1)$. A similar argument gives $\frac{1}{NT}\sum_{i,t}\partial_{\beta_k\beta_l}\ell_{it}=\mathcal{O}_P(1)$.

(ii) For $\xi_{it}(\beta,\phi) = \partial_{\beta_k\pi}\ell_{it}(\beta,\pi_{it})$ or $\xi_{it}(\beta,\phi) = \partial_{\beta_k\beta_l\pi}\ell_{it}(\beta,\pi_{it})$,

$$\begin{aligned}
&\mathbb{E}_\phi\left[\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)}\sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)}\frac{1}{T}\sum_t\left|\frac{1}{N}\sum_i\xi_{it}(\beta,\phi)\right|^q\right] \\
&\le \mathbb{E}_\phi\left[\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)}\sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)}\frac{1}{T}\sum_t\left(\frac{1}{N}\sum_i|\xi_{it}(\beta,\phi)|\right)^q\right] \\
&\le \mathbb{E}_\phi\left[\frac{1}{T}\sum_t\left(\frac{1}{N}\sum_i M(Z_{it})\right)^q\right] \le \mathbb{E}_\phi\left[\frac{1}{T}\sum_t\frac{1}{N}\sum_i M(Z_{it})^q\right] \\
&= \frac{1}{T}\sum_t\frac{1}{N}\sum_i \mathbb{E}_\phi M(Z_{it})^q = \mathcal{O}_P(1),
\end{aligned}$$

i.e. $\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)}\sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)}\frac{1}{T}\sum_t\left|\frac{1}{N}\sum_i\xi_{it}(\beta,\phi)\right|^q = \mathcal{O}_P(1)$. Analogously, it follows that $\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)}\sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)}\frac{1}{N}\sum_i\left|\frac{1}{T}\sum_t\xi_{it}(\beta,\phi)\right|^q = \mathcal{O}_P(1)$.

(iii) For $\xi_{it}(\beta,\phi) = \partial_{\pi^r}\ell_{it}(\beta,\pi_{it})$, with $r\in\{3,4\}$, or $\xi_{it}(\beta,\phi) = \partial_{\beta_k\pi^r}\ell_{it}(\beta,\pi_{it})$, with $r\in\{2,3\}$, or $\xi_{it}(\beta,\phi) = \partial_{\beta_k\beta_l\pi^2}\ell_{it}(\beta,\pi_{it})$,

$$\begin{aligned}
&\mathbb{E}_\phi\left[\left(\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)}\sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)}\max_i\frac{1}{T}\sum_t|\xi_{it}(\beta,\phi)|\right)^{(8+\nu)}\right] \\
&= \mathbb{E}_\phi\left[\max_i\left(\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)}\sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)}\frac{1}{T}\sum_t|\xi_{it}(\beta,\phi)|\right)^{(8+\nu)}\right] \\
&\le \mathbb{E}_\phi\left[\sum_i\left(\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)}\sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)}\frac{1}{T}\sum_t|\xi_{it}(\beta,\phi)|\right)^{(8+\nu)}\right] \le \mathbb{E}_\phi\left[\sum_i\left(\frac{1}{T}\sum_t M(Z_{it})\right)^{(8+\nu)}\right] \\
&\le \mathbb{E}_\phi\left[\sum_i\frac{1}{T}\sum_t M(Z_{it})^{(8+\nu)}\right] = \sum_i\frac{1}{T}\sum_t\mathbb{E}_\phi M(Z_{it})^{(8+\nu)} = \mathcal{O}_P(N).
\end{aligned}$$

Thus, $\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)}\sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)}\max_i\frac{1}{T}\sum_t|\xi_{it}(\beta,\phi)| = \mathcal{O}_P\left(N^{1/(8+\nu)}\right) = \mathcal{O}_P\left(N^{2\epsilon}\right)$. Analogously, it follows that $\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)}\sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)}\max_t\frac{1}{N}\sum_i|\xi_{it}(\beta,\phi)| = \mathcal{O}_P\left(N^{2\epsilon}\right)$.

(iv) Let $\chi_t = \frac{1}{\sqrt{N}}\sum_i\partial_\pi\ell_{it}$. By cross-sectional independence and $\mathbb{E}_\phi(\partial_\pi\ell_{it})^8 \le \mathbb{E}_\phi M(Z_{it})^8 = \mathcal{O}_P(1)$, $\mathbb{E}_\phi\chi_t^8 = \mathcal{O}_P(1)$ uniformly over t. Thus, $\mathbb{E}_\phi\frac{1}{T}\sum_t\chi_t^8 = \mathcal{O}_P(1)$ and therefore $\frac{1}{T}\sum_t\left|\frac{1}{\sqrt{N}}\sum_i\partial_\pi\ell_{it}\right|^q = \mathcal{O}_P(1)$, with $q=8$.

Let $\chi_i = \frac{1}{\sqrt{T}}\sum_t\partial_\pi\ell_{it}(\beta^0,\pi^0_{it})$. By Lemma S.2 and $\mathbb{E}_\phi(\partial_\pi\ell_{it})^{8+\nu} \le \mathbb{E}_\phi M(Z_{it})^{8+\nu} = \mathcal{O}_P(1)$, $\mathbb{E}_\phi\chi_i^8 = \mathcal{O}_P(1)$ uniformly over i. Here we use $\mu > 4/[1-8/(8+\nu)] = 4(8+\nu)/\nu$ that is imposed in Assumption B.1. Thus, $\mathbb{E}_\phi\frac{1}{N}\sum_i\chi_i^8 = \mathcal{O}_P(1)$ and therefore $\frac{1}{N}\sum_i\left|\frac{1}{\sqrt{T}}\sum_t\partial_\pi\ell_{it}\right|^q = \mathcal{O}_P(1)$, with $q=8$.

The proofs for $\frac{1}{T}\sum_t\left|\frac{1}{\sqrt{N}}\sum_i\partial_{\beta_k\pi}\ell_{it} - \mathbb{E}_\phi\left[\partial_{\beta_k\pi}\ell_{it}\right]\right|^2 = \mathcal{O}_P(1)$ and $\frac{1}{N}\sum_i\left|\frac{1}{\sqrt{T}}\sum_t\partial_{\beta_k\pi}\ell_{it} - \mathbb{E}_\phi\left[\partial_{\beta_k\pi}\ell_{it}\right]\right|^2 = \mathcal{O}_P(1)$ are analogous.

(v) It follows by the independence of $\{(\ell_{i1},\ldots,\ell_{iT}) : 1\le i\le N\}$ across i, conditional on ϕ, in Assumption B.1(ii).

(vi) Let $\xi_{it} = \partial_{\pi^r}\ell_{it}(\beta^0,\pi^0_{it}) - \mathbb{E}_\phi\left[\partial_{\pi^r}\ell_{it}\right]$, with $r\in\{2,3\}$, or $\xi_{it} = \partial_{\beta_k\pi^2}\ell_{it}(\beta^0,\pi^0_{it}) - \mathbb{E}_\phi\left[\partial_{\beta_k\pi^2}\ell_{it}\right]$. For

$\tilde{\nu} = \nu$, $\max_i \mathbb{E}_\phi \left[\xi_{it}^{8+\tilde{\nu}}\right] = \mathcal{O}_P(1)$ by assumption. By Lemma S.1,

$$\begin{aligned}\left|\sum_s \mathbb{E}_\phi\left[\xi_{it}\xi_{is}\right]\right| &= \sum_s \left|\mathrm{Cov}_\phi\left(\xi_{it},\, \xi_{is}\right)\right| \\ &\leq \sum_s \left[8\; a(|t-s|)\right]^{1-2/(8+\nu)}\; \left[\mathbb{E}_\phi|\xi_t|^{8+\nu}\right]^{1/(8+\nu)}\; \left[\mathbb{E}_\phi|\xi_s|^{8+\nu}\right]^{1/(8+\nu)} \\ &= \tilde{C}\sum_{m=1}^{\infty} m^{-\mu[1-2/(8+\nu)]} \leq \tilde{C}\sum_{m=1}^{\infty} m^{-4} = \tilde{C}\pi^4/90,\end{aligned}$$

where $\tilde{C}$ is a constant. Here we use that $\mu > 4(8+\nu)/\nu$ implies $\mu[1-2/(8+\nu) > 4$. We thus have shown $\max_i \max_t \sum_s \mathbb{E}_\phi\left[\xi_{it}\xi_{js}\right] \leq \tilde{C}\pi^4/90 =: C$.

Analogous to the proof of part (iv), we can use Lemma S.2 to obtain $\max_i \mathbb{E}_\phi\left\{\left[\frac{1}{\sqrt{T}}\sum_t \xi_{it}\right]^8\right\} \leq C$, and independence across i to obtain $\max_t \mathbb{E}_\phi\left\{\left[\frac{1}{\sqrt{N}}\sum_i \xi_{it}\right]^8\right\} \leq C$. Similarly, by Lemma S.2

$$\max_{i,j} \mathbb{E}_\phi\left\{\left[\frac{1}{\sqrt{T}}\sum_t \left[\xi_{it}\xi_{jt} - \mathbb{E}_\phi\left(\xi_{it}\xi_{jt}\right)\right]\right]^4\right\} \leq C,$$

which requires $\mu > 2/[1-4/(4+\nu/2)]$, which is implied by the assumption that $\mu > 4(8+\nu)/\nu$.

(vii) We have already shown that $\left\|\overline{\mathcal{H}}^{-1}\right\|_q = \mathcal{O}_P(1)$.

Therefore, we can apply Lemma S.7, which shows that Assumption B.1(v) and (vi) hold. We have already shown that Assumption B.1(i), (ii), (iv), (v) and (vi) hold. One can also check that $(NT)^{-1/4+1/(2q)} = o_P(r_\phi)$ and $(NT)^{1/(2q)}r_\beta = o_P(r_\phi)$ are satisfied. In addition, $\mathcal{L}(\beta,\phi)$ is strictly concave. We can therefore invoke Theorem B.3 to show that Assumption B.1(iii) holds and that $\|\widehat{\beta}-\beta^0\| = \mathcal{O}_P((NT)^{-1/4})$. ■

Proof of Theorem C.1, Part (ii)**.** For any $N \times T$ matrix A we define the $N\times T$ matrix $\mathbb{P}A$ as follows

$$(\mathbb{P}A)_{it} = \alpha_i^* + \gamma_t^*, \qquad (\alpha^*,\gamma^*) \in \operatorname*{argmin}_{\alpha,\gamma} \sum_{i,t} \mathbb{E}_\phi(-\partial_{\pi^2}\ell_{it})\left(A_{it}-\alpha_i-\gamma_t\right)^2. \tag{S.1}$$

Here, the minimization is over $\alpha \in \mathbb{R}^N$ and $\gamma \in \mathbb{R}^T$. The operator $\mathbb{P}$ is a linear projection, i.e. we have $\mathbb{P}\mathbb{P} = \mathbb{P}$. It is also convenient to define

$$\widetilde{\mathbb{P}}A = \mathbb{P}\tilde{A}, \quad \text{where} \quad \tilde{A}_{it} = \frac{A_{it}}{\mathbb{E}_\phi(-\partial_{\pi^2}\ell_{it})}. \tag{S.2}$$

$\widetilde{\mathbb{P}}$ is a linear operator, but not a projection. Note that Λ and Ξ defined in (C.1) and (4.3) can be written as $\Lambda = \widetilde{\mathbb{P}}A$ and $\Xi_k = \widetilde{\mathbb{P}}B_k$, where $A_{it} = -\partial_\pi\ell_{it}$ and $B_{k,it} = -\mathbb{E}_\phi(\partial_{\beta_k\pi}\ell_{it})$, for $k = 1,\ldots,\dim\beta$.[5]

By Lemma S.8(ii),

$$\overline{W} = -\frac{1}{\sqrt{NT}}\left(\partial_{\beta\beta'}\overline{\mathcal{L}} + [\partial_{\beta\phi'}\overline{\mathcal{L}}]\;\overline{\mathcal{H}}^{-1}\;[\partial_{\phi\beta'}\overline{\mathcal{L}}]\right) = -\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \left[\mathbb{E}_\phi\left(\partial_{\beta\beta'}\ell_{it}\right) + \mathbb{E}_\phi\left(-\partial_{\pi^2}\ell_{it}\right)\Xi_{it}\Xi_{it}'\right].$$

[5] B_k and Ξ_k are $N\times T$ matrices with entries $B_{k,it}$ and $\Xi_{k,it}$, respectively, while B_{it} and Ξ_{it} are $\dim\beta$-vectors with entries $B_{k,it}$ and $\Xi_{k,it}$.

By Lemma S.8(i),

$$U^{(0)} = \partial_\beta \mathcal{L} + [\partial_{\beta\phi'} \overline{\mathcal{L}}] \, \overline{\mathcal{H}}^{-1} \mathcal{S} = \frac{1}{\sqrt{NT}} \sum_{i,t} \left(\partial_\beta \ell_{it} - \Xi_{it} \, \partial_\pi \ell_{it} \right) = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} D_\beta \ell_{it}.$$

We decompose $U^{(1)} = U^{(1a)} + U^{(1b)}$, with

$$U^{(1a)} = [\partial_{\beta\phi'} \widetilde{\mathcal{L}}] \overline{\mathcal{H}}^{-1} \mathcal{S} - [\partial_{\beta\phi'} \overline{\mathcal{L}}] \, \overline{\mathcal{H}}^{-1} \, \widetilde{\mathcal{H}} \, \overline{\mathcal{H}}^{-1} \, \mathcal{S},$$

$$U^{(1b)} = \sum_{g=1}^{\dim \phi} \left(\partial_{\beta\phi'\phi_g} \overline{\mathcal{L}} + [\partial_{\beta\phi'} \overline{\mathcal{L}}] \, \overline{\mathcal{H}}^{-1} [\partial_{\phi\phi'\phi_g} \overline{\mathcal{L}}] \right) \overline{\mathcal{H}}^{-1} \mathcal{S} [\overline{\mathcal{H}}^{-1} \mathcal{S}]_g / 2.$$

By Lemma S.8(i) and (iii),

$$U^{(1a)} = -\frac{1}{\sqrt{NT}} \sum_{i,t} \Lambda_{it} \left(\partial_{\beta\pi} \tilde{\ell}_{it} + \Xi_{it} \, \partial_{\pi^2} \tilde{\ell}_{it} \right) = -\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \Lambda_{it} \, \left[D_{\beta\pi} \ell_{it} - \mathbb{E}_\phi (D_{\beta\pi} \ell_{it}) \right],$$

and

$$U^{(1b)} = \frac{1}{2\sqrt{NT}} \sum_{i,t} \Lambda_{it}^2 \left[\mathbb{E}_\phi (\partial_{\beta\pi^2} \ell_{it}) + [\partial_{\beta\phi'} \overline{\mathcal{L}}] \, \overline{\mathcal{H}}^{-1} \mathbb{E}_\phi (\partial_\phi \partial_{\pi^2} \ell_{it}) \right],$$

where for each i, t, $\partial_\phi \partial_{\pi^2} \ell_{it}$ is a $\dim \phi$-vector, which can be written as $\partial_\phi \partial_{\pi^2} \ell_{it} = \binom{A 1_T}{A' 1_N}$ for an $N \times T$ matrix A with elements $A_{j\tau} = \partial_{\pi^3} \ell_{j\tau}$ if $j = i$ and $\tau = t$, and $A_{j\tau} = 0$ otherwise. Thus, Lemma S.8(i) gives $[\partial_{\beta\phi'} \overline{\mathcal{L}}] \, \overline{\mathcal{H}}^{-1} \partial_\phi \partial_{\pi^2} \ell_{it} = -\sum_{j,\tau} \Xi_{j\tau} 1(i = j) 1(t = \tau) \partial_{\pi^3} \ell_{it} = -\Xi_{it} \partial_{\pi^3} \ell_{it}$. Therefore

$$U^{(1b)} = \frac{1}{2\sqrt{NT}} \sum_{i,t} \Lambda_{it}^2 \mathbb{E}_\phi \left(\partial_{\beta\pi^2} \ell_{it} - \Xi_{it} \partial_{\pi^3} \ell_{it} \right) = \frac{1}{2\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \Lambda_{it}^2 \, \mathbb{E}_\phi (D_{\beta\pi^2} \ell_{it}).$$

■

Proof of Theorem C.1, Part (iii)**.** Showing that Assumption B.2 is satisfied is analogous to the proof of Lemma S.7 and of part (ii) of this Theorem.

In the proof of Theorem 4.1 we show that Assumption 4.1 implies that $U = \mathcal{O}_P(1)$. This fact together with part (i) of this theorem show that Corollary B.2 is applicable, so that $\sqrt{NT}(\widehat{\beta} - \beta^0) = \overline{W}_\infty^{-1} U + o_P(1) = \mathcal{O}_P(1)$, and we can apply Theorem B.4.

By Lemma S.8 and the result for $\sqrt{NT}(\widehat{\beta} - \beta^0)$,

$$\sqrt{NT} \left[\partial_{\beta'} \overline{\Delta} + (\partial_{\phi'} \overline{\Delta}) \overline{\mathcal{H}}^{-1} (\partial_{\phi\beta'} \overline{\mathcal{L}}) \right] (\widehat{\beta} - \beta^0) = \left[\frac{1}{NT} \sum_{i,t} \mathbb{E}_\phi (D_\beta \Delta_{it}) \right]' \overline{W}_\infty^{-1} \left(U^{(0)} + U^{(1)} \right) + o_P(1). \tag{S.3}$$

We apply Lemma S.8 to $U_\Delta^{(0)}$ and $U_\Delta^{(1)}$ defined in Theorem B.4 to give

$$\sqrt{NT} \, U_\Delta^{(0)} = -\frac{1}{\sqrt{NT}} \sum_{i,t} \mathbb{E}_\phi (\Psi_{it}) \partial_\pi \ell_{it},$$

$$\begin{aligned} \sqrt{NT} \, U_\Delta^{(1)} = {} & \frac{1}{\sqrt{NT}} \sum_{i,t} \Lambda_{it} \left[\Psi_{it} \partial_{\pi^2} \ell_{it} - \mathbb{E}_\phi (\Psi_{it}) \mathbb{E}_\phi (\partial_{\pi^2} \ell_{it}) \right] \\ & + \frac{1}{2\sqrt{NT}} \sum_{i,t} \Lambda_{it}^2 \left[\mathbb{E}_\phi (\partial_{\pi^2} \Delta_{it}) - \mathbb{E}_\phi (\partial_{\pi^3} \ell_{it}) \mathbb{E}_\phi (\Psi_{it}) \right]. \end{aligned} \tag{S.4}$$

The derivation of (S.3) and (S.4) is analogous to the proof of the part (ii) of the Theorem. Combining Theorem B.4 with equations (S.3) and (S.4) gives the result. ■

S.5 Proofs of Appendix D (Lemma D.1)

The following Lemmas are useful to prove Lemma D.1. Let $\mathcal{L}^*(\beta, \phi) = (NT)^{-1/2} \sum_{i,t} \ell_{it}(\beta, \alpha_i + \gamma_t)$.

Lemma S.1. *If the statement of Lemma D.1 holds for some constant $b > 0$, then it holds for any constant $b > 0$.*

Proof of Lemma S.1. Write $\overline{\mathcal{H}} = \overline{\mathcal{H}}^* + \frac{b}{\sqrt{NT}} vv'$, where $\overline{\mathcal{H}}^* = \mathbb{E}_\phi \left[-\frac{\partial^2}{\partial\phi\partial\phi'} \mathcal{L}^* \right]$. Since $\overline{\mathcal{H}}^* v = 0$,

$$\overline{\mathcal{H}}^{-1} = \left(\overline{\mathcal{H}}^* \right)^\dagger + \left(\frac{b}{\sqrt{NT}} vv' \right)^\dagger = \left(\overline{\mathcal{H}}^* \right)^\dagger + \frac{\sqrt{NT}}{b \|vv'\|^2} vv' = \left(\overline{\mathcal{H}}^* \right)^\dagger + \frac{\sqrt{NT}}{b(N+T)^2} vv',$$

where † refers to the Moore-Penrose pseudo-inverse. Thus, if $\overline{\mathcal{H}}_1$ is the expected Hessian for $b = b_1 > 0$ and $\overline{\mathcal{H}}_2$ is the expected Hessian for $b = b_2 > 0$, $\left\| \overline{\mathcal{H}}_1^{-1} - \overline{\mathcal{H}}_2^{-1} \right\|_{\max} = \left\| \left(\frac{1}{b_1} - \frac{1}{b_2} \right) \frac{\sqrt{NT}}{(N+T)^2} vv' \right\|_{\max} = \mathcal{O}\left((NT)^{-1/2} \right)$. ■

Lemma S.2. *Let Assumption 4.1 hold and let $0 < b \leq b_{\min} \left(1 + \frac{\max(N,T)}{\min(N,T)} \frac{b_{\max}}{b_{\min}} \right)^{-1}$. Then,*

$$\left\| \overline{\mathcal{H}}_{(\alpha\alpha)}^{-1} \overline{\mathcal{H}}_{(\alpha\gamma)} \right\|_\infty < 1 - \frac{b}{b_{\max}}, \quad \textit{and} \quad \left\| \overline{\mathcal{H}}_{(\gamma\gamma)}^{-1} \overline{\mathcal{H}}_{(\gamma\alpha)} \right\|_\infty < 1 - \frac{b}{b_{\max}}.$$

Proof of Lemma S.2. Let $h_{it} = \mathbb{E}_\phi(-\partial_{\pi^2} \ell_{it})$, and define

$$\tilde{h}_{it} = h_{it} - b - \frac{1}{b^{-1} + \sum_j \left(\sum_\tau h_{j\tau} \right)^{-1}} \sum_j \frac{h_{jt} - b}{\sum_\tau h_{j\tau}}.$$

By definition, $\overline{\mathcal{H}}_{(\alpha\alpha)} = \overline{\mathcal{H}}_{(\alpha\alpha)}^* + b 1_N 1_N' / \sqrt{NT}$ and $\overline{\mathcal{H}}_{(\alpha\gamma)} = \overline{\mathcal{H}}_{(\alpha\gamma)}^* - b 1_N 1_T' / \sqrt{NT}$. The matrix $\overline{\mathcal{H}}_{(\alpha\alpha)}^*$ is diagonal with elements $\sum_t h_{it} / \sqrt{NT}$. The matrix $\overline{\mathcal{H}}_{(\alpha\gamma)}^*$ has elements $h_{it}/\sqrt{NT}$. The Woodbury identity states that

$$\overline{\mathcal{H}}_{(\alpha\alpha)}^{-1} = \overline{\mathcal{H}}_{(\alpha\alpha)}^{*-1} - \overline{\mathcal{H}}_{(\alpha\alpha)}^{*-1} 1_N \left(\sqrt{NT}\, b^{-1} + 1_N' \overline{\mathcal{H}}_{(\alpha\alpha)}^{*-1} 1_N \right)^{-1} 1_N' \overline{\mathcal{H}}_{(\alpha\alpha)}^{*-1}.$$

Then, $\overline{\mathcal{H}}_{(\alpha\alpha)}^{-1} \overline{\mathcal{H}}_{(\alpha\gamma)} = \overline{\mathcal{H}}_{(\alpha\alpha)}^{*-1} \tilde{H} / \sqrt{NT}$, where $\tilde{H}$ is the $N \times T$ matrix with elements $\tilde{h}_{it}$. Therefore

$$\left\| \overline{\mathcal{H}}_{(\alpha\alpha)}^{-1} \overline{\mathcal{H}}_{(\alpha\gamma)} \right\|_\infty = \max_i \frac{\sum_t \left| \tilde{h}_{it} \right|}{\sum_t h_{it}}.$$

Assumption 4.1(iv) guarantees that $b_{\max} \geq h_{it} \geq b_{\min}$, which implies $h_{jt} - b \geq b_{\min} - b > 0$, and

$$\tilde{h}_{it} > h_{it} - b - \frac{1}{b^{-1}} \sum_j \frac{h_{jt} - b}{\sum_\tau h_{j\tau}} \geq b_{\min} - b \left(1 + \frac{N}{T} \frac{b_{\max}}{b_{\min}} \right) \geq 0.$$

We conclude that

$$\left\| \overline{\mathcal{H}}_{(\alpha\alpha)}^{-1} \overline{\mathcal{H}}_{(\alpha\gamma)} \right\|_\infty = \max_i \frac{\sum_t \tilde{h}_{it}}{\sum_t h_{it}} = 1 - \min_i \frac{1}{\sum_t h_{it}} \sum_t \left(b + \frac{1}{b^{-1} + \sum_j \left(\sum_\tau h_{j\tau} \right)^{-1}} \sum_j \frac{h_{jt} - b}{\sum_\tau h_{j\tau}} \right)$$
$$< 1 - \frac{b}{b_{\max}}.$$

Analogously, $\left\| \overline{\mathcal{H}}_{(\gamma\gamma)}^{-1} \overline{\mathcal{H}}_{(\gamma\alpha)} \right\|_\infty < 1 - \frac{b}{b_{\max}}$. ■

Proof of Lemma D.1. We choose $b < b_{\min}\left(1+\max(\kappa^2,\kappa^{-2})\frac{b_{\max}}{b_{\min}}\right)^{-1}$. Then, $b \leq b_{\min}\left(1+\frac{\max(N,T)}{\min(N,T)}\frac{b_{\max}}{b_{\min}}\right)^{-1}$ for large enough N and T, so that Lemma S.2 becomes applicable. The choice of b has no effect on the general validity of the lemma for all $b>0$ by Lemma S.1.

By the inversion formula for partitioned matrices,

$$\overline{\mathcal{H}}^{-1} = \begin{pmatrix} A & -A\,\overline{\mathcal{H}}_{(\alpha\gamma)}\,\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)} \\ -\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\,\overline{\mathcal{H}}_{(\gamma\alpha)}\,A & \overline{\mathcal{H}}^{-1}_{(\gamma\gamma)} + \overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\,\overline{\mathcal{H}}_{(\gamma\alpha)}\,A\,\overline{\mathcal{H}}_{(\alpha\gamma)}\,\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)} \end{pmatrix},$$

with $A := (\overline{\mathcal{H}}_{(\alpha\alpha)} - \overline{\mathcal{H}}_{(\alpha\gamma)}\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\overline{\mathcal{H}}_{(\gamma\alpha)})^{-1}$. The Woodbury identity states that

$$\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)} = \overline{\mathcal{H}}^{*-1}_{(\alpha\alpha)} - \underbrace{\overline{\mathcal{H}}^{*-1}_{(\alpha\alpha)} 1_N \left(\sqrt{NT}/b + 1_N' \overline{\mathcal{H}}^{*-1}_{(\alpha\alpha)} 1_N\right)^{-1} 1_N' \overline{\mathcal{H}}^{*-1}_{(\alpha\alpha)}}_{=:C_{(\alpha\alpha)}},$$

$$\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)} = \overline{\mathcal{H}}^{*-1}_{(\gamma\gamma)} - \underbrace{\overline{\mathcal{H}}^{*-1}_{(\gamma\gamma)} 1_T \left(\sqrt{NT}/b + 1_T' \overline{\mathcal{H}}^{*-1}_{(\gamma\gamma)} 1_T\right)^{-1} 1_T' \overline{\mathcal{H}}^{*-1}_{(\gamma\gamma)}}_{=:C_{(\gamma\gamma)}}.$$

By Assumption 4.1(v), $\|\overline{\mathcal{H}}^{*-1}_{(\alpha\alpha)}\|_\infty = \mathcal{O}_P(1)$, $\|\overline{\mathcal{H}}^{*-1}_{(\gamma\gamma)}\|_\infty = \mathcal{O}_P(1)$, $\|\overline{\mathcal{H}}^{*}_{(\alpha\gamma)}\|_{\max} = \mathcal{O}_P(1/\sqrt{NT})$. Therefore[6]

$$\|C_{(\alpha\alpha)}\|_{\max} \leq \|\overline{\mathcal{H}}^{*-1}_{(\alpha\alpha)}\|_\infty^2 \, \|1_N 1_N'\|_{\max} \left(\sqrt{NT}/b + 1_N' \overline{\mathcal{H}}^{*-1}_{(\alpha\alpha)} 1_N\right)^{-1} = \mathcal{O}_P(1/\sqrt{NT}),$$

$$\|\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}\|_\infty \leq \|\overline{\mathcal{H}}^{*-1}_{(\alpha\alpha)}\|_\infty + N\|C_{(\alpha\alpha)}\|_{\max} = \mathcal{O}_P(1).$$

Analogously, $\|C_{(\gamma\gamma)}\|_{\max} = \mathcal{O}_P(1/\sqrt{NT})$ and $\|\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\|_\infty = \mathcal{O}_P(1)$. Furthermore, $\|\overline{\mathcal{H}}_{(\alpha\gamma)}\|_{\max} \leq \|\overline{\mathcal{H}}^{*}_{(\alpha\gamma)}\|_{\max} + b/\sqrt{NT} = \mathcal{O}_P(1/\sqrt{NT})$. Define

$$B := \left(\mathbb{1}_N - \overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}\overline{\mathcal{H}}_{(\alpha\gamma)}\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\overline{\mathcal{H}}_{(\gamma\alpha)}\right)^{-1} - \mathbb{1}_N = \sum_{n=1}^{\infty}\left(\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}\overline{\mathcal{H}}_{(\alpha\gamma)}\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\overline{\mathcal{H}}_{(\gamma\alpha)}\right)^{n}.$$

Then, $A = \overline{\mathcal{H}}^{-1}_{(\alpha\alpha)} + \overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}B = \overline{\mathcal{H}}^{*-1}_{(\alpha\alpha)} - C_{(\alpha\alpha)} + \overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}B$. By Lemma S.2, $\|\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}\overline{\mathcal{H}}_{(\alpha\gamma)}\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\overline{\mathcal{H}}_{(\gamma\alpha)}\|_\infty \leq \|\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}\overline{\mathcal{H}}_{(\alpha\gamma)}\|_\infty \|\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\overline{\mathcal{H}}_{(\gamma\alpha)}\|_\infty < \left(1-\frac{b}{b_{\max}}\right)^2 < 1$, and

$$\begin{aligned}\|B\|_{\max} &\leq \sum_{n=0}^{\infty}\left(\|\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}\overline{\mathcal{H}}_{(\alpha\gamma)}\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\overline{\mathcal{H}}_{(\gamma\alpha)}\|_\infty\right)^n \|\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}\|_\infty \|\overline{\mathcal{H}}_{(\alpha\gamma)}\|_\infty \|\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\|_\infty \|\overline{\mathcal{H}}_{(\gamma\alpha)}\|_{\max} \\ &\leq \left[\sum_{n=0}^{\infty}\left(1-\frac{b}{b_{\max}}\right)^{2n}\right] T\,\|\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}\|_\infty \|\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\|_\infty \|\overline{\mathcal{H}}_{(\gamma\alpha)}\|^2_{\max} = \mathcal{O}_P(1/\sqrt{NT}).\end{aligned}$$

By the triangle inequality,

$$\|A\|_\infty \leq \|\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}\|_\infty + N\|\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}\|_\infty \|B\|_{\max} = \mathcal{O}_P(1).$$

Thus, for the different blocks of

$$\overline{\mathcal{H}}^{-1} - \begin{pmatrix} \overline{\mathcal{H}}^{*}_{(\alpha\alpha)} & 0 \\ 0 & \overline{\mathcal{H}}^{*}_{(\gamma\gamma)} \end{pmatrix}^{-1} = \begin{pmatrix} A - \overline{\mathcal{H}}^{*-1}_{(\alpha\alpha)} & -A\,\overline{\mathcal{H}}_{(\alpha\gamma)}\,\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)} \\ -\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\,\overline{\mathcal{H}}_{(\gamma\alpha)}\,A & \overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\,\overline{\mathcal{H}}_{(\gamma\alpha)}\,A\,\overline{\mathcal{H}}_{(\alpha\gamma)}\,\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)} - C_{(\gamma\gamma)} \end{pmatrix},$$

[6] Here and in the following me make use of the inequalities $\|AB\|_{\max} < \|A\|_\infty \|B\|_{\max}$, $\|AB\|_{\max} < \|A\|_{\max}\|B'\|_\infty$, $\|A\|_\infty \leq n\|A\|_{\max}$, which hold for any $m \times n$ matrix A and $n \times p$ matrix B.

we find

$$
\begin{aligned}
\left\| A - \overline{\mathcal{H}}^{*-1}_{(\alpha\alpha)} \right\|_{\max} &= \left\| \overline{\mathcal{H}}^{-1}_{(\alpha\alpha)} B - C_{(\alpha\alpha)} \right\|_{\max} \\
&\leq \| \overline{\mathcal{H}}^{-1}_{(\alpha\alpha)} \|_\infty \|B\|_{\max} - \|C_{(\alpha\alpha)}\|_{\max} = \mathcal{O}_P(1/\sqrt{NT}), \\
\left\| -A\, \overline{\mathcal{H}}_{(\alpha\gamma)}\, \overline{\mathcal{H}}^{-1}_{(\gamma\gamma)} \right\|_{\max} &\leq \|A\|_\infty \|\overline{\mathcal{H}}_{(\alpha\gamma)}\|_{\max} \|\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\|_\infty = \mathcal{O}_P(1/\sqrt{NT}), \\
\left\| \overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\, \overline{\mathcal{H}}_{(\gamma\alpha)}\, A\, \overline{\mathcal{H}}_{(\alpha\gamma)}\, \overline{\mathcal{H}}^{-1}_{(\gamma\gamma)} - C_{(\gamma\gamma)} \right\|_{\max} &\leq \|\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\|^2_\infty \|\overline{\mathcal{H}}_{(\gamma\alpha)}\|_\infty \|A\|_\infty \|\overline{\mathcal{H}}_{(\alpha\gamma)}\|_{\max} + \|C_{(\gamma\gamma)}\|_{\max} \\
&\leq N \|\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\|^2_\infty \|A\|_\infty \|\overline{\mathcal{H}}_{(\alpha\gamma)}\|^2_{\max} + \|C_{(\gamma\gamma)}\|_{\max} = \mathcal{O}_P(1/\sqrt{NT}).
\end{aligned}
$$

The bound $\mathcal{O}_P(1/\sqrt{NT})$ for the max-norm of each block of the matrix yields the same bound for the max-norm of the matrix itself. ■

S.6 Useful Lemmas

S.6.1 Some Properties of Stochastic Processes

Here we collect some known properties of α-mixing processes, which are useful for our proofs.

Lemma S.1. *Let $\{\xi_t\}$ be an α-mixing process with mixing coefficients $a(m)$. Let $\mathbb{E}|\xi_t|^p < \infty$ and $\mathbb{E}|\xi_{t+m}|^q < \infty$ for some $p, q \geq 1$ and $1/p + 1/q < 1$. Then,*

$$
|\text{Cov}\,(\xi_t,\, \xi_{t+m})| \;\leq\; 8\; a(m)^{1/r}\; \left[\mathbb{E}|\xi_t|^p\right]^{1/p} \left[\mathbb{E}|\xi_{t+m}|^q\right]^{1/q},
$$

where $r = (1 - 1/p - 1/q)^{-1}$.

Proof of Lemma S.1. See, for example, Proposition 2.5 in Fan and Yao (2003). ■

The following result is a simple modification of Theorem 1 in Cox and Kim (1995).

Lemma S.2. *Let $\{\xi_t\}$ be an α-mixing process with mixing coefficients $a(m)$. Let $r \geq 1$ be an integer, and let $\delta > 2r$, $\mu > r/(1-2r/\delta)$, $c > 0$ and $C > 0$. Assume that $\sup_t \mathbb{E}\,|\xi_t|^\delta \leq C$ and that $a(m) \leq c\, m^{-\mu}$ for all $m \in \{1, 2, 3, \ldots\}$. Then there exists a constant $B > 0$ depending on r, δ, μ, c and C, but not depending on T or any other distributional characteristics of ξ_t, such that for any $T > 0$,*

$$
\mathbb{E}\left[\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \xi_t \right)^{2r} \right] \leq B.
$$

The following is a central limit theorem for martingale difference sequences.

Lemma S.3. *Consider the scalar process $\xi_{it} = \xi_{NT,it}$, $i = 1, \ldots, N$, $t = 1, \ldots, T$. Let $\{(\xi_{i1}, \ldots, \xi_{iT}) : 1 \leq i \leq N\}$ be independent across i, and be a martingale difference sequence for each i, N, T. Let $\mathbb{E}|\xi_{it}|^{2+\delta}$ be uniformly bounded across i, t, N, T for some $\delta > 0$. Let $\overline{\sigma} = \overline{\sigma}_{NT} > \Delta > 0$ for all sufficiently large NT, and let $\frac{1}{NT}\sum_{i,t} \xi^2_{it} - \overline{\sigma}^2 \to_P 0$ as $NT \to \infty$.*[7] *Then,*

$$
\frac{1}{\overline{\sigma}\,\sqrt{NT}} \sum_{i,t} \xi_{it} \to_d \mathcal{N}(0, 1).
$$

[7] Here can allow for an arbitrary sequence of (N, T) with $NT \to \infty$.

Proof of Lemma S.3. Define $\xi_m = \xi_{M,m} = \xi_{NT,it}$, with $M = NT$ and $m = T(i-1)+t \in \{1,\ldots,M\}$. Then $\{\xi_m,\, m = 1,\ldots,M\}$ is a martingale difference sequence. With this redefinition the statement of the Lemma is equal to Corollary 5.26 in White (2001), which is based on Theorem 2.3 in Mcleish (1974), and which shows that $\frac{1}{\overline{\sigma}\sqrt{M}} \sum_{m=1}^{M} \xi_m \to_d \mathcal{N}(0,1)$. ■

S.6.2 Some Bounds for the Norms of Matrices and Tensors

The following lemma provides bounds for the matrix norm $\|.\|_q$ in terms of the matrix norms $\|.\|_1$, $\|.\|_2$, $\|.\|_\infty$, and a bound for $\|.\|_2$ in terms of $\|.\|_q$ and $\|.\|_{q/(q-1)}$. For sake of clarity we use notation $\|.\|_2$ for the spectral norm in this lemma, which everywhere else is denoted by $\|.\|$, without any index. Recall that $\|A\|_\infty = \max_i \sum_j |A_{ij}|$ and $\|A\|_1 = \|A'\|_\infty$.

Lemma S.4. *For any matrix A we have*

$$\begin{aligned} \|A\|_q &\le \|A\|_1^{1/q} \|A\|_\infty^{1-1/q}, && \text{for } q \ge 1, \\ \|A\|_q &\le \|A\|_2^{2/q} \|A\|_\infty^{1-2/q}, && \text{for } q \ge 2, \\ \|A\|_2 &\le \sqrt{\|A\|_q \|A\|_{q/(q-1)}}, && \text{for } q \ge 1. \end{aligned}$$

Note also that $\|A\|_{q/(q-1)} = \|A'\|_q$ for $q \ge 1$. Thus, for a symmetric matrix A, we have $\|A\|_2 \le \|A\|_q \le \|A\|_\infty$ for any $q \ge 1$.

Proof of Lemma S.4. The statements follow from the fact that $\log \|A\|_q$ is a convex function of $1/q$, which is a consequence of the Riesz-Thorin theorem. For more details and references see e.g. Higham (1992). ■

The following lemma shows that the norm $\|.\|_q$ applied to higher-dimensional tensors with a special structure can be expressed in terms of matrix norms $\|.\|_q$. In our panel application all higher dimensional tensors have such a special structure, since they are obtained as partial derivatives wrt to α and γ from the likelihood function.

Lemma S.5. *Let a be an N-vector with entries a_i, let b be a T-vector with entries b_t, and let c be an $N \times T$ matrix with entries c_{it}. Let A be an $\underbrace{N \times N \times \ldots \times N}_{p \text{ times}}$ tensor with entries*

$$A_{i_1 i_2 \ldots i_p} = \begin{cases} a_{i_1} & \text{if } i_1 = i_2 = \ldots = i_p, \\ 0 & \text{otherwise.} \end{cases}$$

Let B be an $\underbrace{T \times T \times \ldots \times T}_{r \text{ times}}$ tensor with entries

$$B_{t_1 t_2 \ldots t_r} = \begin{cases} b_{t_1} & \text{if } t_1 = t_2 = \ldots = t_r, \\ 0 & \text{otherwise.} \end{cases}$$

Let C be an $\underbrace{N \times N \times \ldots \times N}_{p \text{ times}} \times \underbrace{T \times T \times \ldots \times T}_{r \text{ times}}$ tensor with entries

$$C_{i_1 i_2 \ldots i_p t_1 t_2 \ldots t_r} = \begin{cases} c_{i_1 t_1} & \text{if } i_1 = i_2 = \ldots = i_p \text{ and } t_1 = t_2 = \ldots = t_r, \\ 0 & \text{otherwise.} \end{cases}$$

Let $\widetilde{C}$ *be an* $\underbrace{T \times T \times \ldots \times T}_{r\ times} \times \underbrace{N \times N \times \ldots \times N}_{p\ times}$ *tensor with entries*

$$\widetilde{C}_{t_1 t_2 \ldots t_r i_1 i_2 \ldots i_p} = \begin{cases} c_{i_1 t_1} & \text{if } i_1 = i_2 = \ldots = i_p \text{ and } t_1 = t_2 = \ldots = t_r, \\ 0 & \text{otherwise.} \end{cases}$$

Then,

$$\begin{aligned}
\|A\|_q &= \max_i |a_i|, && \text{for } p \geq 2, \\
\|B\|_q &= \max_t |b_t|, && \text{for } r \geq 2, \\
\|C\|_q &\leq \|c\|_q, && \text{for } p \geq 1,\ r \geq 1, \\
\|\widetilde{C}\|_q &\leq \|c'\|_q, && \text{for } p \geq 1,\ r \geq 1,
\end{aligned}$$

where $\|.\|_q$ *refers to the q-norm defined in* (A.1) *with* $q \geq 1$.

Proof of Lemma S.5. Since the vector norm $\|.\|_{q/(q-1)}$ is dual to the vector norm $\|.\|_q$ we can rewrite the definition of the tensor norm $\|C\|_q$ as follows

$$\|C\|_q = \max_{\|u^{(1)}\|_{q/(q-1)}=1} \max_{\substack{\|u^{(k)}\|_q = 1 \\ k=2,\ldots,p}} \max_{\substack{\|v^{(l)}\|_q = 1 \\ l=1,\ldots,r}} \left| \sum_{i_1 i_2 \ldots i_p = 1}^{N} \sum_{t_1 t_2 \ldots t_r = 1}^{T} u_{i_1}^{(1)} u_{i_2}^{(2)} \cdots u_{i_p}^{(p)} v_{i_1}^{(1)} v_{t_2}^{(2)} \cdots v_{t_r}^{(r)} C_{i_1 i_2 \ldots i_p t_1 t_2 \ldots t_r} \right|.$$

The specific structure of C yields

$$\begin{aligned}
\|C\|_q &= \max_{\|u^{(1)}\|_{q/(q-1)}=1} \max_{\substack{\|u^{(k)}\|_q = 1 \\ k=2,\ldots,p}} \max_{\substack{\|v^{(l)}\|_q = 1 \\ l=1,\ldots,r}} \left| \sum_{i=1}^{N} \sum_{t=1}^{T} u_i^{(1)} u_i^{(2)} \cdots u_i^{(p)} v_t^{(1)} v_t^{(2)} \cdots v_t^{(r)} c_{it} \right| \\
&\leq \max_{\|u\|_{q/(q-1)} \leq 1} \max_{\|v\|_q \leq 1} \left| \sum_{i=1}^{N} \sum_{t=1}^{T} u_i v_i c_{it} \right| = \|c\|_q,
\end{aligned}$$

where we define $u \in \mathbb{R}^N$ with elements $u_i = u_i^{(1)} u_i^{(2)} \cdots u_i^{(p)}$ and $v \in \mathbb{R}^T$ with elements $v_t = v_t^{(1)} v_t^{(2)} \cdots v_t^{(r)}$, and we use that $\|u^{(k)}\|_q = 1$, for $k = 2, \ldots, p$, and $\|v^{(l)}\|_q = 1$, for $l = 2, \ldots, r$, implies $|u_i| \leq |u_i^{(1)}|$ and $|v_t| \leq |v_t^{(1)}|$, and therefore $\|u\|_{q/(1-q)} \leq \|u^{(1)}\|_{q/(1-q)} = 1$ and $\|v\|_q \leq \|v^{(1)}\|_q = 1$. The proof of $\|\widetilde{C}\|_q \leq \|c'\|_q$ is analogous.

Let $A^{(p)} = A$, as defined above, for a particular value of p. For $p = 2$, $A^{(2)}$ is a diagonal $N \times N$

matrix with diagonal elements a_i, so that $\|A^{(2)}\|_q \leq \|A^{(2)}\|_1^{1/q} \|A^{(2)}\|_\infty^{1-1/q} = \max_i |a_i|$. For $p > 2$,

$$\begin{aligned}
\left\|A^{(p)}\right\|_q &= \max_{\|u^{(1)}\|_{q/(q-1)}=1} \max_{\substack{\|u^{(k)}\|_q = 1 \\ k=2,\ldots,p}} \left| \sum_{i_1 i_2 \ldots i_p = 1}^N u^{(1)}_{i_1} u^{(2)}_{i_2} \cdots u^{(p)}_{i_p} A_{i_1 i_2 \ldots i_p} \right| \\
&= \max_{\|u^{(1)}\|_{q/(q-1)}=1} \max_{\substack{\|u^{(k)}\|_q = 1 \\ k=2,\ldots,p}} \left| \sum_{i,j=1}^N u^{(1)}_i u^{(2)}_i \cdots u^{(p-1)}_i u^{(p)}_j A^{(2)}_{ij} \right| \\
&\leq \max_{\|u\|_{q/(q-1)} \leq 1} \max_{\|v\|_q = 1} \left| \sum_{i=1}^N \sum_{t=1}^T u_i v_i A^{(2)}_{ij} \right| = \|A^{(2)}\|_q \leq \max_i |a_i|,
\end{aligned}$$

where we define $u \in \mathbb{R}^N$ with elements $u_i = u^{(1)}_i u^{(2)}_i \cdots u^{(p-1)}_i$ and $v = u^{(p)}$, and we use that $\|u^{(k)}\|_p = 1$, for $k = 2, \ldots, p-1$, implies $|u_i| \leq |u^{(1)}_i|$ and therefore $\|u\|_{q/(q-1)} \leq \|u^{(1)}\|_{q/(q-1)} = 1$. We have thus shown $\left\|A^{(p)}\right\| \leq \max_i |a_i|$. From the definition of $\left\|A^{(p)}\right\|_q$ above, we obtain $\left\|A^{(p)}\right\|_q \geq \max_i |a_i|$ by choosing all $u^{(k)}$ equal to the standard basis vector, whose i^*'th component equals one, where $i^* \in \operatorname{argmax}_i |a_i|$. Thus, $\left\|A^{(p)}\right\|_q = \max_i |a_i|$ for $p \geq 2$. The proof for $\|B\|_q = \max_t |b_t|$ is analogous. ■

The following lemma provides an asymptotic bound for the spectral norm of $N \times T$ matrices, whose entries are mean zero, and cross-sectionally independent and weakly time-serially dependent conditional on ϕ.

Lemma S.6. *Let e be an $N \times T$ matrix with entries e_{it}. Let $\bar{\sigma}_i^2 = \frac{1}{T}\sum_{t=1}^T \mathbb{E}_\phi(e_{it}^2)$, let Ω be the $T \times T$ matrix with entries $\Omega_{ts} = \frac{1}{N}\sum_{i=1}^N \mathbb{E}_\phi(e_{it}e_{is})$, and let $\eta_{ij} = \frac{1}{\sqrt{T}}\sum_{t=1}^T [e_{it}e_{jt} - \mathbb{E}_\phi(e_{it}e_{jt})]$. Consider asymptotic sequences where $N, T \to \infty$ such that N/T converges to a finite positive constant. Assume that*

(i) The distribution of e_{it} is independent across i, conditional on ϕ, and satisfies $\mathbb{E}_\phi(e_{it}) = 0$.

(ii) $\frac{1}{N}\sum_{i=1}^N \left(\bar{\sigma}_i^2\right)^4 = \mathcal{O}_P(1)$, $\quad \frac{1}{T}\mathrm{Tr}(\Omega^4) = \mathcal{O}_P(1)$, $\quad \frac{1}{N}\sum_{i=1}^N \mathbb{E}_\phi\left(\eta_{ii}^4\right) = \mathcal{O}_P(1)$, $\quad \frac{1}{N^2}\sum_{i,j=1}^N \mathbb{E}_\phi\left(\eta_{ij}^4\right) = \mathcal{O}_P(1)$.

Then, $\mathbb{E}_\phi\|e\|^8 = \mathcal{O}_P(N^5)$, and therefore $\|e\| = \mathcal{O}_P(N^{5/8})$.

Proof of Lemma S.6. Let $\|.\|_{\mathrm{F}}$ be the Frobenius norm of a matrix, i.e. $\|A\|_F = \sqrt{\mathrm{Tr}(AA')}$. For

$\bar{\sigma}_i^4 = (\bar{\sigma}_i^2)^2$, $\bar{\sigma}_i^8 = (\bar{\sigma}_i^2)^4$ and $\delta_{jk} = 1(j = k)$,

$$\begin{aligned}
\|e\|^8 = \|ee'ee'\|^2 \leq \|ee'ee'\|_{\mathrm{F}}^2 &= \sum_{i,j=1}^{N} \left(\sum_{k=1}^{N} \sum_{t,\tau=1}^{T} e_{it} e_{kt} e_{k\tau} e_{j\tau} \right)^2 \\
&= T^2 \sum_{i,j=1}^{N} \left[\sum_{k=1}^{N} \left(\eta_{ik} + T^{1/2} \delta_{ik} \bar{\sigma}_i^2 \right) \left(\eta_{jk} + T^{1/2} \delta_{jk} \bar{\sigma}_j^2 \right) \right]^2 \\
&= T^2 \sum_{i,j=1}^{N} \left(\sum_{k=1}^{N} \eta_{ik} \eta_{jk} + 2T^{1/2} \eta_{ij} \bar{\sigma}_i^2 + T \delta_{ij} \bar{\sigma}_i^4 \right)^2 \\
&\leq 3T^2 \sum_{i,j=1}^{N} \left[\left(\sum_{k=1}^{N} \eta_{ik} \eta_{jk} \right)^2 + 4T \eta_{ij}^2 \bar{\sigma}_i^4 + T^2 \delta_{ij} \bar{\sigma}_i^8 \right] \\
&= 3T^2 \sum_{i,j=1}^{N} \left(\sum_{k=1}^{N} \eta_{ik} \eta_{jk} \right)^2 + 12T^3 \sum_{i,j=1}^{N} \bar{\sigma}_i^4 \eta_{ij}^2 + 3T^3 \sum_{i=1}^{N} \bar{\sigma}_i^8,
\end{aligned}$$

where we used that $(a + b + c)^2 \leq 3(a^2 + b^2 + c^3)$. By the Cauchy Schwarz inequality,

$$\begin{aligned}
\mathbb{E}_\phi \|e\|^8 &\leq 3T^2 \mathbb{E}_\phi \left[\sum_{i,j=1}^{N} \left(\sum_{k=1}^{N} \eta_{ik} \eta_{jk} \right)^2 \right] + 12T^3 \sqrt{ \left(N \sum_{i=1}^{N} \bar{\sigma}_i^8 \right) \left(\sum_{i,j=1}^{N} \mathbb{E}_\phi(\eta_{ij}^4) \right) } + 3T^3 \sum_{i=1}^{N} \bar{\sigma}_i^8 \\
&= 3T^2 \mathbb{E}_\phi \left[\sum_{i,j=1}^{N} \left(\sum_{k=1}^{N} \eta_{ik} \eta_{jk} \right)^2 \right] + \mathcal{O}_P(T^3 N^2) + \mathcal{O}_P(T^3 N).
\end{aligned}$$

Moreover,

$$\begin{aligned}
&\mathbb{E}_\phi \left[\sum_{i,j=1}^{N} \left(\sum_{k=1}^{N} \eta_{ik} \eta_{jk} \right)^2 \right] = \sum_{i,j,k,l=1}^{N} \mathbb{E}_\phi(\eta_{ik} \eta_{jk} \eta_{il} \eta_{jl}) = \sum_{i,j,k,l=1}^{N} \mathbb{E}_\phi(\eta_{ij} \eta_{jk} \eta_{kl} \eta_{li}) \\
&\leq \left| \sum_{\substack{i,j,k,l \\ \text{mutually different}}} \mathbb{E}_\phi \left(\eta_{ij} \eta_{jk} \eta_{kl} \eta_{li} \right) \right| + 4 \left| \sum_{i,j,k=1}^{N} a_{ijk} \mathbb{E}_\phi(\eta_{ii} \eta_{ij} \eta_{jk} \eta_{ki}) \right|, \\
&\leq \left| \sum_{\substack{i,j,k,l \\ \text{mutually different}}} \mathbb{E}_\phi \left(\eta_{ij} \eta_{jk} \eta_{kl} \eta_{li} \right) \right| + 4 \left\{ \left[\sum_{i,j,k=1}^{N} \mathbb{E}_\phi(\eta_{ii}^4) \right] \left[\sum_{i,j,k=1}^{N} \mathbb{E}_\phi(\eta_{ij}^4) \right]^3 \right\}^{1/4} \\
&= \left| \sum_{\substack{i,j,k,l \\ \text{mutually different}}} \mathbb{E}_\phi \left(\eta_{ij} \eta_{jk} \eta_{kl} \eta_{li} \right) \right| + 4N^3 \left\{ \left[\frac{1}{N} \sum_{i=1}^{N} \mathbb{E}_\phi(\eta_{ii}^4) \right] \left[\frac{1}{N^2} \sum_{i,j=1}^{N} \mathbb{E}_\phi(\eta_{ij}^4) \right]^3 \right\}^{1/4} \\
&= \left| \sum_{\substack{i,j,k,l \\ \text{mutually different}}} \mathbb{E}_\phi \left(\eta_{ij} \eta_{jk} \eta_{kl} \eta_{li} \right) \right| + \mathcal{O}_P(N^3).
\end{aligned}$$

where in the second step we just renamed the indices and used that η_{ij} is symmetric in i, j; and $a_{ijk} \in [0, 1]$ in the second line is a combinatorial pre-factor; and in the third step we applied the Cauchy-Schwarz inequality.

Let Ω_i be the $T \times T$ matrix with entries $\Omega_{i,ts} = \mathbb{E}_\phi(e_{it}e_{is})$ such that $\Omega = \frac{1}{N}\sum_{i=1}^N \Omega_i$. For i,j,k,l mutually different,

$$\mathbb{E}_\phi\left(\eta_{ij}\eta_{jk}\eta_{kl}\eta_{li}\right) = \frac{1}{T^2}\sum_{t,s,u,v=1}^T \mathbb{E}_\phi(e_{it}e_{jt}e_{js}e_{ks}e_{ku}e_{lu}e_{lv}e_{iv})$$
$$= \frac{1}{T^2}\sum_{t,s,u,v=1}^T \mathbb{E}_\phi(e_{iv}e_{it})\mathbb{E}_\phi(e_{jt}e_{js})\mathbb{E}_\phi(e_{ks}e_{ku})\mathbb{E}_\phi(e_{lu}e_{lv}) = \frac{1}{T^2}\mathrm{Tr}(\Omega_i\Omega_j\Omega_k\Omega_l) \geq 0$$

because $\Omega_i \geq 0$ for all i. Thus,

$$\left|\sum_{\substack{i,j,k,l \\ \text{mutually different}}} \mathbb{E}_\phi\left(\eta_{ij}\eta_{jk}\eta_{kl}\eta_{li}\right)\right| = \sum_{\substack{i,j,k,l \\ \text{mutually different}}} \mathbb{E}_\phi\left(\eta_{ij}\eta_{jk}\eta_{kl}\eta_{li}\right) = \frac{1}{T^2}\sum_{\substack{i,j,k,l \\ \text{mut. different}}} \mathrm{Tr}(\Omega_i\Omega_j\Omega_k\Omega_l)$$
$$\leq \frac{1}{T^2}\sum_{i,j,k,l=1}^N \mathrm{Tr}(\Omega_i\Omega_j\Omega_k\Omega_l) = \frac{N^4}{T^2}\mathrm{Tr}(\Omega^4) = \mathcal{O}_P(N^4/T).$$

Combining all the above results gives $\mathbb{E}_\phi\|e\|^8 = \mathcal{O}_P(N^5)$, since N and T are assumed to grow at the same rate. ■

S.6.3 Verifying the Basic Regularity Conditions in Panel Models

The following Lemma provides sufficient conditions under which the panel fixed effects estimators in the main text satisfy the high-level regularity conditions in Assumptions B.1(v) and (vi).

Lemma S.7. *Let $\mathcal{L}(\beta,\phi) = \frac{1}{\sqrt{NT}}\left[\sum_{i,t}\ell_{it}(\beta,\pi_{it}) - \frac{b}{2}(v'\phi)^2\right]$, where $\pi_{it} = \alpha_i + \gamma_t$, $\alpha = (\alpha_1,\ldots,\alpha_N)'$, $\gamma = (\gamma_1,\ldots,\gamma_T)$, $\phi = (\alpha',\gamma')'$, and $v = (1_N', 1_T')'$. Assume that $\ell_{it}(.,.)$ is four times continuously differentiable in an appropriate neighborhood of the true parameter values (β^0,ϕ^0). Consider limits as $N,T\to\infty$ with $N/T \to \kappa^2 > 0$. Let $4 < q \leq 8$ and $0 \leq \epsilon < 1/8 - 1/(2q)$. Let $r_\beta = r_{\beta,NT} > 0$, $r_\phi = r_{\phi,NT} > 0$, with $r_\beta = o\left[(NT)^{-1/(2q)-\epsilon}\right]$ and $r_\phi = o\left[(NT)^{-\epsilon}\right]$. Assume that*

(i) For $k,l,m \in \{1,2,\ldots,\dim\beta\}$,

$$\frac{1}{\sqrt{NT}}\sum_{i,t}\partial_{\beta_k}\ell_{it} = \mathcal{O}_P(1), \quad \frac{1}{NT}\sum_{i,t}\partial_{\beta_k\beta_l}\ell_{it} = \mathcal{O}_P(1), \quad \frac{1}{NT}\sum_{i,t}\left\{\partial_{\beta_k\beta_l}\ell_{it} - \mathbb{E}_\phi\left[\partial_{\beta_k\beta_l}\ell_{it}\right]\right\} = o_P(1),$$
$$\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)}\ \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)} \frac{1}{NT}\sum_{i,t}\partial_{\beta_k\beta_l\beta_m}\ell_{it}(\beta,\pi_{it}) = \mathcal{O}_P(1).$$

(ii) Let $k,l \in \{1,2,\ldots,\dim\beta\}$. For $\xi_{it}(\beta,\phi) = \partial_{\beta_k\pi}\ell_{it}(\beta,\pi_{it})$ or $\xi_{it}(\beta,\phi) = \partial_{\beta_k\beta_l\pi}\ell_{it}(\beta,\pi_{it})$,

$$\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)}\ \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)} \frac{1}{T}\sum_t\left|\frac{1}{N}\sum_i \xi_{it}(\beta,\phi)\right|^q = \mathcal{O}_P(1),$$
$$\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)}\ \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)} \frac{1}{N}\sum_i\left|\frac{1}{T}\sum_t \xi_{it}(\beta,\phi)\right|^q = \mathcal{O}_P(1).$$

(iii) *Let $k, l \in \{1,2,\ldots,\dim\beta\}$. For $\xi_{it}(\beta,\phi) = \partial_{\pi^r}\ell_{it}(\beta,\pi_{it})$, with $r \in \{3,4\}$, or $\xi_{it}(\beta,\phi) = \partial_{\beta_k\pi^r}\ell_{it}(\beta,\pi_{it})$, with $r \in \{2,3\}$, or $\xi_{it}(\beta,\phi) = \partial_{\beta_k\beta_l\pi^2}\ell_{it}(\beta,\pi_{it})$,*

$$\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)} \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)} \max_i \frac{1}{T}\sum_t |\xi_{it}(\beta,\phi)| = \mathcal{O}_P\left(N^{2\epsilon}\right),$$
$$\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)} \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)} \max_t \frac{1}{N}\sum_i |\xi_{it}(\beta,\phi)| = \mathcal{O}_P\left(N^{2\epsilon}\right).$$

(iv) *Moreover,*

$$\frac{1}{T}\sum_t \left|\frac{1}{\sqrt{N}}\sum_i \partial_\pi \ell_{it}\right|^q = \mathcal{O}_P(1), \quad \frac{1}{N}\sum_i \left|\frac{1}{\sqrt{T}}\sum_t \partial_\pi \ell_{it}\right|^q = \mathcal{O}_P(1),$$
$$\frac{1}{T}\sum_t \left|\frac{1}{\sqrt{N}}\sum_i \partial_{\beta_k\pi}\ell_{it} - \mathbb{E}_\phi\left[\partial_{\beta_k\pi}\ell_{it}\right]\right|^2 = \mathcal{O}_P(1),$$
$$\frac{1}{N}\sum_i \left|\frac{1}{\sqrt{T}}\sum_t \partial_{\beta_k\pi}\ell_{it} - \mathbb{E}_\phi\left[\partial_{\beta_k\pi}\ell_{it}\right]\right|^2 = \mathcal{O}_P(1).$$

(v) *The sequence $\{(\ell_{i1},\ldots,\ell_{iT}) : 1 \le i \le N\}$ is independent across i conditional on ϕ.*

(vi) *Let $k \in \{1,2,\ldots,\dim\beta\}$. For $\xi_{it} = \partial_{\pi^r}\ell_{it} - \mathbb{E}_\phi\left[\partial_{\pi^r}\ell_{it}\right]$, with $r \in \{2,3\}$, or $\xi_{it} = \partial_{\beta_k\pi^2}\ell_{it} - \mathbb{E}_\phi\left[\partial_{\beta_k\pi^2}\ell_{it}\right]$, and some $\tilde{\nu} > 0$,*

$$\max_i \mathbb{E}_\phi\left[\xi_{it}^{8+\tilde{\nu}}\right] \le C, \quad \max_i \max_t \sum_s \mathbb{E}_\phi\left[\xi_{it}\xi_{is}\right] \le C, \quad \max_i \mathbb{E}_\phi\left\{\left[\frac{1}{\sqrt{T}}\sum_t \xi_{it}\right]^8\right\} \le C,$$
$$\max_t \mathbb{E}_\phi\left\{\left[\frac{1}{\sqrt{N}}\sum_i \xi_{it}\right]^8\right\} \le C, \quad \max_{i,j} \mathbb{E}_\phi\left\{\left[\frac{1}{\sqrt{T}}\sum_t \left[\xi_{it}\xi_{jt} - \mathbb{E}_\phi\left(\xi_{it}\xi_{jt}\right)\right]\right]^4\right\} \le C,$$

uniformly in N, T, where $C > 0$ is a constant.

(vii) $\left\|\overline{\mathcal{H}}^{-1}\right\|_q = \mathcal{O}_P(1)$.

Then, Assumptions B.1(v) and (vi) are satisfied with the same parameters q, ϵ, $r_\beta = r_{\beta,NT}$ and $r_\phi = r_{\phi,NT}$ used here.

Proof of Lemma S.7. The penalty term $(v'\phi)^2$ is quadratic in ϕ and does not depend on β. This term thus only enters $\partial_\phi\mathcal{L}(\beta,\phi)$ and $\partial_{\phi\phi'}\mathcal{L}(\beta,\phi)$, but it does not effect any other partial derivative of $\mathcal{L}(\beta,\phi)$. Furthermore, the contribution of the penalty drops out of $\mathcal{S} = \partial_\phi\mathcal{L}(\beta^0,\phi^0)$, because we impose the normalization $v'\phi^0 = 0$. It also drops out of $\widetilde{\mathcal{H}}$, because it contributes the same to $\mathcal{H}$ and $\overline{\mathcal{H}}$. We can therefore ignore the penalty term for the purpose of proving the lemma (but it is necessary to satisfy the assumption $\left\|\overline{\mathcal{H}}^{-1}\right\|_q = \mathcal{O}_P(1)$).

Assumption (i) implies that $\|\partial_\beta\mathcal{L}\| = \mathcal{O}_P(1)$, $\|\partial_{\beta\beta'}\mathcal{L}\| = \mathcal{O}_P(\sqrt{NT})$, $\left\|\partial_{\beta\beta'}\widetilde{\mathcal{L}}\right\| = o_P(\sqrt{NT})$, and $\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)} \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)} \|\partial_{\beta\beta\beta}\mathcal{L}(\beta,\,\phi)\| = \mathcal{O}_P\left(\sqrt{NT}\right)$. Note that it does not matter which norms we use here because $\dim\beta$ is fixed.

By Assumption (ii), $\|\partial_{\beta\phi'}\mathcal{L}\|_q = \mathcal{O}_P\left((NT)^{1/(2q)}\right)$ and $\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)} \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)} \|\partial_{\beta\beta\phi}\mathcal{L}(\beta,\,\phi)\|_q = \mathcal{O}_P\left((NT)^{1/(2q)}\right)$. For example, $\partial_{\beta_k\alpha_i}\mathcal{L} = \frac{1}{\sqrt{NT}}\sum_t \partial_{\beta_k\pi}\ell_{it}$ and therefore

$$\|\partial_{\beta_k\alpha}\mathcal{L}\|_q = \left(\sum_i \left|\frac{1}{\sqrt{NT}}\sum_t \partial_{\beta_k\pi}\ell_{it}\right|^q\right)^{1/q} = \mathcal{O}_P\left(N^{1/q}\right) = \mathcal{O}_P\left((NT)^{1/(2q)}\right).$$

Analogously, $\|\partial_{\beta_k\gamma}\mathcal{L}\|_q = \mathcal{O}_P\left((NT)^{1/(2q)}\right)$, and therefore $\|\partial_{\beta_k\phi}\mathcal{L}\|_q \le \|\partial_{\beta_k\alpha}\mathcal{L}\|_q + \|\partial_{\beta_k\gamma}\mathcal{L}\|_q = \mathcal{O}_P\left((NT)^{1/(2q)}\right)$. This also implies that $\|\partial_{\beta\phi'}\mathcal{L}\|_q = \mathcal{O}_P\left((NT)^{1/(2q)}\right)$ because $\dim\beta$ is fixed.

By Assumption (iii), $\|\partial_{\phi\phi\phi}\mathcal{L}\|_q = \mathcal{O}_P\left((NT)^{\epsilon}\right)$, $\|\partial_{\beta\phi\phi}\mathcal{L}\|_q = \mathcal{O}_P((NT)^{\epsilon})$, $\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)} \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)} \|\partial_{\beta\beta\phi\phi}\mathcal{L}(\beta,\,\phi)\|_q = \mathcal{O}_P\left((NT)^{\epsilon}\right)$, $\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)} \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)} \|\partial_{\beta\phi\phi\phi}\mathcal{L}(\beta,\,\phi)\|_q = \mathcal{O}_P\left((NT)^{\epsilon}\right)$, and $\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)} \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)} \|\partial_{\phi\phi\phi\phi}\mathcal{L}(\beta,\phi)\|_q = \mathcal{O}_P\left((NT)^{\epsilon}\right)$. For example,

$$\begin{aligned}
\|\partial_{\phi\phi\phi}\mathcal{L}\|_q &\le \|\partial_{\alpha\alpha\alpha}\mathcal{L}\|_q + \|\partial_{\alpha\alpha\gamma}\mathcal{L}\|_q + \|\partial_{\alpha\gamma\alpha}\mathcal{L}\|_q + \|\partial_{\alpha\gamma\gamma}\mathcal{L}\|_q \\
&\qquad + \|\partial_{\gamma\alpha\alpha}\mathcal{L}\|_q + \|\partial_{\gamma\alpha\gamma}\mathcal{L}\|_q + \|\partial_{\gamma\gamma\alpha}\mathcal{L}\|_q + \|\partial_{\gamma\gamma\gamma}\mathcal{L}\|_q \\
&\le \|\partial_{\pi\alpha\alpha}\mathcal{L}\|_q + \|\partial_{\pi\gamma\gamma}\mathcal{L}\|_q + 3\,\|\partial_{\pi\alpha\gamma}\mathcal{L}\|_q + 3\,\|\partial_{\pi\gamma\alpha}\mathcal{L}\|_q \\
&\le \|\partial_{\pi\alpha\alpha}\mathcal{L}\|_\infty + \|\partial_{\pi\gamma\gamma}\mathcal{L}\|_\infty + 3\,\|\partial_{\pi\alpha\gamma}\mathcal{L}\|_\infty^{1-1/q}\,\|\partial_{\pi\gamma\alpha}\mathcal{L}\|_\infty^{1/q} + 3\,\|\partial_{\pi\alpha\gamma}\mathcal{L}\|_\infty^{1/q}\,\|\partial_{\pi\gamma\alpha}\mathcal{L}\|_\infty^{1-1/q} \\
&= \frac{1}{\sqrt{NT}}\left[\max_i\left|\sum_t \partial_{\pi^3}\ell_{it}\right| + \max_t\left|\sum_i \partial_{\pi^3}\ell_{it}\right| + 3\left(\max_i\sum_t |\partial_{\pi^3}\ell_{it}|\right)^{1-1/q}\left(\max_t\sum_t |\partial_{\pi^3}\ell_{it}|\right)^{1/q}\right. \\
&\qquad\left. + 3\left(\max_i\sum_t |\partial_{\pi^3}\ell_{it}|\right)^{1/q}\left(\max_t\sum_t |\partial_{\pi^3}\ell_{it}|\right)^{1-1/q}\right] \\
&\le \frac{1}{\sqrt{NT}}\left[\max_i\sum_t |\partial_{\pi^3}\ell_{it}| + \max_t\sum_i |\partial_{\pi^3}\ell_{it}| + 3\left(\max_i\sum_t |\partial_{\pi^3}\ell_{it}|\right)^{1-1/q}\left(\max_t\sum_t |\partial_{\pi^3}\ell_{it}|\right)^{1/q}\right. \\
&\qquad\left. + 3\left(\max_i\sum_t |\partial_{\pi^3}\ell_{it}|\right)^{1/q}\left(\max_t\sum_t |\partial_{\pi^3}\ell_{it}|\right)^{1-1/q}\right] = \mathcal{O}_P(N^{2\epsilon}) = \mathcal{O}_P((NT)^{\epsilon}).
\end{aligned}$$

Here, we use Lemma S.5 to bound the norms of the 3-tensors in terms of the norms of matrices, e.g. $\|\partial_{\alpha\alpha\gamma}\mathcal{L}\|_q \le \|\partial_{\pi\alpha\gamma}\mathcal{L}\|_q$, because $\partial_{\alpha_i\alpha_j\gamma_t}\mathcal{L} = 0$ if $i \neq j$ and $\partial_{\alpha_i\alpha_i\gamma_t}\mathcal{L} = (NT)^{-1/2}\partial_{\pi\alpha_i\gamma_t}$.[8] Then, we use Lemma S.4 to bound q-norms in terms of ∞-norms, and then explicitly expressed those ∞-norm in terms of the elements of the matrices. Finally, we use that $|\sum_i \partial_{\pi^3}\ell_{it}| \le \sum_i |\partial_{\pi^3}\ell_{it}|$ and $|\sum_t \partial_{\pi^3}\ell_{it}| \le \sum_t |\partial_{\pi^3}\ell_{it}|$, and apply Assumption (iii).

By Assumption (iv), $\|\mathcal{S}\|_q = \mathcal{O}_P\left((NT)^{-1/4+1/(2q)}\right)$ and $\left\|\partial_{\beta\phi'}\widetilde{\mathcal{L}}\right\| = \mathcal{O}_P(1)$. For example,

$$\|\mathcal{S}\|_q = \frac{1}{\sqrt{NT}}\left(\sum_i\left|\sum_t \partial_\pi\ell_{it}\right|^q + \sum_t\left|\sum_i \partial_\pi\ell_{it}\right|^q\right)^{1/q} = \mathcal{O}_P\left(N^{-1/2+1/q}\right) = \mathcal{O}_P\left((NT)^{-1/4+1/(2q)}\right).$$

By Assumption (v) and (vi), $\|\widetilde{\mathcal{H}}\| = \mathcal{O}_P\left((NT)^{-3/16}\right) = o_P\left((NT)^{-1/8}\right)$ and $\left\|\partial_{\beta\phi\phi}\widetilde{\mathcal{L}}\right\| = \mathcal{O}_P\left((NT)^{-3/16}\right) = o_P\left((NT)^{-1/8}\right)$. We now show it $\|\widetilde{\mathcal{H}}\|$. The proof for $\left\|\partial_{\beta\phi\phi}\widetilde{\mathcal{L}}\right\|$ is analogous.

[8] With a slight abuse of notation we write $\partial_{\pi\alpha\gamma}\mathcal{L}$ for the $N\times T$ matrix with entries $(NT)^{-1/2}\partial_{\pi^3}\ell_{it} = (NT)^{-1/2}\partial_{\pi^3}\ell_{it}$, and analogously for $\partial_{\pi\alpha\alpha}\mathcal{L}$, $\partial_{\pi\gamma\gamma}\mathcal{L}$, and $\partial_{\pi\gamma\alpha}\mathcal{L}$.

By the triangle inequality,

$$\|\widetilde{\mathcal{H}}\| = \|\partial_{\phi\phi'}\mathcal{L} - \mathbb{E}_\phi\left[\partial_{\phi\phi'}\mathcal{L}\right]\| \leq \|\partial_{\alpha\alpha'}\mathcal{L} - \mathbb{E}_\phi\left[\partial_{\alpha\alpha'}\mathcal{L}\right]\| + \|\partial_{\gamma\gamma'}\mathcal{L} - \mathbb{E}_\phi\left[\partial_{\gamma\gamma'}\mathcal{L}\right]\| + 2\left\|\partial_{\alpha\gamma'}\mathcal{L} - \mathbb{E}_\phi\left[\partial_{\alpha\gamma'}\mathcal{L}\right]\right\|.$$

Let $\xi_{it} = \partial_{\pi^2}\ell_{it} - \mathbb{E}_\phi\left[\partial_{\pi^2}\ell_{it}\right]$. Since $\partial_{\alpha\alpha'}\mathcal{L}$ is a diagonal matrix with diagonal entries $\frac{1}{\sqrt{NT}}\sum_t \xi_{it}$, $\|\partial_{\alpha\alpha'}\mathcal{L} - \mathbb{E}_\phi\left[\partial_{\alpha\alpha'}\mathcal{L}\right]\| = \max_i \frac{1}{\sqrt{NT}}\sum_t \xi_{it}$, and therefore

$$\mathbb{E}_\phi \|\partial_{\alpha\alpha'}\mathcal{L} - \mathbb{E}_\phi\left[\partial_{\alpha\alpha'}\mathcal{L}\right]\|^8 = \mathbb{E}_\phi\left[\max_i\left(\frac{1}{\sqrt{NT}}\sum_t \xi_{it}\right)^8\right]$$
$$\leq \mathbb{E}_\phi\left[\sum_i\left(\frac{1}{\sqrt{NT}}\sum_t \xi_{it}\right)^8\right] \leq CN\left(\frac{1}{\sqrt{N}}\right)^8 = \mathcal{O}_P(N^{-3}).$$

Thus, $\|\partial_{\alpha\alpha'}\mathcal{L} - \mathbb{E}_\phi\left[\partial_{\alpha\alpha'}\mathcal{L}\right]\| = \mathcal{O}_P(N^{-3/8})$. Analogously, $\|\partial_{\gamma\gamma'}\mathcal{L} - \mathbb{E}_\phi\left[\partial_{\gamma\gamma'}\mathcal{L}\right]\| = \mathcal{O}_P(N^{-3/8})$.

Let ξ be the $N\times T$ matrix with entries ξ_{it}. We now show that ξ satisfies all the regularity condition of Lemma S.6 with $e_{it} = \xi_{it}$. Independence across i is assumed. Furthermore, $\bar{\sigma}_i^2 = \frac{1}{T}\sum_{t=1}^T \mathbb{E}_\phi(\xi_{it}^2) \leq C^{1/4}$ so that $\frac{1}{N}\sum_{i=1}^N \left(\bar{\sigma}_i^2\right)^4 = \mathcal{O}_P(1)$. For $\Omega_{ts} = \frac{1}{N}\sum_{i=1}^N \mathbb{E}_\phi(\xi_{it}\xi_{is})$,

$$\frac{1}{T}\mathrm{Tr}(\Omega^4) \leq \|\Omega\|^4 \leq \|\Omega\|_\infty^4 = \left(\max_t \sum_s \mathbb{E}_\phi\left[\xi_{it}\xi_{is}\right]\right)^4 \leq C = \mathcal{O}_P(1).$$

For $\eta_{ij} = \frac{1}{\sqrt{T}}\sum_{t=1}^T \left[\xi_{it}\xi_{jt} - \mathbb{E}_\phi(\xi_{it}\xi_{jt})\right]$ we assume $\mathbb{E}_\phi \eta_{ij}^4 \leq C$, which implies $\frac{1}{N}\sum_{i=1}^N \mathbb{E}_\phi\left(\eta_{ii}^4\right) = \mathcal{O}_P(1)$ and $\frac{1}{N^2}\sum_{i,j=1}^N \mathbb{E}_\phi\left(\eta_{ij}^4\right) = \mathcal{O}_P(1)$. Then, Lemma S.6 gives $\|\xi\| = \mathcal{O}_P(N^{5/8})$. Note that $\xi = \frac{1}{\sqrt{NT}}\partial_{\alpha\gamma'}\mathcal{L} - \mathbb{E}_\phi\left[\partial_{\alpha\gamma'}\mathcal{L}\right]$ and therefore $\|\partial_{\alpha\gamma'}\mathcal{L} - \mathbb{E}_\phi\left[\partial_{\alpha\gamma'}\mathcal{L}\right]\| = \mathcal{O}_P(N^{-3/8})$. We conclude that $\|\widetilde{\mathcal{H}}\| = \mathcal{O}_P(N^{-3/8}) = \mathcal{O}_P\left((NT)^{-3/16}\right)$.

Moreover, for $\xi_{it} = \partial_{\pi^2}\ell_{it} - \mathbb{E}_\phi\left[\partial_{\pi^2}\ell_{it}\right]$

$$\mathbb{E}_\phi \|\widetilde{\mathcal{H}}\|_\infty^{8+\tilde{\nu}} = \mathbb{E}_\phi\left(\frac{1}{\sqrt{NT}}\max_i \sum_t |\xi_{it}|\right)^{8+\tilde{\nu}} = \mathbb{E}_\phi \max_i\left(\frac{1}{\sqrt{NT}}\sum_t |\xi_{it}|\right)^{8+\tilde{\nu}}$$
$$\leq \mathbb{E}_\phi \sum_i\left(\frac{1}{\sqrt{NT}}\sum_t |\xi_{it}|\right)^{8+\tilde{\nu}} \leq \mathbb{E}_\phi \sum_i\left(\frac{T}{\sqrt{NT}}\right)^{8+\tilde{\nu}}\left(\frac{1}{T}\sum_t |\xi_{it}|^{8+\tilde{\nu}}\right) = \mathcal{O}_P(N),$$

and therefore $\|\widetilde{\mathcal{H}}\|_\infty = o_P(N^{1/8})$. Thus, by Lemma S.4

$$\|\widetilde{\mathcal{H}}\|_q \leq \|\widetilde{\mathcal{H}}\|_2^{2/q}\|\widetilde{\mathcal{H}}\|_\infty^{1-2/q} = o_P\left(N^{1/8[-6/q+(1-2/q)]}\right) = o_P\left(N^{-1/q+1/8}\right) = o_P(1),$$

where we use that $q \leq 8$.

Finally we show that $\left\|\sum_{g,h=1}^{\dim\phi} \partial_{\phi\phi_g\phi_h}\widetilde{\mathcal{L}}\,[\overline{\mathcal{H}}^{-1}\mathcal{S}]_g[\overline{\mathcal{H}}^{-1}\mathcal{S}]_h\right\| = o_P\left((NT)^{-1/4}\right)$. First,

$$\left\|\sum_{g,h=1}^{\dim\phi} \partial_{\phi\phi_g\phi_h}\widetilde{\mathcal{L}}\,[\overline{\mathcal{H}}^{-1}\mathcal{S}]_g[\overline{\mathcal{H}}^{-1}\mathcal{S}]_h\right\|$$
$$\leq \left\|\sum_{g,h=1}^{\dim\phi} \partial_{\alpha\phi_g\phi_h}\widetilde{\mathcal{L}}\,[\overline{\mathcal{H}}^{-1}\mathcal{S}]_g[\overline{\mathcal{H}}^{-1}\mathcal{S}]_h\right\| + \left\|\sum_{g,h=1}^{\dim\phi} \partial_{\gamma\phi_g\phi_h}\widetilde{\mathcal{L}}\,[\overline{\mathcal{H}}^{-1}\mathcal{S}]_g[\overline{\mathcal{H}}^{-1}\mathcal{S}]_h\right\|.$$

Let $(v,w)' := \overline{\mathcal{H}}^{-1}\mathcal{S}$, where v is a N-vector and w is a T-vector. We assume $\left\|\overline{\mathcal{H}}^{-1}\right\|_q = \mathcal{O}_P(1)$. By Lemma S.1 this also implies $\left\|\overline{\mathcal{H}}^{-1}\right\| = \mathcal{O}_P(1)$ and $\|\mathcal{S}\| = \mathcal{O}_P(1)$. Thus, $\|v\| \leq \left\|\overline{\mathcal{H}}^{-1}\right\| \|\mathcal{S}\| = \mathcal{O}_P(1)$, $\|w\| \leq \left\|\overline{\mathcal{H}}^{-1}\right\| \|\mathcal{S}\| = \mathcal{O}_P(1)$, $\|v\|_\infty \leq \|v\|_q \leq \left\|\overline{\mathcal{H}}^{-1}\right\|_q \|\mathcal{S}\|_q = \mathcal{O}_P\left((NT)^{-1/4+1/(2q)}\right)$, $\|w\|_\infty \leq \|w\|_q \leq \left\|\overline{\mathcal{H}}^{-1}\right\|_q \|\mathcal{S}\|_q = \mathcal{O}_P\left((NT)^{-1/4+1/(2q)}\right)$. Furthermore, by an analogous argument to the above proof for $\|\widetilde{\mathcal{H}}\|$, Assumption (v) and (vi) imply that $\left\|\partial_{\pi\alpha\alpha'}\widetilde{\mathcal{L}}\right\| = \mathcal{O}_P(N^{-3/8})$, $\left\|\partial_{\pi\alpha\gamma'}\widetilde{\mathcal{L}}\right\| = \mathcal{O}_P(N^{-3/8})$, $\left\|\partial_{\pi\gamma\gamma'}\widetilde{\mathcal{L}}\right\| = \mathcal{O}_P(N^{-3/8})$. Then,

$$\begin{aligned}\sum_{g,h=1}^{\dim\phi} \partial_{\alpha_i\phi_g\phi_h}\widetilde{\mathcal{L}}\,[\overline{\mathcal{H}}^{-1}\mathcal{S}]_g[\overline{\mathcal{H}}^{-1}\mathcal{S}]_h &= \sum_{j,k=1}^{N}(\partial_{\alpha_i\alpha_j\alpha_k}\widetilde{\mathcal{L}})v_jv_k + 2\sum_{j=1}^{N}\sum_{t=1}^{T}(\partial_{\alpha_i\alpha_j\gamma_t}\widetilde{\mathcal{L}})v_jw_t + \sum_{t,s=1}^{T}(\partial_{\alpha_i\gamma_t\gamma_s}\widetilde{\mathcal{L}})w_tw_s \\ &= \sum_{j=1}^{N}(\partial_{\pi^2\alpha_i}\widetilde{\mathcal{L}})v_i^2 + 2\sum_{t=1}^{T}(\partial_{\pi\alpha_i\gamma_t}\widetilde{\mathcal{L}})v_iw_t + \sum_{t=1}^{T}(\partial_{\pi\alpha_i\gamma_t}\widetilde{\mathcal{L}})w_t^2,\end{aligned}$$

and therefore

$$\begin{aligned}\left\|\sum_{g,h=1}^{\dim\phi} \partial_{\alpha\phi_g\phi_h}\widetilde{\mathcal{L}}\,[\overline{\mathcal{H}}^{-1}\mathcal{S}]_g[\overline{\mathcal{H}}^{-1}\mathcal{S}]_h\right\| &\leq \left\|\partial_{\pi\alpha\alpha'}\widetilde{\mathcal{L}}\right\|\|v\|\|v\|_\infty + 2\left\|\partial_{\pi\alpha\gamma'}\widetilde{\mathcal{L}}\right\|\|w\|\|v\|_\infty + \left\|\partial_{\pi\alpha\gamma'}\widetilde{\mathcal{L}}\right\|\|w\|\|w\|_\infty \\ &= \mathcal{O}_P(N^{-3/8})\mathcal{O}_P\left((NT)^{-1/4+1/(2q)}\right) = \mathcal{O}_P\left((NT)^{-1/4-3/16+1/(2q)}\right) = o_P\left((NT)^{-1/4}\right),\end{aligned}$$

where we use that $q > 4$. Analogously, $\left\|\sum_{g,h=1}^{\dim\phi}\partial_{\gamma\phi_g\phi_h}\widetilde{\mathcal{L}}\,[\overline{\mathcal{H}}^{-1}\mathcal{S}]_g[\overline{\mathcal{H}}^{-1}\mathcal{S}]_h\right\| = o_P\left((NT)^{-1/4}\right)$ and thus also $\left\|\sum_{g,h=1}^{\dim\phi}\partial_{\phi\phi_g\phi_h}\widetilde{\mathcal{L}}\,[\overline{\mathcal{H}}^{-1}\mathcal{S}]_g[\overline{\mathcal{H}}^{-1}\mathcal{S}]_h\right\| = o_P\left((NT)^{-1/4}\right)$.[9] ■

S.6.4 A Useful Algebraic Result

Let $\widetilde{\mathbb{P}}$ be the linear operator defined in equation (S.2), and and let $\mathbb{P}$ be the related projection operator defined in (S.1). Lemma S.8 shows how in the context of panel data models some expressions that appear in the general expansion of Appendix B can be conveniently expressed using the operator $\widetilde{\mathbb{P}}$. This lemma is used extensively in the proof of part (ii) of Theorem C.1.

Lemma S.8. *Let A, B and C be $N \times T$ matrices, and let the expected incidental parameter Hessian $\overline{\mathcal{H}}$ be invertible. Define the $N+T$ vectors $\mathcal{A}$ and $\mathcal{B}$ and the $(N+T)\times(N+T)$ matrix $\mathcal{C}$ as follows*[10]

$$\mathcal{A} = \frac{1}{NT}\begin{pmatrix} A1_T \\ A'1_N\end{pmatrix}, \qquad \mathcal{B} = \frac{1}{NT}\begin{pmatrix} B1_T \\ B'1_N\end{pmatrix}, \qquad \mathcal{C} = \frac{1}{NT}\begin{pmatrix} \operatorname{diag}(C1_T) & C \\ C' & \operatorname{diag}(C'1_N)\end{pmatrix}.$$

Then,

[9] Given the structure of this last part of the proof of Lemma S.7 one might wonder why, instead of $\left\|\sum_{g,h=1}^{\dim\phi}\partial_{\phi\phi_g\phi_h}\widetilde{\mathcal{L}}\,[\overline{\mathcal{H}}^{-1}\mathcal{S}]_g[\overline{\mathcal{H}}^{-1}\mathcal{S}]_h\right\| = o_P\left((NT)^{-1/4}\right)$, we did not directly impose $\sum_g\left\|\partial_{\phi_g\phi\phi'}\widetilde{\mathcal{L}}\right\| = o_P\left((NT)^{-1/(2q)}\right)$ as a high-level condition in Assumption B.1(vi). While this alternative high-level assumption would indeed be more elegant and sufficient to derive our results, it would not be satisfied for panel models, because it involves bounding $\sum_i\left\|\partial_{\alpha_i\gamma\gamma'}\widetilde{\mathcal{L}}\right\|$ and $\sum_t\left\|\partial_{\gamma_t\alpha\alpha'}\widetilde{\mathcal{L}}\right\|$, which was avoided in the proof of Lemma S.7.

[10] Note that $A1_T$ is simply the N-vectors with entries $\sum_t A_{it}$ and $A'1_N$ is simply the T-vector with entries $\sum_i A_{it}$, and analogously for B and C.

$$(i) \qquad \mathcal{A}' \, \overline{\mathcal{H}}^{-1} \, \mathcal{B} = \frac{1}{(NT)^{3/2}} \sum_{i,t} (\widetilde{\mathbb{P}} A)_{it} B_{it} = \frac{1}{(NT)^{3/2}} \sum_{i,t} (\widetilde{\mathbb{P}} B)_{it} A_{it},$$

$$(ii) \qquad \mathcal{A}' \, \overline{\mathcal{H}}^{-1} \, \mathcal{B} = \frac{1}{(NT)^{3/2}} \sum_{i,t} \mathbb{E}_\phi(-\partial_{\pi^2} \ell_{it}) (\widetilde{\mathbb{P}} A)_{it} (\widetilde{\mathbb{P}} B)_{it},$$

$$(iii) \qquad \mathcal{A}' \, \overline{\mathcal{H}}^{-1} \, \mathcal{C} \, \overline{\mathcal{H}}^{-1} \, \mathcal{B} = \frac{1}{(NT)^2} \sum_{i,t} (\widetilde{\mathbb{P}} A)_{it} C_{it} (\widetilde{\mathbb{P}} B)_{it}.$$

Proof. Let $\tilde{\alpha}_i^* + \tilde{\gamma}_t^* = (\mathbb{P}\tilde{A})_{it} = (\widetilde{\mathbb{P}} A)_{it}$, with $\tilde{A}$ as defined in equation (S.2). The first order condition of the minimization problem in the definition of $(\mathbb{P}\tilde{A})_{it}$ can be written as $\frac{1}{\sqrt{NT}} \overline{\mathcal{H}}^* \binom{\tilde{\alpha}^*}{\tilde{\gamma}^*} = \mathcal{A}$. One solution to this equation is $\binom{\tilde{\alpha}^*}{\tilde{\gamma}^*} = \sqrt{NT}\, \overline{\mathcal{H}}^{-1} \mathcal{A}$ (this is the solution that imposes the normalization $\sum_i \tilde{\alpha}_i^* = \sum_t \tilde{\gamma}^*$, but this is of no importance in the following). Thus,

$$\sqrt{NT} \, \mathcal{A}' \, \overline{\mathcal{H}}^{-1} \, \mathcal{B} = \begin{pmatrix} \tilde{\alpha}^* \\ \tilde{\gamma}^* \end{pmatrix}' \mathcal{B} = \frac{1}{NT} \left[\sum_{i,t} \tilde{\alpha}_i^* B_{it} + \sum_{i,t} \tilde{\gamma}_t^* B_{it} \right] = \frac{1}{NT} \sum_{i,t} (\widetilde{\mathbb{P}} A)_{it} B_{it}.$$

This gives the first equality of Statement (i). The second equality of Statement (i) follows by symmetry. Statement (ii) is a special case of of Statement (iii) with $\mathcal{C} = \frac{1}{\sqrt{NT}} \overline{\mathcal{H}}^*$, so we only need to prove Statement (iii).

Let $\alpha_i^* + \gamma_t^* = (\mathbb{P}\tilde{B})_{it} = (\widetilde{\mathbb{P}} B)_{it}$, where $\tilde{B}_{it} = \frac{B_{it}}{\mathbb{E}_\phi(-\partial_{\pi^2} \ell_{it})}$. By an argument analogous to the one given above, we can choose $\binom{\alpha^*}{\gamma^*} = \sqrt{NT}\, \overline{\mathcal{H}}^{-1} \mathcal{B}$ as one solution to the minimization problem. Then,

$$\begin{aligned} NT \, \mathcal{A}' \, \overline{\mathcal{H}}^{-1} \, \mathcal{C} \, \overline{\mathcal{H}}^{-1} \, \mathcal{B} &= \frac{1}{NT} \sum_{i,t} \left[\tilde{\alpha}_i^* C_{it} \alpha_i^* + \tilde{\alpha}_i^* C_{it} \gamma_t^* + \tilde{\gamma}_t^* C_{it} \alpha_i^* + \tilde{\gamma}_t^* C_{it} \gamma_t^* \right] \\ &= \frac{1}{NT} \sum_{i,t} (\widetilde{\mathbb{P}} A)_{it} C_{it} (\widetilde{\mathbb{P}} B)_{it}. \end{aligned}$$

■

References

Aghion, P., Bloom, N., Blundell, R., Griffith, R., and Howitt, P. (2005). Competition and innovation: an inverted-U relationship. *The Quarterly Journal of Economics*, 120(2):701.

Cox, D. D. and Kim, T. Y. (1995). Moment bounds for mixing random variables useful in nonparametric function estimation. *Stochastic processes and their applications*, 56(1):151–158.

Fan, J. and Yao, Q. (2003). Nonlinear time series: nonparametric and parametric methods.

Hahn, J. and Kuersteiner, G. (2011). Bias reduction for dynamic nonlinear panel models with fixed effects. *Econometric Theory*, 1(1):1–40.

Higham, N. J. (1992). Estimating the matrix p-norm. *Numerische Mathematik*, 62(1):539–555.

Horn, R. A. and Johnson, C. R. (1985). *Matrix analysis*. Cambridge university press.

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

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

Table S1: Poisson model for patents

Dependent variable: citation-weighted patents	(1)	(2)	(3)	(4)	(5)	(6)
Static model						
Competition	165.12	152.81	387.46	389.99	401.88	401.51
	(54.77)	(55.74)	(67.74)			
	-20.00	*-6.43*	*-5.98*	*-5.49*	*-6.25*	*-4.74*
	(7.74)	*(8.61)*	*(19.68)*			
Competition squared	-88.55	-80.99	-204.55	-205.84	-212.15	-214.03
	(29.08)	(29.61)	(36.17)			
Dynamic model						
Lag-patents	1.05	1.07	0.46	0.48	0.50	0.70
	(0.02)	(0.03)	(0.05)			
	0.86	*0.87*	*0.36*	*0.38*	*0.39*	*0.56*
	(0.02)	*(0.03)*	*(0.07)*			
Competition	62.95	95.70	199.68	184.70	184.64	255.44
	(62.68)	(65.08)	(76.66)			
	-12.78	*-9.03*	*-1.68*	*-0.15*	*-0.43*	*-18.45*
	(7.54)	*(8.18)*	*(15.53)*			
Competition squared	-34.15	-51.09	-105.24	-97.23	-97.22	-136.97
	(33.21)	(34.48)	(40.87)			
Year effects		Yes	Yes	Yes	Yes	Yes
Industry effects			Yes	Yes	Yes	Yes
Bias correction				A	A	J
(number of lags)				1	2	

Notes: Data set obtained from ABBGH. Competition is measured by (1-Lerner index) in the industry-year. All columns are estimated using an unbalanced panel of seventeen industries over the period 1973 to 1994. First year available used as initial condition in dynamic model. The estimates of the coefficients for the static model in columns (2) and (3) replicate the results in ABBGH. A is the bias corrected estimator that uses an analytical correction with a number lags to estimate the spectral expectations specified at the bottom cell. J is the jackknife bias corrected estimator that uses split panel jackknife in both the individual and time dimensions. Standard errors in parentheses and average partial effects in italics.

Table S.2: Homogeneity test for the jackknife

	Cross section	Time series
Static Model	10.49	13.37
	(0.01)	(0.00)
Dynamic Model	1.87	12.41
	(0.60)	(0.01)

Notes: Wald test for equality of common parameters across sub panels.

P-values in parentheses

Table S3: Finite-sample properties in static Poisson model

	Coefficient of Z_{it}					Coefficient of Z_{it}^2					APE of Z_{it}				
	Bias	Std. Dev.	RMSE	SE/SD	p; .95	Bias	Std. Dev.	RMSE	SE/SD	p; .95	Bias	Std. Dev.	RMSE	SE/SD	p; .95
						N = 17, T = 22, unbalanced									
MLE	-59	14	60	1.04	0.01	-58	14	60	1.03	0.01	222	113	248	1.15	0.60
MLE-TE	-62	14	64	1.01	0.01	-62	14	64	1.01	0.01	-9	139	139	1.04	0.94
MLE-FETE	-2	17	17	1.02	0.96	-2	17	17	1.02	0.96	-15	226	226	1.49	1.00
Analytical (L=1)	-1	17	17	1.02	0.96	-1	17	17	1.02	0.96	-9	225	225	1.50	1.00
Analytical (L=2)	-1	17	17	1.02	0.96	-1	17	17	1.02	0.96	-6	225	225	1.50	1.00
Jackknife	-3	25	25	0.69	0.83	-3	25	25	0.70	0.83	-15	333	333	1.01	0.95
						N = 34, T = 22, unbalanced									
MLE	-58	10	59	1.03	0.00	-57	10	58	1.03	0.00	226	81	240	0.98	0.20
MLE-TE	-61	10	62	1.00	0.00	-61	10	62	1.00	0.00	-3	97	97	0.95	0.94
MLE-FETE	0	12	12	0.99	0.96	0	13	13	0.99	0.96	-6	158	158	1.12	0.98
Analytical (L=1)	0	12	12	0.99	0.96	0	13	13	0.99	0.96	0	159	158	1.11	0.98
Analytical (L=2)	1	13	13	0.99	0.96	1	13	13	0.99	0.96	3	159	159	1.11	0.98
Jackknife	-1	14	14	0.90	0.93	-1	14	14	0.90	0.93	-15	208	208	0.85	0.90
						N = 51, T = 22, unbalanced									
MLE	-58	8	58	1.00	0.00	-57	8	57	1.00	0.00	228	66	238	0.96	0.06
MLE-TE	-61	8	61	1.00	0.00	-61	8	61	1.00	0.00	-1	77	77	0.95	0.94
MLE-FETE	0	10	10	0.97	0.94	0	11	11	0.97	0.94	-4	128	128	1.04	0.96
Analytical (L=1)	0	10	10	0.97	0.94	0	11	11	0.97	0.94	2	129	128	1.04	0.96
Analytical (L=2)	1	10	11	0.96	0.94	1	11	11	0.96	0.94	5	129	129	1.04	0.96
Jackknife	0	11	11	0.90	0.93	0	11	11	0.90	0.94	-12	169	170	0.79	0.88

Notes: All the entries are in percentage of the true parameter value. 500 repetitions. The data generating process is: $Y_{it} \sim \text{Poisson}(\exp\{\beta_1 X_{it} + \beta_2 X_{it}^2 + \alpha_i + \gamma_t\})$ with all the variables and coefficients calibrated to the dataset of ABBGH. Average effect is $E[(\beta_1 + 2\beta_2 X_{it})\exp(\beta_1 X_{it} + \beta_2 X_{it}^2 + \alpha_i + \gamma_t)]$. MLE is the Poisson maximum likelihood estimator without individual and time fixed effects; MLE-TE is the Poisson maximum likelihood estimator with time fixed effects; MLE-FETE is the Poisson maximum likelihood estimator with individual and time fixed effects;Analytical (L = l) is the bias corrected estimator that uses an analytical correction with l lags to estimate the spectral expectations; and Jackknife is the bias corrected estimator that uses split panel jackknife in both the individual and time dimension.

Table S4: Finite-sample properties in dynamic Poisson model: lagged dependent variable

	Coefficient of $Y_{i,t-1}$					APE of $Y_{i,t-1}$				
	Bias	Std. Dev.	RMSE	SE/SD	p; .95	Bias	Std. Dev.	RMSE	SE/SD	p; .95
					N = 17,T = 21, unbalanced					
MLE	135	3	135	1.82	0.00	158	2	158	3.75	0.00
MLE-TE	142	3	142	1.95	0.00	163	3	163	4.17	0.00
MLE-FETE	-17	15	23	0.96	0.78	-17	15	22	1.38	0.89
Analytical (L=1)	-7	15	17	0.98	0.91	-8	14	16	1.41	0.97
Analytical (L=2)	-5	15	16	0.96	0.92	-5	15	16	1.38	0.98
Jackknife	4	20	21	0.73	0.85	4	20	20	1.03	0.95
					N = 34, T = 21, unbalanced					
MLE	135	2	135	1.76	0.00	158	2	158	2.82	0.00
MLE-TE	141	2	141	1.77	0.00	162	2	162	2.69	0.00
MLE-FETE	-16	11	19	0.93	0.65	-16	10	19	1.05	0.71
Analytical (L=1)	-7	11	13	0.95	0.89	-7	10	12	1.08	0.92
Analytical (L=2)	-4	11	12	0.93	0.91	-4	10	11	1.05	0.94
Jackknife	3	13	14	0.77	0.85	3	13	13	0.86	0.89
					N = 51, T = 21, unbalanced					
MLE	135	2	135	1.81	0.00	158	1	158	2.58	0.00
MLE-TE	141	2	141	1.79	0.00	162	2	162	2.41	0.00
MLE-FETE	-15	8	17	0.97	0.55	-15	8	17	1.03	0.55
Analytical (L=1)	-6	8	10	0.99	0.90	-6	8	10	1.05	0.91
Analytical (L=2)	-3	8	9	0.97	0.93	-4	8	9	1.03	0.93
Jackknife	3	11	11	0.77	0.87	3	10	11	0.80	0.88

Notes: All the entries are in percentage of the true parameter value. 500 repetitions. The data generating process is: $Y_{it} \sim \text{Poisson}(\exp\{\beta_Y \log(1 + Y_{i,t-1}) + \beta_1 Z_{it} + \beta_2 Z_{it}^2 + \alpha_i + \gamma_t\})$, where all the exogenous variables, initial condition and coefficients are calibrated to the application of ABBGH. Average effect is $\beta_Y \, E[\exp\{((\beta_Y - 1)\log(1 + Y_{i,t-1}) + \beta_1 Z_{it} + \beta_2 Z_{it}^2 + \alpha_i + \gamma_t\}]$. MLE is the Poisson maximum likelihood estimator without individual and time fixed effects; MLE-TE is the Poisson maximum likelihood estimator with time fixed effects; MLE-FETE is the Poisson maximum likelihood estimator with individual and time fixed effects; Analytical (L = l) is the bias corrected estimator that uses an analytical correction with l lags to estimate the spectral expectations; and Jackknife is the bias corrected estimator that uses split panel jackknife in both the individual and time dimension.

Table S5: Finite-sample properties in dynamic Poisson model: exogenous regressor

	Coefficient of Z_{it}					Coefficient of Z_{it}^2					APE of Z_{it}				
	Bias	Std. Dev.	RMSE	SE/SD	p; .95	Bias	Std. Dev.	RMSE	SE/SD	p; .95	Bias	Std. Dev.	RMSE	SE/SD	p; .95
						N = 17, T = 21, unbalanced									
MLE	-76	27	81	1.13	0.29	-76	27	80	1.13	0.30	760	351	837	1.65	0.89
MLE-TE	-65	28	71	1.12	0.44	-65	29	71	1.12	0.45	541	356	647	1.75	0.99
MLE-FETE	9	40	41	0.95	0.92	9	41	42	0.95	0.92	-3	1151	1150	1.08	0.99
Analytical (L=1)	4	40	40	0.97	0.94	4	40	40	0.97	0.94	11	1117	1116	1.11	0.99
Analytical (L=2)	3	39	39	0.97	0.94	3	40	40	0.97	0.94	15	1110	1109	1.12	0.99
Jackknife	3	57	57	0.68	0.82	3	57	57	0.68	0.81	24	1653	1651	0.75	0.86
						N = 34, T = 21, unbalanced									
MLE	-75	19	77	1.18	0.04	-74	19	77	1.18	0.05	777	252	817	1.47	0.42
MLE-TE	-65	19	67	1.18	0.15	-64	19	67	1.18	0.15	534	248	589	1.65	0.88
MLE-FETE	6	28	28	0.97	0.94	6	28	29	0.97	0.94	-68	734	736	1.03	0.94
Analytical (L=1)	2	27	27	0.99	0.95	2	28	28	0.99	0.95	-51	713	714	1.06	0.95
Analytical (L=2)	0	27	27	0.99	0.95	0	27	27	1.00	0.95	-47	706	707	1.07	0.95
Jackknife	2	31	31	0.87	0.92	2	31	31	0.87	0.92	-38	1012	1012	0.74	0.85
						N = 51, T = 21, unbalanced									
MLE	-74	15	76	1.17	0.00	-73	15	75	1.17	0.00	768	201	794	1.48	0.18
MLE-TE	-63	16	65	1.15	0.05	-63	16	65	1.15	0.05	535	197	570	1.68	0.74
MLE-FETE	8	22	23	1.01	0.93	8	22	24	1.01	0.93	-27	606	606	0.99	0.95
Analytical (L=1)	4	21	22	1.02	0.95	4	22	22	1.02	0.95	-11	588	587	1.02	0.96
Analytical (L=2)	2	21	21	1.03	0.95	2	22	22	1.03	0.95	-5	581	580	1.03	0.96
Jackknife	3	25	25	0.89	0.91	4	25	25	0.89	0.91	8	838	837	0.71	0.83

Notes: All the entries are in percentage of the true parameter value. 500 repetitions. The data generating process is: $Y_{it} \sim \text{Poisson}(\exp\{\beta_Y \log(1 + Y_{i,t-1}) + \beta_1 Z_{it} + \beta_2 Z_{it}^2 + \alpha_i + \gamma_t\})$, where all the exogenous variables, initial condition and coefficients are calibrated to the application of ABBGH. Average effect is $E[(\beta_1 + 2\beta_2 Z_{it}) \exp\{\beta_Y \log(1 + Y_{i,t-1}) + \beta_1 Z_{it} + \beta_2 Z_{it}^2 + \alpha_i + \gamma_t\}]$. MLE is the Poisson maximum likelihood estimator without individual and time fixed effects; MLE-TE is the Poisson maximum likelihood estimator with time fixed effects; MLE-FETE is the Poisson maximum likelihood estimator with individual and time fixed effects; Analytical (L = l) is the bias corrected estimator that uses an analytical correction with l lags to estimate the spectral expectations; and Jackknife is the bias corrected estimator that uses split panel jackknife in both the individual and time dimension.