Econometrica Supplementary Material

SUPPLEMENT TO "DISCRETIZING UNOBSERVED HETEROGENEITY"

(*Econometrica*, Vol. 90, No. 2, March 2022, 625–643)

STÉPHANE BONHOMME
Kenneth C. Griffin Department of Economics, University of Chicago

THIBAUT LAMADON
Kenneth C. Griffin Department of Economics, University of Chicago

ELENA MANRESA
Department of Economics, New York University

IN THIS APPENDIX we provide the proofs of various technical lemmas, present several extensions, and give details about the Monte Carlo simulations.

S1. PROOFS OF TECHNICAL LEMMAS

PROOF OF LEMMA A1: From Assumption 3(iii)–(iv) or 4(iii)–(iv), both $\frac{\partial\overline{\alpha}^j(\theta,\xi)}{\partial\theta'}$ and $\frac{\partial\overline{\alpha}^j(\theta,\xi)}{\partial\xi'}$ are uniformly bounded (in probability in the time-varying case). Let $a^j(k,\theta)=\overline{\alpha}^j(\theta,\psi(\widehat{h}(k)))$. We thus have, using Lemmas 1 and 2,

$$\sup_{\theta\in\Theta}\frac{1}{Np}\sum_{i,j}\left\|a^j(\widehat{k}_i,\theta)-\overline{\alpha}^j(\theta,\xi_{i0})\right\|^2=\sup_{\theta\in\Theta}\frac{1}{Np}\sum_{i,j}\left\|\overline{\alpha}^j\big(\theta,\psi\big(\widehat{h}(\widehat{k}_i)\big)\big)-\overline{\alpha}^j\big(\theta,\psi\big(\varphi(\xi_{i0})\big)\big)\right\|^2$$

$$=O_p\left(\frac{1}{N}\sum_i\left\|\widehat{h}(\widehat{k}_i)-\varphi(\xi_{i0})\right\|^2\right)=O_p(\delta). \tag{S1}$$

Let $\theta\in\Theta$. Expanding: $\sum_{i,j}\ell_{ij}(a^j(\widehat{k}_i,\theta),\theta)\leq\sum_{i,j}\ell_{ij}(\widehat{\alpha}^j(\widehat{k}_i,\theta),\theta)$ to second order around $\overline{\alpha}^j(\theta,\xi_{i0})$, and using

$$\max_{i,j}\sup_{(\alpha,\theta)}\left\|v_{ij}^{\alpha}(\alpha,\theta)\right\|=O_p(1), \tag{S2}$$

we have, for some $a_{ij}(\theta)$ between $\widehat{\alpha}^j(\widehat{k}_i,\theta)$ and $\overline{\alpha}^j(\theta,\xi_{i0})$,

$$\begin{aligned}&\frac{1}{2Np}\sum_{i,j}\big(\widehat{\alpha}^j(\widehat{k}_i,\theta)-\overline{\alpha}^j(\theta,\xi_{i0})\big)'\big[-v_{ij}^{\alpha}\big(a_{ij}(\theta),\theta\big)\big]\big(\widehat{\alpha}^j(\widehat{k}_i,\theta)-\overline{\alpha}^j(\theta,\xi_{i0})\big)\\&\leq\frac{1}{Np}\sum_{i,j}v_{ij}\big(\overline{\alpha}^j(\theta,\xi_{i0}),\theta\big)'\big(\widehat{\alpha}^j(\widehat{k}_i,\theta)-a^j(\widehat{k}_i,\theta)\big)+O_p(\delta)\\&=\frac{1}{Np}\sum_{i,j}\overline{v}_j(\widehat{k}_i,\theta)'\big(\widehat{\alpha}^j(\widehat{k}_i,\theta)-a^j(\widehat{k}_i,\theta)\big)+O_p(\delta),\end{aligned} \tag{S3}$$

Stéphane Bonhomme: sbonhomme@uchicago.edu
Thibaut Lamadon: lamadon@uchicago.edu
Elena Manresa: elena.manresa@nyu.edu

 https://doi.org/10.3982/ECTA15238

where $\overline{v}_j(k,\theta)$ denotes the mean over i of $v_{ij}(\overline{\alpha}^j(\theta,\xi_{i0}),\theta)$ in group $\widehat{k}_i=k$, and the $O_p(\delta)$ terms are uniform in θ by (S1).

Now, by Assumption 3(ii) or 4(ii) there exists a constant $\underline{c}>0$ such that

$$\min_{i,j}\inf_{(\alpha,\theta)}\text{mineig}\big[-v_{ij}^{\alpha}(\alpha,\theta)\big]\geq\underline{c}+o_p(1), \tag{S4}$$

where mineig(M) is the minimum eigenvalue of M. Let $A=\frac{1}{Np}\sum_{i,j}\|\widehat{\alpha}^j(\widehat{k}_i,\theta)-\overline{\alpha}^j(\theta,\xi_{i0})\|^2$. By (S3) and the Cauchy–Schwarz inequality, we have

$$A\leq O_p\left[\left(\frac{1}{Np}\sum_{i,j}\big\|\overline{v}_j(\widehat{k}_i,\theta)\big\|^2\right)^{\frac{1}{2}}\left(\frac{1}{Np}\sum_{i,j}\big\|\widehat{\alpha}^j(\widehat{k}_i,\theta)-a^j(\widehat{k}_i,\theta)\big\|^2\right)^{\frac{1}{2}}\right]+O_p(\delta).$$

By (S1) and the triangle inequality: $(\frac{1}{Np}\sum_{i,j}\|\widehat{\alpha}^j(\widehat{k}_i,\theta)-a^j(\widehat{k}_i,\theta)\|^2)^{\frac{1}{2}}\leq A^{\frac{1}{2}}+O_p(\delta^{\frac{1}{2}})$. Hence, $A=O_p[(\frac{1}{Np}\sum_{i,j}\|\overline{v}_j(\widehat{k}_i,\theta)\|^2)^{\frac{1}{2}}(A^{\frac{1}{2}}+O_p(\delta^{\frac{1}{2}}))]+O_p(\delta)$, which implies

$$A=O_p\left(\frac{1}{Np}\sum_{i,j}\big\|\overline{v}_j(\widehat{k}_i,\theta)\big\|^2\right)+O_p(\delta). \tag{S5}$$

We are now going to show that, for all $\theta\in\Theta$,

$$\frac{1}{Np}\sum_{i,j}\big\|\overline{v}_j(\widehat{k}_i,\theta)\big\|^2=O_p(\delta). \tag{S6}$$

Using (S5) and (S6) will then imply (A1). Under the conditions of Theorem 1, it is easy to see that (S6) holds. We are now going to show (S6) under the conditions of Theorem 2. Let, for all j, θ, h, ξ, λ: $\rho_j(h,\xi,\lambda,\theta)=\mathbb{E}_{h_i=h,\xi_{i0}=\xi,\lambda_0=\lambda}(v_{ij}(\overline{\alpha}^j(\theta,\xi),\theta))$, and, for all i, j, θ: $\zeta_{ij}(\theta)=v_{ij}(\overline{\alpha}^j(\theta,\xi_{i0}),\theta)-\rho_j(h_i,\xi_{i0},\lambda_0,\theta)$. By Assumption 4(v), and letting $h_i=\varphi(\xi_{i0})+\varepsilon_i$, we can expand $\rho_j(h_i,\xi_{i0},\lambda_0,\theta)$ twice around $\varphi(\xi_{i0})$ as $\rho_j(\varphi(\xi_{i0}),\xi_{i0},\lambda_0,\theta)+\frac{\partial\rho_j(\varphi(\xi_{i0}),\xi_{i0},\lambda_0,\theta)}{\partial h'}\varepsilon_i+\frac{1}{2}\varepsilon_i'\frac{\partial^2\rho_j(a_{i\theta}^j,\xi_{i0},\lambda_0,\theta)}{\partial h\partial h'}\varepsilon_i$, where $a_{i\theta}^j$ lies between h_i and $\varphi(\xi_{i0})$. Hence, taking expectations, using that $\mathbb{E}_{\xi_{i0},\lambda_0}[\rho_j(h_i,\xi_{i0},\lambda_0,\theta)]=0$, and using Assumptions 2 and 4(v), we have

$$\frac{1}{Np}\sum_{i,j}\big\|\rho_j\big(\varphi(\xi_{i0}),\xi_{i0},\lambda_0,\theta\big)\big\|^2=\frac{1}{Np}\sum_{i,j}\left\|\frac{\partial\rho_j\big(\varphi(\xi_{i0}),\xi_{i0},\lambda_0,\theta\big)}{\partial h'}\mathbb{E}_{\xi_{i0},\lambda_0}[\varepsilon_i]\right\|^2+o_p\left(\frac{1}{T}\right),$$

which is $O_p(\frac{1}{T})$. Hence, $\frac{1}{Np}\sum_{i,j}\|\rho_j(h_i,\xi_{i0},\lambda_0,\theta)\|^2=O_p(\frac{1}{T})$. It thus follows from the triangle inequality that

$$\frac{1}{Np}\sum_{i,j}\big\|\overline{v}_j(\widehat{k}_i,\theta)\big\|^2\leq O_p\left(\frac{1}{T}\right)+\frac{2}{Np}\sum_{i,j}\big\|\overline{\zeta}_j(\widehat{k}_i,\theta)\big\|^2, \tag{S7}$$

where $\overline{\zeta}_j(k,\theta)$ denotes the mean of $\zeta_{ij}(\theta)$ in group $\widehat{k}_i = k$. Now, using that $\widehat{k}_1,\dots,\widehat{k}_N$ are functions of $h_1,\dots,h_N$, we have

$$\mathbb{E}\left[\frac{1}{Np}\sum_{i,j}\left\|\overline{\zeta}_j(\widehat{k}_i,\theta)\right\|^2\right] = \frac{1}{Np}\sum_{k,j}\mathbb{E}\left[\frac{\sum_{i,i'}\mathbf{1}\{\widehat{k}_i=k\}\mathbf{1}\{\widehat{k}_{i'}=k\}\mathbb{E}_{h_1,\dots,h_N,\xi_{10},\dots,\xi_{N0},\lambda_0}\big(\zeta_{ij}(\theta)'\zeta_{i'j}(\theta)\big)}{\sum_i \mathbf{1}\{\widehat{k}_i=k\}}\right].$$

Furthermore, since observations are independent across i given λ_0,

$$\mathbb{E}_{h_1,\dots,h_N,\xi_{10},\dots,\xi_{N0},\lambda_0}\big(\zeta_{i_1,j}(\theta)'\zeta_{i_2,j}(\theta)\big) = \mathbb{E}_{h_{i_1},\xi_{i_10},\lambda_0}\big(\zeta_{i_1,j}(\theta)\big)'\mathbb{E}_{h_{i_2},\xi_{i_20},\lambda_0}\big(\zeta_{i_2,j}(\theta)\big) = 0 \quad \text{for all } i_1 \neq i_2 \text{ and } j.$$

Hence,

$$\mathbb{E}\left[\frac{1}{Np}\sum_{i,j}\left\|\overline{\zeta}_j(\widehat{k}_i,\theta)\right\|^2\right] = \frac{1}{Np}\sum_{k,j}\mathbb{E}\left[\frac{\sum_{i}\mathbf{1}\{\widehat{k}_i=k\}\mathbb{E}_{h_i,\xi_{i0},\lambda_0}\big(\zeta_{ij}(\theta)'\zeta_{ij}(\theta)\big)}{\sum_i \mathbf{1}\{\widehat{k}_i=k\}}\right].$$

Finally, using that $\mathbb{E}_{h_i,\xi_{i0},\lambda_0}(\zeta_{ij}(\theta)) = 0$, and using part (v) in Assumption 4,

$$\mathbb{E}_{h_i=h,\xi_{i0}=\xi,\lambda_0=\lambda}\big(\zeta_{ij}(\theta)'\zeta_{ij}(\theta)\big) = \mathrm{Tr}\big[\mathrm{Var}_{h_i=h,\xi_{i0}=\xi,\lambda_0=\lambda}\big(v_{ij}\big(\overline{\alpha}^j(\theta,\xi_{i0}),\theta\big)\big)\big] = O(1),$$

uniformly in h, ξ, λ.[1] This implies that $\mathbb{E}[\frac{1}{Np}\sum_{i,j}\|\overline{\zeta}_j(\widehat{k}_i,\theta)\|^2] = O(\frac{K}{N})$, and shows (S6) and (A1).

We are now going to show

$$\sup_{\theta\in\Theta}\frac{1}{Np}\sum_{i,j}\left\|\overline{v}_j(\widehat{k}_i,\theta)\right\|^2 = o_p(1). \tag{S8}$$

Using a bounding argument similar to the one we used to show (A1), (A2) will then follow. To see that (S8) holds, let $Z(\theta) = \frac{1}{Np}\sum_{i,j}\|\overline{v}_j(\widehat{k}_i,\theta)\|^2$. By (S6), $Z(\theta) = O_p(\delta)$ for all $\theta\in\Theta$. Moreover, $\frac{\partial Z(\theta)}{\partial\theta} = \frac{2}{Np}\sum_{i,j}\overline{v}_j^{\theta}(\widehat{k}_i,\theta)\overline{v}_j(\widehat{k}_i,\theta) = O_p(\sqrt{\sup_{\theta\in\Theta}Z(\theta)})$ uniformly in θ, using the Cauchy–Schwarz inequality with either Assumption 3(ii) or 4(ii), where $\overline{v}_j^{\theta}(k,\widetilde{\theta})$ is the mean of $\frac{\partial}{\partial\theta}|_{\theta=\widetilde{\theta}}v_{ij}(\overline{\alpha}^j(\theta,\xi_{i0}),\theta)'$ in group $\widehat{k}_i = k$. Since Θ is compact, it follows that $\sup_{\theta\in\Theta}Z(\theta) = o_p(1)$.[2] *Q.E.D.*

[1]Note that the dimension of v_{ij} is fixed throughout, independent of the sample size.

[2]Let $v>0$, $\epsilon>0$. There is $M>0$ such that $\Pr(\sup_{\theta\in\Theta}\|\frac{\partial Z(\theta)}{\partial\theta}\| > M\sqrt{\sup_{\theta\in\Theta}Z(\theta)}) < \frac{\epsilon}{2}$. Take a finite cover of $\Theta = B_1\cup\dots\cup B_R$, where B_r are balls with centers θ_r and diameters $\operatorname{diam} B_r \leq \frac{1}{2M}\sqrt{v}$. Since $\sup_{\theta\in\Theta}Z(\theta) \leq \max_r Z(\theta_r) + \sup_\theta\|\frac{\partial Z(\theta)}{\partial\theta}\|\frac{1}{2M}\sqrt{v}$, and since $a>v \Rightarrow a-\sqrt{a}\frac{1}{2}\sqrt{v} > \frac{v}{2}$, we have $\Pr(\sup_{\theta\in\Theta}Z(\theta)>v) \leq \frac{\epsilon}{2} + \Pr(\max_r Z(\theta_r) > \frac{v}{2})$, which by (S6), is smaller than ϵ for N, T large enough.

PROOF OF LEMMA A2: Let us omit references to θ_0 and α_{i0}^j throughout, and let

$$A = \frac{1}{Np}\sum_{i=1}^N\sum_{j=1}^p \mathbb{E}_{\xi_{i0},\lambda_0}\big(v_{ij}^\theta\big)\big[\mathbb{E}_{\xi_{i0},\lambda_0}\big(v_{ij}^\alpha\big)\big]^{-1} v_{ij}^\alpha\big(\widehat{\alpha}^j(\widehat{k}_i,\theta_0) - \alpha_{i0}^j + \big(v_{ij}^\alpha\big)^{-1}v_{ij}\big),$$

$$B = \frac{1}{Np}\sum_{i=1}^N\sum_{j=1}^p \big(v_{ij}^\theta\big(v_{ij}^\alpha\big)^{-1} - \mathbb{E}_{\xi_{i0},\lambda_0}\big(v_{ij}^\theta\big)\big[\mathbb{E}_{\xi_{i0},\lambda_0}\big(v_{ij}^\alpha\big)\big]^{-1}\big) v_{ij}^\alpha\big(\widehat{\alpha}^j(\widehat{k}_i,\theta_0) - \alpha_{i0}^j\big).$$

We first bound A. Expanding $\sum_i \mathbf{1}\{\widehat{k}_i = k\}v_{ij}(\widehat{\alpha}^j(k)) = 0$ for all k, j, we have, for a_{ij} between α_{i0}^j and $\widehat{\alpha}^j(\widehat{k}_i)$,

$$\sum_i \mathbf{1}\{\widehat{k}_i = k\}v_{ij}\big(\alpha_{i0}^j\big) + \sum_i \mathbf{1}\{\widehat{k}_i = k\}v_{ij}^\alpha\big(\alpha_{i0}^j\big)\big(\widehat{\alpha}^j(\widehat{k}_i) - \alpha_{i0}^j\big)$$
$$+ \frac{1}{2}\sum_i \mathbf{1}\{\widehat{k}_i = k\}v_{ij}^{\alpha\alpha}(a_{ij})\big(\widehat{\alpha}^j(\widehat{k}_i) - \alpha_{i0}^j\big)\otimes\big(\widehat{\alpha}^j(\widehat{k}_i) - \alpha_{i0}^j\big) = 0.$$

It follows that $\widehat{\alpha}^j(\widehat{k}_i) = \widetilde{\alpha}_j(\widehat{k}_i) + \widetilde{v}_j(\widehat{k}_i) + \widetilde{w}_j(\widehat{k}_i)$, where

$$\widetilde{\alpha}_j(k) = \Big(\sum_i \mathbf{1}\{\widehat{k}_i = k\}\big(-v_{ij}^\alpha\big)\Big)^{-1}\Big(\sum_i \mathbf{1}\{\widehat{k}_i = k\}\big(-v_{ij}^\alpha\big)\alpha_{i0}^j\Big),$$

$$\widetilde{v}_j(k) = \Big(\sum_i \mathbf{1}\{\widehat{k}_i = k\}\big(-v_{ij}^\alpha\big)\Big)^{-1}\Big(\sum_i \mathbf{1}\{\widehat{k}_i = k\}v_{ij}\Big),$$

$$\widetilde{w}_j(k) = \frac{1}{2}\Big(\sum_i \mathbf{1}\{\widehat{k}_i = k\}\big(-v_{ij}^\alpha\big)\Big)^{-1}\Big(\sum_i \mathbf{1}\{\widehat{k}_i = k\}v_{ij}^{\alpha\alpha}(a_{ij})\big(\widehat{\alpha}^j(\widehat{k}_i) - \alpha_{i0}^j\big)^{\otimes 2}\Big),$$

where $a^{\otimes 2} = a\otimes a$. Hence, we have

$$A = \frac{1}{Np}\sum_{i,j}\mathbb{E}_{\xi_{i0},\lambda_0}\big(v_{ij}^\theta\big)\big[\mathbb{E}_{\xi_{i0},\lambda_0}\big(v_{ij}^\alpha\big)\big]^{-1}v_{ij}^\alpha\big(\widetilde{w}_j(\widehat{k}_i) + \widetilde{\alpha}_j(\widehat{k}_i) - \alpha_{i0}^j + \widetilde{v}_j(\widehat{k}_i) + \big(v_{ij}^\alpha\big)^{-1}v_{ij}\big).$$

Note first that

$$\frac{1}{Np}\sum_{i,j}\mathbb{E}_{\xi_{i0},\lambda_0}\big(v_{ij}^\theta\big)\big[\mathbb{E}_{\xi_{i0},\lambda_0}\big(v_{ij}^\alpha\big)\big]^{-1}v_{ij}^\alpha\widetilde{w}_j(\widehat{k}_i) = O_p\bigg(\frac{1}{Np}\sum_{i,j}\big\|\widehat{\alpha}^j(\widehat{k}_i) - \alpha_{i0}^j\big\|^2\bigg) = O_p(\delta),$$

where we have used (S2), (A1), and either Assumption 3(ii) or Assumption 4(ii).

Next, let $z_j(\xi_{i0})' = \mathbb{E}_{\xi_{i0},\lambda_0}(v_{ij}^\theta)[\mathbb{E}_{\xi_{i0},\lambda_0}(v_{ij}^\alpha)]^{-1}$. We have

$$\frac{1}{Np}\sum_{i,j}\mathbb{E}_{\xi_{i0},\lambda_0}\big(v_{ij}^\theta\big)\big[\mathbb{E}_{\xi_{i0},\lambda_0}\big(v_{ij}^\alpha\big)\big]^{-1}v_{ij}^\alpha\big(\widetilde{\alpha}_j(\widehat{k}_i) - \alpha_{i0}^j\big)$$
$$= \frac{1}{Np}\sum_{i,j}\big(z_j(\xi_{i0})' - \widetilde{z}_j(\widehat{k}_i)'\big)v_{ij}^\alpha\big(\widetilde{\alpha}_j(\widehat{k}_i) - \alpha_{i0}^j\big), \tag{S9}$$

where, for all k, j,

$$\widetilde{z}_j(k) = \left(\sum_i \mathbf{1}\{\widehat{k}_i = k\}\left(-v_{ij}^{\alpha}\right)\right)^{-1}\left(\sum_i \mathbf{1}\{\widehat{k}_i = k\}\left(-v_{ij}^{\alpha}\right) z_j(\xi_{i0})\right). \tag{S10}$$

Now we have, using that $\alpha^j \mapsto \sum_i (\alpha^j(\widehat{k}_i) - \alpha_{i0}^j)'(-v_{ij}^{\alpha})(\alpha^j(\widehat{k}_i) - \alpha_{i0}^j)$ is minimized at $\alpha^j = \widetilde{\alpha}_j$, and using (S2) and (S4):

$$\begin{aligned}\frac{1}{Np}\sum_{i,j}\left\|\widetilde{\alpha}_j(\widehat{k}_i) - \alpha_{i0}^j\right\|^2 &= O_p\left(\frac{1}{Np}\sum_{i,j}\left(\widetilde{\alpha}_j(\widehat{k}_i) - \alpha_{i0}^j\right)'\left(-v_{ij}^{\alpha}\right)\left(\widetilde{\alpha}_j(\widehat{k}_i) - \alpha_{i0}^j\right)\right) \\ &= O_p\left(\frac{1}{Np}\sum_{i,j}\left(\widehat{\alpha}^j(\widehat{k}_i) - \alpha_{i0}^j\right)'\left(-v_{ij}^{\alpha}\right)\left(\widehat{\alpha}^j(\widehat{k}_i) - \alpha_{i0}^j\right)\right) \\ &= O_p\left(\frac{1}{Np}\sum_{i,j}\left\|\widehat{\alpha}^j(\widehat{k}_i) - \alpha_{i0}^j\right\|^2\right),\end{aligned}$$

where the last expression is $O_p(\delta)$ by (A1). Likewise, since by Assumption 3(iv) or 4(iv) $\frac{\partial \operatorname{vec} z_j(\xi)}{\partial \xi'}$ is bounded (in probability) uniformly in j and ξ, we have

$$\begin{aligned}&\frac{1}{Np}\sum_{i,j}\left\|\widetilde{z}_j(\widehat{k}_i) - z_j(\xi_{i0})\right\|^2 \\ &\quad = O_p\left(\frac{1}{Np}\sum_{i,j}\left(\widetilde{z}_j(\widehat{k}_i) - z_j(\xi_{i0})\right)'\left(-v_{ij}^{\alpha}\right)\left(\widetilde{z}_j(\widehat{k}_i) - z_j(\xi_{i0})\right)\right) \\ &\quad = O_p\left(\frac{1}{Np}\sum_{i,j}\left(z_j\left(\psi\left(\widehat{h}(\widehat{k}_i)\right)\right) - z_j(\xi_{i0})\right)'\left(-v_{ij}^{\alpha}\right)\left(z_j\left(\psi\left(\widehat{h}(\widehat{k}_i)\right)\right) - z_j(\xi_{i0})\right)\right) \\ &\quad = O_p\left(\frac{1}{Np}\sum_{i,j}\left\|\widehat{h}(\widehat{k}_i) - \varphi(\xi_{i0})\right\|^2\right) = O_p(\delta),\end{aligned} \tag{S11}$$

where we have used (S2), (S4), Lemmas 1 and 2, and that ψ is Lipschitz-continuous. Combining results, and using the Cauchy–Schwarz inequality in (S9), we obtain

$$\frac{1}{Np}\sum_{i,j}\mathbb{E}_{\xi_{i0},\lambda_0}\left(v_{ij}^{\theta}\right)\left[\mathbb{E}_{\xi_{i0},\lambda_0}\left(v_{ij}^{\alpha}\right)\right]^{-1} v_{ij}^{\alpha}\left(\widetilde{\alpha}_j(\widehat{k}_i) - \alpha_{i0}^j\right) = O_p(\delta).$$

The last term in A is

$$A_3 = \frac{1}{Np}\sum_{i,j}\mathbb{E}_{\xi_{i0},\lambda_0}\left(v_{ij}^{\theta}\right)\left[\mathbb{E}_{\xi_{i0},\lambda_0}\left(v_{ij}^{\alpha}\right)\right]^{-1}\left(-v_{ij}^{\alpha}\right)\left(\left(-v_{ij}^{\alpha}\right)^{-1} v_{ij} - \widetilde{v}_j(\widehat{k}_i)\right).$$

Since $\widetilde{v}_j(k)=(\sum_i \mathbf{1}\{\widehat{k}_i=k\}(-v^{\alpha}_{ij}))^{-1}(\sum_i \mathbf{1}\{\widehat{k}_i=k\}(-v^{\alpha}_{ij})(-v^{\alpha}_{ij})^{-1}v_{ij})$, we have

$$A_3=\frac{1}{Np}\sum_{i,j}\big(z_j(\xi_{i0})'-\widetilde{z}_j(\widehat{k}_i)'\big)\big(-v^{\alpha}_{ij}\big)\big(-v^{\alpha}_{ij}\big)^{-1}v_{ij}=\frac{1}{Np}\sum_{i,j}\big(z_j(\xi_{i0})'-\widetilde{z}_j(\widehat{k}_i)'\big)v_{ij}$$

$$=\frac{1}{Np}\sum_{i,j}\big(z_j(\xi_{i0})'-z^*_j(\widehat{k}_i)'\big)v_{ij}+\frac{1}{Np}\sum_{i,j}\big(z^*_j(\widehat{k}_i)'-\widetilde{z}_j(\widehat{k}_i)'\big)v_{ij}, \tag{S12}$$

where $\widetilde{z}_j(k)$ is given by (S10), and

$$z^*_j(k)=\left(\sum_i \mathbf{1}\{\widehat{k}_i=k\}\mathbb{E}_{\xi_{i0},\lambda_0}\big(-v^{\alpha}_{ij}\big)\right)^{-1}\left(\sum_i \mathbf{1}\{\widehat{k}_i=k\}\mathbb{E}_{\xi_{i0},\lambda_0}\big(-v^{\alpha}_{ij}\big)z_j(\xi_{i0})\right). \tag{S13}$$

Under the conditions of Theorem 1, it is easy to see that $A_3=O_p(\delta)$. We are now going to show that $A_3=O_p(\delta)$ under the conditions of Theorem 2. To see that the first term on the right-hand side of (S12) is $O_p(\delta)$, we use an argument similar to the one we used to show (S6). Let $\zeta_{ij}=v_{ij}-\mathbb{E}_{h_i,\xi_{i0},\lambda_0}(v_{ij})$. Following the same steps as the ones leading to (S7), we obtain

$$\frac{1}{Np}\sum_{i,j}\big\|\mathbb{E}_{h_i,\xi_{i0},\lambda_0}(v_{ij})\big\|^2=O_p\left(\frac{1}{T}\right). \tag{S14}$$

Moreover, by an argument similar to (S11), since $\mathbb{E}_{\xi_{i0},\lambda_0}(-v^{\alpha}_{ij})$ is bounded away from zero with probability one, we have

$$\frac{1}{Np}\sum_{i,j}\big\|z_j(\xi_{i0})-z^*_j(\widehat{k}_i)\big\|^2=O_p(\delta). \tag{S15}$$

Let $z'=(z'_1,\dots,z'_p)$, and $z^*(k)'=(z^*_1(k)',\dots,z^*_p(k)')$. Since ζ_{ij} are independent across i, with zero mean, conditional on $h_1,\dots,h_N$, $\xi_{10},\dots,\xi_{N0}$, λ_0, we thus have, denoting $\zeta_i=(\zeta'_{i1},\dots,\zeta'_{ip})'$,

$$\begin{aligned}
&\mathbb{E}\left[\left\|\frac{1}{Np}\sum_{i,j}\big(z_j(\xi_{i0})'-z^*_j(\widehat{k}_i)'\big)v_{ij}\right\|^2\right]\\
&\le 2O\left(\frac{1}{T}\right)\mathbb{E}\left[\frac{1}{Np}\sum_{i,j}\big\|z_j(\xi_{i0})-z^*_j(\widehat{k}_i)\big\|^2\right]\\
&\quad+2\mathbb{E}\left[\left\|\frac{1}{Np}\sum_{i,j}\big(z_j(\xi_{i0})'-z^*_j(\widehat{k}_i)'\big)\zeta_{ij}\right\|^2\right]\\
&=O\left(\frac{\delta}{T}\right)+\frac{2}{N^2p^2}\sum_i\mathbb{E}\big[\big(z'_i-z^*(\widehat{k}_i)'\big)\mathbb{E}_{h_i,\xi_{i0},\lambda_0}\big[\zeta_i\zeta'_i\big]\big(z_i-z^*(\widehat{k}_i)\big)\big]\\
&=O\left(\frac{\delta}{T}\right)+O\left(\frac{\delta p}{NT}\right)=O(\delta^2),
\end{aligned}$$

where we have used, in turn, the triangle and Cauchy–Schwarz inequalities, (S14), (S15), conditional independence of the ζ_i across i, part (v) in Assumption 4, and (S15) one more time. Note that, by part (v) in Assumption 4, $\|\mathbb{E}_{h_i,\xi_{i0},\lambda_0}[\zeta_i\zeta_i']\| \leq \operatorname{Tr}\mathbb{E}_{h_i,\xi_{i0},\lambda_0}[\zeta_i\zeta_i'] \leq p\max_j \operatorname{Tr}\mathbb{E}_{h_i,\xi_{i0},\lambda_0}[\zeta_{ij}\zeta_{ij}'] = O_p(p^2/T)$.

Turning to the second term in (S12), we have

$$\frac{1}{Np}\sum_{i,j}\big(z_j^*(\widehat{k}_i)' - \widetilde{z}_j(\widehat{k}_i)'\big)v_{ij} = \frac{1}{Np}\sum_{i,j}\big(z_j^*(\widehat{k}_i)' - \widetilde{z}_j(\widehat{k}_i)'\big)\overline{v}_j(\widehat{k}_i),$$

where by (S6) we have: $\frac{1}{Np}\sum_{i,j}\|\overline{v}_j(\widehat{k}_i)\|^2 = O_p(\delta)$. Moreover,

$$\frac{1}{Np}\sum_{i,j}\big\|z_j^*(\widehat{k}_i) - \widetilde{z}_j(\widehat{k}_i)\big\|^2 \leq \frac{2}{Np}\sum_{i,j}\big\|z_j(\xi_{i0}) - z_j^*(\widehat{k}_i)\big\|^2 + \frac{2}{Np}\sum_{i,j}\big\|z_j(\xi_{i0}) - \widetilde{z}_j(\widehat{k}_i)\big\|^2,$$

where the second term on the right-hand side is $O_p(\delta)$ due to (S11), and the first term is $O_p(\delta)$ due to (S15). This shows that $A_3 = O_p(\delta)$, hence that $A = O_p(\delta)$.

Let us now turn to B. Letting: $\pi_{ij}' = v_{ij}^\theta(v_{ij}^\alpha)^{-1} - \mathbb{E}_{\xi_{i0},\lambda_0}(v_{ij}^\theta)[\mathbb{E}_{\xi_{i0},\lambda_0}(v_{ij}^\alpha)]^{-1}$, we have

$$B = \frac{1}{Np}\sum_{i,j}\pi_{ij}'v_{ij}^\alpha\big(\widetilde{w}_j(\widehat{k}_i) + \widetilde{v}_j(\widehat{k}_i) + \widetilde{\alpha}_j(\widehat{k}_i) - \alpha_{i0}^j\big).$$

First, we have $\frac{1}{Np}\sum_{i,j}\pi_{ij}'v_{ij}^\alpha\widetilde{w}_j(\widehat{k}_i) = O_p(\delta)$. Next, we have $\frac{1}{Np}\sum_{i,j}\pi_{ij}'v_{ij}^\alpha\widetilde{v}_j(\widehat{k}_i) = \frac{1}{Np} \times \sum_{i,j}\widetilde{\pi}_j(\widehat{k}_i)'v_{ij}^\alpha\widetilde{v}_j(\widehat{k}_i)$, where $\widetilde{\pi}_j(k)$ is defined similar to $\widetilde{\alpha}_j(k)$. To see that this quantity is $O_p(\delta)$, note that, by the definition of $\widetilde{v}_j(k)$ and using (S4) and (S6):

$$\frac{1}{Np}\sum_{i,j}\big\|\widetilde{v}_j(\widehat{k}_i)\big\|^2 = O_p\left(\frac{1}{Np}\sum_{i,j}\big\|\overline{v}_j(\widehat{k}_i)\big\|^2\right) = O_p(\delta).$$

Moreover, letting $\tau_{ij} = \pi_{ij}'v_{ij}^\alpha$, we have

$$\frac{1}{Np}\sum_{i,j}\big\|\widetilde{\pi}_j(\widehat{k}_i)\big\|^2 = O_p\left(\frac{1}{Np}\sum_{i,j}\big\|\overline{\tau}_j(\widehat{k}_i)\big\|^2\right).$$

Now, the τ_{ij} are conditionally independent across i, with zero conditional mean given ξ_{i0}, λ_0:

$$\mathbb{E}_{\xi_{i0},\lambda_0}\big(\pi_{ij}'v_{ij}^\alpha\big) = \mathbb{E}_{\xi_{i0},\lambda_0}\Big(\big(v_{ij}^\theta\big(v_{ij}^\alpha\big)^{-1} - \mathbb{E}_{\xi_{i0},\lambda_0}\big(v_{ij}^\theta\big)\big[\mathbb{E}_{\xi_{i0},\lambda_0}\big(v_{ij}^\alpha\big)\big]^{-1}\big)v_{ij}^\alpha\Big) = 0.$$

Using an argument similar to the one we used to show (S6), and using Assumption 4(v) in the time-varying case, it thus follows that $\frac{1}{Np}\sum_{i,j}\|\widetilde{\pi}_j(\widehat{k}_i)\|^2 = O_p(\delta)$. Hence, by the Cauchy–Schwarz inequality: $\frac{1}{Np}\sum_{i,j}\pi_{ij}'v_{ij}^\alpha\widetilde{v}_j(\widehat{k}_i) = O_p(\delta)$.

We lastly bound the third term B_3 in B:

$$\frac{1}{Np}\sum_{i,j}\pi_{ij}'v_{ij}^\alpha\big(\widetilde{\alpha}_j(\widehat{k}_i) - \alpha_{i0}^j\big) = \frac{1}{Np}\sum_{i,j}\pi_{ij}'v_{ij}^\alpha\big[\big(\alpha_j^*(\widehat{k}_i) - \alpha_{i0}^j\big) + \big(\widetilde{\alpha}_j(\widehat{k}_i) - \alpha_j^*(\widehat{k}_i)\big)\big],$$

where $\widetilde{\alpha}_j(k)$ and $\alpha_j^*(k)$ are given by expressions similar to (S10) and (S13), with α_{i0}^j in place of $z_j(\xi_{i0})$ in those formulas. The first term is $O_p(\delta)$ since, similar to (S15): $\frac{1}{Np}\sum_{i,j}\|\alpha_j^*(\widehat{k}_i)-\alpha_{i0}^j\|^2=O_p(\delta)$, and the $\tau_{ij}=\pi_{ij}'v_{ij}^\alpha$ are conditionally independent across i with zero mean given ξ_{i0} and λ_0 (using a similar argument to the first term in (S12)). The second term is

$$\frac{1}{Np}\sum_{i,j}\pi_{ij}'v_{ij}^\alpha\big(\widetilde{\alpha}_j(\widehat{k}_i)-\alpha_j^*(\widehat{k}_i)\big)=\frac{1}{Np}\sum_{i,j}\widetilde{\pi}_j(\widehat{k}_i)'v_{ij}^\alpha\big(\widetilde{\alpha}_j(\widehat{k}_i)-\alpha_j^*(\widehat{k}_i)\big).$$

We have already shown that $\frac{1}{Np}\sum_{i,j}\|\widetilde{\pi}_j(\widehat{k}_i)\|^2=O_p(\delta)$. Moreover, using similar arguments to the ones we used to bound $\frac{1}{Np}\sum_{i,j}\|z_j^*(\widehat{k}_i)-\widetilde{z}_j(\widehat{k}_i)\|^2$ above, we have $\frac{1}{Np}\sum_{i,j}\|\widetilde{\alpha}_j(\widehat{k}_i)-\alpha_j^*(\widehat{k}_i)\|^2=O_p(\delta)$. This shows that $B_3=O_p(\delta)$, hence that $B=O_p(\delta)$. *Q.E.D.*

PROOF OF LEMMA A3: For given k, j, θ-differentiating: $\sum_i\mathbf{1}\{\widehat{k}_i=k\}v_{ij}(\widehat{\alpha}^j(k,\theta),\theta)=0$, and using (S4), we obtain

$$\frac{\partial\widehat{\alpha}^j(k,\theta)}{\partial\theta'}=\left(\sum_i\mathbf{1}\{\widehat{k}_i=k\}\big(-v_{ij}^\alpha\big(\widehat{\alpha}^j(\widehat{k}_i,\theta),\theta\big)\big)\right)^{-1}\sum_i\mathbf{1}\{\widehat{k}_i=k\}v_{ij}^\theta\big(\widehat{\alpha}^j(\widehat{k}_i,\theta),\theta\big)'. \quad \text{(S16)}$$

Let us define, at $\theta=\theta_0$ (and omitting θ_0 and α_{i0}^j from the notation),

$$\frac{\partial\widetilde{\alpha}^j(k)}{\partial\theta'}=\left(\sum_i\mathbf{1}\{\widehat{k}_i=k\}\big(-v_{ij}^\alpha\big)\right)^{-1}\sum_i\mathbf{1}\{\widehat{k}_i=k\}\big(v_{ij}^\theta\big)',$$

$$\frac{\partial\widetilde{\alpha}_*^j(k)}{\partial\theta'}=\left(\sum_i\mathbf{1}\{\widehat{k}_i=k\}\big(-v_{ij}^\alpha\big)\right)^{-1}\sum_i\mathbf{1}\{\widehat{k}_i=k\}\big(-v_{ij}^\alpha\big)\underbrace{\big[\mathbb{E}_{\xi_{i0},\lambda_0}\big(-v_{ij}^\alpha\big)\big]^{-1}\mathbb{E}_{\xi_{i0},\lambda_0}\big(v_{ij}^\theta\big)'}_{=\frac{\partial\overline{\alpha}^j(\xi_{i0})}{\partial\theta'}}.$$

Using (A1) and (S4), we have $\frac{1}{Np}\sum_{i,j}\|\frac{\partial\widehat{\alpha}^j(\widehat{k}_i)}{\partial\theta'}-\frac{\partial\widetilde{\alpha}^j(\widehat{k}_i)}{\partial\theta'}\|^2=o_p(1)$. Moreover,

$$\frac{\partial\widetilde{\alpha}^j(k)}{\partial\theta'}-\frac{\partial\widetilde{\alpha}_*^j(k)}{\partial\theta'}=\left(\frac{\sum_i\mathbf{1}\{\widehat{k}_i=k\}\big(-v_{ij}^\alpha\big)}{\sum_i\mathbf{1}\{\widehat{k}_i=k\}}\right)^{-1}\left(\frac{\sum_i\mathbf{1}\{\widehat{k}_i=k\}\tau_{ij}'}{\sum_i\mathbf{1}\{\widehat{k}_i=k\}}\right),$$

where $\tau_{ij}'=(v_{ij}^\theta)'-(-v_{ij}^\alpha)[\mathbb{E}_{\xi_{i0},\lambda_0}(-v_{ij}^\alpha)]^{-1}\mathbb{E}_{\xi_{i0},\lambda_0}(v_{ij}^\theta)'$ are conditionally independent across i, with zero mean given ξ_{i0} and λ_0. Hence, using (S4), and a similar argument to the one we used to show (S6), we have $\frac{1}{Np}\sum_{i,j}\|\frac{\partial\widetilde{\alpha}^j(\widehat{k}_i)}{\partial\theta'}-\frac{\partial\widetilde{\alpha}_*^j(\widehat{k}_i)}{\partial\theta'}\|^2=o_p(1)$. Lastly, using (S4) we have, as in (S11): $\frac{1}{Np}\sum_{i,j}\|\frac{\partial\widetilde{\alpha}_*^j(\widehat{k}_i)}{\partial\theta'}-\frac{\partial\overline{\alpha}^j(\xi_{i0})}{\partial\theta'}\|^2=o_p(1)$. Combining results shows (A5). *Q.E.D.*

PROOF OF LEMMA A4: In the following, we evaluate all functions at θ_0, and omit θ_0 from the notation. In particular, $\widehat{\alpha}_i$ is a shorthand for $\widehat{\alpha}_i(\theta_0)$. We will use the notation

$\widehat{w}_i = -v_i^{\alpha}(\widehat{\alpha}_i)$. The choice of $K = \widehat{K}$ with $\gamma = o(1)$ implies that

$$\frac{1}{N}\sum_i \big\| h_i - \widehat{h}(\widehat{k}_i)\big\|^2 = o_p\left(\frac{1}{T}\right). \tag{S17}$$

We also have $\frac{1}{N}\sum_i \|\widehat{\alpha}(\widehat{k}_i) - \widehat{\alpha}_i\|^2 = O_p(\frac{1}{T})$. Let, for all k,

$$\widetilde{\alpha}(k) = \left(\sum_i \mathbf{1}\{\widehat{k}_i = k\}\widehat{w}_i\right)^{-1}\sum_i \mathbf{1}\{\widehat{k}_i = k\}\widehat{w}_i\widehat{\alpha}_i. \tag{S18}$$

Expanding $\sum_i \mathbf{1}\{\widehat{k}_i = k\}v_i(\widehat{\alpha}(k)) = 0$ around $\widehat{\alpha}_i$, using that $v_i(\widehat{\alpha}_i) = 0$, we obtain

$$\widehat{\alpha}(k) = \widetilde{\alpha}(k) + \frac{1}{2}\left[\sum_i \mathbf{1}\{\widehat{k}_i = k\}\widehat{w}_i\right]^{-1}\sum_i \mathbf{1}\{\widehat{k}_i = k\}v_i^{\alpha\alpha}\big(a_i(k)\big)\big(\widehat{\alpha}(\widehat{k}_i) - \widehat{\alpha}_i\big)^{\otimes 2},$$

where $a_i(k)$ lies between $\widehat{\alpha}_i$ and $\widehat{\alpha}(k)$, and $v_i^{\alpha\alpha}(a_i(k))$ is a matrix of third derivatives with $(\dim \alpha_{i0})^2$ columns.

To see that (A6) holds, we rely on the following decomposition:

$$\begin{aligned}
\frac{\partial}{\partial\theta}\bigg|_{\theta_0}\Delta L(\theta) &= \frac{1}{N}\sum_i \frac{\partial \ell_i\big(\widehat{\alpha}(\widehat{k}_i)\big)}{\partial\theta} - \frac{1}{N}\sum_i \frac{\partial \ell_i(\widehat{\alpha}_i)}{\partial\theta} \\
&= \frac{1}{N}\sum_i v_i^{\theta}(\widehat{\alpha}_i)\big(\widehat{\alpha}(\widehat{k}_i) - \widehat{\alpha}_i\big) + \frac{1}{2N}\sum_i v_i^{\theta\alpha}(a_i)\big(\widehat{\alpha}(\widehat{k}_i) - \widehat{\alpha}_i\big)^{\otimes 2} \\
&= \frac{1}{N}\sum_i v_i^{\theta}(\widehat{\alpha}_i)\big(\widetilde{\alpha}(\widehat{k}_i) - \widehat{\alpha}_i\big) + \frac{1}{2N}\sum_i v_i^{\theta\alpha}(a_i)\big(\widehat{\alpha}(\widehat{k}_i) - \widehat{\alpha}_i\big)^{\otimes 2} \\
&\quad + \frac{1}{2N}\sum_i v_i^{\theta}(\widehat{\alpha}_i)\big(\mathbb{E}_{\widehat{k}_i}[\widehat{w}_{i'}]\big)^{-1}\mathbb{E}_{\widehat{k}_i}\big[v_{i'}^{\alpha\alpha}\big(a_{i'}(\widehat{k}_{i'})\big)\big(\widehat{\alpha}(\widehat{k}_{i'}) - \widehat{\alpha}_{i'}\big)^{\otimes 2}\big] \\
&= \underbrace{\frac{1}{N}\sum_i v_i^{\theta}(\widehat{\alpha}_i)\big(\widetilde{\alpha}(\widehat{k}_i) - \widehat{\alpha}_i\big)}_{=A_1} + \underbrace{\frac{1}{2N}\sum_i v_i^{\theta\alpha}(a_i)\big(\widetilde{\alpha}(\widehat{k}_i) - \widehat{\alpha}_i\big)^{\otimes 2}}_{=A_2} + o_p\left(\frac{1}{T}\right) \\
&\quad \underbrace{+ \frac{1}{2N}\sum_i v_i^{\theta}(\widehat{\alpha}_i)\big(\mathbb{E}_{\widehat{k}_i}[\widehat{w}_{i'}]\big)^{-1}\mathbb{E}_{\widehat{k}_i}\big[v_{i'}^{\alpha\alpha}\big(a_{i'}(\widehat{k}_{i'})\big)\big(\widetilde{\alpha}(\widehat{k}_{i'}) - \widehat{\alpha}_{i'}\big)^{\otimes 2}\big]}_{=A_3},
\end{aligned}$$

where a_i and $a_i(\widehat{k}_i)$ lie between $\widehat{\alpha}_i$ and $\widehat{\alpha}(\widehat{k}_i)$, and $\mathbb{E}_k$ denotes a mean in group $\widehat{k}_i = k$. Let $\gamma(h) = \{\mathbb{E}_{h_i=h}(\widehat{w}_i)\}^{-1}\mathbb{E}_{h_i=h}(\widehat{w}_i\widehat{\alpha}_i)$ and $\nu_i = \widehat{\alpha}_i - \gamma(h_i)$. Let $\widehat{g}_i = v_i^{\theta}(\widehat{\alpha}_i)(\widehat{w}_i)^{-1}$, $\lambda(h) = \mathbb{E}_{h_i=h}(\widehat{g}_i\widehat{w}_i)\{\mathbb{E}_{h_i=h}(\widehat{w}_i)\}^{-1}$, and $\tau_i = \widehat{g}_i' - \lambda(h_i)'$. Using (S17) we can show, using that γ is Lipschitz-continuous, that $\frac{1}{N}\sum_i \|\gamma(h_i) - \widetilde{\gamma}(\widehat{k}_i)\|^2 = o_p(\frac{1}{T})$.[3] Moreover, we

[3]Here, $\widetilde{\gamma}(k)$, $\widetilde{\lambda}(k)$, $\widetilde{\nu}(k)$, and $\widetilde{\tau}(k)$ are defined similar to $\widetilde{\alpha}(k)$ in (S18), with $\gamma(h_i)$, $\lambda(h_i)$, ν_i, and τ_i, respectively, replacing $\widehat{\alpha}_i$ in that formula.

have $\mathbb{E}[\widehat{w}_i\nu_i \mid h_1,\dots,h_N]=\mathbb{E}_{h_i}[\widehat{w}_i\alpha_i]-\mathbb{E}_{h_i}[\widehat{w}_i\widehat{\alpha}_i]=0$. Similar arguments to the proof of Lemma A1 give $\frac{1}{N}\sum_i\|\mathbb{E}_{\widehat{k}_i}[\widehat{w}_{i'}\nu_{i'}]\|^2=O_p(\frac{K}{NT})=o_p(\frac{1}{T})$. Hence, $\frac{1}{N}\sum_i\|\widetilde{\nu}(\widehat{k}_i)\|^2=\frac{1}{N}\sum_i\|(\mathbb{E}_{\widehat{k}_i}[\widehat{w}_{i'}])^{-1}\mathbb{E}_{\widehat{k}_i}[\widehat{w}_{i'}\nu_{i'}]\|^2=o_p(\frac{1}{T})$. Likewise, we have $\frac{1}{N}\sum_i\|\lambda(h_i)-\widetilde{\lambda}(\widehat{k}_i)\|^2=o_p(\frac{1}{T})$, and: $\frac{1}{N}\sum_i\|\widetilde{\tau}(\widehat{k}_i)\|^2=o_p(\frac{1}{T})$.

Let us now expand the three terms A_1, A_2, A_3 in the above decomposition:

$$
\begin{aligned}
A_1&=\frac{1}{N}\sum_i\widehat{g}_i\widehat{w}_i\big(\widetilde{\alpha}(\widehat{k}_i)-\widehat{\alpha}_i\big)=\frac{1}{N}\sum_i\big(\widehat{g}_i-\widetilde{g}(\widehat{k}_i)\big)\widehat{w}_i\big(\widetilde{\alpha}(\widehat{k}_i)-\widehat{\alpha}_i\big)\\
&=-\frac{1}{N}\sum_i\big(\lambda(h_i)-\widetilde{\lambda}(\widehat{k}_i)+\tau_i'-\widetilde{\tau}(\widehat{k}_i)'\big)\widehat{w}_i\big(\gamma(h_i)-\widetilde{\gamma}(\widehat{k}_i)+\nu_i-\widetilde{\nu}(\widehat{k}_i)\big)\\
&=-\frac{1}{N}\sum_i\tau_i'\widehat{w}_i\nu_i+o_p\Big(\frac{1}{T}\Big)=-\frac{1}{N}\sum_i\tau_i'\mathbb{E}_{\xi_{i0}}\big(-v_i^{\alpha}(\alpha_{i0})\big)\nu_i+o_p\Big(\frac{1}{T}\Big),\\
A_2&=\frac{1}{2N}\sum_i\mathbb{E}_{\xi_{i0}}\big(v_i^{\theta\alpha}(\alpha_{i0})\big)\big(\widetilde{\alpha}(\widehat{k}_i)-\widehat{\alpha}_i\big)^{\otimes 2}+o_p\Big(\frac{1}{T}\Big)\\
&=\frac{1}{2N}\sum_i\mathbb{E}_{\xi_{i0}}\big(v_i^{\theta\alpha}(\alpha_{i0})\big)\big(\widetilde{\gamma}(\widehat{k}_i)-\gamma(h_i)+\widetilde{\nu}(\widehat{k}_i)-\nu_i\big)^{\otimes 2}+o_p\Big(\frac{1}{T}\Big)\\
&=\frac{1}{2N}\sum_i\mathbb{E}_{\xi_{i0}}\big(v_i^{\theta\alpha}(\alpha_{i0})\big)\nu_i^{\otimes 2}+o_p\Big(\frac{1}{T}\Big),\\
A_3&=\frac{1}{2N}\sum_i\mathbb{E}_{\xi_{i0}}\big(v_i^{\theta}(\alpha_{i0})\big)\big[\mathbb{E}_{\xi_{i0}}\big(-v_i^{\alpha}(\alpha_{i0})\big)\big]^{-1}\mathbb{E}_{\xi_{i0}}\big[v_i^{\alpha\alpha}(\alpha_{i0})\big]\nu_i^{\otimes 2}+o_p\Big(\frac{1}{T}\Big).
\end{aligned}
$$

Combining, we get

$$
\begin{aligned}
&\frac{\partial}{\partial\theta}\bigg|_{\theta_0}\Delta L(\theta)\\
&\quad=-\frac{1}{N}\sum_i\tau_i'\mathbb{E}_{\xi_{i0}}\big(-v_i^{\alpha}(\alpha_{i0})\big)\nu_i+o_p\Big(\frac{1}{T}\Big)\\
&\qquad+\frac{1}{2N}\sum_i\big[\mathbb{E}_{\xi_{i0}}\big(v_i^{\theta\alpha}(\alpha_{i0})\big)+\mathbb{E}_{\xi_{i0}}\big(v_i^{\theta}(\alpha_{i0})\big)\big[\mathbb{E}_{\xi_{i0}}\big(-v_i^{\alpha}(\alpha_{i0})\big)\big]^{-1}\mathbb{E}_{\xi_{i0}}\big[v_i^{\alpha\alpha}(\alpha_{i0})\big]\big]\nu_i^{\otimes 2}.
\end{aligned}
$$

Now, $\frac{\partial\widehat{\alpha}_i(\theta_0)}{\partial\theta'}=\widehat{g}_i'$, and

$$
\begin{aligned}
&\frac{\partial}{\partial\theta'}\bigg|_{\theta_0}\operatorname{vec}\mathbb{E}_{\xi_{i0}}\big[-v_i^{\alpha}\big(\overline{\alpha}(\theta,\xi_{i0}),\theta\big)\big]\\
&\quad=-\big(\mathbb{E}_{\xi_{i0}}\big(v_i^{\theta\alpha}(\alpha_{i0})\big)+\mathbb{E}_{\xi_{i0}}\big(v_i^{\theta}(\alpha_{i0})\big)\big[\mathbb{E}_{\xi_{i0}}\big(-v_i^{\alpha}(\alpha_{i0})\big)\big]^{-1}\mathbb{E}_{\xi_{i0}}\big[v_i^{\alpha\alpha}(\alpha_{i0})\big]\big)'.
\end{aligned}
$$

Let $\omega_i = \{\mathbb{E}_{h_i}(\widehat{w}_i)\}^{-1}\widehat{w}_i$ and $\widetilde{\nu}_i(\theta) = \widehat{\alpha}_i(\theta) - \mathbb{E}_{h_i}(\omega_i \widehat{\alpha}_i(\theta))$. Combining the above with the expression of the bias of the FE score, we obtain

$$\frac{\partial}{\partial\theta}\bigg|_{\theta_0} \Delta L(\theta) = -\frac{\partial}{\partial\theta}\bigg|_{\theta_0} \frac{1}{2N}\sum_i \widetilde{\nu}_i(\theta)'\mathbb{E}_{\xi_{i0}}\big[-v_i^{\alpha}\big(\overline{\alpha}(\theta, \xi_{i0}), \theta\big)\big]\widetilde{\nu}_i(\theta) + o_p\left(\frac{1}{T}\right). \tag{S19}$$

Lastly, let $\widehat{\alpha}_i(\theta) = \mathbb{E}_{h_i}(\widehat{\alpha}_i(\theta)) + \nu_i(\theta)$, and $\omega_i = \mathbb{E}_{h_i}(\omega_i) + \eta_i = 1 + \eta_i$. We have $\widetilde{\nu}_i(\theta) = \nu_i(\theta) - \mathbb{E}_{h_i}(\eta_i \nu_i(\theta))$, from which it follows that $\frac{1}{N}\sum_i \|\widetilde{\nu}_i(\theta_0) - \nu_i(\theta_0)\|^2 = o_p(1/T)$. Likewise $\frac{1}{N}\sum_i \|\frac{\partial \widetilde{\nu}_i(\theta_0)}{\partial\theta'} - \frac{\partial \nu_i(\theta_0)}{\partial\theta'}\|^2 = o_p(1/T)$. Hence, (S19) implies (A6). *Q.E.D.*

S2. COMPLEMENTS AND EXTENSIONS

S2.1. *Average Effects*

Let $m_i(\alpha_i, \theta) = \frac{1}{T}\sum_{t=1}^T m(X_{it}, \alpha_i, \theta)$ in the time-invariant case, and $m_i(\alpha_i, \theta) = \frac{1}{T} \times \sum_{t=1}^T m(X_{it}, \alpha_{it}, \theta)$ in the time-varying case. Let $\widehat{M} = \frac{1}{N}\sum_i m_i(\widehat{\alpha}(\widehat{k}_i), \widehat{\theta})$ be the GFE estimator of $M_0 = \frac{1}{N}\sum_i m_i(\alpha_{i0}, \theta_0)$. We use a common notation as in the proofs of Theorems 1 and 2, and denote $m_{ij}(\alpha_i^j, \theta) = m_i(\alpha_i, \theta)$ in the time-invariant case, and $m_{ij}(\alpha_i^j, \theta) = m(X_{it}, \alpha_{it}, \theta)$ in the time-varying case.

ASSUMPTION S1—Average Effects:
(i) *$m_{ij}(\alpha, \theta)$ is twice differentiable in both its arguments, for all i, j.*
(ii) $\max_{i,j}\sup_{\alpha,\theta}\|m_{ij}(\alpha, \theta)\| = O_p(1)$, *and similarly for the first two derivatives of m_{ij}; $\max_j \sup_{\widetilde{\xi},\lambda}\|\frac{\partial}{\partial\xi'}|_{\xi=\widetilde{\xi}}\mathbb{E}_{\xi_{i0}=\xi,\lambda_0=\lambda}(\frac{\partial m_{ij}(\alpha_{i0}^j, \theta_0)}{\partial\alpha})\| = O(1)$; and, letting $\tau_{ij}^m = \frac{\partial m_{ij}(\alpha_{i0}^j, \theta_0)}{\partial\alpha'} - \mathbb{E}_{\xi_{i0},\lambda_0}[\frac{\partial m_{ij}(\alpha_{i0}^j, \theta_0)}{\partial\alpha'}]\mathbb{E}_{\xi_{i0},\lambda_0}[v_{ij}^{\alpha}(\alpha_{i0}^j, \theta_0)]v_{ij}^{\alpha}(\alpha_{i0}^j, \theta_0)$, the function $\mathbb{E}_{h_i=h,\xi_{i0}=\xi,\lambda_0=\lambda}(\mathrm{vec}\,\tau_{ij}^m)$ is twice differentiable with respect to h, with first and second derivatives that are uniformly bounded in j, ξ, λ, and h, and $\|\mathrm{Var}_{h_i=h,\xi_{i0}=\xi,\lambda_0=\lambda}(\mathrm{vec}\,\tau_{ij}^m)\| = O(\frac{p}{T})$, uniformly in j, ξ, λ, and h.*

Let s_i and H as in Theorem 1 or 2, and let $\overline{s} = \frac{1}{N}\sum_i s_i$. Define

$$s_i^m = \frac{1}{p}\sum_j\Bigg\{\mathbb{E}_{\xi_{i0},\lambda_0}\left(\frac{\partial m_{ij}}{\partial\alpha'}\right)\left[\mathbb{E}_{\xi_{i0},\lambda_0}\left(-\frac{\partial^2\ell_{ij}}{\partial\alpha\partial\alpha'}\right)\right]^{-1}\frac{\partial\ell_{ij}}{\partial\alpha} + \mathbb{E}_{\xi_{i0},\lambda_0}\left(\frac{\partial m_{ij}}{\partial\theta'}\right)H^{-1}\overline{s}$$
$$+ \mathbb{E}_{\xi_{i0},\lambda_0}\left(\frac{\partial m_{ij}}{\partial\alpha'}\right)\left[\mathbb{E}_{\xi_{i0},\lambda_0}\left(-\frac{\partial^2\ell_{ij}}{\partial\alpha\partial\alpha'}\right)\right]^{-1}\mathbb{E}_{\xi_{i0},\lambda_0}\left(\frac{\partial^2\ell_{ij}}{\partial\alpha\partial\theta'}\right)H^{-1}\overline{s}\Bigg\}.$$

COROLLARY S1: *Let the conditions of Theorem 1 or 2 hold, and let Assumption S1 hold. Then, as N, T, K tend to infinity such that $Kp/(NT)$ tends to zero*:

$$\widehat{M} = M_0 + \frac{1}{N}\sum_i s_i^m + O_p\left(\frac{1}{T}\right) + O_p\left(\frac{Kp}{NT}\right) + O_p(K^{-\frac{2}{d}}) + o_p\left(\frac{1}{\sqrt{NT}}\right).$$

PROOF: We have, by a Taylor expansion,

$$\begin{aligned}\widehat{M}-M_0 &= \frac{1}{Np}\sum_{i,j} m_{ij}\big(\widehat{\alpha}^j(\widehat{k}_i,\widehat{\theta}),\widehat{\theta}\big)-\frac{1}{Np}\sum_{i,j} m_{ij}\big(\alpha_{i0}^j,\theta_0\big)\\ &= \frac{1}{Np}\sum_{i,j}\frac{\partial m_{ij}\big(\alpha_{i0}^j,\theta_0\big)}{\partial\alpha'}\big(\widehat{\alpha}^j(\widehat{k}_i,\widehat{\theta})-\alpha_{i0}^j\big)\\ &\quad+\frac{1}{Np}\sum_{i,j}\frac{\partial m_{ij}\big(\alpha_{i0}^j,\theta_0\big)}{\partial\theta'}(\widehat{\theta}-\theta_0)+O_p(\delta),\end{aligned}$$

where δ is defined as in the proofs of Theorems 1 and 2.

Using similar arguments to the ones we used to establish Lemma A2, under Assumption S1 we have (recall that $\overline{\alpha}^j(\theta_0,\xi_{i0})=\alpha_{i0}^j$):

$$\begin{aligned}&\frac{1}{Np}\sum_{i,j}\frac{\partial m_{ij}\big(\alpha_{i0}^j,\theta_0\big)}{\partial\alpha'}\big(\widehat{\alpha}^j(\widehat{k}_i,\theta_0)-\overline{\alpha}^j(\theta_0,\xi_{i0})\big)\\ &\quad+\frac{1}{Np}\sum_{i,j}\mathbb{E}_{\xi_{i0},\lambda_0}\left[\frac{\partial m_{ij}\big(\alpha_{i0}^j,\theta_0\big)}{\partial\alpha'}\right]\mathbb{E}_{\xi_{i0},\lambda_0}\big[v_{ij}^{\alpha}\big(\alpha_{i0}^j,\theta_0\big)\big]^{-1}v_{ij}\big(\alpha_{i0}^j,\theta_0\big)=O_p(\delta).\end{aligned}$$

Moreover, using (A5) and Assumption S1 we obtain:

$$\begin{aligned}&\frac{1}{Np}\sum_{i,j}\frac{\partial m_{ij}\big(\alpha_{i0}^j,\theta_0\big)}{\partial\alpha'}\big\{\big(\widehat{\alpha}^j(\widehat{k}_i,\widehat{\theta})-\overline{\alpha}^j(\widehat{\theta},\xi_{i0})\big)-\big(\widehat{\alpha}^j(\widehat{k}_i,\theta_0)-\overline{\alpha}^j(\theta_0,\xi_{i0})\big)\big\}\\ &\quad=o_p\big(\|\widehat{\theta}-\theta_0\|\big)+O_p(\delta)=o_p\left(\frac{1}{\sqrt{NT}}\right)+O_p(\delta).\end{aligned}$$

Combining, we obtain

$$\begin{aligned}\widehat{M}-M_0 &= \frac{1}{Np}\sum_{i,j}\frac{\partial m_{ij}\big(\alpha_{i0}^j,\theta_0\big)}{\partial\alpha'}\big(\widehat{\alpha}^j(\widehat{k}_i,\widehat{\theta})-\widehat{\alpha}^j(\widehat{k}_i,\theta_0)\big)\\ &\quad+\frac{1}{Np}\sum_{i,j}\frac{\partial m_{ij}\big(\alpha_{i0}^j,\theta_0\big)}{\partial\alpha'}\big(\widehat{\alpha}^j(\widehat{k}_i,\theta_0)-\alpha_{i0}^j\big)\\ &\quad+\frac{1}{Np}\sum_{i,j}\frac{\partial m_{ij}\big(\alpha_{i0}^j,\theta_0\big)}{\partial\theta'}(\widehat{\theta}-\theta_0)+O_p(\delta)\\ &= \frac{1}{Np}\sum_{i,j}\frac{\partial m_{ij}\big(\alpha_{i0}^j,\theta_0\big)}{\partial\alpha'}\big(\overline{\alpha}^j(\widehat{\theta},\xi_{i0})-\overline{\alpha}^j(\theta_0,\xi_{i0})\big)\end{aligned}$$

$$+\frac{1}{Np}\sum_{i,j}\mathbb{E}_{\xi_{i0},\lambda_0}\left[\frac{\partial m_{ij}(\alpha_{i0}^j,\theta_0)}{\partial\alpha'}\right]\mathbb{E}_{\xi_{i0},\lambda_0}\left[-v_{ij}^{\alpha}(\alpha_{i0}^j,\theta_0)\right]^{-1}v_{ij}(\alpha_{i0}^j,\theta_0)$$

$$+\frac{1}{Np}\sum_{i,j}\frac{\partial m_{ij}(\alpha_{i0}^j,\theta_0)}{\partial\theta'}(\widehat{\theta}-\theta_0)+O_p(\delta)+o_p\left(\frac{1}{\sqrt{NT}}\right).$$

The result comes from expanding $\overline{\alpha}^j(\widehat{\theta},\xi_{i0})$ around θ_0, and then substituting $\widehat{\theta}-\theta_0$ using Theorem 1 or 2. *Q.E.D.*

S2.2. *Two-Way GFE*

We have the following lemma, whose proof is analogous to that of Lemma 1.

LEMMA S1: *Suppose that there exist random vectors $h_i=\frac{1}{T}\sum_t h(Y_{it},X_{it})$ and $w_t=\frac{1}{N}\sum_i w(Y_{it},X_{it})$, with fixed dimensions, and Lipschitz-continuous functions φ and ϕ, such that $h_i=\varphi(\xi_{i0})+o_p(1)$, $\frac{1}{N}\sum_i\|h_i-\varphi(\xi_{i0})\|^2=O_p(1/T)$, $w_t=\phi(\lambda_{t0})+o_p(1)$, and $\frac{1}{T}\sum_t\|w_t-\phi(\lambda_{t0})\|^2=O_p(1/N)$ as N, T tend to infinity. Then we have, as N, T, K tend to infinity: $\frac{1}{N}\sum_i\|\widehat{h}(\widehat{k}_i)-\varphi(\xi_{i0})\|^2=O_p(\frac{1}{T})+O_p(B_\xi(K))$, and, as N, T, L tend to infinity: $\frac{1}{T}\sum_t\|\widehat{w}(\widehat{l}_t)-\phi(\lambda_{t0})\|^2=O_p(\frac{1}{N})+O_p(B_\lambda(L))$, where $B_\lambda(L)$ is defined analogously to $B_\xi(K)$.*

For all θ, ξ, and λ, let $\overline{\alpha}(\theta,\xi,\lambda)=\operatorname{argmax}_\alpha\mathbb{E}_{\xi_{i0}=\xi,\lambda_{t0}=\lambda}(\ell_{it}(\alpha,\theta))$. In addition, let $\xi_0=(\xi_{10}',\dots,\xi_{N0}')'$, $\lambda_0=(\lambda_{10}',\dots,\lambda_{T0}')'$, $\underline{h}=(h_1',\dots,h_N')'$, and $\underline{w}=(w_1',\dots,w_T')'$.

ASSUMPTION S2—regularity, two-way:

(i) $(Y_{it}',X_{it}')'$, $i=1,..,N$, $t=1,\dots,T$, *are i.i.d. given* ξ_0 *and* λ_0, ξ_{i0} *are i.i.d., and* λ_{t0} *are i.i.d.*; $\ell_{it}(\alpha,\theta)$ *is three times differentiable in* (θ,α); Θ *is compact, the supports of* ξ_{i0} *and* λ_{t0} *are compact, and* θ_0 *belongs to the interior of* Θ.

(ii) N, T *tend jointly to infinity*; $\sup_{\xi,\lambda,\alpha,\theta}|\mathbb{E}_{\xi_{i0}=\xi,\lambda_{t0}=\lambda}(\ell_{it}(\alpha,\theta))|=O(1)$, *and similarly for the first three derivatives of* ℓ_{it}; *the minimum (resp., maximum) eigenvalue of* $(-\frac{\partial^2\ell_{it}(\alpha,\theta)}{\partial\alpha\partial\alpha'})$ *is bounded away from zero (resp., infinity) with probability one uniformly in* i, t, α, θ, *and the third derivatives of* $\ell_{it}(\alpha,\theta)$ *are* $O_p(1)$, *uniformly in* i, t, α, θ; $\frac{1}{NT}\sum_{i,t}[\ell_{it}(\alpha_{it0},\theta_0)-\mathbb{E}_{\xi_{i0},\lambda_{t0}}(\ell_{it}(\alpha_{it0},\theta_0))]^2=O_p(1)$, *and similarly for the first three derivatives of* ℓ_{it}.

(iii) $\inf_{\xi,\lambda,\theta}\mathbb{E}_{\xi_{i0}=\xi,\lambda_{t0}=\lambda}(-\frac{\partial^2\ell_{it}(\overline{\alpha}(\theta,\xi,\lambda),\theta)}{\partial\alpha\partial\alpha'})>0$; $\mathbb{E}[\frac{1}{NT}\sum_{i,t}\ell_{it}(\overline{\alpha}(\theta,\xi_{i0},\lambda_{t0}),\theta)]$ *has a unique maximum at* θ_0 *on* Θ, *and its second derivative is* $-H<0$.

(iv) $\frac{\partial}{\partial\xi'}|_{\widetilde{\xi}}\mathbb{E}_{\xi_{i0}=\xi,\lambda_{t0}=\lambda}(\operatorname{vec}\frac{\partial^2\ell_{it}(\alpha,\theta_0)}{\partial\theta\partial\alpha'})=O(1)$; $\frac{\partial}{\partial\lambda'}|_{\widetilde{\lambda}}\mathbb{E}_{\xi_{i0}=\xi,\lambda_{t0}=\lambda}(\operatorname{vec}\frac{\partial^2\ell_{it}(\alpha,\theta_0)}{\partial\theta\partial\alpha'})=O(1)$; $\frac{\partial}{\partial\xi'}|_{\widetilde{\xi}}\mathbb{E}_{\xi_{i0}=\xi,\lambda_{t0}=\lambda}(\operatorname{vec}\frac{\partial^2\ell_{it}(\alpha,\theta_0)}{\partial\alpha\partial\alpha'})=O(1)$; $\frac{\partial}{\partial\lambda'}|_{\widetilde{\lambda}}\mathbb{E}_{\xi_{i0}=\xi,\lambda_{t0}=\lambda}(\operatorname{vec}\frac{\partial^2\ell_{it}(\alpha,\theta_0)}{\partial\alpha\partial\alpha'})=O(1)$; $\frac{\partial}{\partial\xi'}|_{\widetilde{\xi}}\mathbb{E}_{\xi_{i0}=\xi,\lambda_{t0}=\lambda}(\frac{\partial\ell_{it}(\overline{\alpha}(\theta,\xi,\lambda),\theta)}{\partial\alpha})=O(1)$; $\frac{\partial}{\partial\lambda'}|_{\widetilde{\lambda}}\mathbb{E}_{\xi_{i0}=\xi,\lambda_{t0}=\lambda}(\frac{\partial\ell_{it}(\overline{\alpha}(\theta,\xi,\lambda),\theta)}{\partial\alpha})=O(1)$, *uniformly in* ξ, $\widetilde{\xi}$, λ, $\widetilde{\lambda}$, α, θ.

(v) *Let* $U_{it}(\theta)\in\{\frac{\partial\ell_{it}(\overline{\alpha}(\theta,\xi,\lambda),\theta)}{\partial\alpha},\operatorname{vec}\frac{\partial}{\partial\theta'}|_{\theta_0}\frac{\partial\ell_{it}(\overline{\alpha}(\theta,\xi,\lambda),\theta)}{\partial\alpha}\}$. $\mathbb{E}_{h_i=h,\xi_{i0}=\xi,w_t=w,\lambda_{t0}=\lambda}(U_{it}(\theta))$ *is twice differentiable with respect to* h, w, *with first and second derivatives uniformly bounded in* h, w, ξ, λ, θ; $\|\operatorname{Cov}_{\underline{h}=\widetilde{h},\xi_0=\widetilde{\xi},\underline{w}=\widetilde{w},\lambda_0=\widetilde{\lambda}}(U_{it}(\theta),U_{i's}(\theta))\|=\mathbf{1}\{i=i',t=s\}O(1)+\mathbf{1}\{i=i'\}O(1/T)+\mathbf{1}\{t=s\}O(1/N)+O(1/(NT))$ *uniformly in* i, i', t, s, $\widetilde{h}$, $\widetilde{w}$, $\widetilde{\xi}$, $\widetilde{\lambda}$, θ; $\frac{1}{NT}\sum_{i,t}\|\mathbb{E}_{\underline{h},\xi_0,\underline{w},\lambda_0}(U_{it}(\theta))-\mathbb{E}_{h_i,\xi_{i0},w_t,\lambda_{t0}}(U_{it}(\theta))\|^2=O_p(\frac{1}{T}+\frac{1}{N})$ *uniformly in* θ.

COROLLARY S2: *Let the conditions in Lemma S1 hold. Suppose that* $B_\xi(K) = O_p(K^{-\frac{2}{d}})$ *and* $B_\lambda(L) = O_p(L^{-\frac{2}{d_\lambda}})$. *Suppose that* α *and* μ *are Lipschitz-continuous in both arguments, and that there exist two Lipschitz-continuous functions* ψ *and* Ψ *such that* $\xi_{i0} = \psi(\varphi(\xi_{i0}))$ *and* $\lambda_{t0} = \Psi(\phi(\lambda_{t0}))$. *Lastly, let Assumption S2 hold. Then, as* N, T, K, L *tend to infinity such that* $KL/(NT)$ *tends to zero, we have*

$$\widehat{\theta} = \theta_0 + H^{-1}\frac{1}{N}\sum_i s_i + O_p\left(\frac{1}{T} + \frac{1}{N} + \frac{KL}{NT}\right) + O_p\left(K^{-\frac{2}{d}} + L^{-\frac{2}{d_\lambda}}\right) + o_p\left(\frac{1}{\sqrt{NT}}\right).$$

PROOF: The proof closely follows the steps of that of Theorem 2. Here, we outline the main differences. Let $\delta = \frac{1}{T} + \frac{1}{N} + \frac{KL}{NT} + K^{-\frac{2}{d}} + L^{-\frac{2}{d_\lambda}}$. To show consistency, a key step is to show, for all $\theta \in \Theta$:

$$\frac{1}{NT}\sum_{i,t}\left\|\overline{v}(\widehat{k}_i, \widehat{l}_t, \theta)\right\|^2 = O_p(\delta), \tag{S20}$$

where $\overline{v}(k, l, \theta)$ denotes the mean of $v_{it}(\overline{\alpha}(\theta, \xi_{i0}, \lambda_{t0}), \theta)$ in the intersection of groups $\widehat{k}_i = k$ and $\widehat{l}_t = l$. Let $\rho_{it}(\widetilde{h}, \widetilde{\xi}, \widetilde{w}, \widetilde{\lambda}, \theta) = \mathbb{E}_{\underline{h}=\widetilde{h}, \xi_0=\widetilde{\xi}, \underline{w}=\widetilde{w}, \lambda_0=\widetilde{\lambda}}(v_{it}(\overline{\alpha}(\theta, \xi_{i0}, \lambda_{t0}), \theta))$, and let, for all i, t, θ: $\zeta_{it}(\theta) = v_{it}(\overline{\alpha}(\theta, \xi_{i0}, \lambda_{t0}), \theta) - \rho_{it}(\underline{h}, \xi_0, \underline{w}, \lambda_0, \theta)$. Moreover, let $\rho^*(h, \xi, w, \lambda, \theta) = \mathbb{E}_{h_i=h, \xi_{i0}=\xi, w_t=w, \lambda_{t0}=\lambda}(v_{it}(\overline{\alpha}(\theta, \xi_{i0}, \lambda_{t0}), \theta))$. Using part (v) in Assumption S2, we have

$$\frac{1}{NT}\sum_{i,t}\left\|\rho_{it}(\underline{h}, \xi_0, \underline{w}, \lambda_0, \theta)\right\|^2 \leq \frac{2}{NT}\sum_{i,t}\left\|\rho^*(h_i, \xi_{i0}, w_t, \lambda_{t0}, \theta)\right\|^2 + O_p\left(\frac{1}{T} + \frac{1}{N}\right),$$

which can be shown to be $O_p(\frac{1}{T} + \frac{1}{N})$ using similar arguments to the proof of Lemma A1. We then have

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{NT}\sum_{i,t}\left\|\overline{\zeta}(\widehat{k}_i, \widehat{l}_t, \theta)\right\|^2\right] \\
&\quad = \mathbb{E}\left[\frac{1}{NT}\sum_{k,l}\left(\sum_{i,t}\mathbf{1}\{\widehat{k}_i = k, \widehat{l}_t = l\}\right)\mathbb{E}_{\underline{h}, \xi_0, \underline{w}, \lambda_0}\left(\left\|\overline{\zeta}(k, l, \theta)\right\|^2\right)\right] \\
&\quad \leq \mathbb{E}\left[\frac{1}{NT}\sum_{k,l}\frac{\sum_{i,i',t,s}\mathbf{1}\{\widehat{k}_i = \widehat{k}'_i = k, \widehat{l}_t = \widehat{l}_s = l\}\left\|\mathbb{E}_{\underline{h}, \xi_0, \underline{w}, \lambda_0}\left(\zeta_{it}(\theta)\zeta_{i's}(\theta)\right)\right\|}{\sum_{i,t}\mathbf{1}\{\widehat{k}_i = k, \widehat{l}_t = l\}}\right] = O_p\left(\frac{KL}{NT}\right),
\end{aligned}$$

where we have used part (v) in Assumption S2. We thus obtain (S20).

Similarly to the proof of Lemma A2, we then show

$$\frac{1}{NT}\sum_{i,t}\left\{v_{it}^{\theta}\left(\widehat{\alpha}(\widehat{k}_i, \widehat{l}_t) - \alpha_{it0}\right) + \mathbb{E}_{\xi_{i0}, \lambda_{t0}}\left(v_{it}^{\theta}\right)\left[\mathbb{E}_{\xi_{i0}, \lambda_{t0}}\left(v_{it}^{\alpha}\right)\right]^{-1}v_{it}\right\} = O_p(\delta), \tag{S21}$$

where we omit references to θ_0 and α_{it0}. The first key term is

$$A_3 = \frac{1}{NT}\sum_{i,t}\mathbb{E}_{\xi_{i0}, \lambda_{t0}}\left(v_{it}^{\theta}\right)\left[\mathbb{E}_{\xi_{i0}, \lambda_{t0}}\left(v_{it}^{\alpha}\right)\right]^{-1}\left(-v_{it}^{\alpha}\right)\left(\left(-v_{it}^{\alpha}\right)^{-1}v_{it} - \widetilde{v}(\widehat{k}_i, \widehat{l}_t)\right),$$

where $\widetilde{v}$ is defined analogously to the proof of Lemma A2. To show that $A_3 = O_p(\delta)$, we use similar arguments to the proof of Lemma A2, in combination with part (v) in Assumption S2 (as in the proof of (S20)).

Let $\pi'_{it} = v^\theta_{it}(v^\alpha_{it})^{-1} - \mathbb{E}_{\xi_{i0},\lambda_{t0}}(v^\theta_{it})[\mathbb{E}_{\xi_{i0},\lambda_{t0}}(v^\alpha_{it})]^{-1}$. The second key term is

$$B_3 = \frac{1}{NT}\sum_{i,t}\pi'_{it}v^\alpha_{it}\big(\widetilde{\alpha}(\widehat{k}_i,\widehat{l}_t)-\alpha_{it0}\big)$$
$$= \frac{1}{NT}\sum_{i,t}\pi'_{it}v^\alpha_{it}\big(\alpha^*(\widehat{k}_i,\widehat{l}_t)-\alpha_{it0}\big) + \frac{1}{NT}\sum_{i,t}\pi'_{it}v^\alpha_{it}\big(\widetilde{\alpha}(\widehat{k}_i,\widehat{l}_t)-\alpha^*(\widehat{k}_i,\widehat{l}_t)\big),$$

where $\widetilde{\alpha}(k,l)$ and $\alpha^*(k,l)$ are defined analogously to the proof of Lemma A2. To show that $B_3 = O_p(\delta)$, we use similar arguments to the proof of Lemma A2, again in combination with part (v) in Assumption S2.

The final step, as in the proof of Lemma A3, is to show that

$$\frac{1}{NT}\sum_{i,t}\left\|\frac{\partial\widehat{\alpha}(\widehat{k}_i,\widehat{l}_t,\theta_0)}{\partial\theta'} - \frac{\partial\overline{\alpha}(\theta_0,\xi_{i0},\lambda_{t0})}{\partial\theta'}\right\|^2 = o_p(1). \tag{S22}$$

The proof of (S22) follows similar arguments to the proof of Lemma A3. *Q.E.D.*

S2.3. *GFE Based on Conditional Moments*

ASSUMPTION S3—Heterogeneity, Conditional Case: *There exist vectors ξ_{i0} of fixed dimension d, and ν_{i0} of dimension d_ν, and functions α and μ Lipschitz-continuous in ξ, such that $\alpha_{i0} = \alpha(\xi_{i0})$ and $\mu_{i0} = \mu(\xi_{i0},\nu_{i0})$.*

Differently from Assumption 1, here μ_{i0} depends on an additional heterogeneity component ν_{i0}, and by Assumption 2 the moment h_i is only injective for ξ_{i0}.

ASSUMPTION S4—regularity, conditional case:

(i) *$(Y'_i, X'_i, \xi'_{i0}, \nu'_{i0}, h'_i)'$ are i.i.d.; $(Y'_{it}, X'_{it})'$ are stationary for all i; $\ell_{it}(\alpha,\theta)$ is three times differentiable in both its arguments for all i, t; and Θ is compact, the supports of ξ_{i0} and α_{i0} are compact, and θ_0 belongs to the interior of Θ.*

(ii) *N, T tend jointly to infinity*; $\sup_{\xi,\nu,\alpha,\theta}|\mathbb{E}_{\xi_{i0}=\xi,\nu_{i0}=\nu}(\ell_{it}(\alpha,\theta))| = O(1)$, *and similarly for the first three derivatives of ℓ_{it}*; $\inf_{\xi,\nu,\alpha,\theta}\mathbb{E}_{\xi_{i0}=\xi,\nu_{i0}=\nu}(-\frac{\partial^2\ell_{it}(\alpha,\theta)}{\partial\alpha\partial\alpha'})$ *is positive definite*; *and* $\max_i\sup_{\alpha,\theta}|\ell_i(\alpha,\theta) - \mathbb{E}_{\xi_{i0},\nu_{i0}}(\ell_i(\alpha,\theta))| = o_p(1)$, *and similarly for the first three derivatives of ℓ_i.*

(iii) $\inf_{\xi,\nu,\theta}\mathbb{E}_{\xi_{i0}=\xi,\nu_{i0}=\nu}(-\frac{\partial^2\ell_{it}(\overline{\alpha}(\theta,\xi),\theta)}{\partial\alpha\partial\alpha'}) > 0$; $\mathbb{E}[\frac{1}{T}\sum_{t=1}^T \ell_{it}(\overline{\alpha}(\theta,\xi_{i0}),\theta)]$ *has a unique maximum at θ_0 on Θ, and its matrix of second derivatives is $-H^{\text{cond}} < 0$; and* $\sup_\theta \frac{1}{NT}\sum_{i,t}\|\frac{\partial^2\ell_{it}(\overline{\alpha}(\theta,\xi_{i0}),\theta)}{\partial\theta\partial\alpha'}\|^2 = O_p(1)$.

(iv) $\sup_{\widetilde{\xi},\alpha}\|\frac{\partial}{\partial\xi'}|_{\xi=\widetilde{\xi}}\mathbb{E}_{\xi_{i0}=\xi}(\text{vec}\,\frac{\partial^2\ell_{it}(\alpha,\theta_0)}{\partial\theta\partial\alpha'})\|$; $\sup_{\widetilde{\xi},\alpha}\|\frac{\partial}{\partial\xi'}|_{\xi=\widetilde{\xi}}\mathbb{E}_{\xi_{i0}=\xi}(\text{vec}\,\frac{\partial^2\ell_{it}(\alpha,\theta_0)}{\partial\alpha\partial\alpha'})\|$; *and* $\sup_{\widetilde{\xi},\theta}\|\frac{\partial}{\partial\xi'}|_{\xi=\widetilde{\xi}}\mathbb{E}_{\xi_{i0}=\xi}(\frac{\partial\ell_{it}(\overline{\alpha}(\theta,\widetilde{\xi}),\theta)}{\partial\alpha})\|$ *are $O(1)$.*

(v) $\mathbb{E}_{h_i=h,\xi_{i0}=\xi}(\frac{\partial\ell_{it}(\overline{\alpha}(\theta,\xi),\theta)}{\partial\alpha})$ *is twice differentiable with respect to h and ξ, with first and second derivatives that are uniformly bounded in ξ, h, and θ; and* $\|\text{Var}_{h_i=h,\xi_{i0}=\xi}(\frac{\partial\ell_i(\overline{\alpha}(\theta,\xi),\theta)}{\partial\alpha})\| = O(1)$, *uniformly in ξ, h, and θ.*

COROLLARY S3: *Let Assumptions* 2, S3, *and* S4 *hold*. *Let K be given by* (6), *with* $\gamma = O(1)$. *Then*, *as* N, T *tend to infinity such that* $T^{1+\frac{d}{2}} = O(N)$ *we have*

$$\widehat{\theta} = \theta_0 + O_p\left(\frac{1}{T}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right). \tag{S23}$$

PROOF: Let $\delta = \frac{1}{T} + \frac{K}{N} + K^{-\frac{2}{d}}$.[4] To show consistency, the key step is to show

$$\frac{1}{N}\sum_i \left\| \overline{v}(\widehat{k}_i, \theta)\right\|^2 = O_p(\delta), \quad \forall \theta \in \Theta. \tag{S24}$$

Let, for all θ, h, ξ: $\rho(h, \xi, \theta) = \mathbb{E}_{h_i=h,\xi_{i0}=\xi}(v_i(\overline{\alpha}(\theta, \xi), \theta))$, and let, for all i, θ: $\zeta_i(\theta) = v_i(\overline{\alpha}(\theta, \xi_{i0}), \theta) - \rho(h_i, \xi_{i0}, \theta)$. One can show, using similar techniques to the proof of Lemma A1, that $\frac{1}{N}\sum_i \|\overline{\zeta}(\widehat{k}_i, \theta)\|^2 = O_p(\frac{K}{N})$, and that this implies (S24).[5]

We then show: $\frac{1}{N}\sum_i \frac{\partial \ell_i(\widehat{\alpha}(\widehat{k}_i, \theta_0), \theta_0)}{\partial \theta} = O_p(\delta)$, which will follow from

$$\frac{1}{N}\sum_i v_i^\theta\left(\widehat{\alpha}(\widehat{k}_i) - \alpha_{i0}\right) = O_p(\delta), \tag{S25}$$

where from now on we omit references to θ_0 and α_{i0}. We have

$$\frac{1}{N}\sum_i v_i^\theta\left(\widehat{\alpha}(\widehat{k}_i) - \alpha_{i0}\right) = \frac{1}{N}\sum_i v_i^\theta\left(\widetilde{\alpha}(\widehat{k}_i) - \alpha_{i0} + \widetilde{v}(\widehat{k}_i)\right) + O_p(\delta),$$

where $\widetilde{\alpha}(k)$ and $\widetilde{v}(k)$ are as in the proof of Lemma A2; that is, denoting $w_i = (-v_i^\alpha)$, we have $\widetilde{\alpha}(k) = \overline{w}(k)^{-1}\overline{w\alpha}(k)$ and $\widetilde{v}(k) = \overline{w}(k)^{-1}\overline{v}(k)$.

Let $\gamma_v(h_i) = \mathbb{E}_{h_i}(v_i)$, $\zeta_i^v = v_i - \gamma_v(h_i)$, $\gamma_w(h_i) = \mathbb{E}_{h_i}(w_i)$, $\zeta_i^w = w_i - \gamma_w(h_i)$, $\gamma_{v^\theta}(h_i) = \mathbb{E}_{h_i}(v_i^\theta)$, and $\zeta_i^{v^\theta} = v_i^\theta - \gamma_{v^\theta}(h_i)$. First, we have

$$\begin{aligned}
\frac{1}{N}\sum_i v_i^\theta \widetilde{v}(\widehat{k}_i) &= \frac{1}{N}\sum_i v_i^\theta \overline{w}(\widehat{k}_i)^{-1}\overline{v}(\widehat{k}_i) = \frac{1}{N}\sum_i \overline{v}^\theta(\widehat{k}_i)\overline{w}(\widehat{k}_i)^{-1} v_i \\
&= \frac{1}{N}\sum_i \left(\overline{\gamma}_{v^\theta}(\widehat{k}_i) + \overline{\zeta}^{v^\theta}(\widehat{k}_i)\right)\left(\overline{\gamma}_w(\widehat{k}_i) + \overline{\zeta}^w(\widehat{k}_i)\right)^{-1} v_i \\
&= \frac{1}{N}\sum_i \overline{\gamma}_{v^\theta}(\widehat{k}_i)\overline{\gamma}_w(\widehat{k}_i)^{-1} v_i + O_p(\delta),
\end{aligned}$$

where, for example, $\overline{\gamma}_w(k)$ is the mean of $\gamma_w(h_i)$ in group $\widehat{k}_i = k$, and we have used that $\frac{1}{N}\sum_i \|\overline{\zeta}^{v^\theta}(\widehat{k}_i)\|^2 = O_p(K/N)$, $\frac{1}{N}\sum_i \|\overline{\zeta}^w(\widehat{k}_i)\|^2 = O_p(K/N)$, and $\frac{1}{N}\sum_i \|v_i\|^2 = O_p(1/T)$. Moreover,

$$\frac{1}{N}\sum_i \overline{\gamma}_{v^\theta}(\widehat{k}_i)\overline{\gamma}_w(\widehat{k}_i)^{-1} v_i = \frac{1}{N}\sum_i \overline{\gamma}_{v^\theta}(\widehat{k}_i)\overline{\gamma}_w(\widehat{k}_i)^{-1}\gamma_v(h_i) + O_p(\delta),$$

---

[4] Note that if $K = \widehat{K}$ is given by (6) with $\gamma = O(1)$, then $K = O_p(T^{\frac{d}{2}})$ and $\delta = O_p(\frac{1}{T} + \frac{T^{\frac{d}{2}}}{N})$, so if $T^{1+\frac{d}{2}} = O(N)$ then $\delta = O_p(\frac{1}{T})$.

[5] Note that, in the case of Theorem 1 (i.e., in the absence of additional heterogeneity ν_{i0}), the left-hand side in (S24) is $O_p(\frac{1}{T})$.

where we have used that $\frac{1}{N}\sum_i \|\overline{\zeta}^v(\widehat{k}_i)\|^2 = O_p(K/(NT))$. Lastly, we have

$$\frac{1}{N}\sum_i \overline{\gamma}_{v^\theta}(\widehat{k}_i)\overline{\gamma}_w(\widehat{k}_i)^{-1}\gamma_v(h_i) = \frac{1}{N}\sum_i \gamma_{v^\theta}(h_i)\gamma_w(h_i)^{-1}\gamma_v(h_i) + \frac{1}{N}\sum_i \big[\overline{\gamma}_{v^\theta}(\widehat{k}_i)\overline{\gamma}_w(\widehat{k}_i)^{-1} - \gamma_{v^\theta}(h_i)\gamma_w(h_i)^{-1}\big]\gamma_v(h_i),$$

where the first term is $O_p(\delta)$ since it is a mean of i.i.d. terms with mean $O(1/T)$ and variance $O(1/T)$, and the second term is $O_p(\delta)$ since $\frac{1}{N}\sum_i \|h_i - \overline{h}(\widehat{k}_i)\|^2 = O_p(\delta)$ and the γ functions are Lipschitz-continuous.

Second, let $v_i^\theta w_i^{-1} = \eta(h_i, \xi_{i0}) + e_i$, where $\mathbb{E}_{h_i=h,\xi_{i0}=\xi}(e_i w_i) = 0$. We have

$$\frac{1}{N}\sum_i v_i^\theta\big(\widetilde{\alpha}(\widehat{k}_i) - \alpha_{i0}\big) = \frac{1}{N}\sum_i \eta(h_i, \xi_{i0}) w_i\big(\widetilde{\alpha}(\widehat{k}_i) - \alpha_{i0}\big) + \frac{1}{N}\sum_i e_i w_i\big(\widetilde{\alpha}(\widehat{k}_i) - \alpha_{i0}\big),$$

where the first term is $O_p(\delta)$ since $\frac{1}{N}\sum_i \|h_i - \overline{h}(\widehat{k}_i)\|^2 = O_p(\delta)$, $\frac{1}{N}\sum_i \|\xi_{i0} - \overline{\xi}(\widehat{k}_i)\|^2 = O_p(\delta)$, $\frac{1}{N}\sum_i \|\widetilde{\alpha}(\widehat{k}_i) - \alpha_{i0}\|^2 = O_p(\delta)$, η is Lipschitz-continuous, and w_i is uniformly bounded (as in the proof of Lemma A2); and the second term is

$$\begin{aligned}\frac{1}{N}\sum_i e_i w_i\big(\widetilde{\alpha}(\widehat{k}_i) - \alpha_{i0}\big) &= \frac{1}{N}\sum_i e_i w_i\big(\widetilde{\alpha}(\widehat{k}_i) - \overline{\alpha}(\widehat{k}_i)\big) + \frac{1}{N}\sum_i e_i w_i\big(\overline{\alpha}(\widehat{k}_i) - \alpha_{i0}\big) \\ &= \frac{1}{N}\sum_i \overline{ew}(\widehat{k}_i)\big(\widetilde{\alpha}(\widehat{k}_i) - \overline{\alpha}(\widehat{k}_i)\big) + O_p(\delta) = O_p(\delta),\end{aligned}$$

where we have used that the $(e_i w_i)$'s have zero mean given $h_1, \ldots, h_N, \xi_{10}, \ldots, \xi_{N0}$ with bounded conditional variance, and $\frac{1}{N}\sum_i \|\overline{ew}(\widehat{k}_i)\|^2 = O_p(K/N) = O_p(\delta)$.

Finally, to show $\frac{1}{N}\sum_i \frac{\partial^2}{\partial\theta\partial\theta'}\big|_{\theta_0}(\ell_i(\widehat{\alpha}(\widehat{k}_i, \theta), \theta) - \ell_i(\overline{\alpha}(\theta, \xi_{i0}), \theta)) = o_p(1)$, we use similar arguments to the proof of Lemma A3.[6] *Q.E.D.*

Example: A Linear Homoskedastic Model. Consider the model $Y_{it} = X_{it}\theta_0 + \alpha_{i0} + U_{it}$, where X_{it} are scalar and U_{it} are i.i.d. with mean zero and variance σ^2 given $X_{i1}, \ldots, X_{iT}$, α_{i0}. Let $\widehat{\theta}$ be the GFE estimator based on a moment $h_i = \varphi(\alpha_{i0}) + \varepsilon_i$ that satisfies Assumptions 1 and 2 for $\xi_{i0} = \alpha_{i0}$; that is, h_i is only informative about α_{i0}, but not about the heterogeneity in X_{it}. Let $\zeta_i^X = \overline{X}_i - \mathbb{E}_{h_i}(\overline{X}_i)$, $\zeta_i^\alpha = \alpha_{i0} - \mathbb{E}_{h_i}(\alpha_{i0})$, and $\zeta_i^U = \overline{U}_i - \mathbb{E}_{h_i}(\overline{U}_i)$. We assume that K is large enough for the approximation error to be of smaller order, and that K/N tends to zero, as in Corollary 2. Under appropriate conditions in the regression

[6] Although the arguments are as in the proof of Lemma A3, the target log-likelihood is different since here $\overline{\alpha}(\theta, \xi_{i0})$ only depends on ξ_{i0}, not on $(\xi_{i0}', \nu_{i0}')'$. In particular, the matrix H^{cond} in Assumption S4 differs from the matrix H in Assumption 3; see (S26) for an example.

model, using similar arguments to the proof of Corollary S3 (though with no need for any restriction on the relative rates of N and T), one can show that $\widehat{\theta}$ admits the following expansion:

$$\widehat{\theta} = \theta_0 + \frac{\frac{1}{N}\sum_i \zeta_i^X(\zeta_i^\alpha + \zeta_i^U) + \frac{1}{NT}\sum_{i,t}(X_{it} - \overline{X}_i)(U_{it} - \overline{U}_i)}{\mathbb{E}\left[(X_{it} - \overline{X}_i)^2\right] + \operatorname{Var}(\zeta_i^X)} + o_p\left(\frac{1}{T}\right) + o_p\left(\frac{1}{\sqrt{NT}}\right). \tag{S26}$$

Notice two differences between (S26) and the expansion of the FE estimator: the presence of $\operatorname{Var}(\zeta_i^X)$ in the denominator, and the presence of $\frac{1}{N}\sum_i \zeta_i^X(\zeta_i^\alpha + \zeta_i^U)$ in the numerator. In addition, notice that (S26) simplifies to the expression in Corollary 2 in the absence of additional heterogeneity ν_{i0}.

S3. SIMULATIONS

Model of Wages and Participation (See (2)). We model the initial condition as: $Y_{i0} = \mathbf{1}\{u(\alpha_{i0}) \geq c(1; \theta_0) + U_{i0}\}$, with U_{i0} standard normal, independent of α_{i0}. We set $c(0; \theta_0) = 0$ and $c(1; \theta_0) = -1$. We set α_{i0} and V_{it} to be independent standard normals. In the simulations based on models (2) and (3), we weight the moments by the share of between-i variance to total variance.[7] To compute the variance $\widehat{V}_h$ to set the number of groups in this dynamic model, we use a Newey–West expression with one lag. Lastly, for kmeans computation we use Lloyd's algorithm with 100 random starting values. Table SI shows additional simulation results for this model.

Probit Model With Time-Varying Heterogeneity (See (3)). The U_{it}'s are standard normal independent of the X_{it}'s and the α_{it0}'s. The data generating process (DGP) for the scalar covariate is $X_{it} = \mu_{it0} + V_{it}$, where V_{it} are i.i.d. standard normal independent of the U_{it}'s, α_{it0}'s, and μ_{it0}'s, and $\mu_{it0} = \alpha_{it0}$. We set $\theta_0 = 1$, and set ξ_{i0} and λ_{t0} to be i.i.d. Gamma(1, 1) draws, independent of each other. Table SII shows additional simulation results for this model, including for the two-way GFE estimator based on both the cross-sectional moments $(\frac{1}{N}\sum_i Y_{it}, \frac{1}{N}\sum_i X_{it})'$, and the individual-specific moments $(\overline{Y}_i, \overline{X}_i)'$.

Conditional Moments: An Example. Consider the following probit model: $Y_{it} = \mathbf{1}\{X_{it}'\theta_0 + \alpha_{i0} + U_{it} \geq 0\}$, where the U_{it} are i.i.d. standard normal independent of the X_{it}'s and α_{i0}, and θ_0 is a vector of ones. The DGP for the kth covariate is $X_{itk} = \mathbf{1}\{\mu_{i0k} + V_{itk} \geq 0\}$, where V_{itk} are i.i.d. standard normal independent of the U_{it}'s, α_{i0}, and the μ_{i0k}'s, and α_{i0} and the μ_{i0k}'s follow independent standard normals. We vary the number of covariates between 1 and 3, so the total dimension of heterogeneity varies between 2 and 4. In this model, we expect the bias of FE to be moderate given the time horizon we consider

[7] Specifically, we demean and rescale h_i so that all its components $h_{i\ell}$ have zero mean and unit variance, and multiply each component $h_{i\ell}$ by: $\max(\frac{\sum_i h_{i\ell}^2 - \frac{1}{T^2}\sum_{i,t}(h_{it\ell} - h_{i\ell})^2}{\sum_i h_{i\ell}^2}, 0)$. Using equal weights instead has small effects in these simulations, however, we observed that this particular weighting can improve performance when some moments are substantially less informative about the heterogeneity than others.

TABLE SI

MODEL (2) OF WAGES AND PARTICIPATION.

T	Bias	std	RMSE	se/std	Bias	std	RMSE	se/std
	GFE, $\eta = 1$				FE, $\eta = 1$			
5	−0.569	0.056	0.572	1.114	−0.834	0.061	0.836	1.121
10	−0.204	0.040	0.208	0.983	−0.415	0.041	0.417	1.008
15	−0.119	0.033	0.124	0.951	−0.280	0.032	0.282	1.030
20	−0.088	0.028	0.093	0.981	−0.211	0.027	0.212	1.035
25	−0.070	0.024	0.074	0.987	−0.169	0.024	0.171	1.034
30	−0.055	0.023	0.059	0.974	−0.140	0.022	0.142	1.026
50	−0.032	0.017	0.037	0.990	−0.085	0.017	0.087	1.016
	GFE, $\eta = 2$				FE, $\eta = 2$			
5	−0.519	0.060	0.522	1.111	−0.878	0.064	0.880	1.138
10	−0.164	0.043	0.170	0.983	−0.443	0.042	0.445	1.046
15	−0.077	0.034	0.084	0.980	−0.298	0.033	0.300	1.059
20	−0.049	0.032	0.059	0.895	−0.225	0.029	0.227	1.008
25	−0.040	0.028	0.049	0.921	−0.181	0.025	0.183	1.027
30	−0.031	0.026	0.040	0.903	−0.152	0.024	0.154	0.990
50	−0.014	0.018	0.023	0.985	−0.091	0.017	0.092	1.048

Note: 1000 simulations, $N = 1000$. "RMSE" is root mean squared error, "se" is the average of standard error estimates across simulations, "std" is the standard deviation of the estimator across simulations. η is the risk aversion parameter.

($T = 20$), since α_{i0} is scalar and FE is a conditional approach. The question we ask here is how much the use of conditional moments can help reduce the bias of GFE due to the presence of additional heterogeneity in the covariates and the increased dimensionality of heterogeneity (see Section 4.2).

Consider first using $h_i = (\overline{Y}_i, \overline{X}_i')'$ as moments. In Table SIII we show the biases, standard deviations, and root mean squared errors of FE and GFE among 1000 simulations, for N = 1000 and T = 20. In the top panel we report GFE estimates as a function of the number of groups K. We see that, while the bias of GFE remains moderate with one covariate, the bias increases substantially with the dimension of heterogeneity, in agreement with our theory. By comparison, the bias of FE in the bottom panel is indeed quite small, and it only increases moderately with the number of covariates.

The situation is rather different when using *conditional moments* in GFE. In the middle panel in Table SIII we show simulation results for GFE based on covariates-specific conditional means $\overline{Y}_i(x) = \sum_{t=1}^T \mathbf{1}\{X_{it} = x\} Y_{it} / \sum_{t=1}^T \mathbf{1}\{X_{it} = x\}$. Importantly, in large samples these moments are only informative about α_{i0}, not μ_{i0}. We see that the bias of GFE with conditional moments increases only moderately with the number of covariates, and that FE and GFE with conditional moments have comparable—and quite small—biases.

Regarding implementation, note that, for a given i, all moments $\overline{Y}_i(x)$ may not be available since i's covariates may never take the value x in the sample. In Table SIII, whenever $\overline{Y}_i(x)$ is not available, we set the moment to an imputed value, the overall conditional mean $\overline{Y}(x) = \sum_{i,t} \mathbf{1}\{X_{it} = x\} Y_{it} / \sum_{i,t} \mathbf{1}\{X_{it} = x\}$. The imputation does not affect the theory, provided the event that any of the $\overline{Y}_i(x)$'s is not available tends to zero with probabil-

TABLE SII

PROBIT MODEL (3) WITH TIME-VARYING HETEROGENEITY.

T	Bias	std	RMSE	se/std	Bias	std	RMSE	se/std	Bias	std	RMSE	se/std	Bias	std	RMSE	se/std
	2-way GFE, $\sigma = -10$				GFE, $\sigma = -10$				FE, $\sigma = -10$				IFE, $\sigma = -10$			
5	−0.045	0.035	0.057	0.917	−0.045	0.035	0.057	0.917	0.443	0.071	0.448	0.705	0.116	0.067	0.134	0.456
10	−0.015	0.022	0.027	1.015	−0.014	0.022	0.026	1.015	0.198	0.034	0.201	0.817	0.101	0.035	0.107	0.506
15	−0.006	0.019	0.020	0.936	−0.004	0.019	0.020	0.937	0.130	0.025	0.132	0.842	0.100	0.026	0.103	0.530
20	−0.003	0.016	0.017	0.961	−0.000	0.016	0.016	0.961	0.098	0.020	0.100	0.862	0.087	0.021	0.089	0.568
25	−0.001	0.014	0.014	1.022	0.002	0.014	0.014	1.019	0.080	0.016	0.082	0.939	0.072	0.016	0.074	0.640
30	−0.001	0.013	0.013	1.009	0.003	0.013	0.013	1.011	0.069	0.015	0.070	0.931	0.059	0.015	0.061	0.646
50	0.002	0.010	0.010	1.009	0.006	0.010	0.012	1.011	0.048	0.011	0.049	0.949	0.036	0.011	0.038	0.670
	2-way GFE, $\sigma = 0$				GFE, $\sigma = 0$				FE, $\sigma = 0$				IFE, $\sigma = 0$			
5	−0.045	0.040	0.061	0.907	−0.045	0.040	0.060	0.906	0.486	0.089	0.494	0.675	0.150	0.088	0.173	0.402
10	−0.022	0.026	0.034	0.976	−0.020	0.026	0.033	0.976	0.228	0.043	0.232	0.746	0.119	0.040	0.125	0.505
15	−0.013	0.021	0.024	0.991	−0.010	0.021	0.023	0.988	0.147	0.030	0.150	0.803	0.117	0.030	0.121	0.522
20	−0.008	0.018	0.019	0.993	−0.004	0.018	0.018	0.992	0.110	0.023	0.112	0.869	0.101	0.024	0.104	0.567
25	−0.006	0.016	0.017	0.988	−0.002	0.016	0.016	0.982	0.086	0.020	0.089	0.868	0.082	0.020	0.085	0.583
30	−0.004	0.014	0.015	1.003	0.000	0.014	0.014	1.000	0.073	0.017	0.074	0.927	0.069	0.017	0.071	0.639
50	−0.001	0.011	0.011	0.989	0.005	0.011	0.012	0.983	0.046	0.012	0.047	0.934	0.040	0.012	0.042	0.635
	2-way GFE, $\sigma = 1$				GFE, $\sigma = 1$				FE, $\sigma = 1$				IFE, $\sigma = 1$			
5	−0.050	0.056	0.075	0.752	−0.049	0.056	0.074	0.752	0.565	0.149	0.585	0.497	0.212	0.126	0.246	0.339
10	−0.030	0.028	0.041	0.994	−0.028	0.028	0.040	0.993	0.253	0.060	0.260	0.623	0.144	0.044	0.150	0.515
15	−0.019	0.022	0.029	1.046	−0.015	0.022	0.027	1.045	0.159	0.036	0.163	0.754	0.147	0.031	0.150	0.552
20	−0.013	0.019	0.023	1.016	−0.009	0.019	0.021	1.016	0.116	0.027	0.119	0.841	0.128	0.028	0.131	0.526
25	−0.009	0.018	0.020	0.978	−0.005	0.018	0.019	0.977	0.091	0.023	0.094	0.845	0.106	0.025	0.109	0.514
30	−0.007	0.016	0.017	0.999	−0.002	0.016	0.016	0.998	0.074	0.019	0.076	0.897	0.086	0.021	0.088	0.563
50	−0.003	0.012	0.012	1.022	0.004	0.012	0.013	1.016	0.045	0.013	0.047	0.967	0.053	0.014	0.054	0.638
	2-way GFE, $\sigma = 10$				GFE, $\sigma = 10$				FE, $\sigma = 10$				IFE, $\sigma = 10$			
5	−0.016	0.075	0.076	0.692	−0.015	0.075	0.076	0.692	0.705	0.261	0.752	0.387	0.310	0.261	0.405	0.214
10	−0.014	0.036	0.039	0.903	−0.011	0.037	0.038	0.903	0.321	0.097	0.335	0.481	0.184	0.057	0.192	0.466
15	−0.006	0.027	0.028	0.965	−0.002	0.028	0.028	0.964	0.202	0.053	0.209	0.629	0.184	0.041	0.189	0.498
20	−0.002	0.023	0.023	0.979	0.003	0.024	0.024	0.975	0.148	0.037	0.153	0.725	0.169	0.036	0.173	0.472
25	−0.000	0.020	0.020	0.999	0.006	0.021	0.021	0.998	0.119	0.030	0.122	0.765	0.146	0.035	0.150	0.431
30	0.003	0.019	0.019	0.988	0.009	0.019	0.021	0.992	0.100	0.025	0.103	0.829	0.123	0.029	0.126	0.480
50	0.003	0.014	0.014	1.012	0.010	0.014	0.018	1.010	0.062	0.017	0.064	0.913	0.076	0.017	0.078	0.591

Note: See notes to Table SI. IFE is interacted fixed-effects with one factor. σ is the substitution parameter.

TABLE SIII

PROBIT MODEL WITH BINARY COVARIATES.

K	Bias	std	RMSE	Bias	std	RMSE	Bias	std	RMSE
	GFE, 1 covariate			GFE, 2 covariates			GFE, 3 covariates		
5	−0.188	0.029	0.191	−0.293	0.031	0.294	−0.362	0.042	0.365
10	−0.083	0.027	0.088	−0.205	0.032	0.207	−0.275	0.035	0.277
20	−0.017	0.028	0.033	−0.118	0.030	0.122	−0.206	0.033	0.209
30	0.006	0.029	0.030	−0.081	0.030	0.086	−0.165	0.032	0.168
40	0.018	0.029	0.035	−0.056	0.030	0.064	−0.135	0.033	0.139
50	0.026	0.030	0.039	−0.040	0.030	0.051	−0.115	0.033	0.120
	Cond. GFE, 1 covariate			Cond. GFE, 2 covariates			Cond. GFE, 3 covariates		
5	−0.060	0.035	0.069	−0.085	0.037	0.092	−0.110	0.039	0.117
10	−0.045	0.033	0.056	−0.074	0.042	0.085	−0.099	0.044	0.109
20	−0.015	0.034	0.038	−0.046	0.037	0.059	−0.074	0.045	0.087
30	0.008	0.035	0.036	−0.031	0.036	0.048	−0.061	0.043	0.075
40	0.025	0.036	0.043	−0.020	0.036	0.041	−0.050	0.041	0.065
50	0.033	0.035	0.048	−0.012	0.036	0.038	−0.040	0.040	0.057
	FE, 1 covariate			FE, 2 covariates			FE, 3 covariates		
–	0.062	0.031	0.069	0.074	0.034	0.081	0.089	0.039	0.097

Note: 1000 simulations, $N = 1000$, $T = 20$. In the top panel we show GFE estimates based on unconditional moments for different K values, in the middle panel we show GFE estimates based on conditional moments for different K values, in the bottom row we show FE estimates.

ity approaching one in large samples.[8] Moreover, we have obtained similar results using an alternative conditional first step implementation that does not rely on imputations.[9]

Co-editor Guido Imbens handled this manuscript.

Manuscript received 10 April, 2017; final version accepted 4 June, 2021; available online 29 June, 2021.

[8]To provide intuition in a simple case, suppose that X_{it} are binary, i.i.d. over time given μ_{i0}, with $\Pr(X_{it} = 1 \mid \mu_{i0} = \mu) \in (\epsilon, 1-\epsilon)$ for all μ, for some $\epsilon > 0$. Then $\Pr(\exists i : X_{i1} = \cdots = X_{iT} = 0) \leq N(1-\epsilon)^T$, which tends to zero whenever $(\ln N)/T \to 0$.

[9]This implementation is as follows. Let $I_i(x)$ be the indicator that there exists a t such that $X_{it} = x$, and let $x_1, \ldots, x_M$ denote the points of support of X_{it}. In the first step, we use a Lloyd's-like algorithm to minimize the function $\sum_{i=1}^N \sum_{m=1}^M I_i(x_m)(\overline{Y}_i(x_m) - g(x_m, k_i))^2$, with respect to $k_1, \ldots, k_N$ and $g(x_1, 1), \ldots, g(x_M, K)$.