Econometric Theory, **21**, 2005, 455–469. Printed in the United States of America.
DOI: 10.1017/S0266466605050243

TIME-INVARIANT REGRESSOR IN NONLINEAR PANEL MODEL WITH FIXED EFFECTS

JINYONG HAHN AND JUERGEN MEINECKE
UCLA

This paper generalizes the intuition of Hausman and Taylor (1981, *Econometrica* 49, 1377–1398) and develops a method of dealing with a time-invariant regressor in nonlinear panel models with fixed effects. We illustrate the usefulness of our result by discussing the implication for some nonlinear models of social interactions.

1. INTRODUCTION

Panel data allow the possibility of controlling for unobserved individual specific effects. In linear models, such "fixed effects" are usually eliminated by differencing. An unintended consequence of differencing is that it also eliminates the time-invariant regressor, which renders its coefficient unidentified. Hausman and Taylor (1981) used an instrumental variables (IV) approach to overcome such a problem. They show that a variable that is uncorrelated with individual fixed effects can be used as a valid instrument in estimating such a coefficient.

In this paper, we generalize their intuition and develop a method of dealing with a time-invariant regressor in the nonlinear framework. This method requires a large number of observations per individual (T), so its applicability is limited to the case where T is large. Because the IV estimation requires a large number of individuals (n), we adopt an asymptotic framework where both n and T grow to infinity at the same rate. This result is made possible by recent technical progress of panel analysis under such alternative asymptotics. See, e.g., Hahn and Kuersteiner (2002, 2003), Hahn and Newey (2004), and Woutersen (2002). We illustrate the usefulness of our result by discussing the implication for some nonlinear models of social interactions.

2. IV ESTIMATION

Suppose that we are given a set of moment restrictions

$$E[\varphi(y_{it}, \gamma_{i0} + w_i'\delta_0 + x_{it}'\theta_0)] = 0, \qquad i = 1, \ldots, n; \qquad t = 1, \ldots, T \tag{1}$$

We are grateful to Dan Ackerberg and Jerry Hausman for helpful comments. The first author gratefully acknowledges financial support from NSF grant SES-0313651. Address correspondence to Jinyong Hahn, Department of Economics, UCLA, Box 951477, Los Angeles, CA 90095-1477, USA; e-mail: hahn@econ.ucla.edu.

© 2005 Cambridge University Press 0266-4666/05 $12.00

https://doi.org/10.1017/S0266466605050243 Published online by Cambridge University Press

for some vector-valued function φ, where y_{it}, x_{it}, and w_i denote the dependent variable in the tth period, time-varying regressor in the tth period, and time-invariant regressor. Unobserved individual specific effects are summarized by the scalar variable γ_i. For example, in the case of the linear model

$$y_{it} = \gamma_{i0} + w_i'\delta_0 + x_{it}'\theta_0 + \varepsilon_{it} \tag{2}$$

with (w_i', x_{it}') strictly exogenous such that

$$\varphi(y_{it}, \gamma_{i0} + w_i'\delta_0 + x_{it}'\theta_0) = \begin{pmatrix} w_i \\ x_{it} \\ 1 \end{pmatrix}(y_{it} - \gamma_{i0} - w_i'\delta_0 - x_{it}'\theta_0).$$

Our primary focus is to estimate the coefficient δ_0 of the time-invariant regressor w_i when γ_i is possibly correlated with w_i and x_{it}. We should note that estimation of θ_0 does not present any substantive conceptual challenge. If both n and T grow to infinity at the same rate, θ_0 can be $\sqrt{nT}$-consistently estimated. Letting $\alpha_{i0} \equiv \gamma_{i0} + w_i'\delta_0$, we can rewrite the model as $E[\varphi(y_{it}, \alpha_{i0} + x_{it}'\theta_0)] = 0$, to which we can apply variants of recently developed methods discussed in Hahn and Kuersteiner (2002, 2003), Hahn and Newey (2004), and Woutersen (2002).

To understand how δ_0 could be estimated, suppose for a moment that we observe α_{i0}. Also suppose that we observe an additional variable z_i with $\dim(z_i) = \dim(w_i)$[1] and such that the following condition applies.

CONDITION 1. *(i)* $E[z_i\gamma_{i0}] = 0$*; (ii)* $E[z_i w_i']$ *is nonsingular.*

Note that z_i is required to be uncorrelated with the individual fixed effects. It is clear that we can consistently estimate δ_0 by

$$\tilde{\delta} \equiv \left(\sum_{i=1}^{n} z_i w_i'\right)^{-1}\left(\sum_{i=1}^{n} z_i \alpha_{i0}\right) \tag{3}$$

under the mild condition that the data are independent and identically distributed (i.i.d.) over i.

CONDITION 2. $(\{y_{i1}, y_{i2}, \ldots\}, \{x_{i1}, x_{i2}, \ldots\}, z_i, w_i, \gamma_{i0})$ *is i.i.d. over i.*

Hausman and Taylor (1981) noted that $\tilde{\delta}$ would remain consistent even if we replace α_{i0} by an unbiased estimate. In the nonlinear context, it is difficult to come up with such an unbiased estimator for α_{i0}. Therefore, Hausman and Taylor's method cannot be directly applied. The basic intuition in this paper is that, when both n and T grow to infinity at the same rate, we can come up with a $\sqrt{T}$-consistent estimator for α_{i0}, say, $\hat{\alpha}_i$.

https://doi.org/10.1017/S0266466605050243 Published online by Cambridge University Press

CONDITION 3. *$n, T \to \infty$ such that $n/T \to \rho$, where $0 < \rho < \infty$.*

Because the estimation error becomes very small as the sample size increases, the IV estimator

$$\hat{\delta} \equiv \left(\sum_{i=1}^{n} z_i w_i' \right)^{-1} \left(\sum_{i=1}^{n} z_i \hat{\alpha}_i \right) \tag{4}$$

is expected to be consistent for δ_0 in general. This is quite intuitive. Note that $\hat{\delta}$ is an IV estimator of $\hat{\alpha}_i$ on w_i. Because $\hat{\alpha}_i$ is a proxy for α_{i0}, and because the "measurement error" disappears as $T \to \infty$, we should expect that the distribution of $\sqrt{n}(\hat{\delta} - \delta_0)$ should be similar to that of $\sqrt{n}(\tilde{\delta} - \delta_0)$ if T is large.

To come up with a $\sqrt{T}$-consistent estimator $\hat{\alpha}_i$, we will assume that

$$E[u(X_{it}; \theta_0, \alpha_{i0})] = 0, \qquad E[v(X_{it}; \theta_0, \alpha_{i0})] = 0$$

for some functions u and v, where $X_{it} \equiv (y_{i1}, \ldots, y_{iT}, x_{i1}, \ldots, x_{iT})$ and where $\dim(\theta) = \dim(u) = p$ and $\dim(\alpha) = \dim(v) = 1$. We will consider the estimator that solves

$$0 = \sum_{i=1}^{n} \sum_{t=1}^{T} u(X_{it}; \hat{\theta}, \hat{\alpha}_i), \qquad 0 = \sum_{t=1}^{T} v(X_{it}; \hat{\theta}, \hat{\alpha}_i). \tag{5}$$

This indicates that (i) the first component u is used throughout the sample for estimation of θ_0; and (ii) the second component v is used only for the ith individual to estimate α_{i0}. We do not expect this separation to be constraining in practice. For example, in the linear model (2), we may take $u(X_{it}; \theta, \alpha_i) = x_{it} \cdot (y_{it} - \alpha_i - x_{it}'\theta)$, $v(X_{it}; \theta, \alpha_i) = y_{it} - \alpha_i - x_{it}'\theta$, which will result in the usual fixed effects estimator.

Under regularity conditions discussed in Appendix A, it can be shown that the $\hat{\alpha}_i$, which solves (5), is uniformly consistent over i.[2] Because $\hat{\alpha}_i$ is a proxy for α_{i0}, there is a "measurement error." If there is a correlation between the measurement error and the instrument z_i, the resultant estimator $\hat{\delta}$ may be biased. Condition 4 rules this out.

CONDITION 4. $E[v(X_{it}; \theta, \alpha_i) | z_i] = 0$.[3]

Theorem 1 establishes that the IV estimator $\hat{\delta}$ in (4) based on the proxy $\hat{\alpha}_i$ of α_{i0} has the same asymptotic distribution as $\tilde{\delta}$ in (3).

THEOREM 1. *Assume Conditions 1–7. Further assume that $E[|z_i w_i'|] < \infty$ and $E[|\gamma_{i0}^2 z_i z_i'|] < \infty$. We then have*

$$\sqrt{n}(\hat{\delta} - \delta_0) \to N(0, (E[z_i w_i'])^{-1} E[\gamma_{i0}^2 z_i z_i'] (E[w_i z_i'])^{-1}).$$

Proof. See Appendix E.

https://doi.org/10.1017/S0266466605050243 Published online by Cambridge University Press

3. APPLICATION: NONLINEAR MODEL OF SOCIAL INTERACTIONS

Identification and estimation of various social effects in the nonlinear model of social interactions based on the preceding framework are straightforward and provide one way of dealing with the typical identification problems that are peculiar to these models.

For a grouped cross-section of data, a model of social interactions can have the form of a conditional likelihood $f(y_{gi}, \gamma_{g0} + E_g[y_{gi}]\beta + E_g[s'_{gi}]\phi + s'_{gi}\zeta + x'_{gi}\theta_0)$. Here, y_{gi} is the outcome/behavior of interest for the ith individual in the gth group, and $E_g[\cdot]$ denotes the mean for the gth group. Following the classification of Manski (1993), the coefficient on $E_g[y_{gi}]$ determines the strength of *endogenous* social effects in explaining individual outcomes. In addition to y_{gi}, we observe s_{gi}, a vector of individual characteristics that also generate *exogenous (contextual)* social effects, and x_{gi}, a vector of individual characteristics that operate at the individual level only. Finally γ_{g0}, which is not observed by the econometrician, captures the presence of *correlated* group effects.

The focus of Graham and Hahn (2003) is on the linear-in-means model without contextual effects. They exploit the idea of Hausman and Taylor (1981) of using IVs to identify the parameters. Identification of the endogenous effect is made possible by an instrument that exogenously explains the between-group variation of the individual characteristics x_{gi}. To be more specific, they consider the simplified model

$$y_{gi} = \gamma_{g0} + E_g[y_{gi}]\beta + x'_{gi}\theta_0 + \varepsilon_{gi}$$

and examine whether the endogenous effects can be identified in the presence of correlated effects. For such purpose, they considered the social equilibrium

$$E_g[y_{gi}] = E_g[x'_{gi}]\frac{\theta_0}{1-\beta} + \frac{\gamma_{g0}}{1-\beta}. \tag{6}$$

They show that $\theta_0 = \text{plim}[(1/N)\sum_{g=1}^{N}(\sum_{i=1}^{M^g}\tilde{x}_{gi}\tilde{x}'_{gi})]^{-1}[(1/N)\sum_{g=1}^{N}(\sum_{i=1}^{M^g}\tilde{x}_{gi}\tilde{y}_{gi})]$ under a standard strict exogeneity condition, where for any vector s_{gi}, $\tilde{s}_{gi} \equiv s_{gi} - \bar{s}_g$. They also note that $\theta_0/(1-\beta)$ is the limit of the two-stage least squares estimator applied to the social equilibrium (6) if (i) $E_g[y_{gi}]$ and $E_g[x_{gi}]$ are observed; and (ii) there exists an instrument z_g such that $E[z_g\gamma_{g0}] = 0$ and $E[z_g E_g[x'_{gi}]] \neq 0$.

Brock and Durlauf (2001a, 2001b) are concerned with nonlinear models *without* correlated group effects. To be specific, they considered a logit model[4] with

$$\Pr(y_{gi} = 1 | E_g[y_{gi}], h_g, x_{gi}) = \frac{\exp(k + E_g[y_{gi}]\beta + h'_g\xi + x'_{gi}\theta_0)}{1 + \exp(k + E_g[y_{gi}]\beta + h'_g\xi + x'_{gi}\theta_0)},$$

where $h_g = (E_g[s'_{gi}], s'_{gi})'$. They show that the parameter $(k, \beta, \xi, \theta_0)$ is identified and can be consistently estimated by maximum likelihood estimation. Exploiting nonlinearity, they established that the endogenous effects β can be

https://doi.org/10.1017/S0266466605050243 Published online by Cambridge University Press

identified from ξ. Because h_g contains $E_g[s_{gi}]$, the contextual effects are identified from the endogenous effects in nonlinear models in general.[5]

The result in the previous section can be used to identify social effects in the presence of correlated group effects. Let γ_{g0} denote the group characteristic that may be correlated with observed variables such as $E_g[y_{gi}]$ or x_{gi}. Assume that the conditional likelihood given as $\gamma_{g0}, E_g[y_{gi}], h_g, x_{gi}$ takes the form $f(y_{gi}, \gamma_{g0} + E_g[y_{gi}]\beta + h_g'\xi + x_{gi}'\theta_0)$. Note that the explanatory variables affect the outcome through the linear index $\gamma_{g0} + E_g[y_{gi}]\beta + h_g'\xi + x_{gi}'\theta_0$. For example, we may have

$$\Pr(y_{gi} = 1 | \gamma_{g0}, E_g[y_{gi}], h_g, x_{gi}) = \frac{\exp(\gamma_{g0} + E_g[y_{gi}]\beta + h_g'\xi + x_{gi}'\theta_0)}{1 + \exp(\gamma_{g0} + E_g[y_{gi}]\beta + h_g'\xi + x_{gi}'\theta_0)},$$

which is a generalized version of Brock and Durlauf's logit model that allows correlated group effects. Interpreting $E_g[y_{gi}]$ as just one of the regular time invariant regressors and writing $w_g = (E_g[y_{gi}], h_g')'$ and $w_g'\delta_0 = \beta E_g[y_{gi}] + h_g'\xi$, we get the conditional likelihood as $f(y_{gi}, \gamma_{g0} + w_g'\delta_0 + x_{gi}'\theta_0)$. This model can be understood to be a nonlinear panel model with group fixed effects and some individual-invariant regressor w_g, for which identification results have been established earlier in this paper. We note that consistent estimation of $\gamma_{g0} + w_g'\delta_0$ and θ_0 can be achieved by considering the maximum likelihood estimator, i.e., by taking

$$u(X_{gi}; \theta_0, \alpha_{i0}) = \partial \log f(y_{gi}, \alpha_g + x_{gi}'\theta_0)/\partial\theta_0,$$

$$v(X_{gi}; \theta_0, \alpha_{i0}) = \partial \log f(y_{gi}, \alpha_g + x_{gi}'\theta_0)/\partial\alpha_{i0}.$$

We may therefore conclude that the social effects are identified as such.

4. SUMMARY AND EXTENSION

In this paper, we generalized the result of Hausman and Taylor (1981) to nonlinear panel models with fixed effects. The usefulness of the result is illustrated with some nonlinear models of social interactions. It would be interesting to generalize Hausman and Taylor's specification test to the nonlinear setup, which is left for future research.

NOTES

1. It is easy to generalize the discussion to the overidentified case where $\dim(z_i) > \dim(w_i)$. Because the primary purpose of this paper is identification and consistent estimation of δ_0, we focus on the exactly identified case.

2. See Appendix D.

3. We have $v(X_{it}; \theta, \alpha_i) = \varepsilon_{it}$ in the linear model (2), and Hausman and Taylor's instrument z_i for such a linear model is required to satisfy $E[z_i \cdot \varepsilon_{it}] = 0$.

https://doi.org/10.1017/S0266466605050243 Published online by Cambridge University Press

4. They actually considered the model

$$\Pr(y_{gi}=1|m_g^e,x_{gi},h_g)=\frac{\exp[k+m_g^e\beta+h_g'\xi+x_{gi}'\theta_0]}{1+\exp[k+m_g^e\beta+h_g'\xi+x_{gi}'\theta_0]},$$

where m_g^e denotes the (common) expectation of $\bar{y}_g$ among agents in group g. Under some auxiliary assumption including rational expectations and common knowledge, the model is reduced to the simpler form presented here.

5. They note that these identification results are still valid in the presence of multiple equilibria.

REFERENCES

Brock, W.A. & S.N. Durlauf (2001a) Interactions-based models. In J. Heckman & E. Leamer (eds.), *Handbook of Econometrics*, vol. 5, pp. 3297–3380. North-Holland.

Brock, W.A. & S.N. Durlauf (2001b) Discrete choice with social interactions. *Review of Economic Studies* 68, 235–260.

Graham, B.S. & J. Hahn (2003) Identification and Estimation of the Linear-in-Means Model of Social Interactions. Manuscript, UCLA.

Hahn, J. & G. Kuersteiner (2002) Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large. *Econometrica* 70, 1639–1657.

Hahn, J. & G. Kuersteiner (2003) Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects. Manuscript, UCLA.

Hahn, J. & W.K. Newey (2004) Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica* 72, 1295–1319.

Hausman, J. & W. Taylor (1981) Panel data and unobservable individual effects. *Econometrica* 49, 1377–1398.

Lahiri, S. (1992) Edgeworth correction by moving block bootstrap for stationary and nonstationary data. In R. LePage & L. Billard (eds.), *Exploring the Limits of Bootstrap*, pp. 183–214. Wiley.

Manski, C. (1993) Identification of endogenous social effects: The reflection problem. *Review of Economic Studies* 60, 531–542.

Woutersen, T. (2002) Robustness against Incidental Parameters. Manuscript, University of Western Ontario.

APPENDIX A: REGULARITY CONDITIONS

Condition 5. (i) Given time-invariant variables (α_{i0},z_i,w_i), (y_{it},x_{it}) is i.i.d. over t; (ii) for every i, $G_{(i)}(\theta_0,\alpha_{i0})=0$; (iii) for each $\eta>0$, $\inf_i \inf_{\{(\theta,\alpha):|(\theta,\alpha)-(\theta_0,\alpha_{i0})|>\eta\}}|G_{(i)}(\theta,\alpha)|>0$, where $\hat{G}_{(i)}(\theta,\alpha_i)\equiv T^{-1}\sum_{t=1}^{T}g(X_{it};\theta,\alpha_i)$, $G_{(i)}(\theta,\alpha_i)\equiv E[g(X_{it};\theta,\alpha_i)]$, and $g(\cdot;\theta,\alpha)\equiv(u(\cdot;\theta,\alpha)',v(\cdot;\theta,\alpha)')'$.

Remark 1. We are assuming α_i are deterministic sequence of numbers, i.e., all the results in this paper are results conditional on α_i.

Condition 6. (i) The function $g(\cdot;\theta,\alpha)$ is continuous in $(\theta,\alpha)\in\Upsilon$; (ii) the parameter space Υ is compact; (iii) there exists a function $M(X_{it})$ such that

$$\left|\frac{\partial^{m_1+m_2}g(X_{it};\theta,\alpha_i)}{\partial\theta^{m_1}\partial\alpha_i^{m_2}}\right|\le M(X_{it})\qquad 0\le m_1+m_2\le 1,\ldots,6$$

and $\sup_i E[M(X_{it})^Q]<\infty$ for some $Q>64$.

Condition 7. (i) $\min_i E[v(X_{it};\theta_0,\alpha_{i0})^2]>0$; (ii) $|\lim_{n\to\infty}(1/n)\sum_{i=1}^{n}\mathcal{I}_i|>0$, where $U(X_{it};\theta,\alpha_i)\equiv u(X_{it};\theta,\alpha_i)-\rho_{i0}v(X_{it};\theta,\alpha_i)$, $\rho_{i0}\equiv(E[\partial v(X_{it};\theta,\alpha_i)/\partial\alpha_i])^{-1}E[\partial u(X_{it};\theta,\alpha_i)/\partial\alpha_i]$, and $\mathcal{I}_i\equiv -E[\partial U(X_{it};\theta,\alpha_i)/\partial\theta']$.

https://doi.org/10.1017/S0266466605050243 Published online by Cambridge University Press

APPENDIX B: CONSISTENCY

LEMMA 1. *Assume that W_t are i.i.d. with $E[W_t] = 0$ and $E[W_t^{2k}] < \infty$. Then, $E[(\sum_{t=1}^{T} W_t)^{2k}] = C(k)T^k + o(T^k)$ for some constant $C(k)$.*

Proof. By adopting an argument in the proof of Lemma 5.1 in Lahiri (1992), we have

$$E\left[\left(\sum_{t=1}^{T} W_t\right)^{2k}\right] = \sum_{j=1}^{2k} \sum_{\alpha} C(\alpha_1,\ldots,\alpha_j) \sum_{I} E\left[\prod_{s=1}^{j} W_{t_s}^{\alpha_s}\right], \tag{B.1}$$

where for each fixed $j \in \{1,\ldots,2k\}$, $\sum_\alpha$ extends over all j-tuples of positive integers $(\alpha_1,\ldots,\alpha_j)$ such that $\alpha_1 + \cdots + \alpha_j = 2k$ and $\sum_I$ extends over all ordered j-tuples $(t_1,\ldots,t_j)$ of integers such that $1 \le t_j \le T$. Also, $C(\alpha_1,\ldots,\alpha_j)$ stands for a bounded constant. Note that if $j > k$ then at least one of the indices $\alpha_j = 1$. By independence and the fact that $E[W_t] = 0$ it follows that $E[\prod_{s=1}^{j} W_{t_s}^{\alpha_s}] = 0$ whenever $j > k$. This shows that $|E[(\sum_{t=1}^{T} W_t)^{2k}]| \le C(k)T^k \cdot E[W_t^{2k}]$ for some constant $C(k)$. ■

LEMMA 2. *Suppose that, for each i, $\{\xi_{it}, t = 1,2,\ldots\}$ is a sequence of zero mean i.i.d. random variables. We assume that $\{\xi_{it}, t = 1,2,3\}$ are independent across i. We also assume that $\max_i E[|\xi_{it}|^{16}] < \infty$. Finally, we assume that $n = O(T)$. We then have $\max_i \Pr[|(1/T)\sum_{t=1}^{T} \xi_{it}| > \eta] = o(T^{-2})$ for every $\eta > 0$.*

Proof. Using Lemma 1, we obtain $E[|\sum_{t=1}^{T} \xi_{it}|^{16}] \le CT^8 \cdot E[\xi_{it}^{16}]$, where $C > 0$ is a constant. Therefore, we have $T^2 \Pr[|(1/T)\sum_{t=1}^{T} \xi_{it}| > \eta] \le T^2(CT^8/T^{16}\eta^{16})E[\xi_{it}^{16}]$, or $\max_i T^2 \Pr[|(1/T)\sum_{t=1}^{T} \xi_{it}| > \eta] \le (C/T^6\eta^{16})\max_i E[\xi_{it}^{16}] = o(1)$. ■

LEMMA 3. *Suppose that Conditions 3 and 6 hold. We then have for all $\eta > 0$ that*

$$\Pr\left[\max_{1\le i\le n} \sup_{(\theta,\alpha)} |\hat{G}_{(i)}(\theta,\alpha) - G_{(i)}(\theta,\alpha)| \ge \eta\right] = o(T^{-1}).$$

Proof. Let $\eta > 0$ be given. We note that

$$\Pr\left[\max_{1\le i\le n} \sup_{(\theta,\alpha)} |\hat{G}_{(i)}(\theta,\alpha) - G_{(i)}(\theta,\alpha)| \ge \eta\right]$$

$$\le \sum_{i=1}^{n} \Pr\left[\sup_{(\theta,\alpha)} |\hat{G}_{(i)}(\theta,\alpha) - G_{(i)}(\theta,\alpha)| \ge \eta\right]. \tag{B.2}$$

Let $\varepsilon > 0$ be chosen such that $2\varepsilon \max_i E[M(X_{it})] < \eta/3$. Divide Υ into subsets $\Upsilon_1, \Upsilon_2, \ldots, \Upsilon_{M(\varepsilon)}$ such that $|(\theta,\alpha) - (\theta',\alpha')| < \varepsilon$ whenever (θ,α) and (θ',α') are in the same subset. Let (θ_j,α_j) denote *some* point in Υ_j for each j. Then,

$$\sup_{(\theta,\alpha)} |\hat{G}_{(i)}(\theta,\alpha) - G_{(i)}(\theta,\alpha)| = \max_j \sup_{\Upsilon_j} |\hat{G}_{(i)}(\theta,\alpha) - G_{(i)}(\theta,\alpha)|,$$

https://doi.org/10.1017/S0266466605050243 Published online by Cambridge University Press

and therefore

$$\Pr\left[\sup_{(\theta,\alpha)} |\hat{G}_{(i)}(\theta,\alpha) - G_{(i)}(\theta,\alpha)| \geq \eta\right] \leq \sum_{j=1}^{M(\varepsilon)} \Pr\left[\sup_{\Upsilon_j} |\hat{G}_{(i)}(\theta,\alpha) - G_{(i)}(\theta,\alpha)| \geq \eta\right]. \tag{B.3}$$

For $(\theta,\alpha) \in \Upsilon_j$, we have

$$|\hat{G}_{(i)}(\theta,\alpha) - G_{(i)}(\theta,\alpha)| \leq |\hat{G}_{(i)}(\theta_j,\alpha_j) - G_{(i)}(\theta_j,\alpha_j)| + \frac{\varepsilon}{T}\left|\sum_{t=1}^{T}(M(X_{it}) - E[M(X_{it})])\right| + 2\varepsilon E[M(X_{it})],$$

and therefore

$$\Pr\left[\sup_{\Upsilon_j} |\hat{G}_{(i)}(\theta,\alpha) - G_{(i)}(\theta,\alpha)| \geq \eta\right] \leq \Pr\left[|\hat{G}_{(i)}(\theta_j,\alpha_j) - G_{(i)}(\theta_j,\alpha_j)| \geq \frac{\eta}{3}\right] + \Pr\left[\frac{1}{T}\left|\sum_{t=1}^{T}(M(X_{it}) - E[M(X_{it})])\right| \geq \frac{\eta}{3\varepsilon}\right] = o(T^{-2}) \tag{B.4}$$

by Lemma 2. Combining (B.2)–(B.4) and $n = O(T)$, we obtain the desired conclusion. ■

APPENDIX C: EXPANSION

Let $\hat{\alpha}_i(\theta)$ be such that $\sum_{t=1}^{T} v(X_{it};\theta,\hat{\alpha}_i(\theta)) = 0$. Letting $U_i(X_{it};\theta,\alpha_i) \equiv u(X_{it};\theta,\alpha_i) - \rho_{i0}v(X_{it};\theta,\alpha_i)$, $\rho_{i0} \equiv E[\partial u(X_{it};\theta,\alpha_i)/\partial\alpha_i'](E[\partial v(X_{it};\theta,\alpha_i)/\partial\alpha_i'])^{-1}$ (in the likelihood framework, U is the efficient score for θ), we can recognize that $\hat{\theta}$ is a solution to

$$0 = \sum_{i=1}^{n}\sum_{t=1}^{T} U_i(X_{it};\hat{\theta},\hat{\alpha}_i(\hat{\theta})).$$

Let $F \equiv (F_1,\ldots,F_n)$ denote the collection of distribution functions F_i, where each F_i denotes the distribution function of X_{it}. Let $\hat{F} \equiv (\hat{F}_1,\ldots,\hat{F}_n)$, where $\hat{F}_i$ denotes the empirical distribution function for the stratum i. Define $F(\epsilon) \equiv F + \epsilon\Delta_{iT}$ for $\epsilon \in [0, T^{-1/2}]$, where $\Delta_{iT} \equiv \sqrt{T}(\hat{F}_i - F_i)$. For each fixed θ and ϵ, let $V(X_{it};\theta,\alpha_i) \equiv v(X_{it};\theta,\alpha_i)$ and $\alpha_i(\theta,F_i(\epsilon))$ be the solution to the estimating equation $0 = \int V_i[\theta,\alpha_i(\theta,F_i(\epsilon))]\,dF_i(\epsilon)$ and let $\theta(\epsilon)$ be the solution to the estimating equation $0 = \sum_{i=1}^{n}\int U_i(X_{it};\theta(\epsilon), \alpha_i(\theta(\epsilon),F_i(\epsilon)))\,dF_i(\epsilon)$. By Taylor series expansion, we have $\theta(\hat{F}) - \theta(F) = (1/\sqrt{T})\theta^{\epsilon}(0) + \frac{1}{2}(1/\sqrt{T})^2\theta^{\epsilon\epsilon}(0) + \frac{1}{6}(1/\sqrt{T})^3\theta^{\epsilon\epsilon\epsilon}(\tilde{\epsilon})$, where $\theta^{\epsilon}(\epsilon) \equiv d\theta(F(\epsilon))/d\epsilon$, $\theta^{\epsilon\epsilon}(\epsilon) \equiv d^2\theta(F(\epsilon))/d\epsilon^2,\ldots$, and $\tilde{\epsilon}$ is somewhere in between 0 and $T^{-1/2}$. We therefore have

$$\sqrt{nT}(\theta(\hat{F}) - \theta(F)) = \sqrt{nT}\,\frac{1}{\sqrt{T}}\,\theta^{\epsilon}(0) + \sqrt{nT}\,\frac{1}{2}\left(\frac{1}{\sqrt{T}}\right)^2\theta^{\epsilon\epsilon}(0) + \frac{1}{6}\sqrt{\frac{n}{T}}\,\frac{1}{\sqrt{T}}\,\theta^{\epsilon\epsilon\epsilon}(\tilde{\epsilon}). \tag{C.1}$$

https://doi.org/10.1017/S0266466605050243 Published online by Cambridge University Press

The last term in (C.1) can be shown to be $o_p(1)$ by the same method as in Hahn and Newey (2004).

LEMMA 4. *For every* $\eta > 0$, *we have* $\Pr[\max_{0\le\epsilon\le 1/\sqrt{T}}|\theta(\epsilon)-\theta_0| \ge \eta] = o(T^{-1})$ *and* $\Pr[\max_{1\le i\le n}\max_{0\le\epsilon\le 1/\sqrt{T}}|\alpha_i(\theta(\epsilon),F_i(\epsilon))-\alpha_{i0}| \ge \eta] = o(T^{-1})$.

Proof. Only the first assertion is proved. The second assertion can be proved similarly. Let η be given. Recall that $F(\epsilon) \equiv F + \epsilon\sqrt{T}(\hat{F}-F) = (1-\epsilon\sqrt{T})F + \epsilon\sqrt{T}\hat{F}$. We therefore have $\int g(\cdot;\theta,\alpha_i(\theta))\,dF_i(\epsilon) = (1-\epsilon\sqrt{T})G_{(i)}(\theta,\alpha_i) + \epsilon\sqrt{T}\hat{G}_{(i)}(\theta,\alpha_i)$, from which we find $|\int g(\cdot;\theta,\alpha_i(\theta))\,dF_i(\epsilon) - G_{(i)}(\theta,\alpha_i)| \le \epsilon\sqrt{T}|\hat{G}_{(i)}(\theta,\alpha) - G_{(i)}(\theta,\alpha)| \le |\hat{G}_{(i)}(\theta,\alpha) - G_{(i)}(\theta,\alpha)|$. By Lemma 3, we also have $\Pr[\max_{0\le\epsilon\le 1/\sqrt{T}}\max_{1\le i\le n}\sup_{(\theta,\alpha)}|\int g(\cdot;\theta,\alpha_i(\theta))\,dF_i(\epsilon) - G_{(i)}(\theta,\alpha)| \ge \eta] = o(T^{-1})$. Therefore, for every $0 \le \epsilon \le 1/\sqrt{T}$ with probability equal to $1-o(T^{-1})$, we have

$$
\begin{aligned}
&\min_{|\theta-\theta_0|>\eta,\alpha_1,\ldots,\alpha_n} n^{-1}\sum_{i=1}^{n}\left|\int g(\cdot;\theta,\alpha_i(\theta))\,dF_i(\epsilon)\right| \\
&\quad\ge \min_{|(\theta,\alpha)-(\theta_0,\alpha_{i0})|>\eta} n^{-1}\sum_{i=1}^{n}\left|\int g(\cdot;\theta,\alpha_i(\theta))\,dF_i(\epsilon)\right| \\
&\quad> \min_{|(\theta,\alpha)-(\theta_0,\alpha_{i0})|>\eta} n^{-1}\sum_{i=1}^{n}|G_{(i)}(\theta,\alpha_i)| - \frac{1}{2}\varepsilon > \frac{1}{2}\varepsilon > 0 \\
&\quad= n^{-1}\sum_{i=1}^{n}\left|\int g(\cdot;\hat{\theta}(\epsilon),\alpha_i(\hat{\theta}(\epsilon)))\,dF_i(\epsilon)\right|,
\end{aligned}
$$

where $\varepsilon \equiv \inf_i \inf_{\{(\theta,\alpha):|(\theta,\alpha)-(\theta_0,\alpha_{i0})|>\eta\}}|G_{(i)}(\theta,\alpha)| > 0$. It follows that

$\Pr[\max_{0\le\epsilon\le 1/\sqrt{T}}|\hat{\theta}(\epsilon)-\theta_0| \ge \eta] = o(T^{-1})$. ■

LEMMA 5. *Suppose that, for each* i, $\{\xi_{it}(\phi), t = 1,2,\ldots\}$ *is a sequence of zero mean i.i.d. random variables indexed by some parameter* $\phi \in \Phi$. *We assume that* $\{\xi_{it}(\phi), t = 1,2,\ldots\}$ *are independent across* i. *We also assume that* $\sup_{\phi\in\Phi}|\xi_{it}(\phi)| \le B_{it}$ *for some sequence of random variables* B_{it} *that is i.i.d. across* t *and independent across* i. *Finally, we assume that* $\max_i E[|B_{it}|^{64}] < \infty$ *and* $n = O(T)$. *We then have* $\Pr[\max_i|(1/\sqrt{T})\sum_{t=1}^{T}\xi_{it}(\phi_i)| > T^{(1/10)-\upsilon}] = o(T^{-1})$ *for every* υ *such that* $0 \le \upsilon < \frac{11}{160}$. *Here,* $\{\phi_i\}$ *is an arbitrary sequence in* Φ.

Proof. By Markov's inequality, we obtain

$$
\begin{aligned}
&T\Pr\left[\max_i\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi_{it}(\phi_i)\right| > T^{(1/10)-\upsilon}\right] \le T\sum_{i=1}^{n}\Pr\left[\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi_{it}(\phi_i)\right| > T^{(1/10)-\upsilon}\right] \\
&\quad= nT\cdot\Pr\left[\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi_{it}(\phi_i)\right| > T^{(3/5)-\upsilon}\right] \le nT\,\frac{E\left[\left|\sum_{t=1}^{T}\xi_{it}(\phi_i)\right|^{64}\right]}{T^{(3/5)\times 64-64\upsilon}}.
\end{aligned}
$$

https://doi.org/10.1017/S0266466605050243 Published online by Cambridge University Press

By Lemma 1, we have $E[|\sum_{t=1}^{T}\xi_{it}(\phi_i)|^{64}] \le CT^{32}\cdot E[|\xi_{it}(\phi_i)|^{64}] \le T^{32}\cdot C\max_i E[|B_{it}|^{64}]$ for some C. Therefore, we have

$$T\Pr\left[\max_i\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi_{it}(\phi_i)\right| > T^{(1/10)-\nu}\right] \le nT\,\frac{O(T^{32})}{T^{(3/5)\times 64-64\nu}} = o(1). \qquad \blacksquare$$

LEMMA 6. *Suppose that* $K_i(\cdot;\theta(\epsilon),\alpha_i(\theta(\epsilon),\epsilon))$ *is equal to* $\partial^{m_1+m_2}g(X_{it};\theta(\epsilon),\alpha_i(\theta(\epsilon),\epsilon))/\partial\theta^{m_1}\partial\alpha_i^{m_2}$ *for some* $m_1+m_2 \le 1,\ldots,5$. *Then, for any* $\eta > 0$, *we have*

$$\Pr\left[\max_{0\le\epsilon\le\frac{1}{\sqrt{T}}}\left|\frac{1}{n}\sum_{i=1}^{n}\int K_i(\cdot;\theta(\epsilon),\alpha_i(\theta(\epsilon),\epsilon))\,dF_i(\epsilon) - \frac{1}{n}\sum_{i=1}^{n}E[K_i(X_{it};\theta_0,\alpha_{i0})]\right| > \eta\right]$$

$$= o(T^{-1}),$$

$$\Pr\left[\max_i\max_{0\le\epsilon\le\frac{1}{\sqrt{T}}}\left|\int K_i(\cdot;\theta(\epsilon),\alpha_i(\theta(\epsilon),\epsilon))\,dF_i(\epsilon) - E[K_i(X_{it};\theta_0,\alpha_{i0})]\right| > \eta\right]$$

$$= o(T^{-1}).$$

Also,

$$\Pr\left[\max_i\max_{0\le\epsilon\le\frac{1}{\sqrt{T}}}\left|\int K_i(\cdot;\theta(\epsilon),\alpha_i(\theta(\epsilon),\epsilon))\,d\Delta_{iT}\right| > CT^{(1/10)-\nu}\right] = o(T^{-1})$$

for some constant $C > 0$ *and* $0 \le \nu < \frac{11}{160}$.

Proof. Note that we have

$$\left|\int K_i(\cdot;\theta(\epsilon),\alpha_i(\theta(\epsilon),\epsilon))\,dF_i(\epsilon) - \int K_i(\cdot;\theta(\epsilon),\alpha_i(\theta(\epsilon),\epsilon))\,dF_i\right|$$

$$\le \left|\int K_i(\cdot;\theta(\epsilon),\alpha_i(\theta(\epsilon),\epsilon))\,dF_i(\epsilon) - \int K_i(\cdot;\theta(0),\alpha_i(\theta(0),0))\,dF_i(\epsilon)\right|$$

$$+ \left|\int K_i(\cdot;\theta(0),\alpha_i(\theta(0),0))\,dF_i(\epsilon) - \int K_i(\cdot;\theta(0),\alpha_i(\theta(0),0))\,dF_i\right|$$

$$\le \int M(X_{it})(\|\theta(\epsilon)-\theta\| + |\alpha_i(\theta(\epsilon),\epsilon)-\alpha_i|)\,d|F_i(\epsilon)|$$

$$+ \epsilon\sqrt{T}\left|\int K_i(\cdot;\theta(0),\alpha_i(\theta(0),0))\,d(\hat{F}_i - F_i)\right|.$$

https://doi.org/10.1017/S0266466605050243 Published online by Cambridge University Press

Therefore, we have

$$\left| \frac{1}{n}\sum_{i=1}^{n} \int K_i(\cdot;\theta(\epsilon),\alpha_i(\theta(\epsilon),\epsilon))\, dF_i(\epsilon) - \frac{1}{n}\sum_{i=1}^{n} E[K_i(X_{it};\theta_0,\alpha_{i0})] \right|$$

$$\leq \|\theta(\epsilon)-\theta\| \cdot \frac{1}{n}\sum_{i=1}^{n}\left(E[M(X_{it})] + \frac{1}{T}\sum_{t=1}^{T} M(X_{it}) \right)$$

$$+ \left(\frac{1}{n}\sum_{i=1}^{n} (\alpha_i(\theta(\epsilon),\epsilon)-\alpha_i)^2 \right)^{1/2} \left(\frac{1}{n}\sum_{i=1}^{n}\left(E[M(X_{it})] + \frac{1}{T}\sum_{t=1}^{T} M(X_{it}) \right)^2 \right)^{1/2}$$

$$+ \left| \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T} K_i(X_{it};\theta(0),\alpha_i(\theta(0),0)) - E[K_i(X_{it};\theta(0),\alpha_i(\theta(0),0))] \right) \right|,$$

the right-hand side of which can be bounded by using Lemmas 4 and 5 in absolute value by some $\eta > 0$ with probability $1 - o(T^{-1})$, which proves the first claim. The second claim can be proved similarly.

As for the third claim, we can show using Lemma 5 that $\max_i |\int K_i(\cdot;\theta(\epsilon), \alpha_i(\theta(\epsilon),\epsilon))\, d\Delta_{iT}|$ can be bounded by in absolute value by $CT^{(1/10)-\nu}$ for some constant $C > 0$ and ν such that $0 \leq \nu < \frac{11}{160}$ with probability $1 - o(T^{-1})$. ■

LEMMA 7. $\Pr[\max_i \max_{0\leq\epsilon\leq 1/\sqrt{T}} |\alpha_i^{\theta}(\epsilon)| > C] = o(T^{-1})$ *and* $\Pr[\max_i \max_{0\leq\epsilon\leq 1/T} |\alpha_i^{\epsilon}(\epsilon)| > CT^{(1/10)-\nu}] = o(T^{-1})$ *for some constant* $C > 0$ *and* $0 \leq \nu < \frac{11}{160}$.

Proof. In Hahn and Kuersteiner (2003), it is shown that

$$\frac{\partial \alpha_i(\theta,F_i(\epsilon))}{\partial\theta} = -\left(\int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i}\, dF_i(\epsilon) \right)^{-1} \left(\int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\theta}\, dF_i(\epsilon) \right),$$

$$\frac{\partial \alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon} = -\left(\int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i}\, dF_i(\epsilon) \right)^{-1} \left(\int V_i(\cdot,\theta,\epsilon)\, d\Delta_{iT} \right).$$

Using Lemma 6, we can see that $(\int [\partial V_i(\cdot,\theta,\epsilon)/\partial\alpha_i]\, dF_i(\epsilon))^{-1}$ is uniformly bounded away from zero with probability $1 - o(T^{-1})$. We can also see that, with probability $1 - o(T^{-1})$, $\int [\partial V_i(\cdot,\theta,\epsilon)/\partial\theta]\, dF_i(\epsilon)$ is uniformly bounded by some constant C and $\int V_i(\cdot,\theta,\epsilon)\, d\Delta_{iT}$ is uniformly bounded by $CT^{(1/10)-\nu}$. ■

LEMMA 8. $\Pr[\max_{0\leq\epsilon\leq 1/\sqrt{T}} |\theta^{\epsilon}(\epsilon)| > CT^{(1/10)-\nu}] = o(T^{-1})$ *for some constant* $C > 0$ *and* $0 \leq \nu < \frac{11}{160}$.

Proof. In Hahn and Kuersteiner (2003), it is shown that $\theta^{\epsilon}(\epsilon)$ is equal to

$$\theta^{\epsilon}(\epsilon) = -\left[\frac{1}{n}\sum_{i=1}^{n} \int \left(\frac{\partial h_i(\cdot,\epsilon)}{\partial\theta'} + \frac{\partial h_i(\cdot,\epsilon)}{\partial\alpha_i} \frac{\partial\alpha_i}{\partial\theta'} \right) dF_i(\epsilon) \right]^{-1}$$

$$\times \left[\frac{1}{n}\sum_{i=1}^{n} \frac{\partial\alpha_i}{\partial\epsilon} \left(\int \frac{\partial h_i(\cdot,\epsilon)}{\partial\alpha_i}\, dF_i(\epsilon) \right) + \frac{1}{n}\sum_{i=1}^{n} \int h_i(\cdot,\epsilon)\, d\Delta_{iT} \right]. \qquad \textbf{(C.2)}$$

https://doi.org/10.1017/S0266466605050243 Published online by Cambridge University Press

Using Lemmas 6 and 7 we can bound the denominator of $\theta^{\epsilon}(\epsilon)$ by some $C > 0$ and the numerator by some $CT^{(1/10)-\nu}$ with probability $1 - o(T^{-1})$. ■

LEMMA 9.

$$\Pr\left[\max_i \max_{0\le\epsilon\le\frac{1}{\sqrt{T}}} |\alpha_i^{\theta_r\theta_{r'}}(\epsilon)| > C\right] = o(T^{-1}),$$

$$\Pr\left[\max_i \max_{0\le\epsilon\le\frac{1}{\sqrt{T}}} |\alpha_i^{\theta_r\epsilon}(\epsilon)| > CT^{(1/10)-\nu}\right] = o(T^{-1}),$$

$$\Pr\left[\max_i \max_{0\le\epsilon\le\frac{1}{\sqrt{T}}} |\alpha_i^{\epsilon\epsilon}(\epsilon)| > C(T^{(1/10)-\nu})^2\right] = o(T^{-1})$$

for some constant $C > 0$ *and* $0 \le \nu < \frac{11}{160}$. *Here,* $\alpha_i^{\theta_r\theta_{r'}} \equiv \partial^2\alpha_i/\partial\theta_r\partial\theta_{r'}$. *We similarly define* $\alpha_i^{\theta_r\epsilon}$.

Proof. By repeatedly differentiating $0 = \int V_i(\cdot,\theta,\epsilon)\,dF_i(\epsilon)$ with respect to ϵ, we obtain

$$\begin{aligned}
0 = {} & \int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\theta\partial\theta'}\,dF_i(\epsilon) + \frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\theta}\left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i\partial\theta'}\,dF_i(\epsilon)\right) \\
& + \left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i\partial\theta}\,dF_i(\epsilon)\right)\frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\theta'} \\
& + \left(\int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i}\,dF_i(\epsilon)\right)\frac{\partial^2\alpha_i(\theta,F_i(\epsilon))}{\partial\theta\partial\theta'} \\
& + \left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i^2}\,dF_i(\epsilon)\right)\frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\theta}\,\frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\theta'}, \\
0 = {} & \left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\theta\partial\alpha_i}\,dF_i(\epsilon)\right)\frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon} \\
& + \left(\int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i}\,dF_i(\epsilon)\right)\frac{\partial^2\alpha_i(\theta,F_i(\epsilon))}{\partial\theta\partial\epsilon} \\
& + \left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i^2}\,dF_i(\epsilon)\right)\frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\theta}\,\frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon} \\
& + \int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\theta}\,d\Delta_{iT} + \left(\int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i}\,d\Delta_{iT}\right)\frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\theta},
\end{aligned}$$

and

$$\begin{aligned}
0 = {} & \left(\int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i}\,dF_i(\epsilon)\right)\frac{\partial^2\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon^2} \\
& + \left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i^2}\,dF_i(\epsilon)\right)\left(\frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon}\right)^2 \\
& + 2\left(\int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i}\,d\Delta_{iT}\right)\frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon}.
\end{aligned}$$

https://doi.org/10.1017/S0266466605050243 Published online by Cambridge University Press

The result then follows by applying the same argument as in the proof of Lemma 7. ■

LEMMA 10. $\Pr[\max_{0\le\epsilon\le 1/\sqrt{T}}|\theta^{\epsilon\epsilon}(\epsilon)| > C(T^{(1/10)-\upsilon})^2] = o(T^{-1})$ *for some constant* $C > 0$ *and* $0 \le \upsilon < \frac{11}{160}$.

Sketch of Proof. By repeatedly differentiating $0 = \sum_{i=1}^{n}\int U_i(X_{it};\theta(\epsilon),\alpha_i(\theta(\epsilon),F_i(\epsilon)))\,dF_i(\epsilon)$ with respect to ϵ, we obtain a characterization of $\theta^{\epsilon\epsilon}(\epsilon)$. (For more detailed characterization, see Hahn and Kuersteiner, 2003.) The conclusion follows by combining it with Lemmas 6–9. ■

APPENDIX D: UNIFORM CONSISTENCY OF $\hat{\alpha}_i$

THEOREM 2. *Under Conditions 3, 5, 6, and 7, we have*

$$\sqrt{T}(\hat{\alpha}_i - \alpha_{i0}) = \lambda_i + \kappa_i,$$

where

$$\lambda_i \equiv -\left(E\left[\frac{\partial v_{it}}{\partial \alpha_i}\right]\right)^{-1}\left(T^{-1/2}\sum_{t=1}^{T} v_{it}\right) = O_p(1)$$

and $\Pr[\max_i|\kappa_i| \ge \eta] = o(1)$ *for every* $\eta > 0$. *Here,* $v_{it} \equiv v_{it}(X_{it},\theta_0,\alpha_{i0})$.

Let $\hat{\alpha}_i(\epsilon)$ denote α_i that sets $\int V_i(\cdot;\hat{\theta}(\epsilon),\alpha_i)\,dF_i(\epsilon)$ equal to zero. (In other words, let $\hat{\alpha}_i(\epsilon) \equiv \alpha_i(\theta(\epsilon),F_i(\epsilon))$.) We then have the expansion

$$\hat{\alpha}_i(\epsilon) - \alpha_{i0} = \frac{1}{\sqrt{T}}\,\hat{\alpha}_i^{\epsilon}(0) + \frac{1}{2T}\,\hat{\alpha}_i^{\epsilon\epsilon}(\tilde{\epsilon}) \tag{D.1}$$

for some $\tilde{\epsilon} \in [0, T^{-1/2}]$. Let $v_i(\cdot,\epsilon) \equiv V_i(\theta(F(\epsilon)),\alpha_i(F_i(\epsilon)))$. The first-order condition may be written as $0 = \int v_i(\cdot,\epsilon)\,dF_i(\epsilon)$. Differentiating repeatedly with respect to ϵ, we obtain

$$0 = \int \frac{dv_i(\cdot,\epsilon)}{d\epsilon}\,dF_i(\epsilon) + \int v_i(\cdot,\epsilon)\,d\Delta_{iT}, \tag{D.2}$$

$$0 = \int \frac{d^2v_i(\cdot,\epsilon)}{d\epsilon^2}\,dF_i(\epsilon) + 2\int \frac{dv_i(\cdot,\epsilon)}{d\epsilon}\,d\Delta_{iT}. \tag{D.3}$$

Because $dv_i(\cdot,\epsilon)/d\epsilon = [\partial v_i(\cdot,\epsilon)/\partial\theta'](\partial\theta/\partial\epsilon) + [\partial v_i(\cdot,\epsilon)/\partial\alpha_i](\partial\alpha_i/\partial\epsilon)$, (D.2) implies that

$$\hat{\alpha}_i^{\epsilon}(\epsilon) = -\left(\int \frac{\partial v_i(\cdot,\epsilon)}{\partial\alpha_i}\,dF_i(\epsilon)\right)^{-1}\left[\int\left(\frac{\partial v_i(\cdot,\epsilon)}{\partial\theta'}\right)dF_i(\epsilon)\theta^{\epsilon}(\epsilon) + \int v_i(\cdot,\epsilon)\,d\Delta_{iT}\right].$$

Evaluating at $\epsilon = 0$ we obtain

$$\hat{\alpha}_i^{\epsilon}(0) = -(E[V_i^{\alpha}])^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} v_i(\cdot,\theta_0,\alpha_{i0}) + E[V_i^{\theta}]\theta^{\epsilon}(0)\right), \tag{D.4}$$

https://doi.org/10.1017/S0266466605050243 Published online by Cambridge University Press

where $V_i^{\alpha} \equiv \partial v(X_{it};\theta,\alpha_i)/\partial\alpha_i$, $V_i^{\theta} \equiv \partial v(X_{it};\theta,\alpha_i)/\partial\theta$, and $\theta^{\epsilon}(0)$ can be deduced from (C.2). Next, consider

$$\frac{d^2 v_i(\cdot,\epsilon)}{d\epsilon^2} = \frac{\partial\theta'}{\partial\epsilon}\frac{\partial v_i(\cdot,\epsilon)}{\partial\theta\partial\theta'}\frac{\partial\theta}{\partial\epsilon} + \frac{\partial v_i(\cdot,\epsilon)}{\partial\theta'}\frac{\partial^2\theta}{(\partial\epsilon)^2} + 2\frac{\partial v_i(\cdot,\epsilon)}{\partial\theta'\partial\alpha_i}\frac{\partial\theta}{\partial\epsilon}\frac{\partial\alpha_i}{\partial\epsilon} + \frac{\partial^2 v_i(\cdot,\epsilon)}{(\partial\alpha_i)^2}\left(\frac{\partial\alpha_i}{\partial\epsilon}\right)^2 + \frac{\partial v_i(\cdot,\epsilon)}{\partial\alpha_i}\frac{\partial^2\alpha_i}{(\partial\epsilon)^2}$$

such that $\hat{\alpha}_i^{\epsilon\epsilon}(\epsilon)$ is characterized by

$$\begin{aligned} 0 = {} & \theta^{\epsilon}(\epsilon)' \int \frac{\partial v_i(\cdot,\epsilon)}{\partial\theta\partial\theta'} dF_i(\epsilon)\theta^{\epsilon}(\epsilon) + \int \frac{\partial v_i(\cdot,\epsilon)}{\partial\theta'} dF_i(\epsilon)\theta^{\epsilon\epsilon}(\epsilon) \\ & + 2\int \frac{\partial v_i(\cdot,\epsilon)}{\partial\theta'\partial\alpha_i} dF_i(\epsilon)\theta^{\epsilon}(\epsilon)\hat{\alpha}_i^{\epsilon}(\epsilon) + \int \frac{\partial^2 v_i(\cdot,\epsilon)}{(\partial\alpha_i)^2} dF_i(\epsilon)(\hat{\alpha}_i^{\epsilon}(\epsilon))^2 \\ & + \int \frac{\partial v_i(\cdot,\epsilon)}{\partial\alpha_i} dF_i(\epsilon)\hat{\alpha}_i^{\epsilon\epsilon}(\epsilon) + \int \frac{\partial v_i(\cdot,\epsilon)}{\partial\theta'} d\Delta_{iT}\theta^{\epsilon}(\epsilon) \\ & + \int \frac{\partial v_i(\cdot,\epsilon)}{\partial\alpha_i} d\Delta_{iT}^{\epsilon}\hat{\alpha}i(\epsilon). \end{aligned} \tag{D.5}$$

We now combine (D.1), (D.4), and

$$\theta^{\epsilon}(0) = \frac{1}{\sqrt{n}}\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\left(\frac{1}{\sqrt{n}\sqrt{T}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_i\right) \tag{D.6}$$

and obtain

$$\begin{aligned} \sqrt{T}(\hat{\alpha}_i - \alpha_{i0}) = {} & -(E[V_i^{\alpha}])^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} v_i(\cdot,\theta_0,\alpha_{i0})\right) \\ & - \frac{1}{\sqrt{n}}(E[V_i^{\alpha}])^{-1}E[V_i^{\theta}]\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\left(\frac{1}{\sqrt{n}\sqrt{T}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_i\right) \\ & + \frac{1}{2\sqrt{T}}\hat{\alpha}_i^{\epsilon\epsilon}(\tilde{\epsilon}). \end{aligned} \tag{D.7}$$

Here, (D.6) can be obtained by evaluating (C.2) at $\epsilon = 0$.

In light of (D.7), Theorem 2 can be obtained by showing that

$$\Pr\left[\max_i \sup_{\epsilon\in[0,1/\sqrt{T}]} |\hat{\alpha}_i^{\epsilon\epsilon}(\epsilon)| > (T^{(1/10)-\upsilon})^2\right] = o(T^{-1}) \tag{D.8}$$

for any $0 \le \upsilon < \frac{11}{160}$. This follows from representation (D.5) and also from Lemmas 6, 8, and 10.

https://doi.org/10.1017/S0266466605050243 Published online by Cambridge University Press

APPENDIX E: PROOF OF THEOREM 1

Theorem 2 implies that we have

$$\sqrt{n}(\hat{\delta} - \delta_0) = \left(n^{-1} \sum_{i=1}^{n} z_i w_i' \right)^{-1} \left(n^{-1/2} \sum_{i=1}^{n} z_i \gamma_{i0} + n^{-1/2} T^{-1/2} \sum_{i=1}^{n} z_i \lambda_i \right) + o_p(1).$$

Under Conditions 4 and 5, the term $n^{-1/2} T^{-1/2} \sum_{i=1}^{n} z_i \lambda_i$ is of order $o_p(1)$, and substitution of $\hat{\alpha}_i$ for α_{i0} in (3) does not affect the asymptotic distribution of the resultant estimator (4):

$$\sqrt{n}(\hat{\delta} - \delta_0) = \left(n^{-1} \sum_{i=1}^{n} z_i w_i' \right)^{-1} \left(n^{-1/2} \sum_{i=1}^{n} z_i \gamma_{i0} \right) + o_p(1). \quad \blacksquare$$

https://doi.org/10.1017/S0266466605050243 Published online by Cambridge University Press