Econometric Theory, **27**, 2011, 1152–1191.
doi:10.1017/S0266466611000028

BIAS REDUCTION FOR DYNAMIC NONLINEAR PANEL MODELS WITH FIXED EFFECTS

JINYONG HAHN
UCLA

GUIDO KUERSTEINER
Georgetown University

The fixed effects estimator of panel models can be severely biased because of well-known incidental parameter problems. It is shown that this bias can be reduced in nonlinear dynamic panel models. We consider asymptotics where n and T grow at the same rate as an approximation that facilitates comparison of bias properties. Under these asymptotics, the bias-corrected estimators we propose are centered at the truth, whereas fixed effects estimators are not. We discuss several examples and provide Monte Carlo evidence for the small sample performance of our procedure.

1. INTRODUCTION

An intuitive method of dealing with unobserved individual effects in panel data is to treat each effect as a separate parameter to be estimated. Unfortunately, such estimators are typically severely biased due to the incidental parameters problem noted by Neyman and Scott (1948). Hahn and Kuersteiner (2002) showed that this type of bias in a panel AR(1) model can be alleviated by considering an alternative approximation where the number of individuals (n) and the number of time series observations (T) grow to infinity at the same rate.[1] Hahn and Newey (2004) showed how the bias correction can be implemented in a static nonlinear panel data model with fixed effects under the same asymptotics. In this paper we examine asymptotic biases of general *dynamic nonlinear* panel models with fixed effects and develop methods to remove them.

We are grateful for helpful comments by Gary Chamberlain, Shakeeb Khan, Whitney Newey, the editor, and three anonymous referees. We also benefited from comments by workshop participants at AUEB, Boston University, Columbia University, Federal Reserve Board, Georgetown University, Harvard/MIT, Johns Hopkins University, Iowa State University, Panel Conference sponsored by NSF-NBER-UCLA, Penn State University, Linz Time Series Workshop, Triangle Seminar, Université de Montréal, UC Irvine, UC San Diego, University of Cyprus, University of Pittsburgh, USC, and Yale University. We thank Ekaterini Kyriazidou for making her Gauss code available. The first author gratefully acknowledges financial support from NSF Grant SES-0921187 and SES-0819638. The second author gratefully acknowledges financial support from NSF Grant SES-0095132 and SES-0523186. Address correspondence to Guido Kuersteiner, Georgetown University, Department of Economics, ICC 572, 37th and O Streets, Washington, D.C. 20057; e-mail:gk232@georgetown.edu.

 © Cambridge University Press 2011

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press

One advantage of our approach is that it is quite flexible, and that it only requires a Taylor series expansion of the first-order condition. We discuss examples that show how our method can be applied to a variety of nonlinear models.

Our analysis could provide an alternative to estimators that have desirable properties under different asymptotics where n grows to infinity with T fixed. Such estimators tend to be model-specific and difficult to generalize. For example, Honoré and Kyriazidou's (2000) identification and estimation methodology does not seem to generalize to models other than the logit case. On the other hand, it is expected that our approach requires moderately large T.

A potential drawback of any bias correction method is that the estimators of the fixed effects are needed as preliminary estimators for estimates of the bias and may affect the quality of the latter. Our Monte Carlo experiments shed some light on the finite sample performance of our method. Our results indicate that the bias corrections we propose are effective at removing bias and do not lead to any significant deterioration of estimator variance.[2]

2. EXAMPLES AND REGULARITY CONDITIONS

In this section we formally introduce the fixed effects estimator and our bias-corrected estimator. We also provide an asymptotic argument to justify our method.

Suppose that we are given a panel data model with a common parameter of interest θ_0 and individual specific fixed effects γ_{i0}, $i = 1, \ldots, n$. We define a maximization estimator by

$$\left(\hat{\theta}, \hat{\gamma}_1, \ldots, \hat{\gamma}_n\right) = \underset{\theta, \gamma_1, \ldots, \gamma_n}{\operatorname{argmax}} \sum_{i=1}^{n} \sum_{t=1}^{T} \psi\left(x_{it}; \theta, \gamma_i\right) \tag{1}$$

for some criterion function $\psi(\cdot)$ that does not depend on T. We assume that ψ is a sensible function in the sense that, if n is fixed, and $T \to \infty$, the estimator $\left(\hat{\theta}, \hat{\gamma}_1, \ldots, \hat{\gamma}_n\right)$ is consistent for $(\theta_0, \gamma_{10}, \ldots, \gamma_{n0})$. In practice ψ is often chosen to be the likelihood function for a particular parametric family of distributions.[3] To simplify notation, we assume $\dim(\gamma_i) = 1$. We use the following notation throughout the paper:

$$U_i\left(x_{it}; \theta, \gamma_i\right) \equiv \frac{\partial \psi\left(x_{it}; \theta, \gamma_i\right)}{\partial \theta} - \rho_{i0} \cdot \frac{\partial \psi\left(x_{it}; \theta, \gamma_i\right)}{\partial \gamma_i},$$

$$V_i\left(x_{it}; \theta, \gamma_i\right) \equiv \frac{\partial \psi\left(x_{it}; \theta, \gamma_i\right)}{\partial \gamma_i},$$

$$\rho_{i0} \equiv \mathrm{E}\left[\frac{\partial^2 \psi\left(x_{it}; \theta_0, \gamma_{i0}\right)}{\partial \theta \partial \gamma_i}\right] \Bigg/ \mathrm{E}\left[\frac{\partial^2 \psi\left(x_{it}; \theta_0, \gamma_{i0}\right)}{\partial \gamma_i^2}\right],$$

$$\mathcal{I}_i \equiv -\mathrm{E}\left[\frac{\partial U_i\left(x_{it}; \theta_0, \gamma_{i0}\right)}{\partial \theta'}\right].$$

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press

We use the short-hand notation $U_{it} \equiv U_i(x_{it}; \theta_0, \gamma_{i0})$ and $V_{it} \equiv V_i(x_{it}; \theta_0, \gamma_{i0})$. We will denote by $U_{it}^{\gamma_i}$ and $U_{it}^{\gamma_i \gamma_i}$ the first and second derivatives of U_{it} with respect to γ_i. Likewise, we will denote by $V_{it}^{\gamma_i}$ the derivative of V_{it} with respect to γ_i.

We now discuss a number of models that fall within the scope of our method.

Example 1
The binary panel model with fixed effects is characterized by $y_{it} = 1(\gamma_{i0} + z_{it}'\theta_0 + \varepsilon_{it} \geq 0)$, where ε_{it} conditional on z_{it} either has a logistic or standard normal distribution. The maximum likelihood estimator (MLE) which is consistent as $T \to \infty$, is a special case of (1), with $x_{it} = \left(y_{it}, z_{it}'\right)'$ and $\psi(x_{it}; \theta, \gamma_i) = y_{it} \log \Lambda\left(\gamma_i + z_{it}'\theta\right) + (1 - y_{it}) \log\left(1 - \Lambda\left(\gamma_i + z_{it}'\theta\right)\right)$, where Λ denotes the cumulative distribution function (CDF) of a logistic or standard normal distribution. Extensions to multinomial logit follow easily in the same way and have been applied by, for example, Hendel and Nevo (2006).

Example 2
The dynamic binary panel model with fixed effects is characterized by the forward solution of $y_{it} = 1(\gamma_{i0} + z_{it}'\zeta_0 + \tau_0 y_{i,t-1} + \varepsilon_{it} \geq 0)$ with initial condition y_{i0} drawn from the stationary distribution. Here ε_{it} either has a logistic or standard normal distribution. The MLE is a special case of (1), with $x_{it} = \left(y_{it}, z_{it}', y_{it-1}\right)'$, $\theta = \left(\zeta', \tau\right)'$, and $\psi(x_{it}; \theta, \gamma_i) = y_{it} \log \Lambda\left(\gamma_i + z_{it}'\zeta + \tau y_{it-1}\right) + (1 - y_{it}) \log\left(1 - \Lambda\left(\gamma_i + z_{it}'\zeta + \tau y_{it-1}\right)\right)$. This model has been applied by Chintagunta, Kyriazidou, and Perktold (2001) to household brand choices.

Example 3
Tobit models with lagged dependent variables and fixed effects are characterized by the forward solution to $y_{it} = \max\left(0, z_{it}'\zeta_0 + \tau_0 y_{it-1} + \gamma_{i0} + \varepsilon_{it}\right)$ with initial condition y_{i0} drawn from the stationary distribution. If ε_{it} is independent and identically distributed (i.i.d.) $N\left(0, \sigma_\varepsilon^2\right)$ as in the previous example, we obtain the MLE by setting

$$\psi(x_{it}; \theta, \gamma_i) = 1(y_{it} = 0) \log \Lambda\left(\left(\tau y_{it-1} + z_{it}'\zeta + \gamma_i\right)/\sigma_\varepsilon^2\right) + 1(y_{it} > 0)\, \lambda\left(\left(y_{it} - \tau y_{it-1} - z_{it}'\zeta - \gamma_i\right)/\sigma_\varepsilon^2\right) \Big/ \sigma_\varepsilon,$$

where Λ is the CDF of the standard normal distribution and λ is the corresponding density. It is a special case of (1), with $x_{it} = \left(y_{it}, z_{it}', y_{it-1}\right)'$, and $\theta = \left(\zeta', \tau, \sigma_\varepsilon^2\right)'$. For the special case where $\tau_0 = 0$, we have a Tobit model with unobserved effects that was considered in Heckman and Macurdy (1980) and Honoré (1992).

Our theoretical results hold under a number of regularity conditions that are listed next. Below we verify the regularity conditions for each of the previous examples.

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press

Condition 1. For each $\eta > 0$, $\inf_i \left[G_{(i)}(\theta_0, \gamma_{i0}) - \sup_{\{(\theta,\gamma):\|(\theta,\gamma)-(\theta_0,\gamma_{i0})\|>\eta\}} G_{(i)}(\theta,\gamma) \right] > 0$, where $G_{(i)}(\theta,\gamma_i) \equiv \mathrm{E}[\psi(x_{it};\theta,\gamma_i)]$, and $\|\cdot\|$ denotes the euclidean norm.

Condition 2. $n, T \to \infty$ such that $\frac{n}{T} \to \kappa$, where $0 < \kappa < \infty$.

Condition 3. (i) For each i, $\{x_{it}, t = 1, 2, \ldots\}$ is a stationary mixing sequence; (ii) $\{x_{it}, t = 1, 2, \ldots\}$ are independent across i; (iii) $\sup_i |\alpha_i(m)| \leq Ca^m$ for some a such that $0 < a < 1$ and some $C > 0$, where $\mathcal{A}_t^i \equiv \sigma(x_{it}, x_{it-1}, x_{it-2}, \ldots)$, $\mathcal{B}_t^i \equiv \sigma(x_{it}, x_{it+1}, x_{it+2}, \ldots)$, and

$$\alpha_i(m) \equiv \sup_t \sup_{A\in\mathcal{A}_t^i, B\in\mathcal{B}_{t+m}^i} |P(A\cap B) - P(A)P(B)|.$$

Condition 4. Let $\psi(x_{it},\phi)$ be a function indexed by the parameter $\phi = (\theta,\gamma) \in \operatorname{int}\Phi$, where Φ is a compact, convex subset of $\mathbb{R}^p$, $p \equiv \dim(\phi) = R + 1$, and $R = \dim(\theta)$. Let $\nu = (\nu_1, \ldots, \nu_p)$ be a vector of nonnegative integers ν_i, $|\nu| = \sum_{j=1}^p \nu_j$ and $D^\nu \psi(x_{it},\phi) = \partial^{|\nu|}\psi(x_{it},\phi) / \left(\partial\phi_1^{\nu_1} \ldots \partial\phi_p^{\nu_p}\right)$. There exists a function $M(x_{it})$ such that $\|D^\nu\psi(x_{it},\phi_1) - D^\nu\psi(x_{it},\phi_2)\| \leq M(x_{it})\|\phi_1 - \phi_2\|$ for all $\phi_1, \phi_2 \in \Phi$ and $|\nu| \leq 5$. The function $M(x_{it})$ satisfies $\sup_{\phi\in\Phi}\|D^\nu\psi(x_{it},\phi)\| \leq M(x_{it})$ and

$$\sup_i \mathrm{E}\left[|M(x_{it})|^{(10+10q)/(1-10\upsilon)+\delta}\right] < \infty \tag{2}$$

for some integer $q \geq p/2+2$, some $\delta > 0$, and $0 < \upsilon < 1/10$.

Condition 5. Let λ_{iT} denote the smallest eigenvalue of $\Sigma_{iT} = \mathrm{Var}\left(T^{-1/2}\sum_{t=1}^T U_i(x_{it};\theta_0,\gamma_{i0})\right)$. We assume that $\inf_i \inf_T \lambda_{iT} > 0$.

Condition 6. $\inf_i \left|\mathrm{E}\left[\partial V_i(x_{it};\theta_0,\gamma_{i0}) / \partial\gamma_i\right]\right| > 0$.

Condition 7. Let $\mu_{i1} \leq \ldots \leq \mu_{ik} \leq \ldots \leq \mu_{iR}$ be the eigenvalues of $\mathcal{I}_i$ in ascending order. Assume that (i) $0 < \inf_i \mu_{i1} \leq \sup_i \mu_{iR} < \infty$; (ii) $\lim_{n\to\infty} n^{-1} \sum_{i=1}^n \mathcal{I}_i$ exists; (iii) letting $\mathcal{I} \equiv \lim_{n\to\infty} n^{-1}\sum_{i=1}^n \mathcal{I}_i$, we assume that $\mathcal{I}$ is positive definite.

Remark 1. Condition 3 implicitly imposes that the fixed effects γ_i are constant and uniformly bounded in i. In applications where a model with randomly realized individual effects γ_i that are possibly correlated with covariates is more desirable, our analysis can be interpreted as being conditional on the realizations for γ_i. In this case the stationarity requirement needs to be restated as follows: "For each i, $\{x_{it}, t = 1, 2, \ldots\}$ is a stationary mixing sequence conditional on γ_i." Of course, if the sequence $\{x_{it}, t = 1, 2, \ldots\}$ is stationary conditionally, it is also stationary unconditionally. Similarly, Condition 3(iii) needs to be modified as follows: $\sup_i |\alpha_i(m|\gamma_i)| \leq Ca^m$ a.s. (almost surely) where $\alpha_i(m|\gamma_i) \equiv \sup_t \sup_{A\in\mathcal{A}_t^i, B\in\mathcal{B}_{t+m}^i}$

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press

$|P(A \cap B|\gamma_i) - P(A|\gamma_i)\,P(B|\gamma_i)|$ a.s. Remarks 3 and 4 below discuss specific examples of models with random effects.

Remark 2. The uniform bound $\sup_i |\alpha_i(m|\gamma_i)| \leq Ca^m$ a.s. discussed in Remark 1 as well as Conditions 3(i) and 4 impose implicit restrictions on the marginal distribution of γ_i. A sufficient restriction for Conditions 3 and 4 to hold in our examples is that the support of the distribution of γ_i is bounded uniformly in i.

Condition 1 is a sufficient condition that guarantees that the parameters are identified based on time series variation. This is a typical condition usually invoked to prove consistency of extremum estimators. Condition 2 formalizes our asymptotic approximation, where n and T go to infinity at the same rate. Condition 3 restricts the serial dependence in each time series $\{x_{it}, t = 1, 2, \ldots\}$. Condition 3 also imposes stationarity, which rules out time trends. This is one important drawback of our approach. Stationarity is sometimes used as a way to solve the initial conditions problem associated with a random effects approach. Therefore, there might be less of a case of taking a fixed effects approach under stationarity. Finally, Condition 3 imposes independence but no homogeneity assumptions on the behavior of x_{it} across i. In particular, we impose a uniform upper bound on the decay rate of temporal dependence without requiring that dependence be homogeneous across i. We assume that this uniform upper bound is decaying exponentially with lag length. Exponential rates of decay are not necessary to establish first-order asymptotic results. At the cost of more complicated notation in the assumptions and proofs, sharper upper bounds decaying at geometric rates could be given but all of our examples are mixing at exponential rates. Condition 4 imposes a moment condition on the envelope function $M(x_{it})$. The moment condition is chosen so that Hansen's (1991) maximal inequality can be applied. Hahn and Newey (2004) require somewhat stronger restrictions on moments because they analyze the jackknife estimator, which requires expansions of higher orders than is the case here. Conditions 5, 6, and 7 are put in place to rule out the possibility that the properties of any estimator can be influenced by only a small number of i's.

We discuss in turn the additional restrictions necessary such that our examples satisfy Conditions 1–7. In Example 1, we can first see from Newey and McFadden (1994, p. 2125) that the probit likelihood has a unique maximum and $\mathrm{E}[|\psi(x_{it}; \theta, \gamma_i)|] < \infty$ if $\mathrm{E}\left[w_{it} w_{it}'\right]$ is positive definite where $w_{it} = \left(1, z_{it}'\right)'$. Also, $\psi(x_{it}; \theta, \gamma_i)$ is globally concave such that $G_{(i)}(\theta, \gamma_i)$ has a unique global optimum. Similar arguments hold for the logit case. Condition 3 is satisfied if z_{it} is a stationary mixing sequence with mixing coefficients α_i such that $\sup_i |\alpha_i(m)| \leq Ca^m$ for some a such that $0 < a < 1$ and some $C > 0$ because y_{it} is a measurable function of z_{it} and thus inherits the mixing and stationarity properties from z_{it}. For logit and probit models all finite order derivatives of $\log \Lambda(v)$ exist and are continuous. This implies that the derivatives are uniformly bounded on a compact parameter space Φ and satisfy the Lipschitz condition of Condition 4. For the

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press

logit model with $\Lambda(x_{it},\phi)=(1+\exp(-(\gamma_i+z_{it}\theta)))^{-1}$, it follows that all the derivatives of $\log\Lambda(x_{it},\phi)$ are of the form $D^v\left[\left(z_{it}'\phi\right)^{|v|}\right]g_{|v|}(z_{it}'\phi)$ with $g_{|v|}(\cdot)$ a uniformly bounded function. The moment restrictions of Condition 4 are then satisfied if $\sup_i \mathrm{E}\left[\left|z_{it}^j\right|^{|v|+(10+10q)/(1-10v)+\delta}\right]<\infty$ for $j=1,\ldots,R$ where z_{it}^j is the jth element of z_{it}. Similar results hold for the probit model where we also require the same moment bounds on z_{it}^j. If the regressors are assumed to have bounded support, this requirement is automatically satisfied. For the probit model Conditions 5, 6, and 7 hold if $\mathrm{E}\left[w_{it}w_{it}'\right]$ is positive definite uniformly in i by the results of Newey and McFadden (1994, p. 2147 and fn 29).

For Example 2 we note that De Jong and Woutersen (2011) discuss conditions for consistency such that for $w_{it}=\left(1,y_{it-1},z_{it}'\right)'$ we need $\mathrm{E}\left[w_{it}w_{it}'\right]$ nonsingular. Stationarity and exponential mixing conditions hold if z_{it} is strictly stationary with mixing coefficients α_i such that $\sup_i|\alpha_i(m)|\leq Ca^m$ for some a such that $0<a<1$ and some $C>0$ and ε_{it} is i.i.d. Then, by de Jong and Woutersen (2011, Thm. 2), it follows that (y_{it},z_{it}) is strong mixing with strong mixing coefficients $\beta_i=C_0(C_1\exp(-C_2m)+\alpha_i(m))$ for positive constants C_0,C_1,C_2, and is strictly stationary. The moment bounds are satisfied as before if $\sup_i \mathrm{E}\left[\left|z_{it}^j\right|^{|v|+(10+10q)/(1-10v)+\delta}\right]<\infty$, and the remaining conditions hold under the same assumptions as previously discussed.

For Example 3 we first note that identification as in Condition 1 follows if $\left(\zeta_0,\tau_0,\gamma_{i0},\sigma_{i\varepsilon}^2\right)\in\operatorname{int}\Phi$, $\mathrm{E}\left[w_{it}w_{it}'\right]$ is positive definite for $w_{it}=(1,y_{it-1},z_{it}')'$, and ε_{it} is i.i.d. Gaussian from Amemiya (1973) and Olsen (1978). We establish stationarity and exponential mixing properties in Hahn and Kuersteiner (2010b) under some regularity conditions. These are $|\tau_0|<1$, ε_{it} is i.i.d. across i and t with $\varepsilon_{it}\sim N\left(0,\sigma_\varepsilon^2\right)$ and $z_{it}=\Pi z_{it-1}+\eta_{it}$ where η_{it} is i.i.d. and independent of ε_{js} for all j and s, η_{it} has density $p_\eta(\eta)$ such that $\int_\eta\left|p_\eta(\eta+x)-p_\eta(\eta)\right|d\eta\leq C_\eta\|x\|$ for some $0<C_\eta<\infty$, $\mathrm{E}\left[\|\eta_{it}\|^r\right]<\infty$ for some $r>2$, and $\max_j\left|\pi_j\right|<1$ where π_j are the eigenvalues of Π. Let y_{it} and $\psi(x_{it};\theta,\gamma_i)$ be as defined before in Example 3. Then, Condition 4 is satisfied if (y_{it},z_{it}) is strictly stationary and the moment bounds $\sup_i \mathrm{E}\left[\left|\eta_{it}^j\right|^{|v|+(10+10q)/(1-10v)+\delta}\right]<\infty$ hold where η_{it}^j is the jth element of η_{it}. Conditions 5, 6, and 7 hold if $\mathrm{E}\left[w_{it}w_{it}'\right]$ is positive definite uniformly in i by the results of Amemiya (1973).

Remark 3. Mixing when individual effects γ_i are random can be handled as follows. For Example 1 the conditions on z_{it} need to be strengthened to the requirement that z_{it} is a stationary mixing sequence with mixing coefficients α_i such that $\sup_i|\alpha_i(m|\gamma_i)|\leq Ca^m$ for some a such that $0<a<1$ and some $C>0$ conditional on γ_i. A model for z_{it} that satisfies this condition is $z_{it}=\Pi z_{it-1}+\gamma_i+\eta_{it}$, where Π and η_{it} satisfy the conditions discussed above. Example 2 is a special case of the model in de Jong and Woutersen (2011). A condition for their Theorem 1, modified to our example, reads

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press

$$P\left(\tau_0+\gamma_i+z_{it}'\zeta_0+\varepsilon_{it}>0|y_{it-K-1},y_{it-K-2},\ldots,\gamma_i\right)$$
$$-P\left(\gamma_i+z_{it}'\zeta_0+\varepsilon_{it}>0|y_{it-K-1},y_{it-K-2},\ldots,\gamma_i\right)\leq 1-\delta \text{ a.s.}$$

for some $\delta>0$ and K a positive integer where we explicitly condition on γ_i. The above condition is satisfied if z_{it} is uniform mixing, K is chosen large enough, and

$$P\left(\tau_0+\gamma_i+z_{it}'\zeta_0+\varepsilon_{it}>0|\gamma_i\right)-P\left(\gamma_i+z_{it}'\zeta_0+\varepsilon_{it}>0|\gamma_i\right)\leq 1-\delta \text{ a.s.},$$

which holds if $|\gamma_i|\leq B<\infty$ a.s. It then follows from de Jong and Woutersen (2011, Thm. 2) that y_{it} is strong mixing with mixing coefficients $C\left(\upsilon_i(m)+\alpha_i(m)\right)$, where $\upsilon_i(m)\leq(1-\delta)^{m/2K}$.

For Example 3 it follows from Hahn and Kuersteiner (2010b) that γ_i only affects the constant C of the upper bound for $\alpha_i(m)$. The constant C in turn is bounded as long as $\sup_i \mathrm{E}\left[|\gamma_i|^r\right]<\infty$ for r sufficiently large. Clearly, when γ_i has bounded support, this condition holds for any $r<\infty$.

Remark 4. Dependence between observable variables, including covariates, and γ_i arises almost invariably in dynamic models. In Example 1, assuming that $\dim(z_{it})=1$ to simplify exposition, stationarity is achieved for $z_{i0}|\gamma_i\sim N\left(\gamma_i/(1-\Pi),\sigma_\eta^2/\left(1-\Pi^2\right)\right)$ a.s., $\eta_{it}\sim N\left(0,\sigma_\eta^2\right)$, γ_i with bounded support and $\eta_{it},\varepsilon_{it}$, and γ_i mutually independent. It then follows that

$$z_{it}=\frac{1-\Pi^t}{1-\Pi}\gamma_i+\Pi^t z_{i0}+\textstyle\sum_{j=0}^{t-1}\Pi^j\eta_{it-j}.$$

Conditional on γ_i, $z_{it}|\gamma_i\sim N\left(\gamma_i/(1-\Pi),\sigma_\eta^2/\left(1-\Pi^2\right)\right)$, and is strictly stationary with mixing coefficients that do not depend on γ_i. This follows because, conditional on γ_i, z_{it} is a Gaussian process with an autocovariance function that only depends on Π. This implies that $\sup_i|\alpha_i(m|\gamma_i)|\leq Ca^m$. At the same time, $\mathrm{Cov}(\gamma_i,z_{it})=\frac{1-\Pi^t}{1-\Pi}\sigma_\gamma^2+\frac{\Pi^t}{1-\Pi}\sigma_\gamma^2=\frac{1}{1-\Pi}\sigma_\gamma^2$, where $\sigma_\gamma^2=\mathrm{Var}(\gamma_i)$ such that γ_i is correlated with z_{it} despite the fact that it is drawn independently of η_{it}. In Examples 2 and 3 stationary processes with random effects could be constructed in a similar fashion. In both cases γ_i is not independent of x_{it} because it affects the transition probabilities for y_{it}. As the arguments in Remark 3 show, this dependence does not materially affect the bounds for the mixing coefficients.

3. BIAS APPROXIMATION AND CORRECTION

The estimator $\hat{\theta}$ is commonly called the fixed effects estimator. It is well known that, when the number of individuals n is large relative to T, the estimator $\hat{\theta}$ tends to be severely biased. This is the incidental parameters problem, which can be analyzed under asymptotics where n and T grow to infinity at the same rate. Under these asymptotics, the incidental parameters problem manifests itself as the asymptotic bias of the limiting distribution of $\hat{\theta}$.

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press

Under the regularity conditions summarized in Section 2, we obtain the following theorem as a characterization of the approximate bias of the fixed effects estimator.

THEOREM 1. *Assume that Conditions 1–7 hold. Then*

$$\sqrt{nT}\left(\hat{\theta}-\theta_0\right) \to N\left(\beta\sqrt{\kappa}, \mathcal{I}^{-1}\Omega\left(\mathcal{I}'\right)^{-1}\right),$$

where

$$f_i^{VU^\gamma} \equiv \Sigma_{l=-\infty}^{\infty} \mathrm{Cov}\left(V_{it}, U_{it-l}^{\gamma_i}\right), \qquad f_i^{VV} \equiv \Sigma_{l=-\infty}^{\infty} \mathrm{Cov}\left(V_{it}, V_{it-l}\right),$$
$$\varphi^{VU^\gamma} \equiv \lim n^{-1}\Sigma_{i=1}^{n}\left(\mathrm{E}\left[V_{it}^{\gamma_i}\right]\right)^{-1} f_i^{VU^\gamma},$$
$$\varphi^{VV} \equiv \frac{1}{2}\lim n^{-1}\Sigma_{i=1}^{n}\left(\mathrm{E}\left[V_{it}^{\gamma_i}\right]\right)^{-2}\mathrm{E}\left[U_{it}^{\gamma_i\gamma_i}\right] f_i^{VV},$$
$$\Psi \equiv \varphi^{VU^\gamma} - \varphi^{VV}, \qquad \beta \equiv -\mathcal{I}^{-1}\Psi, \qquad \Omega \equiv \lim_n n^{-1}\Sigma_{i=1}^{n}\Sigma_{iT}.$$

Proof. See Appendix, Section E. ■

Our bias-corrected estimator $\hat{\hat{\theta}}$, which removes the asymptotic bias $\beta\sqrt{\kappa}$ in Theorem 1, is computed as follows.

Step 1. Compute the fixed effects estimator $\hat{\theta}, \hat{\gamma}_1, \ldots, \hat{\gamma}_n$.

Step 2. Compute sample analogs $\hat{V}_{it}, \hat{V}_{it}^{\gamma_i}, \hat{U}_{it}, \hat{U}_{it}^{\gamma_i}, \hat{U}_{it}^{\gamma_i\gamma_i}, \hat{U}_{it}^{\theta}$ of $V_{it}, V_{it}^{\gamma_i}, U_{it}, U_{it}^{\gamma_i}, U_{it}^{\gamma_i\gamma_i}, U_{it}^{\theta}$ by evaluating the latter at $\hat{\theta}, \hat{\gamma}_1, \ldots, \hat{\gamma}_n$, e.g., $\hat{V}_{it} \equiv \partial\psi\left(x_{it}; \hat{\theta}, \hat{\gamma}_i\right) \Big/ \partial\gamma_i$.[4]

Step 3. Compute sample analog estimators for $\mathrm{E}\left[V_{it}^{\gamma_i}\right]$, $\mathrm{E}\left[U_{it}^{\gamma_i\gamma_i}\right]$, and $\mathcal{I}_i$ by $\hat{\mathrm{E}}\left[V_{it}^{\gamma_i}\right] \equiv T^{-1}\Sigma_{t=1}^{T}\hat{V}_{it}^{\gamma_i}$, $\hat{\mathrm{E}}\left[U_{it}^{\gamma_i\gamma_i}\right] \equiv T^{-1}\Sigma_{t=1}^{T}\hat{U}_{it}^{\gamma_i\gamma_i}$, and $\hat{\mathcal{I}}_i \equiv -T^{-1}\Sigma_{t=1}^{T}\hat{U}_{it}^{\theta}$.

Step 4. Fix a bandwidth m, and compute $\hat{f}_i^{VU^\gamma} \equiv T^{-1}\Sigma_{l=-m}^{m}\Sigma_{t=\max(1,l)}^{\min(T,T+l)}\hat{V}_{it}\hat{U}_{it-l}^{\gamma_i}$ and $\hat{f}_i^{VV} \equiv T^{-1}\Sigma_{l=-m}^{m}\Sigma_{t=\max(1,l)}^{\min(T,T+l)}\hat{V}_{it}\hat{V}_{it-l}$. (Here, $\hat{f}_i^{VU^\gamma}$ and $\hat{f}_i^{VV}$ are estimators of $f_i^{VU^\gamma}$ and f_i^{VV}, which are cross-spectra and spectra at zero frequency of the processes U_{it}^{γ} and V_{it}.)

Step 5. Let

$$\hat{\beta} \equiv -\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\mathcal{I}}_i\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\hat{f}_i^{VU^\gamma}}{\hat{\mathrm{E}}\left[V_{it}^{\gamma_i}\right]} - \frac{\hat{\mathrm{E}}\left[U_{it}^{\gamma_i\gamma_i}\right]\hat{f}_i^{VV}}{2\left(\hat{\mathrm{E}}\left[V_{it}^{\gamma_i}\right]\right)^2}\right], \tag{3}$$

and compute the bias-corrected estimator $\hat{\hat{\theta}}$ by

$$\hat{\hat{\theta}} \equiv \hat{\theta} - \frac{1}{T}\hat{\beta}.$$

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press

Our proposal is based on the following idea. If a reasonably precise estimator $\hat{\beta}$ of β is available, we would expect that $\hat{\tilde{\theta}} \equiv \hat{\theta} - T^{-1}\hat{\beta}$ would be less biased than the fixed effects estimator $\hat{\theta}$. Theorem 2 below formalizes the bias correction procedure under asymptotics where n and T grow to ∞ at the same rate.

The main result of our paper can then be stated as follows.

THEOREM 2. *Assume that Conditions 1–7 hold. Let* $m, T \to \infty$ *such that* $m/T \to 0$. *Then,* $\sqrt{nT}\left(\hat{\tilde{\theta}} - \theta\right) \to N\left(0, \mathcal{I}^{-1}\Omega\mathcal{I}^{-1}\right)$.

Proof. See Appendix, Section F. ■

4. DISCUSSION

In this section we provide some intuition for Theorems 1 and 2. In order to understand the nature of the bias introduced by the fixed effects estimator $\hat{\gamma}_i$, we first note that the fixed effects estimator for $\hat{\theta}$ is obtained formally by concentrating out the fixed effects γ_i from the criterion function. Letting $\hat{\gamma}_i(\theta) \equiv \operatorname{argmax}_a \sum_{t=1}^{T} \psi(x_{it}; \theta, a)$, we can characterize $\hat{\theta}$ as the solution to the first-order condition $0 = \sum_{i=1}^{n}\sum_{t=1}^{T} U_i\left(x_{it}; \hat{\theta}, \hat{\gamma}_i\left(\hat{\theta}\right)\right)$. The intuition behind the bias of the fixed effects estimator can be gained by considering an infeasible estimator $\tilde{\theta}$ based on $\hat{\gamma}_i(\theta_0)$ rather than $\hat{\gamma}_i\left(\hat{\theta}\right)$, where $\tilde{\theta}$ solves the first-order conditions

$$0 = \sum_{i=1}^{n}\sum_{t=1}^{T} U_i\left(x_{it}; \tilde{\theta}, \hat{\gamma}_i(\theta_0)\right).$$

For $\tilde{\theta}$, standard arguments suggest that $\tilde{\theta} - \theta_0 \approx \left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1} \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} U_i\left(x_{it}; \theta_0, \hat{\gamma}_i(\theta_0)\right)$. Using a second-order Taylor series expansion to approximate $U_i\left(x_{it}; \theta_0, \hat{\gamma}_i(\theta_0)\right)$ around γ_{i0}, we obtain

$$\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T} U_i\left(x_{it}; \theta_0, \hat{\gamma}_i(\theta_0)\right) \approx \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T} U_{it}$$

$$+ \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T} U_{it}^{\gamma_i}\left(\hat{\gamma}_i(\theta_0) - \gamma_{i0}\right) + \frac{1}{2\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T} U_{it}^{\gamma_i\gamma_i}\left(\hat{\gamma}_i(\theta_0) - \gamma_{i0}\right)^2.$$

The first term on the right satisfies a central limit theorem because $\mathrm{E}[U_{it}] = 0$. As for the second and third terms, we note that $\hat{\gamma}_i(\theta_0) - \gamma_{i0} \approx -T^{-1}\sum_{t=1}^{T} V_{it}\left(\mathrm{E}\left[V_{it}^{\gamma_i}\right]\right)^{-1}$, and substituting for $\hat{\gamma}_i(\theta_0) - \gamma_{i0}$ in the approximation for $U_i\left(x_{it}; \theta_0, \hat{\gamma}_i(\theta_0)\right)$ leads to

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press

$$\sqrt{nT}\left(\tilde{\theta}-\theta_0\right) \approx \left(\frac{1}{n}\Sigma_{i=1}^{n}\mathcal{I}_i\right)^{-1}\left(\frac{1}{\sqrt{nT}}\Sigma_{i=1}^{n}\Sigma_{t=1}^{T}U_{it}\right) - \sqrt{\frac{n}{T}}\left(\frac{1}{n}\Sigma_{i=1}^{n}\mathcal{I}_i\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\Sigma_{t=1}^{T}V_{it}}{\sqrt{T}\mathrm{E}\left[V_{it}^{\gamma_i}\right]}\right] \times\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(U_{it}^{\gamma_i}-\frac{\mathrm{E}\left[U_{it}^{\gamma_i\gamma_i}\right]}{2\mathrm{E}\left[V_{it}^{\gamma_i}\right]}V_{it}\right)\right].$$

The preceding discussion indicates the existence of two sources for the asymptotic bias. The correlation between V_{it} and $U_{is}^{\gamma_i}$ is a generalized form of an endogeneity bias. Note that in linear panel models this correlation is zero if the model only contains exogenous regressors. With the more general nonlinear models considered here it may be nonzero even if all the regressors are strictly exogenous. The second term involves the variance and autocovariance of V_{it}. This second term in general is zero in linear models irrespective of whether the regressors are endogenous or exogenous because for linear models $U_{it}^{\gamma_i\gamma_i}=0$.

The expectation of the second term on the right determines the approximate bias of $\tilde{\theta}$. Letting $\beta_n \equiv -\left(n^{-1}\Sigma_{i=1}^{n}\mathcal{I}_i\right)^{-1}\Psi_n$, where

$$\Psi_n \equiv \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\mathrm{E}\left[V_{it}^{\gamma_i}\right]}\mathrm{E}\left[\frac{1}{T}\left(\sum_{t=1}^{T}V_{it}\right)\left(\sum_{t=1}^{T}U_{it}^{\gamma_i}\right)\right] - \frac{\mathrm{E}\left[U_{it}^{\gamma_i\gamma_i}\right]}{2\left(\mathrm{E}\left[V_{it}^{\gamma_i}\right]\right)^2}\mathrm{E}\left[\frac{1}{T}\left(\sum_{t=1}^{T}V_{it}\right)^2\right],$$

we can write

$$\tilde{\theta} \approx \theta_0 + \left(\frac{1}{n}\Sigma_{i=1}^{n}\mathcal{I}_i\right)^{-1}\left(\frac{1}{nT}\Sigma_{i=1}^{n}\Sigma_{t=1}^{T}U_{it}\right) + T^{-1}\beta_n,$$

where $T^{-1}\beta_n$ can be interpreted as an approximation to the bias of $\tilde{\theta}$.

5. MONTE CARLO RESULTS

In order to evaluate the quality of our bias-corrected estimator, we conduct some Monte Carlo experiments. We consider a binary response model with lagged dependent variables. Specifically, we consider a model where

$$y_{i0} = 1\left(\gamma_{i0} + z_{i0}'\zeta_0 + \varepsilon_{i0} \geq 0\right),$$
$$y_{it} = 1\left(\gamma_{i0} + z_{it}'\zeta_0 + y_{i,t-1}\cdot\tau_0 + \varepsilon_{it} \geq 0\right), \qquad t=1,\ldots,T-1; \qquad i=1,\ldots,n.$$

This model received attention from the empirical literature because it can shed light on the nature of time persistence of individual decisions typically observed in many panel data sets. Heckman (1978, 1981a, 1981b) discusses alternative sources of this persistence. Persistence may be due to "state dependence," where

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press

the current period decision is changed because of past decisions, or it may be due to unobserved individual heterogeneity, which impacts the current decision regardless of the past decisions. Here, state dependence is captured by τ_0, and unobserved heterogeneity is captured by the fixed effects γ_{i0}. See Robert and Tybout (1997), Hyslop (1999), or Chay and Hyslop (2000), e.g., for more on the empirical relevance of this model in industrial organization (IO) or labor applications.

In our Monte Carlo study we generate ε_{it} from two different distributions, thereby generating panel probit and logit models. Our $\hat{\theta}$ is the fixed effects MLE. Our Monte Carlo results are based on 1,000 replications.

It is difficult to interpret the Monte Carlo results without a comparison to competing estimators. For the logit model, Honoré and Kyriazidou (2000) proposed an estimator that is consistent under large n, fixed T asymptotics. We therefore compare our estimator for the logit model to the fixed effects MLE and Honoré and Kyriazidou's estimator.[5] We generate the strictly exogenous regressor z_{it} and the fixed effect γ_{i0} such that $z_{it} \sim N\left(0, \left(\pi^2/3\right) \cdot I_{\dim(z_{it})}\right)$ and $\gamma_{i0} = (z_{i0} + z_{i1} + z_{i2} + z_{i3})/4$. When $\dim(z_{it}) > 1$, we chose γ_{i0} to be the sample mean of the first components of $z_{i0}, z_{i1}, z_{i2}, z_{i3}$. The z_{it} are independent over t.

Implementation of our bias correction requires that we choose the bandwidth parameter m. We chose m according to the automatic method proposed by Hahn and Kuersteiner (2006). Details of the implementation of the method are given in Section 6 in the Appendix. For the logit model, the method chose m around 0 for $T = 4$, and around 1 for $T = 8$ and $T = 16$. For the probit model, the method chose m around 1 for $T = 8$, and around 2 for $T = 16$.

The results are summarized in Table 1.[6] The bias-corrected MLE removes about half of the bias and root mean squared error (RMSE). When $\dim(z_{it}) = 1$, our bias-corrected estimator is slightly inferior to Honoré and Kyriazidou's (2000) estimator. On the other hand, when $\dim(z_{it}) = 2$, our bias-corrected estimator strictly dominates Honoré and Kyriazidou's. The latter is known to have a slower rate of convergence when $\dim(z_{it})$ is large, which may explain this result.

For the panel probit model with fixed effects, there does not seem to exist any estimator in the literature that attempts to reduce the bias of the MLE. We therefore limit our analysis to a comparison of the fixed effects MLE with our bias-corrected MLE. We generate the strictly exogenous regressor z_{it} and the fixed effect γ_{i0} such that it is $z_{it} \sim N\left(0, I_{\dim(z_{it})}\right)$ and $\gamma_{i0} = 3 \cdot (z_{i0} + z_{i1} + z_{i2} + z_{i3})/4$. Table 2 summarizes the results for the cases where z_{it} are independent over t. Our bias-corrected estimator reduces about half of the bias of the fixed effects MLE, and the RMSE is reduced by about the same amount. We also considered the case where z_{it} follows an AR(1) process with the first-order serial correlation equal to 1/2. The results are summarized in Table 3, where we find similar results as in Table 2.

Our estimator is motivated by asymptotics where T grows to infinity at the same rate as n. We therefore expect that it may perform poorly when T is very small. Table 4 summarizes the result for $T = 4$. We find that there is no substantial improvement over the fixed effects MLE for this sample size.

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press

TABLE 1. Monte Carlo results for logit models: Benchmark design

A. Benchmark design with dim $(z_{it}) = 1$, $\zeta = 1$, $z_{it} \sim N(0,1)$

			Estimator proposed here				Fixed effects MLE				Honore-Kyriazidou's (2000) estimator			
			Estimate of ζ		Estimate of τ		Estimate of ζ		Estimate of τ		Estimate of ζ		Estimate of τ	
τ	n	T	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
0.5	250	8	0.08	0.11	−0.25	0.29	0.26	0.27	−0.75	0.77	0.02	0.08	−0.07	0.18
	250	16	0.01	0.04	−0.06	0.11	0.10	0.11	−0.31	0.32	0.00	0.04	−0.08	0.13
	500	8	0.07	0.09	−0.24	0.27	0.25	0.26	−0.75	0.76	0.01	0.05	−0.07	0.14
	500	16	0.01	0.03	−0.06	0.09	0.10	0.10	−0.31	0.31	0.00	0.03	−0.08	0.11
1	250	8	0.08	0.11	−0.29	0.33	0.26	0.27	−0.71	0.73	0.03	0.08	−0.14	0.23
	250	16	0.02	0.04	−0.09	0.13	0.10	0.11	−0.30	0.31	0.00	0.04	−0.16	0.19
	500	8	0.08	0.09	−0.29	0.31	0.25	0.26	−0.71	0.72	0.02	0.06	−0.14	0.20
	500	16	0.02	0.03	−0.09	0.11	0.10	0.11	−0.30	0.31	0.00	0.03	−0.16	0.17

Table continues on overleaf

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press

TABLE 1. *continued*

B. Benchmark design with dim $(z_{it}) = 2$, $\zeta_1 = \zeta_2 = 1$, $\zeta_{it} \sim N(0,1)$

			Estimator proposed here				Fixed effects MLE				Honore−Kyriazidou's (2000) estimator			
			Estimate of ζ_1		Estimate of τ		Estimate of ζ_1		Estimate of τ		Estimate of ζ_1		Estimate of τ	
τ	n	T	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
0.5	250	8	0.09	0.13	−0.24	0.30	0.32	0.34	−0.73	0.76	−1.09	1.10	−0.29	0.32
	250	16	0.02	0.05	−0.06	0.12	0.12	0.13	−0.29	0.31	−1.04	1.04	−0.49	0.26
	500	8	0.09	0.11	−0.23	0.26	0.32	0.32	−0.73	0.74	−1.10	1.10	−0.29	0.31
	500	16	0.02	0.04	−0.06	0.09	0.12	0.12	−0.29	0.30	−1.04	1.04	−0.25	0.26
1	250	8	0.09	0.13	−0.26	0.32	0.32	0.34	−0.64	0.67	−1.10	1.10	−0.52	0.54
	250	16	0.02	0.05	−0.07	0.13	0.12	0.13	−0.25	0.28	−1.04	1.04	−0.49	0.49
	500	8	0.09	0.11	−0.25	0.29	0.32	0.33	−0.64	0.66	−1.10	1.10	−0.52	0.53
	500	16	0.02	0.04	−0.07	0.10	0.12	0.13	−0.25	0.27	−1.04	1.04	−0.49	0.49

Note: "Bias" denotes the mean bias of the estimator. "RMSE" denotes the root mean squared error of the estimator. All results are based on 1,000 Monte Carlo runs.

TABLE 2. Monte Carlo results for probit models: Benchmark design

A. Benchmark design with $\dim(z_{it}) = 1, \zeta = 1, z_{it} \sim N(0,1)$

			Estimator proposed here				Fixed effects MLE			
			Estimate of ζ		Estimate of τ		Estimate of ζ		Estimate of τ	
τ	n	T	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
0.5	250	8	0.14	0.17	−0.24	0.27	0.24	0.26	−0.46	0.48
	250	16	0.03	0.06	−0.08	0.11	0.09	0.10	−0.21	0.23
	500	8	0.13	0.15	−0.23	0.25	0.23	0.24	−0.46	0.47
	500	16	0.03	0.04	−0.08	0.09	0.09	0.09	−0.21	0.22
1	250	8	0.14	0.18	−0.31	0.36	0.25	0.28	−0.45	0.47
	250	16	0.04	0.07	−0.11	0.13	0.10	0.12	−0.22	0.24
	500	8	0.13	0.15	−0.32	0.35	0.24	0.26	−0.45	0.46
	500	16	0.04	0.05	−0.10	0.12	0.10	0.11	−0.22	0.23

B. Benchmark design with $\dim(z_{it}) = 2, \zeta_1 = \zeta_2 = 1, z_{it} \sim N(0,1)$

			Estimator proposed here				Fixed effects MLE			
			Estimate of ζ_1		Estimate of τ		Estimate of ζ_1		Estimate of τ	
τ	n	T	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
0.5	250	8	0.21	0.24	−0.20	0.24	0.33	0.35	−0.41	0.44
	250	16	0.05	0.07	−0.07	0.10	0.12	0.13	−0.19	0.20
	500	8	0.21	0.22	−0.20	0.22	0.32	0.33	−0.42	0.43
	500	16	0.04	0.06	−0.07	0.09	0.12	0.12	−0.19	0.19
1	250	8	0.22	0.25	−0.19	0.26	0.34	0.36	−0.34	0.37
	250	16	0.06	0.08	−0.07	0.10	0.13	0.15	−0.16	0.18
	500	8	0.21	0.23	−0.20	0.25	0.33	0.34	−0.34	0.36
	500	16	0.06	0.07	−0.07	0.09	0.13	0.14	−0.16	0.17

C. Benchmark design with $\dim(z_{it}) = 5, \zeta_1 = 1, \zeta_2 = \ldots = \zeta_5 = 0, z_{it} \sim N(0,1)$

			Estimator proposed here				Fixed effects MLE			
			Estimate of ζ_1		Estimate of τ		Estimate of ζ_1		Estimate of τ	
τ	n	T	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
0.5	250	8	0.14	0.17	−0.24	0.27	0.25	0.27	−0.46	0.48
	250	16	0.03	0.06	−0.08	0.11	0.09	0.10	−0.21	0.23
	500	8	0.14	0.15	−0.23	0.25	0.24	0.25	−0.46	0.47
	500	16	0.03	0.04	−0.08	0.09	0.09	0.10	−0.21	0.22
1	250	8	0.15	0.19	−0.31	0.35	0.27	0.29	−0.44	0.46
	250	16	0.04	0.07	−0.10	0.13	0.10	0.12	−0.22	0.23
	500	8	0.14	0.16	−0.32	0.36	0.25	0.26	−0.45	0.46
	500	16	0.04	0.06	−0.10	0.11	0.10	0.11	−0.22	0.23

Note: "Bias" denotes the mean bias of the estimator. "RMSE" denotes the root mean squared error of the estimator. All results are based on 1,000 Monte Carlo runs.

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press

TABLE 3. Monte Carlo results for probit models: Serially correlated

A. Serially Correlated z_{it} with $\dim(z_{it}) = 1, \zeta = 1, z_{it} \sim N(0,1)$

			Estimator proposed here				Fixed effects MLE			
			Estimate of ζ		Estimate of τ		Estimate of ζ		Estimate of τ	
τ	n	T	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
0.5	250	8	0.16	0.21	−0.28	0.32	0.29	0.32	−0.51	0.53
	250	16	0.06	0.08	−0.11	0.14	0.13	0.15	−0.26	0.27
	500	8	0.16	0.18	−0.27	0.29	0.28	0.29	−0.51	0.52
	500	16	0.06	0.07	−0.11	0.13	0.13	0.14	−0.25	0.26
1	250	8	0.19	0.24	−0.38	0.42	0.32	0.36	−0.50	0.53
	250	16	0.07	0.10	−0.15	0.18	0.15	0.17	−0.27	0.28
	500	8	0.17	0.20	−0.40	0.43	0.30	0.32	−0.51	0.52
	500	16	0.07	0.09	−0.14	0.16	0.15	0.16	−0.27	0.28

B. Serially Correlated z_{it} with $\dim(z_{it}) = 2, \zeta_1 = \zeta_2 = 1, z_{it} \sim N(0,1)$

			Estimator proposed here				Fixed effects MLE			
			Estimate of ζ_1		Estimate of τ		Estimate of ζ_1		Estimate of τ	
τ	n	T	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
0.5	250	8	0.23	0.27	−0.26	0.31	0.37	0.40	−0.48	0.51
	250	16	0.08	0.10	−0.11	0.14	0.16	0.17	−0.24	0.25
	500	8	0.22	0.24	−0.27	0.30	0.35	0.37	−0.48	0.50
	500	16	0.07	0.09	−0.11	0.13	0.16	0.16	−0.24	0.25
1	250	8	0.26	0.30	−0.28	0.34	0.39	0.43	−0.41	0.44
	250	16	0.09	0.12	−0.11	0.14	0.18	0.19	−0.21	0.23
	500	8	0.24	0.27	−0.30	0.34	0.38	0.40	−0.42	0.44
	500	16	0.09	0.10	−0.11	0.13	0.17	0.18	−0.21	0.22

C. Serially Correlated z_{it} with $\dim(z_{it}) = 5, \zeta_1 = 1, \zeta_2 = \ldots = \zeta_5 = 0, z_{it} \sim N(0,1)$

			Estimator proposed here				Fixed effects MLE			
			Estimate of ζ_1		Estimate of τ		Estimate of ζ_1		Estimate of τ	
τ	n	T	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
0.5	250	8	0.18	0.22	−0.30	0.34	0.31	0.34	−0.52	0.54
	250	16	0.06	0.09	−0.12	0.15	0.14	0.15	−0.26	0.27
	500	8	0.17	0.19	−0.28	0.31	0.29	0.31	−0.51	0.52
	500	16	0.06	0.07	−0.11	0.13	0.13	0.14	−0.26	0.26
1	250	8	0.21	0.26	−0.37	0.42	0.35	0.38	−0.50	0.53
	250	16	0.08	0.10	−0.15	0.18	0.16	0.17	−0.27	0.28
	500	8	0.19	0.21	−0.39	0.43	0.32	0.34	−0.51	0.52
	500	16	0.07	0.09	−0.14	0.16	0.15	0.16	−0.27	0.28

Note: "Bias" denotes the mean bias of the estimator. "RMSE" denotes the root mean squared error of the estimator. All results are based on 1,000 Monte Carlo runs.

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press

TABLE 4. Monte Carlo results for logit models: Benchmark design with $\dim(z_{it}) = 1, \zeta = 1\ z_{it} \sim N(0, \pi^2/3)$

			Estimator proposed here				Fixed effects MLE				Honore-Kyriazidou's (2000) estimator			
			Estimate of ζ		Estimate of τ		Estimate of ζ		Estimate of τ		Estimate of ζ		Estimate of τ	
τ	n	T	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
0.5	250	4	−1.00	3.12	−0.25	3.41	0.83	0.90	−2.61	2.68	0.12	0.29	−0.04	0.59
	500	4	−0.97	1.72	−0.29	1.88	0.80	0.83	−2.57	2.60	0.07	0.18	−0.04	0.37
1	250	4	−0.85	2.29	−0.88	2.02	0.83	0.90	−2.41	2.48	0.18	0.38	−0.05	0.65
	500	4	−0.85	1.55	−0.88	1.61	0.79	0.83	−2.38	2.41	0.11	0.23	−0.09	0.42

Note: "Bias" denotes the mean bias of the estimator. "RMSE" denotes the root mean squared error of the estimator. All results are based on 1,000 Monte Carlo runs.

6. SUMMARY

This paper develops an asymptotic expansion for nonlinear dynamic panel data models with fixed effects. The conditions needed involve mixing assumptions and bounds on higher-order moments. The results are useful to develop bias-corrected fixed effects estimators for nonlinear dynamic panel models and have been applied in this paper and other related work. We expect these expansions to be useful in other contexts, such as establishing the properties of bootstrap-based procedures. Questions for future research include the treatment of time effects as well as the effects of cross-sectional dependence.

NOTES

1. Bias correction for linear dynamic panel models was also considered by Moon and Phillips (2004), Phillips and Sul (2003), and Han, Phillips, and Sul (2010).

2. After a first draft of this paper was circulated, a number of papers extended our approach in various directions. Bester and Hansen (2009) develop similar corrections to the profile likelihood function instead of correcting the estimator directly. Lee, Yu, and de Jong (2008) extend bias corrections similar to the ones developed here to the estimation of linear spatial models. Lee (2007b) considers nonparameteric estimators for nonlinear dynamic panel models with fixed effects, and Lee (2007a) investigates the effects of dynamic misspecification on estimator bias in linear dynamic panel models with fixed effects.

3. When the model contains lagged endogenous regressors and ψ is a likelihood function, then it is the likelihood function conditional on the lagged endogenous regressors. The estimators $\hat{\theta}, \hat{\gamma}_1, \ldots, \hat{\gamma}_n$ then are the MLEs conditional on y_{i0}.

4. See Appendix Section A for precise definitions of $\hat{V}_{it}^{\gamma_i}$, $\hat{U}_{it}$, $\hat{U}_{it}^{\gamma_i}$, $\hat{U}_{it}^{\gamma_i \gamma_i}$, and $\hat{U}_{it}^{\theta}$.

5. The method of Honoré and Kyriazidou (2000) depends on a bandwidth choice. In our tables we report results obtained with their preferred choice, $c = 8$ (their notation). Especially for the case $\dim(z_{it}) = 2$, which is not reported in their paper, we experimented with c ranging from 2 to 128 but find no significant effect on the performance of the estimator.

6. To conserve space, we only report results for the first component of ζ when $\dim(z_{it}) > 1$.

REFERENCES

Amemiya, T. (1973) Regression analysis when the dependent variable is truncated normal. *Econometrica* 41, 997–1016.

Andrews, D.W.K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.

Bester, A.C & C. Hansen (2009) A penalty function approach to bias reduction in nonlinear panel models with fixed effects. *Journal of Business and Economic Statistics* 27, 131–148.

Chay, K.Y. & D.R. Hyslop (2000) Identification and Estimation of Dynamic Binary Response Panel Data Models: Empirical Evidence Using Alternative Approaches. Unpublished manuscript.

Chintagunta, P., E. Kyriazidou, & J. Perktold (2001) Panel data analysis of household brand choices. *Journal of Econometrics* 103, 111–153.

De Jong, R. & T. Woutersen (2011) Dynamic time series binary choice. *Econometric Theory* 27, 623–702.

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.

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press

Hahn, J. & G. Kuersteiner (2006) Bandwidth Choice for Bias Estimators in Dynamic Nonlinear Panel Models. Unpublished manuscript.

Hahn, J. & G. Kuersteiner (2010a) Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects: Supplementary Appendix. http://www9.georgetown.edu/faculty/gk232/research/Hahn-Kuersteiner-Supp-Appendix2.pdf.

Hahn, J. & G. Kuersteiner (2010b) Stationarity and mixing properties of the dynamic Tobit model. *Economics Letters* 107, 105–111.

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

Hall, P. & C. Heyde (1980) *Martingale Limit Theory and Its Application*. Academic Press.

Hall, P. & J. Horowitz (1996) Bootstrap critical values for tests based on generalized-method-of-moments estimators. *Econometrica* 64, 891–916.

Han, C., P.C.B. Phillips, & D. Sul (2010) X-Differencing and Dynamic Panel Model Estimation. Cowles Foundation Discussion Paper No. 1747.

Hansen, B.E. (1991) Strong laws for dependent heterogenous processes. *Econometric Theory* 7, 213–221.

Heckman, J.J. (1978) Simple statistical models for discrete panel data developed and applied to test the hypothesis of true state dependence against the hypothesis of spurious state dependence. *Annales de l'INSEE,* 227–269.

Heckman, J.J. (1981a) Heterogeneity and state dependence. In S. Rosen (ed.), *Studies in Labor Markets*. University of Chicago Press.

Heckman, J.J. (1981b) Statistical models for discrete panel data. In C. Manski & D. McFadden (eds.), *Structural Analysis of Discrete Data*. MIT Press.

Heckman, J.J. & T.E. MaCurdy (1980) A life cycle model of female labour supply. *Review of Economic Studies* 47, 47–74.

Hendel, I. & A. Nevo (2006) Sales and consumer inventory. *RAND Journal of Economics* 37, 543–561.

Honoré, B.E. (1992) Trimmed LAD and least squares estimation of truncated and censored regression models with fixed effects. *Econometrica* 60, 533–565.

Honoré, B. & E. Kyriazidou (2000) Panel data discrete choice models with lagged dependent variables. *Econometrica* 68, 839–874.

Hyslop, D.R. (1999) State dependence, serial correlation, and heterogeneity in intertemporal labor force participation of married women. *Econometrica* 67, 1255–1294.

Lee, L., J. Yu, & R. de Jong (2008) Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. *Journal of Econometrics* 146, 118–134.

Lee, Y. (2007a) Bias Correction in Dynamic Panels under Time Series Misspecification. Unpublished manuscript.

Lee, Y. (2007b) Nonparametric Estimation of Dynamic Panel Models with Fixed Effects. Unpublished manuscript.

Magnus, J. & H. Neudecker (1988) *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley.

McLeish, D.L. (1975) A maximal inequality and dependent strong laws. *Annals of Probability* 3, 829–839.

Moon, H.R. & P.C.B Phillips (2004) GMM estimation of autoregressive roots near unity with panel data. *Econometrica* 72, 467–522.

Newey, W.K. & D. McFadden (1994) Large sample estimation and hypothesis testing. In R.F. Engle & D.L. McFadden (eds.), *The Handbook of Econometrics*, vol. IV, pp. 2111–2245. Elsevier.

Neyman, J. & E. Scott (1948) Consistent estimates based on partially consistent observations. *Econometrica* 16, 1–31.

Olsen, R.J. (1978) Note on the uniqueness of the maximum likelihood estimator for the Tobit model. *Econometrica* 46, 1211–1215.

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press

Phillips, P.C.B. & D. Sul (2003) Dynamic panel estimation and homogeneity testing under cross section dependence. *Econometrics Journal* 6, 217–259.

Robert, M.J. & J.R. Tybout (1997) The decision to export in Colombia: An empirical model of entry with sunk costs. *American Economic Review* 87, 545–564.

Van der Vaart, A.W. & J.A. Wellner (1996) *Weak Convergence and Empirical Processes*. Springer Verlag.

APPENDIX

Throughout this Appendix, we will let $F \equiv (F_1, \ldots, F_n)$ denote the collection of (marginal) distribution functions of x_{it} and $\hat{F} \equiv \left(\hat{F}_1, \ldots, \hat{F}_n\right)$, where $\hat{F}_i$ denotes the empirical distribution function for the ith observation. Define $F(\epsilon) \equiv F + \epsilon\sqrt{T}\left(\hat{F} - F\right)$ for $\epsilon \in \left[0, T^{-1/2}\right]$, and $\Delta_{iT} \equiv \sqrt{T}\left(\hat{F}_i - F_i\right)$.

A. Sample Analogs.

$$\hat{V}_{it}^{\gamma_i} \equiv \partial^2\psi\left(x_{it};\hat{\theta},\hat{\gamma}_i\right) \Big/ \partial\gamma_i^2,$$

$$\hat{\rho}_i \equiv \left(T^{-1}\sum_{t=1}^{T}\frac{\partial^2\psi\left(x_{it};\hat{\theta},\hat{\gamma}_i\right)}{\partial\gamma_i^2}\right)^{-1}\left(T^{-1}\sum_{t=1}^{T}\frac{\partial^2\psi\left(x_{it};\hat{\theta},\hat{\gamma}_i\right)}{\partial\theta\partial\gamma_i}\right),$$

$$\hat{U}_{it} \equiv \frac{\partial\psi\left(x_{it};\hat{\theta},\hat{\gamma}_i\right)}{\partial\theta} - \hat{\rho}_i \cdot \frac{\partial\psi\left(x_{it};\hat{\theta},\hat{\gamma}_i\right)}{\partial\gamma_i},$$

$$\hat{U}_{it}^{\gamma_i} \equiv \frac{\partial^2\psi\left(x_{it};\hat{\theta},\hat{\gamma}_i\right)}{\partial\gamma_i\partial\theta} - \hat{\rho}_i \cdot \frac{\partial^2\psi\left(x_{it};\hat{\theta},\hat{\gamma}_i\right)}{\partial\gamma_i^2},$$

$$\hat{U}_{it}^{\gamma_i\gamma_i} \equiv \frac{\partial^3\psi\left(x_{it};\hat{\theta},\hat{\gamma}_i\right)}{\partial\gamma_i^2\partial\theta} - \hat{\rho}_i \cdot \frac{\partial^3\psi\left(x_{it};\hat{\theta},\hat{\gamma}_i\right)}{\partial\gamma_i^3},$$

$$\hat{U}_{it}^{\theta} \equiv \frac{\partial^2\psi\left(x_{it};\hat{\theta},\hat{\gamma}_i\right)}{\partial\theta\partial\theta'} - \hat{\rho}_i \cdot \frac{\partial^2\psi\left(x_{it};\hat{\theta},\hat{\gamma}_i\right)}{\partial\gamma_i\partial\theta'}.$$

B. Some Lemmas.

LEMMA 1. *Suppose that, for each i, $\{\xi_{it}, t = 1, 2, \ldots\}$ is a mixing sequence with $\mathrm{E}\left[\xi_{it}\right] = 0$ for all i, t. Let $\mathcal{A}_t^i = \sigma\left(\xi_{it}, \xi_{it-1}, \xi_{it-2}, \ldots\right)$ and $\mathcal{B}_t^i = \sigma\left(\xi_{it}, \xi_{it+1}, \xi_{it+2}, \ldots\right)$. Let*

$$\alpha_i(m) = \sup_t \sup_{A\in\mathcal{A}_t^i, B\in\mathcal{B}_{t+m}^i} \left|P(A\cap B) - P(A)P(B)\right|.$$

Assume that $\sup_i |\alpha_i(m)| \leq Ca^m$ for some a such that $0 < a < 1$ and some $0 < C < \infty$. We assume that $\{\xi_{it}, t = 1, 2, 3, \ldots\}$ are independent across i. We also assume that $n = O(T)$. Finally, assume that $\mathrm{E}\left[\|\xi_{it}\|^{4+\delta}\right] < \infty$ for some $\delta > 0$. We then have

$$\mathrm{P}\left[\max_{1\leq i\leq n}\left\|\frac{1}{T}\Sigma_{t=1}^{T}\xi_{it}\right\| > \eta\right] = o\left(T^{-1}\right)$$

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press

for every $\eta > 0$. *Now assume that* $\mathrm{E}\left[\|\xi_{it}\|^{(10+10q)/(1-10\upsilon)+\delta}\right] < \infty$ *for some* $\delta > 0$, $0 < \upsilon < 1/10$ *and some integer* $q \geq 1$. *Then,*

$$\mathrm{P}\left[\max_{1\leq i\leq n}\left\|\frac{1}{\sqrt{T}}\Sigma_{t=1}^{T}\xi_{it}\right\| > \eta T^{1/10-\upsilon}\right] = o\left(T^{-q}\right)$$

for every $\eta > 0$.

Proof. By the Markov inequality,

$$\begin{aligned}\mathrm{P}\left[\max_{1\leq i\leq n}\left\|\frac{1}{T}\Sigma_{t=1}^{T}\xi_{it}\right\| > \eta\right] &= \mathrm{P}\left[\max_{1\leq i\leq n}\left\|\Sigma_{t=1}^{T}\xi_{it}\right\|^{4+\delta/2} > \eta^{4+\delta/2}T^{4+\delta/2}\right]\\ &\leq T^{-(4+\delta/2)}\eta^{-(4+\delta/2)}\mathrm{E}\left[\max_{1\leq i\leq n}\left\|\Sigma_{t=1}^{T}\xi_{it}\right\|^{4+\delta/2}\right],\end{aligned}$$

and by an inequality for the maximum of random variables (van der Vaart and Wellner, 1996, p. 96) one obtains

$$\mathrm{E}\left[\max_{1\leq i\leq n}\left\|\Sigma_{t=1}^{T}\xi_{it}\right\|^{4+\delta/2}\right] \leq n\max_{i}\mathrm{E}\left[\left\|\Sigma_{t=1}^{T}\xi_{it}\right\|^{4+\delta/2}\right].$$

By Lemma 2.1 of McLeish (1975) it follows that ξ_{it} satisfies Assumptions 1 and 2 in Hansen (1991). From Corollary 3 of Hansen it follows that

$$\mathrm{E}\left[\left\|\Sigma_{t=1}^{T}\xi_{it}\right\|^{4+\delta/2}\right] \leq C\left(\mathrm{E}\left[\|\xi_{it}\|^{4+\delta}\right]\right)^{(8+\delta)/(4+\delta)}T^{2+\delta/4}.$$

Then, for a constant C not necessarily the same as before,

$$\begin{aligned}\mathrm{P}\left[\max_{1\leq i\leq n}\left\|\frac{1}{T}\Sigma_{t=1}^{T}\xi_{it}\right\| > \eta\right] &\leq nT^{-(4+\delta/2)}\eta^{-4+\delta/2}CT^{2+\delta/4}\\ &= O(T^{-(1+\delta/4)}) = o\left(T^{-1}\right).\end{aligned}$$

For the second part of Lemma 1, note that by previous arguments,

$$\begin{aligned}&T^{q}\,\mathrm{P}\left[\max_{1\leq i\leq n}\left\|\frac{1}{\sqrt{T}}\Sigma_{t=1}^{T}\xi_{it}\right\| > \eta T^{1/10-\upsilon}\right]\\ &\quad= T^{q}\,\mathrm{P}\left[\max_{1\leq i\leq n}\left\|\Sigma_{t=1}^{T}\xi_{it}\right\| > \eta T^{3/5-\upsilon}\right]\\ &\quad= T^{q}\,\mathrm{P}\left[\max_{1\leq i\leq n}\left\|\Sigma_{t=1}^{T}\xi_{it}\right\|^{(10+10q)/(1-10\upsilon)+\delta/2}\right.\\ &\qquad\left.> \eta^{(10+10q)/(1-10\upsilon)+\delta/2}T^{(3/5-\upsilon)((10+10q)/(1-10\upsilon)+\delta/2)}\right]\\ &\quad\leq T^{q}T^{-(3/5-\upsilon)((10+10q)/(1-10\upsilon)+\delta/2)}\eta^{-(10+10q)/(1-10\upsilon)+\delta/2}\\ &\qquad\times\mathrm{E}\left[\max_{1\leq i\leq n}\left\|\Sigma_{t=1}^{T}\xi_{it}\right\|^{(10+10q)/(1-10\upsilon)+\delta/2}\right]\\ &\quad= O\left[T^{q}T^{(-6q-6+10\upsilon q+10\upsilon)/(1-10\upsilon)-(3/5-\upsilon)\delta/2}n\cdot CT^{(5q+5)/(1-10\upsilon)+\delta/4}\right]\\ &\quad= O\left(T^{-(1-10\upsilon)\delta/20}\right) = o(1)\end{aligned}$$

for $\upsilon > 0$ sufficiently small. ■

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press

LEMMA 2. *Let $\xi(x_{it},\phi)$ be a function indexed by the parameter $\phi \in \Phi$, where Φ is a convex subset of $\mathbb{R}^p$ with $\mathrm{E}\left[\xi(x_{it},\phi)\right]=0$ for all i,t and $\phi \in \Phi$. Assume that there exists a function $\mathbf{M}(x_{it})$ such that $\|\xi(x_{it},\phi_1)-\xi(x_{it},\phi_2)\| \le \mathbf{M}(x_{it})\|\phi_1-\phi_2\|$ for all $\phi_1,\phi_2 \in \Phi$ and $\sup_\phi \|\xi(x_{it},\phi)\| \le \mathbf{M}(x_{it})$. For each i, let x_{it} be an α-mixing process with exponentially decaying mixing coefficients $\underline{\alpha}_i(m)$ satisfying $\sup_i \left|\underline{\alpha}_i(m)\right| \le Ca^m$ for some a such that $0<a<1$ and some $0<C<\infty$. Let q denote a positive integer such that $q \ge \frac{p+4}{2}$, where $p=\dim\phi$. We also assume that $\mathrm{E}\left[|\mathbf{M}(x_{it})|^{(10+10q)/(1-10\upsilon)+\delta}\right]<\infty$ for some $\delta>0$. Finally, assume that $n=O(T)$. We then have $\mathrm{P}\left[\max_i \left\|\frac{1}{\sqrt{T}}\sum_{t=1}^T \xi(x_{it},\phi_i)\right\| > T^{1/10-\upsilon}\right]=o\left(T^{-1}\right)$ for $0<\upsilon<1/10$. Here, $\{\phi_i\}$ is an arbitrary nonstochastic sequence in Φ.*

Proof. Note that we have

$$T\,\mathrm{P}\left[\max_i \left\|\frac{1}{\sqrt{T}}\sum_{t=1}^T \xi(x_{it},\phi_i)\right\| > T^{1/10-\upsilon}\right] \le T\sum_{i=1}^n \mathrm{P}\left[\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^T \xi(x_{it},\phi_i)\right\| > T^{1/10-\upsilon}\right].$$

Adapting an argument in Hall and Horowitz (1996) we choose $\varepsilon>0$ and divide Φ into subsets $\Phi_1,\Phi_2,\ldots\Phi_{\vartheta(\varepsilon)}$ such that $\|\phi_1-\phi_2\|<\frac{\varepsilon}{\sqrt{T}}$ whenever ϕ_1,ϕ_2 are in the same subset Φ_i. Then

$$\mathrm{P}\left[\sup_{\phi\in\Phi}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^T \xi(x_{it},\phi)\right\| > T^{1/10-\upsilon}\right] \le \sum_{j=1}^{\vartheta(\varepsilon)} \mathrm{P}\left[\sup_{\phi\in\Phi_j}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^T \xi(x_{it},\phi)\right\| > T^{1/10-\upsilon}\right],$$

and for some $\phi_j \in \Phi_j$ and any $\phi \in \Phi_j$,

$$\begin{aligned}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^T \xi(x_{it},\phi)\right\| &\le \left\|\frac{1}{\sqrt{T}}\sum_{t=1}^T \xi\left(x_{it},\phi_j\right)\right\| + \frac{1}{\sqrt{T}}\sum_{t=1}^T \left\|\xi(x_{it},\phi)-\xi\left(x_{it},\phi_j\right)\right\| \\ &\le \left\|\frac{1}{\sqrt{T}}\sum_{t=1}^T \xi\left(x_{it},\phi_j\right)\right\| + \left|\frac{\varepsilon}{T}\sum_{t=1}^T \left(M(x_{it})-\mathrm{E}\left[M(x_{it})\right]\right)\right| \\ &\quad + 2\varepsilon\mathrm{E}\left[M(x_{it})\right]\end{aligned}$$

such that

$$\begin{aligned}&\mathrm{P}\left[\sup_{\phi\in\Phi_j}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^T \xi(x_{it},\phi)\right\| > T^{1/10-\upsilon}\right] \\ &\quad\le \mathrm{P}\left[\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^T \xi\left(x_{it},\phi_j\right)\right\| > \frac{T^{1/10-\upsilon}}{3}\right] \\ &\qquad + \mathrm{P}\left[\frac{1}{T}\left|\sum_{t=1}^T \left(M(x_{it})-\mathrm{E}\left[M(x_{it})\right]\right)\right| > \frac{T^{1/10-\upsilon}}{3}\right].\end{aligned}$$

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press

By Lemma 1, it follows that both terms on the right-hand side are of order $o\left(T^{-q}\right)$, where the orders are uniform in i. Since $\vartheta(\varepsilon) = O\left(T^{p/2}\right)$ it follows that

$$T\,\mathrm{P}\left[\max_i \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \xi(x_{it}, \phi_i) \right\| > T^{1/10-v}\right] = nT \cdot o\left(T^{-q}\right) \cdot O\left(T^{p/2}\right)$$
$$= o\left(T^{2-q+p/2}\right) = o(1). \qquad \blacksquare$$

LEMMA 3. *Assume that x_{it} satisfies Condition 3, and let $\xi(x_{it}, \phi)$ be a function indexed by the parameter $\phi \in \operatorname{int}\Phi$, where Φ is a convex subset of $\mathbb{R}^p$. For any sequence $\phi_i \in \operatorname{int}\Phi$, assume* $\mathrm{E}\left[\xi(x_{it}, \phi_i)\right] = 0$. *Further assume that* $\sup_\phi \|\xi(x_{it}, \phi)\| \le \mathbf{M}(x_{it})$ *for some* $\mathbf{M}(x_{it})$ *such that* $\mathrm{E}\left[\mathbf{M}(x_{it})^4\right] < \infty$. *Let* $\Sigma_{nT} = \sum_{i=1}^{n} \Sigma_{iT}^{\xi\xi}$ *with* $\Sigma_{iT}^{\xi\xi} = \mathrm{Var}\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \xi(x_{it}, \phi_i)\right)$. *Denote the smallest eigenvalue of* $\Sigma_{iT}^{\xi\xi}$ *by* λ_{iT}^{ξ}, *and assume that* $\inf_i \inf_T \lambda_{iT}^{\xi} > 0$. *Then,* $\frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} \xi(x_{it}, \phi_i) \xrightarrow{d} N\left(0, f^{\xi\xi}\right)$, *and* $\sup_i \left\| \Sigma_{iT}^{\xi\xi} - f_i^{\xi\xi} \right\| \to 0$, *where* $f^{\xi\xi} \equiv \lim n^{-1} \sum_{i=1}^{n} f_i^{\xi\xi}$ *and* $f_i^{\xi\xi} \equiv \sum_{j=-\infty}^{\infty} \mathrm{E}\left[\xi(x_{it}, \phi_i) \xi\left(x_{it-j}, \phi_i\right)'\right]$.

Proof. First note that by Condition 3 and the fact that the mixing property is preserved by measurable transformations of finitely many elements x_{it}, $\xi(x_{it}, \phi_i)$ is mixing with exponentially decaying mixing coefficients. Let $p' \equiv \dim(\xi)$. By the Cramer-Wold theorem it is enough to consider $v_{i,n,T} = \ell' \Sigma_{nT}^{-1/2} Z_{iT} = \ell' \Sigma_{nT}^{-1/2} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \xi(x_{it}, \phi_i)$ for all $\ell \in \mathbb{R}^{p'}, \|\ell\| = 1$ where $\Sigma_{nT} = \sum_{i=1}^{n} \Sigma_{iT}^{\xi\xi}$ with $\Sigma_{iT}^{\xi\xi} = \mathrm{Var}\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \xi(x_{it}, \phi_i)\right)$. The Lindeberg-Feller condition requires that for any $\varepsilon > 0$,

$$\sum_{i=1}^{n} \mathrm{E}\left[v_{i,n,T}^2 \mathbf{1}\left\{\left|v_{i,n,T}\right| > \varepsilon\right\}\right] \to 0.$$

In order to verify this condition, let $\xi_{it} \equiv \xi(x_{it}, \phi_i)$, and note that

$$\mathrm{E}\left[v_{i,n,T}^2 \mathbf{1}\left\{\left|v_{i,n,T}\right| > \varepsilon\right\}\right] = \mathrm{E}\left[\left\|v_{i,n,T}\right\|^2 \mathbf{1}\left\{\left|v_{i,n,T}\right| > \varepsilon\right\}\right]$$
$$\le \left\|\ell' \Sigma_{nT}^{-1} \ell\right\|^2 \mathrm{E}\left[\|Z_{iT}\|^2 \mathbf{1}\left\{\left|v_{i,n,T}\right| > \varepsilon\right\}\right]$$
$$\le \varepsilon^{-2} \left\|\ell' \Sigma_{nT}^{-1} \ell\right\|^4 \mathrm{E}\left[\|Z_{iT}\|^4\right]$$
$$= \varepsilon^{-2} \left\|\ell' \Sigma_{nT}^{-1} \ell\right\|^4 T^{-2} \sum_{t_1,..,t_4} \mathrm{E}\left[\xi_{it_1}' \xi_{it_2} \xi_{it_3}' \xi_{it_4}\right],$$

where $\mathbf{1}\left\{\left|v_{i,n,T}\right| > \varepsilon\right\} \le \mathbf{1}\left\{\|Z_{iT}\|^2 > \varepsilon^2 / \left\|\ell' \Sigma_{nT}^{-1} \ell\right\|^2\right\}$. Observe that (i) $\ell' \Sigma_{nT}^{-1} \ell \le \max_k \lambda_k^{-1} = 1/\min_k \lambda_k$ with λ_k eigenvalues of Σ_{nT}; and (ii) $\min_k \lambda_k \ge n \inf_i \inf_T \lambda_{iT}^{\xi}$ by Magnus and Neudecker (1988, p. 204). It follows that $\left\|\ell' \Sigma_{nT}^{-1} \ell\right\|^4 \le c_2 n^{-2}$ for some constant $c_2 < \infty$.

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press

We now take care of $\mathrm{E}\left[\xi'_{it_1}\xi_{it_2}\xi'_{it_3}\xi_{it_4}\right]$. For this purpose, let the jth element of ξ_{it} be denoted by $\xi_{j,it}$. We then have

$$\begin{aligned}\mathrm{E}\left[\xi'_{it_1}\xi_{it_2}\xi'_{it_3}\xi_{it_4}\right] &= \sum_{j_1,j_2}^{p'} \mathrm{E}\left[\xi_{j_1,it_1}\xi_{j_1,it_2}\xi_{j_2,it_3}\xi_{j_2,it_4}\right] \\ &= \sum_{j_1,j_2}^{p'} \left\{\mathrm{E}\left[\xi_{j_1,it_1}\xi_{j_1,it_2}\right]\mathrm{E}\left[\xi_{j_2,it_3}\xi_{j_2,it_4}\right]\right. \\ &\quad + \mathrm{E}\left[\xi_{j_2,it_3}\xi_{j_1,it_2}\right]\mathrm{E}\left[\xi_{j_1,it_1}\xi_{j_2,it_4}\right] + \mathrm{E}\left[\xi_{j_2,it_3}\xi_{j_1,it_1}\right]\mathrm{E}\left[\xi_{j_1,it_2}\xi_{j_2,it_4}\right] \\ &\quad \left. + \mathrm{Cum}\left(\xi_{j_1,it_1},\xi_{j_1,it_2},\xi_{j_2,it_3},\xi_{j_2,it_4}\right)\right\},\end{aligned}$$

where $\mathrm{Cum}\left(\xi_{j_1,it_1},\xi_{j_1,it_2},\xi_{j_2,it_3},\xi_{j_2,it_4}\right)$ is a fourth-order cumulant term. The terms on the right-hand side can be bounded by making the following observations. First, from Andrews (1991, Lem. 1), it follows that $\sum_{t_2,..,t_4}\sup_{t_1}\mathrm{Cum}\left(\xi_{j_1,it_1},\xi_{j_1,it_2},\xi_{j_2,it_3},\xi_{j_2,it_4}\right) < \infty$. Second, from Hall and Heyde (1980, Cor. A.2), we have $\mathrm{E}\left[\xi_{j_1,it_1}\xi_{j_1,it_2}\right] \le 8\,\mathrm{E}\left[\left|\xi_{j_1,it_1}\right|^4\right]$ $\mathrm{E}\left[\left|\xi_{j_1,it_2}\right|^4\right]\alpha\left(t_1-t_2\right)^{1/2}$ such that $T^{-1}\sum_{t_1,t_2}\mathrm{E}\left[\xi_{j_1,it_1}\xi_{j_1,it_2}\right] \le c_1\sum_{l=0}^{T}(1-l/T)$ $\left(\sqrt{a}\right)^l < \infty$ for all T and some constant $c_1 < \infty$.

These arguments show that

$$\mathrm{E}\left[v_{i,n,T}^2\mathbf{1}\left\{\left|v_{i,n,T}\right| > \varepsilon\right\}\right] \le cn^{-2}$$

uniformly in i, T for some constant $c < \infty$. Therefore, the Lindeberg-Feller condition is satisfied. The result then follows from the fact that

$$\begin{aligned}\sup_i\left\|\Sigma_{iT}^{\xi\xi} - f_i^{\xi\xi}\right\| &\le \frac{1}{T}\sum_{l=-T}^{T}|l|\sup_i\left\|\mathrm{Cov}\left(\xi_{it},\xi'_{it-l}\right)\right\| + \sum_{|l|\ge T+1}\sup_i\left\|\mathrm{Cov}\left(\xi_{it},\xi'_{it-l}\right)\right\| \\ &\le \frac{1}{T}\sum_{l=-\infty}^{\infty}|l|\sup_i\left\|\mathrm{Cov}\left(\xi_{it},\xi'_{it-l}\right)\right\| + c_1\sum_{|l|\ge T+1}^{\infty}\left(\sqrt{a}\right)^l \\ &\to 0 \text{ as } T\to\infty,\end{aligned}$$

such that iterated and joint limits are the same and $\Sigma_{nT} \to f^{\xi\xi}$. ■

C. Consistency. Recall that $G_{(i)}(\theta,\gamma) = \mathrm{E}\left[\psi\left(x_{it};\theta,\gamma\right)\right]$, and define $\hat{G}_{(i)}(\theta,\gamma) \equiv T^{-1}\sum_{t=1}^{T}\psi\left(x_{it};\theta,\gamma\right)$.

LEMMA 4. *For all $\eta > 0$ it follows that*

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

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

$$\begin{aligned}&\mathrm{P}\left[\max_{1\le i\le n}\sup_{(\theta,\gamma)}\left|\hat{G}_{(i)}(\theta,\gamma) - G_{(i)}(\theta,\gamma)\right| \ge \eta\right] \\ &\quad\le \sum_{i=1}^{n}\mathrm{P}\left[\sup_{(\theta,\gamma)}\left|\hat{G}_{(i)}(\theta,\gamma) - G_{(i)}(\theta,\gamma)\right| \ge \eta\right]. \qquad \textbf{(A.1)}\end{aligned}$$

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press

Let $\varepsilon > 0$ be chosen such that $2\varepsilon \max_i \mathrm{E}\left[M\left(x_{it}\right)\right] < \frac{\eta}{3}$. Divide Υ into subsets $\Upsilon_1, \Upsilon_2, \ldots, \Upsilon_{\vartheta(\varepsilon)}$ such that $\left|(\theta,\gamma) - \left(\theta',\gamma'\right)\right| < \varepsilon$ whenever (θ,γ) and $\left(\theta',\gamma'\right)$ are in the same subset. Let $\left(\theta_j,\gamma_j\right)$ denote *some* point in Υ_j for each j. Then

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

and therefore

$$\mathrm{P}\left[\sup_{(\theta,\gamma)}\left|\hat{G}_{(i)}(\theta,\gamma) - G_{(i)}(\theta,\gamma)\right| > \eta\right] \leq \sum_{j=1}^{\vartheta(\varepsilon)} \mathrm{P}\left[\sup_{\Upsilon_j}\left|\hat{G}_{(i)}(\theta,\gamma) - G_{(i)}(\theta,\gamma)\right| > \eta\right]. \tag{A.2}$$

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

$$\begin{aligned}\left|\hat{G}_{(i)}(\theta,\gamma) - G_{(i)}(\theta,\gamma)\right| \leq{}& \left|\hat{G}_{(i)}\left(\theta_j,\gamma_j\right) - G_{(i)}\left(\theta_j,\gamma_j\right)\right| \\ &+ \frac{\varepsilon}{T}\left|\sum_{t=1}^{T}\left(M\left(x_{it}\right) - \mathrm{E}\left[M\left(x_{it}\right)\right]\right)\right| + 2\varepsilon\mathrm{E}\left[M\left(x_{it}\right)\right].\end{aligned}$$

Then

$$\begin{aligned}\mathrm{P}\left[\sup_{\Upsilon_j}\left|\hat{G}_{(i)}(\theta,\gamma) - G_{(i)}(\theta,\gamma)\right| > \eta\right] \leq{}& \mathrm{P}\left[\left|\hat{G}_{(i)}\left(\theta_j,\gamma_j\right) - G_{(i)}\left(\theta_j,\gamma_j\right)\right| > \frac{\eta}{3}\right] \\ &+ \mathrm{P}\left[\frac{1}{T}\left|\sum_{t=1}^{T}\left(M\left(x_{it}\right) - \mathrm{E}\left[M\left(x_{it}\right)\right]\right)\right| > \frac{\eta}{3\varepsilon}\right] \\ ={}& o\left(T^{-2}\right),\end{aligned} \tag{A.3}$$

by Lemma 1. Combining (A.1), (A.2), (A.3), and $n = O(T)$, we obtain the desired conclusion. ■

THEOREM 3. $\mathrm{P}\left[\left\|\hat{\theta} - \theta_0\right\| \geq \eta\right] = o\left(T^{-1}\right)$ *for every* $\eta > 0$.

Proof. Let η be given, and let

$$\varepsilon \equiv \inf_i\left[G_{(i)}(\theta_0,\gamma_{i0}) - \sup_{\{(\theta,\gamma):\|(\theta,\gamma)-(\theta_0,\gamma_{i0})\|>\eta\}} G_{(i)}(\theta,\gamma)\right] > 0.$$

With probability equal to $1 - o\left(\frac{1}{T}\right)$, we have

$$\begin{aligned}\max_{\|\theta-\theta_0\|>\eta,\gamma_1,\ldots,\gamma_n} n^{-1}\sum_{i=1}^{n}\hat{G}_{(i)}(\theta,\gamma_i) \leq{}& \max_{\|(\theta,\gamma_i)-(\theta_0,\gamma_{i0})\|>\eta} n^{-1}\sum_{i=1}^{n}\hat{G}_{(i)}(\theta,\gamma_i) \\ <{}& \max_{\|(\theta,\gamma_i)-(\theta_0,\gamma_{i0})\|>\eta} n^{-1}\sum_{i=1}^{n}G_{(i)}(\theta,\gamma_i) + \frac{1}{3}\varepsilon \\ <{}& n^{-1}\sum_{i=1}^{n}G_{(i)}(\theta_0,\gamma_{i0}) - \frac{2}{3}\varepsilon \\ <{}& n^{-1}\sum_{i=1}^{n}\hat{G}_{(i)}(\theta_0,\gamma_{i0}) - \frac{1}{3}\varepsilon,\end{aligned}$$

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press

where the second and fourth inequalities are based on Lemma 4. Because

$$\max_{\theta,\gamma_1,\ldots,\gamma_n} n^{-1}\sum_{i=1}^{n}\hat{G}_{(i)}(\theta,\gamma_i) \geq n^{-1}\sum_{i=1}^{n}\hat{G}_{(i)}(\theta_0,\gamma_{i0}),$$

by definition we can conclude that $\mathrm{P}\left[\left\|\hat{\theta}-\theta_0\right\| \geq \eta\right] = o\left(T^{-1}\right)$. ■

THEOREM 4. $\mathrm{P}\left[\max_{1\leq i\leq n}\left|\hat{\gamma}_i-\gamma_{i0}\right| \geq \eta\right] = o\left(T^{-1}\right)$.

Proof. We first prove that

$$T\,\mathrm{P}\left[\max_{1\leq i\leq n}\sup_{\gamma}\left|\hat{G}_{(i)}\left(\hat{\theta},\gamma\right)-G_{(i)}(\theta_0,\gamma)\right| \geq \eta\right] = o(1) \tag{A.4}$$

for every $\eta > 0$. Note that

$$\begin{aligned}
&\max_{1\leq i\leq n}\sup_{\gamma}\left|\hat{G}_{(i)}\left(\hat{\theta},\gamma\right)-G_{(i)}(\theta_0,\gamma)\right| \\
&\quad\leq \max_{1\leq i\leq n}\sup_{\gamma}\left|\hat{G}_{(i)}\left(\hat{\theta},\gamma\right)-G_{(i)}\left(\hat{\theta},\gamma\right)\right| + \max_{1\leq i\leq n}\sup_{\gamma}\left|G_{(i)}\left(\hat{\theta},\gamma\right)-G_{(i)}(\theta_0,\gamma)\right| \\
&\quad\leq \max_{1\leq i\leq n}\sup_{(\theta,\gamma)}\left|\hat{G}_{(i)}(\theta,\gamma)-G_{(i)}(\theta,\gamma)\right| + \max_{1\leq i\leq n}\mathrm{E}\left[M(x_{it})\right]\cdot\left\|\hat{\theta}-\theta_0\right\|.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&T\,\mathrm{P}\left[\max_{1\leq i\leq n}\sup_{\gamma}\left|\hat{G}_{(i)}\left(\hat{\theta},\gamma\right)-G_{(i)}(\theta_0,\gamma)\right| \geq \eta\right] \\
&\quad\leq T\,\mathrm{P}\left[\max_{1\leq i\leq n}\sup_{(\theta,\gamma)}\left|\hat{G}_{(i)}(\theta,\gamma)-G_{(i)}(\theta,\gamma)\right| \geq \frac{\eta}{2}\right] \\
&\qquad + T\,\mathrm{P}\left[\left\|\hat{\theta}-\theta_0\right\| \geq \frac{\eta}{2\left(1+\max_{1\leq i\leq n}\mathrm{E}\left[M(x_{it})\right]\right)}\right] \\
&\quad= o(1),
\end{aligned}$$

by Lemma 4 and Theorem 3.

We now get back to the proof of Theorem 4. It suffices to prove that

$$T\,\mathrm{P}\left[\max_{1\leq i\leq n}\left|\hat{\gamma}_i-\gamma_{i0}\right| \geq \eta\right] = o(1)$$

for every $\eta > 0$. Let η be given, and let

$$\varepsilon \equiv \inf_i\left[G_{(i)}(\theta_0,\gamma_{i0}) - \sup_{\{\gamma_i:|\gamma_i-\gamma_{i0}|>\eta\}}G_{(i)}(\theta_0,\gamma_i)\right] > 0.$$

Condition on the event $\left\{\max_{1\leq i\leq n}\sup_{\gamma}\left|\hat{G}_{(i)}\left(\hat{\theta},\gamma\right)-G_{(i)}(\theta_0,\gamma)\right| \leq \frac{1}{3}\varepsilon\right\}$, which has probability equal to $1-o\left(\frac{1}{T}\right)$ by (A.4). We then have

$$\max_{|\gamma_i-\gamma_{i0}|>\eta}\hat{G}_{(i)}\left(\hat{\theta},\gamma_i\right) < \max_{|\gamma_i-\gamma_{i0}|>\eta}G_{(i)}(\theta_0,\gamma_i)+\frac{1}{3}\varepsilon < G_{(i)}(\theta_0,\gamma_{i0})-\frac{2}{3}\varepsilon < \hat{G}_{(i)}\left(\hat{\theta},\gamma_{i0}\right)-\frac{1}{3}\varepsilon.$$

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press

This is inconsistent with $\hat{G}_{(i)}\left(\hat{\theta},\hat{\gamma}_i\right) \geq \hat{G}_{(i)}\left(\hat{\theta},\gamma_{i0}\right)$, and therefore, $\left|\hat{\gamma}_i - \gamma_{i0}\right| \leq \eta$ for every i. ■

D. Proof of Expansion (A.30). Let $\hat{\gamma}_i(\theta) \equiv \text{argmax}_a \sum_{t=1}^{T} \psi(x_{it};\theta,a)$. Notice that $\hat{\theta}$ can be given an alternative characterization: $\hat{\theta} \equiv \text{argmax}_c \sum_{i=1}^{n}\sum_{t=1}^{T} \psi\left(x_{it};c,\hat{\gamma}_i(c)\right)$. Then, $\hat{\gamma}_i(\theta)$ solves

$$\sum_{t=1}^{T} \frac{\partial \psi\left(x_{it};\theta,\hat{\gamma}_i(\theta)\right)}{\partial \gamma_i} = 0 \tag{A.5}$$

and, similarly, $\hat{\theta}$ solves

$$0 = \sum_{i=1}^{n}\sum_{t=1}^{T} \frac{\partial \psi\left(x_{it};\hat{\theta},\hat{\gamma}_i\left(\hat{\theta}\right)\right)}{\partial \theta}. \tag{A.6}$$

Now multiply (A.5) by ρ_{i0}, evaluate it at $\theta = \hat{\theta}$, and add it to (A.6), such that

$$0 = \sum_{i=1}^{n}\sum_{t=1}^{T} \frac{\partial \psi\left(x_{it};\hat{\theta},\hat{\gamma}_i\left(\hat{\theta}\right)\right)}{\partial \theta} + \sum_{i=1}^{n}\left(\rho_{i0}\sum_{t=1}^{T} \frac{\partial \psi\left(x_{it};\hat{\theta},\hat{\gamma}_i\left(\hat{\theta}\right)\right)}{\partial \gamma_i}\right).$$

This shows that $\hat{\theta}$ solves

$$0 = \sum_{i=1}^{n}\sum_{t=1}^{T} U_i\left(x_{it};\hat{\theta},\hat{\gamma}_i\left(\hat{\theta}\right)\right).$$

Let $F \equiv (F_1,\ldots,F_n)$ denote the collection of (marginal) distribution functions of x_{it}. Let $\hat{F} \equiv \left(\hat{F}_1,\ldots,\hat{F}_n\right)$, where $\hat{F}_i$ denotes the empirical distribution function for the observation i. Define $F(\epsilon) \equiv F + \epsilon\sqrt{T}\left(\hat{F} - F\right)$ for $\epsilon \in \left[0, T^{-1/2}\right]$. For each fixed θ and ϵ, let $\gamma_i(\theta, F_i(\epsilon))$ be the solution to the estimating equation

$$0 = \int V_i\left[\cdot;\theta,\gamma_i(\theta,F_i(\epsilon))\right] dF_i(\epsilon),$$

and let

$$\hat{\theta}(\epsilon) = \theta(F(\epsilon)) \tag{A.7}$$

be the solution to the estimating equation

$$0 = \sum_{i=1}^{n} \int U_i\left(x_{it};\theta(F(\epsilon)),\gamma_i(\theta(F(\epsilon)),F_i(\epsilon))\right) dF_i(\epsilon).$$

By a Taylor series expansion, we have

$$\theta\left(\hat{F}\right) - \theta(F) = \frac{1}{\sqrt{T}}\theta^{\epsilon}(0) + \frac{1}{2}\left(\frac{1}{\sqrt{T}}\right)^2 \theta^{\epsilon\epsilon}(0) + \frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^3 \theta^{\epsilon\epsilon\epsilon}(\tilde{\epsilon}), \tag{A.8}$$

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press

where $\theta^{\epsilon}(\epsilon) \equiv d\theta(F(\epsilon))/d\epsilon$, $\theta^{\epsilon\epsilon}(\epsilon) \equiv d^2\theta(F(\epsilon))/d\epsilon^2$, ..., and $\tilde{\epsilon}$ is somewhere in between 0 and $T^{-1/2}$. Theorem 5 below allows us to ignore the last term. We will therefore work with the expansion

$$\sqrt{nT}\left(\theta\left(\hat{F}\right)-\theta(F)\right)=\sqrt{nT}\frac{1}{\sqrt{T}}\theta^{\epsilon}(0)+\sqrt{nT}\frac{1}{2}\left(\frac{1}{\sqrt{T}}\right)^2\theta^{\epsilon\epsilon}(0)+o_p(1). \tag{A.9}$$

Let

$$h_i(\cdot,\epsilon)\equiv U_i(\cdot;\theta(F(\epsilon)),\gamma_i(\theta(F(\epsilon)),F_i(\epsilon))). \tag{A.10}$$

The first-order condition may be written as

$$0=\frac{1}{n}\sum_{i=1}^{n}\int h_i(\cdot,\epsilon)\,dF_i(\epsilon). \tag{A.11}$$

Differentiating repeatedly with respect to ϵ, we obtain

$$0=\frac{1}{n}\sum_{i=1}^{n}\int\frac{dh_i(\cdot,\epsilon)}{d\epsilon}dF_i(\epsilon)+\frac{1}{n}\sum_{i=1}^{n}\int h_i(\cdot,\epsilon)\,d\Delta_{iT}, \tag{A.12}$$

$$0=\frac{1}{n}\sum_{i=1}^{n}\int\frac{d^2h_i(\cdot,\epsilon)}{d\epsilon^2}dF_i(\epsilon)+2\frac{1}{n}\sum_{i=1}^{n}\int\frac{dh_i(\cdot,\epsilon)}{d\epsilon}d\Delta_{iT}, \tag{A.13}$$

$$0=\frac{1}{n}\sum_{i=1}^{n}\int\frac{d^3h_i(\cdot,\epsilon)}{d\epsilon^3}dF_i(\epsilon)+3\frac{1}{n}\sum_{i=1}^{n}\int\frac{d^2h_i(\cdot,\epsilon)}{d\epsilon^2}d\Delta_{iT}, \tag{A.14}$$

where $\Delta_{iT}\equiv\sqrt{T}\left(\hat{F}_i-F_i\right)$.

D.1. $\theta^{\epsilon}(0)$. Evaluating (A.12) at $\epsilon=0$, and noting that $\mathrm{E}\left[U_i^{\gamma_i}\right]=0$, we obtain

$$\theta^{\epsilon}(0)=\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\int U_i\,d\Delta_{iT}\right). \tag{A.15}$$

D.2. $\gamma_i^{\theta},\gamma_i^{\epsilon}$. In the ith observation, $\gamma_i(\theta,F_i(\epsilon))$ solves the estimating equation

$$\int V_i(\cdot;\theta,\gamma_i(\theta,F_i(\epsilon)))\,dF_i(\epsilon)=0. \tag{A.16}$$

Differentiating the left-hand side (LHS) with respect to θ and ϵ, we obtain

$$0=\int\frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\theta}dF_i(\epsilon)+\left(\int\frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\gamma_i}dF_i(\epsilon)\right)\frac{\partial\gamma_i(\theta,F_i(\epsilon))}{\partial\theta}, \tag{A.17}$$

$$0=\left(\int\frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\gamma_i}dF_i(\epsilon)\right)\frac{\partial\gamma_i(\theta,F_i(\epsilon))}{\partial\epsilon}+\int V_i(\cdot,\theta,\epsilon)\,d\Delta_{iT}. \tag{A.18}$$

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press

Equating these equations to zero and solving for derivatives of γ_i evaluated at $\epsilon = 0$ gives

$$\gamma_i^{\theta} = -\frac{\mathrm{E}\left[\frac{\partial V_i}{\partial \theta}\right]}{\mathrm{E}\left[\frac{\partial V_i}{\partial \gamma_i}\right]}, \tag{A.19}$$

$$\gamma_i^{\epsilon} = -\frac{\Sigma_{t=1}^{T} V_{it}}{\sqrt{T}\mathrm{E}\left[\frac{\partial V_i}{\partial \gamma_i}\right]} = -\frac{T^{-1/2}\Sigma_{t=1}^{T} V_{it}}{\mathrm{E}\left[\frac{\partial V_i}{\partial \gamma_i}\right]}, \tag{A.20}$$

where $\gamma_i^{\theta} \equiv \frac{\partial \gamma_i(\theta, F_i(0))}{\partial \theta}$ and $\gamma_i^{\epsilon} \equiv \frac{\partial \gamma_i(\theta, F_i(0))}{\partial \epsilon}$. Note that they are evaluated at $\theta = \hat{\theta}(0) = \theta_0$, and $\epsilon = 0$.

D.3. $\theta^{\epsilon\epsilon}(0)$. Now, evaluating each term of (A.13) at $\epsilon = 0$, and noting that $\mathrm{E}\left[U_i^{\gamma_i}\right] = 0$, we obtain

$$\begin{aligned}
0 = {} & \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left[\frac{\partial U_i}{\partial \theta'}\right]\theta^{\epsilon\epsilon}(0) \\
& + \frac{2}{n}\sum_{i=1}^{n} \gamma_i^{\epsilon} \cdot \mathrm{E}\left[\frac{\partial^2 U_i}{\partial \theta' \partial \gamma_i}\right]\theta^{\epsilon}(0) + \frac{2}{n}\sum_{i=1}^{n} \gamma_i^{\epsilon}\left(\theta^{\epsilon}(0)'\gamma_i^{\theta}\right) \cdot \mathrm{E}\left[\frac{\partial^2 U_i}{\partial \gamma_i^2}\right] \\
& + \mathcal{G} + \frac{2}{n}\sum_{i=1}^{n} \theta^{\epsilon}(0)'\gamma_i^{\theta} \cdot \mathrm{E}\left[\frac{\partial^2 U_i}{\partial \theta' \partial \gamma_i}\right]\theta^{\epsilon}(0) + \frac{1}{n}\sum_{i=1}^{n}\left(\theta^{\epsilon}(0)'\gamma_i^{\theta}\right)^2 \cdot \mathrm{E}\left[\frac{\partial^2 U_i}{\partial \gamma_i^2}\right] \\
& + \frac{1}{n}\sum_{i=1}^{n}\left(\gamma_i^{\epsilon}\right)^2 \cdot \mathrm{E}\left[\frac{\partial^2 U_i}{\partial \gamma_i^2}\right] \\
& + \frac{2}{n}\sum_{i=1}^{n}\left(\int \frac{\partial U_i}{\partial \theta'} d\Delta_{iT}\right)\theta^{\epsilon}(0) + \frac{2}{n}\sum_{i=1}^{n}\left(\theta^{\epsilon}(0)'\gamma_i^{\theta}\right) \cdot \int \frac{\partial U_i}{\partial \gamma_i} d\Delta_{iT} \\
& + \frac{2}{n}\sum_{i=1}^{n} \gamma_i^{\epsilon} \cdot \int \frac{\partial U_i}{\partial \gamma_i} d\Delta_{iT},
\end{aligned}$$

where

$$\mathcal{G} \equiv \left[\theta^{\epsilon}(0)'\left(\tfrac{1}{n}\Sigma_{i=1}^{n}\mathrm{E}\left[\frac{\partial^2 U_i^{(1)}}{\partial \theta \partial \theta'}\right]\right)\theta^{\epsilon}(0), \ldots, \theta^{\epsilon}(0)'\left(\tfrac{1}{n}\Sigma_{i=1}^{n}\mathrm{E}\left[\frac{\partial^2 U_i^{(R)}}{\partial \theta \partial \theta'}\right]\right)\theta^{\epsilon}(0)\right]',$$

from which we obtain

$$\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)\theta^{\epsilon\epsilon}(0) = \frac{1}{n}\sum_{i=1}^{n}\left(\gamma_i^{\epsilon}\right)^2 \cdot \mathrm{E}\left[\frac{\partial^2 U_i}{\partial \gamma_i^2}\right] + \frac{2}{n}\sum_{i=1}^{n}\gamma_i^{\epsilon} \cdot \int \frac{\partial U_i}{\partial \gamma_i} d\Delta_{iT} + \mathcal{T} \tag{A.21}$$

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press

for

$$\begin{aligned}\mathcal{T} \equiv & 2\left(\frac{1}{n}\sum_{i=1}^{n}\gamma_i^{\epsilon}\cdot \mathrm{E}\left[\frac{\partial^2 U_i}{\partial\theta'\partial\gamma_i}\right]\right)\theta^{\epsilon}(0)+\left(\frac{2}{n}\sum_{i=1}^{n}\gamma_i^{\epsilon}\mathrm{E}\left[\frac{\partial^2 U_i}{\partial\gamma_i^2}\right]\left(\gamma_i^{\theta}\right)'\right)\theta^{\epsilon}(0)\\ & +2\left(\frac{1}{n}\sum_{i=1}^{n}\int\frac{\partial U_i}{\partial\theta'}d\Delta_{iT}\right)\theta^{\epsilon}(0)+\left(\frac{2}{n}\sum_{i=1}^{n}\int\frac{\partial U_i}{\partial\gamma_i}d\Delta_{iT}\cdot\left(\gamma_i^{\theta}\right)'\right)\theta^{\epsilon}(0)\\ & +\mathcal{G}+\frac{2}{n}\sum_{i=1}^{n}\theta^{\epsilon}(0)'\gamma_i^{\theta}\cdot\mathrm{E}\left[\frac{\partial^2 U_i}{\partial\theta'\partial\gamma_i}\right]\theta^{\epsilon}(0)\\ & +\frac{1}{n}\sum_{i=1}^{n}\left(\theta^{\epsilon}(0)'\gamma_i^{\theta}\right)^2\cdot\mathrm{E}\left[\frac{\partial^2 U_i}{\partial\gamma_i^2}\right]. \end{aligned} \tag{A.22}$$

Next note that

$$\frac{1}{n}\sum_{i=1}^{n}(\gamma_i^{\epsilon})^2\cdot\mathrm{E}\left[\frac{\partial^2 U_i}{\partial\gamma_i^2}\right]=\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[U_i^{\gamma_i\gamma_i}\right]\left(\frac{T^{-1/2}\Sigma_{t=1}^{T}V_{it}}{\mathrm{E}\left[\frac{\partial V_i}{\partial\gamma_i}\right]}\right)^2,$$

$$\sum_{i=1}^{n}\gamma_i^{\epsilon}\cdot\int\frac{\partial U_i}{\partial\gamma_i}d\Delta_{iT}=-\frac{1}{n}\sum_{i=1}^{n}\frac{T^{-1/2}\Sigma_{t=1}^{T}V_{it}}{\mathrm{E}\left[\frac{\partial V_i}{\partial\gamma_i}\right]}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(U_i^{\gamma_i}-\mathrm{E}\left[U_i^{\gamma_i}\right]\right)\right),$$

and that by Hahn and Kuersteiner (2010a, Lem. 3) it follows that

$$\left(\frac{1}{n}\sum_{i=1}^{n}\gamma_i^{\epsilon}\cdot\mathrm{E}\left[\frac{\partial^2 U_i}{\partial\theta'\partial\gamma_i}\right]\right)\theta^{\epsilon}(0)=O_p\left(\frac{1}{\sqrt{n}}\right)O_p\left(\frac{1}{\sqrt{n}}\right)=O_p\left(\frac{1}{n}\right), \tag{A.23}$$

$$\left(\frac{1}{n}\sum_{i=1}^{n}\gamma_i^{\epsilon}\mathrm{E}\left[\frac{\partial^2 U_i}{\partial\gamma_i^2}\right]\left(\gamma_i^{\theta}\right)'\right)\theta^{\epsilon}(0)=O_p\left(\frac{1}{\sqrt{n}}\right)O_p\left(\frac{1}{\sqrt{n}}\right)=O_p\left(\frac{1}{n}\right), \tag{A.24}$$

$$\left(\frac{1}{n}\sum_{i=1}^{n}\int\frac{\partial U_i}{\partial\theta'}d\Delta_{iT}\right)\theta^{\epsilon}(0)=O_p\left(\frac{1}{\sqrt{n}}\right)O_p\left(\frac{1}{\sqrt{n}}\right)=O_p\left(\frac{1}{n}\right), \tag{A.25}$$

$$\left(\frac{1}{n}\sum_{i=1}^{n}\int\frac{\partial U_i}{\partial\gamma_i}d\Delta_{iT}\cdot\left(\gamma_i^{\theta}\right)'\right)\theta^{\epsilon}(0)=O_p\left(\frac{1}{\sqrt{n}}\right)O_p\left(\frac{1}{\sqrt{n}}\right)=O_p\left(\frac{1}{n}\right), \tag{A.26}$$

and

$$\theta^{\epsilon}(0)'\left(\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[\frac{\partial^2 U_i^{(r)}}{\partial\theta\partial\theta'}\right]\right)\theta^{\epsilon}(0)=O_p\left(\frac{1}{\sqrt{n}}\right)O(1)\,O_p\left(\frac{1}{\sqrt{n}}\right)=O_p\left(\frac{1}{n}\right), \tag{A.27}$$

$$\frac{1}{n}\sum_{i=1}^{n}\theta^{\epsilon}(0)'\gamma_i^{\theta}\cdot\mathrm{E}\left[\frac{\partial^2 U_i}{\partial\theta'\partial\gamma_i}\right]\theta^{\epsilon}(0)=O_p\left(\frac{1}{\sqrt{n}}\right)O(1)\,O_p\left(\frac{1}{\sqrt{n}}\right)=O_p\left(\frac{1}{n}\right), \tag{A.28}$$

$$\frac{1}{n}\sum_{i=1}^{n}\left(\theta^{\epsilon}(0)'\gamma_i^{\theta}\right)^2\cdot\mathrm{E}\left[\frac{\partial^2 U_i}{\partial\gamma_i^2}\right]=O_p\left(\frac{1}{\sqrt{n}}\right)^2O(1)=O_p\left(\frac{1}{n}\right). \tag{A.29}$$

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press

These arguments establish that

$$\begin{aligned}\theta^{\epsilon\epsilon}(0) &= \left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[U_i^{\gamma_i\gamma_i}\right]\left(\frac{T^{-1/2}\sum_{t=1}^{T}V_{it}}{\mathrm{E}\left[\frac{\partial V_i}{\partial\gamma_i}\right]}\right)^2 \\ &\quad - 2\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\frac{T^{-1/2}\sum_{t=1}^{T}V_{it}}{\mathrm{E}\left[\frac{\partial V_i}{\partial\gamma_i}\right]}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(U_i^{\gamma_i}-\mathrm{E}\left[U_i^{\gamma_i}\right]\right)\right) \\ &\quad + O_p\left(\frac{1}{n}\right) \\ &= -2\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\sum_{t=1}^{T}V_{it}}{\sqrt{T}\mathrm{E}\left[\frac{\partial V_i}{\partial\gamma_i}\right]}\right]\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(U_i^{\gamma_i}-\frac{\mathrm{E}\left[U_i^{\gamma_i\gamma_i}\right]}{2\mathrm{E}\left[\frac{\partial V_i}{\partial\gamma_i}\right]}V_{it}\right)\right]+o_p(1).\end{aligned}$$

It follows that

$$\begin{aligned}\sqrt{nT}\frac{1}{2}\left(\frac{1}{\sqrt{T}}\right)^2\theta^{\epsilon\epsilon}(0) &= -\sqrt{\frac{n}{T}}\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{V_{it}}{\mathrm{E}\left[\frac{\partial V_i}{\partial\gamma_i}\right]}\right]\right. \\ &\quad \left.\times\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(U_i^{\gamma_i}-\frac{\mathrm{E}\left[U_i^{\gamma_i\gamma_i}\right]}{2\mathrm{E}\left[\frac{\partial V_i}{\partial\gamma_i}\right]}V_{it}\right)\right]\right\}+o_p(1).\end{aligned}$$

D.4. $\gamma_i^{\theta\theta}$, $\gamma_i^{\theta\epsilon}$, *and* $\gamma_i^{\epsilon\epsilon}$.

Remark 5. All the lemmas below are used to establish

$$\sqrt{nT}\frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^3\theta^{\epsilon\epsilon\epsilon}\left(\widetilde{\epsilon}\right)=o_p(1).$$

This is achieved by establishing properties of the expansion uniformly in $\widetilde{\epsilon}\in\left[0,\frac{1}{\sqrt{T}}\right]$ in the lemmas that follow.

Second-order differentiation $\left(\frac{\partial^2}{\partial\theta^2},\frac{\partial^2}{\partial\theta\partial\epsilon},\frac{\partial^2}{\partial\epsilon^2}\right)$ of (A.16) yields

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

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press

$$0 = \left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\theta\partial\gamma_i} dF_i(\epsilon) \right) \frac{\partial \gamma_i(\theta, F_i(\epsilon))}{\partial\epsilon} + \left(\int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\gamma_i} dF_i(\epsilon) \right) \frac{\partial^2 \gamma_i(\theta, F_i(\epsilon))}{\partial\theta\partial\epsilon} + \left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\gamma_i^2} dF_i(\epsilon) \right) \frac{\partial \gamma_i(\theta, F_i(\epsilon))}{\partial\theta} \frac{\partial \gamma_i(\theta, F_i(\epsilon))}{\partial\epsilon} + \int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\theta} d\Delta_{iT} + \left(\int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\gamma_i} d\Delta_{iT} \right) \frac{\partial \gamma_i(\theta, F_i(\epsilon))}{\partial\theta},$$

and

$$0 = \left(\int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\gamma_i} dF_i(\epsilon) \right) \frac{\partial^2 \gamma_i(\theta, F_i(\epsilon))}{\partial\epsilon^2} + \left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\gamma_i^2} dF_i(\epsilon) \right) \times \left(\frac{\partial \gamma_i(\theta, F_i(\epsilon))}{\partial\epsilon} \right)^2 + 2 \left(\int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\gamma_i} d\Delta_{iT} \right) \frac{\partial \gamma_i(\theta, F_i(\epsilon))}{\partial\epsilon}.$$

These three equalities characterize $\frac{\partial^2 \gamma_i(\theta, F_i(\epsilon))}{\partial\theta\partial\theta'}$, $\frac{\partial^2 \gamma_i(\theta, F_i(\epsilon))}{\partial\theta\partial\epsilon}$, and $\frac{\partial^2 \gamma_i(\theta, F_i(\epsilon))}{\partial\epsilon^2}$.

LEMMA 5. *Let* $\widehat{\theta}(\epsilon) \equiv \operatorname{argmax}_c \sum_{i=1}^n \int \psi(x_{it}; c, \gamma_i(c, F_i(\epsilon)))\, dF(\epsilon)$, *where*

$$\gamma_i(\theta, F_i(\epsilon)) \equiv \operatorname*{argmax}_a \int \psi(x_{it}; \theta, a)\, dF_i(\epsilon).$$

Then

$$T\,\mathrm{P}\left[\max_{0 \le \epsilon \le \frac{1}{\sqrt{T}}} \left| \widehat{\theta}(\epsilon) - \theta_0 \right| \ge \eta \right] = o(1),$$

and for $\widehat{\gamma}_i(\epsilon) = \gamma_i\left(\widehat{\theta}(\epsilon), F(\epsilon)\right)$,

$$T\,\mathrm{P}\left[\max_{1 \le i \le n} \max_{0 \le \epsilon \le \frac{1}{\sqrt{T}}} \left| \widehat{\gamma}_i(\epsilon) - \gamma_{i0} \right| \ge \eta \right] = o(1),$$

for every $\eta > 0$.

Remark 6. Using the notation of this Lemma, $\widehat{\gamma}_i(\theta)$ in (A.5) is defined as $\widehat{\gamma}_i(\theta) = \gamma_i\left(\theta, F\left(T^{-1/2}\right)\right)$. Similarly, $\hat{\theta}$ in Theorem 3 is defined as $\hat{\theta} = \widehat{\theta}\left(T^{-1/2}\right)$ and $\hat{\gamma}_i$ in Theorem 4 is defined as $\hat{\gamma}_i = \gamma_i\left(\widehat{\theta}, F\left(T^{-1/2}\right)\right)$.

Proof. See Hahn and Kuersteiner (2010a). ■

LEMMA 6. *Suppose that* $K_i\left(\cdot; \widehat{\theta}(\epsilon), \gamma_i\left(\hat{\theta}(\epsilon), F(\epsilon)\right)\right)$ *is equal to*

$$\frac{\partial^{m_1+m_2} \psi\left(x_{it}; \hat{\theta}(\epsilon), \gamma_i\left(\hat{\theta}(\epsilon), F(\epsilon)\right)\right)}{\partial\theta^{m_1} \partial\gamma_i^{m_2}}$$

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press

for some $m_1 + m_2 \leq 1, \ldots, 5$. *Then, for any* $\eta > 0$, *we have*

$$\mathrm{P}\left[\max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}}\left|\frac{1}{n}\sum_{i=1}^{n}\int K_i\left(\cdot;\hat{\theta}(\epsilon),\gamma_i\left(\hat{\theta}(\epsilon),F(\epsilon)\right)\right)dF_i(\epsilon)\right.\right.$$
$$\left.\left.-\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[K_i\left(x_{it};\theta_0,\gamma_{i0}\right)\right]\right|>\eta\right]=o\left(T^{-1}\right)$$

and

$$\mathrm{P}\left[\max_{i}\max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}}\left|\int K_i\left(\cdot;\hat{\theta}(\epsilon),\gamma_i\left(\hat{\theta}(\epsilon),F(\epsilon)\right)\right)dF_i(\epsilon)\right.\right.$$
$$\left.\left.-\mathrm{E}\left[K_i\left(x_{it};\theta_0,\gamma_{i0}\right)\right]\right|>\eta\right]=o\left(T^{-1}\right).$$

Also,

$$\mathrm{P}\left[\max_{i}\max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}}\left|\int K_i\left(\cdot;\hat{\theta}(\epsilon),\gamma_i\left(\hat{\theta}(\epsilon),F(\epsilon)\right)\right)d\Delta_{iT}\right|>CT^{\frac{1}{10}-\upsilon}\right]=o\left(T^{-1}\right)$$

for some constant $C > 0$ *and* $0 < \upsilon < 1/10$.

Proof. See Hahn and Kuersteiner (2010a). ■

LEMMA 7. *For* $\partial\gamma_i(\theta, F_i(\epsilon))/\partial\theta$ *and* $\partial\gamma_i(\theta, F_i(\epsilon))/\partial\epsilon$ *implicitly defined by (A.17) and (A.18), let* $\gamma_i^{\theta}(\epsilon) = \partial\gamma_i\left(\hat{\theta}(\epsilon), F_i(\epsilon)\right)/\partial\theta$ *and* $\gamma_i^{\epsilon}(\epsilon) = \partial\gamma_i\left(\hat{\theta}(\epsilon), F_i(\epsilon)\right)/\partial\epsilon$. *Then it follows that*

$$\mathrm{P}\left[\max_{i}\max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}}\left|\gamma_i^{\theta}(\epsilon)\right|>C\right]=o\left(T^{-1}\right),$$
$$\mathrm{P}\left[\max_{i}\max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}}\left|\gamma_i^{\epsilon}(\epsilon)\right|>CT^{1/10-\upsilon}\right]=o\left(T^{-1}\right),$$

for some constant $C > 0$ *and* $0 < \upsilon < 1/10$.

Proof. See Hahn and Kuersteiner (2010a). ■

LEMMA 8. *For* $\theta^{\epsilon}(\epsilon) \equiv d\theta(F(\epsilon))/d\epsilon$ *where* $\theta(F(\epsilon))$ *is defined in (A.7), it follows that*

$$\mathrm{P}\left[\max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}}\left|\theta^{\epsilon}(\epsilon)\right|>CT^{1/10-\upsilon}\right]=o\left(T^{-1}\right),$$

for some constant $C > 0$ *and* $0 < \upsilon < 1/10$.

Proof. See Hahn and Kuersteiner (2010a). ■

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press

LEMMA 9. *Let* $\gamma_i^{\theta_r\theta_{r'}}(\epsilon) \equiv \frac{\partial^2 \gamma_i\left(\hat{\theta}(\epsilon), F_i(\epsilon)\right)}{\partial\theta_r \partial\theta_{r'}}$, *where* $\frac{\partial^2 \gamma_i(\theta, F_i(\epsilon))}{\partial\theta_r \partial\theta_{r'}}$ *is defined in Section D.4. We define* $\gamma_i^{\theta_r \epsilon}(\epsilon)$ *and* $\gamma_i^{\epsilon\epsilon}(\epsilon)$ *similarly. Then it follows that*

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

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

$$\mathrm{P}\left[\max_i \max_{0\le\epsilon\le\frac{1}{\sqrt{T}}} \left|\gamma_i^{\epsilon\epsilon}(\epsilon)\right| > C\left(T^{1/10-\upsilon}\right)^2\right] = o\left(T^{-1}\right),$$

for some constant $C > 0$ *and* $0 < \upsilon < 1/10$.

Proof. See Hahn and Kuersteiner (2010a). ■

LEMMA 10.

$$\mathrm{P}\left[\max_{0\le\epsilon\le\frac{1}{\sqrt{T}}} \left|\theta^{\epsilon\epsilon}(\epsilon)\right| > C\left(T^{1/10-\upsilon}\right)^2\right] = o\left(T^{-1}\right),$$

for some constant $C > 0$ *and* $0 < \upsilon < 1/10$.

Proof. The conclusion follows by using the characterization of $\theta^{\epsilon\epsilon}(\epsilon)$, and Lemmas 6, 7, 8, and 9. ■

LEMMA 11. *Let* $\gamma_i^{\theta_r\theta_{r'}\theta_{r''}}(\epsilon) \equiv \frac{\partial^3 \gamma_i\left(\hat{\theta}(\epsilon), F_i(\epsilon)\right)}{\partial\theta_r \partial\theta_{r'} \partial\theta_{r''}}$, *where* $\frac{\partial^3 \gamma_i(\theta, F_i(\epsilon))}{\partial\theta_r \partial\theta_{r'} \partial\theta_{r''}}$ *is defined in the proof. We similarly define* $\gamma_i^{\theta_r\theta_{r'}\epsilon}(\epsilon)$, $\gamma_i^{\theta_r\epsilon\epsilon}(\epsilon)$, *and* $\gamma_i^{\epsilon\epsilon\epsilon}(\epsilon)$. *Then it follows that*

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

$$\mathrm{P}\left[\max_i \max_{0\le\epsilon\le\frac{1}{\sqrt{T}}} \left|\gamma_i^{\theta_r\theta_{r'}\epsilon}(\epsilon)\right| > CT^{1/10-\upsilon}\right] = o\left(T^{-1}\right),$$

$$\mathrm{P}\left[\max_i \max_{0\le\epsilon\le\frac{1}{\sqrt{T}}} \left|\gamma_i^{\theta_r\epsilon\epsilon}(\epsilon)\right| > C\left(T^{1/10-\upsilon}\right)^2\right] = o\left(T^{-1}\right),$$

$$\mathrm{P}\left[\max_i \max_{0\le\epsilon\le\frac{1}{\sqrt{T}}} \left|\gamma_i^{\epsilon\epsilon\epsilon}(\epsilon)\right| > C\left(T^{1/10-\upsilon}\right)^3\right] = o\left(T^{-1}\right),$$

for some constant $C > 0$ *and* $0 < \upsilon < 1/10$.

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press

Proof. See Hahn and Kuersteiner (2010a). ■

THEOREM 5.

$$\mathrm{P}\left[\max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}}\left|\theta^{\epsilon\epsilon\epsilon}(\epsilon)\right|>C\left(T^{1/10-\upsilon}\right)^3\right]=o\left(T^{-1}\right),$$

for some constant $C>0$ *and* $0<\upsilon<1/10$.

Proof. From (A.14), we have

$$0=\frac{1}{n}\sum_{i=1}^{n}\int\frac{d^3h_i(\cdot,\epsilon)}{d\epsilon^3}dF_i(\epsilon)+3\frac{1}{n}\sum_{i=1}^{n}\int\frac{d^2h_i(\cdot,\epsilon)}{d\epsilon^2}d\Delta_{iT}.$$

Combining Lemmas 6, 7, 8, 9, and 10, we can bound $\frac{1}{n}\sum_{i=1}^{n}\int\frac{d^2h_i(\cdot,\epsilon)}{d\epsilon^2}d\Delta_{iT}$ by $C\left(T^{1/10-\upsilon}\right)^3$ with probability $1-o\left(T^{-1}\right)$. Recall that the rth component $\frac{d^2h_i^{(r)}(\cdot,\epsilon)}{d\epsilon^2}$ of $\frac{d^2h_i(\cdot,\epsilon)}{d\epsilon^2}$ is equal to

$$\begin{aligned}\frac{d^2h_i^{(r)}(\cdot,\epsilon)}{d\epsilon^2}=&\frac{\partial\theta(\cdot,\epsilon)'}{\partial\epsilon}\frac{\partial^2h_i^{(r)}(\cdot,\epsilon)}{\partial\theta\partial\theta'}\frac{\partial\theta(\cdot,\epsilon)}{\partial\epsilon}+\frac{\partial^2h_i^{(r)}(\cdot,\epsilon)}{\partial\theta'\partial\gamma_i}\left(\frac{\partial\gamma_i}{\partial\theta'}\frac{\partial\theta}{\partial\epsilon}\right)\frac{\partial\theta}{\partial\epsilon}+\frac{\partial^2h_i^{(r)}(\cdot,\epsilon)}{\partial\theta'\partial\gamma_i}\frac{\partial\gamma_i}{\partial\epsilon}\frac{\partial\theta}{\partial\epsilon}\\&+\frac{\partial h_i^{(r)}(\cdot,\epsilon)}{\partial\theta'}\frac{\partial^2\theta}{\partial\epsilon^2}\\&+\frac{\partial^2h_i^{(r)}(\cdot,\epsilon)}{\partial\theta'\partial\gamma_i}\frac{\partial\theta}{\partial\epsilon}\frac{\partial\gamma_i}{\partial\theta'}\frac{\partial\theta}{\partial\epsilon}+\frac{\partial^2h_i^{(r)}(\cdot,\epsilon)}{\partial\gamma_i^2}\left(\frac{\partial\gamma_i}{\partial\theta'}\frac{\partial\theta}{\partial\epsilon}\right)^2+\frac{\partial^2h_i^{(r)}(\cdot,\epsilon)}{\partial\gamma_i^2}\left(\frac{\partial\gamma_i}{\partial\theta'}\frac{\partial\theta}{\partial\epsilon}\right)\frac{\partial\gamma_i}{\partial\epsilon}\\&+\frac{\partial h_i^{(r)}(\cdot,\epsilon)}{\partial\gamma_i}\left(\frac{\partial\theta'}{\partial\epsilon}\frac{\partial^2\gamma_i}{\partial\theta\partial\theta'}\right)\frac{\partial\theta}{\partial\epsilon}+\frac{\partial h_i^{(r)}(\cdot,\epsilon)}{\partial\gamma_i}\frac{\partial^2\gamma_i}{\partial\theta'\partial\epsilon}\frac{\partial\theta}{\partial\epsilon}+\frac{\partial h_i^{(r)}(\cdot,\epsilon)}{\partial\gamma_i}\frac{\partial\gamma_i}{\partial\theta'}\frac{\partial^2\theta}{\partial\epsilon^2}\\&+\frac{\partial^2h_i^{(r)}(\cdot,\epsilon)}{\partial\theta'\partial\gamma_i}\frac{\partial\theta}{\partial\epsilon}\frac{\partial\gamma_i}{\partial\epsilon}+\frac{\partial^2h_i^{(r)}(\cdot,\epsilon)}{\partial\gamma_i^2}\left(\frac{\partial\gamma_i}{\partial\theta'}\frac{\partial\theta}{\partial\epsilon}\right)\frac{\partial\gamma_i}{\partial\epsilon}+\frac{\partial^2h_i^{(r)}(\cdot,\epsilon)}{\partial\gamma_i^2}\left(\frac{\partial\gamma_i}{\partial\epsilon}\right)^2\\&+\frac{\partial h_i^{(r)}(\cdot,\epsilon)}{\partial\gamma_i}\frac{\partial^2\gamma_i}{\partial\epsilon\partial\theta'}\frac{\partial\theta}{\partial\epsilon}+\frac{\partial h_i^{(r)}(\cdot,\epsilon)}{\partial\gamma_i}\frac{\partial^2\gamma_i}{\partial\epsilon^2}.\end{aligned}$$

Using Lemmas 6, 7, 8, 9, and 10 again, we can conclude that $\frac{1}{n}\sum_{i=1}^{n}\int\frac{d^3h_i(\cdot,\epsilon)}{d\epsilon^3}dF_i(\epsilon)$ is equal to $\left(\frac{1}{n}\sum_{i=1}^{n}\int\frac{\partial h_i(\cdot,\epsilon)}{\partial\theta'}dF_i(\epsilon)\right)\frac{\partial^3\theta}{\partial\epsilon^3}$ plus terms that can all be bounded by $\frac{1}{n}\sum_{i=1}^{n}\int\frac{d^2h_i(\cdot,\epsilon)}{d\epsilon^2}d\Delta_{iT}$, which is bounded by $C\left(T^{1/10-\upsilon}\right)^3$ with probability $1-o\left(T^{-1}\right)$. Because $\left(\frac{1}{n}\sum_{i=1}^{n}\int\frac{\partial h_i(\cdot,\epsilon)}{\partial\theta'}dF_i(\epsilon)\right)^{-1}$ is bounded away from 0 by Lemma 6, we obtain the desired conclusion. ■

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press

E. Proof of Theorem 1. Using the Taylor series expansion as in Hahn and Newey (2004), it is shown in Section D that

$$\begin{aligned}\sqrt{nT}\left(\hat{\theta}-\theta_0\right) = &\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}\right)\\ &-\sqrt{\frac{n}{T}}\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\\ &\times\left\{\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{V_{it}}{\mathrm{E}\left[\frac{\partial V_i}{\partial\gamma_i}\right]}\right]\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(U_i^{\gamma_i}-\frac{\mathrm{E}\left[U_{it}^{\gamma_i\gamma_i}\right]}{2\mathrm{E}\left[\frac{\partial V_i}{\partial\gamma_i}\right]}V_{it}\right)\right]\right\}\\ &+o_p(1). \end{aligned}\tag{A.30}$$

In light of Lemma 3, it suffices to establish the limiting behavior of the second term on the right. Let

$$Z_{iT}\equiv\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{V_{it}}{\mathrm{E}\left[\frac{\partial V_i}{\partial\gamma_i}\right]}\right]\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(U_{it}^{\gamma_i}-\frac{\mathrm{E}\left[U_{it}^{\gamma_i\gamma_i}\right]}{2\mathrm{E}\left[\frac{\partial V_i}{\partial\gamma_i}\right]}V_{it}\right)\right].$$

Then, Z_{iT} are independent across i, and

$$\mathrm{E}\left[Z_{iT}\right]=\frac{\Sigma_{iT}^{VU}}{\mathrm{E}\left[\frac{\partial V_i}{\partial\gamma_i}\right]}-\frac{\mathrm{E}\left[U_{it}^{\gamma_i\gamma_i}\right]}{2\left(\mathrm{E}\left[\frac{\partial V_i}{\partial\gamma_i}\right]\right)^2}\Sigma_{iT}^{VV},$$

where $\Sigma_{iT}^{VU}\equiv T^{-1}\sum_{t,s=1}^{T}\mathrm{E}\left[V_{it}U_{is}^{\gamma_i\prime}\right]$ and $\Sigma_{iT}^{VV}\equiv T^{-1}\sum_{t,s=1}^{T}\mathrm{E}\left[V_{it}V_{is}'\right]$. Next note that

$$\begin{aligned}\mathrm{Var}(Z_{iT}) = T^{-2}\sum_{t_1,\ldots,t_4=1}^{T}\Bigg\{&\frac{\mathrm{E}\left[V_{it_1}V_{it_4}U_{it_2}^{\gamma_i}U_{it_3}^{\gamma_i\prime}\right]-\Sigma_{iT}^{VU}\Sigma_{iT}^{VU\prime}}{\left(\mathrm{E}\left[\frac{\partial V_i}{\partial\gamma_i}\right]\right)^2}\\ &-\frac{\mathrm{E}\left[U_{it_3}^{\gamma_i\gamma_i}\right]\left(\mathrm{E}\left[V_{it_1}V_{it_4}V_{it_3}U_{it_2}^{\gamma_i\prime}\right]-\Sigma_{iT}^{VV}\Sigma_{iT}^{VU\prime}\right)}{2\left(\mathrm{E}\left[\frac{\partial V_i}{\partial\gamma_i}\right]\right)^3}\\ &-\frac{\left(\mathrm{E}\left[V_{it_1}V_{it_4}V_{it_2}U_{it_3}^{\gamma_i}\right]-\Sigma_{iT}^{VV}\Sigma_{iT}^{VU}\right)\mathrm{E}\left[U_{it_2}^{\gamma_i\gamma_i}\right]}{2\left(\mathrm{E}\left[\frac{\partial V_i}{\partial\gamma_i}\right]\right)^3}\\ &+\frac{\mathrm{E}\left[U_{it_2}^{\gamma_i\gamma_i}\right]\mathrm{E}\left[U_{it_3}^{\gamma_i\gamma_i}\right]'\left(\mathrm{E}\left[V_{it_1}V_{it_2}V_{it_3}V_{it_4}\right]-\Sigma_{iT}^{VV}\Sigma_{iT}^{VV}\right)}{4\left(\mathrm{E}\left[\frac{\partial V_i}{\partial\gamma_i}\right]\right)^4}\Bigg\}.\end{aligned}$$

Note that $V_{it_k}, U_{it_{k'}}^{\gamma_i}$ are random variables measurable with respect to the filtration generated by x_{it}. By Condition 4, sufficient moments exist to apply Corollary A.2 of Hall

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press

and Heyde (1980, p. 278) as well as Lemma 1 of Andrews (1991). First note that for any element (j_1, j_2) we have

$$\begin{aligned}
&\mathrm{E}\left[V_{it_1}V_{it_4}U_{it_2}^{\gamma_i,j_1}U_{it_3}^{\gamma_i' j_2}\right] - \left[\Sigma_{iT}^{VU}\,\Sigma_{iT}^{VU\prime}\right]_{j_1,j_2} \\
&\quad = \mathrm{E}\left[V_{it_1}V_{it_4}\right]\mathrm{E}\left[U_{it_2}^{\gamma_i,j_1}U_{it_3}^{\gamma_i' j_2}\right] + \mathrm{E}\left[V_{it_1}U_{it_3}^{\gamma_i' j_2}\right]\mathrm{E}\left[V_{it_4}U_{it_2}^{\gamma_i,j_1}\right] \\
&\qquad + \mathrm{Cum}\left(V_{it_1}, V_{it_4}, U_{it_2}^{\gamma_i,j_1}, U_{it_3}^{\gamma_i' j_2}\right),
\end{aligned}$$

where the fourth-order cumulant $\mathrm{Cum}\left(V_{it_1}, V_{it_4}, U_{it_2}^{\gamma_i,j_1}, U_{it_3}^{\gamma_i' j_2}\right)$ is uniformly summable over three indices. This implies that $T^{-2}\sum_{t_1,\ldots,t_4=1}^{T}\left|\mathrm{Cum}\left(V_{it_1}, V_{it_4}, U_{it_2}^{\gamma_i,j_1}, U_{it_3}^{\gamma_i' j_2}\right)\right| = O\left(T^{-1}\right)$ uniformly in i. For $\delta > 0$ and some constant $0 < c < \infty$, it follows from Corollary A.2 of Hall and Heyde (1980, p. 278) and Condition 3 that

$$\sup_i\left|\mathrm{E}\left[V_{it_1}U_{it_2}^{\gamma_i,j_1}\right]\right| \le 8c\left(\mathrm{E}\left[\left|V_{it_1}\right|^{2+\delta}\right]\right)^{1/2+\delta}\left(\mathrm{E}\left[\left|U_{it_2}^{\gamma_i,j_1}\right|^{2+\delta}\right]\right)^{1/2+\delta}\left(a^{\delta/2+\delta}\right)^{|t_1-t_2|},$$

with similar inequalities holding for the remaining second moments. These arguments establish that

$$\sup_i\left\|T^{-2}\sum_{t_1,\ldots,t_4=1}^{T}\mathrm{E}\left[V_{it_1}V_{it_4}U_{it_2}^{\gamma_i}U_{it_3}^{\gamma_i'}\right] - \Sigma_{iT}^{VU}\,\Sigma_{iT}^{VU\prime}\right\| = O(1),$$

and the same can be established for the remaining terms in $\mathrm{Var}(Z_{iT})$. By the Markov inequality it follows that

$$\mathrm{P}\left[\left\|\frac{1}{n}\Sigma_{i=1}^{n}\left(Z_{iT} - \mathrm{E}\left[Z_{iT}\right]\right)\right\| > \eta\right] < \frac{\sup_i \mathrm{Var}(Z_{iT})}{n\eta^2} \to 0,$$

such that $\frac{1}{n}\Sigma_{i=1}^{n}Z_{iT} = \frac{1}{n}\Sigma_{i=1}^{n}\mathrm{E}\left[Z_{iT}\right] + O_p(n^{-1/2})$. By applying Lemma 3 to $\xi = \left(V, U^{\gamma}\right)$, we obtain $\sup_i\left|\Sigma_{iT}^{VU^{\gamma}} - f_i^{VU^{\gamma}}\right| \to 0$ and $\sup_i\left|\Sigma_{iT}^{VV} - f_i^{VV}\right| \to 0$ as $T \to \infty$. Uniformity of convergence then implies that joint and iterated limits exist and agree such that $\frac{1}{n}\Sigma_{i=1}^{n}\mathrm{E}\left[Z_{iT}\right] \to \Psi$. Therefore, the second term on the right can be written as

$$-\sqrt{\frac{n}{T}}\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\Psi + o_p(1).$$

F. Proof of Theorem 2. We first prove two lemmas.

LEMMA 12. *Let $k_{it} = k(x_{it};\theta_0,\gamma_{i0})$ and $\hat{k}_{it} = k\left(x_{it};\hat{\theta},\hat{\gamma}_i\right)$, where x_{it} satisfies Condition 3, k_{it} satisfies Condition 4, and $\hat{\theta}$, $\hat{\gamma}_i$ are defined in (1). Assume that $\mathrm{E}\left[k_{it}\right] = 0$ for i,t. Let A_i be a conformable matrix of constants such that $\max_i\|A_i\| < \infty$. Let*

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press

$f_i^{kk} = \Sigma_{l=-\infty}^{\infty} \mathrm{E}\left[k_{it}k'_{it-l}\right]$ *and* $f^{kk} = \lim_{n\to\infty} n^{-1}\Sigma_{i=1}^{n} A_i f_i^{kk}$. *Then,* $\frac{1}{n}\Sigma_{i=1}^{n} A_i \left(\frac{1}{T}\Sigma_{l=-m}^{m}\Sigma_{t=\max(1,l)}^{\min(T,T+l)} \hat{k}_{it}\hat{k}'_{it-l}\right) - f^{kk} = o_p(1)$, *where* $m,n,T \to \infty$ *such that* $m = o(T)$ *and* $n/T \to \kappa$.

Proof. Let $r_1 = \max(1,l)$ and $r_2 = \min(T, T+l)$, and define $K_{i,m} = \frac{1}{T}\Sigma_{l=-m}^{m}\Sigma_{t=r_1}^{r_2} k_{it}k'_{it-l}$. From Lemmas 3 and 4 of Hahn and Kuersteiner (2006) it follows that $\frac{1}{n}\Sigma_{i=1}^{n} A_i K_{i,m} - f^{kk} = O_p(T^{-1}) + \left(a^{\delta/2+\delta}\right)^m = o_p(1)$ as $m,n,T \to \infty$ such that $m = o(T)$ and $n/T \to \kappa$.

Next, we turn to showing that

$$\frac{1}{n}\Sigma_{i=1}^{n} A_i \frac{1}{T}\Sigma_{l=-m}^{m}\Sigma_{t=r_1}^{r_2}\left(\hat{k}_{it}\hat{k}'_{it-l} - k_{it}k'_{it-l}\right) = o_p(m/T).$$

We use the decomposition

$$\begin{aligned}
&\frac{1}{T}\Sigma_{l=-m}^{m}\Sigma_{t=r_1}^{r_2}\left(\hat{k}_{it}\hat{k}'_{it-l} - k_{it}k'_{it-l}\right)\\
&\quad= \frac{1}{T}\Sigma_{l=-m}^{m}\Sigma_{t=r_1}^{r_2}\left(\hat{k}_{it} - k_{it}\right)\left(\hat{k}_{it-l} - k_{it-l}\right)'\\
&\qquad+ \frac{1}{T}\Sigma_{l=-m}^{m}\Sigma_{t=r_1}^{r_2} k_{it}\left(\hat{k}_{it-l} - k_{it-l}\right)' + \frac{1}{T}\Sigma_{l=-m}^{m}\Sigma_{t=r_1}^{r_2}\left(\hat{k}_{it} - k_{it}\right)k'_{it-l},
\end{aligned}$$

and define

$$\begin{aligned}
k_{it}^{\theta}(x_{it};\theta,\gamma_i) &= \partial k(x_{it};\theta,\gamma_i)/\partial\theta',\\
k_{it}^{\gamma}(x_{it};\theta,\gamma_i) &= \partial k(x_{it};\theta,\gamma_i)/\partial\gamma,\\
k_{it}^{\gamma\gamma}(x_{it};\theta,\gamma_i) &= \partial^2 k(x_{it};\theta,\gamma_i)/(\partial\gamma)^2,\\
k_{it}^{\gamma\theta}(x_{it};\theta,\gamma_i) &= \partial^2 k(x_{it};\theta,\gamma_i)/\left(\partial\gamma\,\partial\theta'\right),\\
k_{it}^{\theta\theta}(x_{it};\theta,\gamma_i) &= \partial\left(\mathrm{vec}\,\partial k(x_{it};\theta,\gamma_i)/\partial\theta'\right)/\partial\theta',
\end{aligned}$$

as well as $k_{0,it}^{\theta} = k_{it}^{\theta}(x_{it};\theta_0,\gamma_{i0})$, $k_{0,it}^{\gamma} = k_{it}^{\gamma}(x_{it};\theta_0,\gamma_{i0})$, $k_{it}^{\gamma\gamma} = k_{it}^{\gamma\gamma}\left(x_{it};\tilde{\theta},\tilde{\gamma}_i\right)$, $k_{it}^{\gamma\theta} = k_{it}^{\gamma\theta}\left(x_{it};\tilde{\theta}',\tilde{\gamma}_i'\right)$, and $k_{it}^{\theta\theta} = k_{it}^{\theta\theta}\left(x_{it};\tilde{\theta}'',\tilde{\gamma}_i''\right)$. Use a second-order Taylor approximation to

$$\begin{aligned}
\hat{k}_{it} - k_{it} &= k_{0,it}^{\theta}\left(\hat{\theta}-\theta_0\right) + k_{0,it}^{\gamma}\left(\hat{\gamma}_i - \gamma_{i0}\right) + \frac{1}{2}\left(\left(\hat{\theta}-\theta_0\right)'\otimes I\right)k_{it}^{\theta\theta}\left(\hat{\theta}-\theta_0\right)\\
&\quad+ \frac{1}{2}k_{it}^{\gamma\gamma}\left(\hat{\gamma}_i - \gamma_{i0}\right)^2 + k_{it}^{\gamma\theta}\left(\hat{\theta}-\theta_0\right)\left(\hat{\gamma}_i - \gamma_{i0}\right),
\end{aligned}$$

where $\tilde{\theta}, \tilde{\gamma}_i, \tilde{\theta}', \tilde{\gamma}_i', \tilde{\theta}'', \tilde{\gamma}_i''$ such that $\left\|\tilde{\theta}-\theta_0\right\| \le \left\|\hat{\theta}-\theta_0\right\|$, $\left\|\tilde{\theta}'-\theta_0\right\| \le \left\|\hat{\theta}-\theta_0\right\|$, etc. by the multivariate version of the mean value theorem. Note that each row of $\partial k\left(x_{it};\tilde{\theta},\tilde{\gamma}_i\right)/\partial\theta'$ needs to be evaluated at a different $\tilde{\theta}$, but in slight abuse of notation we do not make this explicit. Then

$$\frac{1}{T}\Sigma_{l=-m}^{m}\Sigma_{t=r_1}^{r_2}\left(\hat{k}_{it}\hat{k}'_{it-l} - k_{it}k'_{it-l}\right) = \Sigma_{j=3}^{35} T_j,$$

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press

where T_j is defined in Hahn and Kuersteiner (2006). By Lemmas 5–20 in Hahn and Kuersteiner (2006) it follows that $\sum_{j=3}^{35} T_j = O_p(m/T)$, which establishes the result. ■

LEMMA 13. *Under Conditions 1–7, we have (i)* $\max_i \left\| \hat{\mathrm{E}}\left[V_i^{\gamma_i}\right] - \mathrm{E}\left[V_i^{\gamma_i}\right] \right\| = o_p(1)$*; (ii)* $\max_i \left\| \hat{\mathrm{E}}\left[V_i^{\theta}\right] - \mathrm{E}\left[V_i^{\theta}\right] \right\| = o_p(1)$*; (iii)* $\max_i \left\| \hat{\mathrm{E}}\left[U_i^{\gamma_i\gamma_i}\right] - \mathrm{E}\left[U_i^{\gamma_i\gamma_i}\right] \right\| = o_p(1)$*; (iv)* $n^{-1}\sum_{i=1}^n \hat{\mathcal{I}}_i - \mathcal{I} = o_p(1)$.

Proof. We note that $\hat{\beta}$ depends on the estimators $\hat{\theta}$ and $\hat{\gamma}_i$. We obtain feasible versions of $\hat{\gamma}_i$ by substituting $\hat{\theta}$ in the criterion function and maximizing with respect to γ_i such that $\hat{\gamma}_i(\epsilon)$ solves $\hat{\gamma}_i(\epsilon) \equiv \operatorname{argmax}_a \int \psi\left(x_{it}; \hat{\theta}(\epsilon), a\right) dF_i(\epsilon)$. First consider $\hat{\mathrm{E}}\left[V_i^{\gamma_i}\right] = \frac{1}{T}\sum_{t=1}^T V_{it}^{\gamma_i}\left(x_{it}; \hat{\theta}, \hat{\gamma}_i\right)$. We have

$$\begin{aligned}
\left\| \hat{\mathrm{E}}\left[V_i^{\gamma_i}\right] - \mathrm{E}\left[V_i^{\gamma_i}\right] \right\| &\leq \left\| \frac{1}{T}\sum_{t=1}^T V_{it}^{\gamma_i}\left(x_{it}; \hat{\theta}, \hat{\gamma}_i\right) - \frac{1}{T}\sum_{t=1}^T V_{it}^{\gamma_i}\left(x_{it}; \theta_0, \gamma_{i0}\right) \right\| \\
&\quad + \left\| \frac{1}{T}\sum_{t=1}^T V_{it}^{\gamma_i}\left(x_{it}; \theta_0, \gamma_{i0}\right) - \mathrm{E}\left[V_{it}^{\gamma_i}\left(x_{it}; \theta_0, \gamma_{i0}\right)\right] \right\| \\
&\leq \left(\frac{1}{T}\sum_{t=1}^T \|M(x_{it})\| \right) \left(\left\| \hat{\theta} - \theta \right\| + \max_i \left| \hat{\gamma}_i - \gamma_{i0} \right| \right) \\
&\quad + \max_i \left\| \frac{1}{T}\sum_{t=1}^T V_{it}^{\gamma_i}\left(x_{it}; \theta_0, \gamma_{i0}\right) - \mathrm{E}\left[V_{it}^{\gamma_i}\left(x_{it}; \theta_0, \gamma_{i0}\right)\right] \right\|,
\end{aligned}$$

so that

$$\begin{aligned}
\max_i \left\| \hat{\mathrm{E}}\left[V_i^{\gamma_i}\right] - \mathrm{E}\left[V_i^{\gamma_i}\right] \right\| &\leq \left(\max_i \frac{1}{T}\sum_{t=1}^T \|M(x_{it})\| \right) \left(\left\| \hat{\theta} - \theta \right\| + \max_i \left| \hat{\gamma}_i - \gamma_{i0} \right| \right) \\
&\quad + \max_i \left\| \frac{1}{T}\sum_{t=1}^T V_{it}^{\gamma_i}\left(x_{it}; \theta_0, \gamma_{i0}\right) - \mathrm{E}\left[V_{it}^{\gamma_i}\left(x_{it}; \theta_0, \gamma_{i0}\right)\right] \right\|.
\end{aligned}$$

By Lemma 2, the second term tends to zero with probability $1 - o(T^{-1})$. Applying Theorems 3 and 4 to the first term, we obtain

$$\max_i \left\| \hat{\mathrm{E}}\left[V_i^{\gamma_i}\right] - \mathrm{E}\left[V_i^{\gamma_i}\right] \right\| = o_p(1).$$

In the same way, we obtain

$$\max_i \left\| \hat{\mathrm{E}}\left[V_i^{\theta}\right] - \mathrm{E}\left[V_i^{\theta}\right] \right\| = o_p(1)$$

and

$$\max_i \left\| \hat{\mathrm{E}}\left[U_i^{\gamma_i\gamma_i}\right] - \mathrm{E}\left[U_i^{\gamma_i\gamma_i}\right] \right\| = o_p(1).$$

Let $\hat{\mathrm{E}}\left[V_i^{\gamma_i}\right] = \mathrm{E}\left[V_i^{\gamma_i}\right] + o_p(1)$, which holds uniformly in i. Thus,

$$\max_i \left\| \hat{\mathcal{I}}_i - \mathcal{I}_i \right\| \leq \left(\max_i \frac{1}{T}\sum_{t=1}^T \|M(x_{it})\| \right) \left(\left\| \hat{\theta} - \theta \right\| + \max_i \left| \hat{\gamma}_i - \gamma_{i0} \right| \right) + o_p(1).$$

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press

Since $\left\|\hat{\theta}-\theta\right\| = o_p(1)$ and $\max_i \left|\hat{\gamma}_i - \gamma_{i0}\right| = o_p(1)$ by Theorems 3 and 4, and $\max_i \frac{1}{T}\sum_{t=1}^{T} \|M(x_{it})\| = O_p(1)$ by Lemma 1, it follows that $\max_i \left\|\hat{\mathcal{I}}_i - \mathcal{I}_i\right\| = o_p(1)$ such that

$$n^{-1}\sum_{i=1}^{n} \hat{\mathcal{I}}_i - \mathcal{I} = o_p(1).$$

■

The proof of Theorem 2 now follows from Lemmas 12 and 13. By Lemma 13, we have

$$\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\hat{f}_i^{VU^\gamma}}{\hat{\mathrm{E}}\left[V_{it}^{\gamma_i}\right]} - \frac{\hat{f}_i^{VU^\gamma}}{\mathrm{E}\left[V_{it}^{\gamma_i}\right]}\right] = o_p(1),$$

$$\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\hat{\mathrm{E}}\left[U_i^{\gamma_i\gamma_i}(x_{it};\theta,\gamma_i)\right]\hat{f}_i^{VV}}{2\left(\hat{\mathrm{E}}\left[V_{it}^{\gamma_i}\right]\right)^2} - \frac{\mathrm{E}\left[U_i^{\gamma_i\gamma_i}(x_{it};\theta,\gamma_i)\right]\hat{f}_i^{VV}}{2\left(\mathrm{E}\left[V_{it}^{\gamma_i}\right]\right)^2}\right] = o_p(1).$$

In order to establish the result we thus need to apply Lemma 12 to show that

$$\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\hat{f}_i^{VU^\gamma} - f_i^{VU^\gamma}}{\mathrm{E}\left[V_{it}^{\gamma_i}\right]}\right] = o_p(1), \qquad \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\mathrm{E}\left[U_i^{\gamma_i\gamma_i}(x_{it};\theta,\gamma_i)\right]\left(\hat{f}_i^{VV} - f_i^{VV}\right)}{\left(\mathrm{E}\left[V_{it}^{\gamma_i}\right]\right)^2}\right] = o_p(1).$$

The result follows by Lemma 12, since $\inf\left|\mathrm{E}\left[V_{it}^{\gamma_i}\right]\right| > 0$ by Condition 6 and $\left\|\mathrm{E}\left[U_i^{\gamma_i\gamma_i}(x_{it};\theta,\gamma_i)\right]\right\| \le \mathrm{E}\left[M(x_{it})\right] < \infty$.

G. Bandwidth Selection: AR(1) Approximation. We define

$$\hat{k}_{it} = k\left(x_{it};\hat{\theta},\hat{\gamma}_i\right) = \begin{bmatrix} V_i\left(x_{it};\hat{\theta},\hat{\gamma}_i\right) \\ U_i^{\gamma_i}\left(x_{it};\hat{\theta},\hat{\gamma}_i\right) \end{bmatrix}$$

and let

$$a_{i,1} = -\frac{\mathrm{E}\left[U_i^{\gamma_i\gamma_i}(x_{it};\theta,\gamma_i)\right]}{2\left(\mathrm{E}\left[\frac{\partial V_i(x_{it};\theta,\gamma_i)}{\partial\gamma_i}\right]\right)^2}, a_{i,2} = \frac{1}{\mathrm{E}\left[\frac{\partial V_i(x_{it};\theta,\gamma_i)}{\partial\gamma_i}\right]},$$

where a_1 is an $R\times 1$ vector where $R = \dim\theta$ and a_2 is a scalar. Then define the $2R\times(R+1)$ matrix

$$\tilde{A} = \begin{bmatrix} a_1 & 0_{R\times 1} & \cdots & 0_{R\times 1} \\ 0 & a_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & a_2 \end{bmatrix},$$

and let

$$A_i = -\mathcal{I}^{-1}[I_R, I_R]\tilde{A},$$

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press

where I_R is the R-dimensional identity matrix such that

$$\tilde{\beta}_m \equiv \frac{1}{n}\sum_{i=1}^{n} A_i \frac{1}{T}\sum_{l=-m}^{m}\sum_{t=\max(1,l)}^{\min(T,T+l)} \hat{k}_{it}\hat{k}'_{it-l}B,$$

where

$$B = \begin{bmatrix} 1 \\ 0_{R\times 1} \end{bmatrix}.$$

Let $k_{it}^{\gamma}(x_{it};\theta,\gamma_i) = \partial k(x_{it};\theta,\gamma_i)/\partial\gamma$ and $\psi_{it}^{\gamma,v} = -\left(\mathrm{E}\left[V_i^{\gamma}\right]\right)^{-1}\frac{\partial\psi(x_{it};\theta_0,\gamma_{i0})}{\partial\gamma}$. We approximate the spectral density matrix of k_{it} with a parametric model. We use $\hat{k}_{it}^{\gamma} = k_{it}^{\gamma}\left(x_{it};\hat{\theta},\hat{\gamma}_i\right)$ and so on to denote estimated quantities. Let

$$\tilde{H}_i = \begin{bmatrix} \tilde{h}_{i1} & & 0 \\ & \ddots & \\ 0 & & \tilde{h}_{iR+1} \end{bmatrix}$$

be the pseudo-true value that minimizes $\mathrm{E}\left\|k_{it} - H_i k_{it-1}\right\|^2$ under the restriction that the off-diagonal elements of H_i are zero. The parameters $\tilde{h}_i$ are estimated as

$$\hat{h}_{ij} = \frac{\left(T^{-1}\Sigma_{t=1}^{T-1}\hat{k}_{j,it+1}\hat{k}'_{j,it}\right)}{\left(T^{-1}\Sigma_{t=1}^{T-1}\hat{k}_{j,it}\hat{k}'_{j,it}\right)},$$

where $\hat{k}_{j,it+1}$ is the jth element of $\hat{k}_{it+1}$. Then define

$$\hat{H}_i = \begin{bmatrix} \hat{h}_{i1} & & 0 \\ & \ddots & \\ 0 & & \hat{h}_{iR+1} \end{bmatrix}$$

and the function

$$\hat{\mathfrak{T}}_0(m) = \frac{1}{n}\Sigma_{i=1}^{n}\hat{A}_i\left[\left(\hat{H}_i\right)^{m+1}\left(I-\hat{H}_i\right)^{-1}\left(T^{-1}\Sigma_{t=1}^{T}\hat{k}_{it}\hat{k}'_{it}\right) + \left(T^{-1}\Sigma_{t=1}^{T}\hat{k}_{it}\hat{k}'_{it}\right)\left(I-\hat{H}_i\right)^{-1}\left(\hat{H}_i\right)^{m+1}\right]. \tag{A.31}$$

Note that $\left(\hat{H}_i\right)^m$ is the diagonal matrix with elements $\left(\hat{h}_{ij}\right)^m$ on the diagonal.

Select the bandwidth $\hat{m}$ by solving

$$\hat{m} = \arg\min_m B'\left(\hat{\mathfrak{T}}_0(m) + \frac{m}{T}\mathbf{1}_R\mathbf{1}'_{R+1}\right)'\left(\hat{T}_0(m) + \frac{m}{T}\mathbf{1}_R\mathbf{1}'_{R+1}\right)B,$$

where $\mathbf{1}_R$ is an $R\times 1$ matrix consisting of all elements equal to one. Note that $\mathbf{1}'_{R+1}B = 1$ such that the bias contribution to $\hat{\beta}_m$ that is increasing in m takes on the form $\frac{m}{T}\mathbf{1}_R$; i.e., all elements of β are affected in the same way.

https://doi.org/10.1017/S0266466611000028 Published online by Cambridge University Press