Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects

Jinyong Hahn
UCLA

Guido Kuersteiner
MIT

May, 2004

Acknowledgment: We are grateful for helpful comments by Gary Chamberlain, Shakeeb Khan, Whitney Newey, and workshop participants of AUEB, Boston University, Columbia University, Federal Reserve Board, Georgetown University, Harvard/MIT, Iowa State University, Panel Conference sponsored by NSF-NBER-UCLA, Penn State University, Université de Montréal, UC Irvine, University of Cyprus, University of Pittsburgh, USC and Yale University. We thank Ekaterini Kyriazidou for making her Gauss code available. The second author gratefully acknowledges Financial support from NSF Grant SES 0095132.

Abstract

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 as T grows with n. We consider asymptotics where n and T grow at the same rate as an approximation that allows us to compare bias properties. Under these asymptotics, bias corrected estimators we propose are centered at the truth, whereas fixed effects estimators are not. Our methods are applicable to a wide variety of non-linear dynamic panel models. We discuss several examples and provide Monte Carlo evidence for the small sample performance of our procedure.

1 Introduction

Panel data, consisting of observations across time for different individual economic agents, allows the possibility of controlling for unobserved individual heterogeneity. Failure to control for heterogeneity can result in misleading inferences. One method to deal with unobserved individual effects is to treat each effect as a separate parameter to be estimated. Unfortunately, these estimators are typically subject to the incidental parameters problem noted by Neyman and Scott (1948). The estimators of the parameters of interest will be inconsistent if the number of individuals goes to infinity while the number of time periods is held fixed, which suggests that fixed effects estimators may be severely biased.

Hahn and Kuersteiner (2002) recently showed that the bias in a panel AR(1) model can be alleviated substantially 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. Hahn and Newey (2002) showed how the bias correction can be implemented in a *static* nonlinear panel data model with fixed effects under the same asymptotics. The static assumption there is so strong that both dependent and explanatory variables are required to be independently distributed over time. In other words, even explanatory variables are required to be independent and identically distributed over time, which is clearly violated for many applications. In this paper, we examine asymptotic biases of general dynamic nonlinear panel models with fixed effects, and develop methods to remove them.

Our analysis could provide a useful alternative to papers that propose estimators that have desirable properties with T fixed such as Honoré and Kyriazidou (2000), who examined a dynamic binary response model. Their identification and estimation methodology requires conditioning on covariates taking identical values over time, which is not required in our approach. On the other hand, it is expected that our approach requires reasonably large T to be a good approximation. We investigate in Monte Carlo experiments how large T needs to be in practice for our approach to work. For the dynamic Probit model discussed in Example 2, we find that the bias can be reduced by half when $T = 16$ in samples of 250 and 500 individuals. For the dynamic Logit model similar bias reductions are available already for $T = 8$, again for 250 and 500 individuals. As far as the root mean squared error (RMSE) of the estimators is concerned, we find that for the MLE the bias component by far is the dominating factor of the RMSE. Our simulation results also show that bias correction does not seem to come at the cost of increased variances and consequently the reductions in RMSE of our bias corrected estimators relative to the MLE are of the same order of magnitude as the bias reductions. The sample sizes we consider in our Monte Carlo experiments are relevant for many real life applications in labor economics and consumer choice.

There exist a number of methods for bias correction in the literature. The jackknife and bootstrap are examples. In an iid context it was shown by Hahn, Kuersteiner and Newey (2002) that these methods are all equivalent. With temporal dependence this equivalence is not likely to carry over in general because the need to employ blocking schemes reduces relevant rates of convergence for bias estimators. More specifically it is not clear how the jackknife discussed in Hahn and Newey (2002) could be adapted to the situation where observations are dependent. A potential drawback of any bias correction method is that the fixed effect estimators are needed as preliminary estimators for estimates of the bias, and may strongly affect the quality of the latter, as discussed by Kiviet (1995) in a slightly different context. Our Monte Carlo experiments shed some light on the finite sample performance of our method.

One advantage of our approach is that it is quite flexible. We discuss examples that show how our method can be applied to a variety of, mainly parametric, non-linear models. For some of these models there do not seem to exist any adequate alternative estimators which are less biased than the MLE in finite samples.

2 Bias Corrected Estimator: Intuition

In this section, we characterize the approximate bias of the fixed effects estimator for a dynamic nonlinear panel model. We consider an asymptotic approximation where n and T grow to infinity at the same rate. We show that the fixed effects estimator is consistent and asymptotically normal, but has an asymptotic bias. We provide a formula for the asymptotic bias under our asymptotic approximation.

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

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

for some criterion function $\psi(\cdot)$ that does not depend on T. We assume that ψ is a sensible function that a time-series econometrician would use: If n is fixed, and $T \to \infty$, the estimator is such that $\left(\widehat{\theta}, \widehat{\gamma}_1, \ldots, \widehat{\gamma}_n\right)$ is consistent for $(\theta_0, \gamma_{10}, \ldots, \gamma_{n0})$. For simplicity of notation, we will assume $\dim(\theta) = R$ and $\dim(\gamma) = 1$.

Example 1 *The binary panel model with fixed effects is characterized by* $y_{it} = 1\left(\gamma_{i0} + z_{it}'\theta_0 + e_{it}\right)$ *where* e_{it} *conditional on* z_{it} *either has a logistic or standard normal distribution. The MLE, defined as the maximizer of*

$$\textstyle\sum_{i=1}^{n} \sum_{t=1}^{T} \left[y_{it} \log \Lambda\left(\gamma_i + z_{it}'\theta\right) + \left(1 - y_{it}\right) \log\left(1 - \Lambda\left(\gamma_i + z_{it}'\theta\right)\right)\right],$$

where Λ *denotes the Logit or Probit CDF, is consistent as* $T \to \infty$*. The MLE is a special case of (1), with* $x_{it} = \left(y_{it}, z_{it}'\right)'$ *and* $\psi\left(x_{it}; \theta, \gamma_i\right) = y_{it} \log \Lambda\left(\gamma_i + z_{it}'\theta\right) + \left(1 - y_{it}\right) \log\left(1 - \Lambda\left(\gamma_i + z_{it}'\theta\right)\right)$*. Extensions to multinomial Logit follow easily in the same way and have been applied for example by Hendel and Nevo (2002).*

Example 2 *The dynamic binary panel model with fixed effects is defined as* $y_{it} = 1\left(\gamma_{i0} + z_{it}'\beta_0 + \tau_0 y_{i,t-1} + \varepsilon_{it}\right)$ *where the MLE is again obtained from maximizing the conditional likelihood based on the Logit or Probit CDF depending on whether* ε_{it} *is conditionally logistic or standard normal. The criterion then takes on the form*

$$\psi\left(x_{it}; \theta, \gamma_i\right) = y_{it} \log \Lambda\left(\gamma_i + z_{it}'\beta + \tau y_{it-1}\right) + \left(1 - y_{it}\right) \log\left(1 - \Lambda\left(\gamma_i + z_{it}'\beta + \tau y_{it-1}\right)\right)$$

with $x_{it} = \left(y_{it}, z_{it}', y_{it-1}\right)'$ *and* $\theta = \left(\beta', \tau\right)'$*. 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 have been considered by Honoré (1993) who obtains orthogonality conditions and constructs a method of moments estimator. The model can be written as* $y_{it} = \max(0, z_{it}'\beta_0 + \tau_0 y_{it-1} + \gamma_{i0} + \varepsilon_{it})$*. If* ε_{it} *is iid Gaussian as in the previous example, we obtain*

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

where Λ is the cumulative distribution function of the standard normal distribution and λ is the corresponding density. Also, for the special case where $\tau_0 = 0$, we have an unobserved effects Tobit model that was considered in Heckman and Macurdy (1980) and Honoré (1992).

The fixed effects estimator for $\widehat{\theta}$ is obtained formally by concentrating out the fixed effects γ_i from the criterion function. We solve

$$\widehat{\gamma}_i(\theta) \equiv \underset{a}{\operatorname{argmax}} \sum_{t=1}^{T} \psi(x_{it}; \theta, a) \text{ and } \widehat{\theta} \equiv \underset{c}{\operatorname{argmax}} \sum_{i=1}^{n} \sum_{t=1}^{T} \psi(x_{it}; c, \widehat{\gamma}_i(c))$$

where $\widehat{\gamma}_i(\theta)$ is obtained for each individual i. Substituting out the estimator for γ_i in ψ, then leads to the concentrated criterion function and $\widehat{\theta}$ can be characterized as the solution to the first order condition

$$0 = \sum_{i=1}^{n} \sum_{t=1}^{T} U_i\left(x_{it}; \widehat{\theta}, \widehat{\gamma}_i\left(\widehat{\theta}\right)\right), \tag{2}$$

where we use the following notation throughout the paper:

$$U_i(x_{it}; \theta, \gamma_i) \equiv \frac{\partial \psi(x_{it}; \theta, \gamma_i)}{\partial \theta} - \rho_{i0} \cdot \frac{\partial \psi(x_{it}; \theta, \gamma_i)}{\partial \gamma_i}, \quad V_i(x_{it}; \theta, \gamma_i) \equiv \frac{\partial \psi(x_{it}; \theta, \gamma_i)}{\partial \gamma_i},$$
$$\rho_{i0} \equiv E\left[\frac{\partial^2 \psi(x_{it}; \theta_0, \gamma_{i0})}{\partial \theta \partial \gamma_i}\right] \Big/ E\left[\frac{\partial^2 \psi(x_{it}; \theta_0, \gamma_{i0})}{\partial \gamma_i^2}\right], \quad \mathcal{I}_i \equiv -E\left[\frac{\partial U_i(x_{it}; \theta_0, \gamma_{i0})}{\partial \theta'}\right]$$

Note that, in case ψ is the log likelihood, $V_i(x_{it}; \theta, \gamma_i)$, $U_i(x_{it}; \theta, \gamma_i)$, and $\mathcal{I}_i$ denote the score for γ_i, the efficient score for θ, and the Fisher information for θ from the ith observation. For simplicity of notation, we will write $U_{it} \equiv U_i(x_{it}; \theta_0, \gamma_{i0})$ 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. Note that $E\left[U_{it}^{\gamma_i}\right] = 0$.

The key to understanding inconsistency results when T is fixed and corresponding biases when T increases with n is to note that, while $E[U_{it}] = 0$ in general, it does not hold that $E[U(x_{it}; \theta_0, \widehat{\gamma}_i(\theta_0))] = 0$. Replacing γ_{i0} with an estimator $\widehat{\gamma}_i(\theta_0)$ therefore results in biases in the estimation of θ. To understand the nature of the bias introduced by the fixed effect estimator $\widehat{\gamma}_i$, we consider an infeasible estimator $\tilde{\theta}$ based on $\widehat{\gamma}_i(\theta_0)$ rather than $\widehat{\gamma}_i\left(\hat{\theta}\right)$, where $\tilde{\theta}$ solves the first order conditions

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

Our analysis in Section 4 establishes an expansion for the more complicated feasible estimator $\hat{\theta}$. The results in Section 4 imply that $\sqrt{nT}\left(\hat{\theta} - \tilde{\theta}\right) = o_p(1)$, which means that the intuition gained from studying $\tilde{\theta}$ carries over to $\hat{\theta}$. For $\tilde{\theta}$, standard arguments suggest that

$$\sqrt{nT}\left(\widetilde{\theta} - \theta_0\right) \approx \left(\frac{1}{n}\sum_{i=1}^{n} \mathcal{I}_i\right)^{-1} \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} U(x_{it}; \theta_0, \widehat{\gamma}_i(\theta_0)).$$

Because $E[U(x_{it}; \theta_0, \widehat{\gamma}_i(\theta_0))] \neq 0$, we cannot apply the central limit theorem to the numerator on the right side. We use a second order Taylor series expansion to approximate $U(x_{it}; \theta_0, \widehat{\gamma}_i(\theta_0))$ around γ_{i0}:

$$\frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} U(x_{it}; \theta_0, \widehat{\gamma}_i(\theta_0)) \approx \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} U_{it}$$
$$+ \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} U_{it}^{\gamma_i} (\widehat{\gamma}_i(\theta_0) - \gamma_{i0}) + \frac{1}{2\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} U_{it}^{\gamma_i \gamma_i} (\widehat{\gamma}_i(\theta_0) - \gamma_{i0})^2$$

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

$$\begin{aligned}\sqrt{nT}\left(\widetilde{\theta}-\theta_0\right) \approx & \left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}\right) \\ & -\sqrt{\frac{n}{T}}\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\sum_{t=1}^{T}V_{it}}{\sqrt{T}E\left[V_{it}^{\gamma_i}\right]}\right]\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(U_{it}^{\gamma_i}-\frac{E\left[U_{it}^{\gamma_i\gamma_i}\right]}{2E\left[V_{it}^{\gamma_i}\right]}V_{it}\right)\right]\end{aligned}$$

The probability limit of the second term on the right determines the asymptotic bias of $\widetilde{\theta}$. Let

$$\beta = -\operatorname{plim}\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\sum_{t=1}^{T}V_{it}}{\sqrt{T}E\left[V_{it}^{\gamma_i}\right]}\right]\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(U_{it}^{\gamma_i}-\frac{E\left[U_{it}^{\gamma_i\gamma_i}\right]}{2E\left[V_{it}^{\gamma_i}\right]}V_{it}\right)\right]$$

We can then write

$$\widetilde{\theta} \approx \theta_0 + \left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}\right) + T^{-1}\beta$$

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

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 such as the ones considered by Hahn and Kuersteiner (2002) this correlation is zero if the model only contains exogenous regressors. With the more general non-linear models considered here it may be non-zero 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 $U_{it}^{\gamma_i\gamma_i} = 0$.

To describe the nature of the bias more fully we define spectra and cross spectra at frequency zero in the following way

$$\begin{aligned}&f_i^{VU^\gamma} \equiv \sum_{l=-\infty}^{\infty}\operatorname{Cov}\left(V_{it},U_{it-l}^{\gamma_i}\right), \quad f_i^{VV} \equiv \sum_{l=-\infty}^{\infty}\operatorname{Cov}\left(V_{it},V_{it-l}\right), \\ &\varphi^{VU^\gamma} \equiv \lim n^{-1}\sum_{i=1}^{n}\left(E\left[V_{it}^{\gamma_i}\right]\right)^{-1}f_i^{VU^\gamma}, \quad \varphi^{VV} \equiv \frac{1}{2}\lim n^{-1}\sum_{i=1}^{n}\left(E\left[V_{it}^{\gamma_i}\right]\right)^{-2}E\left[U_{it}^{\gamma_i\gamma_i}\right]f_i^{VV}.\end{aligned}$$

Note that $f_i^{VU^\gamma}$ and f_i^{VV} are cross-spectra and spectra at zero frequency of the processes U_{it}^{γ} and V_{it}. We allow for cross-sectional heterogeneity by allowing $f_i^{VU^\gamma}$ and f_i^{VV} to differ across i. Consequently, φ^{VU^γ} and φ^{VV} are weighted averages of the spectral quantities $f_i^{VU^\gamma}$ and f_i^{VV}. Our bias correction is based on the characterization that

$$\beta \equiv -\mathcal{I}^{-1}\Psi \tag{3}$$

where $\Psi \equiv \varphi^{VU^\gamma} - \varphi^{VV}$ and the existence of $\mathcal{I}^{-1} = \left(\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}$ is justified by Condition 7 in Section 4. We give an asymptotic justification for the approximate bias β in the next section.

If a reasonably precise estimator $\widehat{\beta}$ of β is available, we would expect that $\widehat{\widehat{\theta}} \equiv \widehat{\theta} - T^{-1}\widehat{\beta}$ would be less biased than the fixed effects estimator $\widehat{\theta}$. Hahn and Newey (2002) proposed a method of characterizing and estimating β for the static model when x_{it} is iid. over time. In the context of the binary panel model discussed above, this implies that their bias reduction may be used if $(z_{it}', e_{it})'$ is independent over time.

Unfortunately, this requires that even the explanatory variable z_{it} satisfies the independence restriction over time, which is expected to be violated in many applications. Thus, one important difference between the case considered in Hahn and Newey (2002) and our model is that all the elements of the second order term of the expansion are correlated across different time periods.

Our estimate of the asymptotic bias is based on sample analogs of $\mathcal{I}$ and Ψ. For this purpose, we construct sample analog estimators for $E\left[V_{it}^{\gamma_i}\right]$, $E\left[U_{it}^{\gamma_i\gamma_i}\right]$, $f_i^{VU^\gamma}$, and f_i^{VV}, and plug them into the expression for β. We will use the hat notation to denote these sample analogs.[1] Natural estimators for $E\left[V_{it}^{\gamma_i}\right]$ and $E\left[U_i^{\gamma_i\gamma_i}\right]$ are then given by

$$\widehat{E}\left[V_{it}^{\gamma_i}\right] \equiv \frac{1}{T}\sum_{t=1}^{T}\widehat{V}_{it}^{\gamma_i}, \quad \widehat{E}\left[U_i^{\gamma_i\gamma_i}\right] \equiv T^{-1}\sum_{t=1}^{T}\widehat{U}_{it}^{\gamma_i\gamma_i},$$

The estimators for spectral quantities $f_i^{VU^\gamma}$ and f_i^{VV} are given by

$$\widehat{f}_i^{VU^\gamma} \equiv \sum_{l=-m}^{m}\widehat{\Gamma}_{il}^{VU^\gamma}, \quad \widehat{f}_i^{VV} \equiv \sum_{l=-m}^{m}\widehat{\Gamma}_{il}^{VV},$$

where

$$\widehat{\Gamma}_{il}^{VU^\gamma} \equiv T^{-1}\sum_{t=\max(1,l)}^{\max(T,T+l)}\widehat{V}_{it}\widehat{U}_{it-l}^{\gamma_i}, \quad \widehat{\Gamma}_{il}^{VV} \equiv T^{-1}\sum_{t=\max(1,l)}^{\max(T,T+l)}\widehat{V}_{it}\hat{V}_{it-l}.$$

The parameter m is a bandwidth parameter that needs to be chosen such that $m/T^{1/2} \to 0$ as $T \to \infty$. We thus estimate β by

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

and our bias corrected estimator is given by

$$\widehat{\widehat{\theta}} \equiv \widehat{\theta} - \frac{1}{T}\widehat{\beta}$$

For many microeconometric applications, T is relatively small so a natural choice for m would be 1 in practice, which is the bandwidth considered in our Monte Carlo experiment reported in the next section. Our Monte Carlo experiments also indicate that the bias of the MLE is quite small even with moderately large T, which is further justification to focus on cases where T is so small that $m = 1$ is the only reasonable choice for the bandwidth. Nevertheless, when T is moderately large, which is unlikely in many applications, bandwidth selection methods developed in the time series literature could in principle be used here. The literature on spectral density estimation has focused on three main methods. Kernel smoothing and optimal bandwidth selection was considered by Parzen (1957). The special case of spectral estimation at zero frequency was further analyzed by Newey and West (1987, 1994), Andrews (1991) and Andrews and Monahan (1992). Methods for selecting the bandwidth discussed in these papers are based on minimizing the approximate mean squared error (MSE) as a function of m and on plugging this optimal choice of m into the formula for the spectral density. It follows immediately that our Conditions 3 and 4 discussed in Section 4 imply that Assumption 2 of Newey and West (1994) holds uniformly in i. This suggests that a formal theory of bandwidth choice could be based on selecting $\hat{m}_i$ optimally for each individual time series. This choice is not

[1] Precise definitions can be found in Appendix A.

likely to be optimal for Ψ however, and additional work would be needed to establish uniform convergence of any choice of m that is individual specific. For the aforementioned reasons we do not believe that it is worth while to develop these results in more detail.

A second approach to spectral density estimation is the cross validation procedure of Hurvich (1985), Beltrao and Bloomfield (1987), Robinson (1991) and Velasco (2000) which has the advantage that it does not rely on additional nuisance parameters such as formulas of the approximate MSE that are needed for the plug in procedures. Cross validation is thus fully automatic which is not the case for the plug in methods where the best currently available procedure by Newey and West (1994) still needs a non-automatic bandwidth choice for the estimation of the nuisance parameters. Currently, cross-validation for a single frequency is only considered by Velasco (2000) under conditions that are more restrictive than ours. This indicates that more research would be needed to adapt this procedure to our application. A third method for spectral density estimation is based on infinite order vector autoregressions (VAR(∞)) as flexible parametric forms of the covariance structure. Such methods were proposed by Akaike (1969). Integrated MSE optimal selection of approximating VAR(h) models was analyzed by Shibata (1981) but fully automatic versions implementing this selection procedure do not seem to exist in the literature. Recently, fully automatic versions of asymptotically unbiased VAR(∞) estimators were obtained by Kuersteiner (2004). In principle, VAR approximations could be adapted to our context but this is beyond the scope of this paper.

Our formulations for estimators of $f_i^{VU^\gamma}$and f_i^{VV} rely on the truncated kernel. Newey and West (1994) argue that the choice of the kernel function is of secondary importance in the performance of statistical tests based on spectral estimates and we conjecture that this is even more so in our application. When spectral densities such as $f_i^{VU^\gamma}$and f_i^{VV} are estimated in the context of constructing confidence regions or test statistics, one needs to guarantee that the estimators are positive definite matrices. This is typically achieved by choosing appropriate kernel functions as pointed out by Newey and West (1987) and Andrews (1991). In the current context of bias correction, positivity of the estimates is of no concern and the main motivation for using any kernel other than a truncated kernel disappears. On the other hand, plug-in methods for bandwidth selection are mostly formulated for kernel functions other than the truncated kernel. If such automated procedures were to be used then a kernel based estimate may be preferable even in our context. As explained earlier, these problems can be neglected for the sample sizes we think are most relevant to the bias correction problem.

3 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

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

By generating ε_{it} from two different distributions, we generate panel Probit and Logit models. Robert and Tybout (1997) employed a panel Probit model, in which the coefficient of the lagged dependent variable

measured the importance of sunk costs in an empirical model of entry.[2] We generate the strictly exogenous regressor z_{it} such that it is iid. $N\left(0, I_{\dim(z_{it})}\right)$ for Probit models, and $N\left(0, \left(\pi^2/3\right) \cdot I_{\dim(z_{it})}\right)$ for Logit models. Our $\widehat{\theta}$ is the maximum likelihood estimator.

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 MLE with the bias corrected MLE for the bandwidth choice $m = 1$. The results are reported in Table 1. We can see that the bias corrected MLE removes about half of the bias and RMSE. To investigate the robustness of our bandwidth selection we compare the performance of the bias corrected estimator for alternative choices of m where we set $m = 0$ or $m = 2$. We find that $m = 1$ gives the best results over all in terms of bias reduction and RMSE.

For the panel Logit model with fixed effects, Honoré and Kyriazidou (2000) proposed an estimator that is consistent under large n, fixed T asymptotics. We therefore compare their estimator to the MLE, and the bias corrected MLE with $m = 1$. As in the panel Probit case, the bias corrected MLE removes about half of the bias and RMSE. When $\dim(z_{it}) = 1$, our bias corrected estimator is slightly inferior to Honoré and Kyriazidou's estimator. The latter is known to have a slower rate of convergence when $\dim(z_{it})$ is large. We therefore compare the properties of their estimator with ours when $\dim(z_{it}) = 2$. We find that, when the dimension of z_{it} is as small as 2, our bias corrected estimator strictly dominates Honoré and Kyriazidou's (2000). We conjecture that the poor performance of the Honoré and Kyriazidou estimator is due to the fact that the z_{it}'s are continuously distributed.[3]

It should be pointed out that the initial observation y_{i0} was generated in such a way that the stationarity assumption is violated. Monte Carlo evidence therefore suggests that our bias correction is robust to mild violations of the stationarity assumption.

4 Asymptotic Theory

We assume the following:

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* $\widehat{G}_{(i)}(\theta, \gamma_i) \equiv T^{-1}\sum_{t=1}^{T} \psi(x_{it}; \theta, \gamma_i)$ *and* $G_{(i)}(\theta, \gamma_i) \equiv E[\psi(x_{it}; \theta, \gamma_i)]$.

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

Condition 3 *Suppose that, for each* i, $\{x_{it}, t = 1, 2, \ldots\}$ *is a stationary mixing sequence. Let* $\mathcal{A}_t^i = \sigma(x_{it}, x_{it-1}, x_{it-2}, ...)$, $\mathcal{B}_t^i = \sigma(x_{it}, x_{it+1}, x_{it+2}, ...)$ *and* $\alpha_i(m) = \sup_t \sup_{A \in \mathcal{A}_t^i, B \in \mathcal{B}_{t+m}^i} |P(A \cap B) - P(A)P(B)|$. *Assume that* $\sup_i |\alpha_i(m)| \leq Ca^m$ *for some* a *such that* $0 < a < 1$ *and some* $C > 0$. *We assume that* $\{x_{it}, t = 1, 2, 3, ...\}$ *are independent across* i.

[2] It should be pointed out that they assumed the random effects model. Because there did not exist any estimator that dealt with the incidental parameters problem with panel Probit models, it is perhaps unavoidable that they had to assume this particular structure for the individual specific effects.

[3] The method Honoré and Kyriazidou 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 .1 to 8192 but found no significant effect on the performance of the estimator.

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 $\mathbb{R}^p$ and $p\equiv \dim\phi = R+1$. Let $\nu=(\nu_1,...,\nu_k)$ be a vector of non-negative integers v_i, $|v|=\sum_{j=1}^k v_j$ and $D^v\psi(x_{it},\phi)=\partial^{|\nu|}\psi(x_{it},\phi)/\left(\partial\phi_1^{v_1}...\partial\phi_k^{\nu_k}\right)$. Assume that there exists a function $M(x_{it})$ such that $|D^v\psi(x_{it},\phi_1)-D^v\psi(x_{it},\phi_2)|\leq M(x_{it})\|\phi_1-\phi_2\|$ for all $\phi_1,\phi_2\in\Phi$ and $|v|\leq 5$. We also assume that $M(x_{it})$ satisfies $\sup_{\phi\in\Phi}\|D^v\psi(x_{it},\phi)\|\leq M(x_{it})$ and $\sup_i E\left[|M(x_{it})|^{10q+12+\delta}\right]<\infty$ for some integer $q\geq p/2+2$ and for some $\delta>0$.*

Condition 5 $\inf_i \inf_T \lambda_{iT}>0$, *where λ_{iT} is the smallest eigenvalue of* $\Sigma_{iT}=\operatorname{Var}\left(T^{-1/2}\sum_{t=1}^T U_i(x_{it};\theta,\gamma_i)\right)$.

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

Condition 7 *Let $\mu_{i1}\leq ...\leq \mu_{ik}\leq ...\leq \mu_{iR}$ 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.*

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 $\{x_{it}, t=1,2,\ldots\}$ as well as the moments of $D^v\psi$. Condition 3 also imposes stationarity, which rules out time-dummies. 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 only mild assumptions on the behavior of x_{it} across i. In particular, we impose a uniform upperbound on the decay rate of temporal dependence without requiring that dependence be homogeneous across i. 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.

Example 1 cont.: From Newey and McFadden (1994, p.2125) it follows that the Probit likelihood has a unique maximum and $E[|\psi(x_{it};\theta,\gamma_i)|]<\infty$ if $E[w_{it}w_{it}']$ is positive definite where $w_{it}=(1,z_{it}')'$. 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 Lipshitz condition of Condition 4. For the 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 $D^v\left[(z_{it}'\phi)^{|v|}\right]g_{|v|}(z_{it}'\phi)$ with $g_{|v|}(\cdot)$ a uniformly bounded function. The moment restrictions of Condition 4 are then satisfied if $\sup_i E\left[\left|z_{it}^j\right|^{|v|+10q+12+\delta}\right]<\infty$ for $j=1,...,R$ where z_{it}^j is the j-th 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 $E[w_{it}w_{it}']$ is positive definite uniformly in i by the results of Newey and McFadden (1994, p.2147 and Footnote 29).

Example 2 cont.: De Jong and Woutersen (2003) discuss conditions for consistency which are analogous to the case in Example 1 such that for $w_{it} = (1, y_{it-1}, z'_{it})'$ we need $E[w_{it}w'_{it}]$ 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 iid. Then, by de Jong and Woutersen (2003, Theorem 2), it follows that (y_{it}, z_{it}) is strong mixing with strong mixing coefficients $\beta_i = C_0(C_1 \exp(-C_2 m) + \alpha_i(m))$ for positive constants C_0, C_1, C_2 and can be assumed to be strictly stationary. The moment bounds are satisfied as before if $\sup_i E\left[\left|z_{it}^j\right|^{|v|+10q+12+\delta}\right] < \infty$ and the remaining conditions hold under the same assumptions as previously discussed.

Example 3 cont.: Identification as in Condition 1 follows if $(\beta_0, \tau_0, \gamma_{i0}, \sigma^2_{i\varepsilon}) \in \operatorname{int}\Theta$, $E[w_{it}w'_{it}]$ is positive definite for $w_{it} = (1, y_{it-1}, z'_{it})'$ and ε_{it} is iid Gaussian from Amemiya (1973) and Olsen (1978). In this example it is not clear whether the mixing property can be established as before. We use it to illustrate the role that the mixing condition plays in our proofs and how it can be relaxed. Inspection of our proofs shows that the mixing condition is only used as a convenient way to summarize the decay pattern of autocovariances of various nonlinear functions of x_{it}. In particular, our results depend on the moment inequality of Hall and Heyde (1980, Corollary A.2). Assume that $|\tau_0| < 1$, ε_{it} is iid across i and t with $\varepsilon_{it} \sim N(0, \sigma^2_\varepsilon)$ and z_{it} is strictly stationary and mixing with $\sup_i |\alpha_i(m)| \leq Ca^m$ for some a such that $0 < a < 1$ and some $C > 0$ and $\sup_i E[|z_{it}|^r] < \infty$ for some $r > 7 + 10q + 12 + \delta$ with $q \geq p/2 + 2$ and for some $\delta > 0$. Let y_{it} and $\psi(x_{it}; \theta, \gamma_i)$ be as defined before in Example 3. Then, Condition 4 is satisfied, x_{it} can be assumed to be strictly stationary and for $|v| \leq 5$ autocovariances of $D^v\psi(x_{it}, \phi)$ decay exponentially uniformly for $\phi \in \operatorname{int}\Theta$.[4] Conditions 5, 6 and 7 hold if $E[w_{it}w'_{it}]$ is positive definite uniformly in i by the results of Amemiya (1973).

It can be shown that the parameter estimates are consistent under the alternative asymptotics: For every $\eta > 0$,

$$\Pr\left[\left|\widehat{\theta} - \theta_0\right| \geq \eta\right] = o\left(T^{-1}\right) \tag{4}$$

and

$$\Pr\left[\max_{1 \leq i \leq n} |\widehat{\gamma}_i - \gamma_{i0}| \geq \eta\right] = o\left(T^{-1}\right). \tag{5}$$

See Appendix C for a proof. Although these consistency results are not directly useful in understanding the asymptotic bias of $\widehat{\theta}$ discussed below, establishing uniform consistency for the parameter estimates is the key to constructing consistent estimates of the asymptotic bias. This is because the asymptotic bias implicitly depends on $\widehat{\theta}$ and $\hat{\gamma}_i$, and the natural estimator using sample analogs requires substituting $\widehat{\theta}$ and $\widehat{\gamma}_i$ for θ_0 and γ_{i0}. Uniform consistency of the fixed effects can also be useful when estimates of the fixed effects are necessary to evaluate the average partial effects of z_{it} on y_{it} in the panel Probit example.

We now present a theoretical justification of the asymptotic bias formula, which is based on a higher order expansion. The expansion is defined in terms of functional derivatives of the population distribution of x_{it} in the direction of the empirical distribution function of the sample. To describe the expansion let $F \equiv (F_1, \ldots, F_n)$ denote the collection of (marginal) distribution functions of x_{it} and let

[4] A proof is available upon request.

$\widehat{F} \equiv \left(\widehat{F}_1, \ldots, \widehat{F}_n\right)$, where $\widehat{F}_i$ denotes the empirical distribution function for the i-th observation. Define $F(\epsilon) \equiv F + \epsilon\sqrt{T}\left(\widehat{F} - F\right)$ 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 $0 = \int V_i\left[\theta, \gamma_i(\theta, F_i(\epsilon))\right] dF_i(\epsilon)$, and let $\theta(F(\epsilon))$ be the solution to the estimating equation $0 = \sum_{i=1}^n \int U_i\left(x_{it}; \theta(F(\epsilon)), \gamma_i\left(\theta(F_i(\epsilon)), F_i(\epsilon)\right)\right) dF_i(\epsilon)$. By a Taylor series expansion, we have

$$\theta\left(\widehat{F}\right) - \theta(F) = \frac{1}{\sqrt{T}}\theta^{\epsilon}(0) + \frac{1}{2}\left(\frac{1}{\sqrt{T}}\right)^2 \theta^{\epsilon\epsilon}(0) + \frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^3 \theta^{\epsilon\epsilon\epsilon}\left(\widetilde{\epsilon}\right), \tag{6}$$

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 $\widetilde{\epsilon}$ is somewhere in between 0 and $T^{-1/2}$. Lemma 2 in the Appendix allows us to ignore the last term. We will therefore work with the expansion

$$\sqrt{nT}\left(\theta\left(\widehat{F}\right) - \theta(F)\right) = \sqrt{nT}\frac{1}{\sqrt{T}}\theta^{\epsilon}(0) + \sqrt{nT}\frac{1}{2}\left(\frac{1}{\sqrt{T}}\right)^2 \theta^{\epsilon\epsilon}(0) + o_p(1). \tag{7}$$

The term

$$\sqrt{n}\theta^{\epsilon}(0) = \left(\frac{1}{n}\sum_{i=1}^n \mathcal{I}_i\right)^{-1}\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^n\sum_{t=1}^T U_{it}\right)$$

is the efficient score for θ evaluated at γ_{i0}. It admits an asymptotically normal distribution centered at zero because it is essentially an average over independent, mean zero random variables. It is shown in Appendix C that the second order term in our expansion takes the form

$$\frac{1}{2}\sqrt{n/T}\theta^{\epsilon\epsilon}(0) = -\sqrt{\frac{n}{T}}\left(\frac{1}{n}\sum_{i=1}^n \mathcal{I}_i\right)^{-1}\frac{1}{n}\sum_{i=1}^n\left[\frac{\sum_{t=1}^T V_{it}}{\sqrt{T}E\left[V_{it}^{\gamma_i}\right]}\right]\left[\frac{1}{\sqrt{T}}\sum_{t=1}^T\left(U_{it}^{\gamma_i} - \frac{E\left[U_{it}^{\gamma_i\gamma_i}\right]}{2E\left[V_{it}^{\gamma_i}\right]}V_{it}\right)\right] + o_p(1).$$

It turns out that under our asymptotics where n and T tend to infinity jointly, the first term on the right of (7) determines the asymptotic distribution of the estimator, while the second term turns out to be a pure bias term. In the proof of Theorem 1 in the Appendix, it is shown that

$$\frac{1}{\sqrt{n}\sqrt{T}}\sum_{i=1}^n\sum_{t=1}^T U_{it} \to N(0, \Omega), \tag{8}$$

where $\Omega \equiv \lim_n n^{-1}\sum_{i=1}^n \Sigma_{iT}$ and $\Sigma_{iT} \equiv \text{Var}\left(T^{-1/2}\sum_{t=1}^T U_i(x_{it}; \theta, \gamma_i)\right)$. It may be interesting to point out that the Central Limit result in (8) is based on the independence across individuals i of the triangular array $v_{iT} = T^{-1/2}\sum_{t=1}^T U_i(x_{it}; \theta, \gamma_i)$ where v_{iT} has uniformly (in i and T) bounded variance Σ_{iT}. In other words, the mixing and stationarity properties of the process x_{it} do not play a major role in this result, except to guarantee the uniform boundedness of Σ_{iT}. We also point out that we do not need to impose homogeneity of the distribution of x_{it} across individuals, as long as moments are uniformly bounded across i.

In the same way we show that

$$\frac{1}{n}\sum_{i=1}^n\left[\frac{\sum_{t=1}^T V_{it}}{\sqrt{T}E\left[V_{it}^{\gamma_i}\right]}\right]\left[\frac{1}{\sqrt{T}}\sum_{t=1}^T\left(U_{it}^{\gamma_i} - \frac{E\left[U_i^{\gamma_i\gamma_i}\right]}{2E\left[V_{it}^{\gamma_i}\right]}V_{it}\right)\right] = \Psi + o_p(1). \tag{9}$$

In this case the stationarity assumptions for x_{it} do help to simplify the form of Ψ but it is clear that more general results could be obtained without this restriction.

While stationarity assumptions do not play a crucial role in the representation of our main result in Theorem 1 they are used in bounding the error terms in the approximation of the first order condition.

Our results rely on an extension of a Lemma by Lahiri (1992) where we use the stationarity assumption to simplify the argument. It is possible that a different proof strategy could be used to relax this restriction but this question will be left for future research.

The asymptotic distribution of the fixed effects estimator $\widehat{\theta}$ is obtained by combining (8) and (9):

Theorem 1 *Assume that Conditions 1, 2, 3, 4, 5, 6, and 7 hold. Then*

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

where $\beta \equiv -\mathcal{I}^{-1}\Psi$.

Proof. See Appendix C. ■

We can further show that the bias corrected estimator removes the asymptotic bias:

Theorem 2 *Assume that Conditions 1, 2, 3, 4, 5 and 7 hold. Let* $m, T \to \infty$ *such that* $m/T^{1/2} \to 0$. *Then,* $\sqrt{nT}\left(\widehat{\widehat{\theta}}-\theta\right) \to N\left(0, \mathcal{I}^{-1}\Omega\mathcal{I}^{-1}\right)$, *where* $\widehat{\widehat{\theta}} \equiv \widehat{\theta} - \frac{1}{T}\widehat{\beta}$.

Proof. See Appendix C. ■

5 Summary

In this paper, we provide a simple characterization of the asymptotic bias of a fixed effects estimator for dynamic nonlinear panel models with fixed effects. The asymptotic bias was based on the "large n, large T" asymptotics adopted by, e.g., Hahn and Kuersteiner (2002) and Hahn and Newey (2002). A method of reducing bias based on these expansions was developed.

The method we propose is quite flexible and our examples show that it can be applied to a variety of, mainly parametric, non-linear models. For some of these models there do not seem to exist any adequate alternative estimators which are less biased than the MLE in finite samples. The dynamic panel probit model is an example of such a model. We investigate the quality of our bias correction for this particular example and the closely related dynamic panel Logit model. For the latter there does exist an alternative estimator by Honoré and Kyriazidou (2000). We expect our procedure to perform relatively well compared to this particular alternative procedure in situations where there is a reasonable amount of time series variation in the data, the dimension of the regressor space is relatively large and at least some of the regressors are continuously distributed. We conduct a Monte Carlo experiment to shed some light on the data-requirements for our method to perform well.

As is apparent from the previous discussion, the quality of our approximation may be poor in some models or for certain type of data sets, especially when T is much smaller than in our Monte Carlo experiments. For linear models there are a number of alternative approaches to the estimation of models with individual specific fixed effects. One example of a method with desirable finite sample properties is the quasi MLE developed for panel vector autoregressive models by Binder, Hsiao, and Pesaran (2000). It might be interesting to develop a nonlinear analog of this procedure.

Appendix

A Sample Analogs

$$\begin{aligned}
\widehat{V}_{it} &\equiv \frac{\partial \psi\left(x_{it};\widehat{\theta},\widehat{\gamma}_i\right)}{\partial \gamma_i}, \quad \widehat{V}_{it}^{\gamma_i} \equiv \frac{\partial^2 \psi\left(x_{it};\widehat{\theta},\widehat{\gamma}_i\right)}{\partial \gamma_i^2}, \\
\widehat{\rho}_i &\equiv \left(T^{-1}\sum_{t=1}^{T}\frac{\partial^2 \psi\left(x_{it};\widehat{\theta},\widehat{\gamma}_i\right)}{\partial \gamma_i^2}\right)^{-1}\left(T^{-1}\sum_{t=1}^{T}\frac{\partial^2 \psi\left(x_{it};\widehat{\theta},\widehat{\gamma}_i\right)}{\partial \theta \partial \gamma_i}\right), \\
\widehat{U}_{it} &\equiv \frac{\partial \psi\left(x_{it};\widehat{\theta},\widehat{\gamma}_i\right)}{\partial \theta} - \widehat{\rho}_i \cdot \frac{\partial \psi\left(x_{it};\widehat{\theta},\widehat{\gamma}_i\right)}{\partial \gamma_i}, \\
\widehat{U}_{it}^{\gamma_i} &\equiv \frac{\partial^2 \psi\left(x_{it};\widehat{\theta},\widehat{\gamma}_i\right)}{\partial \gamma_i \partial \theta} - \widehat{\rho}_i \cdot \frac{\partial^2 \psi\left(x_{it};\widehat{\theta},\widehat{\gamma}_i\right)}{\partial \gamma_i^2}, \\
\widehat{U}_{it}^{\gamma_i\gamma_i} &\equiv \frac{\partial^3 \psi\left(x_{it};\widehat{\theta},\widehat{\gamma}_i\right)}{\partial \gamma_i^2 \partial \theta} - \widehat{\rho}_i \cdot \frac{\partial^3 \psi\left(x_{it};\widehat{\theta},\widehat{\gamma}_i\right)}{\partial \gamma_i^3}.
\end{aligned}$$

B Auxiliary Lemmas

We present a few Lemmas needed in the proof of the main results of Section 4. All the proofs are available upon request.

Lemma 1 *For all $\eta > 0$ it follows that*

$$\Pr\left[\max_{1\leq i\leq n}\sup_{(\theta,\gamma)}\left|\widehat{G}_{(i)}(\theta,\gamma) - G_{(i)}(\theta,\gamma)\right| \geq \eta\right] = o\left(T^{-1}\right)$$

Lemma 2 $\Pr\left[\max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}}\left|\theta^{\epsilon\epsilon\epsilon}(\epsilon)\right| > C\left(T^{\frac{1}{10}-\upsilon}\right)^3\right] = o\left(T^{-1}\right)$ *for some constant $C > 0$ and $0 < \upsilon < (100q+120)^{-1}$.*

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 $E[\xi(x_{it},\phi_i)] = 0$. Moreover $\sup_\phi \|\xi(x_{it},\phi)\| \leq \mathbf{M}(x_{it})$ for some $\mathbf{M}(x_{it})$ such that $E\left[\mathbf{M}(x_{it})^4\right] < \infty$. Let $\Sigma_{nT} = \sum_{i=1}^n \Sigma_{iT}^{\xi\xi}$ with $\Sigma_{iT}^{\xi\xi} = \operatorname{Var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T \xi(x_{it},\phi_i)\right)$. 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 $f^{\xi\xi} \equiv \lim n^{-1}\sum_{i=1}^n f_i^{\xi\xi}$ with $f_i^{\xi\xi} \equiv \sum_{j=-\infty}^{\infty} E\left[\xi(x_{it},\phi_i)\,\xi(x_{it-j},\phi_i)'\right]$.*

Lemma 4 *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 $E[\xi(x_{it},\phi)] = 0$ for all i,t and $\phi \in \Phi$. Assume that there exists a function $\mathbf{M}(x_{it})$ such that $|\xi(x_{it},\phi_1) - \xi(x_{it},\phi_2)| \leq \mathbf{M}(x_{it})\|\phi_1 - \phi_2\|$ for all $\phi_1,\phi_2 \in \Phi$ and $\sup_\phi|\xi(x_{it},\phi)| \leq \mathbf{M}(x_{it})$. For each i, let x_{it} be a α-mixing process with exponentially decaying mixing coefficients $\underline{\alpha}_i(m)$ satisfying $\sup_i|\underline{\alpha}_i(m)| \leq Ca^m$ for some a such that $0 < a < 1$ and some $0 < C < \infty$. Let q denote a positive integer such that $q \geq \frac{p+4}{2}$,*

where $p = \dim \phi$. *We also assume that* $E\left[\left|\mathbf{M}\left(x_{it}\right)\right|^{10q+12+\delta}\right] < \infty$ *for some* $\delta > 0$. *Finally, assume that* $n = O\left(T\right)$. *We then have* $\Pr\left[\max_i \left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi\left(x_{it},\phi_i\right)\right| > T^{\frac{1}{10}-\upsilon}\right] = o\left(T^{-1}\right)$ *for* $0 < \upsilon < \left(100q+120\right)^{-1}$. *Here,* $\{\phi_i\}$ *is an arbitrary nonstochastic sequence in* Φ.

Lemma 5 *Let Conditions 1, 2, 3, 4 and 5 be satisfied. Then* $\Pr\left[\max_i \sup_{\epsilon\in\left[0,1/\sqrt{T}\right]}\left|\widehat{\gamma}_i^{\epsilon}\left(\epsilon\right)\right| > T^{\frac{1}{10}-\upsilon}\right] = o\left(T^{-1}\right)$, $\Pr\left[\max_i \left|\widehat{\gamma}_i^{\epsilon}\left(0\right)\right| > T^{\frac{1}{10}-\upsilon}\right] = o\left(T^{-1}\right)$, $\Pr\left[\max_i \sup_{\epsilon\in\left[0,1/\sqrt{T}\right]}\left|\widehat{\gamma}_i^{\epsilon\epsilon}\left(\epsilon\right)\right| > \left(T^{\frac{1}{10}-\upsilon}\right)^2\right] = o\left(T^{-1}\right)$, *and* $\Pr\left[\max_i \left|\sqrt{T}\left(\widehat{\gamma}_i - \gamma_{i0}\right)\right| > T^{1/10-\upsilon}\right] = o\left(T^{-1}\right)$ *for* $0 < \upsilon < \left(100q+120\right)^{-1}$.

Lemma 6 *Let* $k_{it} = k\left(x_{it};\theta_0,\gamma_{i0}\right)$ *and* $\widehat{k}_{it} = k\left(x_{it};\widehat{\theta},\widehat{\gamma}_i\right)$ *where* x_{it} *satisfies Condition 3,* k_{it} *satisfies Condition 4 and* $\widehat{\theta}$, $\widehat{\gamma}_i$ *are defined in (1). Assume that* $E\left[k_{it}\right] = 0$ *for* i,t. *Let* A_i *be conformable matrix of constants such that* $\max_i \left\|A_i\right\| < \infty$. *Let* $f_i^{kk} = \sum_{l=-\infty}^{\infty} E\left[k_{it}k_{it-l}'\right]$ *and* $f^{kk} = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} A_i f_i^{kk}$. *Then,* $\frac{1}{n}\sum_{i=1}^{n} A_i\left(\frac{1}{T}\sum_{l=-m}^{m}\sum_{t=\max(1,l)}^{\min(T,T+l)}\widehat{k}_{it}\widehat{k}_{it-l}'\right) - f^{kk} = o_p(1)$, *where* $m, T \to \infty$ *such that* $m = o\left(T^{1/2}\right)$.

C Proof of Main Results

Proposition 1 *Assume that Conditions 1, 2, 3, 4, 5, 6, and 7 hold. Then, for every* $\eta > 0$, $\Pr\left[\left|\widehat{\theta} - \theta_0\right| \geq \eta\right] = o\left(T^{-1}\right)$ *and* $\Pr\left[\max_{1\leq i\leq n}\left|\widehat{\gamma}_i - \gamma_{i0}\right| \geq \eta\right] = o\left(T^{-1}\right)$.

Proof. The result follows by standard arguments from Lemma 1. ■

Proof of Theorem 1:. We first obtain a Taylor series expansion in ϵ for $\theta\left(F\left(\epsilon\right)\right)$ as defined in Section 4. By Lemma 2 we only need to consider the first two terms in the Taylor expansion of $\theta\left(F\left(\epsilon\right)\right)$ formally stated in equation (6). Let

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

such that $\theta\left(F\left(\epsilon\right)\right)$ solves the first order condition which may be written as

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

Differentiating repeatedly with respect to ϵ, we obtain

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

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

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

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

Evaluating (12) at $\epsilon = 0$, and noting that $E\left[U_i^{\gamma_i}\right] = 0$, we solve[5] for $\theta^{\epsilon}\left(0\right) = d\theta\left(F\left(\epsilon\right)\right)/d\epsilon$ such that

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

[5] More detailed steps of this derivation are available on request.

Using Lemma 3, we show that

$$\frac{1}{\sqrt{n}\sqrt{T}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}\rightarrow N\left(0,\Omega\right)$$

and

$$\sqrt{nT}\frac{1}{\sqrt{T}}\theta^{\epsilon}\left(0\right)=\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_{i}\right)^{-1}\left(\frac{1}{\sqrt{n}\sqrt{T}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{i}\right)\rightarrow N\left(0,\mathcal{I}^{-1}\Omega\left(\mathcal{I}'\right)^{-1}\right)$$

We now turn to the analysis of the term $\theta^{\epsilon\epsilon}\left(0\right)=d^{2}\theta\left(F\left(\epsilon\right)\right)/\left(d\epsilon\right)^{2}$ which depends on the estimator for $\hat{\gamma}_{i}$ and reflects the impact of incidental parameters on the bias of $\hat{\theta}$. In order to evaluate this impact we need to analyze the first order conditions for $\hat{\gamma}_{i}$. In the ith observation, $\gamma_{i}\left(\theta,F_{i}\left(\epsilon\right)\right)$ solves the estimating equation

$$\int V_{i}\left(\cdot;\theta,\gamma_{i}\left(\theta,F_{i}\left(\epsilon\right)\right)\right)dF_{i}\left(\epsilon\right)=0 \tag{16}$$

Differentiating the LHS with respect to θ and ϵ, we obtain

$$\begin{aligned}0 &= \int\frac{\partial V_{i}\left(\cdot,\theta,\epsilon\right)}{\partial\theta}dF_{i}\left(\epsilon\right)+\left(\int\frac{\partial V_{i}\left(\cdot,\theta,\epsilon\right)}{\partial\gamma_{i}}dF_{i}\left(\epsilon\right)\right)\frac{\partial\gamma_{i}\left(\theta,F_{i}\left(\epsilon\right)\right)}{\partial\theta},\\ 0 &= \left(\int\frac{\partial V_{i}\left(\cdot,\theta,\epsilon\right)}{\partial\gamma_{i}}dF_{i}\left(\epsilon\right)\right)\frac{\partial\gamma_{i}\left(\theta,F_{i}\left(\epsilon\right)\right)}{\partial\epsilon}+\int V_{i}\left(\cdot,\theta,\epsilon\right)d\Delta_{iT}.\end{aligned}$$

Equating these equations to zero and solving for derivatives of γ_{i} evaluated at $\epsilon=0$ gives

$$\gamma_{i}^{\theta} = -\frac{E\left[\frac{\partial V_{i}}{\partial\theta}\right]}{E\left[\frac{\partial V_{i}}{\partial\gamma_{i}}\right]}, \tag{17}$$

$$\gamma_{i}^{\epsilon} = -\frac{\sum_{t=1}^{T}V_{it}}{\sqrt{T}E\left[\frac{\partial V_{i}}{\partial\gamma_{i}}\right]}=-\frac{T^{-1/2}\sum_{t=1}^{T}V_{it}}{E\left[\frac{\partial V_{i}}{\partial\gamma_{i}}\right]}, \tag{18}$$

where $\gamma_{i}^{\theta}\equiv\frac{\partial\gamma_{i}(\theta,F_{i}(0))}{\partial\theta}$, and $\gamma_{i}^{\epsilon}\equiv\frac{\partial\gamma_{i}(\theta,F_{i}(0))}{\partial\epsilon}$.

Now, evaluating each term of (13) at $\epsilon=0$, and noting that $E\left[U_{i}^{\gamma_{i}}\right]=0$, we obtain

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

where

$$\mathcal{G}\equiv\left[\begin{array}{ccc}\theta^{\epsilon}\left(0\right)'\left(\frac{1}{n}\sum_{i=1}^{n}E\left[\frac{\partial^{2}U_{i}^{(1)}}{\partial\theta\partial\theta'}\right]\right)\theta^{\epsilon}\left(0\right) & \cdots & \theta^{\epsilon}\left(0\right)'\left(\frac{1}{n}\sum_{i=1}^{n}E\left[\frac{\partial^{2}U_{i}^{(R)}}{\partial\theta\partial\theta'}\right]\right)\theta^{\epsilon}\left(0\right)\end{array}\right]'$$

from which we obtain

$$\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)\theta^{\epsilon\epsilon}(0)=\frac{1}{n}\sum_{i=1}^{n}(\gamma_i^{\epsilon})^2\cdot E\left[\frac{\partial^2 U_i}{\partial\gamma_i^2}\right]+\frac{2}{n}\sum_{i=1}^{n}\gamma_i^{\epsilon}\cdot\int\frac{\partial U_i}{\partial\gamma_i}d\Delta_{iT}+\mathcal{T} \tag{19}$$

for some $\mathcal{T}=O_p\left(n^{-1}\right)$. (Proof is available upon request.) It follows that

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

It remains to establish the limiting behavior of $\theta^{\epsilon\epsilon}(0)$. Let $Z_{iT}\equiv\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{V_{it}}{E\left[\frac{\partial V_i}{\partial\gamma_i}\right]}\right]\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(U_i^{\gamma_i}-\frac{E\left[U_i^{\gamma_i\gamma_i}\right]}{2E\left[\frac{\partial V_i}{\partial\gamma_i}\right]}V_{it}\right)\right]$. Then, Z_{iT} are independent across i such that

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

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

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

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 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

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

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. 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

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

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

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

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

$$\Pr\left[\left|\frac{1}{n}\sum_{i=1}^{n}\left(Z_{iT} - E\left[Z_{iT}\right]\right)\right| > \eta\right] < \frac{\sup_i \text{Var}(Z_{iT})}{n\eta} \to 0$$

such that $\frac{1}{n}\sum_{i=1}^{n} Z_{iT} = \frac{1}{n}\sum_{i=1}^{n} E\left[Z_{iT}\right] + O_p(n^{-1/2})$. By applying Lemma 3 to $\xi = (V, U^{\gamma})$, 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 $\frac{1}{n}\sum_{i=1}^{n} E\left[Z_{iT}\right] \to \Psi$. Therefore, we have

$$\sqrt{nT}\frac{1}{2}\left(\frac{1}{\sqrt{T}}\right)^2 \theta^{\epsilon\epsilon}(0) = -\sqrt{\frac{n}{T}}\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\Psi + o_p(1)$$

and the result of the Theorem follows immediately upon combining the results for $\theta^{\epsilon}(0)$ and $\theta^{\epsilon\epsilon}(0)$. ■

Proof of Theorem 2:. 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 $\widehat{\gamma}_i(\epsilon) \equiv \text{argmax}_a \int \psi\left(x_{it}; \widehat{\theta}(\epsilon), a\right) dF_i(\epsilon)$. Using the same arguments as earlier, we are looking for the expansion $\widehat{\gamma}_i(\epsilon) - \gamma_{i0} = \frac{1}{\sqrt{T}}\widehat{\gamma}_i^{\epsilon}(0) + \frac{1}{2T}\widehat{\gamma}_i^{\epsilon\epsilon}(\tilde{\epsilon})$ for some $\tilde{\epsilon} \in \left[0, T^{-1/2}\right]$, which can be derived from the first order condition $0 = \int v_i(\cdot, \epsilon)\, dF_i(\epsilon)$, where $v_i(\cdot, \epsilon) \equiv V_i\left(\theta\left(F(\epsilon)\right), \gamma_i\left(F_i(\epsilon)\right)\right)$. Considering the expansion for $\widehat{\gamma}_i(\epsilon) - \gamma_{i0}$ is useful in establishing uniform rates of convergence for $\hat{\gamma}_i$ across i which in turn are needed to prove that $\hat{\beta} - \beta = o_p(1)$. Differentiating the first order condition repeatedly with respect to ϵ, we obtain

$$0 = \int \frac{dv_i(\cdot, \epsilon)}{d\epsilon} dF_i(\epsilon) + \int v_i(\cdot, \epsilon)\, d\Delta_{iT} \tag{20}$$

$$0 = \int \frac{d^2 v_i(\cdot, \epsilon)}{d\epsilon^2} dF_i(\epsilon) + 2\int \frac{dv_i(\cdot, \epsilon)}{d\epsilon} d\Delta_{iT} \tag{21}$$

From (20), we obtain

$$\widehat{\gamma}_i^{\epsilon}(0) = -\left(E\left[V_i^{\gamma}\right]\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial\psi\left(x_{it}; \theta_0, \gamma_{i0}\right)}{\partial\gamma} + E\left[V_i^{\theta}\right]\theta^{\epsilon}(0)\right) \tag{22}$$

$$\widehat{\gamma}_i^{\epsilon}(\epsilon) = -\left(\int \frac{\partial v_i(\cdot, \epsilon)}{\partial\gamma_i} dF_i(\epsilon)\right)^{-1}\left[\int\left(\frac{\partial v_i(\cdot, \epsilon)}{\partial\theta'}\right) dF_i(\epsilon)\,\theta^{\epsilon}(\epsilon) + \int v_i(\cdot, \epsilon)\, d\Delta_{iT}\right] \tag{23}$$

where $\theta^{\epsilon}(0)$ is defined in (15). From (21), we also obtain a characterization of $\widehat{\gamma}_i^{\epsilon\epsilon}(\epsilon)$

$$\begin{aligned} 0 = {} & \theta^{\epsilon}(\epsilon)' \int \frac{\partial v_i(\cdot, \epsilon)}{\partial\theta\partial\theta'} dF_i(\epsilon)\,\theta^{\epsilon}(\epsilon) + \int \frac{\partial v_i(\cdot, \epsilon)}{\partial\theta'} dF_i(\epsilon)\,\theta^{\epsilon\epsilon}(\epsilon) \\ & + 2\int \frac{\partial v_i(\cdot, \epsilon)}{\partial\theta'\partial\gamma_i} dF_i(\epsilon)\,\theta^{\epsilon}(\epsilon)\,\widehat{\gamma}_i^{\epsilon}(\epsilon) + \int \frac{\partial^2 v_i(\cdot, \epsilon)}{\left(\partial\gamma_i\right)^2} dF_i(\epsilon)\left(\widehat{\gamma}_i^{\epsilon}(\epsilon)\right)^2 \\ & + \int \frac{\partial v_i(\cdot, \epsilon)}{\partial\gamma_i} dF_i(\epsilon)\,\widehat{\gamma}_i^{\epsilon\epsilon}(\epsilon) + \int \frac{\partial v_i(\cdot, \epsilon)}{\partial\theta'} d\Delta_{iT}\theta^{\epsilon}(\epsilon) + \int \frac{\partial v_i(\cdot, \epsilon)}{\partial\gamma_i} d\Delta_{iT}^{\epsilon}\widehat{\gamma} i(\epsilon) \end{aligned} \tag{24}$$

We use these results to show that $\widehat{\beta}-\beta=o_p(1)$. First consider $\widehat{E}\left[V_i^{\gamma_i}\right]=\frac{1}{T}\sum_{t=1}^{T}V_{it}^{\gamma_i}\left(x_{it};\widehat{\theta},\widehat{\gamma}_i\right)$. We have

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

so that

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

By Lemma 4 the second term tends to zero with probability $1-o(T^{-1})$. Applying Lemmas 4 and 5 to the first term, we obtain $\max_i\left\|\widehat{E}\left[V_i^{\gamma_i}\right]-E\left[V_i^{\gamma_i}\right]\right\|=o_p\left(1\right)$. In the same way, we obtain $\max_i\left\|\widehat{E}\left[V_i^{\theta}\right]-E\left[V_i^{\theta}\right]\right\|=o_p\left(1\right)$ and $\max_i\left\|\widehat{E}\left[U_i^{\gamma_i\gamma_i}\right]-E\left[U_i^{\gamma_i\gamma_i}\right]\right\|=o_p\left(1\right)$. Let $\widehat{E}\left[V_i^{\gamma_i}\right]=E\left[V_i^{\gamma_i}\right]+o_p(1)$, which holds uniformly in i. Thus, $\max_i\left\|\widehat{\mathcal{I}}_i-\mathcal{I}_i\right\|\leq\sup_i E\left[\left\|M(x_{it})\right\|\right]\left(\left\|\widehat{\theta}-\theta\right\|+\max_i\left|\widehat{\gamma}_i-\gamma_{i0}\right|\right)+o_p\left(1\right)$. Since $\left|\widehat{\gamma}_i-\gamma_{i0}\right|\leq\frac{1}{\sqrt{T}}\left|\widehat{\gamma}_i^{\epsilon}\left(0\right)\right|+\frac{1}{2T}\left|\widehat{\gamma}_i^{\epsilon\epsilon}\left(\tilde{\epsilon}\right)\right|$ with $\max_i T^{-\frac{1}{10}}\left|\widehat{\gamma}_i^{\epsilon}\left(0\right)\right|=o_p(1)$ and $\max_i T^{-\frac{2}{10}}\left|\widehat{\gamma}_i^{\epsilon\epsilon}\left(\tilde{\epsilon}\right)\right|=o_p(1)$ by Lemma 5, it follows that $\max_i\left\|\widehat{\mathcal{I}}_i-\mathcal{I}_i\right\|=o_p(1)$ such that $n^{-1}\sum_{i=1}^{n}\widehat{\mathcal{I}}_i-\mathcal{I}=o_p(1)$.

Using these results we now have

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\widehat{f}_i^{VU^{\gamma}}}{\widehat{E}\left[V_{it}^{\gamma_i}\right]}-\frac{\widehat{f}_i^{VU^{\gamma}}}{E\left[V_{it}^{\gamma_i}\right]}\right] &= o_p\left(1\right), \\ \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\widehat{E}\left[U_i^{\gamma_i\gamma_i}\left(x_{it};\theta,\gamma_i\right)\right]\widehat{f}_i^{VV}}{2\left(\widehat{E}\left[V_{it}^{\gamma_i}\right]\right)^2}-\frac{E\left[U_i^{\gamma_i\gamma_i}\left(x_{it};\theta,\gamma_i\right)\right]\widehat{f}_i^{VV}}{2\left(E\left[V_{it}^{\gamma_i}\right]\right)^2}\right] &= o_p\left(1\right).\end{aligned}$$

In order to establish the result we thus need to apply Lemma 6 to show that

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

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

References

[1] Akaike, H. (1969): "Power Spectrum Estimation Through Autoregressive Model Fitting," *Ann. Inst. Statist. Math,* 21, 407-419

[2] Amemiya, T. (1973): "Rergression Analysis when the Dependent Variable Is Truncated Normal," *Econometrica,* 41, 997-1016.

[3] Andrews, D.W.K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, pp. 817–858.

[4] Andrews, D.W.K. and J. C. Monahan (1992) "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica,* 60, 953-966.

[5] Arellano, M. (2000): "Discrete Choices with Panel Data", Investigaciones Económicas Lecture, XXV Simposio de Análisis Económico, Bellaterra.

[6] Beltrao, K. I., and P. Bloomfield (1987): "Determining the Bandwidth of a Kernel Spectrum Estimate," *Journal of Time Series Analysis,* 8, 21-38.

[7] Billingsley, P. (1986): *Probability and Measure*. John Wiley and Sons, New York.

[8] Binder, J., Hsiao, C., and M.H. Pesaran (2000): "Estimation and Inference in Short Panel Vector Autoregressions with Unit Roots and Cointegration," *unpublished manuscript.*

[9] Chanda, K.C. (1974): "Strong Mixing Properties of Linear Stochastic Processes," *J. Appl. Prob.,* 11, 401-408.

[10] De Jong, R. and T. Woutersen (2003): "Dynamic Time Series Binary Choice," *unpublished manuscript.*

[11] Hahn, J. and 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.

[12] Hahn, J., G. Kuersteiner, and W.K. Newey (2002): "Higher Order Properties of Bootstrap and Jackknife Bias Corrected Maximum Likelihood Estimators", *unpublished manuscript.*

[13] Hahn, J., and J. Meinecke (2003): "Time Invariant Regressor in Nonlinear Panel Model with Fixed Effects", *unpublished manuscript.*

[14] Hahn, J. and W.K. Newey (2002): "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models", *unpublished manuscript.*

[15] Hall, P., and C. Heyde (1980): *Martingale Limit Theory and its Application*. Academic Press.

[16] Hall, P. and J. Horowitz (1996): "Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators", *Econometrica* 64, 891-916.

[17] Heckman, J.J., and T. E. Macurdy (1980): "A Life Cycle Model of Female Labour Supply," *The Review of Economic Studies,* 47, 47-74.

[18] Hendel, I. and A. Nevo (2002): "Sales and Consumer Inventory," *unpublished manuscript.*

[19] Honore, B.E. (1992): "Trimmed Lad and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects," *Econometrica,* 60, 533-565.

[20] Honore, B., and E. Kyriazidou (2000): "Panel Data Discrete Choice Models with Lagged Dependent Variables", *Econometrica* 68, pp. 839 - 874.

[21] Hurvich, C.M. (1985): "Data Driven Choice of a Spectrum Estimate: Extending the Applicability of Cross-Validation methods," *Journal of the American Statistical Association,* 80, 933-940.

[22] Kiviet, J.F. (1995): "On Bias, Inconsistency and Efficiency of Various Estimators in Dynamic Panel Data Models", *Journal of Econometrics* 68, pp. 53 - 78.

[23] Kuersteiner, G. (2004): "Automatic Inference for Infinite Order Vector Autoregressions," forthcoming *Econometric Theory.*

[24] Lahiri, S. (1992): "Edgeworth Correction by Moving Block Bootstrap for Stationary and Nonstationary Data," in *Exploring the Limits of Bootstrap*, ed. by R. LePage, and L. Billard, pp. 183–214. Wiley.

[25] Newey, W. K., and K. D. West (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, pp. 703–708.

[26] Newey, W.K., and K. D. West (1994): "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies,* 61, 631-654.

[27] Newey, W.K., and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in: The Handbook of Econometrics, Volume IV, edited by R.F. Engle and D.L. McFadden, Elsevier Science B.V., 2111-2245.

[28] Neyman, J., and E. Scott (1948): "Consistent Estimates Based on Partially Consistent Observations", *Econometrica* 16, pp. 1 -31.

[29] Olsen, R.J. (1978): "Note on the Uniqueness of the Maximum Likelihood Estimator for the Tobit Model," *Econometrica* 46, pp. 1211-1215.

[30] Parzen, E. (1957): "On Consistent Estimates of the Spectrum of a Stationary Time Series," *Annals of Mathematical Statistics,* 41, 44-58.

[31] Robert, M.J., and J.R. Tybout (1997): "The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs", *American Economic Review* 87, pp. 545-564.

[32] Robinson, P.M. (1991): "Automatic Frequency Domain Inference on Semiparametric and Nonparametric Models," *Econometrica,* 59, 1329-1363.

[33] Shibata, R. (1981): "An Optimal Autoregressive Spectral Estimate," *Annals of Statistics,* 9, 300-306.

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

[35] Velasco, C. (2000): "Local cross-validation for spectrum bandwidth choice," *J. Time Ser. Anal.,* 21, 329–361.

Table 1: Monte Carlo Results for Probit Models

DGP					Mean Bias						RMSE					
					MLE			Bias-Corrected MLE			MLE			Bias-Corrected MLE		
n	*T*	ζ	τ	*dim(z)*	ζ		τ	ζ		τ	ζ		τ	ζ		τ
250	8	1	0.5	1	0.262		-0.471	0.155		-0.219	0.282		0.488	0.183		0.247
250	8	(1,1)'	0.5	2	0.312	0.318	-0.424	0.201	0.206	-0.195	0.330	0.334	0.448	0.225	0.229	0.236
500	8	1	0.5	1	0.261		-0.470	0.155		-0.218	0.271		0.478	0.168		0.232
500	8	(1,1)'	0.5	2	0.301	0.308	-0.420	0.192	0.199	-0.193	0.311	0.318	0.432	0.205	0.211	0.214
250	16	1	0.5	1	0.117		-0.203	0.043		-0.064	0.127		0.217	0.062		0.095
250	16	(1,1)'	0.5	2	0.128	0.131	-0.177	0.043	0.046	-0.052	0.139	0.141	0.193	0.065	0.065	0.089
500	16	1	0.5	1	0.116		-0.203	0.043		-0.063	0.121		0.210	0.054		0.080
500	16	(1,1)'	0.5	2	0.124	0.126	-0.180	0.041	0.042	-0.055	0.130	0.131	0.188	0.054	0.053	0.075
250	8	1	1	1	0.275		-0.431	0.170		-0.234	0.298		0.453	0.199		0.266
250	8	(1,1)'	1	2	0.315	0.325	-0.340	0.206	0.216	-0.168	0.335	0.344	0.375	0.233	0.241	0.222
500	8	1	1	1	0.273		-0.434	0.168		-0.237	0.285		0.445	0.183		0.253
500	8	(1,1)'	1	2	0.303	0.312	-0.341	0.197	0.205	-0.170	0.314	0.322	0.359	0.212	0.219	0.198
250	16	1	1	1	0.126		-0.198	0.051		-0.093	0.137		0.214	0.071		0.118
250	16	(1,1)'	1	2	0.132	0.134	-0.152	0.047	0.048	-0.067	0.143	0.145	0.174	0.069	0.070	0.103
500	16	1	1	1	0.125		-0.196	0.050		-0.091	0.131		0.205	0.061		0.105
500	16	(1,1)'	1	2	0.127	0.128	-0.155	0.044	0.044	-0.071	0.135	0.135	0.166	0.059	0.057	0.089

Note 1: $z_{it} \sim N(0,1)$
Note 2: Results based on 1000 runs.

Table 2: Monte Carlo Results for Logit Models

DGP					Mean Bias									RMSE							
					MLE			Bias-Corrected MLE			Honore-Kyriazidou			MLE			Bias-Corrected MLE			Honore-Kyriazidou		
n	*T*	ζ	τ	*dim(z)*	ζ		τ	ζ		τ	ζ		τ	ζ		τ	ζ		τ	ζ		τ
250	8	1	0.5	1	0.255		-0.746	0.076		-0.243	0.016		-0.069	0.269		0.767	0.107		0.287	0.076		0.182
250	8	(1,1)'	0.5	2	0.321	0.326	-0.728	0.088	0.092	-0.229	-1.095	-1.088	-0.287	0.337	0.342	0.757	0.130	0.132	0.291	1.095	1.089	0.316
500	8	1	0.5	1	0.248		-0.747	0.071		-0.242	0.011		-0.074	0.256		0.757	0.089		0.265	0.053		0.144
500	8	(1,1)'	0.5	2	0.316	0.320	-0.727	0.086	0.089	-0.228	-1.097	-1.088	-0.291	0.324	0.328	0.742	0.108	0.110	0.260	1.097	1.088	0.305
250	16	1	0.5	1	0.098		-0.307	0.013		-0.060	0.004		-0.081	0.108		0.323	0.042		0.110	0.044		0.127
250	16	(1,1)'	0.5	2	0.119	0.119	-0.286	0.018	0.017	-0.054	-1.039	-1.035	-0.486	0.129	0.129	0.308	0.047	0.047	0.119	1.039	1.038	0.264
500	16	1	0.5	1	0.098		-0.307	0.013		-0.059	0.004		-0.081	0.103		0.315	0.031		0.088	0.030		0.106
500	16	(1,1)'	0.5	2	0.119	0.118	-0.286	0.017	0.016	-0.054	-1.038	-1.037	-0.251	0.124	0.123	0.297	0.035	0.034	0.091	1.038	1.037	0.257
250	8	1	1	1	0.259		-0.711	0.081		-0.287	0.026		-0.142	0.273		0.734	0.111		0.327	0.084		0.232
250	8	(1,1)'	1	2	0.322	0.330	-0.638	0.091	0.097	-0.246	-1.097	-1.084	-0.518	0.340	0.346	0.673	0.136	0.139	0.306	1.098	1.084	0.535
500	8	1	1	1	0.251		-0.711	0.075		-0.285	0.018		-0.144	0.259		0.723	0.093		0.306	0.058		0.197
500	8	(1,1)'	1	2	0.317	0.324	-0.640	0.088	0.093	-0.247	-1.099	-1.083	-0.524	0.326	0.332	0.657	0.110	0.114	0.278	1.099	1.083	0.533
250	16	1	1	1	0.101		-0.298	0.015		-0.087	0.005		-0.160	0.111		0.315	0.044		0.127	0.044		0.189
250	16	(1,1)'	1	2	0.122	0.122	-0.253	0.019	0.019	-0.069	-1.038	-1.035	-0.487	0.133	0.133	0.279	0.049	0.049	0.127	1.039	1.035	0.493
500	16	1	1	1	0.101		-0.296	0.015		-0.085	0.004		-0.158	0.106		0.305	0.033		0.109	0.032		0.173
500	16	(1,1)'	1	2	0.122	0.120	-0.255	0.019	0.017	-0.069	-1.039	-1.035	-0.486	0.127	0.125	0.267	0.036	0.035	0.102	1.039	1.035	0.489

Note 1: $z_{it} \sim N(0, \pi^2/3)$
Note 2: Honore-Kyriaizoud estimator is based on bandwidth parameter = 8
Note 3: Results based on 1000 runs.