JACKKNIFE AND ANALYTICAL BIAS REDUCTION FOR NONLINEAR PANEL MODELS

Jinyong Hahn
Whitney Newey

THE INSTITUTE FOR FISCAL STUDIES
DEPARTMENT OF ECONOMICS, UCL
cemmap working paper CWP17/03

Jackknife and Analytical Bias Reduction for Nonlinear Panel Models

Jinyong Hahn
UCLA

Whitney Newey
MIT

October, 2003

ACKNOWLEDGMENT: Newey's research was partially supported by the NSF under Grant No. 0136869. Able research assistance was provided by C. Hansen. Helpful comments were given by G. Chamberlain, V. Chernozhukov, I. Fernandez-Val, J. Hausman, G. Kuersteiner, and seminar participants of University of Arizona, University of Bristol, CAM Workshop at University of Copenhagen, University of Exeter, Harvard/MIT, UC Berkeley, UC Riverside, and UC San Diego are appreciated.

Abstract

Fixed effects estimators of panel models can be severely biased because of the well-known incidental parameters problem. We show that this bias can be reduced by using a panel jackknife or an analytical bias correction motivated by large T. We give bias corrections for averages over the fixed effects, as well as model parameters. We find large bias reductions from using these approaches in examples. We consider asymptotics where T grows with n, as an approximation to the properties of the estimators in econometric applications. We show that if T grows at the same rate as n the fixed effects estimator is asymptotically biased, so that asymptotic confidence intervals are incorrect, but that they are correct for the panel jackknife. We show T growing faster than $n^{1/3}$ suffices for correctness of the analytic correction, a property we also conjecture for the jackknife.

1 Introduction

Panel data, consisting of observations across time for different individual economic agents, allows the possibility of controlling for unobserved time invariant individual heterogeneity. Such heterogeneity can be an important phenomenon, and failure to control for it can result in misleading inferences. This problem is particularly severe when the unobserved heterogeneity is correlated with explanatory variables. Such a situation arises naturally when some of the explanatory variables are decision variables and the unobserved heterogeneity represents variation in tastes or technology.

Models and methods of controlling for unobserved heterogeneity in linear models are well established; see Chamberlain (1984) and Arellano and Honoré (2001) for references and discussion. Controlling for unobserved heterogeneity is much more difficult in nonlinear models. Conditional maximum likelihood can be used in the rare instance that there is a sufficient statistic for the individual effect (Anderson, 1970). In a few cases clever estimators that do not depend on individual effects have been found (e.g. Manski, 1987, Honoré, 1992, and Horowitz and Lee, 2003). However, such cases are the exception rather than the rule. In most models no sufficient statistic exists and no clever estimator has been found. Further, such clever estimators usually provide no guidance towards estimating averages over individual effects.

One way of attempting to control for individual effects in nonlinear models is to treat each effect as a separate parameter to be estimated. Unfortunately, such 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 n goes to infinity while the number of time periods T is held fixed. This inconsistency occurs because only a finite number of observations are available to estimate each individual effect. Hence the estimation error for the individual effects does not vanish as the sample size grows, and this error contaminates the estimates of parameters of interest.

Even in static models the incidental parameters bias can be severe, e.g. see Chamberlain (1980) and Abrevaya (1997), although the size of the bias varies greatly with the parameter of interest, e.g. see Greene (2002). Increasing T need not fully solve this problem, because fixed effects estimators are asymptotically biased even if T grows at the same rate as n. For dynamic linear models this result was shown by Hahn and Kuersteiner (2002) and we show it here for nonlinear models.

We consider two approaches to reducing the bias from fixed effects estimation in nonlinear models. The panel jackknife uses the variation in the fixed effects estimators as each time period is dropped, one at a time, to form a bias corrected estimator using the Quenouille (1956) and Tukey (1958) jackknife formula as in Newey (1989). The analytic bias correction uses the bias formula obtained from an asymptotic expansion as T grows, similarly to Hahn and Kuersteiner (2002). This correction is obtained by generalizing a bias formula from Waterman et. al. (2000), estimating the bias term, and using this to correct the estimator. We also relate the analytical bias-correction formula to an interesting approach developed by Woutersen (2002).

In nonlinear models the object of interest may be an average over the fixed effects rather than the parameters of the likelihood. Such object are of interest because they can be used to summarize results for the whole population. For example, Chamberlain (1984) considers the effect of changing regressors on the average over the fixed effect of a binary choice probability. Here we also give bias corrections for such averages.

In numerical and Monte Carlo examples we find that these large T bias corrections can substantially reduce incidental parameters bias. We find that this bias is particularly small for averages over fixed effects of probit choice probabilities. Surprisingly, there is no incidental parameters bias at all for these average probabilities in a Gaussian linear model.

We analyze the properties of both jackknife and analytical bias corrections under asymptotics as n and T grow. We show that if T grows faster than $n^{1/3}$, the analytic bias correction yields an estimator that is asymptotically normal and centered at the truth. Thus, the analytic bias correction may have good properties in economic applications, where n tends to be much greater than T. Woutersen (2002) gives an analogous result for his bias correction. We also show that the jackknife correction is asymptotically normal centered at the truth when n and T grow at the same rate. We conjecture that the jackknife correction also has this limiting distribution when T grows faster than $n^{1/3}$, although we do not show it here because it is difficult to do. As usual, the asymptotic theory is primarily intended as an approximation to the finite sample distribution of the estimator, here especially to its bias. Under our asymptotic approximation, the bias correction does not increase the asymptotic variance. This suggests that the benefit of bias reduction substantially dominates any potential increase of variance.

In Section 2 we use some comparatively simple calculations to describe how the bias corrections work. Section 3 gives two numerical examples. Section 4 describes and interprets the analytical bias corrections. Section 5 presents bias corrections for averages over fixed effects. Section 6 gives the Monte Carlo results. Asymptotic theory is given in Section 7.

2 Fixed Effects and Bias Correction

We first describe fixed effects estimators. Let the data observations be denoted by $z_{it}, (t = 1, ..., T; i = 1, ..., n)$. Let θ denote a $p \times 1$ vector of parameters, α_i a scalar individual effect, and $f(z|\theta, \alpha)$ a density function with respect to some measure. The setting where $z = (y, x)$ and f is a conditional density of y given x is included as a special case when the dominating measure for f is a product of the marginal distribution for x with some measure for y. The fixed effects estimator is obtained by doing maximum likelihood treating each α_i as a parameter to be estimated. Assuming that z_{it} are independent across i and t, and concentrating out the α_i leads to

$$\widehat{\theta}_T \equiv \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{n} \sum_{t=1}^{T} \ln f\left(z_{it}|\theta, \widehat{\alpha}_i(\theta)\right), \quad \widehat{\alpha}_i(\theta) \equiv \underset{\alpha}{\operatorname{argmax}} \sum_{t=1}^{T} \ln f\left(z_{it}|\theta, \alpha\right).$$

Here the $\widehat{\alpha}_i(\theta)$ depends on the data only through $Z_{iT} \equiv (z_{i1}, \ldots, z_{iT})'$. For any function $m(Z_{iT}, \alpha_i)$, let $\overline{E}[m(Z_{iT}, \alpha_i)]$ $\equiv \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} E[m(Z_{iT}, \alpha_i)]$, where here and in the asymptotics we treat each α_i as fixed (i.e. we condition on them). It will follow from the usual extremum estimator properties (e.g. Amemiya, 1973) that as $n \to \infty$ with T fixed,

$$\widehat{\theta}_T \xrightarrow{p} \theta_T, \quad \theta_T \equiv \underset{\theta}{\operatorname{argmax}}\, \overline{E}\left[\sum_{t=1}^{T} \ln f\left(z_{it}|\theta, \widehat{\alpha}_i(\theta)\right)\right]. \tag{1}$$

When z_{it} has pdf $f(z_{it}; \theta_0, \alpha_i)$, generally $\theta_T \neq \theta_0$. This is the incidental parameters problem noted by Neyman and Scott (1948). The source of this problem is the estimation error of $\widehat{\alpha}_i(\theta)$. If each $\widehat{\alpha}_i(\theta)$ were replaced by $\alpha_i(\theta)$ maximizing $E\left[\sum_{t=1}^{T} \ln f(z_{it}|\theta, \alpha)\right]$, then $\theta_T = \theta_0$. We consider corrections for the asymptotic bias $\theta_T - \theta_0$.

Some expansions can be used to explain our results. Note that the bias should be small for large enough T, i.e., $\lim_{T\to\infty} \theta_T = \theta_0$. Furthermore, for smooth likelihoods we generally have $\theta_T = \theta_0 + \frac{B}{T} + O\left(\frac{1}{T^2}\right)$ for some B. Intuitively, expanding in $\widehat{\alpha}_i(\theta)$ around $\alpha_i(\theta)$ leads to bias terms of order $1/T$ from three sources: a) the asymptotic bias of $\widehat{\alpha}_i(\theta)$ as T grows; b) the correlation resulting from $\widehat{\alpha}_i(\theta)$ and $\widehat{\theta}$ depending on the same data, and; c) the variance of $\widehat{\alpha}_i(\theta)$. We will make this intuition precise below, and derive the formula for B. We will also generally have, as $n, T \to \infty$, $\sqrt{nT}\left(\widehat{\theta} - \theta_T\right) \xrightarrow{d} N(0, \Omega)$. This is just asymptotic normality of $\widehat{\theta}$ when centered at its probability limit, a result known to hold even under misspecification (e.g. White, 1982).

Under these general conditions the fixed effects estimator is asymptotically biased even if T grows at the same rate as n. For $n/T \to \rho$,

$$\sqrt{nT}\left(\widehat{\theta} - \theta_0\right) = \sqrt{nT}\left(\widehat{\theta} - \theta_T\right) + \sqrt{nT}\frac{B}{T} + O\left(\sqrt{\frac{n}{T^3}}\right) \xrightarrow{d} N\left(B\sqrt{\rho}, \Omega\right).$$

Thus, even when T grows as fast as n, asymptotic confidence intervals based on the fixed effects estimator will be incorrect, due to the limiting distribution of $\sqrt{nT}\left(\widehat{\theta} - \theta_0\right)$ not being centered at 0.

The panel jackknife is an automatic method of bias correction. To describe it, let $\widehat{\theta}_{(t)}$ be the fixed effects estimator based on the subsample excluding the observations of the tth period. The jackknife estimator is

$$\widetilde{\theta} \equiv T\widehat{\theta} - (T-1)\sum_{t=1}^{T}\widehat{\theta}_{(t)}/T. \tag{2}$$

To explain the bias correction from this estimator it is helpful to consider a further expansion

$$\theta_T = \theta_0 + \frac{B}{T} + \frac{D}{T^2} + O\left(\frac{1}{T^3}\right). \tag{3}$$

The limit of $\widetilde{\theta}$ for fixed T and how it changes with T shows the effect of the bias correction. The estimator $\widetilde{\theta}$ will converge in probability to

$$T\theta_T - (T-1)\theta_{T-1} = \theta_0 + \left(\frac{1}{T} - \frac{1}{T-1}\right)D + O\left(\frac{1}{T^2}\right) = \theta_0 + O\left(\frac{1}{T^2}\right).$$

Thus, we see that the asymptotic bias of the jackknife corrected estimator is of order $1/T^2$. Consequently, as we show below, this estimator will have an asymptotic distribution centered at 0 when $n/T \to \rho$. We conjecture that this same result will hold with $T/n^{1/3} \to \infty$, although it would be difficult to show, due to the presence of the T and $(T-1)$ multiplicative factors in equation (2).

The panel jackknife is less complicated than it may seem, although it does require $T+1$ fixed effects estimations. As recently noted by Greene (2002), careful construction of an algorithm can make fixed effects computation quite straightforward. Furthermore, good starting values for $\widehat{\theta}_{(t)}$ and the fixed effects are provided by $\widehat{\theta}$ and $\widehat{\alpha}_i$. It may even be possible to obtain the bias reduction by using a one or two step estimator in place of $\widehat{\theta}_{(t)}$, starting at $\widehat{\theta}$.

An analytical bias correction is to plug into the formula for B estimators of its unknown components to construct $\widehat{B}$, and then form a bias corrected estimator

$$\widehat{\theta}^1 \equiv \widehat{\theta} - \frac{\widehat{B}}{T}. \tag{4}$$

This estimator removes enough of the incidental parameters bias that it may give correct confidence intervals as long as T grows faster than $n^{1/3}$. To see this, suppose that $\sqrt{nT}\left(\widehat{B} - B\right)/T \xrightarrow{p} 0$ (e.g. which holds if $\widehat{B} = B\left(\widehat{\theta}\right)$ for a smooth function $B(\theta)$ of θ). Then for $T/n^{1/3} \longrightarrow \infty$,

$$\sqrt{nT}\left(\widehat{\theta}^1 - \theta_0\right) = \sqrt{nT}\left(\widehat{\theta} - \theta_T\right) - \sqrt{\frac{n}{T}}\left(\widehat{B} - B\right) + O\left(\sqrt{\frac{n}{T^3}}\right) \xrightarrow{d} N(0, \Omega). \tag{5}$$

We will give precise regularity conditions for this result to hold.

For small T this estimator may be sensitive to the choice of $\widehat{B}$. If $\widehat{\theta}$ is heavily biased and it is used in the construction of $\widehat{B}$ it may adversely affect the properties of $\widehat{\theta}^1$. One way to deal with this problem is to use $\widehat{\theta}^1$ in the construction of another $\widehat{B}$, and then form a new bias corrected as in equation (4).[1] One could even iterate this procedure, updating $\widehat{B}$ several times using the previous estimator of $\widehat{\theta}$. To be precise, let $\widetilde{B}(\theta)$ denote an estimator of B depending on θ, and suppose that $\widehat{B} = \widetilde{B}\left(\widehat{\theta}\right)$. Then $\widehat{\theta}^1 = \widehat{\theta} - \widetilde{B}\left(\widehat{\theta}\right)/T$. Iterating gives $\widehat{\theta}^k = \widehat{\theta} - \widetilde{B}\left(\widehat{\theta}^{k-1}\right)/T$, $(k = 2, 3, ...)$. If this estimator were iterated to convergence it would give $\widehat{\theta}^\infty$ solving

$$\widehat{\theta}^\infty = \widehat{\theta} - \widetilde{B}\left(\widehat{\theta}^\infty\right)/T \tag{6}$$

In general this estimator will not have improved asymptotic properties, but may have lower bias for small T, as it does in some of the examples to follow.

The bootstrap is another way of reducing bias. We conjecture that some version of a bootstrap bias correction would also remove the asymptotic bias (e.g. with truncation as in Hahn, Kuersteiner, and Newey, 2002).

3 Numerical Examples

The practical importance of these bias corrections depends on how much bias is removed for the relatively small T that is often relevant in econometric applications. In this section we provide some simple examples showing that these corrections can remove the great majority of the bias even with small T. We do this by calculating probability limits as n grows with T fixed.

The first example has z_{it} distributed $N(\alpha_i, \theta_0)$. This example was considered in Neyman and Scott (1948), who showed that $\theta_T = \theta_0 - \theta_0/T$. Since the formula for θ_T is linear in $1/T$, we have $B = -\theta_0$. Thus, B can be estimated by $\widehat{B} = B\left(\widehat{\theta}\right)$ for $B(\theta) = -\theta$. Using this $\widehat{B}$ for the bias correction gives $\widehat{\theta}^1 = \widehat{\theta} - \widehat{B}/T = \widehat{\theta} + \widehat{\theta}/T$, which converges to $\theta_T^1 = \theta_0[(T-1)/T](1+1/T) = \theta_0 - \theta_0/T^2$. Iterating the bias correction to convergence gives, as in equation (6), the solution to $\widehat{\theta}^\infty = \widehat{\theta} + \widehat{\theta}^\infty/T$, which is $\widehat{\theta}^\infty = [T/(T-1)]\,\widehat{\theta}$, which is consistent. It can also be shown that the jackknife bias correction yields $\widetilde{\theta} = [T/(T-1)]\,\widehat{\theta}$, and so is consistent for θ_0. To summarize and illustrate these results, Table One presents limits $\theta_T, \theta_T^1, \theta_T^\infty, \theta_{TJ}$ of the fixed effects, one-step bias correction, fully iterated bias correction, and jackknife estimators as $n \to \infty$ with T fixed. We find that even the one-step bias correction is a substantial improvement, although the iterated and jackknife bias corrections are perfect.

Table One: $N(\alpha_i, 1)$				
T	θ_T	θ_T^1	θ_T^∞	θ_{TJ}
3	.67	.89	1	1
4	.75	.94	1	1
5	.80	.96	1	1
7	.86	.98	1	1
10	.90	.99	1	1

The second example is estimation of the survivor function in a regression version of the previous model, which corresponds to the choice probability in a probit model. In a binary choice model Chamberlain (1984) proposed averaging over the marginal distribution of the fixed effects to measure how regressors affect choice probabilities.

[1] This iterated bias correction was suggested to us by V. Chernozhukov.

Here we consider a local version for a linear model, the derivative of the probability of the survivor function of y with respect to x, averaged over the fixed effect. Specifically, suppose that $y_{it} = x_{it}'\beta_0 + \alpha_i + \varepsilon_{it}$, where $\varepsilon_{it} \sim N(0, \sigma^2)$. As is well known, the fixed effects MLE $\hat{\beta}$ is consistent but $\hat{\sigma}^2$ converges to $\sigma^2 - \sigma^2/T$, as in the previous example. The prediction we consider is

$$\begin{aligned}\mu_0 &= \int [\partial \Pr(y > r|x, \alpha)/\partial x]\, f(\alpha)d\alpha = \int \{\partial\Phi([x'\beta_0 - r + \alpha]/\sigma)/\partial x\}\, f(\alpha)d\alpha \\ &= \frac{\beta_0}{\sigma}\int \phi\left(\frac{x'\beta_0 - r + \alpha}{\sigma}\right) f(\alpha)d\alpha.\end{aligned}$$

The fixed effects estimator of μ_0 is obtained by plugging in the fixed effects estimators of β_0, α_i, and σ^2 and averaging over $\hat{\alpha}_i$, as in

$$\hat{\mu} = \sum_{i=1}^{n} \frac{\hat{\beta}}{\hat{\sigma}}\phi\left(\frac{x'\hat{\beta} - r + \hat{\alpha}_i}{\hat{\sigma}}\right)/n.$$

Surprisingly, this estimator is consistent at $n \to \infty$ for fixed T, despite the inconsistency of $\hat{\sigma}$ and the estimation error in $\hat{\alpha}_i$. Intuitively, consistency occurs because the $-\sigma^2/T$ downward bias in $\hat{\sigma}^2$ is exactly compensated for by the estimation error in $\hat{\alpha}_i$ being Gaussian with variance σ^2/T. Furthermore, the estimated average survivor function $\hat{S} = \sum_{i=1}^{n} \Phi\left(\frac{x'\hat{\beta} - r + \hat{\alpha}_i}{\hat{\sigma}}\right)/n$, will also be consistent, because $\hat{\mu}$ is its derivative and is consistent for each x.[2]

Because the fixed effects estimator $\hat{\mu}$ is consistent for any T, the jackknife bias corrected estimator will also be consistent for every T. The analytical bias correction need not be consistent when the estimator is, but will be very close for large T. In the tables below, the analytical bias correction is nearly consistent. The analytical bias corrected estimator is

$$\tilde{\mu}^1 = \frac{\tilde{\beta}}{\tilde{\sigma}}\sum_{i=1}^{n}\left[\phi\left(\frac{x'\tilde{\beta} - r + \hat{\alpha}_i}{\tilde{\sigma}}\right) - \phi''\left(\frac{x'\tilde{\beta} - r + \hat{\alpha}_i}{\tilde{\sigma}}\right)/2T\right]/n$$

where we take $\tilde{\sigma} = T\hat{\sigma}^2/(T-1)$ to be the fully iterated bias corrected estimator and $\tilde{\beta}$ to be the fixed effects estimator. Table Two shows the ratio of the probability limit of this estimator to the truth for $\alpha_i \sim N(0, \sigma^2)$ and $m = (x'\beta_0 - r)/\sigma$ equal to 0, 1, and 2.

Table Two: Survivor Derivative

T	$m = 0$	$m = 1$	$m = 2$
3	.991	.989	.976
4	.995	.991	.978
5	.997	.993	.980
7	.998	.994	.984
10	.999	.997	.988

We see that there is very little bias in the estimated survivor function with an analytical bias correction. These results, as well as the consistency of fixed effects and jackknife bias correction help explain the low biases we find in the estimation of average choice probability derivatives in the Monte Carlo results below. Although we do not expect fixed effects to be consistent outside this simple example, it appears that in more than one case that probabilities, when they are averaged over fixed effects, have very small bias.

[2] To see consistency, note that $\hat{S}$ converges to $S_0 = E[\Phi((m + \hat{\alpha}_i)/\sigma_T^2)]$ where $\hat{\alpha}_i = \alpha_i + v_i$, $v_i \sim N(0, \sigma^2/T)$, $\sigma_T^2 = \sigma^2 - \sigma^2/T$, and $m = x'\beta - r$. Since $E[\Phi((m + \hat{\alpha}_i)/\sigma_T)|\alpha_i]$ is the CDF of a normal mixture of v_i with $N(0, \sigma_T^2)$, which CDF is $\Phi((m + \alpha_i)/\sigma)$, it follows by interated expectations that $S_0 = E[\Phi((m + \alpha_i)/\sigma)]$.

4 Analytical Bias Correction

A little more notation is useful for describing the analytical bias correction. Let

$$u_{it}(\theta,\alpha) \equiv \frac{\partial}{\partial\theta}\log f(z_{it}|\theta,\alpha), \quad v_{it}(\theta,\alpha) \equiv \frac{\partial}{\partial\alpha_i}\log f(z_{it}|\theta,\alpha),$$

and additional subscripts denote partial derivatives, e.g. $u_{it\theta}(\theta,\alpha) \equiv \partial u_{it}(\theta,\alpha)/\partial\theta$. Also, define

$$\begin{aligned}
\widehat{v}_{it}(\theta) &\equiv v_{it}(\theta,\widehat{\alpha}_i(\theta)),\ \widehat{v}_{it\alpha}(\theta) \equiv v_{it\alpha}(\theta,\widehat{\alpha}_i(\theta)),\ \widehat{v}_{it\alpha\alpha}(\theta) \equiv v_{it\alpha\alpha}(\theta,\widehat{\alpha}_i(\theta)),\ \widehat{u}_{it}(\theta) \equiv u_{it}(\theta,\widehat{\alpha}_i(\theta)),\\
\widehat{u}_{it\alpha}(\theta) &\equiv \widehat{u}_{it\alpha}(\theta,\widehat{\alpha}_i(\theta)),\ \widehat{u}_{it\alpha\alpha}(\theta) \equiv u_{it\alpha\alpha}(\theta,\widehat{\alpha}_i(\theta)),\ \widehat{\psi}_{it}(\theta) \equiv -\widehat{v}_{it}(\theta)/\left(\sum_{s=1}^{T}\widehat{v}_{is\alpha}(\theta)/T\right),\ \widehat{\sigma}_i^2(\theta) \equiv \sum_{s=1}^{T}\widehat{\psi}_{it}(\theta)^2/T
\end{aligned}$$

Here $\widehat{\psi}_{it}(\theta)$ and $\widehat{\sigma}_i^2(\theta)$ are estimators of the influence function and asymptotic variance, respectively, of $\widehat{\alpha}_i(\theta)$ as T grows. Let

$$\begin{aligned}
\widehat{\beta}_i(\theta) &= -\left(\sum_{s=1}^{T}\widehat{v}_{is\alpha}(\theta)\right)^{-1}\sum_{s=1}^{T}\left\{\widehat{v}_{is\alpha}(\theta)\widehat{\psi}_{is}(\theta) + \widehat{v}_{is\alpha\alpha}(\theta)\widehat{\sigma}_i^2(\theta)/2\right\},\\
\widehat{H}(\theta) &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{u_{it\theta}(\theta,\widehat{\alpha}_i(\theta)) - \widehat{u}_{it\alpha}(\theta)\left[\sum_{s=1}^{T}v_{is\theta}(\theta,\widehat{\alpha}_i(\theta))/\sum_{s=1}^{T}\widehat{v}_{is\alpha}(\theta)\right]\right\}
\end{aligned}$$

Here $\widehat{\beta}_i(\theta)$ is an estimator of the higher-order asymptotic bias of $\widehat{\beta}_i(\theta)$ from a stochastic expansion of $\widehat{\alpha}_i(\theta)$ as T grows. An analytic, i.e. closed form estimator, of the bias term can be formed as

$$\widehat{B}(\theta) = -\widehat{H}(\theta)^{-1}\widehat{b}(\theta), \widehat{b}(\theta) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{\widehat{u}_{it\alpha}(\theta)\left[\widehat{\beta}_i(\theta)+\widehat{\psi}_{it}(\theta)\right] + \widehat{u}_{it\alpha\alpha}(\theta)\widehat{\sigma}_i^2(\theta)/2\right\}.$$

The bias corrected estimator, either one step or iterated, can then be formed as described above, with $\widetilde{B}(\theta) = \widehat{B}(\theta)$. We give an interpretation of the formula for $\widehat{B}(\theta)$ below.

It is interesting to note that the bias correction $\widehat{B}(\theta)$ does not depend on the likelihood setting, and so would be valid for any fixed effects m-estimator based on the moment equations $E[u_{it}(\theta_0,\alpha_i)] = 0$ and $E[v_{it}(\theta_0,\alpha_i)] = 0$. For simplicity though, we will only give asymptotic theory for the maximum likelihood setup.

The likelihood setting can be used to simplify the bias correction using Bartlett identities. Let

$$\begin{aligned}
\widehat{U}_{it}(\theta) &\equiv \widehat{u}_{it}(\theta) - \widehat{v}_{it}(\theta)\sum_{s=1}^{T}\widehat{u}_{is}(\theta)\widehat{v}_{is}(\theta)/\left(\sum_{s=1}^{T}\widehat{v}_{is}(\theta)^2\right), \widehat{V}_{2it}(\theta) \equiv \widehat{v}_{it}(\theta)^2 + \widehat{v}_{it\alpha}(\theta).\\
\overline{B}(\theta) &= -\overline{H}(\theta)^{-1}\overline{b}(\theta),\quad \overline{H}(\theta) = \frac{-1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\widehat{U}_{it}(\theta)\widehat{U}_{it}(\theta)',\quad \overline{b}(\theta) = \frac{-1}{2n}\sum_{i=1}^{n}\sum_{t=1}^{T}\widehat{U}_{it}(\theta)\widehat{V}_{2it}(\theta)/\left(\sum_{s=1}^{T}\widehat{v}_{is}(\theta)^2\right).
\end{aligned}$$

The bias corrected estimator can then be formed with $\widetilde{B}(\theta) = \bar{B}(\theta)$. This correction uses outer products in place of some derivatives. In the Monte Carlo example we find that this replacement has very little effect on the properties of bias corrected estimators.

The analytical bias correction has an interpretation that explains its form. The matrix $\hat{H}(\theta)$ is an estimator of the Jacobian of $\sum_{i,t}\hat{u}_{it}(\theta)/(nT)$ with respect to θ, the first term (involving $u_{it\theta}$) being the direct derivative and the second being the derivative acting through $\hat{\alpha}_i(\theta)$. The vector $\hat{b}(\theta)$ is an estimator of $\bar{E}[\sum_{i,t}\hat{u}_{it}(\theta_0)/(nT)] = E[\hat{u}_{it}(\theta_0)]$. Its form can be explained by an expansion. For notational convenience suppress the θ_0 and α_i arguments of $u_{it}(\theta,\alpha)$, $v_{it}(\theta,\alpha)$, and their partial derivatives evaluated at (θ_0,α_i) (e.g. $v_{it\alpha} = v_{it\alpha}(\theta_0,\alpha_i)$). Then, using independence across t, standard higher-order asymptotics gives (e.g. Rilstone et. al., 1996), for $T \to \infty$,

$$\widehat{\alpha}_i = \alpha_i + \beta_i/T + \sum_{t=1}^{T} \psi_{it}/T + o_p(1/T), \ \psi_{it} = -E[v_{it\alpha}]^{-1} v_{it}, \ \beta_i = -E[v_{it\alpha}]^{-1}\left\{E[v_{it\alpha}\psi_{it}] + \frac{1}{2}E[v_{it\alpha\alpha}]E\left[\psi_{it}^2\right]\right\}$$

Then, expanding around α_i, and assuming orders in probability correspond to orders in expectation, we obtain

$$\begin{aligned} E[u_{it}(\theta_0,\widehat{\alpha}_i)] &= E[u_{it}] + E[u_{it\alpha}(\widehat{\alpha}_i - \alpha_i)] + E\left[u_{it\alpha\alpha}(\widehat{\alpha}_i-\alpha_i)^2\right]/2 + o(1/T) \\ &= 0 + \frac{1}{T}\{E[u_{it\alpha}]\beta_i + E[u_{it\alpha}\psi_{it}] + E[u_{it\alpha\alpha}]E\left[\psi_{it}^2\right]/2\} + o(1/T). \end{aligned}$$

Thus we see that there are three terms in the bias of the moment function, all arising from the randomness of $\widehat{\alpha}_i$. The first term comes from the asymptotic bias of $\widehat{\alpha}_i$. The second term comes from correlation of $u_{it\alpha}$ with ψ_{it}, and is present because $\hat{\alpha}_i$ is computed from the same observations as $\hat{\theta}$. The third term is pure random parameters bias, that corresponds to the standard bias formula for nonlinear functions of sample averages. The expression $\widehat{b}(\theta)$ is just a sample analog of this formula.

The likelihood based bias correction uses second and third order Bartlett identities in its construction. The identities used are, for $V_{2it} = v_{it}^2 + v_{it\alpha}$,

$$\begin{aligned} E[v_{it\alpha}] &= -E[v_{it}^2],\ E[u_{it\alpha}] = -E[u_{it}v_{it}],\ E[u_{it\theta}] = -E[u_{it}u_{it}'],\ E[v_{it\theta}] = -E[v_{it}u_{it}'], \\ E[v_{it\alpha}v_{it}] + E[v_{it\alpha\alpha}]/2 &= -E[v_{it}V_{2it}]/2,\ E[u_{it\alpha}v_{it}] + E[u_{it\alpha\alpha}]/2 = -E[u_{it}V_{2it}]/2. \end{aligned}$$

These identities imply that $\beta_i = -E[v_{it}^2]^{-2}E[v_{it}V_{2it}]/2$, so that for $U_{it} = u_{it} - v_{it}E[v_{it}^2]^{-1}E[v_{it}u_{it}]$, we have

$$\begin{aligned} & E[u_{it\alpha}]\beta_i + E[u_{it\alpha}\psi_{it}] + E[u_{it\alpha\alpha}]E[\psi_{it}^2]/2 \\ = \ & E[u_{it}v_{it}]E[v_{it}^2]^{-2}E[v_{it}V_{2it}]/2 - E[v_{it}^2]^{-1}E[u_{it}V_{2it}]/2 = -E[v_{it}^2]^{-1}E[U_{it}V_{2it}]/2. \end{aligned}$$

The formula for $\bar{b}(\theta)$ is a sample analog of this expression. Also, the term $\bar{H}(\theta)$ is an outer product approximation to the Jacobian based on the Bartlett identities.

Another approach to bias correction for fixed effects, that helps explain the relationship of our estimators to those of Lancaster (2002) and Woutersen (2002), is to construct the estimator as the solution to a bias corrected version of the estimating equation (i.e. first-order conditions). Firth (1993) suggests this approach to asymptotic bias correction with a fixed number of parameters and Fernandez (2003) for nonlinear panel data models. To describe it, let $\widehat{m}(\theta) = \sum_{i=1}^{n}\sum_{t=1}^{T}\widehat{u}_{it}(\theta)/(nT)$, so that the fixed effects estimator solves $\widehat{m}(\widehat{\theta}) = 0$. Also, let $\check{b}(\theta)/T$ be an estimator of the order $1/T$ bias of $\widehat{m}(\theta)$ when θ is the true parameter, such as $\check{b}(\theta) = \hat{b}(\theta)$ or $\check{b}(\theta) = \bar{b}(\theta)$. Consider an estimator $\bar{\theta}$ obtained by solving

$$0 = \widehat{m}(\theta) - \widetilde{b}(\theta)/T. \tag{8}$$

This estimator also uses an analytical bias correction, but to the estimating equation rather than the estimator. As an example, when $\check{b}(\theta) = \bar{b}(\theta)$, the estimator $\bar{\theta}$ would solve

$$0 = \sum_{i=1}^{n}\sum_{t=1}^{T}\left[\widehat{u}_{it}(\theta) + \frac{1}{2}\widehat{U}_{it}(\theta)\widehat{V}_{2it}(\theta) / \left(\sum_{s=1}^{T}\widehat{v}_{is}(\theta)^2\right)\right]. \tag{9}$$

As shown in Appendix I, this estimating equation is asymptotically equivalent to Woutersen's (2002) Laplace approximation to the integral of a parameter orthogonalization. In this sense, an estimating equation version of our fixed effects bias correction is equivalent to Woutersen's (2002) approach. Unlike Woutersen's (2002), this bias correction requires no integration, and so may be easier to implement in practice.

Bias corrections of estimating equations are related to the iterated bias correction discussed in Section 2. To describe this relationship, let $\hat{m}(\theta)$ be some vector of functions of parameters and data and consider an estimator $\hat{\theta}$ satisfying $\hat{m}(\hat{\theta}) = 0$. Suppose that $\widehat{\theta}^{\infty}$ satisfies equation (6), where $\check{B}(\theta) = \check{H}(\theta)^{-1}\check{b}(\theta)$, $\check{H}(\theta)$ is an estimator of $\partial\hat{m}(\theta)/\partial\theta$ at the true parameter, and $\check{b}(\theta)/T$ is an order $1/T$ bias estimator for $\hat{m}(\theta)$ when θ is true. Then equation (6) implies that $\widehat{\theta}^{\infty}$ solves

$$0 = \check{H}(\theta)\left(\widehat{\theta} - \theta\right) - \widetilde{b}(\theta)/T = \widehat{m}\left(\widehat{\theta}\right) + \check{H}(\theta)\left(\widehat{\theta} - \theta\right) - \widetilde{b}(\theta)/T. \tag{10}$$

If $\widehat{m}(\theta)$ is linear in θ and $\check{H}(\theta) = \partial\hat{m}(\theta)/\partial\theta$, this equation is the same as equation (8). Otherwise, this equation is an approximation to equation (8). Thus we find that the fully iterated bias corrected estimator $\widehat{\theta}^{\infty}$ can be interpreted as the solution to an approximation to equation (8). In particular, the fully iterated bias corrected estimator based on the likelihood correction $\hat{B}(\theta) = \bar{H}(\theta)^{-1}\bar{b}(\theta)$ is an approximation to the solution to equation (9), and hence an approximation to the Woutersen (2002) estimator.

The one-step bias correction $\hat{\theta}^{1} = \hat{\theta} - \hat{B}(\hat{\theta})/T$ may have computational advantages over the solution to an estimating equation and over the iterated bias correction. In a number of important models the likelihood is concave in θ and $\alpha_1, ..., \alpha_n$, so that the fixed effects estimator $\hat{\theta}$ is straightforward to calculate. Also, the formula for the bias correction has a closed form, so that the bias correction can easily be computed. In contrast, the estimating equation and fully iterated bias correction require solution to a nonlinear equation, which may be more difficult to calculate.

5 Bias Corrections for Fixed Effect Averages

The type of object we consider is an average over the fixed effects of the form

$$\mu(w) = \overline{E}\left[m\left(w, z_{it}, \theta_0, \alpha_i\right)\right].$$

An example is Chamberlain's (1984) average probability difference for probit, where y is the probit binary dependent variable, $w = (x, \tilde{x})$ is a pair of regressor values, and $m(w, z, \theta, \alpha) = \Phi(x'\theta + \alpha) - \Phi(\tilde{x}'\theta + \alpha)$. A local version of this would be its derivative with respect to x, where $m(w, z, \theta, \alpha) = \theta\phi(x'\theta + \alpha)$.

We can construct a bias corrected estimator of $\mu(w)$ by a panel jackknife or by an analytical method. To describe the panel jackknife here let $\widehat{\alpha}_{i(t)}$ be the fixed effect with the t^{th} observation excluded and

$$\begin{aligned}
\widehat{\mu}(w) &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} m\left(w, z_{it}, \widehat{\theta}, \widehat{\alpha}_i\right), \quad \widehat{\mu}_{(t)}(w) = \frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{s\neq t} m\left(w, z_{is}, \widehat{\theta}_{(t)}, \widehat{\alpha}_{i(t)}\right), \\
\widetilde{\mu}(w) &= T\widehat{\mu}(w) - (T-1)\sum_{t=1}^{T}\widehat{\mu}_{(t)}(w)/T.
\end{aligned}$$

Here $\widetilde{\mu}(w)$ is a jackknife bias corrected estimator of $\mu(w)$.

It is also possible to construct an analytic bias correction. Let $\widetilde{\theta}$ be a bias corrected estimator (either jackknife or analytical) of θ and $\widetilde{\alpha}_i = \widehat{\alpha}_i\left(\widetilde{\theta}\right)$, $(i = 1, ..., n)$ the corresponding fixed effects. Also, let $\bar{\sigma}_i^2(\theta) = T/\sum_{s=1}^T \hat{v}_{it}(\theta)^2$, $\bar{\beta}_i(\theta) = -\bar{\sigma}_i^2(\theta)^2 \sum_{s=1}^T \hat{v}_{it}(\theta)\hat{V}_{2it}(\theta)/2T$, and $\bar{\psi}_{it}(\theta) = \bar{\sigma}_i^2(\theta)\hat{v}_{it}(\theta)$ denote maximum likelihood versions of estimated bias terms. Then a bias corrected estimator of μ that uses the likelihood structure is

$$\begin{aligned} \widetilde{\mu}(w) &= \sum_{i=1}^{n}\sum_{t=1}^{T} m\left(w, z_{it}, \widetilde{\theta}, \widetilde{\alpha}_i\right)/nT - \widehat{\Delta}(w)/T, \\ \widehat{\Delta}(w) &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\{m_\alpha\left(w, z_{it}, \widetilde{\theta}, \widetilde{\alpha}_i\right)\left[\bar{\beta}_i\left(\widetilde{\theta}\right) + \bar{\psi}_{it}\left(\widetilde{\theta}\right)\right] + m_{\alpha\alpha}\left(w, z_{it}, \widetilde{\theta}, \widetilde{\alpha}_i\right)\bar{\sigma}_i^2\left(\widetilde{\theta}\right)/2\}. \end{aligned}$$

When m_α does not depend on z_{it}, the $\bar{\psi}_{it}\left(\widetilde{\theta}\right)$ term can be dropped. We could also give a form of the bias correction that does not depend on the likelihood structure, which we omit for brevity.

6 A Monte Carlo Study

For a Monte Carlo example we consider fixed effects probit estimation, using a design similar to that of Heckman (1981). The model design is:

$$\begin{aligned} y_{it} &= 1(x_{it}\theta_0 + \alpha_i + \varepsilon_{it} > 0), \alpha_i \sim N(0,1), \varepsilon_{it} \sim N(0,1), \\ x_{it} &= t/10 + x_{i,t-1}/2 + u_{it}, x_{i0} = u_{i0},\ u_{it} = U(-1/2, 1/2). \\ N &= 100, T = 8; \beta = 1, -1. \end{aligned}$$

Although this design does not fit completely within our framework, because x_{it} is correlated over time, we use it because it is similar to a widely cited design, and so helps us compare our results with previous ones. Also, y_{it} is independent over time conditional on α_i and $(x_{i1}, ..., x_{iT})'$, so there is no autocorrelation in that sense.

We give results for estimators of θ_0. We also consider estimation of the average of the derivative of the choice probability $\Phi(x'\theta + \alpha)$ with respect to x at a particular $x = w$, which is

$$\mu = \theta_0 \bar{E}[\phi(w'\theta_0 + \alpha_i)].$$

The fixed effects estimator of this object is $\hat{\mu} = \hat{\theta}\sum_{i=1}^n \phi\left(w'\hat{\theta} + \hat{\alpha}_i\right)/n$ and the jackknife bias correction is described in Section 5. The analytical bias corrected estimator is

$$\hat{\mu}^1 = \check{\theta}\sum_{i=1}^{n}[\phi(w'\check{\theta} + \check{\alpha}_i) - \phi'(w'\check{\theta} + \check{\alpha}_i)\bar{\beta}_i(\check{\theta})/T - \phi''(w'\check{\theta} + \check{\alpha}_i)\bar{\sigma}_i^2(\check{\theta})/2T]/n,$$

where $\check{\theta} = \hat{\theta}^1$ is the bias corrected estimator of θ, $\check{\alpha}_i$ is the corresponding estimator of the fixed effect, $\bar{\beta}_i(\theta)$ and $\bar{\sigma}_i^2(\theta)$ are the MLE bias and asymptotic variance estimators, respectively, described in Section 5.

Table Three gives the Monte Carlo results for the estimator of θ_0 when $T = 8$ and $\theta_0 = 1$. Here SD is the standard deviation of the estimator, $\hat{p}$ denotes a rejection frequency with nominal value given at the column top, and SE/SD is the ratio of the average standard error to standard deviation. Analytic refers to the MLE (outer product) bias correction and Analytic-M to that based on general estimating equations, i.e. to $\hat{B}(\theta)$ in Section 4. We find little difference in performance between these two types of analytic correction. We also find that iterating the bias correction seemed not to matter much, although for brevity we do not report those results. We see here that the fixed effects MLE is biased upward by around 20 percent, confirming the results of Greene (2002) and Woutersen (2002). The bias corrections all reduce the bias to about 5 percent. In addition, they reduce the dispersion. An explanation for this reduction is that the probit coefficients are analogous to the ratio of a regression coefficient to a scale parameter. If the primary fixed effect bias is as if the scale parameter were too small, as it would be in a linear model, then the correction would raise the scale parameter and shrink the estimator towards zero, and so reduce dispersion. We also find large improvements in rejection frequencies, which are much closer to their nominal values after bias correction.

Table Three: Properties of $\hat{\theta}$, $T = 8$.						
Estimator of θ_0	Mean	Med.	SD	$\hat{p}$; .05	$\hat{p}$; .10	SE/SD
MLE	1.18	1.17	.151	.267	.370	.901
Jackknife	.953	.950	.119	.056	.102	1.08
Analytic	1.05	1.05	.134	.062	.135	.985
Analytic-M	1.05	1.05	.132	.060	.126	1.00

Table Four gives the ratio of estimators to the truth for the estimator $\hat{\mu}$ of the average probability derivative evaluated at the mean of the regressor. We find quite small fixed effect biases. They are not quite as small as in the linear model, but are very small relative to standard errors. We also find little improvement in rejection frequencies here. The source of this problem may be underestimation of dispersion, as shown by the last column.

Table Four: Properties of $\hat{\mu}$, $T = 8$.						
Estimator of μ/μ_0	Mean	Med.	SD	$\hat{p}$;.05	$\hat{p}$;.10	SE/SD
MLE	1.02	1.02	.131	.078	.140	.893
Jackknife	1.00	.992	.130	.086	.159	.869
Analytic	1.02	1.02	.133	.090	.153	.857
Analytic-M	1.02	1.02	.131	.087	.154	.870

Tables Five and Six give corresponding results to Tables Three and Four respectively, for $T = 4$. The results are similar, but with larger biases and dispersion and less accurate standard errors.

Table Five: Properties of $\hat{\theta}$, $T = 4$						
Estimator of θ_0	Mean	Med.	SD	$\hat{p}$; .05	$\hat{p}$; .10	SE/SD
MLE	1.42	1.41	.397	.269	.373	.814
Jackknife	.752	.743	.262	.100	.177	1.15
Analytic	1.12	1.11	.306	.055	.101	1.02
Analytic-M	1.21	1.20	.335	.102	.172	.940

Table Six: Properties of $\hat{\mu}$, $T = 4$.						
Estimator of μ/μ_0	Mean	Med.	SD	$\hat{p}$; .05	$\hat{p}$; .10	SE/SD
MLE	1.00	1.00	.257	.103	.168	.829
Jackknife	1.06	1.05	.307	.159	.224	.704
Analytic	.996	.994	.265	.113	.178	.811
Analytic-M	1.05	1.05	.266	.117	.185	.805

7 Asymptotic Theory

The first result we consider is the asymptotic distribution of the fixed effects MLE when $n, T \to \infty$ at the same rate. We impose the following conditions:

Condition 1 *$n, T \to \infty$ such that $\frac{n}{T} \to \rho$, where $0 < \rho < \infty$.*

Condition 2 *(i) The function $\log f(\cdot; \theta, \alpha)$ is continuous in $(\theta, \alpha) \in \Upsilon$; (ii) The parameter space Υ is compact; (iii) There exists a function $M(z_{it})$ such that $|\log f(z_{it}; \theta, \alpha_i)| \leq M(z_{it})$, $\left|\frac{\partial \log f(z_{it}; \theta, \alpha_i)}{\partial(\theta, \alpha_i)}\right| \leq M(z_{it})$, and $\sup_i E\left[M(z_{it})^{33}\right] < \infty$.*

Condition 3 *For each $\eta > 0$, $\inf_i \left[G_{(i)}(\theta_0, \alpha_{i0}) - \sup_{\{(\theta,\alpha):|(\theta,\alpha)-(\theta_0,\alpha_{i0})|>\eta\}} G_{(i)}(\theta, \alpha)\right] > 0$, where*

$$\widehat{G}_{(i)}(\theta, \alpha_i) \equiv T^{-1}\sum_{t=1}^{T} \log f(z_{it}; \theta, \alpha_i) \equiv T^{-1}\sum_{t=1}^{T} g(z_{it}; \theta, \alpha_i), \quad G_{(i)}(\theta, \alpha_i) \equiv E[\log f(z_{it}; \theta, \alpha_i)]$$

Let $\mathcal{I}_i \equiv E[U_{it}U_{it}']$.

Condition 4 *(i) There exists some $M(z_{it})$ such that*

$$\left|\frac{\partial^{m_1+m_2} \log f(z_{it}; \theta, \alpha_i)}{\partial\theta^{m_1}\partial\alpha_i^{m_2}}\right| \leq M(z_{it}) \qquad 0 \leq m_1 + m_2 \leq 1, \ldots, 6$$

and $\sup_i E\left[M(z_{it})^Q\right] < \infty$ for some $Q > 64$; (ii) $\bar{E}[\mathcal{I}_i] > 0$; (iii) $\min_i E\left[v_{it}^2\right] > 0$.

Under these conditions we can obtain the asymptotic distribution of the fixed effects MLE as n and T grow at the same rate.

Theorem 1 *Under Conditions 1, 2, 3, and 4, we have*

$$\sqrt{nT}\left(\widehat{\theta} - \theta_0\right) \xrightarrow{d} N\left(B\sqrt{\rho}, \{\bar{E}[\mathcal{I}_i]\}^{-1}\right), \qquad B \equiv -\frac{1}{2}\left(\bar{E}[\mathcal{I}_i]\right)^{-1} \bar{E}[E[U_{it}V_{2it}]/E[v_{it}^2]].$$

Proof. See Appendix A. ■

Waterman et al (2000) derive the asymptotic distribution of $\sqrt{nT}\left(\widehat{\theta} - \theta_0\right)$ assuming that $\widehat{\theta}$ and $\widehat{\alpha}_i$ are consistent for θ_0 and α_{i0}, and $\dim(\theta) = 1$. Theorem 1 is a rigorous result that includes consistency for $\widehat{\theta}$ and $\widehat{\alpha}_i$, in the sense that,

$$\Pr\left[\left|\widehat{\theta} - \theta_0\right| \geq \eta\right] = o\left(T^{-1}\right), \qquad \Pr\left[\max_{1\leq i\leq n} |\widehat{\alpha}_i - \alpha_{i0}| \geq \eta\right] = o\left(T^{-1}\right).$$

We also allow for $\dim(\theta)$ to be bigger than 1.

Next, we show that the analytic bias correction eliminates the asymptotic bias term.

Theorem 2 *If $T/n^{1/3} \to \infty$, and conditions 2, 3, and 4 are satisfied, then $\widehat{\theta}^1 = \hat{\theta} - \bar{B}(\hat{\theta})/T$ satisfies*

$$\sqrt{nT}\left(\widehat{\theta}^1 - \theta_0\right) \xrightarrow{d} N\left(0, \{\bar{E}[\mathcal{I}_i]\}^{-1}\right).$$

Proof. See Appendix B. ■

By equation (23) in Appendix B, we can see that $\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\widehat{U}_{it}\widehat{U}_{it}'\right)^{-1}$ is a consistent estimator of the asymptotic variance $(\bar{E}[\mathcal{I}_i])^{-1}$ of $\sqrt{nT}\left(\widehat{\theta}^{1}-\theta_0\right)$, which can be used in the usual way for asymptotic inference about θ_0. Also, it is straightforward to show asymptotic normality of any of the iterated bias corrected estimators, $\hat{\theta}^{k}=\hat{\theta}-\bar{B}(\hat{\theta}^{k-1})/T$.

Under $n/T\rightarrow\rho$, Waterman et. al. (2000) derive the asymptotic distribution of a bias corrected estimator that is not feasible. Let $\widetilde{V}_{it}\equiv(v_{it},V_{2it})'$ and

$$\widetilde{\rho}_i\equiv(0,1)\left(E\left[\widetilde{V}_{it}\widetilde{V}_{it}'\right]\right)^{-1}E\left[\widetilde{V}_{it}U_{it}\right].$$

Their estimator, for scalar θ, is a solution to

$$\sum_{i=1}^{n}\sum_{t=1}^{T}\{\hat{u}_{it}(\theta)-\hat{V}_{2it}(\theta)\widetilde{\rho}_i\}=0.$$

This estimator is based on a modified estimating equation that is different than Woutersen's (2002). It is not feasible because $\widetilde{\rho}_i$ is unknown. In contrast, our estimator (and that of Woutersen, 2002) is based on a feasible bias correction, i.e. one that is computed from the data. Also, Theorem 2 is rigorous, allows for T to grow slower than n, and allows θ to be a vector.

The jackknife bias corrected estimator has the same limiting distribution as the analytic bias corrected estimator when T grows at the same rate as n. In order to simplify the proof we only show this for scalar θ.

Theorem 3 *Suppose that* $\dim(\theta)=1$. *Also suppose that Conditions 1, 2, 3, and 4, hold. We then have*

$$\sqrt{nT}\left(\widetilde{\theta}-\theta_0\right)\xrightarrow{d}N\left(0,\{\bar{E}[\mathcal{I}_i]\}^{-1}\right).$$

Proof. See Appendix C. ■

It seems plausible that $\hat{\theta}^{1}$ and $\widetilde{\theta}$ would be asymptotically efficient with respect to our asymptotics. Formal proof of such efficiency would require a Hajék-type convolution theorem, as established by Hahn and Kuersteiner (2002) for the case of panel AR(1) model with Gaussian innovation. Such a result is outside the scope of the current paper.

8 Summary

We developed two methods of reducing bias of the fixed effects maximum likelihood estimator for nonlinear panel models with fixed effects, an analytic method and an automatic method based on a panel jackknife. It would be useful to have some measure of how well these bias corrections work in a particular application. One could do this by estimating the extra term D in the expansion of equation (3). One could also seek to remove D by analytical and/or jackknife methods. Such analysis is expected to be substantially complicated, and we leave it to future research.

REFERENCES

References

[1] ABREVAYA, J. (1997), "The Equivalence of Two Estimators of the Fixed-Effects Logit Model," *Economics Letters* 55, 41-43.

[2] AMEMIYA, T. (1973), "Regression Analysis When the Dependent Variable is Truncated Normal," *Econometrica* 41, 997-1016.

[3] ANDERSON, E. (1970), "Asymptotic Properties of Conditional Maximum Likelihood Estimators," *Journal of the Royal Statistical Society*, Series B, 32, 283-301.

[4] ARELLANO, M. AND B. HONORÉ (2001), "Panel Data Models: Some Recent Developments," in J. J. Heckman and E. Leamer, eds., *Handbook of Econometrics, Vol. 5*, Amsterdam: North-Holland

[5] CHAMBERLAIN, G. (1980), "Analysis of Covariance with Qualitative Data," *Review of Economic Studies* 47, 225-238.

[6] CHAMBERLAIN, G. (1984), "Panel Data," in Z. GRILICHES AND M. INTRILIGATOR, eds., *Handbook of Econometrics, Vol. 2.* Amsterdam: North-Holland.

[7] FERNANDEZ-VAL, I. (2003), "Higher-Order Properties of GMM and GEL Estimators in Panel Data Models," working paper, MIT Department of Economics.

[8] FIRTH, D. (1993), "Bias Reduction of Maximum Likelihood Estimates," *Biometrika* 80, 27-38.

[9] GREENE, W.H. (2002), "The Bias of the Fixed Effects Estimator in Nonlinear Models," *unpublished manuscript*, New York University.

[10] 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.

[11] HAHN, J., G. KUERSTEINER, AND W. NEWEY (2002), "Higher Order Properties of Bootstrap and Jackknife Bias Corrected Maximum Likelihood Estimators," *unpublished manuscript.*

[12] HECKMAN, J. (1981), "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process and Some Monte Carlo Evidence," in MCFADDEN, D. AND C. MANSKI eds *Structural Analysis of Discrete Data*, Cambridge: MIT Press.

[13] HONORÉ, B. (1992), "Trimmed LAD and Least Squares Estimation of Truncated and Censored Models with Fixed Effects," *Econometrica* 60, 533-565.

[14] Horowitz, J.L., and S. Lee (2003), "Semiparametric Estimation of a Panel Data Proportional Hazard Model with Fixed Effects," *Journal of Econometrics*, forthcoming.

[15] LANCASTER, T. (2002), "Orthogonal Parameters and Panel Data," *Review of Economic Studies* 69, 647-666.

[16] MANSKI, C. (1987), "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data," Econometrica 55, 357-362.

[17] NEWEY, W.K. (1987), "Estimation of Nonlinear Models in Panel Data," NSF Proposal.

[18] NEYMAN, J., AND E.L. SCOTT, (1948), "Consistent Estimates Based on Partially Consistent Observations," *Econometrica* 16, 1-32.

[19] NICKELL, S. (1981), "Biases in Dynamic Models with Fixed Effects", *Econometrica* 49, pp. 1417 - 1426.

[20] QUENOUILLE, M.H. (1956), "Notes on Bias in Estimation," *Biometrika* 43, 353-360.

[21] RILSTONE, P., V. K. SRIVASTAVA, AND A. ULLAH (1996) "The Second-order Bias and Mean Squared Error of Nonlinear Estimators," *Journal of Econometrics* 75, 369-395.

[22] TUKEY, J.W. (1958), "Bias and Confidence in Not-Quite Large Samples (Abstract)," *Annals of Mathematical Statistics* 29, 614.

[23] WATERMAN, R., B. LINDSAY, H. LI, AND B. LI (2000), "Projected Score Methods for Nuisance Parameters: Asymptotics and Neyman-Scott Problems," *unpublished manuscript*.

[24] WHITE, H. (1982), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica* 50, 1-26.

[25] WOUTERSEN, T.M. (2002), "Robustness Against Incidental Parameters," *unpublished manuscript*, University of Western Ontario.

Appendix I: Relationship to Woutersen (2002)

by

Ivan Fernandez-Val

Let $E_{\theta,\alpha_i}[\bullet]$ denote the expectation with respect to the joint density $\Pi_{t=1}^{T} f(z_t|\theta,\alpha_i)$. Let

$$g_{it}(\theta,\alpha_i) = u_{it}(\theta,\alpha_i) - \{E_{\theta,\alpha_i}[u_{it\alpha}(\theta,\alpha_i)]/E_{\theta,\alpha_i}[v_{it\alpha}(\theta,\alpha_i)]\}v_{it}(\theta,\alpha_i),$$

Note that $g_{it} = U_{it}$, $E_{\theta,\alpha_i}[u_{it}(\theta,\alpha_i)] \equiv 0$, and $E_{\theta,\alpha_i}[v_{it}(\theta,\alpha_i)] \equiv 0$, so that $E_{\theta,\alpha_i}[g_{it}(\theta,\alpha_i)] \equiv 0$. Differentiating this identity twice with respect to α, assuming that differentiation inside the integral is allowed, and evaluating at the truth gives

$$0 = E[g_{it\alpha}] + E[U_{it}v_{it}], \qquad 0 = E[g_{it\alpha\alpha}] + 2E[g_{it\alpha}v_{it}] + E[U_{it}v_{it\alpha}] + E[U_{it}v_{it}^2].$$

Furthermore, by construction of g_{it} we have $E_{\theta,\alpha_i}[g_{it\alpha}(\theta,\alpha_i)] \equiv 0$, so that differentiating this identity with respect to α gives $E[g_{it\alpha\alpha}] + E[g_{it\alpha}v_{it}] = 0$. Plugging this in, and using the form of V_{2it}, gives $E[g_{it\alpha\alpha}] = E[U_{it}V_{2it}]$. By this equality it follows that another form of bias corrected estimating equation is

$$0 = \sum_{i=1}^{n}\sum_{t=1}^{T}\left[\widehat{u}_{it}(\theta) + \frac{1}{2}g_{it\alpha\alpha}(\theta,\hat{\alpha}_i(\theta))/\left(\sum_{s=1}^{T}\widehat{v}_{is}(\theta)^2\right)\right].$$

This estimating equation is exactly like that of Woutersen (2002), except that it is missing a term that is multiplied by $\sum_{t=1}^{T} g_{it\alpha}(\theta,\hat{\alpha}_i(\theta))/T$. Because $E[g_{it\alpha}] = 0$, this additional term will converge in probability to zero, so that dropping it does not affect the bias correction asymptotically. Thus, we find that the bias corrected estimating equation (9) is a modified version of Woutersen's (2002) estimating equation.

Appendix II

A Proof of Theorem 1

The proof consists of two steps. First, it is shown that $\widehat{\theta}$ is consistent. Second, asymptotic normality is established. Consistency proof is rather standard, and is presented in a supplementary appendix available upon request. There, it is shown that, under Conditions 1, 2, and 3, $\Pr\left[\left|\widehat{\theta}-\theta_0\right|\geq\eta\right]=o\left(T^{-1}\right)$, and $\Pr\left[\max_{1\leq i\leq n}\left|\widehat{\alpha}_i-\alpha_{i0}\right|\geq\eta\right]=o\left(T^{-1}\right)$ for every $\eta>0$. Proof of asymptotic normality starts with the observation that MLE $\widehat{\theta}$ also solves $0=\sum_{i=1}^n\sum_{t=1}^T U\left(z_{it};\widehat{\theta},\widehat{\alpha}_i\left(\widehat{\theta}\right)\right)$, where $U\left(z_{it};\theta,\alpha_i\right)\equiv u_{it}\left(\theta,\alpha_i\right)-\rho_{i0}\cdot v_{it}\left(\theta,\alpha_i\right)$, and $\rho_{i0}=E\left[u_{it}v_{it}\right]/E\left[v_{it}^2\right]$. For accounting purpose, it turns out to be convenient to attach the subscript i to U and V when they are evaluated at z_{it}. Let $F\equiv\left(F_1,\ldots,F_n\right)$ denote the collection of distribution functions. Let $\widehat{F}\equiv\left(\widehat{F}_1,\ldots,\widehat{F}_n\right)$, where $\widehat{F}_i$ denotes the empirical distribution function for the stratum i. Define $F\left(\epsilon\right)\equiv F+\epsilon\sqrt{T}\left(\widehat{F}-F\right)$ for $\epsilon\in\left[0,T^{-1/2}\right]$. For each fixed θ and ϵ, let $\alpha_i\left(\theta,F_i\left(\epsilon\right)\right)$ and $\theta\left(F\left(\epsilon\right)\right)$ be the solutions to the estimating equations

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

By Taylor series expansion, we have $\theta\left(\widehat{F}\right)-\theta\left(F\right)=\frac{1}{\sqrt{T}}\theta^{\epsilon}\left(0\right)+\frac{1}{2}\left(\frac{1}{\sqrt{T}}\right)^2\theta^{\epsilon\epsilon}\left(0\right)+\frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^3\theta^{\epsilon\epsilon\epsilon}\left(\widetilde{\epsilon}\right)$, where $\theta^{\epsilon}\left(\epsilon\right)\equiv d\theta\left(F\left(\epsilon\right)\right)/d\epsilon$, $\theta^{\epsilon\epsilon}\left(\epsilon\right)\equiv d^2\theta\left(F\left(\epsilon\right)\right)/d\epsilon^2,\ldots$, and $\widetilde{\epsilon}$ is somewhere in between 0 and $T^{-1/2}$. We therefore have

$$\sqrt{nT}\left(\theta\left(\widehat{F}\right)-\theta\left(F\right)\right)=\sqrt{nT}\frac{1}{\sqrt{T}}\theta^{\epsilon}\left(0\right)+\sqrt{nT}\frac{1}{2}\left(\frac{1}{\sqrt{T}}\right)^2\theta^{\epsilon\epsilon}\left(0\right)+\frac{1}{6}\sqrt{\frac{n}{T}}\frac{1}{\sqrt{T}}\theta^{\epsilon\epsilon\epsilon}\left(\widetilde{\epsilon}\right). \tag{11}$$

In a supplementary appendix available upon request, it is shown that the last term in (11) is $o_p\left(1\right)$ under Condition 4. It is shown below that

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

from which Theorem 1 follows.

Let

$$h_i\left(\cdot,\epsilon\right)\equiv U_i\left(\cdot;\theta\left(F\left(\epsilon\right)\right),\alpha_i\left(\theta\left(F\left(\epsilon\right)\right),F_i\left(\epsilon\right)\right)\right) \tag{12}$$

The first order condition may be written as

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

Differentiating repeatedly with respect to ϵ, we obtain

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

$$0=\frac{1}{n}\sum_{i=1}^n\int\frac{d^2h_i\left(\cdot,\epsilon\right)}{d\epsilon^2}dF_i\left(\epsilon\right)+2\frac{1}{n}\sum_{i=1}^n\int\frac{dh_i\left(\cdot,\epsilon\right)}{d\epsilon}d\Delta_{iT} \tag{15}$$

$$0=\frac{1}{n}\sum_{i=1}^n\int\frac{d^3h_i\left(\cdot,\epsilon\right)}{d\epsilon^3}dF_i\left(\epsilon\right)+3\frac{1}{n}\sum_{i=1}^n\int\frac{d^2h_i\left(\cdot,\epsilon\right)}{d\epsilon^2}d\Delta_{iT} \tag{16}$$

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

A.1 $\theta^{\epsilon}(0)$

Because $\frac{dh_i(\cdot,\epsilon)}{d\epsilon} = \frac{\partial h_i(\cdot,\epsilon)}{\partial\theta'}\frac{\partial\theta}{\partial\epsilon} + \frac{\partial h_i(\cdot,\epsilon)}{\partial\alpha_i}\frac{\partial\alpha_i}{\partial\theta'}\frac{\partial\theta}{\partial\epsilon} + \frac{\partial h_i(\cdot,\epsilon)}{\partial\alpha_i}\frac{\partial\alpha_i}{\partial\epsilon}$, we may rewrite (14) as

$$0 = \frac{1}{n}\sum_{i=1}^{n}\int\left(\frac{\partial h_i(\cdot,\epsilon)}{\partial\theta'}\frac{\partial\theta}{\partial\epsilon} + \frac{\partial h_i(\cdot,\epsilon)}{\partial\alpha_i}\frac{\partial\alpha_i}{\partial\theta'}\frac{\partial\theta}{\partial\epsilon} + \frac{\partial h_i(\cdot,\epsilon)}{\partial\alpha_i}\frac{\partial\alpha_i}{\partial\epsilon}\right)dF_i(\epsilon) + \frac{1}{n}\sum_{i=1}^{n}\int h_i(\cdot,\epsilon)\,d\Delta_{iT} \tag{17}$$

Evaluating at $\epsilon = 0$, and noting that $E[U_i^{\alpha_i}] = 0$, we obtain

$$\theta^{\epsilon}(0) = \left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\int U_i d\Delta_{iT}\right) \tag{18}$$

We therefore have

$$\sqrt{nT}\frac{1}{\sqrt{T}}\theta^{\epsilon}(0) = \left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\left(\frac{1}{\sqrt{n}\sqrt{T}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_i\right) \to N\left(0,\left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\right)$$

A.2 α_i^{θ} and α_i^{ϵ}

In the ith stratum, $\alpha_i(\theta, F_i(\epsilon))$ solves the estimating equation

$$\int V_i(\cdot;\theta,\alpha_i(\theta,F_i(\epsilon)))\,dF_i(\epsilon) = 0 \tag{19}$$

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

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

Observe that

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

Evaluating the derivatives of α_i evaluated at $\epsilon = 0$ gives

$$\alpha_i^{\theta} = -\frac{E\left[\frac{\partial V_i}{\partial\theta}\right]}{E\left[\frac{\partial V_i}{\partial\alpha_i}\right]} = \frac{E\left[\frac{\partial V_i}{\partial\theta}\right]}{E[V_i^2]} = O(1), \tag{20}$$

$$\alpha_i^{\epsilon} = -\frac{\sum_{t=1}^{T}V_{it}}{\sqrt{T}E\left[\frac{\partial V_i}{\partial\alpha_i}\right]} = \frac{T^{-1/2}\sum_{t=1}^{T}V_{it}}{E[V_i^2]} = O_p(1), \tag{21}$$

where

$$\alpha_i^{\theta} \equiv \frac{\partial\alpha_i(\theta,F_i(0))}{\partial\theta}, \qquad \alpha_i^{\epsilon} \equiv \frac{\partial\alpha_i(\theta,F_i(0))}{\partial\epsilon}.$$

A.3 $\theta^{\epsilon\epsilon}(0)$

Note that

$$\begin{aligned}
\frac{d^2 h_i(\cdot,\epsilon)}{d\epsilon^2} &= \mathcal{G}_i(\cdot,\epsilon) + \frac{\partial^2 h_i(\cdot,\epsilon)}{\partial\theta'\partial\alpha_i}\left(\frac{\partial\alpha_i}{\partial\theta'}\frac{\partial\theta}{\partial\epsilon}\right)\frac{\partial\theta}{\partial\epsilon} + \frac{\partial^2 h_i(\cdot,\epsilon)}{\partial\theta'\partial\alpha_i}\frac{\partial\alpha_i}{\partial\epsilon}\frac{\partial\theta}{\partial\epsilon} + \frac{\partial h_i(\cdot,\epsilon)}{\partial\theta'}\frac{\partial^2\theta}{\partial\epsilon^2} \\
&+ \frac{\partial^2 h_i(\cdot,\epsilon)}{\partial\theta'\partial\alpha_i}\frac{\partial\theta}{\partial\epsilon}\frac{\partial\alpha_i}{\partial\theta'}\frac{\partial\theta}{\partial\epsilon} + \frac{\partial^2 h_i(\cdot,\epsilon)}{\partial\alpha_i^2}\left(\frac{\partial\alpha_i}{\partial\theta'}\frac{\partial\theta}{\partial\epsilon}\right)^2 + \frac{\partial^2 h_i(\cdot,\epsilon)}{\partial\alpha_i^2}\left(\frac{\partial\alpha_i}{\partial\theta'}\frac{\partial\theta}{\partial\epsilon}\right)\frac{\partial\alpha_i}{\partial\epsilon} \\
&+ \frac{\partial h_i(\cdot,\epsilon)}{\partial\alpha_i}\left(\frac{\partial\theta'}{\partial\epsilon}\frac{\partial^2\alpha_i}{\partial\theta\partial\theta'}\right)\frac{\partial\theta}{\partial\epsilon} + \frac{\partial h_i(\cdot,\epsilon)}{\partial\alpha_i}\frac{\partial^2\alpha_i}{\partial\theta'\partial\epsilon}\frac{\partial\theta}{\partial\epsilon} + \frac{\partial h_i(\cdot,\epsilon)}{\partial\alpha_i}\frac{\partial\alpha_i}{\partial\theta'}\frac{\partial^2\theta}{\partial\epsilon^2} \\
&+ \frac{\partial^2 h_i(\cdot,\epsilon)}{\partial\theta'\partial\alpha_i}\frac{\partial\theta}{\partial\epsilon}\frac{\partial\alpha_i}{\partial\epsilon} + \frac{\partial^2 h_i(\cdot,\epsilon)}{\partial\alpha_i^2}\left(\frac{\partial\alpha_i}{\partial\theta'}\frac{\partial\theta}{\partial\epsilon}\right)\frac{\partial\alpha_i}{\partial\epsilon} + \frac{\partial^2 h_i(\cdot,\epsilon)}{\partial\alpha_i^2}\left(\frac{\partial\alpha_i}{\partial\epsilon}\right)^2 \\
&+ \frac{\partial h_i(\cdot,\epsilon)}{\partial\alpha_i}\frac{\partial^2\alpha_i}{\partial\epsilon\partial\theta'}\frac{\partial\theta}{\partial\epsilon} + \frac{\partial h_i(\cdot,\epsilon)}{\partial\alpha_i}\frac{\partial^2\alpha_i}{\partial\epsilon^2},
\end{aligned}$$

where $\mathcal{G}_i(\cdot,\epsilon)$ is a p-dimensional column vector such that its r-th element $\mathcal{G}_i^{(r)}(\cdot,\epsilon)$ is equal to $\mathcal{G}_i^{(r)}(\cdot,\epsilon) = \frac{\partial\theta(\cdot,\epsilon)'}{\partial\epsilon}\frac{\partial^2 h_i^{(r)}(\cdot,\epsilon)}{\partial\theta\partial\theta'}\frac{\partial\theta(\cdot,\epsilon)}{\partial\epsilon}$, and $h_i^{(r)}(\cdot,\epsilon)$ denotes the r-th element of h_i.

Evaluating each term of (15) at $\epsilon = 0$, and noting that $E[U_i^{\alpha_i}] = 0$, we obtain

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

where

$$\mathcal{G} = \left[\theta^\epsilon(0)'\left(\frac{1}{n}\sum_{i=1}^n E\left[\frac{\partial^2 U_i^{(1)}}{\partial\theta\partial\theta'}\right]\right)\theta^\epsilon(0),\cdots,\theta^\epsilon(0)'\left(\frac{1}{n}\sum_{i=1}^n E\left[\frac{\partial^2 U_i^{(p)}}{\partial\theta\partial\theta'}\right]\right)\theta^\epsilon(0)\right]'$$

from which we obtain

$$\begin{aligned}
\left(\frac{1}{n}\sum_{i=1}^n\mathcal{I}_i\right)\theta^{\epsilon\epsilon}(0) &= \frac{1}{n}\sum_{i=1}^n(\alpha_i^\epsilon)^2\cdot E\left[\frac{\partial^2 U_i}{\partial\alpha_i^2}\right] + \frac{2}{n}\sum_{i=1}^n\alpha_i^\epsilon\cdot\int\frac{\partial U_i}{\partial\alpha_i}d\Delta_{iT} \\
&+ 2\left(\frac{1}{n}\sum_{i=1}^n\alpha_i^\epsilon\cdot E\left[\frac{\partial^2 U_i}{\partial\theta'\partial\alpha_i}\right]\right)\theta^\epsilon(0) + \left(\frac{1}{n}\sum_{i=1}^n\alpha_i^\epsilon E\left[\frac{\partial^2 U_i}{\partial\alpha_i^2}\right]\left(\alpha_i^\theta\right)'\right)\theta^\epsilon(0) \\
&+ 2\left(\frac{1}{n}\sum_{i=1}^n\int\frac{\partial U_i}{\partial\theta'}d\Delta_{iT}\right)\theta^\epsilon(0) + \left(\frac{1}{n}\sum_{i=1}^n\int\frac{\partial U_i}{\partial\alpha_i}d\Delta_{iT}\cdot\left(\alpha_i^\theta\right)'\right)\theta^\epsilon(0) + \mathcal{G} \\
&+ \frac{2}{n}\sum_{i=1}^n\theta^\epsilon(0)'\alpha_i^\theta\cdot E\left[\frac{\partial^2 U_i}{\partial\theta'\partial\alpha_i}\right]\theta^\epsilon(0) + \frac{1}{n}\sum_{i=1}^n\left(\theta^\epsilon(0)'\alpha_i^\theta\right)^2\cdot E\left[\frac{\partial^2 U_i}{\partial\alpha_i^2}\right] \qquad (22)
\end{aligned}$$

Because

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n(\alpha_i^\epsilon)^2\cdot E\left[\frac{\partial^2 U_i}{\partial\alpha_i^2}\right] &= \frac{1}{n}\sum_{i=1}^n E\left[U_i^{\alpha_i\alpha_i}\right]\left(\frac{\sum_{t=1}^T V_{it}}{\sqrt{T}E\left[V_i^2\right]}\right)^2 \\
\sum_{i=1}^n\alpha_i^\epsilon\cdot\int\frac{\partial U_i}{\partial\alpha_i}d\Delta_{iT} &= \frac{1}{n}\sum_{i=1}^n\frac{\sum_{t=1}^T V_{it}}{\sqrt{T}E\left[V_i^2\right]}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T\left(U_i^{\alpha_i} - E\left[U_i^{\alpha_i}\right]\right)\right)
\end{aligned}$$

$$\left(\frac{1}{n}\sum_{i=1}^{n}\alpha_i^{\epsilon}\cdot E\left[\frac{\partial^2 U_i}{\partial\theta'\partial\alpha_i}\right]\right)\theta^{\epsilon}(0) = O_p\left(\frac{1}{\sqrt{n}}\right)O_p\left(\frac{1}{\sqrt{n}}\right)=O_p\left(\frac{1}{n}\right)$$

$$\left(\frac{1}{n}\sum_{i=1}^{n}\alpha_i^{\epsilon}E\left[\frac{\partial^2 U_i}{\partial\alpha_i^2}\right]\left(\alpha_i^{\theta}\right)'\right)\theta^{\epsilon}(0) = O_p\left(\frac{1}{\sqrt{n}}\right)O_p\left(\frac{1}{\sqrt{n}}\right)=O_p\left(\frac{1}{n}\right)$$

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

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

and

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

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

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

we may write

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

Therefore, we have

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

The terms in square parentheses have joint limiting normal distributions by CLT. Their product is a quadratic form with mean

$$\frac{E\left[V_{it}U_i^{\alpha_i}\right]}{E\left[V_i^2\right]}+\frac{E\left[U_i^{\alpha_i\alpha_i}\right]}{2E\left[V_i^2\right]}=\frac{E\left[V_{it}U_i^{\alpha_i}\right]}{E\left[V_i^2\right]}-\frac{E\left[V_{it}U_i^{\alpha_i}\right]}{2E\left[V_i^2\right]}=\frac{E\left[V_{it}U_i^{\alpha_i}\right]}{2E\left[V_i^2\right]}=-\frac{E\left[V_{2it}U_i\right]}{2E\left[V_i^2\right]}.$$

Therefore, we have

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

B Proof of Theorem 2

Proof of the theorem under the stronger condition where $n=O(T)$ is be provided here. (More general proof under the condition $n=o\left(T^3\right)$ is available as Supplementary Appendix II upon request.) It suffices to show that

$\widehat{\beta} = \beta + o_p(1)$. The conclusion easily follows from $\sqrt{nT} \cdot \frac{1}{T}\widehat{\beta} = \sqrt{\frac{n}{T}}\widehat{\beta} = \beta\sqrt{\rho} + o_p(1)$. Note that for $M_{it} = M_{it}(z_{it})$

$$\left|\frac{1}{T}\sum_{t=1}^{T}\hat{V}_{2it}\left(\widehat{\theta}\right)\hat{u}_{it}\left(\widehat{\theta}\right) - E\left[V_{2it}u_{it}\right]\right| \leq \left|\frac{1}{T}\sum_{t=1}^{T}\hat{V}_{2it}\left(\widehat{\theta}\right)\hat{u}_{it}\left(\widehat{\theta}\right) - \frac{1}{T}\sum_{t=1}^{T}V_{2it}u_{it}\right| + \left|\frac{1}{T}\sum_{t=1}^{T}V_{2it}u_{it} - E\left[V_{2it}u_{it}\right]\right|$$
$$\leq \left(\max_i E\left[M_{it}^2\right] + \left|\frac{1}{T}\sum_{t=1}^{T}\left(M_{it}^2 - E\left[M_{it}^2\right]\right)\right|\right)\left(\left|\widehat{\theta} - \theta_0\right| + \max_i\left|\widehat{\alpha}_i - \alpha_{i0}\right|\right) + \left|\frac{1}{T}\sum_{t=1}^{T}V_{2it}u_{it} - E\left[V_{2it}u_{it}\right]\right|$$

By Lemma 2 in Supplementary Appendix I, we obtain

$$\max_i\left|\frac{1}{T}\sum_{t=1}^{T}\left(M_{it}^2 - E\left[M_{it}^2\right]\right)\right| = o_p(1), \qquad \max_i\left|\frac{1}{T}\sum_{t=1}^{T}V_{2it}u_{it} - E\left[V_{2it}u_{it}\right]\right| = o_p(1).$$

Combined with Lemmas 4 and 5 in Supplementary Appendix I, we obtain

$$\max_i\left|\frac{1}{T}\sum_{t=1}^{T}\hat{V}_{2it}\left(\widehat{\theta}\right)u_{it}\left(\hat{\theta}\right) - E\left[V_{2it}u_{it}\right]\right| = o_p(1)$$

We can similarly show that

$$\max_i\left|\frac{1}{T}\sum_{t=1}^{T}\hat{v}_{it}\left(\widehat{\theta}\right)^2 - E\left[v_{it}^2\right]\right| = o_p(1), \qquad \max_i\left|\frac{1}{T}\sum_{t=1}^{T}\hat{V}_{2it}\left(\widehat{\theta}\right)\hat{v}_{it}\left(\widehat{\theta}\right) - E\left[V_{2it}v_{it}\right]\right| = o_p(1).$$

$$\max_i\left|\frac{1}{T}\sum_{t=1}^{T}\hat{u}_{it}\left(\widehat{\theta}\right)^2 - E\left[u_{it}^2\right]\right| = o_p(1), \qquad \max_i\left|\frac{1}{T}\sum_{t=1}^{T}u_{it}\left(\widehat{\theta}\right)v_{it}\left(\widehat{\theta}\right) - E[u_{it}v_{it}]\right| = o_p(1).$$

It follows that

$$\left|\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\widehat{U}_{it}\widehat{U}_{it}' - \frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right| = o_p(1), \left|\frac{1}{n}\sum_{i=1}^{n}\frac{\frac{1}{T}\sum_{t=1}^{T}\widehat{U}_{it}\widehat{V}_{2it}}{\frac{1}{T}\sum_{t=1}^{T}\widehat{V}_{it}^2} - \frac{1}{n}\sum_{i=1}^{n}\frac{E\left[V_{2it}U_{it}\right]}{E\left[V_{it}^2\right]}\right| = o_p(1)$$

from which we obtain

$$\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\widehat{U}_{it}\widehat{U}_{it}' = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i + o_p(1) \tag{23}$$

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\frac{1}{T}\sum_{t=1}^{T}\widehat{U}_{it}\widehat{V}_{2it}}{\frac{1}{T}\sum_{t=1}^{T}\widehat{V}_{it}^2} = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\frac{E\left[V_{2it}U_{it}\right]}{E\left[V_{it}^2\right]} + o_p(1) \tag{24}$$

C Proof of Theorem 3

By Theorem 7 in Supplementary Appendix I, we have $\Pr\left[\left|\frac{1}{\sqrt{T}}\theta^{\epsilon\epsilon\epsilon\epsilon\epsilon\epsilon}\left(\widetilde{\epsilon}\right)\right| \geq \eta\right] = o\left(\frac{1}{T}\right)$. Therefore,

$$T \cdot \frac{1}{120}\sqrt{\frac{n}{T}}\frac{1}{T\sqrt{T}}\theta^{\epsilon\epsilon\epsilon\epsilon\epsilon\epsilon}\left(\widetilde{\epsilon}\right) = o_p(1)$$

and

$$\left|T \cdot \frac{1}{T}\sum_{t=1}^{T}\frac{1}{120}\sqrt{\frac{n}{T-1}}\frac{1}{(T-1)\sqrt{T-1}}\theta_{(t)}^{\epsilon\epsilon\epsilon\epsilon\epsilon\epsilon}\left(\widetilde{\epsilon}_{(t)}\right)\right| \leq \frac{1}{120}\sqrt{\frac{n}{T-1}}\frac{T}{(T-1)\sqrt{T-1}}\max_t\left|\theta_{(t)}^{\epsilon\epsilon\epsilon\epsilon\epsilon\epsilon}\left(\widetilde{\epsilon}_{(t)}\right)\right|$$
$$= O\left(\max_t\left|\frac{1}{\sqrt{T}}\theta_{(t)}^{\epsilon\epsilon\epsilon\epsilon\epsilon\epsilon}\left(\widetilde{\epsilon}_{(t)}\right)\right|\right) = o_p(1),$$

where $\theta_{(t)}^{\epsilon\epsilon\epsilon\epsilon\epsilon\epsilon}$ denotes the delete-t version of $\theta_{(t)}^{\epsilon\epsilon\epsilon\epsilon\epsilon\epsilon}$, and the second equality is based on

$$\Pr\left[\max_t\left|\frac{1}{\sqrt{T}}\theta_{(t)}^{\epsilon\epsilon\epsilon\epsilon\epsilon\epsilon}\left(\widetilde{\epsilon}_{(t)}\right)\right|\geq\eta\right]\leq\sum_t\Pr\left[\left|\frac{1}{\sqrt{T}}\theta_{(t)}^{\epsilon\epsilon\epsilon\epsilon\epsilon\epsilon}\left(\widetilde{\epsilon}_{(t)}\right)\right|\geq\eta\right]=T\cdot o\left(\frac{1}{T}\right)=o\left(1\right).$$

We may therefore write

$$\begin{aligned}\sqrt{nT}\left(\widetilde{\theta}-\theta_0\right) &= T\left(\sqrt{nT}\left(\widehat{\theta}-\theta_0\right)-\sqrt{\frac{T}{T-1}}\frac{1}{T}\sum_{t=1}^{T}\sqrt{n\left(T-1\right)}\left(\widehat{\theta}_{(t)}-\theta_0\right)\right)\\ &\quad+\sqrt{\frac{T}{T-1}}\frac{1}{T}\sum_{t=1}^{T}\sqrt{n\left(T-1\right)}\left(\widehat{\theta}_{(t)}-\theta_0\right)\end{aligned}$$

or

$$\begin{aligned}&\sqrt{nT}\left(\widetilde{\theta}-\theta_0\right)\\ &= T\left(\sqrt{n}\theta^{\epsilon}\left(0\right)-\sqrt{\frac{T}{T-1}}\frac{1}{T}\sum_{t=1}^{T}\sqrt{n}\theta_{(t)}^{\epsilon}\left(0\right)\right)+\frac{1}{2}T\left(\sqrt{\frac{n}{T}}\theta^{\epsilon\epsilon}\left(0\right)-\sqrt{\frac{T}{T-1}}\frac{1}{T}\sum_{t=1}^{T}\sqrt{\frac{n}{T-1}}\theta_{(t)}^{\epsilon\epsilon}\left(0\right)\right)\\ &\quad+\frac{1}{6}T\left(\sqrt{\frac{n}{T}}\frac{1}{\sqrt{T}}\theta^{\epsilon\epsilon\epsilon}\left(0\right)-\sqrt{\frac{T}{T-1}}\frac{1}{T}\sum_{t=1}^{T}\sqrt{\frac{n}{T-1}}\frac{1}{\sqrt{T-1}}\theta_{(t)}^{\epsilon\epsilon\epsilon}\left(0\right)\right)\\ &\quad+\frac{1}{24}T\left(\sqrt{\frac{n}{T}}\frac{1}{T}\theta^{\epsilon\epsilon\epsilon\epsilon}\left(0\right)-\sqrt{\frac{T}{T-1}}\frac{1}{T}\sum_{t=1}^{T}\sqrt{\frac{n}{T-1}}\frac{1}{T-1}\theta_{(t)}^{\epsilon\epsilon\epsilon\epsilon}\left(0\right)\right)\\ &\quad+\sqrt{\frac{T}{T-1}}\frac{1}{T}\sum_{t=1}^{T}\sqrt{n}\theta_{(t)}^{\epsilon}\left(0\right)+\frac{1}{2}\sqrt{\frac{T}{T-1}}\frac{1}{T}\sum_{t=1}^{T}\sqrt{\frac{n}{T-1}}\theta_{(t)}^{\epsilon\epsilon}\left(0\right)\\ &\quad+\frac{1}{6}\sqrt{\frac{T}{T-1}}\frac{1}{T}\sum_{t=1}^{T}\sqrt{\frac{n}{T-1}}\frac{1}{\sqrt{T-1}}\theta_{(t)}^{\epsilon\epsilon\epsilon}\left(0\right)+\frac{1}{24}\sqrt{\frac{T}{T-1}}\frac{1}{T}\sum_{t=1}^{T}\sqrt{\frac{n}{T-1}}\frac{1}{T-1}\theta_{(t)}^{\epsilon\epsilon\epsilon\epsilon}\left(0\right)+o_p\left(1\right)\end{aligned}$$

In a supplementary appendix available upon request, it is shown that

$$T\left(\sqrt{n}\theta^{\epsilon}\left(0\right)-\sqrt{\frac{T}{T-1}}\frac{1}{T}\sum_{t=1}^{T}\sqrt{n}\theta_{(t)}^{\epsilon}\left(0\right)\right)+\sqrt{\frac{T}{T-1}}\frac{1}{T}\sum_{t=1}^{T}\sqrt{n}\theta_{(t)}^{\epsilon}\left(0\right)=\sqrt{n}\theta^{\epsilon}\left(0\right)$$

$$\frac{1}{2}T\left(\sqrt{\frac{n}{T}}\theta^{\epsilon\epsilon}\left(0\right)-\sqrt{\frac{T}{T-1}}\frac{1}{T}\sum_{t=1}^{T}\sqrt{\frac{n}{T-1}}\theta_{(t)}^{\epsilon\epsilon}\left(0\right)\right)+\frac{1}{2}\sqrt{\frac{T}{T-1}}\frac{1}{T}\sum_{t=1}^{T}\sqrt{\frac{n}{T-1}}\theta_{(t)}^{\epsilon\epsilon}\left(0\right)=o_p\left(1\right)$$

$$T\left(\sqrt{\frac{n}{T}}\frac{1}{\sqrt{T}}\theta^{\epsilon\epsilon\epsilon}\left(0\right)-\sqrt{\frac{T}{T-1}}\frac{1}{T}\sum_{t=1}^{T}\sqrt{\frac{n}{T-1}}\frac{1}{\sqrt{T-1}}\theta_{(t)}^{\epsilon\epsilon\epsilon}\left(0\right)\right)+\sqrt{\frac{T}{T-1}}\frac{1}{T}\sum_{t=1}^{T}\sqrt{\frac{n}{T-1}}\frac{1}{\sqrt{T-1}}\theta_{(t)}^{\epsilon\epsilon\epsilon}\left(0\right)=o_p\left(1\right)$$

$$T\left(\sqrt{\frac{n}{T}}\frac{1}{T}\theta^{\epsilon\epsilon\epsilon\epsilon}\left(0\right)-\sqrt{\frac{T}{T-1}}\frac{1}{T}\sum_{t=1}^{T}\sqrt{\frac{n}{T-1}}\frac{1}{T-1}\theta_{(t)}^{\epsilon\epsilon\epsilon\epsilon}\left(0\right)\right)+\sqrt{\frac{T}{T-1}}\frac{1}{T}\sum_{t=1}^{T}\sqrt{\frac{n}{T-1}}\frac{1}{T-1}\theta_{(t)}^{\epsilon\epsilon\epsilon\epsilon}\left(0\right)=o_p\left(1\right)$$

from which the conclusion follows.