A likelihood-based approximate solution to the incidental parameter problem in dynamic nonlinear models with multiple effects

Manuel Arellano
CEMFI

Jinyong Hahn
UCLA

October 2006

1 Introduction

There is a body of well understood nonlinear models in econometrics, which are routinely estimated by maximum likelihood or related methods using cross-sectional or time series data. These include, to name a few, discrete choice, conditional volatility, or duration models. In panel data applications of these models, a leading motivation is to exploit the time series variability to allow for heterogeneity in some of the coefficients, which is a powerful way of addressing endogeneity concerns.

Unfortunately, when the time series dimension T is small relative to the cross-sectional dimension n, ML estimates of the common parameters or other average effects can be severely biased, specially in dynamic models. This is reflected in asymptotic results such as the fixed-T inconsistency of the ML estimator for some models, or the lack of identification of the model's parameters in a large n fixed-T population for others. Sometimes it is possible to obtain fixed T large n consistent estimators of certain common parameters, based on features of the distribution of the data that do not depend on individual-specific parameters. Nevertheless, situations of this type are more the exception than the rule from the point of view of the needs of applied work.

A useful question is to ask how much heterogeneity can be given empirical content in a particular panel model and data set. One could, for example, expect time series of size 10 to 20 to be statistically informative for up to two or three different coefficients for certain processes. From this perspective, it is natural to choose a population framework that does not rule out the possibility of statistical learning from individual time series in panel data, so that both T and n tend to infinity. If T is statistically informative but much smaller than n, as is often the case with micropanels, this should be reflected in the choice of methods of estimation and inference. For example, by seeking estimators with biases of order $1/T^2$ or less as opposed to the standard magnitude of $1/T$, and asymptotic approximations where n/T or n/T^3 converge to a constant.

Such is the goal of the recent literature on bias-adjusted estimation methods for nonlinear panel data models with fixed effects. Three different approaches can be distinguished in this literature. One approach is to construct and analytical or numerical bias correction of a fixed effects estimator. Hahn and Newey (2002) considered corrections of this type for static nonlinear panel data models when n and T increase at the same rate, and Hahn and Kuersteiner (2004) provided a similar analysis for dynamic models. A second approach is to consider estimators from bias corrected moment equations. Estimators of this type have been discussed in Woutersen (2002), Arellano (2003), Carro (2004), and Fernandez-Val (2005), amongst others. Finally, a third approach is to consider estimation from a bias corrected objective function relative to some target criterion. Adjustments of this type were discussed in Pace and Salvan (2005) for a generic concentrated likelihood with independent observations, and in Arellano and Hahn (2006) for static nonlinear panel models.[1]

In this paper we consider a modified objective function strategy to obtain estimators without bias

[1] See also Arellano and Hahn (2006) for a review of the literature.

to order $1/T$ in nonlinear dynamic panel models with multiple effects. We consider two approaches to bias correct the objective function, both of which depend on a Hessian term and an outer product of score term, the latter depending on the dynamic dependence of the score. One approach uses a determinant based correction, which we argue later is appropriate in likelihood settings. When the model fully specifies the distribution of the data, it is possible to obtain the expected outer product term and we discuss this possibility. The other approach uses a trace based correction, which we show later is not restricted to the likelihood setup, and is based on a trimmed outer product matrix of the sample score vector. The trace based approach has been independently discussed in a recent paper by Bester and Hansen (2005) as the integral of a bias-corrected moment equation.

Aside from being criterion based, an advantage of these estimators is the great simplicity and transparency of the required corrections by comparison with bias corrections of estimators or moment equations, specially in models with multiple effects. Another benefit of our approach is that bias corrected objective functions can be related to various modifications of the concentrated likelihood suggested in the statistical literature as approximations to conditional or marginal likelihood functions. For example, the determinant based approach is analogous to the Cox and Reid (1987)'s adjusted profile likelihood approach when fixed effects are information orthogonal to common parameters.

We analyze the asymptotic properties of both trace based and determinant based estimators when n and T grow at the same rate, and show that they are asymptotically normal and centered at the truth. Our strategy is to develop a theory for general bias corrected estimating equations, so that we can obtain asymptotic results for a specific bias correction method using the first order conditions.

The paper is organized as follows. Section 2 explains how bias correction of the objective function works. Section 3 presents some examples. Section 4 gives the asymptotic theory. Finally, a brief conclusion is in Section 5. Proofs and technical details are given in the Appendix.

2 Correcting the Objective Function

Let the data be denoted by x_{it} $(t = 1, ..., T; i = 1, ..., n)$. Suppose that we are given a panel data model with a common parameter of interest θ_0 and potentially vector-valued individual specific fixed effects γ_{i0}, $i = 1, \ldots, n$. We consider a maximization estimator defined by

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

for some criterion function $\psi(\cdot)$ that does not depend on T. Here, ψ is a sensible function in the sense that, if n is fixed, and $T \to \infty$, the estimator $\left(\widehat{\theta}, \widehat{\gamma}_1, \ldots, \widehat{\gamma}_n\right)$ is consistent for $(\theta_0, \gamma_{10}, \ldots, \gamma_{n0})$.

In a likelihood setup, we assume that $x_{it} = (y_{it}, y_{i,t-1}, \ldots, y_{i,t-q})$ and

$$\psi\left(x_{it}; \theta, \gamma_i\right) = \ln p_c\left(y_{it} \mid y_{i,t-1}, \ldots, y_{i,t-q}; \theta, \gamma_i\right),$$

where p_c denotes the conditional density of y_{it}.[2]

Letting $\widehat{\gamma}_i(\theta) \equiv \operatorname{argmax}_a \sum_{t=1}^T \psi(x_{it};\theta,a)$, we can characterize $\widehat{\theta}$ as the estimator that maximizes the concentrated objective function

$$\widehat{\theta} = \underset{\theta}{\operatorname{argmax}} \frac{1}{n}\sum_{i=1}^n \overline{\psi}_i(\theta,\widehat{\gamma}_i(\theta))$$

where

$$\overline{\psi}_i(\theta,\gamma_i) \equiv \frac{1}{T}\sum_{t=1}^T \psi(x_{it};\theta,\gamma_i).$$

Now let θ_T be the value that maximizes the limiting expected concentrated objective function for fixed T:

$$\theta_T \equiv \underset{\theta}{\operatorname{argmax}} \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n E\left[\overline{\psi}_i(\theta,\widehat{\gamma}_i(\theta))\right].$$

Due to the noise in estimating $\widehat{\gamma}_i(\theta)$, in general $\theta_T \neq \theta_0$ (Neyman and Scott (1948)'s incidental parameters problem). This problem would not occur if the quantities $\widehat{\gamma}_i(\theta)$ were replaced by $\gamma_i(\theta)$ defined as[3]

$$\gamma_i(\theta) \equiv \underset{c}{\operatorname{argmax}} \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^T E\left[\psi(x_{it};\theta,c)\right]. \tag{2}$$

So we could think of the infeasible concentrated objective function $\sum_{i=1}^n \overline{\psi}_i(\theta,\gamma_i(\theta))/n$ as a target criterion and $\sum_{i=1}^n \overline{\psi}_i(\theta,\widehat{\gamma}_i(\theta))/n$ as a plug-in estimate with a bias of order $1/T$. The source of incidental parameter bias is that the concentrated objective function is itself a biased estimate of the target criterion. This suggests maximizing a modified objective function that has no bias up to a certain order in T.

For smooth objective functions, the bias in the expected concentrated function at an arbitrary θ can be usually expanded in orders of magnitude of T:

$$\lim_{n\to\infty} E\left[\frac{1}{n}\sum_{i=1}^n \overline{\psi}_i(\theta,\widehat{\gamma}_i(\theta)) - \frac{1}{n}\sum_{i=1}^n \overline{\psi}_i(\theta,\gamma_i(\theta))\right] = \frac{1}{T}B(\theta) + o\left(\frac{1}{T}\right) \tag{3}$$

for some $B(\theta)$.

A bias corrected concentrated objective function is to plug into the formula for $B(\theta)$ estimators of its unknown components to construct $\widehat{B}(\theta)$, and then obtain an estimator that maximizes the adjusted criterion:

$$\widetilde{\theta} \equiv \arg\max_{\theta} \left(\frac{1}{n}\sum_{i=1}^n \overline{\psi}_i(\theta,\widehat{\gamma}_i(\theta)) - \frac{1}{T}\widehat{B}(\theta)\right). \tag{4}$$

[2] We abstract away from strictly exogenous regressors. For shortness we may write $\psi_{it}(\theta,\gamma_i) = \psi(x_{it};\theta,\gamma_i)$.

[3] Note that $\gamma_i(\theta_0) = \gamma_{i0}$ and that in the likelihood setup $\gamma_i(\theta)$ is fully determined by θ and the true values, θ_0 and γ_{i0}.

The resulting estimator removes the leading term of the incidental parameters bias and, unlike $\widehat{\theta}$, it may give correct asymptotic confidence intervals when T grows as fast as n.

To see this, consider an expansion for the first order conditions around the truth

$$\left(-\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta\partial\theta'}\overline{\psi}_i\left(\theta_0,\widehat{\gamma}_i\left(\theta_0\right)\right)\right)\sqrt{nT}\left(\widetilde{\theta}-\theta_0\right)\approx\sqrt{nT}\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\overline{\psi}_i\left(\theta_0,\widehat{\gamma}_i\left(\theta_0\right)\right)-\sqrt{\frac{n}{T}}\frac{\partial\widehat{B}\left(\theta_0\right)}{\partial\theta},$$

and suppose that n/T tends to a constant, $\sqrt{nT}\sum_{i=1}^{n}\left(\partial/\partial\theta\right)\overline{\psi}_i\left(\theta_0,\gamma_i\left(\theta_0\right)\right)/n\xrightarrow{d}\mathcal{N}\left(0,\Omega\right)$,

$$\sqrt{nT}\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\overline{\psi}_i\left(\theta_0,\widehat{\gamma}_i\left(\theta_0\right)\right)=\sqrt{nT}\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\overline{\psi}_i\left(\theta_0,\gamma_i\left(\theta_0\right)\right)+\sqrt{\frac{n}{T}}\frac{\partial B\left(\theta_0\right)}{\partial\theta}+o_p\left(1\right)$$

and that

$$\frac{\partial\widehat{B}\left(\theta_0\right)}{\partial\theta}=\frac{\partial B\left(\theta_0\right)}{\partial\theta}+o_p\left(1\right).$$

Thus, also

$$\sqrt{nT}\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\overline{\psi}_i\left(\theta_0,\widehat{\gamma}_i\left(\theta_0\right)\right)-\sqrt{\frac{n}{T}}\frac{\partial\widehat{B}\left(\theta_0\right)}{\partial\theta}\xrightarrow{d}\mathcal{N}\left(0,\Omega\right),$$

which suggests that as $n,T\rightarrow\infty$, $\sqrt{nT}\left(\widetilde{\theta}-\theta_0\right)$ is asymptotically normal with zero mean and the same asymptotic variance as the fixed effects estimator. We will give precise conditions for this result to hold.

2.1 Formulae for the Bias Correction

Let us introduce the notation:

$$\begin{aligned}\overline{V}_i\left(\theta,\gamma_i\right)&\equiv\frac{\partial\overline{\psi}_i\left(\theta,\gamma_i\right)}{\partial\gamma_i},\\ \overline{H}_i\left(\theta\right)&\equiv-\lim_{T\rightarrow\infty}E\left[\frac{\partial\overline{V}_i\left(\theta,\gamma_i\left(\theta\right)\right)}{\partial\gamma_i'}\right],\\ \overline{\Upsilon}_i\left(\theta\right)&\equiv\lim_{T\rightarrow\infty}TE\left[\overline{V}_i\left(\theta,\gamma_i\left(\theta\right)\right)\overline{V}_i\left(\theta,\gamma_i\left(\theta\right)\right)'\right].\end{aligned}$$

A first-order stochastic expansion for an arbitrary fixed θ gives

$$\widehat{\gamma}_i\left(\theta\right)-\gamma_i\left(\theta\right)=\overline{H}_i\left(\theta\right)^{-1}\overline{V}_i\left(\theta,\gamma_i\left(\theta\right)\right)+O_p\left(\frac{1}{T}\right).$$

Next, expanding $\overline{\psi}_i\left(\theta,\widehat{\gamma}_i\left(\theta\right)\right)$ around $\gamma_i\left(\theta\right)$ for fixed θ we get

$$\begin{aligned}\overline{\psi}_i\left(\theta,\widehat{\gamma}_i\left(\theta\right)\right)-\overline{\psi}_i\left(\theta,\gamma_i\left(\theta\right)\right)&=\overline{V}_i\left(\theta,\gamma_i\left(\theta\right)\right)'\left[\widehat{\gamma}_i\left(\theta\right)-\gamma_i\left(\theta\right)\right]\\&\quad-\frac{1}{2}\left[\widehat{\gamma}_i\left(\theta\right)-\gamma_i\left(\theta\right)\right]'\overline{H}_i\left(\theta\right)\left[\widehat{\gamma}_i\left(\theta\right)-\gamma_i\left(\theta\right)\right]+O_p\left(\frac{1}{T^{3/2}}\right),\end{aligned}$$

and combining the two expansions,

$$\overline{\psi}_i(\theta,\widehat{\gamma}_i(\theta))-\overline{\psi}_i(\theta,\gamma_i(\theta))=\frac{1}{2}\overline{V}_i(\theta,\gamma_i(\theta))'\,\overline{H}_i(\theta)^{-1}\,\overline{V}_i(\theta,\gamma_i(\theta))+O_p\left(\frac{1}{T^{3/2}}\right).$$

Finally, taking expectations and assuming that the expectations operator and the stochastic order symbols can be interchanged, we obtain

$$E\left[\overline{\psi}_i(\theta,\widehat{\gamma}_i(\theta))-\overline{\psi}_i(\theta,\gamma_i(\theta))\right]=\frac{1}{T}\beta_i(\theta)+O\left(\frac{1}{T^{3/2}}\right)$$

where

$$\beta_i(\theta)\equiv\frac{1}{2}\operatorname{trace}\left[\overline{H}_i(\theta)^{-1}\,\overline{\Upsilon}_i(\theta)\right]=\frac{1}{2}\operatorname{trace}\left\{\overline{H}_i(\theta)\operatorname{Var}\left(\sqrt{T}\left[\widehat{\gamma}_i(\theta)-\gamma_i(\theta)\right]\right)\right\}. \tag{5}$$

In the likelihood setup the information identity is satisfied at the truth so that $\overline{H}_i(\theta_0)^{-1}\,\overline{\Upsilon}_i(\theta_0)=I$. Moreover, $V_i(x_{it};\theta_0,\gamma_i(\theta_0))$ is a martingale sequence with the implication that

$$\overline{\Upsilon}_i(\theta_0)=\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}E\left[V_i(x_{it};\theta_0,\gamma_{i0})\,V_i(x_{it};\theta_0,\gamma_{i0})'\right].$$

When evaluated at other values of θ, the score vector $V_i(x_{it};\theta,\gamma_i(\theta))$ still has zero mean but in general it will be serially correlated:

$$\overline{\Upsilon}_i(\theta)=\sum_{l=-\infty}^{\infty}\overline{\Gamma}_l(\theta)$$

where $\overline{\Gamma}_l(\theta)$ denotes the steady-state covariance matrix between $V_i(x_{it};\theta,\gamma_i(\theta))$ and $V_i(x_{it-l};\theta,\gamma_i(\theta))$:

$$\overline{\Gamma}_l(\theta)\equiv\lim_{T\to\infty}\frac{1}{T}\sum_{t=l+1}^{T}E\left[V_i(x_{it};\theta,\gamma_i(\theta))\,V_i(x_{it-l};\theta,\gamma_i(\theta))'\right]\qquad l>0.$$

2.2 Estimation of the Bias

An estimator for the bias term in the modified concentrated likelihood (4) can be formed using $\widehat{B}(\theta)=\sum_{i=1}^{n}\widehat{\beta}_i(\theta)/n$, where $\widehat{\beta}_i(\theta)$ is a sample counterpart of the previous formulae.

Trace Based Approach One possibility is

$$\widehat{\beta}_i(\theta)=\frac{1}{2}\operatorname{trace}\left[H_i(\theta,\widehat{\gamma}_i(\theta))^{-1}\,\Upsilon_i(\theta,\widehat{\gamma}_i(\theta))\right] \tag{6}$$

where

$$H_i(\theta,\gamma)\equiv-\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2\psi_{it}(\theta,\gamma)}{\partial\gamma\partial\gamma'} \tag{7}$$

$$\Upsilon_i(\theta,\gamma)\equiv\textstyle\sum_{l=-m}^{m}w_{T,l}\Gamma_l(\theta,\gamma) \tag{8}$$

$$\Gamma_l(\theta,\gamma)\equiv\frac{1}{T}\sum_{t=\max(1,l+1)}^{\min(T,T+l)}\frac{\partial\psi_{it}(\theta,\gamma)}{\partial\gamma_i}\frac{\partial\psi_{it-l}(\theta,\gamma)}{\partial\gamma_i'}. \tag{9}$$

The quantity m is a bandwidth parameter and $w_{T,l}$ denotes a weight that guarantees positive definiteness of $\Upsilon_i(\theta,\gamma)$, e.g., a Bartlett kernel weight such that $w_{T,l} = 1 - \frac{l}{m+1}$.[4] Note that with $m = T-1$ and $w_{T,l} = 1$, $\Upsilon_i(\theta,\gamma) \equiv \overline{V}_i(\theta,\gamma_i(\theta))\,\overline{V}_i(\theta,\gamma_i(\theta))'$, so that in such case $\Upsilon_i(\theta,\widehat{\gamma}_i(\theta)) \equiv 0$.

The adjustment term $\widehat{\beta}_i(\theta)$ does not depend on the likelihood setting, and so it is valid for any fixed effects estimation problem based on the objective function $\sum_{i=1}^n \sum_{t=1}^T \psi(x_{it};\theta,\gamma_i)$. The trace-based approach can be regarded as an objective-function and estimating equation counterpart to the approach of bias-correction of the estimator in Hahn and Kuersteiner (2004).

Determinant Based Approach In the likelihood setting we can consider a local version of the estimated bias constructed as an expansion of $\widehat{\beta}_i(\theta)$ at θ_0 using that at the truth $\overline{H}_i(\theta_0)^{-1}\,\overline{\Upsilon}_i(\theta_0) = I$. To see this, note that

$$\widehat{\beta}_i(\theta) = \frac{1}{2}\sum_{j=1}^{p}\left[\widehat{\lambda}_j(\theta) - 1\right] + \frac{1}{2}p = \frac{1}{2}\sum_{j=1}^{p}\ln\widehat{\lambda}_j(\theta) + \frac{1}{2}p + O\left(\frac{1}{T}\right)$$

where $\widehat{\lambda}_j(\theta)$ denotes the j-th eigenvalue of $H_i(\theta,\widehat{\gamma}_i(\theta))^{-1}\,\Upsilon_i(\theta,\widehat{\gamma}_i(\theta))$ and $p = \dim(\theta)$. Since $\sum_{j=1}^p \ln\widehat{\lambda}_j(\theta) = \ln\det\left[H_i(\theta,\widehat{\gamma}_i(\theta))^{-1}\,\Upsilon_i(\theta,\widehat{\gamma}_i(\theta))\right]$, discarding constants, we can consider the alternative adjustment

$$\widetilde{\beta}_i(\theta) = -\frac{1}{2}\ln\det\left[H_i(\theta,\widehat{\gamma}_i(\theta))\right] + \frac{1}{2}\ln\det\left[\Upsilon_i(\theta,\widehat{\gamma}_i(\theta))\right]. \tag{10}$$

The resulting modified concentrated likelihood function is

$$L_D(\theta) = \sum_{i=1}^{n}\sum_{t=1}^{T}\psi(x_{it};\theta,\widehat{\gamma}_i(\theta)) + \frac{1}{2}\sum_{i=1}^{n}\ln\det\left[H_i(\theta,\widehat{\gamma}_i(\theta))\right] - \frac{1}{2}\sum_{i=1}^{n}\ln\det\left[\Upsilon_i(\theta,\widehat{\gamma}_i(\theta))\right] \tag{11}$$

where $\psi(x_{it};\theta,\gamma_i) = \ln p_c(y_{it} \mid y_{i,t-1},\ldots,y_{i,t-q};\theta,\gamma_i)$.

The criterion $L_D(\theta)$ is a multivariate and dynamic version of the adjusted concentrated likelihood considered by DiCiccio and Stern (1993), and DiCiccio, Martin, Stern, and Young (1996).

Using the arguments in Pace and Salvan (2005), it can be related to the adjusted profile likelihood considered by Cox and Reid (1987) as an approximation to the likelihood conditioned on the ML estimates of the fixed effects. In a model with independent observations, Ferguson, Reid, and Cox (1991) showed that such a modification led to bias reduction when the nuisance parameters were information orthogonal to the parameters of interest.

In our context, the Cox–Reid approach maximizes

$$L_{CR}(\theta) = \sum_{i=1}^{n}\sum_{t=1}^{T}\psi(x_{it};\theta,\widehat{\gamma}_i(\theta)) - \frac{1}{2}\sum_{i=1}^{n}\ln\det\left[H_i(\theta,\widehat{\gamma}_i(\theta))\right],$$

[4] For simplicity of exposition, we will assume that the $w_{T,l}$ are indeed Bartlett weights throughout the rest of the paper.

and the connection with $L_D(\theta)$ can be expressed as

$$L_D(\theta) = L_{CR}(\theta) - \frac{1}{2}\sum_{i=1}^{n} \ln\det \widehat{\mathrm{Var}}\left[\sqrt{nT}\left(\widehat{\gamma}_i(\theta) - \gamma_i(\theta)\right)\right]$$

where the variance term is given by the sandwich formula

$$\widehat{\mathrm{Var}}\left[\sqrt{nT}\left(\widehat{\gamma}_i(\theta) - \gamma_i(\theta)\right)\right] = \left[H_i\left(\theta, \widehat{\gamma}_i(\theta)\right)\right]^{-1} \Upsilon_i\left(\theta, \widehat{\gamma}_i(\theta)\right)\left[H_i\left(\theta, \widehat{\gamma}_i(\theta)\right)\right]^{-1}.$$

The conclusion is that $L_D(\theta)$ can be regarded as a generalized Cox–Reid function with an additional term to account for non-orthogonality. Under orthogonality the extra term is not needed because the variance of $\widehat{\gamma}_i(\theta)$ does not change much with θ.

Determinant Approach Using Expected Quantities In the likelihood setting, an expected outer product function can be calculated for given values of (θ, γ_i) and (θ_0, γ_{i0}) analytically or numerically, because the density of the data is available. Specifically, we may consider

$$\Upsilon_{Ti}(\theta, \gamma; \theta_0, \gamma_{i0}) \equiv \textstyle\sum_{l=-m}^{m} w_{T,l}\Gamma_{Tl}(\theta, \gamma; \theta_0, \gamma_{i0}) \tag{12}$$

where, for $l > 0$, we have

$$\Gamma_{Tl}(\theta, \gamma; \theta_0, \gamma_{i0}) = \frac{1}{T-l}\sum_{t=l+1}^{T} E_{\theta_0, \gamma_{i0}}\left[V_i(x_{it}; \theta, \gamma) V_i(x_{it-l}; \theta, \gamma)'\right]. \tag{13}$$

Alternatively, a centered covariance could be calculated:

$$\Gamma^{*}_{Tl}(\theta, \gamma; \theta_0, \gamma_{i0}) = \Gamma_{Tl}(\theta, \gamma; \theta_0, \gamma_{i0}) - \mu_{T0}(\theta, \gamma; \theta_0, \gamma_{i0})\,\mu_{Tl}(\theta, \gamma; \theta_0, \gamma_{i0})' \tag{14}$$

where $\mu_{Tl}(\theta, \gamma; \theta_0, \gamma_{i0}) = (T-l)^{-1}\sum_{t=l+1}^{T} E_{\theta_0, \gamma_{i0}}\left[V_i(x_{it-l}; \theta, \gamma)\right]$. Note that when evaluated at $\gamma = \gamma_i(\theta)$ for arbitrary θ we have $\mu_{Tl}(\theta, \gamma_i(\theta); \theta_0, \gamma_{i0}) = 0$, so that centered and non-centered quantities coincide.

This leads to an alternative modified concentrated likelihood of the form

$$L_{ED}\left(\theta; \widehat{\theta}\right) = \sum_{i=1}^{n}\sum_{t=1}^{T} \psi\left(x_{it}; \theta, \widehat{\gamma}_i(\theta)\right) + \frac{1}{2}\sum_{i=1}^{n} \ln\det H_i\left(\theta, \widehat{\gamma}_i(\theta)\right) - \frac{1}{2}\sum_{i=1}^{n} \ln\det \Upsilon_{Ti}\left(\theta, \widehat{\gamma}_i(\theta); \widehat{\theta}, \widehat{\gamma}_i\left(\widehat{\theta}\right)\right). \tag{15}$$

Iterated Adjusted Likelihood Estimation An undesirable feature of the estimator $\widehat{\theta}_1 = \arg\max_\theta L_{ED}\left(\theta; \widehat{\theta}\right)$ is its dependence on $\widehat{\theta}$, which may have a large bias. This problem can be avoided by considering an iterative procedure. That is, once we have $\widehat{\theta}_1$, we use it to evaluate the expectations required in calculating a new estimate. Pursuing the iteration

$$\widehat{\theta}_K = \arg\max_\theta L_{ED}\left(\theta; \widehat{\theta}_{K-1}\right) \tag{16}$$

until convergence, we obtain an estimator $\widehat{\theta}_\infty$ that solves

$$S_{ED}\left(\widehat{\theta}_\infty;\widehat{\theta}_\infty\right)=0 \tag{17}$$

where $S_{ED}(\theta;\theta_*)$ denotes the score of $L_{ED}(\theta;\theta_*)$ for fixed θ_*. Note that, in contrast with the iterated procedure, a continuously updated method will not work in this case (that is, maximizing a criterion of the form $L_{ED}(\theta;\theta)$).

Discussion Both likelihood and pseudo likelihood settings are important in applications. For example, there are nonlinear likelihood models whose parameters are no longer interpretable when the likelihood is only regarded as a pseudo likelihood.

In a likelihood situation it seems natural to use the determinant form of the correction, but also an expectation based estimate of the outer product term, specially if an analytical calculation is available, hence avoiding semiparametric kernel estimation. However, if expectations need to be evaluated by simulation, the conceptual advantage of the expectation–based adjustment is less clear, because the number of simulations to be chosen is an issue.

In contrast, in a pseudo likelihood or incomplete model setting it is natural to use the trace form of the correction and a kernel–based estimate of $\overline{\Upsilon}_i(\theta)$, which is the only possibility available.

3 Examples

We consider three examples. The first one is static and linear, but illustrates the differences between the two approaches in a familiar context. The second is a conditional volatility model, and the last one is a dynamic binary choice formulation.

Example 1 Consider a simple multivariate model for an unconditional covariance structure with heterogeneous means, where

$$\psi(x_{it};\theta,\gamma_i)=C-\frac{1}{2}\ln\det\Omega(\theta)-\frac{1}{2}(x_{it}-\gamma_i)'\,\Omega(\theta)^{-1}(x_{it}-\gamma_i).$$

If $\Omega(\theta)$ is unrestricted then $\theta=vech\,[\Omega(\theta)]$. We have $\widehat{\gamma}_i(\theta)=\overline{x}_i$ and

$$\frac{\partial\psi(x_{it};\theta,\gamma_i)}{\partial\gamma_i}=\Omega(\theta)^{-1}(x_{it}-\gamma_i),\qquad\frac{\partial^2\psi(x_{it};\theta,\gamma_i)}{\partial\gamma_i\partial\gamma_i'}=-\Omega(\theta)^{-1}$$

$$\begin{aligned}
H_i(\theta,\gamma) &\equiv -\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2\psi_{it}(\theta,\gamma)}{\partial\gamma\partial\gamma'}=\Omega(\theta)^{-1}\\
\Upsilon_i(\theta,\widehat{\gamma}_i(\theta)) &\equiv \textstyle\sum_{l=-m}^{m}w_{T,l}\Gamma_l(\theta,\widehat{\gamma}_i(\theta))\\
\Gamma_l(\theta,\widehat{\gamma}_i(\theta)) &\equiv \Omega(\theta)^{-1}\left[\frac{1}{T}\sum_{t=\max(1,l+1)}^{\min(T,T+l)}(x_{it}-\overline{x}_i)(x_{it-l}-\overline{x}_i)'\right]\Omega(\theta)^{-1}.
\end{aligned}$$

The determinant approach with $m=0$ gives

$$\begin{aligned} L_D(\theta) &= C - \frac{nT}{2}\ln\det\Omega(\theta) - \frac{1}{2}\sum_{i=1}^{n}\sum_{t=1}^{T}(x_{it}-\overline{x}_i)'\,\Omega(\theta)^{-1}(x_{it}-\overline{x}_i) \\ &\quad + \frac{n}{2}\ln\det\left[\Omega(\theta)^{-1}\right] - \frac{1}{2}\ln\det\left(\Omega(\theta)^{-1}\frac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T}(x_{it}-\overline{x}_i)(x_{it}-\overline{x}_i)'\,\Omega(\theta)^{-1}\right). \end{aligned}$$

Finally, collecting terms and discarding constants we get

$$L_D(\theta) = C - \frac{n(T-1)}{2}\ln\det\Omega(\theta) - \frac{nT}{2}\,\text{trace}\left[\Omega(\theta)^{-1}\widehat{\Omega}\right]$$

where $\widehat{\Omega}$ is the unrestricted fixed effects estimate:

$$\widehat{\Omega} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(x_{it}-\overline{x}_i)(x_{it}-\overline{x}_i)'.$$

Thus, the information adjustment performs the required degrees of freedom correction (i.e. the corrected unrestricted estimate is $\widetilde{\Omega} = \frac{T}{T-1}\widehat{\Omega}$).

The trace-based approach should provide bias reduction in the presence of neglected serial correlation. It gives

$$\widehat{\beta}_i(\theta) = \frac{1}{2}\,\text{trace}\left[\widetilde{\Gamma}_i\Omega(\theta)^{-1}\right]$$

where

$$\widetilde{\Gamma}_i = \frac{1}{T}\textstyle\sum_{l=-m}^{m} w_{T,l}\displaystyle\sum_{t=\max(1,l+1)}^{\min(T,T+l)}(x_{it}-\overline{x}_i)(x_{it-l}-\overline{x}_i)'.$$

Letting $\widetilde{\Gamma} = n^{-1}\sum_{i=1}^{n}\widetilde{\Gamma}_i$, we obtain

$$L_{TR}(\theta) = C - \frac{nT}{2}\ln\det\Omega(\theta) - \frac{nT}{2}\,\text{trace}\left[\Omega(\theta)^{-1}\widehat{\Omega}\right] - \frac{n}{2}\,\text{trace}\left[\Omega(\theta)^{-1}\widetilde{\Gamma}\right].$$

Note that with $m=0$, $\widetilde{\Gamma} = \widehat{\Omega}$, so that in this case the corrected unrestricted estimate is $\widetilde{\Omega}_{TR} = \frac{T+1}{T}\widehat{\Omega}$, which removes the bias of order T^{-1}, but is not fully unbiased. In general, the trace-based unrestricted estimate is given by

$$\widetilde{\Omega}_{TR} = \widehat{\Omega} + \frac{1}{T}\widetilde{\Gamma}.$$

Example 2 The next example is a heteroskedastic autoregressive model with two fixed effects, one in the conditional mean and another in the conditional variance. Letting $\theta = (\theta_1, \theta_2)$ and $\gamma_i = (\gamma_{1i}, \gamma_{2i})$, we have

$$\psi(x_{it};\theta,\gamma_i) = -\frac{1}{2}\ln h(y_{it-1},\gamma_{2i}) - \frac{1}{2}\frac{(y_{it}-\theta_1 y_{it-1}-\gamma_{1i})^2}{h(y_{it-1},\gamma_{2i})}$$

where

$$h\left(y_{it-1},\gamma_{2i}\right)=\left(\gamma_{2i}+\theta_2 y_{it-1}\right)^2.$$

A model of this type, but with an exponential ARCH formulation of the conditional variance, is developed in Hospido (2006), where some of the estimators considered in this paper, as well as simulation-based alternatives, are implemented and applied to study individual wage dynamics.

Example 3 A third example is an autoregressive binary formulation of the form

$$\psi\left(x_{it};\theta,\gamma_i\right)=y_{it}\ln\Lambda\left(\gamma_{1i}+\gamma_{2i}y_{it-1}+\theta y_{it-2}\right)+\left(1-y_{it}\right)\ln\left[1-\Lambda\left(\gamma_{1i}+\gamma_{2i}y_{it-1}+\theta y_{it-2}\right)\right]$$

where $\Lambda(r)$ is the logit or probit *cdf.*

This model was suggested in Chamberlain (1985) as a framework for testing duration dependence from binary panel data, by testing the restriction $\theta=0$. Chamberlain showed that, in the absence of exogenous variables, a simple fixed-T consistent estimator for θ is available for the logistic specification of this model. A random effects formulation of a model of this type has been recently applied by Card and Hyslop (2005) to study the effects of earnings subsidies on welfare participation.

4 Asymptotic Theory

We first consider general conditions for a bias corrected estimating equation to deliver an asymptotic normality theorem for the estimation error centered at the truth.

Notation 1 *We use the following additional notation throughout:*

$$U_i\left(x_{it};\theta,\gamma_i\right)\equiv\frac{\partial\psi\left(x_{it};\theta,\gamma_i\right)}{\partial\theta}-\rho_{i0}\cdot\frac{\partial\psi\left(x_{it};\theta,\gamma_i\right)}{\partial\gamma_i},\quad V_i\left(x_{it};\theta,\gamma_i\right)\equiv\frac{\partial\psi\left(x_{it};\theta,\gamma_i\right)}{\partial\gamma_i},$$
$$\rho_{i0}\equiv E\left[\frac{\partial^2\psi\left(x_{it};\theta_0,\gamma_{i0}\right)}{\partial\theta\partial\gamma_i'}\right]\left(E\left[\frac{\partial^2\psi\left(x_{it};\theta_0,\gamma_{i0}\right)}{\partial\gamma_i\partial\gamma_i'}\right]\right)^{-1},\quad\mathcal{I}_i\equiv-E\left[\frac{\partial U_i\left(x_{it};\theta_0,\gamma_{i0}\right)}{\partial\theta'}\right],$$
$$\widetilde{V}_{it}\equiv-\left(E\left[\frac{\partial V_i}{\partial\gamma_i'}\right]\right)^{-1}V_{it}.$$

For simplicity of notation, we will occasionally write $U_{it}\equiv U_i\left(x_{it};\theta_0,\gamma_{i0}\right)$ and $V_{it}\equiv V_i\left(x_{it};\theta_0,\gamma_{i0}\right)$. We will denote by $U_{it}^{\gamma_i}\equiv\partial U_{it}/\partial\gamma_i'$ and $U_{it}^{\gamma_i\gamma_i}\equiv\partial^2 U_{it}/\left(\partial\gamma_i'\otimes\partial\gamma_i'\right)$ the first and second derivatives of U_{it} with respect to γ_i. Likewise, we will denote by $V_{it}^{\gamma_i}$ the derivative $\partial V_{it}/\partial\gamma_i'$ of V_{it} with respect to γ_i.

Using this notation, we can characterize $\widehat{\theta}$ as the solution to the first order condition

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

The normalized score $\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}U\left(x_{it};\theta_0,\widehat{\gamma}_i\left(\theta_0\right)\right)$ has an asymptotic bias, which renders the fixed effects estimator $\widehat{\theta}$ biased. The asymptotic bias of the normalized score can be shown[5] to be equal to $\frac{1}{T}$ times $\Psi_0\left(\theta_0,\left\{\gamma_{10},\gamma_{20},\ldots\right\}\right)$, where

$$\begin{aligned}\Psi_0\left(\theta_0,\left\{\gamma_{10},\gamma_{20},\ldots\right\}\right) &= \operatorname{plim}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}U_{it}^{\gamma_i}\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widetilde{V}_{it}\right)\\ &\quad+\operatorname{plim}\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}E\left[U_i^{\gamma_i\gamma_i}\right]\left[\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widetilde{V}_{it}\right)\otimes\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widetilde{V}_{it}\right)\right].\end{aligned}$$

Note that $\gamma_{i0}\equiv\operatorname{argmax}_c E\left[\psi\left(x_{it};\theta_0,c\right)\right]$. Therefore, using $\gamma_i\left(\theta\right)\equiv\operatorname{argmax}_c E\left[\psi\left(x_{it};\theta,c\right)\right]$, we can write

$$\Psi_0\left(\theta_0,\left\{\gamma_{10},\gamma_{20},\ldots\right\}\right)=\Psi_0\left(\theta_0,\left\{\gamma_1\left(\theta_0\right),\gamma_2\left(\theta_0\right),\ldots\right\}\right),$$

which can be regarded as a function in θ_0. Such function will be written as $\Psi_0\left(\theta\right)$ without loss of generality. We will approximate it by $\Psi_n\left(\theta_0\right)\equiv\Psi_0\left(\theta_0,\left\{\widehat{\gamma}_1\left(\theta_0\right),\widehat{\gamma}_2\left(\theta_0\right),\ldots\right\}\right)$. Letting $S_n\left(\theta_0\right)$ denote some sample counter-part of $\Psi_n\left(\theta_0\right)$, we may consider solving

$$0=\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}U\left(x_{it};\widetilde{\theta},\widehat{\gamma}_i\left(\widetilde{\theta}\right)\right)-\frac{1}{T}S_n\left(\widetilde{\theta}\right)\tag{18}$$

instead. We will assume that there exists some B_n such that $S_n\left(\theta\right)=\partial B_n\left(\theta\right)/\partial\theta$, in which case our estimator $\widetilde{\theta}$ can be understood as a solution to

$$\operatorname*{argmax}_{\theta,\gamma_1,\ldots,\gamma_n}\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\psi\left(x_{it};\theta,\gamma_i\right)-\frac{1}{T}B_n\left(\theta\right)\tag{19}$$

We impose the following conditions:

Condition 1 $\Pr\left[\sup_\theta\left|\frac{1}{T}B_n\left(\theta\right)\right|\geq\eta\right]=o\left(T^{-1}\right)$ *for every* $\eta>0$.

Condition 2 $\sup_\theta\frac{1}{T}\left|\partial S_n\left(\theta\right)/\partial\theta'\right|=o_p\left(1\right)$.

Condition 3

$$\begin{aligned}S_n\left(\theta_0\right) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{l=-\infty}^{\infty}E\left[U_{it}^{\gamma_i}\widetilde{V}_{it-l}\right]\\ &\quad+\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}E\left[U_i^{\gamma_i\gamma_i}\right]\operatorname{vec}\left(\sum_{l=-\infty}^{\infty}E\left[\widetilde{V}_{it}\widetilde{V}_{it-l}'\right]\right)+o_p\left(1\right).\end{aligned}$$

Under these conditions and the regularity conditions in Appendix A, we can obtain the asymptotic distribution of $\widetilde{\theta}$ as n and T grow at the same rate.

[5] This is a standard result, but we do provide a rigorous derivation in Supplementary Appendix, which is available upon request.

Theorem 2 *Assume that Conditions 1, 2, and 3 hold. Further assume that the regularity conditions in Appendix A hold. Finally, assume that $n/T \to \kappa$, where $0 < \kappa < \infty$. Then*

$$\sqrt{nT}\left(\widetilde{\theta} - \theta_0\right) \Rightarrow N\left(0, \mathcal{I}^{-1}\Omega\left(\mathcal{I}'\right)^{-1}\right)$$

Proof. See Appendix B. ■

4.1 Determinant Based Approach

We now assume that $x_{it} = (y_{it}, y_{i,t-1}, \ldots, y_{i,t-q})$ and $\psi(x_{it}; \theta, \gamma_i) = \ln p_c(y_{it}|\, y_{i,t-1}, \ldots, y_{i,t-q}; \theta, \gamma_i)$, where p_c here denotes the conditional density of y_{it}. We propose to estimate θ_0 by

$$\widetilde{\theta} = \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{n}\sum_{t=1}^{T} \psi\left(x_{it}; \theta, \widehat{\gamma}_i(\theta)\right) + \frac{1}{2}\sum_{i=1}^{n} \ln\det H_i\left(\theta, \widehat{\gamma}_i(\theta)\right) - \frac{1}{2}\sum_{i=1}^{n} \ln\det \Upsilon_i\left(\theta, \widehat{\gamma}_i(\theta)\right) \qquad (20)$$

where $H_i(\theta, \gamma)$ and $\Upsilon_i(\theta, \gamma)$ are as defined in (7)–(9).

Comparing (20) with (19), we obtain

$$B_n(\theta) = -\frac{1}{2n}\sum_{i=1}^{n} \ln\det H_i\left(\theta, \widehat{\gamma}_i(\theta)\right) + \frac{1}{2n}\sum_{i=1}^{n} \ln\det \Upsilon_i\left(\theta, \widehat{\gamma}_i(\theta)\right) \qquad (21)$$

By differentiating B_n, we obtain $S_n(\theta_0)$. It can be shown that[6]

Theorem 3 *Assume that the regularity conditions in Appendix A hold. Then, the $B_n(\theta)$ as defined in (21) satisfies Condition 1.*

Theorem 4 *Assume that the regularity conditions in Appendix A hold. Further assume that $m = o\left(T^{1/2}\right)$. Then, the $B_n(\theta)$ as defined in (21) satisfies Condition 2.*

Theorem 5 *Assume that the model is given by the likelihood. Also assume that the regularity conditions in Appendix A hold. Further assume that $m = o\left(T^{2/5}\right)$. Then, the $B_n(\theta)$ as defined in (21) satisfies Condition 3.*

Proof. See Appendix C. ■

Remark 1 *Proof of Theorem 5 uses the information equality, as discussed in Appendix on page 18. This explains why the likelihood setup is required here.*

Conclusion 1 *Theorems 3, 4, and 5 imply that Theorem 2 applies to our new estimator.*

[6] The proofs of Theorems 3 and 4 are in Supplementary Appendix, which is available upon request.

4.2 Expectation-based Determinant Approach

We now consider a variant of the above estimator where instead of $\Upsilon_i(\theta,\gamma)$ we use

$$\overline{\Upsilon}_i(\theta,\gamma) \equiv \textstyle\sum_{l=-m}^{m} w_{T,l} E_{\widehat{\theta},\widehat{\gamma}_i}\left[\frac{\partial\psi_{it}(\theta,\gamma)}{\partial\gamma}\frac{\partial\psi_{it-l}(\theta,\gamma)}{\partial\gamma'}\right]. \tag{22}$$

Here, $E_{\widehat{\theta},\widehat{\gamma}_i}[\cdot]$ denotes an expectation taken with respect to the density evaluated at $\left(\widehat{\theta},\widehat{\gamma}_i\right)$. Note that $B_n(\theta)$ is defined similarly as in (21). As before, B_n, we obtain $S_n(\theta_0)$ by differentiating B_n. It can be shown that[7]

Theorem 6 *Assume that the regularity conditions in Appendix A hold. Then the $B_n(\theta)$ based on (22) satisfies Condition 1 as long as $m\to\infty$ such that $m=o\left(T^{2/5}\right)$. The same result hold even when the preliminary estimates $\left(\widehat{\theta},\widehat{\gamma}_i\right)$ in (22) are replaced by some (θ^*,γ_i^*) such that $\|\theta^*-\theta\|=O_p\left(T^{-2/5}\right)$ and $\sup_i\|\gamma_i^*-\gamma_{i0}\|=O_p\left(T^{-2/5}\right)$.*

Theorem 7 *Assume that the regularity conditions in Appendix A hold. Further assume that $m=o\left(T^{2/5}\right)$. Then, the $B_n(\theta)$ based on (22) satisfies Condition 2. The same result hold even when the preliminary estimates $\left(\widehat{\theta},\widehat{\gamma}_i\right)$ in (22) are replaced by some (θ^*,γ_i^*) such that $\|\theta^*-\theta\|=O_p\left(T^{-2/5}\right)$ and $\sup_i\|\gamma_i^*-\gamma_{i0}\|=O_p\left(T^{-2/5}\right)$.*

Theorem 8 *Assume that the model is given by the likelihood. Also assume that the regularity conditions in Appendix A hold. Further assume that $m=o\left(T^{2/5}\right)$. Then, the $B_n(\theta)$ as defined in (21) satisfies Condition 3. The same result hold even when the preliminary estimates $\left(\widehat{\theta},\widehat{\gamma}_i\right)$ in (22) are replaced by some (θ^*,γ_i^*) such that $\|\theta^*-\theta\|=O_p\left(T^{-2/5}\right)$ and $\sup_i\|\gamma_i^*-\gamma_{i0}\|=O_p\left(T^{-2/5}\right)$.*

Proof. See Appendix G. ■

Remark 2 *Proof of Theorem 5 uses the information equality, as discussed in Supplementary Appendix on page 22. This explains why we required the likelihood setup.*

Conclusion 2 *Theorems 6, 7, and 8 imply that Theorem 2 applies to our new estimator, even when the preliminary estimates $\left(\widehat{\theta},\widehat{\gamma}_i\right)$ in (22) are replaced by some (θ^*,γ_i^*) satisfying some regularity condition.*

4.3 Trace Based Approach

We now consider a slightly different approach where we set

$$B_n(\theta) = \frac{1}{2n}\sum_{i=1}^{n}\operatorname{trace}\left(H_i(\theta,\widehat{\gamma}_i(\theta))^{-1}\Upsilon_i(\theta,\widehat{\gamma}_i(\theta))\right) \tag{23}$$

It can be shown that[8]

[7] The proof of Theorem 6 is in Supplementary Appendix, which is available upon request. The proof of Theorem 7 is similar to that of Theorem 4, and is omitted.

[8] The proof of Theorem 9 is similar to those of Theorems 3 and 4, and is omitted.

Theorem 9 *Assume that the regularity conditions in Appendix A hold. Further assume that* $m = o\left(T^{1/2}\right)$*. Then, the* $B_n(\theta)$ *as defined in (23) satisfies Conditions 1 and 2.*

Theorem 10 *Assume that the regularity conditions in Appendix A hold. Further assume that* $m = o\left(T^{1/2}\right)$*. Then, the* $B_n(\theta)$ *as defined in (23) satisfies Condition Condition 3.*

Proof. See Appendix D. ■

Remark 3 *Proof of Theorem 10 does not use the information equality. We therefore do not require the likelihood setup here.*

5 Concluding Remarks

We discussed a modified objective function strategy to obtain estimators without bias to order $1/T$ in nonlinear dynamic panel models with multiple effects. Estimation proceeds from a bias corrected objective function relative to some target infeasible criterion. We considered a determinant based approach for likelihood settings, and a trace based approach, which is not restricted to the likelihood setup. Both approaches depend exclusively on the Hessian and the outer product of the scores of the fixed effects. They produce simple and transparent corrections even in models with multiple effects.

We analyzed the asymptotic properties of the new estimators when n and T grow at the same rate, and showed that they are asymptotically normal and centered at the truth.

These approaches are likely to be useful in applications where the value of T is not negligible relative to n, as is the case with many household, firm, and country-level panels. However, if T/n is too small, further refinements may be required, because the sampling standard deviation of the $1/T$ bias-corrected estimators will be small by comparison with the bias.

Existing Monte Carlo results and empirical estimates for binary choice and conditional volatility models are very encouraging, but more needs to be known about the properties of the new methods for other models and datasets.

Appendix

A Regularity Conditions

Assumption 1 *For each* $\eta > 0$, $\inf_i \left[G_{(i)}(\theta_0, \gamma_{i0}) - \sup_{\{(\theta,\gamma):|(\theta,\gamma)-(\theta_0,\gamma_{i0})|>\eta\}} G_{(i)}(\theta, \gamma) \right] > 0$.

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

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

Assumption 4 *Let* $\psi(x_{it}, \phi)$ *be a function indexed by the parameter* $\phi = (\theta, \gamma) \in \operatorname{int}\Phi$, *where* Φ *is a compact, convex subset of* $\mathbb{R}^p$, $p \equiv \dim(\phi)$, *and* $R = \dim(\theta)$. *Let* $\nu = (\nu_1, ..., \nu_k)$ *be a vector of non-negative integers* v_i, $|v| = \sum_{j=1}^k v_j$ *and* $D^v \psi(x_{it}, \phi) = \partial^{|\nu|} \psi(x_{it}, \phi) / \left(\partial \phi_1^{v_1} ... \partial \phi_k^{\nu_k}\right)$. *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* $|v| \leq 5$. *The function* $M(x_{it})$ *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* $\delta > 0$.

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

Assumption 6 *(i)* $\inf_i \inf_{\theta, \gamma_i} \left| E\left[\partial^2 \psi(x_{it}; \theta, \gamma_i) / \partial \gamma_i \partial \gamma_i'\right] \right| > 0$;
(ii) $\inf_i \inf_{\theta, \gamma_i} \sum_{l=-\infty}^{\infty} E\left[(\partial \psi(x_{it}; \theta, \gamma_i) / \partial \gamma_i)(\partial \psi(x_{it-l}; \theta, \gamma_i) / \partial \gamma_i')\right] > 0$.

Remark 4 *Assumption 6 is stronger than the one assumed in Hahn and Kuersteiner (2004).*

Assumption 7 *Let* $\mu_{i1} \leq ... \leq \mu_{ik} \leq ... \leq \mu_{iR}$ *be the eigenvalues of* $\mathcal{I}_i$ *in ascending order. We have (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.*

Assumption 8 $\sup_{(\theta,\gamma)\in\Phi} \sup_l E_{\theta,\gamma}\left[M(x_{it}) M(x_{it-l})\right] < \infty$.

B Proof of Theorem 2

We focus on asymptotic normality here, taking consistency result as given. (The consistency result is available in a Supplementary Appendix, which is available upon request.) Because $0 = \sum_{t=1}^T V\left(x_{it}; \widetilde{\theta}, \widehat{\gamma}_i\left(\widetilde{\theta}\right)\right)$ by definition, $\widetilde{\theta}$ can be given the alternative characterization

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

By the Taylor series expansion, we obtain

$$\begin{aligned} 0 &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}U\left(x_{it};\theta_0,\widehat{\gamma}_i\left(\theta_0\right)\right)-\frac{1}{T}S_n\left(\theta_0\right) \\ &\quad+\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}U^{\theta}\left(x_{it};\overline{\theta},\widehat{\gamma}_i\left(\overline{\theta}\right)\right)\right)\left(\widetilde{\theta}-\theta_0\right)-\frac{1}{T}\frac{\partial S_n\left(\overline{\theta}\right)}{\partial\theta'}\left(\widetilde{\theta}-\theta_0\right) \end{aligned}$$

for some $\overline{\theta}$ on the line segment adjoining θ_0 and $\widetilde{\theta}$. Because $\mathcal{I}_i \equiv -E\left[\frac{\partial U_i(x_{it};\theta_0,\gamma_{i0})}{\partial\theta'}\right]$, we may define $\overline{\mathcal{I}}_i \equiv -\frac{1}{T}\sum_{t=1}^{T}U^{\theta}\left(x_{it};\overline{\theta},\widehat{\gamma}_i\left(\overline{\theta}\right)\right)$, which yields

$$\sqrt{nT}\left(\widetilde{\theta}-\theta_0\right)=\left(\frac{1}{n}\sum_{i=1}^{n}\overline{\mathcal{I}}_i+\frac{1}{T}\frac{\partial S_n\left(\overline{\theta}\right)}{\partial\theta'}\right)^{-1}\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U\left(x_{it};\theta_0,\widehat{\gamma}_i\left(\theta_0\right)\right)-\sqrt{\frac{n}{T}}S_n\left(\theta_0\right)\right) \quad (24)$$

It can be shown[9] that $\frac{1}{n}\sum_{i=1}^{n}\overline{\mathcal{I}}_i=\mathcal{I}+o_p(1)$. By Condition 2, we also have $\frac{1}{T}\,\partial S_n\left(\overline{\theta}\right)/\partial\theta'=o_p(1)$. We therefore have

$$\frac{1}{n}\sum_{i=1}^{n}\overline{\mathcal{I}}_i+\frac{1}{T}\frac{\partial S_n\left(\overline{\theta}\right)}{\partial\theta'}=\mathcal{I}+o_p(1) \quad (25)$$

By applying a second order Taylor series approximation to $\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}U\left(x_{it};\theta_0,\widehat{\gamma}_i\left(\theta_0\right)\right)$ around γ_{i0}, and noting that $\widehat{\gamma}_i\left(\theta_0\right)-\gamma_{i0}=-\left(E\left[\frac{\partial V_i}{\partial\gamma_i'}\right]\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}V_{it}\right)+o_p\left(\frac{1}{\sqrt{T}}\right)=\frac{1}{T}\sum_{t=1}^{T}\widetilde{V}_{it}+o_p\left(\frac{1}{\sqrt{T}}\right)$, we can anticipate that[10]

$$\begin{aligned} &\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U\left(x_{it};\theta_0,\widehat{\gamma}_i\left(\theta_0\right)\right) \\ =&\ \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}+\sqrt{\frac{n}{T}}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}U_{it}^{\gamma_i}\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widetilde{V}_{it}\right) \\ &+\sqrt{\frac{n}{T}}\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}E\left[U_i^{\gamma_i\gamma_i}\right]\left[\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widetilde{V}_{it}\right)\otimes\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widetilde{V}_{it}\right)\right]+o_p(1) \end{aligned} \quad (26)$$

It can be shown that by using the same argument as in Hahn and Kuersteiner (2004) that

$$\begin{aligned} &\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}E\left[U_i^{\gamma_i\gamma_i}\right]\left[\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widetilde{V}_{it}\right)\otimes\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widetilde{V}_{it}\right)\right]+\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}U_{it}^{\gamma_i}\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widetilde{V}_{it}\right) \\ &\qquad=\frac{1}{n}\sum_{i=1}^{n}\sum_{l=-\infty}^{\infty}E\left[U_i^{\gamma_i}\widetilde{V}_{it-l}\right]+\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}E\left[U_i^{\gamma_i\gamma_i}\right]\operatorname{vec}\left(\sum_{l=-\infty}^{\infty}E\left[\widetilde{V}_{it}\widetilde{V}_{it-l}'\right]\right)+o_p(1) \end{aligned}$$

which, when combining (24), (25), (26) and Condition 3, yields

$$\sqrt{nT}\left(\widetilde{\theta}-\theta_0\right)=\mathcal{I}^{-1}\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U\left(x_{it};\theta_0,\gamma_{i0}\right)\right)+o_p(1)$$

from which the conclusion follows.

[9] See Lemma 6 in Supplementary Appendix A.

[10] In Supplementary Appendix C, we provide a rigorous proof of the expansion (26).

C Proof of Theorem 5

By differentiating B_n, we obtain that $S_n(\theta) = [2] + \cdots + [5]$, where

$$[2] \equiv -\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^3\psi_{it}}{\partial\theta\left(\partial\gamma'\otimes\partial\gamma'\right)}\right)\operatorname{vec}\left(\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2\psi_{it}}{\partial\gamma\partial\gamma'}\right)^{-1}\right)$$

$$[3] \equiv -\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\frac{\partial\widehat{\gamma}_i'(\theta)}{\partial\theta}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^3\psi_{it}}{\partial\gamma\left(\partial\gamma'\otimes\partial\gamma'\right)}\right)\operatorname{vec}\left(\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2\psi_{it}}{\partial\gamma\partial\gamma'}\right)^{-1}\right)$$

$$\begin{aligned}[4] \equiv\ & \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{T}\textstyle\sum_{l=-m}^{m} w_{T,l}\left(\displaystyle\sum_{t=\max(1,l+1)}^{\min(T,T+l)}\frac{\partial}{\partial\theta}\left(\left(\frac{\partial\psi_{it}}{\partial\gamma'}\right)\otimes\left(\frac{\partial\psi_{it-l}}{\partial\gamma'}\right)\right)\right)\right] \\ & \cdot\operatorname{vec}\left(\left(\frac{1}{T}\textstyle\sum_{l=-m}^{m} w_{T,l}\left(\displaystyle\sum_{t=\max(1,l+1)}^{\min(T,T+l)}\frac{\partial\psi_{it}}{\partial\gamma}\frac{\partial\psi_{it-l}}{\partial\gamma'}\right)\right)^{-1}\right)\end{aligned}$$

and

$$\begin{aligned}[5] \equiv\ & \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\frac{\partial\widehat{\gamma}_i'(\theta)}{\partial\theta}\left[\frac{1}{T}\textstyle\sum_{l=-m}^{m} w_{T,l}\left(\displaystyle\sum_{t=\max(1,l+1)}^{\min(T,T+l)}\frac{\partial}{\partial\gamma}\left(\left(\frac{\partial\psi_{it}}{\partial\gamma'}\right)\otimes\left(\frac{\partial\psi_{it-l}}{\partial\gamma'}\right)\right)\right)\right] \\ & \cdot\operatorname{vec}\left(\left(\frac{1}{T}\textstyle\sum_{l=-m}^{m} w_{T,l}\left(\displaystyle\sum_{t=\max(1,l+1)}^{\min(T,T+l)}\frac{\partial\psi_{it}}{\partial\gamma}\frac{\partial\psi_{it-l}}{\partial\gamma'}\right)\right)^{-1}\right)\end{aligned}$$

We will often use the first order condition for $\widehat{\gamma}_i(\theta)$, which implies that

$$\frac{\partial\widehat{\gamma}_i'(\theta)}{\partial\theta} = -\left(\sum_{t=1}^{T}\frac{\partial^2\psi_{it}\left(\theta,\widehat{\gamma}_i(\theta)\right)}{\partial\theta\partial\gamma'}\right)\left(\sum_{t=1}^{T}\frac{\partial^2\psi_{it}\left(\theta,\widehat{\gamma}_i(\theta)\right)}{\partial\gamma\partial\gamma'}\right)^{-1}. \tag{27}$$

In the discussion below, all the terms $[2], \cdots, [5]$ will be evaluated at θ_0. We first take care of the expansion of $[2]+[3]$. Note first that, by definition of $U_{it}(\theta,\gamma_i)$, we have $\frac{\partial^3\psi_{it}(\theta,\gamma)}{\partial\theta(\partial\gamma'\otimes\partial\gamma')} = U_{it}^{\gamma\gamma} + \rho_i V_{it}^{\gamma\gamma}$, where $V_{it}^{\gamma\gamma}(\theta,\gamma_i) = \frac{\partial^2 V_{it}(\theta,\gamma_i)}{\partial\gamma'\otimes\partial\gamma'}$. It turns out that all the averages over t on the RHS of [2] is uniformly consistent over i.[11] We therefore obtain

$$[2] = -\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\left(E\left[U_{it}^{\gamma\gamma}\right]+\rho_i E\left[V_{it}^{\gamma\gamma}\right]\right)\operatorname{vec}\left(\left(E\left[V_{it}^{\gamma}\right]\right)^{-1}\right) + o_p(1) \tag{28}$$

The uniform consistency over i combined with (27) also implies that

$$\max_i\left|\frac{\partial\widehat{\gamma}_i'(\theta)}{\partial\theta}+\rho_i\right| = o_p(1) \tag{29}$$

Using the uniform consistency and equation (29), we obtain

$$[3] = \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\rho_i E\left[V_{it}^{\gamma\gamma}\right]\operatorname{vec}\left(\left(E\left[V_{it}^{\gamma}\right]\right)^{-1}\right) + o_p(1) \tag{30}$$

[11] See Lemma 6 in Supplementary Appendix.

Combining (28) and (30), we obtain

$$[2]+[3]=-\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}E\left[U_{it}^{\gamma\gamma}\right]\mathrm{vec}\left(\left(E\left[V_{it}^{\gamma}\right]\right)^{-1}\right)+o_p(1) \tag{31}$$

We now take care of the expansion of $[4]+[5]$. Note that

$$\frac{\partial}{\partial\theta}\left(\left(\frac{\partial\psi_{it}(\theta,\gamma)}{\partial\gamma'}\right)\otimes\left(\frac{\partial\psi_{it-l}(\theta,\gamma)}{\partial\gamma'}\right)\right)=\left(\frac{\partial^2\psi_{it}(\theta,\gamma)}{\partial\theta\partial\gamma'}\right)\otimes\left(\frac{\partial\psi_{it-l}(\theta,\gamma)}{\partial\gamma'}\right)$$
$$+\left(\frac{\partial\psi_{it}(\theta,\gamma)}{\partial\gamma'}\right)\otimes\left(\frac{\partial^2\psi_{it-l}(\theta,\gamma)}{\partial\theta\partial\gamma'}\right)=\left(U_{it}^{\gamma}+\rho_i V_{it}^{\gamma}\right)\otimes V_{it-l}'+V_{it}'\otimes\left(U_{it-l}^{\gamma}+\rho_i V_{it-l}^{\gamma}\right)$$

and

$$\frac{\partial}{\partial\gamma}\left(\left(\frac{\partial\psi_{it}(\theta,\gamma)}{\partial\gamma'}\right)\otimes\left(\frac{\partial\psi_{it}(\theta,\gamma)}{\partial\gamma'}\right)\right)=V_{it}^{\gamma}\otimes V_{it-l}'+V_{it}'\otimes V_{it-l}^{\gamma}$$

we can write

$$\begin{aligned}[4]+[5] &= \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{T}\sum_{l=-m}^{m}w_{T,l}\sum_{t=\max(1,l+1)}^{\min(T,T+l)}\left(\begin{array}{c}U_{it}^{\gamma}\left(\theta_0,\widehat{\gamma}_i(\theta_0)\right)\otimes V_{it-l}\left(\theta_0,\widehat{\gamma}_i(\theta_0)\right)'\\ +V_{it}\left(\theta_0,\widehat{\gamma}_i(\theta_0)\right)'\otimes U_{it-l}^{\gamma}\left(\theta_0,\widehat{\gamma}_i(\theta_0)\right)\end{array}\right)\right]\\
&\quad\cdot\mathrm{vec}\left(\left(\frac{1}{T}\sum_{l=-m}^{m}w_{T,l}\sum_{t=\max(1,l+1)}^{\min(T,T+l)}V_{it}\left(\theta_0,\widehat{\gamma}_i(\theta_0)\right)V_{it-l}\left(\theta_0,\widehat{\gamma}_i(\theta_0)\right)'\right)^{-1}\right)+o_p(1)\end{aligned}$$

Using Lemma 5 in Supplementary Appendix, we obtain

$$\max_i\left|\frac{1}{T}\sum_{l=-m}^{m}w_{T,l}\sum_{t=\max(1,l+1)}^{\min(T,T+l)}V_{it}\left(\theta_0,\widehat{\gamma}_i(\theta_0)\right)V_{it-l}\left(\theta_0,\widehat{\gamma}_i(\theta_0)\right)'-\sum_{l=-\infty}^{\infty}E\left[V_{it}V_{it-l}'\right]\right|=o_p(1)$$

Furthermore, if the conditional likelihood is properly defined, then we should have V_{it} serially uncorrelated, which implies that

$$\begin{aligned}&\max_i\left|\frac{1}{T}\sum_{l=-m}^{m}w_{T,l}\sum_{t=\max(1,l+1)}^{\min(T,T+l)}V_{it}V_{it-l}'-E\left[V_{it}V_{it}'\right]\right|\\
&\qquad=\max_i\left|\frac{1}{T}\sum_{l=-m}^{m}w_{T,l}\sum_{t=\max(1,l+1)}^{\min(T,T+l)}V_{it}V_{it-l}'+E\left[V_{it}^{\gamma}\right]\right|=o_p(1)\end{aligned}$$

where the first equality is based on the information equality. Therefore, we obtain

$$\begin{aligned}&[4]+[5]\\
&= -\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{T}\sum_{l=-m}^{m}w_{T,l}\sum_{t=\max(1,l+1)}^{\min(T,T+l)}\left(\begin{array}{c}U_{it}^{\gamma}\left(\theta_0,\widehat{\gamma}_i(\theta_0)\right)\otimes V_{it-l}\left(\theta_0,\widehat{\gamma}_i(\theta_0)\right)'\\ +V_{it}\left(\theta_0,\widehat{\gamma}_i(\theta_0)\right)'\otimes U_{it-l}^{\gamma}\left(\theta_0,\widehat{\gamma}_i(\theta_0)\right)\end{array}\right)\right]\\
&\quad\cdot\mathrm{vec}\left(E\left[V_{it}^{\gamma}\right]^{-1}\right)+o_p(1)\end{aligned}$$

Using Lemma 5 again, we obtain

$$\begin{aligned}[4]+[5] &= -\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\sum_{l=-\infty}^{\infty} E\left[U_{it}^{\gamma}\otimes V_{it-l}'+V_{it}'\otimes U_{it-l}^{\gamma}\right]\operatorname{vec}\left(E\left[V_{it}^{\gamma}\right]^{-1}\right)+o_p(1)\\ &= \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\sum_{l=-\infty}^{\infty} E\left[U_{it}^{\gamma}\widetilde{V}_{it-l}+U_{it-l}^{\gamma}\widetilde{V}_{it}\right]+o_p(1)\\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{l=-\infty}^{\infty} E\left[U_{it}^{\gamma}\widetilde{V}_{it-l}\right]+o_p(1) \qquad (32)\end{aligned}$$

Combining (31) and (32), we obtain

$$S_n(\theta_0) = -\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n} E\left[U_{it}^{\gamma\gamma}\right]\operatorname{vec}\left(\left(E\left[V_{it}^{\gamma}\right]\right)^{-1}\right)+\frac{1}{n}\sum_{i=1}^{n}\sum_{l=-\infty}^{\infty} E\left[U_{it}^{\gamma}\widetilde{V}_{it-l}\right]+o_p(1) \qquad (33)$$

Now, we note that, under correct specification of conditional likelihood, the $\widetilde{V}_{it}$ would have zero serial correlation and we would therefore have $\sum_{l=-\infty}^{\infty} E\left[\widetilde{V}_{it}\widetilde{V}_{it-l}'\right] = E\left[\widetilde{V}_{it}\widetilde{V}_{it}'\right] = \left(E\left[V_i^{\gamma}\right]\right)^{-1} E\left[V_{it}V_{it}'\right]\left(E\left[V_i^{\gamma}\right]\right)^{-1}$. Furthermore, we have $E\left[V_{it}V_{it}'\right] = -E\left[V_i^{\gamma}\right]$ by the information equality. It follows that

$$S_n(\theta_0) = \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n} E\left[U_i^{\gamma_i\gamma_i}\right]\operatorname{vec}\left(\sum_{l=-\infty}^{\infty} E\left[\widetilde{V}_{it}\widetilde{V}_{it-l}'\right]\right)+\frac{1}{n}\sum_{i=1}^{n}\sum_{l=-\infty}^{\infty} E\left[U_{it}^{\gamma}\widetilde{V}_{it-l}\right]+o_p(1)$$

D Proof of Theorem 10

We have

$$\begin{aligned}S_n(\theta) &= \frac{1}{2n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^3\psi_{it}}{\partial\theta\left(\partial\gamma'\otimes\partial\gamma'\right)}\right)\operatorname{vec}\left(H_i^{-1}\Upsilon_i H_i^{-1}\right)\\ &\quad+\frac{1}{2n}\sum_{i=1}^{n}\frac{\partial\widehat{\gamma}_i'(\theta)}{\partial\theta}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^3\psi_{it}}{\partial\gamma\left(\partial\gamma'\otimes\partial\gamma'\right)}\right)\operatorname{vec}\left(H_i^{-1}\Upsilon_i H_i^{-1}\right)\\ &\quad+\frac{1}{2n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{l=-m}^{m}\sum_{t=\max(1,l+1)}^{\min(T,T+l)}\frac{\partial}{\partial\theta}\left(\left(\frac{\partial\psi_{it}}{\partial\gamma'}\right)\otimes\left(\frac{\partial\psi_{it-l}}{\partial\gamma'}\right)\right)\right)\operatorname{vec}\left(H_i^{-1}\right)\\ &\quad+\frac{1}{2n}\sum_{i=1}^{n}\frac{\partial\widehat{\gamma}_i'(\theta)}{\partial\theta}\left(\frac{1}{T}\sum_{l=-m}^{m}\sum_{t=\max(1,l+1)}^{\min(T,T+l)}\frac{\partial}{\partial\gamma}\left(\left(\frac{\partial\psi_{it}}{\partial\gamma'}\right)\otimes\left(\frac{\partial\psi_{it-l}}{\partial\gamma'}\right)\right)\right)\operatorname{vec}\left(H_i^{-1}\right)\end{aligned}$$

Proceeding as in Section C, we can obtain that

$$\begin{aligned}S_n(\theta_0) &= \frac{1}{2n}\sum_{i=1}^{n} E\left[U_{it}^{\gamma\gamma}\right]\operatorname{vec}\left(\left(E\left[V_{it}^{\gamma}\right]\right)^{-1}\left(\sum_{l=-\infty}^{\infty} E\left[V_{it}V_{it-l}'\right]\right)\left(E\left[V_{it}^{\gamma}\right]\right)^{-1}\right)\\ &\quad-\frac{1}{2n}\sum_{i=1}^{n}\sum_{l=-\infty}^{\infty} E\left[U_{it}^{\gamma}\otimes V_{it-l}'+V_{it}'\otimes U_{it-l}^{\gamma}\right]\operatorname{vec}\left(\left(E\left[V_{it}^{\gamma}\right]\right)^{-1}\right)+o_p(1)\\ &= \frac{1}{2n}\sum_{i=1}^{n} E\left[U_{it}^{\gamma\gamma}\right]\operatorname{vec}\left(\sum_{l=-\infty}^{\infty} E\left[\widetilde{V}_{it}\widetilde{V}_{it-l}'\right]\right)+\frac{1}{2n}\sum_{i=1}^{n}\sum_{l=-\infty}^{\infty} E\left[U_{it}^{\gamma}\widetilde{V}_{it-l}+U_{it-l}^{\gamma}\widetilde{V}_{it}\right]+o_p(1)\\ &= \frac{1}{2n}\sum_{i=1}^{n} E\left[U_{it}^{\gamma\gamma}\right]\operatorname{vec}\left(\sum_{l=-\infty}^{\infty} E\left[\widetilde{V}_{it}\widetilde{V}_{it-l}'\right]\right)+\frac{1}{n}\sum_{i=1}^{n}\sum_{l=-\infty}^{\infty} E\left[U_{it}^{\gamma}\widetilde{V}_{it-l}\right]+o_p(1)\end{aligned}$$

References

[1] Arellano, M. (2003): “Discrete Choices with Panel Data”, *Investigaciones Económicas*, 27, 423–458.

[2] Arellano, M. and J. Hahn (2006): “Understanding Bias in Nonlinear Panel Models: Some Recent Developments”. In: R. Blundell, W. Newey, and T. Persson (eds.): *Advances in Economics and Econometrics, Ninth World Congress*, Cambridge University Press, forthcoming.

[3] Bester, C. A. and C. Hansen (2005): “A Penalty Function Approach to Bias Reduction in Non-linear Panel Models with Fixed Effects”, unpublished manuscript.

[4] Card, D. and D. R. Hyslop (2005): “Estimating the Effects of a Time-Limited Earnings Subsidy for Welfare-Leavers”, *Econometrica*, 73, 1723-1770.

[5] Carro, J. (2006): “Estimating Dynamic Panel Data Discrete Choice Models with Fixed Effects”, *Journal of Econometrics*, forthcoming.

[6] Chamberlain, G. (1985): “Heterogeneity, Omitted Variable Bias, and Duration Dependence”, in J. J. Heckman and B. Singer (eds.), *Longitudinal Analysis of Labor Market Data*, Cambridge University Press.

[7] Cox, D. R. and N. Reid (1987): “Parameter Orthogonality and Approximate Conditional Inference” (with discussion), *Journal of the Royal Statistical Society*, Series B, 49, 1–39.

[8] DiCiccio, T. J. and S. E. Stern (1993): “An Adjustment to Profile Likelihood Based on Observed Information”, Technical Report, Department of Statistics, Stanford University.

[9] DiCiccio, T. J., M. A. Martin, S. E. Stern, and G. A. Young (1996): “Information Bias and Adjusted Profile Likelihoods”, *Journal of the Royal Statistical Society*, Series B, 58, 189–203.

[10] Ferguson, H., N. Reid, and D. R. Cox (1991): “Estimating Equations from Modified Profile Likelihood”, in Godambe, V. P. (ed.), *Estimating Functions*, Oxford University Press.

[11] Fernández-Val, I. (2005): “Estimation of Structural Parameters and Marginal Effects in Binary Choice Panel Data Models with Fixed Effects”, unpublished manuscript.

[12] Hahn, J., and G. Kuersteiner (2004): “Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects”, unpublished manuscript.

[13] Hahn, J. and W.K. Newey (2004): “Jackknife and Analytical Bias Reduction for Nonlinear Panel Models”, *Econometrica*, 72, 1295–1319.

[14] Hospido, L. (2006): "Modelling Heterogeneity and Dynamics in the Volatility of Individual Wages", unpublished manuscript.

[15] Neyman, J. and E. L. Scott (1948): "Consistent Estimates Based on Partially Consistent Observations", *Econometrica*, 16, 1–32.

[16] Pace, L. and A. Salvan (2005): "Adjustments of the Profile Likelihood from a New Perspective", *Journal of Statistical Planning and Inference*, forthcoming.

[17] Woutersen, T. (2002): "Robustness against Incidental Parameters", unpublished manuscript.