Global Economic Review, 2016
Vol. 45, No. 3, 251–274, http://dx.doi.org/10.1080/1226508X.2016.1211811

Routledge
Taylor & Francis Group

A likelihood-Based Approximate Solution to the Incidental Parameter Problem in Dynamic Nonlinear Models with Multiple Effects

MANUEL ARELLANO* & JINYONG HAHN**
*CEMFI, Madrid, Spain, **UCLA, Los Angeles, CA, USA

ABSTRACT *We discuss 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 consider 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 analyze the asymptotic properties of both types of 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.*

KEY WORDS: Nonlinear panel data; fixed effects; bias reduction

JEL CLASSIFICATION: C33

1. Introduction

It is now well understood that maximum likelihood (ML) estimates of the panel data models with fixed effects can be severely biased when the time series dimension T is small relative to the cross-sectional dimension n. Such bias is often discussed through asymptotic results such as the fixed-T inconsistency of the ML estimator for some models. A useful practical question is to ask how much heterogeneity can be given empirical content in a particular panel model and data set. 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.

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

Correspondence Address: Jinyong Hahn, UCLA, Los Angeles, USA. Email: hahn@econ.ucla.edu

distinguished in this literature. One approach is to construct and analytical or numerical bias correction of a fixed effects estimator. Hahn and Newey (2004) considered corrections of this type for static nonlinear panel data models, 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 (2007), and Fernández-Val (2005), among 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 (2006) 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 to order $1/T$ in nonlinear dynamic panel models with *multiple effects*. Analytic formulae for estimator correction as in Hahn and Newey (2004) or Hahn and Kuersteiner (2004) do not exist in the literature for models with multiple effects. Although derivation of such formulae does not present any conceptual challenge, the multiplicity of fixed effects is expected to make the derivation tedious and the resultant formulae complicated. Our approach produces estimators with the same asymptotic properties as that of the (yet non-existent) estimator correction approach. The advantage of our approach is the convenience in that it requires only a simple adjustment to the objective function. 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 (2009) as the integral of a bias-corrected moment equation. To be more precise, Bester and Hansen (2009) proposed objective Bayesian priors to reduce the bias in the posterior mode and posterior mean in a time series context. They show that the trace-based approach is asymptotically equivalent to using the data-dependent bias reducing prior, but it is not exactly identical in the finite sample. We show that their intuition carries over to panel context by explicitly adopting the alternative asymptotics.

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, especially 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's (1987) 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.

We acknowledge that correcting the objective function may require solving a highly nonlinear optimization problem, whereas an estimator correction approach would only require to estimate a correction term once. Computational element of our approach can be non-trivial as the adjusted objective functions are generally non-convex. As noted above, though, analytic formula for the estimator correction approach does not yet exist in the literature for models with multiple effects. Although it does not present any conceptual difficulty, we anticipate the correction term to take a rather complicated form. Practitioners should therefore weigh the convenience of simple analytic nature of our correction term against the potential computational difficulty.

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} $(t = 1, \ldots, T;\ i = 1, \ldots, n)$. 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}, $i = 1, \ldots, n$, where the fixed effects γ_{i0} are considered to be nonstochastic constants. We consider a maximization estimator defined by

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

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 $(\hat{\theta}, \hat{\gamma}_1, \ldots, \hat{\gamma}_n)$ 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

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

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

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

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

where

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

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

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

where the expectation is taken with respect to the distribution of x_{it}. Due to the noise in estimating $\hat{\gamma}_i(\theta)$, in general $\theta_T \neq \theta_0$ (Neyman Scott's 1948 incidental parameters problem). This problem would not occur if the quantities $\hat{\gamma}_i(\theta)$ were replaced by $\gamma_i(\theta)$ defined as[3]

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

So we could think of the infeasible concentrated objective function $\sum_{i=1}^{n} \bar{\psi}_i(\theta, \gamma_i(\theta))/n$ as a target criterion and $\sum_{i=1}^{n} \bar{\psi}_i(\theta, \hat{\gamma}_i(\theta))/n$ 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:

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

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 $\hat{B}(\theta)$, and then obtain an estimator that maximizes the adjusted criterion:

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

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

In order to gain intuition, consider an expansion for the first-order conditions around the truth

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

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

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

and that

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

Thus, also

$$\sqrt{nT}\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\bar{\psi}_i(\theta_0, \hat{\gamma}_i(\theta_0)) - \sqrt{\frac{n}{T}}\frac{\partial\hat{B}(\theta_0)}{\partial\theta} d\to\mathcal{N}(0, \Omega),$$

which suggests that as $n, T \to \infty$, $\sqrt{nT}(\tilde{\theta}-\theta_0)$ 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: Intuition

In this section, we provide an intuitive derivation of the bias formulae. The derivation is not meant to be a rigorous asymptotic argument, but is meant to provide a heuristic intuition. For example, we make an assumption that the expectations operator and the stochastic order symbols can be interchanged, which is clearly invalid in general. Rigorous asymptotic argument is provided later in Section 4 and Appendix.

Let us introduce the notation:

$$\bar{V}_i(\theta, \gamma_i) \equiv \frac{\partial\bar{\psi}_i(\theta, \gamma_i)}{\partial\gamma_i},$$

$$\bar{H}_i(\theta) \equiv -\lim_{T\to\infty} E\left[\frac{\partial\bar{V}_i(\theta, \gamma_i(\theta))}{\partial\gamma_i'}\right],$$

$$\bar{\Upsilon}_i(\theta) \equiv \lim_{T\to\infty} TE[\bar{V}_i(\theta, \gamma_i(\theta))\bar{V}_i(\theta, \gamma_i(\theta))'].$$

A first-order stochastic expansion for an arbitrary fixed θ gives[4]

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

Next, expanding $\bar{\psi}_i(\theta, \hat{\gamma}_i(\theta))$ around $\gamma_i(\theta)$ for fixed θ, we get

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

and combining the two expansions,

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

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

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

where

$$\beta_i(\theta) \equiv \frac{1}{2}\text{trace}[\bar{H}_i(\theta)^{-1}\bar{\Upsilon}_i(\theta)] = \frac{1}{2}\text{trace}\{\bar{H}_i(\theta)\text{Var}(\sqrt{T}[\hat{\gamma}_i(\theta) - \gamma_i(\theta)])\}. \tag{5}$$

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

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

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:

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

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

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

2.2. *Estimation of the Bias*

An estimator for the bias term in the modified concentrated likelihood (4) can be formed using

$$\hat{B}(\theta) = \sum_{i=1}^{n} \hat{\beta}_i(\theta)/n,$$

where $\hat{\beta}_i(\theta)$ is a sample counterpart of the previous formulae.

Trace-based approach. One possibility is

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

where

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

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

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

The quantity m is a bandwidth parameter and $w_{T,l}$ denotes a weight that guarantees positive definiteness of $\Upsilon_i(\theta, \gamma)$, for example, a Bartlett kernel weight such that $w_{T,l} = 1 - l/(m+1)$.[5] Note that with $m = T - 1$ and $w_{T,l} = 1$, $\Upsilon_i(\theta, \gamma) \equiv \bar{V}_i(\theta, \gamma_i(\theta)) \bar{V}_i(\theta, \gamma_i(\theta))'$, so that in such a case $\Upsilon_i(\theta, \hat{\gamma}_i(\theta)) \equiv 0$.

The resulting modified concentrated likelihood function is

$$L_T(\theta) = \sum_{i=1}^{n} \sum_{t=1}^{T} \psi(x_{it}; \theta, \hat{\gamma}_i(\theta)) - \frac{1}{2} \sum_{i=1}^{n} \text{trace}(H_i(\theta, \hat{\gamma}_i(\theta))^{-1} \Upsilon_i(\theta, \hat{\gamma}_i(\theta))). \tag{10}$$

For the trace-based approach, we do not need to guarantee positive definiteness of $\Upsilon_i(\theta, \gamma)$, and can, in fact, use the unweighted truncated estimator as was done by Hahn and Kuersteiner (2004). We use the positive-definite version solely for the purpose of facilitating the comparison with other approaches.[6]

The adjustment term $\hat{\beta}_i(\theta)$ 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). The choice of the bandwidth parameter m is an important matter for practical implementation, but it requires a highly complicated analysis, and as such, is beyond the scope of the current paper. We note that Hahn and Kuersteiner (2007) developed such a method for correcting the estimator in a panel model with scalar fixed effects.

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

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

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

$$\tilde{\beta}_i(\theta) = -\frac{1}{2}\ln\det[H_i(\theta,\hat{\gamma}_i(\theta))] + \frac{1}{2}\ln\det[\Upsilon_i(\theta,\hat{\gamma}_i(\theta))]. \tag{11}$$

The resulting modified concentrated likelihood function is

$$L_D(\theta) = \sum_{i=1}^{n}\sum_{t=1}^{T}\psi(x_{it};\theta,\hat{\gamma}_i(\theta)) + \frac{1}{2}\sum_{i=1}^{n}\ln\det[H_i(\theta,\hat{\gamma}_i(\theta))] - \frac{1}{2}\sum_{i=1}^{n}\ln\det[\Upsilon_i(\theta,\hat{\gamma}_i(\theta))], \tag{12}$$

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 et al. (1996).

Using the arguments in Pace and Salvan (2006), 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 et al. (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

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

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

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

where the variance term is given by the sandwich formula:

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

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 $\hat{\gamma}_i(\theta)$ 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

$$\Upsilon_{Ti}(\theta,\gamma;\theta_0,\gamma_{i0}) \equiv \sum_{l=-m}^{m} w_{T,l}\Gamma_{Tl}(\theta,\gamma;\theta_0,\gamma_{i0}), \tag{13}$$

where, for $l > 0$, we have

$$\Gamma_{Tl}(\theta,\gamma;\theta_0,\gamma_{i0}) = \frac{1}{T-l}\sum_{t=l+1}^{\mathrm{T}} E_{\theta_0,\gamma_{i0}}\left[\frac{\partial\psi(x_{it};\theta,\gamma_i)}{\partial\gamma_i}\frac{\partial\psi(x_{it-l};\theta,\gamma_i)}{\partial\gamma_i'}\right]. \tag{14}$$

Alternatively, a centered covariance could be calculated:

$$\Gamma^{*}_{Tl}(\theta,\gamma;\theta_0,\gamma_{i0}) = \Gamma_{Tl}(\theta,\gamma;\theta_0,\gamma_{i0}) - \mu_{T0}(\theta,\gamma;\theta_0,\gamma_{i0})\mu_{Tl}(\theta,\gamma;\theta_0,\gamma_{i0})', \tag{15}$$

where $\mu_{Tl}(\theta,\gamma;\theta_0,\gamma_{i0}) = (T-l)^{-1}\sum_{t=l+1}^{\mathrm{T}} E_{\theta_0,\gamma_{i0}}[V_i(x_{it-l};\theta,\gamma)]$. 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

$$L_{\mathrm{ED}}(\theta;\hat{\theta}) = \sum_{i=1}^{n}\sum_{t=1}^{\mathrm{T}} \psi(x_{it};\theta,\hat{\gamma}_i(\theta)) + \frac{1}{2}\sum_{i=1}^{n}\ln\det H_i(\theta,\hat{\gamma}_i(\theta))$$
$$- \frac{1}{2}\sum_{i=1}^{n}\ln\det\Upsilon_{Ti}(\theta,\hat{\gamma}_i(\theta);\hat{\theta},\hat{\gamma}_i(\hat{\theta})). \tag{16}$$

Note that expectations can also be taken to simplify the expression of the Hessian $H_i(\theta, \hat{\gamma}_i(\theta))$.

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

$$\hat{\theta}_K = \arg\max_\theta L_{\text{ED}}(\theta; \hat{\theta}_{K-1}) \tag{17}$$

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

$$S_{\text{ED}}(\hat{\theta}_\infty; \hat{\theta}_\infty) = 0, \tag{18}$$

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

Although this approach is intuitive, we should remark that we have not developed a rigorous theory to justify this approach.

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, especially 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 an incomplete model setting, it is natural to use the trace form of the correction and a kernel-based estimate of $\bar{\Upsilon}_i(\theta)$, which is the only possibility available.

3. Examples

We consider four examples. The first one is a simple panel AR(1) model with scalar fixed effects. This simple model has been researched heavily in the literature, and multiplicity consideration is irrelevant. We discuss the model only to make it easier to compare various approaches. The rest of the examples do include multiple fixed effects. The second one is static and linear, but illustrates the differences between the two approaches in a familiar context. The third one is a conditional volatility model, and the last one is a dynamic binary choice formulation.

Example 1 Consider the dynamic panel model with fixed effects with known variance of the error term, where

$$\psi(x_{it};\theta,\gamma_i) = -\frac{1}{2}\ln\sigma^2 - \frac{1}{2}\frac{(y_{it}-\theta y_{it-1}-\gamma_i)^2}{\sigma^2}.$$

We will assume that the initial value is taken from a stationary distribution, so that $y_{it} \sim N(\gamma_{i0}/(1-\theta_0), \sigma^2/(1-\theta_0^2))$. In order to make it simple, we assume that $\sigma^2 = 1$, and we obtain

$$\psi(x_{it};\theta,\gamma_i) = -\frac{1}{2}(y_{it}-\theta y_{it-1}-\gamma_i)^2.$$

We note that

$$\begin{aligned}
\gamma_i(\theta) &= E[y_{it}-\theta y_{it-1}] = \frac{1-\theta}{1-\theta_0}\gamma_i,\\
\bar{H}_i(\theta) &= 1,\\
\bar{\Upsilon}_i(\theta) &= \lim_{T\to\infty} \mathrm{T}\mathrm{E}\left[\left(\left(\bar{y}_{i\cdot} - \frac{\gamma_i}{1-\theta_0}\right) - \theta\left(\bar{y}_{i\cdot-1} - \frac{\gamma_i}{1-\theta_0}\right)\right)^2\right] = \frac{(1-\theta)^2}{(1-\theta_0)^2},\\
\beta_i(\theta) &= \frac{1}{2}\bar{\Upsilon}_i(\theta),
\end{aligned}$$

where $\bar{y}_{i\cdot} \equiv \sum_{t=1}^{\mathrm{T}} y_{it}$ and $\bar{y}_{i\cdot-1} \equiv \sum_{t=1}^{\mathrm{T}} y_{it-1}$. We also note that

$$\begin{aligned}
\hat{\gamma}_i(\theta) &= \bar{y}_{i\cdot} - \theta\bar{y}_{i\cdot-1},\\
H_i(\theta,\hat{\gamma}_i(\theta)) &= 1,\\
\Upsilon_i(\theta,\hat{\gamma}_i(\theta)) &= \frac{1}{T}\sum_{l=-m}^{m} w_{T,l}\left(\sum_{t=\max(1,l+1)}^{\min(T,T+l)} (\tilde{y}_{it}-\theta\tilde{x}_{it})(\tilde{y}_{it-l}-\theta\tilde{x}_{it-l})\right),\\
\hat{\beta}_i(\theta) &= \frac{1}{2}\Upsilon_i(\theta,\hat{\gamma}_i(\theta)),
\end{aligned}$$

where $\tilde{y}_{it} \equiv y_{it} - \bar{y}_{i\cdot}$, and $\tilde{x}_{it} \equiv y_{it-1} - \bar{y}_{i\cdot-1}$. It follows that the trace-based approach would maximize

$$\begin{aligned}
&-\frac{1}{2}\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{\mathrm{T}}(\tilde{y}_{it}-\theta\tilde{x}_{it})^2\\
&-\frac{1}{2}\frac{1}{nT}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{l=-m}^{m} w_{T,l}\left(\sum_{t=\max(1,l+1)}^{\min(T,T+l)} (\tilde{y}_{it}-\theta\tilde{x}_{it})(\tilde{y}_{it-l}-\theta\tilde{x}_{it-l})\right)\right).
\end{aligned}$$

On the other hand, the determinant-based approach would maximize

$$-\frac{1}{2}\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{\mathrm{T}}(\tilde{y}_{it}-\theta\tilde{x}_{it})^2$$
$$-\frac{1}{2}\sum_{i=1}^{n}\ln\left(\frac{1}{T}\sum_{l=-m}^{m}w_{T,l}\left(\sum_{t=\max(1,l+1)}^{\min(T,T+l)}(\tilde{y}_{it}-\theta\tilde{x}_{it})(\tilde{y}_{it-l}-\theta\tilde{x}_{it-l})\right)\right).$$

If we use the expected quantities instead, we note that

$$E_{\theta_0,\gamma_{i0}}[y_{it}y_{it-l}]=\left(\frac{\gamma_{i0}}{1-\theta_0}\right)^2+\frac{\theta_0^l}{1-\theta_0^2},$$

from which we obtain

$$\begin{aligned}\Gamma_{Tl}(\theta,\gamma;\theta_0,\gamma_{i0}) &= E_{\theta_0,\gamma_{i0}}[(y_{it}-\theta y_{it-1}-\gamma_i)(y_{it-l}-\theta y_{it-1-l}-\gamma_i)]\\ &=\left(\frac{\gamma_{i0}}{1-\theta_0}\right)^2+\frac{\theta_0^l}{1-\theta_0^2}-\theta\left(\left(\frac{\gamma_{i0}}{1-\theta_0}\right)^2+\frac{\theta_0^{l-1}}{1-\theta_0^2}\right)-\gamma_i\frac{\gamma_{i0}}{1-\theta_0}\\ &\quad -\theta\left(\left(\frac{\gamma_{i0}}{1-\theta_0}\right)^2+\frac{\theta_0^{l+1}}{1-\theta_0^2}\right)+\theta^2\left(\left(\frac{\gamma_{i0}}{1-\theta_0}\right)^2+\frac{\theta_0^l}{1-\theta_0^2}\right)+\gamma_i\theta\frac{\gamma_{i0}}{1-\theta_0}\\ &\quad -\gamma_i\frac{\gamma_{i0}}{1-\theta_0}+\gamma_i\theta\frac{\gamma_{i0}}{1-\theta_0}+\gamma_i^2\\ &=\frac{(1-\theta)^2}{(1-\theta_0)^2}\gamma_{i0}^2+\frac{\theta_0^{l-1}}{1-\theta_0^2}(\theta_0-\theta)(1-\theta_0\theta)+\gamma_i^2-2\gamma_i\frac{\gamma_{i0}}{1-\theta_0}+2\gamma_i\theta\frac{\gamma_{i0}}{1-\theta_0}\end{aligned}$$

and

$$\begin{aligned}\Gamma_{Tl}(\theta,\hat{\gamma}_i(\theta);\hat{\theta},\hat{\gamma}_i(\hat{\theta})) &=\frac{(1-\theta)^2}{(1-\hat{\theta})^2}\hat{\gamma}_i(\hat{\theta})^2+\frac{\hat{\theta}^{l-1}}{1-\hat{\theta}^2}(\hat{\theta}-\theta)(1-\hat{\theta}\theta)\\ &\quad +\hat{\gamma}_i(\theta)^2-2\hat{\gamma}_i(\theta)\frac{\hat{\gamma}_i(\hat{\theta})}{1-\hat{\theta}}+2\hat{\gamma}_i(\theta)\theta\frac{\hat{\gamma}_i(\hat{\theta})}{1-\hat{\theta}},\\ \Upsilon_{Ti}(\theta,\hat{\gamma}_i(\theta);\hat{\theta},\hat{\gamma}_i(\hat{\theta})) &=\sum_{l=-m}^{m}w_{T,l}\Gamma_{Tl}(\theta,\hat{\gamma}_i(\theta);\hat{\theta},\hat{\gamma}_i(\hat{\theta}))\end{aligned}$$

the determinant-based approach would maximize

$$-\frac{1}{2}\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{\mathrm{T}}(\tilde{y}_{it}-\theta\tilde{x}_{it})^2-\frac{1}{2}\sum_{i=1}^{n}\ln\det\Upsilon_{Ti}(\theta,\hat{\gamma}_i(\theta);\hat{\theta},\hat{\gamma}_i(\hat{\theta})).$$

Example 2 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 $\theta = \text{vech}[\Omega(\theta)]$. We have $\hat{\gamma}_i(\theta) = \bar{x}_i$ and

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

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

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

$$\Gamma_l(\theta, \hat{\gamma}_i(\theta)) \equiv \Omega(\theta)^{-1}\left[\frac{1}{T}\sum_{t=\max(1,l+1)}^{\min(T,T+l)}(x_{it} - \bar{x}_i)(x_{it-l} - \bar{x}_i)'\right]\Omega(\theta)^{-1}.$$

The determinant approach with $m = 0$ gives

$$L_D(\theta) = C - \frac{nT}{2}\ln\det\Omega(\theta) - \frac{1}{2}\sum_{i=1}^{n}\sum_{t=1}^{\mathrm{T}}(x_{it} - \bar{x}_i)'\Omega(\theta)^{-1}(x_{it} - \bar{x}_i)$$
$$+ \frac{n}{2}\ln\det[\Omega(\theta)^{-1}] - \frac{1}{2}\sum_{i=1}^{n}\ln\det\left(\Omega(\theta)^{-1}\frac{1}{T}\sum_{t=1}^{\mathrm{T}}(x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)'\Omega(\theta)^{-1}\right).$$

Finally, collecting terms and discarding constants, we get

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

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

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

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

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

$$\hat{\beta}_i(\theta) = \frac{1}{2}\text{trace}[\tilde{\Gamma}_i\Omega(\theta)^{-1}],$$

where

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

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

$$L_{\mathrm{TR}}(\theta) = C - \frac{nT}{2}\ln\det\Omega(\theta) - \frac{nT}{2}\mathrm{trace}[\Omega(\theta)^{-1}\hat{\Omega}] - \frac{n}{2}\mathrm{trace}[\Omega(\theta)^{-1}\tilde{\Gamma}].$$

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

$$\tilde{\Omega}_{\mathrm{TR}} = \hat{\Omega} + \frac{1}{T}\tilde{\Gamma}.$$

Example 3 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

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

A model of this type, but with an exponential ARCH formulation of the conditional variance, is developed in Hospido (2012), 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 4 The next example is an autoregressive binary formulation of the form

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

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:*

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

$$\rho_i \equiv E\left[\frac{\partial^2\psi(x_{it};\theta_0,\gamma_{i0})}{\partial\theta\partial\gamma_i'}\right]\left(E\left[\frac{\partial^2\psi(x_{it};\theta_0,\gamma_{i0})}{\partial\gamma_i\partial\gamma_i'}\right]\right)^{-1}, \quad \mathcal{I}_i \equiv -E\left[\frac{\partial U_i(x_{it};\theta_0,\gamma_{i0})}{\partial\theta'}\right],$$

$$\tilde{V}_{it} \equiv -\left(E\left[\frac{\partial V_i}{\partial\gamma_i'}\right]\right)^{-1} V_{it}.$$

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

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

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

The normalized score $(1/nT)\sum_{i=1}^{n}\sum_{t=1}^{T} U(x_{it};\theta_0,\hat{\gamma}_i(\theta_0))$ has an asymptotic bias, which renders the fixed effects estimator $\hat{\theta}$ biased. The asymptotic bias of the normalized score can be shown[7] to be equal to $1/T$ times $\Psi_0(\theta_0,\{\gamma_{10},\gamma_{20},\ldots\})$, where

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

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

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

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

$$0 = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{\mathrm{T}} U(x_{it}; \tilde{\theta}, \hat{\gamma}_i(\tilde{\theta})) - \frac{1}{T} S_n(\tilde{\theta}) \tag{19}$$

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

$$\underset{\theta, \gamma_1, \ldots, \gamma_n}{\operatorname{argmax}} \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{\mathrm{T}} \psi(x_{it}; \theta, \gamma_i) - \frac{1}{T} B_n(\theta). \tag{20}$$

Remark 1 Comparison of (10), (12), (16) with (20) implies that $B_n(\theta)$ is equal to

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

for a determinant approach,

$$B_n(\theta) = \frac{1}{2n}\sum_{i=1}^{n} \ln\det H_i(\theta, \hat{\gamma}_i(\theta)) - \frac{1}{2n}\sum_{i=1}^{n} \ln\det \Upsilon_{Ti}(\theta, \hat{\gamma}_i(\theta); \hat{\theta}, \hat{\gamma}_i(\hat{\theta})) \tag{22}$$

for an expectation-based determinant approach and

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

for a trace approach, where $H_i(\theta, \gamma)$ and $\Upsilon_i(\theta, \gamma)$ are as defined in Equations (7)–(9).

We impose the following conditions:

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

Condition 2 $\sup_\theta (1/T)|\partial S_n(\theta)/\partial\theta'| = o_p(1)$.

Condition 3 $S_n(\theta_0) = (1/n)\sum_{i=1}^{n}\sum_{l=-\infty}^{\infty} E[U_{it}^{\gamma_i}\tilde{V}_{it-l}]$
$+\frac{1}{2}(1/n)\sum_{i=1}^{n} E[U_i^{\gamma_i\gamma_i}]\operatorname{vec}(\sum_{l=-\infty}^{\infty} E[\tilde{V}_{it}\tilde{V}'_{it-l}]) + o_p(1)$.

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

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

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

where $\mathcal{I} \equiv \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \mathcal{I}_i$ *and* Ω *is the asymptotic variance of* $(1/\sqrt{nT}) \sum_{i=1}^{n} \sum_{t=1}^{T} U(x_{it}; \theta_0, \gamma_{i0})$.

Proof See Appendix B. □

Following three results show that the $B_n(\theta)$ in Equations (21)–(23) satisfy Conditions 1–3, and as such the resultant estimator is asymptotically free of bias.

THEOREM 3 Assume that the model is given by the likelihood. Also assume that the regularity conditions in Appendix A hold. Further assume that $m = o(T^{2/5})$. Then, $B_n(\theta)$ as defined inEquation (21) satisfies Conditions 1–3.

Proof See Appendix C.[8] □

THEOREM 4 *Assume that the model is given by the likelihood. Also assume that the regularity conditions in Appendix A hold. Further assume that* $m = o(T^{2/5})$. *Then,* $B_n(\theta)$ *based on*

$$\bar{\Upsilon}_i(\theta, \gamma) \equiv \sum_{l=-m}^{m} w_{T,l} E_{\hat{\theta},\hat{\gamma}_i}\left[\frac{\partial \psi_{it}(\theta, \gamma)}{\partial \gamma} \frac{\partial \psi_{it-l}(\theta, \gamma)}{\partial \gamma'}\right] \tag{24}$$

satisfies Conditions 1–3. (*Here,* $E_{\hat{\theta},\hat{\gamma}_i}[\cdot]$ *denotes an expectation taken with respect to the density evaluated at* $(\hat{\theta}, \hat{\gamma}_i)$.) *The same result hold even when the preliminary estimates* $(\hat{\theta}, \hat{\gamma}_i)$ *in Equation* (24) *are replaced by some* (θ^*, γ_i^*) *such that* $\|\theta^* - \theta\| = O_p(T^{-2/5})$ *and* $\sup_i \|\gamma_i^* - \gamma_{i0}\| = O_p(T^{-2/5})$.

Proof The proof of Theorem 4 is similar to that of Theorem 3. It is available upon request in Supplementary Appendix.[9] □

THEOREM 5 *Assume that the regularity conditions in Appendix A hold. Further assume that* $m = o(T^{1/2})$. *Then,* $B_n(\theta)$ *as defined in Equation* (23) *satisfies Conditions* 1–3.

Proof See Appendix D.[10] □

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 data sets.

Notes

1. See also Arellano and Hahn (2006) for a review of the literature.
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}.
4. We assume that all the $O_p(\cdot)$ terms are uniform over i in this section.
5. For simplicity of exposition, we will assume that the $w_{T,l}$ are indeed Bartlett weights throughout the rest of the paper.
6. The proof for asymptotic theory goes through with only one change: In place of Lemma 5 in the Supplementary Appendix, we need to use Lemma 6 instead.
7. This is a standard result, but we do provide a rigorous derivation in Supplementary Appendix, which is available upon request.
8. The proof of Theorem 3 uses the information equality. This explains why the likelihood setup is required here.
9. The proof of Theorem 4 uses the information equality. This explains why we required the likelihood setup.
10. The proof of Theorem 5 does not use the information equality. We therefore do not require the likelihood setup here.
11. The consistency result is available as Theorem 8 in a Supplementary Appendix, which is available upon request. Theorem 8 shows that, if Assumptions 1–3, and Condition 1 are satisfied, then $\Pr[|\tilde{\theta} - \theta_0| \geq \eta] = o(T^{-1})$ for every $\eta > 0$.
12. See Lemma 7 in Supplementary Appendix A.
13. In Supplementary Appendix C, we provide a rigorous proof of the expansion (A3).
14. The proofs of Theorems 6 and 7 are in Supplementary Appendix, which is available upon request.
15. See Lemma 7 in Supplementary Appendix.

References

Arellano, M. (2003) Discrete choices with panel data, *Investigaciones Económicas*, 27, 423–458.

Arellano, M., and Hahn, J. (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* (New York: Cambridge University Press).

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

Card, D., and Hyslop, D. R. (2005) Estimating the effects of a time-limited earnings subsidy for welfare-leavers, *Econometrica*, 73, 1723–1770.

Carro, J. (2007) Estimating dynamic panel data discrete choice models with fixed effects, *Journal of Econometrics*, 140, 503–528.

Chamberlain, G. (1985) Heterogeneity, omitted variable bias, and duration dependence, in: J. J. Heckman and B. Singer (Eds.) *Longitudinal Analysis of Labor Market Data* (New York: Cambridge University Press).

Cox, D. R., and Reid, N. (1987) Parameter orthogonality and approximate conditional inference (with discussion), *Journal of the Royal Statistical Society, Series B*, 49, 1–39.

DiCiccio, T. J., and Stern, S. E. (1993) An adjustment to profile likelihood based on observed information, Technical Report, Department of Statistics, Stanford University.

DiCiccio, T. J., Martin, M. A., Stern, S. E., and Young, G. A. (1996) Information bias and adjusted profile likelihoods, *Journal of the Royal Statistical Society, Series B*, 58, 189–203.

Ferguson, H., Reid, N., and Cox, D. R. (1991) Estimating equations from modified profile likelihood, in: V. P. Godambe (Ed.) *Estimating Functions* (New York: Oxford University Press).

Fernández-Val, I. (2005) 'Estimation of Structural Parameters and Marginal Effects in Binary Choice Panel Data Models with Fixed Effects', unpublished manuscript.

Hahn, J., and Kuersteiner, G. (2004) 'Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects', unpublished manuscript.

Hahn, J., and Kuersteiner, G. (2007) 'Bandwidth Choice for Bias Estimators in Dynamic Nonlinear Panel Models', unpublished manuscript.

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

Hospido, L. (2012) Modelling Heterogeneity and Dynamics in the Volatility of Individual Wages, *Journal of Applied Econometrics*, 27, 386–414.

Neyman, J., and Scott, E. L. (1948) Consistent estimates based on partially consistent observations, *Econometrica*, 16, 1–32.

Pace, L., and Salvan, A. (2006) Adjustments of the profile likelihood from a new perspective, *Journal of Statistical Planning and Inference*, 136, 3554–3564.

Woutersen, T. (2002) 'Robustness Against Incidental Parameters', unpublished manuscript.

Appendix A. Regularity Conditions

ASSUMPTION 1 *For each* $\eta > 0$, $\inf_i [G_{(i)}(\theta_0, \gamma_{i0}) - \sup_{\{(\theta,\gamma):|(\theta,\gamma)-(\theta_0,\gamma_{i0})|>\eta\}} G_{(i)}(\theta, \gamma)] > 0$, *where* $G_{(i)}(\theta, \gamma) \equiv E[\psi(x_{it}; \theta, \gamma)]$.

ASSUMPTION 2 $n, T \to \infty$ *such that* $n/T \to \kappa$, *where* $0 < \kappa < \infty$.

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

ASSUMPTION 4 *Let* $\psi(x_{it}, \phi)$ *be a function indexed by the parameter* $\phi = (\theta, \gamma) \in int\ \Phi$, *where* Φ *is a compact, convex subset of* $\mathbb{R}^*$, $p \equiv \dim(\phi)$, *and* $R = \dim(\theta)$. *Let* $\nu = (\nu_1, \ldots, \nu_k)$ *be a vector of non-negative* *integers* v_i, $|v| = \sum_{j=1}^{k} v_j$ *and* $D^v\psi(x_{it}, \phi) = \partial^{|v|}\psi(x_{it}, \phi)/(\partial\phi_1^{v_1}\cdots\partial\phi_k^{v_k})$*. 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* $\sup_i E[|M(x_{it})|^{10q+12+\delta}] < \infty$ *for some integer* $q \geq p/2 + 2$ *and for some* $\delta > 0$*.*

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

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

Remark 2 Assumption 6 is stronger than the one assumed in Hahn and Kuersteiner (2004) in the sense that Assumption 6(ii) was not imposed there.

Assumption 7 *Let* $\mu_{i1} \leq \cdots \leq \mu_{ik} \leq \cdots \leq \mu_{iR}$ *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 $\sup_{(\theta,\gamma)\in\Phi}\sup_l E_{\theta,\gamma}[M(x_{it})M(x_{it-l})] < \infty$.

Appendix B. Proof of Theorem thm2

We focus on asymptotic normality here, taking consistency result as given.[11] Because $0 = \sum_{t=1}^{T} V(x_{it}; \tilde\theta, \hat\gamma_i(\tilde\theta))$ by definition, $\tilde\theta$ can be given the alternative characterization

$$0 = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} U(x_{it}; \tilde\theta, \hat\gamma_i(\tilde\theta)) - \frac{1}{T}S_n(\tilde\theta).$$

By the Taylor series expansion, we obtain

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

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

$$\sqrt{nT}(\tilde\theta - \theta_0) = \left(\frac{1}{n}\sum_{i=1}^{n}\bar{\mathcal{I}}_i + \frac{1}{T}\frac{\partial S_n(\bar\theta)}{\partial\theta'}\right)^{-1}\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T} U(x_{it}; \theta_0, \hat\gamma_i(\theta_0)) - \sqrt{\frac{n}{T}}S_n(\theta_0)\right). \tag{A1}$$

It can be shown[12] that $(1/n)\sum_{i=1}^{n}\bar{\mathcal{I}}_i = \mathcal{I} + o_p(1)$. By Condition 2, we also have $(1/T)\partial S_n(\bar\theta)/\partial\theta' = o_p(1)$. We therefore have

$$\frac{1}{n}\sum_{i=1}^{n}\bar{\mathcal{I}}_i + \frac{1}{T}\frac{\partial S_n(\bar\theta)}{\partial\theta'} = \mathcal{I} + o_p(1). \tag{A2}$$

By applying a second-order Taylor series approximation to $(1/nT)\sum_{i=1}^{n}\sum_{t=1}^{T} U(x_{it}; \theta_0, \hat\gamma_i(\theta_0))$ around γ_{i0}, and noting that $\hat\gamma_i(\theta_0) - \gamma_{i0} = -(E[\partial V_i/\partial\gamma_i'])^{-1}((1/T)\sum_{t=1}^{T} V_{it}) + o_p(1/\sqrt{T}) = (1/T)\sum_{t=1}^{T}\tilde V_{it} + o_p(1/\sqrt{T})$, we can anticipate that[13]

$$\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U(x_{it};\theta_0,\hat{\gamma}_i(\theta_0))=\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}+\sqrt{\frac{n}{T}}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}U_{it}^{\gamma_i}\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{V}_{it}\right)$$
$$+\sqrt{\frac{n}{T}}\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}E[U_i^{\gamma_i\gamma_i}]\left[\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{V}_{it}\right)\otimes\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{V}_{it}\right)\right]+o_p(1). \quad \text{(A3)}$$

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

$$\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}E[U_i^{\gamma_i\gamma_i}]\left[\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{V}_{it}\right)\otimes\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{V}_{it}\right)\right]+\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}U_{it}^{\gamma_i}\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{V}_{it}\right)$$
$$=\frac{1}{n}\sum_{i=1}^{n}\sum_{l=-\infty}^{\infty}E[U_i^{\gamma_i}\tilde{V}_{it-l}]+\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}E[U_i^{\gamma_i\gamma_i}]\mathrm{vec}\left(\sum_{l=-\infty}^{\infty}E[\tilde{V}_{it}\tilde{V}_{it-l}']\right)+o_p(1),$$

which, when combining (A1)–(A3) and Condition 3, yields

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

from which the conclusion follows.

Appendix C. Proof of Theorem 3

We will first state without proof[14] that Conditions 1 and 2 are satisfied:

THEOREM 6 *Assume that the regularity conditions in Appendix A hold. Then, $B_n(\theta)$ as defined in Equation (21) satisfies Condition 1.*

THEOREM 7 *Assume that the regularity conditions in Appendix A hold. Further assume that $m=o(T^{1/2})$. Then, $B_n(\theta)$ as defined in Equation (21) satisfies Condition 2.*

Below, we show that Condition 3 is satisfied. By differentiating B_n, we obtain that $S_n(\theta)=[2]+\cdots+[5]$, where

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

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

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

and

$$[5] \equiv \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\frac{\partial\hat{\gamma}_i'(\theta)}{\partial\theta}\left[\frac{1}{T}\sum_{l=-m}^{m}w_{T,l}\left(\sum_{t=\max(1,l+1)}^{\min(T,T+l)}\frac{\partial}{\partial\gamma}\left(\left(\frac{\partial\psi_{it}}{\partial\gamma'}\right)\otimes\left(\frac{\partial\psi_{it-l}}{\partial\gamma'}\right)\right)\right)\right]$$
$$\cdot\ \mathrm{vec}\left(\left(\frac{1}{T}\sum_{l=-m}^{m}w_{T,l}\left(\sum_{t=\max(1,l+1)}^{\min(T,T+l)}\frac{\partial\psi_{it}}{\partial\gamma}\frac{\partial\psi_{it-l}}{\partial\gamma'}\right)\right)^{-1}\right).$$

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

$$\frac{\partial\hat{\gamma}_i'(\theta)}{\partial\theta} = -\left(\sum_{t=1}^{T}\frac{\partial^2\psi_{it}(\theta,\hat{\gamma}_i(\theta))}{\partial\theta\partial\gamma'}\right)\left(\sum_{t=1}^{T}\frac{\partial^2\psi_{it}(\theta,\hat{\gamma}_i(\theta))}{\partial\gamma\partial\gamma'}\right)^{-1}. \tag{A4}$$

In the discussion below, all the terms [2], ..., [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 $\partial^3\psi_{it}(\theta,\gamma)/\partial\theta(\partial\gamma'\otimes\partial\gamma') = U_{it}^{\gamma\gamma} + \rho_i V_{it}^{\gamma\gamma}$, where $V_{it}^{\gamma\gamma}(\theta,\gamma_i) = \partial^2 V_{it}(\theta,\gamma_i)/\partial\gamma'\otimes\partial\gamma'$. (Recall that $\rho_i \equiv E[\partial^2\psi(x_{it};\theta_0,\gamma_{i0})/\partial\theta\partial\gamma_i'](E[\partial^2\psi(x_{it};\theta_0,\gamma_{i0})/\partial\gamma_i\partial\gamma_i'])^{-1}$.) It turns out that all the averages over t on the RHS of [2] is uniformly consistent over i.[15] We therefore obtain

$$[2] = -\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}(E[U_{it}^{\gamma\gamma}] + \rho_i E[V_{it}^{\gamma\gamma}])\mathrm{vec}((E[V_{it}^{\gamma}])^{-1}) + o_p(1). \tag{A5}$$

The uniform consistency over i combined with Equation (A6) also implies that

$$\max_i\left|\frac{\partial\hat{\gamma}_i'(\theta)}{\partial\theta} + \rho_i\right| = o_p(1). \tag{A6}$$

Using the uniform consistency and Equation (A6), we obtain

$$[3] = \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\rho_i E[V_{it}^{\gamma\gamma}]\mathrm{vec}((E[V_{it}^{\gamma}])^{-1}) + o_p(1). \tag{A7}$$

Combining Equations (A5) and (A7), we obtain

$$[2] + [3] = -\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}E[U_{it}^{\gamma\gamma}]\mathrm{vec}((E[V_{it}^{\gamma}])^{-1}) + o_p(1). \tag{A8}$$

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

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

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

$$[4]+[5]=\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{T}\sum_{l=-m}^{m}w_{T,l}\sum_{t=\max(1,l+1)}^{\min(T,T+l)}\begin{pmatrix}U_{it}^{\gamma}(\theta_0,\hat{\gamma}_i(\theta_0))\otimes V_{it-l}(\theta_0,\hat{\gamma}_i(\theta_0))'\\+V_{it}(\theta_0,\hat{\gamma}_i(\theta_0))'\otimes U_{it-l}^{\gamma}(\theta_0,\hat{\gamma}_i(\theta_0))\end{pmatrix}\right]$$
$$\cdot\operatorname{vec}\left(\left(\frac{1}{T}\sum_{l=-m}^{m}w_{T,l}\sum_{t=\max(1,l+1)}^{\min(T,T+l)}V_{it}(\theta_0,\hat{\gamma}_i(\theta_0))V_{it-l}(\theta_0,\hat{\gamma}_i(\theta_0))'\right)^{-1}\right)+o_p(1).$$

Using Lemma 5 in Supplementary Appendix, we obtain

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

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

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

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

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

Using Lemma 5 again, we obtain

$$\begin{aligned}[4]+[5]&=-\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\sum_{l=-\infty}^{\infty}E[U_{it}^{\gamma}\otimes V_{it-l}'+V_{it}'\otimes U_{it-l}^{\gamma}]\operatorname{vec}(E[V_{it}^{\gamma}]^{-1})+o_p(1)\\&=\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\sum_{l=-\infty}^{\infty}E[U_{it}^{\gamma}\tilde{V}_{it-l}+U_{it-l}^{\gamma}\tilde{V}_{it}]+o_p(1)\\&=\frac{1}{n}\sum_{i=1}^{n}\sum_{l=-\infty}^{\infty}E[U_{it}^{\gamma}\tilde{V}_{it-l}]+o_p(1).\end{aligned}\tag{A9}$$

Combining Equations (A8) and (A9), we obtain

$$S_n(\theta_0)=-\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}E[U_{it}^{\gamma\gamma}]\operatorname{vec}((E[V_{it}^{\gamma}])^{-1})+\frac{1}{n}\sum_{i=1}^{n}\sum_{l=-\infty}^{\infty}E[U_{it}^{\gamma}\tilde{V}_{it-l}]+o_p(1).\tag{A10}$$

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

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

Appendix D. Proof of Theorem 5

Using the same argument as in the proofs for Theorems 6 and 7, it can be shown that $B_n(\theta)$ as defined in Equation (23) satisfies Conditions 1 and 2. We therefore only establish Condition 3.

Our proof is based on the characterization

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

Proceeding as in Appendix C, we can obtain that

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