KU LEUVEN

DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES
DPS16.07

MAY 2016

Likelihood-based inference for nonlinear models with both individual and time effects

Yutao SUN

Econometrics

Faculty of Economics
and Business

Likelihood-based inference for nonlinear models with both individual and time effects

Yutao Sun*
KU Leuven

Abstract

We propose a bias correction method for nonlinear models with both individual and time effects. Under the presence of the incidental parameter problem, the maximum likelihood estimator derived from such models may be severely biased. Our method produces an approximation to an infeasible log-likelihood function that is not exposed to the incidental parameter problem. The maximizer derived from the approximating function serves as a bias-corrected estimator that is asymptotically unbiased when the sequence N/T converges to a constant. The proposed method is general in several perspectives. The method can be extended to models with multiple fixed effects and can be easily modified to accommodate dynamic models.

Keywords: Incidental parameter problem, maximum likelihood, asymptotic bias correction.

1 Introduction

In many panel applications, researchers would like to incorporate heterogeneities that are individual- and time-dependent. When such heterogeneities are correlated with the covariates of the model, a fixed-effect model including both individual and time effects is usually needed. However, certain class of nonlinear fixed-effect models would produce a severely biased estimate of the parameter that is associated with the covariate (also known as the common parameter). This is the incidental parameter problem (IPP) of Neyman and Scott (1948). For models with only individual effects, Lancaster (2000) and Arellano and Hahn (2005) provide extensive reviews. To briefly introduce the problem, consider the density (conditional on covariates) $f(Y_{it};\theta,a_i)$ where Y_{it} is a scalar outcome of the i,tth observation with $i=1,\cdots,N$ indexing the individuals and $t=1,\cdots,T$ indexing the time periods, a_i is the individual-effect parameter, and θ is the common parameter. Under $N\to\infty$ with T fixed, the maximum likelihood (ML) estimator of a_i, $\widehat{\alpha}_i$, remains to be a random variable. The log-likelihood adopts this randomness in the sense that $\widehat{\theta}$, the ML estimator of θ, converges to an incorrect probability limit that is different from θ_0, the true value of θ. When T increases with N, the random variation in $\widehat{\alpha}_i$ vanishes only slowly. In that case, the asymptotic distribution of $\widehat{\theta}-\theta_0$ contains a bias depending on the relative rate at which $N,T\to\infty$ (Hahn and Newey, 2004). There is a substantial body of literature addressing the

*Address: Research Center of Econometrics, Naamsestraat 69 - box 3565, 3000 Leuven, Belgium. Tel.: +32 16 37 62 75. Email: yutao.sun@kuleuven.be.

IPP when only individual effects are present. For example, Cox and Reid (1987), Lancaster (2002), and Moreira (2008) consider certain reparameterizations producing a ML estimator of θ that is invariant to the individual effects; Fernández-Val (2009) considers the estimation for the fixed-effect probit model, while Greene (2004) considers the Tobit model; Hahn and Newey (2004) and Dhaene and Jochmans (2015) propose bias correction methods based on the jackknife; while Arellano and Hahn (2006) and Arellano and Bonhomme (2009) introduce correction techniques in which a bias-corrected estimate is obtained by maximizing a modified objective function. Other related works include, e.g., Honoré (1993), Hsiao et al. (2002), Alvarez and Arellano (2003), and Hahn and Kuersteiner (2011).

When the model contains both individual and time effects, the situation is much more severe. Consider the density $f(Y_{it}; \theta, a_i, c_t)$ where c_t is the additional time-effect parameter. When N increases with T fixed (or vise-verse), no consistent estimator of θ may be constructed in general. When N and T grow at the same rate, the variations in $\widehat{\alpha}_i$ and $\widehat{\gamma}_t$ induce a bias in $\widehat{\theta}$, which is of the same order as the variance such that the asymptotic distribution of $\widehat{\theta} - \theta_0$ is not centered at 0. In many cases, this bias is nonnegligible to the point that a bias correction technique must be considered. The literature related to this type of model is relatively sparse. For example, Charbonneau (2014) considers binary response models with multiple fixed effects; Okui (2010) studies the estimation of the autocovariance and the autocorrelation; Bai (2009) and Chen et al. (2014) study fixed-effect models in which the individual and time effects enter interactively. For models with both effects, certain correction techniques that apply to the single-effect model (a model with individual effects only) may be generalized to accommodate the two sets of effects. For instance, the recent work of Fernández-Val and Weidner (2016) introduces a split-panel jackknife, similar to Dhaene and Jochmans (2015), that incorporates both individual and time effects. In addition, they also derive a technique that can be used to construct a bias-corrected estimate of θ. Their correction technique is implemented on the parameter level similar to Hahn and Newey (2004), i.e., they provide formulas for b and d such that $\widehat{\theta} - \theta_0 = b/T + d/N + o_p\left(T^{-1}\right) + o_p\left(N^{-1}\right)$. Alternative to this, the contribution of our paper to the literature is that we extend the method proposed by Arellano and Hahn (2006) to accommodate models with both individual and time effect (two-effect models). Arellano and Hahn (2006) introduce an approximating log-likelihood function, accurate to the order of $o_p\left(T^{-1}\right)$ (in the single-effect case), to an infeasible log-likelihood that is immune to the IPP. When the approximating function is maximized, the resulting maximizer constitutes a bias-corrected estimate that is unbiased to the order of $O_p\left(T^{-1}\right)$. We generalize their approach to derive an approximating log-likelihood function that is accurate to the order of, in the context of a two-effect model, $o_p\left(T^{-1}\right) + o_p\left(N^{-1}\right)$. Our approach is slightly simpler than Fernández-Val and Weidner (2016) in the sense that we do not require the calculation of the third derivative of the log-likelihood. Our approach is general in the sense that we do not require the fixed-effect parameters to be additive. We focus only on cases where Y_{it} is independent across i and t. We do, however, briefly discuss how dynamic models and models with multiple fixed effects can be treated within the context of our approach.

The rest of the paper is organized into the following sections. Section 2 presents a detailed introduction of the IPP in the context of a two-effect model. In this section, we show that the log-likelihood function possesses an asymptotic bias. Section 3 derives the bias-corrected log-likelihood function for static models containing both individual and time effect. We provide a preliminary discussions on dynamic models in this section. Also in this section, we present the corrected log-likelihood function for models with a general number of fixed effects and explicitly derive the corrected log-likelihood function for models with 3 sets of fixed effects. Section 4 contains several examples of the application of the corrected log-likelihood. We impose the correction on two modified versions of the variance model of Neyman and Scott (1948) that include, respectively, 2 and 3 sets of fixed effects. Additionally, we present suggestive simulation studies on the static logit and the static probit model.

2 Incidental Parameter Problem with Both Individual and Time Effects

We consider a dataset containing a scalar outcome Y_{it} (conditional on certain covariates) where $i = 1, \cdots, N$ and $t = 1, \cdots, T$ for some positive integer N and T. We focus on cases that are static, i.e., Y_{it} is assumed to be independent across i and t. In addition, Y_{it} is assumed to be governed by a distribution with a smooth density $f(Y_{it}; \theta, a_i, c_t)$ that is known up to values for θ, a_i, and c_t, where a_i is a scalar individual-effect parameter that depends only on the ith individual, c_t is a scalar time-effect parameter that depends only on the tth time period, and θ is a vector of parameters of interest that is the same for all i, t. Our specification is similar to the recent paper by Fernández-Val and Weidner (2016) except that we focus only on independent data. We will, however, discuss the way to incorporate dynamic data in section 3.2. In addition, we will discuss the accommodation of models with more than two sets of fixed effects in section 3.3.

In the context of ML and for an arbitrarily given θ, estimators for a_i and c_t can be constructed as

$$\widehat{\alpha}_1(\theta), \cdots, \widehat{\alpha}_N(\theta), \widehat{\gamma}_1(\theta), \cdots, \widehat{\gamma}_T(\theta) \equiv \arg \max_{a_1, \cdots, a_N, c_1, \cdots, c_T} \frac{1}{NT} \sum_{it} \log f(Y_{it}; \theta, a_i, c_t).$$

Note that $\widehat{\alpha}_i(\theta)$ and $\widehat{\gamma}_t(\theta)$ are assumed to be unique and finite, and to be interior to their corresponding parameter space, which is compact. Subsequently, $\widehat{\theta}$, the ML estimator for θ, can be obtained as

$$\widehat{\theta} \equiv \arg \max_{\theta} \frac{1}{NT} \sum_{it} \log f(Y_{it}; \theta, \widehat{\alpha}_i(\theta), \widehat{\gamma}_t(\theta)).$$

For many models such as probit and logit, when $N \to \infty$ with T fixed, $\widehat{\theta}$ is inconsistent, i.e., assuming the expectation exists,

$$\begin{aligned} \operatorname*{plim}_{N\to\infty} \widehat{\theta} &= \theta_T \equiv \arg \max_{\theta} \operatorname*{plim}_{N\to\infty} \frac{1}{NT} \sum_{it} \log f(Y_{it}; \theta, \widehat{\alpha}_i(\theta), \widehat{\gamma}_t(\theta)) \\ &\neq \theta_0 \equiv \arg \max_{\theta} \frac{1}{NT} \sum_{it} \mathbb{E} \log f(Y_{it}; \theta, \alpha_i(\theta), \gamma_t(\theta)) \end{aligned}$$

where

$$\alpha_1(\theta), \cdots, \alpha_N(\theta), \gamma_1(\theta), \cdots, \gamma_T(\theta) \equiv \arg \max_{a_1, \cdots, a_N, c_1, \cdots, c_T} \frac{1}{NT} \sum_{it} \mathbb{E} \log f(Y_{it}; \theta, a_i, c_t)$$

and $\mathbb{E}(\cdot)$ denotes the expectation computed under the true density $f(\cdot; \theta_0, \alpha_i(\theta_0), \gamma_t(\theta_0))$. When N is fixed and $T \to \infty$, a similar result holds for $\widehat{\theta}$. In both cases, no consistent estimator of θ could be constructed in general. On the other hand, when $N/T \to \kappa$ as $N, T \longrightarrow \infty$ where $0 < \kappa < \infty$, $\widehat{\theta}$ is generally consistent, i.e.,

$$\widehat{\theta} \xrightarrow{p} \theta_0.$$

However, the asymptotic distribution of $\sqrt{NT}\left(\widehat{\theta} - \theta_0\right)$ contains a bias in the sense that the distribution is not centered at 0. The presence of this is due to the fact that

$$\frac{1}{NT} \sum_{it} \log f\left(Y_{it}; \theta, \widehat{\alpha}_i(\theta), \widehat{\gamma}_t(\theta)\right)$$

consists of an asymptotic bias away from the infeasible log-likelihood

$$\frac{1}{NT} \sum_{it} \log f\left(Y_{it}; \theta, \alpha_i(\theta), \gamma_t(\theta)\right).$$

To see this, write $\nabla_{a_i} \log f(\cdot)$ and $\nabla_{c_t} \log f(\cdot)$ for the first derivatives, evaluated at $\alpha_i(\theta)$ and $\gamma_t(\theta)$, of $\log f(Y_{it}; \theta, a_i, c_t)$ w.r.t. a_i and c_t respectively. Consider an expansion of $1/\sqrt{NT} \sum_{it} \log f\left(Y_{it}; \theta, \widehat{\alpha}_i(\theta), \widehat{\gamma}_t(\theta)\right)$,

$$\begin{aligned}
& \frac{1}{\sqrt{NT}} \sum_{it} \log f\left(Y_{it}; \theta, \widehat{\alpha}_i(\theta), \widehat{\gamma}_t(\theta)\right) \\
\approx\; & \frac{1}{\sqrt{NT}} \sum_{it} \log f\left(Y_{it}; \theta, \alpha_i(\theta), \gamma_t(\theta)\right) \\
& + \frac{1}{N} \sum_i \left[\left(\frac{1}{T} \sum_t \nabla_{a_i} \log f\left(Y_{it}; \theta, \alpha_i(\theta), \gamma_t(\theta)\right)\right) \sqrt{NT}\left(\widehat{\alpha}_i(\theta) - \alpha_i(\theta)\right)\right] \\
& + \frac{1}{T} \sum_t \left[\left(\frac{1}{N} \sum_i \nabla_{c_t} \log f\left(Y_{it}; \theta, \alpha_i(\theta), \gamma_t(\theta)\right)\right) \sqrt{NT}\left(\widehat{\gamma}_t(\theta) - \gamma_t(\theta)\right)\right]
\end{aligned}$$

where, as

$$\begin{aligned}
\frac{1}{T} \sum_t \nabla_{a_i} \log f\left(Y_{it}; \theta, \alpha_i(\theta), \gamma_t(\theta)\right) &= O_p\left(T^{-\frac{1}{2}}\right), \\
\frac{1}{N} \sum_i \nabla_{c_t} \log f\left(Y_{it}; \theta, \alpha_i(\theta), \gamma_t(\theta)\right) &= O_p\left(N^{-\frac{1}{2}}\right),
\end{aligned}$$

$$N/T \to \kappa, \qquad \widehat{\alpha}_i(\theta) - \alpha_i(\theta) = O_p\left(T^{-\frac{1}{2}}\right), \qquad \widehat{\gamma}_t(\theta) - \gamma_t(\theta) = O_p\left(N^{-\frac{1}{2}}\right);$$

it follows that

$$\frac{1}{\sqrt{NT}} \sum_{it} \log f\left(Y_{it}; \theta, \widehat{\alpha}_i(\theta), \widehat{\gamma}_t(\theta)\right)$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{NT}} \sum_{it} \log f\left(Y_{it}; \theta, \alpha_i(\theta), \gamma_t(\theta)\right) + \sqrt{NT} O_p\left(T^{-1}\right) + \sqrt{NT} O_p\left(N^{-1}\right) \\
&= \frac{1}{\sqrt{NT}} \sum_{it} \log f\left(Y_{it}; \theta, \alpha_i(\theta), \gamma_t(\theta)\right) + \left(\sqrt{\kappa} + \frac{1}{\sqrt{\kappa}}\right) O_p(1)
\end{aligned}
$$

such that, assuming that the stochastic order operator and the expectation can be interchanged,

$$
\begin{aligned}
& \frac{1}{\sqrt{NT}} \sum_{it} \mathbb{E} \log f\left(Y_{it}; \theta, \widehat{\alpha}_i(\theta), \widehat{\gamma}_t(\theta)\right) \\
= & \frac{1}{\sqrt{NT}} \sum_{it} \mathbb{E} \log f\left(Y_{it}; \theta, \alpha_i(\theta), \gamma_t(\theta)\right) + \left(\sqrt{\kappa} + \frac{1}{\sqrt{\kappa}}\right) O(1).
\end{aligned}
$$

On the other hand, the log-likelihood $1/NT \sum_{it} \log f\left(Y_{it}; \theta, \alpha_i(\theta), \gamma_t(\theta)\right)$ is not exposed to the IPP and hence, may be thought of as an infeasible target function to which an approximation

$$
\begin{aligned}
& \frac{1}{NT} \sum_{it} \log f\left(Y_{it}; \theta, \alpha_i(\theta), \gamma_t(\theta)\right) \\
= & \frac{1}{NT} \sum_{it} \log f\left(Y_{it}; \theta, \widehat{\alpha}_i(\theta), \widehat{\gamma}_t(\theta)\right) + \frac{B(\theta)}{T} + \frac{D(\theta)}{N} + o_p\left(T^{-1}\right) + o_p\left(N^{-1}\right)
\end{aligned}
$$

may be constructed for some $B(\theta)$ and $D(\theta)$ evaluated at $\alpha_i(\theta)$ and $\gamma_t(\theta)$. We will present the exact derivation of $B(\theta)$ and $D(\theta)$ in section 3. The approximating log-likelihood function is asymptotically unbiased, i.e.,

$$
\begin{aligned}
& \frac{1}{\sqrt{NT}} \sum_{it} \log f\left(Y_{it}; \theta, \widehat{\alpha}_i(\theta), \widehat{\gamma}_t(\theta)\right) + \sqrt{NT} \frac{B(\theta)}{T} + \sqrt{NT} \frac{D(\theta)}{N} \\
= & \frac{1}{\sqrt{NT}} \sum_{it} \log f\left(Y_{it}; \theta, \alpha_i(\theta), \gamma_t(\theta)\right) + \sqrt{NT} o_p\left(T^{-1}\right) + \sqrt{NT} o_p\left(N^{-1}\right) \\
= & \frac{1}{\sqrt{NT}} \sum_{it} \log f\left(Y_{it}; \theta, \alpha_i(\theta), \gamma_t(\theta)\right) + \left(\sqrt{\kappa} + \frac{1}{\sqrt{\kappa}}\right) o(1).
\end{aligned}
$$

It then follows that, as

$$
\begin{aligned}
& \frac{1}{NT} \sum_{it} \log f\left(Y_{it}; \theta, \widehat{\alpha}_i(\theta), \widehat{\gamma}_t(\theta)\right) + \frac{B(\theta)}{T} + \frac{D(\theta)}{N} \\
= & \frac{1}{NT} \sum_{it} \log f\left(Y_{it}; \theta, \widehat{\alpha}_i(\theta), \widehat{\gamma}_t(\theta)\right) + \frac{\widehat{B}(\theta)}{T} + \frac{\widehat{D}(\theta)}{N} + o_p\left(T^{-1}\right) + o_p\left(N^{-1}\right)
\end{aligned}
$$

where $\widehat{B}(\theta)$ and $\widehat{D}(\theta)$ are, respectively, $B(\theta)$ and $D(\theta)$ evaluated at $\widehat{\alpha}_i(\theta)$ and $\widehat{\gamma}_t(\theta)$;

$$
\widetilde{\theta} \equiv \arg\max_{\theta} \left(\frac{1}{NT} \sum_{it} \log f\left(Y_{it}; \theta, \widehat{\alpha}_i(\theta), \widehat{\gamma}_t(\theta)\right) + \frac{\widehat{B}(\theta)}{T} + \frac{\widehat{D}(\theta)}{N} \right) \tag{2.1}
$$

may serve as a bias-corrected estimator of θ, satisfying, when the asymptotic sequence $N/T \to \kappa$ as $N, T \longrightarrow \infty$,

$$
\sqrt{NT}\left(\widetilde{\theta} - \theta_0\right) \to_d \mathcal{N}(0, \Sigma)
$$

where $\mathcal{N}(0,\Sigma)$ is the normal distribution with mean zero and covariance matrix Σ being the standard ML asymptotic variance. Here an important point to be observed is that, when N and T are small, the maximizer of the infeasible log-likelihood function,

$$\bar{\theta} \equiv \arg\max_{\theta} \frac{1}{NT} \sum_{it} \log f\left(Y_{it}; \theta, \alpha_i(\theta), \gamma_t(\theta)\right)$$

still can be slightly biased, typically to the order of $O_p(1/NT)$. This is due to the fact that the model is nonlinear in θ - see Box (1971) for details.

3 Correcting the Objective Function

3.1 Static Model with Individual and Time Effects

Let

$$c \equiv (c_1, \cdots, c_T), \qquad \widehat{\gamma} \equiv (\widehat{\gamma}_1(\theta), \cdots, \widehat{\gamma}_T(\theta)), \qquad \alpha_i \equiv \alpha_i(\theta), \qquad \widehat{\alpha}_i \equiv \widehat{\alpha}_i(\theta),$$
$$l_i(a_i, c) \equiv \frac{1}{T}\sum_t \log f(Y_{it}; \theta, a_i, c_t), \qquad l_i^{(r)}(a_i, c) \equiv \frac{1}{T}\sum_t \nabla^r_{a_i} \log f(Y_{it}; \theta, a_i, c_t).$$

Note that we write $\nabla^r_{a_i} \log f(Y_{it}; \theta, \widetilde{a}_i, \widetilde{c}_t)$ for $\nabla^r_{a_i} \log f(Y_{it}; \theta, a_i, c_t)$ evaluated as some specific parameter values $a_i = \widetilde{a}_i$ and $c_t = \widetilde{c}_t$. As similar to Cox and Snell (1968), $l_i^{(1)}(\widehat{\alpha}_i, \widehat{\gamma}) = 0$ and hence can be expanded in a_i around α_i,

$$\begin{aligned} 0 &= l_i^{(1)}(\alpha_i, \widehat{\gamma}) + l_i^{(2)}(\alpha_i, \widehat{\gamma})(\widehat{\alpha}_i - \alpha_i) + o_p\left(T^{-\frac{1}{2}}\right) \\ 0 &= l_i^{(1)}(\alpha_i, \widehat{\gamma}) + \mathbb{E}l_i^{(2)}(\alpha_i, \widehat{\gamma})(\widehat{\alpha}_i - \alpha_i) + o_p\left(T^{-\frac{1}{2}}\right) \end{aligned}$$

where, as $(\widehat{\alpha}_i - \alpha_i) = O_p\left(T^{-1/2}\right)$, replacing $l_i^{(2)}(\alpha_i, \widehat{\gamma})$ with $\mathbb{E}l_i^{(2)}(\alpha_i, \widehat{\gamma})$ generates a bias to the negligible order of $o_p\left(T^{-1/2}\right)$. Next,

$$(\widehat{\alpha}_i - \alpha_i) = -\frac{l_i^{(1)}(\alpha_i, \widehat{\gamma})}{\mathbb{E}l_i^{(2)}(\alpha_i, \widehat{\gamma})} + o_p\left(T^{-\frac{1}{2}}\right) \tag{3.1}$$

where, for regular circumstances, $\mathbb{E}l_i^{(2)}(\alpha_i, \widehat{\gamma}) < 0$ such that equation (3.1) is well-defined. Similarly, for an arbitrarily given c, $l_i(\widehat{\alpha}_i, c)$ can also be expanded in a_i around α_i,

$$\begin{aligned} l_i(\widehat{\alpha}_i, c) &= l_i(\alpha_i, c) + l_i^{(1)}(\alpha_i, c)(\widehat{\alpha}_i - \alpha_i) + \frac{1}{2}\mathbb{E}l_i^{(2)}(\alpha_i, c)(\widehat{\alpha}_i - \alpha_i)^2 + o_p\left(T^{-1}\right) \\ l_i(\alpha_i, c) &= l_i(\widehat{\alpha}_i, c) - l_i^{(1)}(\alpha_i, c)(\widehat{\alpha}_i - \alpha_i) - \frac{1}{2}\mathbb{E}l_i^{(2)}(\alpha_i, c)(\widehat{\alpha}_i - \alpha_i)^2 \\ &\quad + o_p\left(T^{-1}\right) \end{aligned} \tag{3.2}$$

in which, similarly to the above, replacing $l_i^{(2)}(\alpha_i, \widehat{\gamma})$ with $\mathbb{E}l_i^{(2)}(\alpha_i, \widehat{\gamma})$ induces a bias to the negligible order of $o_p\left(T^{-1}\right)$. Noticing $l_i^{(1)}(\alpha_i, c) = O_p\left(T^{-1/2}\right)$, combine equation (3.1) and (3.2),

$$l_i(\alpha_i, c)$$

$$
\begin{aligned}
&= l_i\left(\widehat{\alpha}_i, c\right) - l_i^{(1)}\left(\alpha_i, c\right)\left(-\frac{l_i^{(1)}\left(\alpha_i, \widehat{\gamma}\right)}{\mathbb{E}l_i^{(2)}\left(\alpha_i, \widehat{\gamma}\right)}\right) - \frac{1}{2}\mathbb{E}l_i^{(2)}\left(\alpha_i, c\right)\left(-\frac{l_i^{(1)}\left(\alpha_i, \widehat{\gamma}\right)}{\mathbb{E}l_i^{(2)}\left(\alpha_i, \widehat{\gamma}\right)}\right)^2 \\
&\quad + o_p\left(T^{-1}\right).
\end{aligned}
$$

Here, by the definition of $l_i^{(1)}\left(a_i, c\right)$, it is clear that

$$
\begin{aligned}
\left(l_i^{(1)}\left(a_i, c\right)\right)^2 &= \frac{1}{T^2}\sum_t \left[\nabla_{a_i} \log f\left(Y_{it}; \theta, a_i, c_t\right)\right]^2 \\
&\quad + \frac{1}{T^2}\sum_{t \neq t'} \nabla_{a_i} \log f\left(Y_{it}; \theta, a_i, c_t\right) \nabla_{a_i} \log f\left(Y_{it'}; \theta, a_i, c_{t'}\right)
\end{aligned}
$$

such that

$$
\begin{aligned}
& l_i\left(\alpha_i, c\right) \\
= \; & l_i\left(\widehat{\alpha}_i, c\right) \\
& + \frac{1}{T}\frac{1/T \sum_t \nabla_{a_i} \log f\left(Y_{it}; \theta, \alpha_i, c_t\right) \nabla_{a_i} \log f\left(Y_{it}; \theta, \alpha_i, \widehat{\gamma}_t\right)}{\mathbb{E}l_i^{(2)}\left(\alpha_i, \widehat{\gamma}\right)} \\
& + \frac{1}{T}\frac{1/T \sum_{t \neq t'} \nabla_{a_i} \log f\left(Y_{it}; \theta, \alpha_i, c_t\right) \nabla_{a_i} \log f\left(Y_{it'}; \theta, \alpha_i, \widehat{\gamma}_{t'}\right)}{\mathbb{E}l_i^{(2)}\left(\alpha_i, \widehat{\gamma}\right)} \\
& - \frac{1}{2}\frac{1}{T}\frac{1/T \sum_t \left[\nabla_{a_i} \log f\left(Y_{it}; \theta, \alpha_i, \widehat{\gamma}_t\right)\right]^2 \mathbb{E}l_i^{(2)}\left(\alpha_i, c\right)}{\left(\mathbb{E}l_i^{(2)}\left(\alpha_i, \widehat{\gamma}\right)\right)^2} \\
& - \frac{1}{2}\frac{1}{T}\frac{1/T \sum_{t \neq t'} \left[\nabla_{a_i} \log f\left(Y_{it}; \theta, \alpha_i, \widehat{\gamma}_t\right) \nabla_{a_i} \log f\left(Y_{it'}; \theta, \alpha_i, \widehat{\gamma}_{t'}\right)\right] \mathbb{E}l_i^{(2)}\left(\alpha_i, c\right)}{\left(\mathbb{E}l_i^{(2)}\left(\alpha_i, \widehat{\gamma}\right)\right)^2} \\
& + o_p\left(T^{-1}\right)
\end{aligned}
$$

where, by the independency across t,

$$
\begin{aligned}
\mathbb{E}\nabla_{a_i} \log f\left(Y_{it}; \theta, \alpha_i, c_t\right) \nabla_{a_i} \log f\left(Y_{it}; \theta, \alpha_i, \widehat{\gamma}_{t'}\right) = 0, \\
\mathbb{E}\nabla_{a_i} \log f\left(Y_{it}; \theta, \alpha_i, \widehat{\gamma}_t\right) \nabla_{a_i} \log f\left(Y_{it}; \theta, \alpha_i, \widehat{\gamma}_{t'}\right) = 0
\end{aligned}
$$

such that

$$
\begin{aligned}
\mathbb{E}l_i\left(\alpha_i, c\right) &= \mathbb{E}l_i\left(\widehat{\alpha}_i, c\right) + \frac{\mathbb{E}b_i\left(\alpha_i, c\right)}{T} + o\left(T^{-1}\right) \\
&= \mathbb{E}l_i\left(\widehat{\alpha}_i, c\right) + \frac{\mathbb{E}b_i\left(\widehat{\alpha}_i, c\right)}{T} + o\left(T^{-1}\right) \qquad (3.3)
\end{aligned}
$$

with

$$
\begin{aligned}
b_i\left(a_i, c\right) \equiv \; & \frac{1/T \sum_t \nabla_{a_i} \log f\left(Y_{it}; \theta, a_i, c_t\right) \nabla_{a_i} \log f\left(Y_{it}; \theta, a_i, \widehat{\gamma}_t\right)}{\mathbb{E}l_i^{(2)}\left(a_i, \widehat{\gamma}\right)} \\
& - \frac{1}{2}\frac{1/T \sum_t \left[\nabla_{a_i} \log f\left(Y_{it}; \theta, a_i, \widehat{\gamma}_t\right)\right]^2 \mathbb{E}l_i^{(2)}\left(a_i, c\right)}{\left(\mathbb{E}l_i^{(2)}\left(a_i, \widehat{\gamma}\right)\right)^2}.
\end{aligned}
$$

Note that equation (3.3) holds for every c with a slightly embarrassing complication that $b_i(\widehat{\alpha}_i, c)$ depends on $\widehat{\gamma}$. This is because $l_i^{(1)}(a_i, c) = 0$ if and only if $a_i = \widehat{\alpha}_i$ and $c = \widehat{\gamma}$. When evaluated at $\widehat{\gamma}$, $b_i(\widehat{\alpha}_i, c)$ reduces to

$$b_i(\widehat{\alpha}_i, \widehat{\gamma}) = \frac{1}{2} \frac{1/T \sum_t [\nabla_{a_i} \log f(Y_{it}; \theta, \widehat{\alpha}_i, \widehat{\gamma}_t)]^2}{\mathbb{E} l_i^{(2)}(\widehat{\alpha}_i, \widehat{\gamma})}$$

which coincides in structure with the bias term developed by Arellano and Hahn (2006).

In a similar fashion, let

$$a \equiv (a_1, \cdots, a_N), \qquad \widehat{\alpha} \equiv (\widehat{\alpha}_1(\theta), \cdots, \widehat{\alpha}_N(\theta)), \qquad \gamma_t \equiv \gamma_t(\theta), \widehat{\gamma}_t \equiv \widehat{\gamma}_t(\theta),$$
$$l_t(a, c_t) \equiv \frac{1}{N} \sum_i \log f(Y_{it}; \theta, a_i, c_t), \qquad l_t^{(r)}(a, c_t) \equiv \frac{1}{N} \sum_i \nabla_{c_t}^r \log f(Y_{it}; \theta, a_i, c_t).$$

$l_t^{(1)}(\widehat{\alpha}, \widehat{\gamma}_t) = 0$ can be expanded in c_t around γ_t,

$$\begin{aligned} 0 &= l_t^{(1)}(\widehat{\alpha}, \gamma_t) + \mathbb{E} l_t^{(2)}(\widehat{\alpha}, \gamma_t)(\widehat{\gamma}_t - \gamma_t) + o_p\left(N^{-\frac{1}{2}}\right) \\ (\widehat{\gamma}_t - \gamma_t) &= -\frac{l_t^{(1)}(\widehat{\alpha}, \gamma_t)}{\mathbb{E} l_t^{(2)}(\widehat{\alpha}, \gamma_t)} + o_p\left(N^{-\frac{1}{2}}\right) \end{aligned} \tag{3.4}$$

where $\mathbb{E} l_t^{(2)}(\widehat{\alpha}, \gamma_t) < 0$ such that equation (3.4) is well-defined. Next, for an arbitrarily given a, expand $l_t(a, \widehat{\gamma}_t)$ in c_t around γ_t,

$$\begin{aligned} l_t(a, \widehat{\gamma}_t) &= l_t(a, \gamma_t) + l_t^{(1)}(a, \gamma_t)(\widehat{\gamma}_t - \gamma_t) + \frac{1}{2} \mathbb{E} l_t^{(2)}(a, \gamma_t)(\widehat{\gamma}_t - \gamma_t)^2 + o_p\left(N^{-1}\right) \\ l_t(a, \gamma_t) &= l_t(a, \widehat{\gamma}_t) - l_t^{(1)}(a, \gamma_t)(\widehat{\gamma}_t - \gamma_t) - \frac{1}{2} \mathbb{E} l_t^{(2)}(a, \gamma_t)(\widehat{\gamma}_t - \gamma_t)^2 \\ &\quad + o_p\left(N^{-1}\right) \end{aligned} \tag{3.5}$$

such that a combination of equation (3.4) and (3.5) gives,

$$\begin{aligned} & l_t(a, \gamma_t) \\ = \; & l_t(a, \widehat{\gamma}_t) + \frac{l_t^{(1)}(a, \gamma_t) l_t^{(1)}(\widehat{\alpha}, \gamma_t)}{\mathbb{E} l_t^{(2)}(\widehat{\alpha}, \gamma_t)} - \frac{1}{2} \frac{\left(l_t^{(1)}(\widehat{\alpha}, \gamma_t)\right)^2 \mathbb{E} l_t^{(2)}(a, \gamma_t)}{\left(\mathbb{E} l_t^{(2)}(\widehat{\alpha}, \gamma_t)\right)^2} + o_p\left(N^{-1}\right) \\ = \; & l_t(\alpha, \widehat{\gamma}_t) \\ & + \frac{1}{N} \frac{1/N \sum_i \nabla_{c_t} \log f(Y_{it}; \theta, a_i, \gamma_t) \nabla_{c_t} \log f(Y_{it}; \theta, \widehat{\alpha}_i, \gamma_t)}{\mathbb{E} l_t^{(2)}(\widehat{\alpha}, \gamma_t)} \\ & + \frac{1}{N} \frac{1/N \sum_{i \neq i'} \nabla_{c_t} \log f(Y_{it}; \theta, a_i, \gamma_t) \nabla_{c_t} \log f(Y_{i't}; \theta, \widehat{\alpha}_{i'}, \gamma_t)}{\mathbb{E} l_t^{(2)}(\widehat{\alpha}, \gamma_t)} \\ & - \frac{1}{2} \frac{1}{N} \frac{1/N \sum_i [\nabla_{c_t} \log f(Y_{it}; \theta, \widehat{\alpha}_i, \gamma_t)]^2 \mathbb{E} l_t(a, \gamma_t)}{\left(\mathbb{E} l_t^{(2)}(\widehat{\alpha}, \gamma_t)\right)^2} \\ & - \frac{1}{2} \frac{1}{N} \frac{1/N \sum_{i \neq i'} [\nabla_{c_t} \log f(Y_{it}; \theta, \widehat{\alpha}_i, \gamma_t) \nabla_{c_t} \log f(Y_{i't}; \theta, \widehat{\alpha}_{i'}, \gamma_t)] \mathbb{E} l_t^{(2)}(a, \gamma_t)}{\left(\mathbb{E} l_t^{(2)}(\widehat{\alpha}, \gamma_t)\right)^2} \\ & + o_p\left(N^{-1}\right) \end{aligned}$$

where, because of the lack of spatial dependency,

$$\mathbb{E}\nabla_{c_t}\log f\left(Y_{it};\theta,a_i,\gamma_t\right)\nabla_{c_t}\log f\left(Y_{i't};\theta,\widehat{\alpha}_{i'},\gamma_t\right)=0,$$
$$\mathbb{E}\nabla_{c_t}\log f\left(Y_{it};\theta,\widehat{\alpha}_i,\gamma_t\right)\nabla_{c_t}\log f\left(Y_{i't};\theta,\widehat{\alpha}_{i'},\gamma_t\right)=0$$

such that

$$\begin{aligned}\mathbb{E}l_t\left(a,\gamma_t\right) &= \mathbb{E}l_t\left(a,\widehat{\gamma}_t\right)+\frac{\mathbb{E}d_t\left(a,\gamma_t\right)}{N}+o\left(N^{-1}\right)\\ &= \mathbb{E}l_t\left(a,\widehat{\gamma}_t\right)+\frac{\mathbb{E}d_t\left(a,\widehat{\gamma}_t\right)}{N}+o\left(N^{-1}\right)\end{aligned} \tag{3.6}$$

with

$$\begin{aligned}d_t\left(a,c_t\right) \equiv{} & \frac{1/N\sum_i\nabla_{c_t}\log f\left(Y_{it};\theta,a_i,c_t\right)\nabla_{c_t}\log f\left(Y_{it};\theta,\widehat{\alpha}_i,c_t\right)}{\mathbb{E}l_t^{(2)}\left(\widehat{\alpha},c_t\right)}\\ & -\frac{1}{2}\frac{1/N\sum_i\left[\nabla_{c_t}\log f\left(Y_{it};\theta,\widehat{\alpha}_i,c_t\right)\right]^2\mathbb{E}l_t^{(2)}\left(a,c_t\right)}{\left(\mathbb{E}l_t^{(2)}\left(\widehat{\alpha},c_t\right)\right)^2}.\end{aligned}$$

Next, observe that, for every a and c,

$$\frac{1}{T}\sum_t l_t\left(a,c_t\right)\equiv\frac{1}{NT}\sum_{it}\log f\left(Y_{it};\theta,a_i,c_t\right)\equiv\frac{1}{N}\sum_i l_i\left(a_i,c\right), \tag{3.7}$$

i.e., the change of the order of sums does not affect the value of the sum. It follows that averaging equation (3.3) evaluated at γ gives

$$\frac{1}{N}\sum_i\mathbb{E}l_i\left(\alpha_i,\gamma\right)=\frac{1}{N}\sum_i\mathbb{E}l_i\left(\widehat{\alpha}_i,\gamma\right)+\frac{1}{T}\frac{1}{N}\sum_i\mathbb{E}b_i\left(\widehat{\alpha}_i,\gamma\right)+o\left(T^{-1}\right) \tag{3.8}$$

and averaging equation (3.6) evaluated at $\widehat{\alpha}$ gives

$$\frac{1}{T}\sum_t\mathbb{E}l_t\left(\widehat{\alpha},\gamma_t\right)=\frac{1}{T}\sum_t\mathbb{E}l_t\left(\widehat{\alpha},\widehat{\gamma}_t\right)+\frac{1}{N}\frac{1}{T}\sum_t\mathbb{E}d_t\left(\widehat{\alpha},\widehat{\gamma}_t\right)+o\left(N^{-1}\right) \tag{3.9}$$

such that, if equation (3.7) to (3.9) are combined,

$$\begin{aligned}\frac{1}{N}\sum_i\mathbb{E}l_i\left(\alpha_i,\gamma\right) ={} & \frac{1}{T}\sum_t\mathbb{E}l_t\left(\widehat{\alpha},\widehat{\gamma}_t\right)+\frac{1}{N}\frac{1}{T}\sum_t\mathbb{E}d_t\left(\widehat{\alpha},\widehat{\gamma}_t\right)+\frac{1}{T}\frac{1}{N}\sum_i\mathbb{E}b_i\left(\widehat{\alpha}_i,\gamma\right)\\ & +o\left(T^{-1}\right)+o\left(N^{-1}\right)\end{aligned}$$

where, as $\mathbb{E}b_i\left(\widehat{\alpha}_i,\widehat{\gamma}\right)=\mathbb{E}b_i\left(\widehat{\alpha}_i,\gamma\right)+o\left(1\right)$,

$$\mathbb{E}L\left(\theta\right)=\mathbb{E}\widehat{L}\left(\theta\right)+\frac{\mathbb{E}\widehat{B}\left(\theta\right)}{T}+\frac{\mathbb{E}\widehat{D}\left(\theta\right)}{N}+o\left(T^{-1}\right)+o\left(N^{-1}\right) \tag{3.10}$$

with

$$\begin{aligned}L\left(\theta\right) &\equiv \frac{1}{NT}\sum_{it}\log f\left(Y_{it};\theta,\alpha_i,\gamma_t\right), \qquad \widehat{L}\left(\theta\right)\equiv\frac{1}{NT}\sum_{it}\log f\left(Y_{it};\theta,\widehat{\alpha}_i,\widehat{\gamma}_t\right),\\ \widehat{B}\left(\theta\right) &\equiv \frac{1}{N}\sum_i b_i\left(\widehat{\alpha}_i,\widehat{\gamma}\right)=\frac{1}{2}\frac{1}{N}\sum_i\frac{1/T\sum_t\left[\nabla_{\alpha_i}\log f\left(Y_{it};\theta,\widehat{\alpha}_i,\widehat{\gamma}_t\right)\right]^2}{\mathbb{E}l_i^{(2)}\left(\widehat{\alpha}_i,\widehat{\gamma}\right)},\end{aligned}$$

$$\widehat{D}(\theta) \quad \equiv \quad \frac{1}{T}\sum\nolimits_t d_t(\widehat{\alpha},\widehat{\gamma}_t) = \frac{1}{2}\frac{1}{T}\sum\nolimits_t \frac{1/N\sum_i \left[\nabla_{c_t}\log f(Y_{it};\theta,\widehat{\alpha}_i,\widehat{\gamma}_t)\right]^2}{\mathbb{E}l_t^{(2)}(\widehat{\alpha},\widehat{\gamma}_t)}.$$

The corrected log-likelihood can then be constructed as

$$\widetilde{L}(\theta) \equiv \widehat{L}(\theta) + \frac{\widehat{B}(\theta)}{T} + \frac{\widehat{D}(\theta)}{N} \tag{3.11}$$

in which the right-hand side only depends on Y_{it}, the given θ, $\widehat{\alpha}_i$, and $\widehat{\gamma}_t$ and hence, can be constructed in a straightforward way from the data. Here notice that $B(\theta)$ and $D(\theta)$ are symmetric in structure. This reflects the fact that a_i and c_t are interchangeable, which, given their specification, is obvious. Also, when, for instance, c_t disappears, the corresponding $D(\theta)$ drops from equation (3.10) whereas $B(\theta)$ remains unaffected.

3.2 Dynamic Model

When Y_{it} are dynamic, a slight modification to equation (3.10) must be adopted. Such modification is essentially an implementation of the optimal weights introduced by Arellano and Hahn (2006) into the quantities $B(\theta)$ and $D(\theta)$. For this reason, we will only briefly introduce the modification.

Suppose first that Y_{it} are correlated across t but are independent across i. In this case, $D(\theta)$ can be kept intact whereas $b_i(a_i, c)$ should be modified. More specifically,

$$\begin{aligned} & b_i(a_i,c) \\ \equiv \quad & \frac{1}{2}\frac{1/T\sum_t\left[\nabla_{a_i}\log f(Y_{it};\theta,a_i,c_t)\right]^2}{\mathbb{E}l_i^{(2)}(a_i,c)} \\ & +\frac{1}{2}\sum_{\tau=-m,\tau\neq 0}^{m}\frac{\frac{1}{T}\sum_{t=\max(1,\tau+1)}^{\min(T,T+\tau)} w_\tau \nabla_{a_i}\log f(Y_{it};\cdot,c_t)\nabla_{a_i}\log f(Y_{it-\tau};\cdot,c_{t-\tau})}{\mathbb{E}l_i^{(2)}(a_i,c)} \end{aligned}$$

where $w_\tau \equiv 1 - \tau/(m+1)$ (the Bartlett kernel weight) and m may be chosen according to the dynamic, across t, of Y_{it}. In addition, multiple choices of the weight are available - see, e.g., Fernández-Val and Weidner (2016), Hahn et al. (2007), and Hahn and Kuersteiner (2011). For a static model, $w_m = 0$ such that the second term in $b_i(a_i, c)$ drops out. When Y_{it} are correlated across i, a similar modification of $d_t(a, c_t)$ is necessary.

3.3 Multiple Fixed Effects

Models with additional effects can also be treated in a similar fashion. Suppose $j = 1, \cdots, J$ for an arbitrarily fixed positive integer J and consider the density

$$f\left(Y_{i_1\cdots i_J};\theta,g_{i_1}^{(1)},\cdots,g_{i_J}^{(J)}\right),\qquad i_j = 1,\cdots,N_j,\qquad N_j\in\mathbb{N}$$

where $Y_{i_1\cdots i_J}$ are independent across $i_1, \cdots, i_J$, while $g_{i_j}^{(j)}$ is the i_jth fixed-effect parameter belonging to the jth set of fixed effects and θ is the parameter that applies to all $Y_{i_1\cdots i_J}$.

Let

$$
\begin{aligned}
\widehat{\eta} &\equiv \widehat{\eta}_1^{(1)}(\theta), \cdots, \widehat{\eta}_{N_1}^{(1)}(\theta), \cdots, \widehat{\eta}_1^{(J)}(\theta), \cdots, \widehat{\eta}_{N_J}^{(J)}(\theta) \\
&\equiv \arg \max_{g_1^{(1)}, \cdots, g_{N_1}^{(1)}, \cdots, g_1^{(J)}, \cdots, g_{N_J}^{(J)}} \frac{1}{\prod_j N_j} \sum_{i_1 \cdots i_J} \log f\left(Y_{i_1 \cdots i_J}; \theta, g_{i_1}^{(1)}, \cdots, g_{i_J}^{(J)}\right), \\
\eta &\equiv \eta_1^{(1)}(\theta), \cdots, \eta_{N_1}^{(1)}(\theta), \cdots, \eta_1^{(J)}(\theta), \cdots, \eta_{N_J}^{(J)}(\theta) \\
&\equiv \arg \max_{g_1^{(1)}, \cdots, g_{N_1}^{(1)}, \cdots, g_1^{(J)}, \cdots, g_{N_J}^{(J)}} \frac{1}{\prod_j N_j} \sum_{i_1 \cdots i_J} \mathbb{E} \log f\left(Y_{i_1 \cdots i_J}; \theta, g_{i_1}^{(1)}, \cdots, g_{i_J}^{(J)}\right).
\end{aligned}
$$

It follows that, after a similar derivation,

$$
\mathbb{E}L_J(\theta) = \mathbb{E}\widehat{L}_J(\theta) + \sum_j \frac{\mathbb{E}\widehat{K}_j(\theta)}{\prod_{s \neq j} N_s} + \sum_j o\left(\prod_{s \neq j} N_s^{-1}\right) \tag{3.12}
$$

where,

$$
L_J(\theta) \equiv \frac{1}{\prod_j N_j} \sum_{i_1 \cdots i_J} \log f\left(Y_{i_1 \cdots i_J}; \theta, \eta\right), \qquad \widehat{L}_J(\theta) \equiv \frac{1}{\prod_j N_j} \sum_{i_1 \cdots i_J} \log f\left(Y_{i_1 \cdots i_J}; \theta, \widehat{\eta}\right),
$$

$$
\widehat{K}_j(\theta) \equiv \frac{1}{2} \frac{1}{N_j} \sum_{i_j} \frac{\sum_{s \neq j} \sum_{i_s} \left[\nabla_{g_{i_j}^{(j)}} \log f\left(Y_{i_1 \cdots i_J}; \theta, \widehat{\eta}\right)\right]^2}{\sum_{s \neq j} \sum_{i_s} \nabla^2_{g_{i_j}^{(j)}} \log f\left(Y_{i_1 \cdots i_J}; \theta, \widehat{\eta}\right)}.
$$

Some condition regulating $N_j \to \infty$ must be enforced for equation (3.12) to hold; i.e., $N_j / N_{j'} \to \kappa_{j,j'}$, where $0 < \kappa_{j,j'} < \infty$, for all $j \neq j'$, all $N_j \to \infty$ at the same speed. In addition, when the model is dynamic, the modification introduced in section 3.2 may be implemented into the corresponding $K_j(\theta)$.

When $J \to \infty$, an additional condition regulating the speed of convergence of J must be imposed such that the reminder term $\sum_j o\left(\prod_{s \neq j} N_s^{-1}\right)$ still vanishes at a desired rate. Suppose $N_j = N$ for every j and some $N \to \infty$,

$$
\begin{aligned}
& \frac{1}{N^{J/2}} \sum_{i_1 \cdots i_J} \log f\left(Y; \theta, \widehat{\eta}\right) + N^{\frac{J}{2}} \sum_j \frac{\widehat{K}_j(\theta)}{N^{J-1}} \\
= & \frac{1}{N^{J/2}} \sum_{i_1 \cdots i_J} \log f\left(Y; \theta, \eta\right) + J N^{\frac{J}{2}} o_p\left(N^{-(J-1)}\right) \\
= & \frac{1}{N^{J/2}} \sum_{i_1 \cdots i_J} \log f\left(Y; \theta, \eta\right) + J N^{-\frac{J-2}{2}} o_p(1)
\end{aligned}
$$

in which $J N^{-\frac{J-2}{2}} o_p(1) = o_p(1)$ if $J N^{-\frac{J-2}{2}} < \infty$, i.e.,

$$
\frac{J}{N^{(J-2)/2}} \to \kappa'
$$

as $N, J \to \infty$ where $\kappa' < \infty$. Under this condition, the asymptotic distribution of

$$\widetilde{\theta} \equiv \arg\max_{\theta} \left(\frac{1}{N^J} \sum_{i_1 \cdots i_J} \log f\left(Y; \theta, \widehat{\eta}\right) + \sum_j \frac{\widehat{K}_j\left(\theta\right)}{N^{J-1}} \right)$$

is recentered at 0.

Next, we briefly derive the corrected log-likelihood function for $J = 3$. When $J = 3$, we have the density

$$f\left(a_i, c_t, g_s\right) \equiv f\left(Y_{its}; \theta, a_i, c_t, g_s\right)$$

where $i = 1, \cdots, N$, $t = 1, \cdots, T$, $s = 1, \cdots, S$, a_i and c_t are defined as above, and g_s is an additional effect. Let

$$\begin{aligned}
&\widehat{\alpha}_1\left(\theta\right), \cdots, \widehat{\alpha}_N\left(\theta\right), \widehat{\gamma}_1\left(\theta\right), \cdots, \widehat{\gamma}_T\left(\theta\right), \widehat{\eta}_1\left(\theta\right), \cdots, \widehat{\eta}_S\left(\theta\right) \\
&\equiv \arg \max_{a_1, \cdots, a_N, c_1, \cdots, c_T, g_1, \cdots, g_S} \frac{1}{NTS} \sum_{its} \log f\left(a_i, c_t, g_s\right), \\
&\alpha_1\left(\theta\right), \cdots, \alpha_N\left(\theta\right), \gamma_1\left(\theta\right), \cdots, \gamma_T\left(\theta\right), \eta_1\left(\theta\right), \cdots, \eta_S\left(\theta\right) \\
&\equiv \arg \max_{a_1, \cdots, a_N, c_1, \cdots, c_T, g_1, \cdots, g_S} \frac{1}{NTS} \sum_{its} \mathbb{E} \log f\left(a_i, c_t, g_s\right), \\
&l_i\left(a_i, c, g\right) \equiv \frac{1}{TS} \sum_{ts} \log f\left(a_i, c_t, g_s\right), \\
&l_i^{(r)}\left(a_i, c, g\right) \equiv \frac{1}{TS} \sum_{ts} \nabla_{a_i}^r \log f\left(a_i, c_t, g_s\right)
\end{aligned}$$

where $g \equiv (g_1, \cdots, g_S)$. Observing $l_i^{(1)}\left(\widehat{\alpha}_i, \widehat{\gamma}, \widehat{\eta}\right) = 0$ where $\widehat{\eta} \equiv \left(\widehat{\eta}_1\left(\theta\right), \cdots, \widehat{\eta}_S\left(\theta\right)\right)$, an expansion of $l_i^{(1)}\left(\widehat{\alpha}_i, \widehat{\gamma}, \widehat{\eta}\right) = 0$ in a_i around α_i gives

$$\begin{aligned}
0 &= l_i^{(1)}\left(\alpha_i, \widehat{\gamma}, \widehat{\eta}\right) + \mathbb{E} l_i^{(2)}\left(\alpha_i, \widehat{\gamma}, \widehat{\eta}\right)\left(\widehat{\alpha}_i - \alpha_i\right) + o_p\left(\frac{1}{\sqrt{TS}}\right) \\
\left(\widehat{\alpha}_i - \widehat{\alpha}_i\right) &= -\frac{l_i^{(1)}\left(\alpha_i, \widehat{\gamma}, \widehat{\eta}\right)}{\mathbb{E} l_i^{(2)}\left(\alpha_i, \widehat{\gamma}, \widehat{\eta}\right)} + o_p\left(\frac{1}{\sqrt{TS}}\right);
\end{aligned}$$

and a similar expansion of $l_i\left(\widehat{\alpha}_i, c, g\right)$ gives

$$l_i\left(\alpha_i, c, g\right) = l_i\left(\widehat{\alpha}_i, c, g\right) - l_i^{(1)}\left(\alpha_i, c, g\right)\left(\widehat{\alpha}_i - \alpha_i\right) - \frac{1}{2} \mathbb{E} l_i^{(2)}\left(\alpha_i, c, g\right)\left(\widehat{\alpha}_i - \alpha_i\right)^2 + o_p\left(\frac{1}{TS}\right).$$

It follows that

$$\begin{aligned}
l_i\left(\alpha_i, c, g\right) &= l_i\left(\widehat{\alpha}_i, c, g\right) - l_i^{(1)}\left(\alpha_i, c, g\right)\left(-\frac{l_i^{(1)}\left(\alpha_i, \widehat{\gamma}, \widehat{\eta}\right)}{\mathbb{E} l_i^{(2)}\left(\alpha_i, \widehat{\gamma}, \widehat{\eta}\right)}\right) \\
&\quad - \frac{1}{2} \mathbb{E} l_i^{(2)}\left(\alpha_i, c, g\right)\left(\frac{l_i^{(1)}\left(\alpha_i, \widehat{\gamma}, \widehat{\eta}\right)}{\mathbb{E} l_i^{(2)}\left(\alpha_i, \widehat{\gamma}, \widehat{\eta}\right)}\right)^2 + o_p\left(\frac{1}{TS}\right)
\end{aligned}$$

such that, as

$$l_i^{(1)}\left(\alpha_i, \widehat{\gamma}, \widehat{\eta}\right) l_i^{(1)}\left(\alpha_i, c, g\right)$$

$$
\begin{aligned}
&= \frac{1}{(TS)^2}\sum_{ts}\nabla_{a_i}\log f(\alpha_i,\widehat{\gamma}_t,\widehat{\eta}_s)\nabla_{a_i}\log f(\alpha_i,c_t,g_s) \\
&\quad+\frac{1}{(TS)^2}\sum_{(t,t',s,s')\in\mathcal{TS}}\nabla_{a_i}\log f(\alpha_i,\widehat{\gamma}_t,\widehat{\eta}_s)\nabla_{a_i}\log f(\alpha_i,c_{t'},g_{s'}), \\
&\quad\left(l_i^{(1)}(\alpha_i,\widehat{\gamma},\widehat{\eta})\right)^2 \\
&= \frac{1}{(TS)^2}\sum_{ts}\left[\nabla_{a_i}\log f(\alpha_i,\widehat{\gamma}_t,\widehat{\eta}_s)\right]^2 \\
&\quad+\frac{1}{(TS)^2}\sum_{(t,t',s,s')\in\mathcal{TS}}\nabla_{a_i}\log f(\alpha_i,\widehat{\gamma}_t,\widehat{\eta}_s)\nabla_{a_i}\log f(\alpha_i,\widehat{\gamma}_{t'},\widehat{\eta}_{s'})
\end{aligned}
$$

in which

$$
\widehat{\eta}_s\equiv\widehat{\eta}_s(\theta),\qquad \mathcal{TS}\equiv\left\{(t,t',s,s')\,|t\neq t'\vee s\neq s';t,t'=1,\cdots,T;s,s'=1,\cdots,S\right\},
$$
$$
\mathbb{E}\nabla_{a_i}\log f(\alpha_i,\widehat{\gamma}_t,\widehat{\eta}_s)\nabla_{a_i}\log f(\alpha_i,c_{t'},g_{s'})=0,
$$
$$
\mathbb{E}\nabla_{a_i}\log f(\alpha_i,\widehat{\gamma}_t,\widehat{\eta}_s)\nabla_{a_i}\log f(\alpha_i,\widehat{\gamma}_{t'},\widehat{\eta}_{s'})=0.
$$

We then have

$$
\begin{aligned}
\mathbb{E}l_i(\alpha_i,c,g) &= \mathbb{E}l_i(\widehat{\alpha}_i,c,g)+\frac{\mathbb{E}b_i(\alpha_i,c,g)}{TS}+o\left(\frac{1}{TS}\right) \\
&= \mathbb{E}l_i(\widehat{\alpha}_i,c,g)+\frac{\mathbb{E}b_i(\widehat{\alpha}_i,c,g)}{TS}+o\left(\frac{1}{TS}\right)
\end{aligned}
$$

where

$$
\begin{aligned}
b_i(a_i,c,g) &\equiv \frac{1/TS\sum_{ts}\nabla_{a_i}\log f(a_i,\widehat{\gamma}_t,\widehat{\eta}_s)\nabla_{a_i}\log f(a_i,c_t,g_s)}{\mathbb{E}l_i^{(2)}(a_i,\widehat{\gamma},\widehat{\eta})} \\
&\quad-\frac{1}{2}\frac{1/TS\sum_{ts}\left[\nabla_{a_i}\log f(a_i,\widehat{\gamma}_t,\widehat{\eta}_s)\right]^2\mathbb{E}l_i^{(2)}(a_i,c,g)}{\left(\mathbb{E}l_i^{(2)}(a_i,\widehat{\gamma},\widehat{\eta})\right)^2}.
\end{aligned}
$$

By a similar derivation,

$$
\begin{aligned}
\mathbb{E}l_t(a,\gamma_t,g) &= \mathbb{E}l_t(a,\widehat{\gamma}_t,g)+\frac{\mathbb{E}d_t(a,\gamma_t,g)}{NS}+o\left(\frac{1}{NS}\right) \\
&= \mathbb{E}l_t(a,\widehat{\gamma}_t,g)+\frac{\mathbb{E}d_t(a,\widehat{\gamma}_t,g)}{NS}+o\left(\frac{1}{NS}\right) \\
\mathbb{E}l_s(a,c,\eta_s) &= \mathbb{E}l_s(a,c,\widehat{\eta}_s)+\frac{\mathbb{E}k_s(a,c,\eta_s)}{NT}+o\left(\frac{1}{NT}\right) \\
&= \mathbb{E}l_s(a,c,\widehat{\eta}_s)+\frac{\mathbb{E}k_s(a,c,\widehat{\eta}_s)}{NT}+o\left(\frac{1}{NT}\right)
\end{aligned}
$$

where

$$
l_t(a,c_t,g)\equiv\frac{1}{NS}\sum_{is}\log f(a_i,c_t,g_s),\qquad l_t^{(r)}(a,c_t,g)\equiv\frac{1}{NS}\sum_{is}\nabla_{c_t}^r\log f(a_i,c_t,g_s),
$$
$$
l_s(a,c,g_s)\equiv\frac{1}{NT}\sum_{it}\log f(a_i,c_t,g_s),\qquad l_s^{(r)}(a,c,g_s)\equiv\frac{1}{NT}\sum_{it}\nabla_{g_s}^r\log f(a_i,c_t,g_s),
$$

$$d_t(a,c_t,g) \equiv \frac{1/NS\sum_{is}\nabla_{c_t}\log f(\widehat{\alpha}_i,c_t,\widehat{\eta}_s)\nabla_{c_t}\log f(a_i,c_t,g_s)}{\mathbb{E}l_t^{(2)}(\widehat{\alpha},c_t,\widehat{\eta})} - \frac{1}{2}\frac{1/NS\sum_{is}[\nabla_{c_t}\log f(\widehat{\alpha}_i,c_t,\widehat{\eta}_s)]^2\,\mathbb{E}l_t^{(2)}(a,c_t,g)}{\left(\mathbb{E}l_t^{(2)}(\widehat{\alpha},c_t,\widehat{\eta})\right)^2},$$

$$k_s(a,c,g_s) \equiv \frac{1/NT\sum_{it}\nabla_{g_s}\log f(\widehat{\alpha}_i,\widehat{\gamma}_t,g_s)\nabla_{g_s}\log f(a_i,c_t,g_s)}{\mathbb{E}l_s^{(2)}(\widehat{\alpha},\widehat{\gamma},g_s)} - \frac{1}{2}\frac{1/NT\sum_{it}[\nabla_{g_s}\log f(\widehat{\alpha}_i,\widehat{\gamma}_t,g_s)]^2\,\mathbb{E}l_s^{(2)}(a,c,g_s)}{\left(\mathbb{E}l_s^{(2)}(\widehat{\alpha},\widehat{\gamma},g_s)\right)^2}.$$

Next, as

$$\frac{1}{N}\sum_i l_i(a_i,c,g) \equiv \frac{1}{T}\sum_t l_t(a,c_t,g) \equiv \frac{1}{S}\sum_s l_s(a,c,g_s)$$

and, letting $\eta \equiv (\eta_1(\theta),\cdots,\eta_S(\theta))$,

$$\begin{aligned}
\frac{1}{N}\sum_i \mathbb{E}l_i(\alpha_i,\gamma,\eta) &= \frac{1}{N}\sum_i \mathbb{E}l_i(\widehat{\alpha}_i,\gamma,\eta) + \frac{1}{N}\sum_i \frac{\mathbb{E}b_i(\widehat{\alpha}_i,\gamma,\eta)}{TS} + o\left(\frac{1}{TS}\right),\\
\frac{1}{T}\sum_t \mathbb{E}l_t(\widehat{\alpha},\gamma_t,\eta) &= \frac{1}{T}\sum_t \mathbb{E}l_t(\widehat{\alpha},\widehat{\gamma}_t,\eta) + \frac{1}{T}\sum_t \frac{\mathbb{E}d_t(\widehat{\alpha},\widehat{\gamma}_t,\eta)}{NS} + o\left(\frac{1}{NS}\right),\\
\frac{1}{S}\sum_s \mathbb{E}l_s(\widehat{\alpha},\widehat{\gamma},\eta_s) &= \frac{1}{S}\sum_s \mathbb{E}l_s(\widehat{\alpha},\widehat{\gamma},\widehat{\eta}_s) + \frac{1}{S}\sum_s \frac{\mathbb{E}k_s(\widehat{\alpha},\widehat{\gamma},\widehat{\eta}_s)}{NT} + o\left(\frac{1}{NT}\right);
\end{aligned}$$

it follows that

$$\begin{aligned}
&\frac{1}{N}\sum_i \mathbb{E}l_i(\alpha_i,\gamma,\eta)\\
=\;& \frac{1}{T}\sum_t \mathbb{E}l_t(\widehat{\alpha},\widehat{\gamma}_t,\eta) + \frac{1}{T}\sum_t \frac{\mathbb{E}d_t(\widehat{\alpha},\widehat{\gamma}_t,\eta)}{NS} + \frac{1}{N}\sum_i \frac{\mathbb{E}b_i(\widehat{\alpha}_i,\gamma,\eta)}{TS}\\
&+o\left(\frac{1}{TS}\right) + o\left(\frac{1}{NS}\right)\\
=\;& \frac{1}{S}\sum_s \mathbb{E}l_s(\widehat{\alpha},\widehat{\gamma},\widehat{\eta}_s) + \frac{1}{S}\sum_s \frac{\mathbb{E}k_s(\widehat{\alpha},\widehat{\gamma},\widehat{\eta}_s)}{NT} + \frac{1}{T}\sum_t \frac{\mathbb{E}d_t(\widehat{\alpha},\widehat{\gamma}_t,\eta)}{NS}\\
&+\frac{1}{N}\sum_i \frac{\mathbb{E}b_i(\widehat{\alpha}_i,\gamma,\eta)}{TS} + o\left(\frac{1}{TS}\right) + o\left(\frac{1}{NS}\right) + o\left(\frac{1}{NT}\right)\\
=\;& \frac{1}{S}\sum_s \mathbb{E}l_s(\widehat{\alpha},\widehat{\gamma},\widehat{\eta}_s) + \frac{1}{S}\sum_s \frac{\mathbb{E}k_s(\widehat{\alpha},\widehat{\gamma},\widehat{\eta}_s)}{NT} + \frac{1}{T}\sum_t \frac{\mathbb{E}d_t(\widehat{\alpha},\widehat{\gamma}_t,\widehat{\eta})}{NS} + \frac{O(1/NT)}{NS}\\
&+\frac{1}{N}\sum_i \frac{\mathbb{E}b_i(\widehat{\alpha}_i,\widehat{\gamma},\widehat{\eta})}{TS} + \frac{O(1/NS)}{TS} + \frac{O(1/NT)}{TS}\\
&+o\left(\frac{1}{TS}\right) + o\left(\frac{1}{NS}\right) + o\left(\frac{1}{NT}\right)
\end{aligned}$$

in which

$$\frac{O(1/NT)}{NS} = o\left(\frac{1}{NTS}\right), \qquad \frac{O(1/NS)}{TS} = o\left(\frac{1}{NTS}\right), \qquad \frac{O(1/NT)}{TS} = o\left(\frac{1}{NTS}\right).$$

Finally,

$$\mathbb{E}L(\theta) = \mathbb{E}\widehat{L}(\theta) + \frac{\mathbb{E}\widehat{B}(\theta)}{TS} + \frac{\mathbb{E}\widehat{D}(\theta)}{NS} + \frac{\mathbb{E}\widehat{K}(\theta)}{NT} + o\left(\frac{1}{TS}\right) + o\left(\frac{1}{NS}\right) + o\left(\frac{1}{NT}\right) \quad (3.13)$$

where

$$L(\theta) \equiv \frac{1}{NTS}\sum_{its} \log f(\alpha_i, \gamma_t, \eta_s), \qquad \widehat{L}(\theta) = \frac{1}{NTS}\sum_{its} \log f(\widehat{\alpha}_i, \widehat{\gamma}_t, \widehat{\eta}_s),$$
$$\widehat{B}(\theta) \equiv \frac{1}{N}\sum_i b_i(\widehat{\alpha}_i, \widehat{\gamma}, \widehat{\eta}), \qquad \widehat{D}(\theta) \equiv \frac{1}{T}\sum_t d_t(\widehat{\alpha}, \widehat{\gamma}_t, \widehat{\eta}),$$
$$\widehat{K}(\theta) \equiv \frac{1}{S}\sum_s k_s(\widehat{\alpha}, \widehat{\gamma}, \widehat{\eta}_s).$$

We present a simple example in section 4.1 regarding the application of equation (3.13).

4 Application of Correction

4.1 Analytical Correction of Many-normal-mean Model

The first example is a variation of the Neyman and Scott (1948) variance example. Let $Y_{it} \sim \mathcal{N}(\alpha_i + \gamma_t, \theta_0)$ where $\mathcal{N}(\alpha_i + \gamma_t, \theta_0)$ is the normal density with mean $\alpha_i + \gamma_t$ and variance θ_0. The individual log-likelihood for a single Y_{it} follows as

$$\log f(Y_{it}; \theta, a_i, c_t) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\theta - \frac{(Y_{it} - a_i - c_t)^2}{2\theta}$$

and the log-likelihood of all observations follows as

$$\frac{1}{NT}\sum_{it} \log f(Y_{it}; \theta, a_i, c_t) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\theta - \frac{1}{NT}\sum_{it}\frac{(Y_{it} - a_i - c_t)^2}{2\theta}.$$

Here it is obvious that α_i and γ_t, and hence the estimators $\widehat{\alpha}_i$ and $\widehat{\gamma}_t$, are not uniquely identified. This, however, does not affect the analysis, because the following can be set up, similar to Fernández-Val and Weidner (2016),

$$\begin{aligned} \gamma_t &\equiv 0, \\ \widehat{\alpha}_i &\equiv \frac{1}{T}\sum_t Y_{it}, \qquad \widehat{\gamma}_t \equiv \frac{1}{N}\sum_i Y_{it} - \frac{1}{NT}\sum_{it} Y_{it} \end{aligned}$$

from which we have

$$\frac{1}{NT}\sum_{it} \log f(Y_{it}; \theta, \widehat{\alpha}_i, \widehat{\gamma}_t) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\theta - \frac{1}{NT}\sum_{it}\frac{(Y_{it} - \widehat{\alpha}_i - \widehat{\gamma}_t)^2}{2\theta},$$

whose maximum is achieved when

$$0 = \frac{1}{NT}\sum_{it}\frac{\partial \log f(Y_{it}; \theta, \widehat{\alpha}_i, \widehat{\gamma}_t)}{\partial\theta}$$

$$= -\frac{1}{2}\frac{1}{\theta} + \frac{1}{NT}\sum_{it}\frac{(Y_{it} - \widehat{\alpha}_i - \widehat{\gamma}_t)^2}{2\theta^2}$$

where, as $\theta \neq 0$ and under $N, T \to \infty$,

$$\begin{aligned}\widehat{\theta} &= \frac{1}{NT}\sum_{it}(Y_{it} - \widehat{\alpha}_i - \widehat{\gamma}_t)^2 = \theta_0 - \frac{\theta_0}{T} - \frac{\theta_0}{N} + \frac{\theta_0}{NT} \\ &= \theta_0 + O_p\left(N^{-1}\right) + O_p\left(T^{-1}\right).\end{aligned}$$

The corrected log-likelihood defined in equation (3.11) can be applied to this model. Observe that

$$\begin{aligned}\nabla_{a_i}\log f(Y_{it};\theta, a_i, c_t) &= \frac{Y_{it} - a_i - c_t}{\theta}, \qquad \nabla_{c_t}\log f(Y_{it};\theta, a_i, c_t) = \frac{Y_{it} - a_i - c_t}{\theta}, \\ \nabla^2_{a_i}\log f(Y_{it};\theta, a_i, c_t) &= -\frac{1}{\theta}, \qquad \nabla^2_{c_t}\log f(Y_{it};\theta, a_i, c_t) = -\frac{1}{\theta}\end{aligned}$$

such that

$$b_i(\widehat{\alpha}_i, \widehat{\gamma}) = -\frac{1}{T}\sum_t \frac{(Y_{it} - \widehat{\alpha}_i - \widehat{\gamma}_t)^2}{2\theta}, \qquad d_t(\widehat{\alpha}, \widehat{\gamma}_t) = -\frac{1}{N}\sum_i \frac{(Y_{it} - \widehat{\alpha}_i - \widehat{\gamma}_t)^2}{2\theta}$$

and that the corrected profiled log-likelihood is

$$\begin{aligned}\widetilde{L}(\theta) &= -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\theta - \frac{1}{NT}\sum_{it}\frac{(Y_{it} - \widehat{\alpha}_i - \widehat{\gamma}_t)^2}{2\theta} \\ &\quad -\frac{1}{T}\frac{1}{NT}\sum_{it}\frac{(Y_{it} - \widehat{\alpha}_i - \widehat{\gamma}_t)^2}{2\theta} - \frac{1}{N}\frac{1}{NT}\sum_{it}\frac{(Y_{it} - \widehat{\alpha}_i - \widehat{\gamma}_t)^2}{2\theta} \\ &= -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\theta - \left(1 + \frac{1}{T} + \frac{1}{N}\right)\frac{1}{NT}\sum_{it}\frac{(Y_{it} - \widehat{\alpha}_i - \widehat{\gamma}_t)^2}{2\theta},\end{aligned}$$

which is maximized when

$$0 = -\frac{1}{2}\frac{1}{\theta} + \left(1 + \frac{1}{T} + \frac{1}{N}\right)\frac{1}{NT}\sum_{it}\frac{(Y_{it} - \widehat{\alpha}_i - \widehat{\gamma}_t)^2}{2\theta^2}$$

such that

$$\begin{aligned}\widetilde{\theta} &= \left(1 + \frac{1}{T} + \frac{1}{N}\right)\frac{1}{NT}\sum_{it}(Y_{it} - \widehat{\alpha}_i - \widehat{\gamma}_t)^2 = \left(1 + \frac{1}{T} + \frac{1}{N}\right)\widehat{\theta} \\ &= \theta_0\left(1 + \frac{1}{T} + \frac{1}{N}\right)\left(1 - \frac{1}{T} - \frac{1}{N} + \frac{1}{NT}\right) \\ &= \theta_0 - \frac{\theta_0}{NT} - \frac{\theta_0}{T^2} - \frac{\theta_0}{N^2} + \frac{\theta_0}{N^2T} + \frac{\theta_0}{NT^2}\end{aligned}$$

implying, as $N/T \to \kappa$ when $N, T \to \infty$,

$$\begin{aligned}\widetilde{\theta} - \theta_0 &= O_p\left(\frac{1}{NT}\right) \\ &= o_p\left(T^{-1}\right) + o_p\left(N^{-1}\right).\end{aligned}$$

Here one should observe that $\widetilde{\theta}$ possesses a higher-order bias to the order of $o_p(1/NT)$ which does not exist in $\widehat{\theta}$. This is because the correction terms themselves depend on plug-in estimates. This will generate a bias to the higher order, i.e.,

$$\begin{aligned}\frac{1}{T}\mathbb{E}b_i(\widehat{\alpha}_i,\widehat{\gamma}) &= \frac{1}{T}\mathbb{E}b_i(\alpha_i,\gamma)+\frac{1}{T}O\left(T^{-1}\right)+\frac{1}{T}O\left(N^{-1}\right),\\ \frac{1}{N}\mathbb{E}d_t(\widehat{\alpha},\widehat{\gamma}_t) &= \frac{1}{N}\mathbb{E}d_t(\alpha,\gamma_t)+\frac{1}{N}O\left(T^{-1}\right)+\frac{1}{N}O\left(N^{-1}\right).\end{aligned}$$

Second, $\widetilde{\theta}$ would not be fully unbiased even when α and γ were plugged into the correction terms. To see this, suppose

$$b_i(\alpha_i,\gamma)=-\frac{1}{T}\sum_t\frac{(Y_{it}-\alpha_i-\gamma_t)^2}{2\theta},\qquad d_t(\alpha,\gamma_t)=-\frac{1}{N}\sum_i\frac{(Y_{it}-\alpha_i-\gamma_t)^2}{2\theta}$$

were plugged in, the corrected profile log-likelihood would then be

$$\begin{aligned}&-\frac{1}{2}\log(2\pi)-\frac{1}{2}\log\theta-\frac{1}{NT}\sum_{it}\frac{(Y_{it}-\widehat{\alpha}_i-\widehat{\gamma}_t)^2}{2\theta}\\ &-\frac{1}{T}\frac{1}{NT}\sum_{it}\frac{(Y_{it}-\alpha_i-\gamma_t)^2}{2\theta}-\frac{1}{N}\frac{1}{NT}\sum_{it}\frac{(Y_{it}-\alpha_i-\gamma_t)^2}{2\theta},\end{aligned}$$

which is maximized when

$$\begin{aligned}0 &= -\frac{1}{2}\frac{1}{\theta}+\frac{1}{NT}\sum_{it}\frac{(Y_{it}-\widehat{\alpha}_i-\widehat{\gamma}_t)^2}{2\theta^2}+\left(\frac{1}{T}+\frac{1}{N}\right)\frac{1}{NT}\sum_{it}\frac{(Y_{it}-\alpha_i-\gamma_t)^2}{2\theta^2}\\ \widehat{\theta}^* &= \frac{1}{NT}\sum_{it}(Y_{it}-\widehat{\alpha}_i-\widehat{\gamma}_t)^2+\left(\frac{1}{T}+\frac{1}{N}\right)\frac{1}{NT}\sum_{it}(Y_{it}-\alpha_i-\gamma_t)^2,\end{aligned}$$

i.e.,

$$\widehat{\theta}^*=\theta_0+\frac{\theta_0}{NT}=\theta_0+O_p\left(\frac{1}{NT}\right)$$

implying that there is still a bias that is of the order of $O_p(1/NT)$.

Next, let us introduce an additional nuisance parameter such that

$$Y_{its}\sim\mathcal{N}\left(\alpha_i+\gamma_t+\eta_s,\theta_0\right)$$

and that the individual log-likelihood for a single Y_{its} becomes

$$\log f(Y_{its};\theta,a_i,c_t,g_s)=-\frac{1}{2}\log(2\pi)-\frac{1}{2}\log\theta-\frac{(Y_{its}-a_i-c_t-g_s)^2}{2\theta}.$$

This corresponds to the $J=3$ case introduced in section 3.3. For

$$\begin{aligned}\gamma_t &= \eta_s=0,\qquad \widehat{\alpha}_i\equiv\frac{1}{TS}\sum_{ts}Y_{its},\\ \widehat{\gamma}_t &\equiv \frac{1}{NS}\sum_{is}Y_{its}-\frac{1}{NTS}\sum_{its}Y_{its},\qquad \widehat{\eta}_s\equiv\frac{1}{NT}\sum_{it}Y_{its}-\frac{1}{NTS}\sum_{its}Y_{its},\end{aligned}$$

the profiled log-likelihood becomes

$$\log f\left(Y_{its};\theta,\widehat{\alpha}_i,\widehat{\gamma}_t,\widehat{\eta}_s\right)=-\frac{1}{2}\log\left(2\pi\right)-\frac{1}{2}\log\theta-\frac{1}{NTS}\sum_{its}\frac{\left(Y_{its}-\widehat{\alpha}_i-\widehat{\gamma}_t-\widehat{\eta}_s\right)^2}{2\theta},$$

which is maximized when

$$\begin{aligned}0&=-\frac{1}{2}\frac{1}{\theta}+\frac{1}{NTS}\sum_{its}\frac{\left(Y_{its}-\widehat{\alpha}_i-\widehat{\gamma}_t-\widehat{\eta}_s\right)^2}{2\theta}\\ \widehat{\theta}&=\frac{1}{NTS}\sum_{its}\left(Y_{its}-\widehat{\alpha}_i-\widehat{\gamma}_t-\widehat{\eta}_s\right)^2.\end{aligned}$$

Here it can be shown that, after some algebra,

$$\widehat{\theta}=\theta_0-\frac{\theta_0}{TS}-\frac{\theta_0}{NS}-\frac{\theta_0}{NT}+O_p\left(\frac{1}{NTS}\right).$$

For the correction, observe that

$$\begin{aligned}b_i\left(\widehat{\alpha}_i,\widehat{\gamma},\widehat{\eta}\right)&=-\frac{1}{TS}\sum_{ts}\frac{\left(Y_{it}-\widehat{\alpha}_i-\widehat{\gamma}_t-\widehat{\eta}_s\right)^2}{2\theta},\\ d_t\left(\widehat{\alpha},\widehat{\gamma}_t,\widehat{\eta}\right)&=-\frac{1}{NS}\sum_{is}\frac{\left(Y_{it}-\widehat{\alpha}_i-\widehat{\gamma}_t-\widehat{\eta}_s\right)^2}{2\theta},\\ k_s\left(\widehat{\alpha},\widehat{\gamma},\widehat{\eta}_s\right)&=-\frac{1}{NT}\sum_{it}\frac{\left(Y_{it}-\widehat{\alpha}_i-\widehat{\gamma}_t-\widehat{\eta}_s\right)^2}{2\theta}\end{aligned}$$

such that the corrected profiled log-likelihood follows as

$$\begin{aligned}\widetilde{L}\left(\theta\right)&=-\frac{1}{2}\log\left(2\pi\right)-\frac{1}{2}\log\theta\\ &\quad-\left(1+\frac{1}{NS}+\frac{1}{NT}+\frac{1}{TS}\right)\frac{1}{NTS}\sum_{its}\frac{\left(Y_{it}-\widehat{\alpha}_i-\widehat{\gamma}_t-\widehat{\eta}_s\right)^2}{2\theta},\end{aligned}$$

which is maximized when

$$0=-\frac{1}{2}\frac{1}{\theta}+\left(1+\frac{1}{NS}+\frac{1}{NT}+\frac{1}{TS}\right)\frac{1}{NTS}\sum_{its}\frac{\left(Y_{it}-\widehat{\alpha}_i-\widehat{\gamma}_t-\widehat{\eta}_s\right)^2}{2\theta}$$

such that

$$\begin{aligned}\widetilde{\theta}&=\left(1+\frac{1}{NS}+\frac{1}{NT}+\frac{1}{TS}\right)\widehat{\theta}\\ &=\left(1+\frac{1}{NS}+\frac{1}{NT}+\frac{1}{TS}\right)\left(\theta_0-\frac{\theta_0}{TS}-\frac{\theta_0}{NS}-\frac{\theta_0}{NT}\right)+O_p\left(\frac{1}{NTS}\right)\\ &=\theta_0+O_p\left(\frac{1}{NTS}\right),\end{aligned}$$

which indicates, under $N/S\to\kappa_{N,S}$ and $S/T\to\kappa_{S,T}$ as $N,T,S\to\infty$ where $0<\kappa_{N,S}<\infty$ and $0<\kappa_{S,T}<\infty$,

$$\widetilde{\theta}-\theta_0=O_p\left(\frac{1}{NTS}\right).$$

4.2 Correction of Static Logit

The next example is the static logit model. We consider

$$Y_{it} = 1\left(X_{it}\theta_0 + \alpha_i + \gamma_t + \varepsilon_{it} \geq 0\right)$$

where ε_{it} follows a standard logistic distribution and X_{it} is a scalar covariate.

Tables 1, 2, and 3 present simulation results of the logit model under three different designs.

1. $X_{it} \sim \mathcal{N}(0,1)$ and $\alpha_i = \gamma_t = 0$ for all i and t. This represents the case where the model could be consistently estimated by a pooled logit.

2. $X_{it} \sim \mathcal{N}(0,1)$, $\alpha_i \sim \mathcal{N}(0,1/16)$, and $\gamma_t \sim \mathcal{N}(0,1/16)$. This represents the case where the model could be consistently estimated by a random-effect logit.

3. $X_{it} \sim \mathcal{N}(\alpha_i + \gamma_t, 1)$ with $\alpha_i \sim \mathcal{N}(0,1/16)$ and $\gamma_t \sim \mathcal{N}(0,1/16)$. This represents the case where the model must be estimated by a fixed-effect logit.

The number of replications in the Monte Carlo experiment is $1,000$ with N, T, and θ_0 chosen according to the description in the tables. Notice that the IPP occurs when α_i and γ_t are allowed to be estimated. That is, even when $\alpha_i = \gamma_t = 0$, i.e., the underlying model is a pooled logit, estimating a fixed-effect model would induce the IPP.

We find that the correction is generally sufficient given the variation of $\widehat{\theta}$. For example in design 1 and under $\theta_0 = 0.5$ and $N, T = 10$, the correction technique reduces the bias by a percentage of roughly 67%. The RMSEs also improve significantly. Under the same setting, the RMSE is reduced by roughly 24%. This highlights a distinct feature of the analytical correction, i.e., the correction technique typically would not induce a large dispersion to the estimators. The finite-sample properties of the corrected estimators, therefore, are more desirable in terms of the variation. On the other hand in design 3, we find that there are two cases (bold) where the bias in $\widetilde{\theta}$ seems to increase when N, T are increased from 40 to 80. We regard this as a consequence of the variation that is still large.

Figures 1 and 2 present plots of the profiled log-likelihood functions for $N, T = 10$, $N, T = 20$, $N, T = 40$, and $N, T = 80$. The model is $Y_{it} = 1\left(X_{it}\theta_0 + \alpha_i + \gamma_t + \varepsilon_{it} \geq 0\right)$ where ε_{it} is standard logistically distributed, $X_{it} \sim \mathcal{N}(0,1)$, $\alpha_i = \gamma_t = 0$, and $\theta_0 = 0.5$. The plotted quantities are $\widehat{L}(\theta)$ (circle), $\widetilde{L}(\theta)$ (triangle), and $L(\theta)$ (asterisk) computed for $\theta = 0.3, \cdots, 0.7$ with a step of 0.01. The plotted quantities are evaluated on a single simulated dataset. Compared with $\widehat{L}(\theta)$, we find that the approximation of $\widetilde{L}(\theta)$, the corrected profiled log-likelihood, to $L(\theta)$, the infeasible profiled log-likelihood, is dramatically improved for every chosen θ even when N, T are small. In addition, the maximizer, in θ, of $\widetilde{L}(\theta)$ is very close to that of $L(\theta)$.

On the other hand, we find that $L(\theta)$ is still biased in the sense that the maximizer in θ of $L(\theta)$ is not θ_0. This may be due to two facts. First, when N, T are small, $L(\theta)$ remains random with a large variation such that $\overline{\theta} \equiv \arg\max_\theta L(\theta)$ has a large variation. Second, $L(\theta)$ is nonlinear in θ such that $\overline{\theta}$, in general, possesses a bias up to the order of $O_p(1/NT)$, which may not be negligible when N, T are very small.

Table 1: Double IPP - Simulation Result for Logit Model - Design 1

Setting	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE
$N, T = 10$		$\theta_0 = 0.5$			$\theta_0 = -0.5$			$\theta_0 = 1$			$\theta_0 = -1$	
$\widehat{\theta}$	0.6564	0.3129	0.3591	−0.6558	0.3116	0.3539	1.3216	0.3216	0.5262	−1.3224	0.3224	0.5520
$\widetilde{\theta}$	0.5510	0.1021	0.2735	−0.5507	0.1013	0.2697	1.0860	0.0860	0.3369	−1.0849	0.0849	0.3576
$N, T = 20$		$\theta_0 = 0.5$			$\theta_0 = -0.5$			$\theta_0 = 1$			$\theta_0 = -1$	
$\widehat{\theta}$	0.5575	0.1149	0.1398	−0.5640	0.1280	0.1445	1.1359	0.1359	0.2199	−1.1300	0.1300	0.2103
$\widetilde{\theta}$	0.5095	0.0189	0.1166	−0.5152	0.0304	0.1189	1.0300	0.0300	0.1574	−1.0248	0.0248	0.1501
$N, T = 40$		$\theta_0 = 0.5$			$\theta_0 = -0.5$			$\theta_0 = 1$			$\theta_0 = -1$	
$\widehat{\theta}$	0.5280	0.0560	0.0674	−0.5273	0.0546	0.0639	1.0558	0.0558	0.0929	−1.0598	0.0598	0.0928
$\widetilde{\theta}$	0.5037	0.0074	0.0586	−0.5030	0.0060	0.0552	1.0039	0.0039	0.0703	−1.0077	0.0077	0.0676
$N, T = 80$		$\theta_0 = 0.5$			$\theta_0 = -0.5$			$\theta_0 = 1$			$\theta_0 = -1$	
$\widehat{\theta}$	0.5128	0.0256	0.0302	−0.5136	0.0271	0.0327	1.0298	0.0298	0.0455	−1.0282	0.0282	0.0452
$\widetilde{\theta}$	0.5005	0.0009	0.0267	−0.5013	0.0025	0.0291	1.0037	0.0037	0.0336	−1.0022	0.0022	0.0344

Notes: Bias is presented relative to θ_0. The number of replications is $1,000$. Model: $Y_{it} = 1\left(X_{it}\theta_0 + \alpha_i + \gamma_t + \varepsilon_{it} \geq 0\right)$ where ε_{it} is standard-logistically distributed, $X_{it} \sim \mathcal{N}(0,1)$, and $\alpha_i = \gamma_t = 0$. $\widehat{\theta}$ is the original estimate, $\widetilde{\theta}$ is the bias-corrected estimate.

Table 2: Double IPP - Simulation Result for Logit Model - Design 2

Setting	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE
$N, T = 10$	$\theta_0 = 0.5$			$\theta_0 = -0.5$			$\theta_0 = 1$			$\theta_0 = -1$		
$\widehat{\theta}$	0.6314	0.2628	0.3442	-0.6472	0.2945	0.3634	1.3052	0.3052	0.5555	-1.2930	0.2930	0.5193
$\widetilde{\theta}$	0.5298	0.0595	0.2661	-0.5415	0.0830	0.2785	1.0692	0.0692	0.3601	-1.0642	0.0642	0.3372
$N, T = 20$	$\theta_0 = 0.5$			$\theta_0 = -0.5$			$\theta_0 = 1$			$\theta_0 = -1$		
$\widehat{\theta}$	0.5576	0.1151	0.1446	-0.5553	0.1105	0.1422	1.1311	0.1311	0.2047	-1.1360	0.1360	0.2145
$\widetilde{\theta}$	0.5093	0.0186	0.1210	-0.5074	0.0149	0.1195	1.0257	0.0257	0.1430	-1.0299	0.0299	0.1513
$N, T = 40$	$\theta_0 = 0.5$			$\theta_0 = -0.5$			$\theta_0 = 1$			$\theta_0 = -1$		
$\widehat{\theta}$	0.5311	0.0621	0.0672	-0.5283	0.0565	0.0633	1.0596	0.0596	0.0925	-1.0558	0.0558	0.0925
$\widetilde{\theta}$	0.5069	0.0137	0.0572	-0.5042	0.0083	0.0541	1.0077	0.0077	0.0674	-1.0041	0.0041	0.0699
$N, T = 80$	$\theta_0 = 0.5$			$\theta_0 = -0.5$			$\theta_0 = 1$			$\theta_0 = -1$		
$\widehat{\theta}$	0.5124	0.0248	0.0310	-0.5150	0.0299	0.0324	1.0278	0.0278	0.0444	-1.0278	0.0278	0.0437
$\widetilde{\theta}$	0.5002	0.0004	0.0277	-0.5027	0.0054	0.0282	1.0017	0.0017	0.0337	-1.0017	0.0017	0.0327

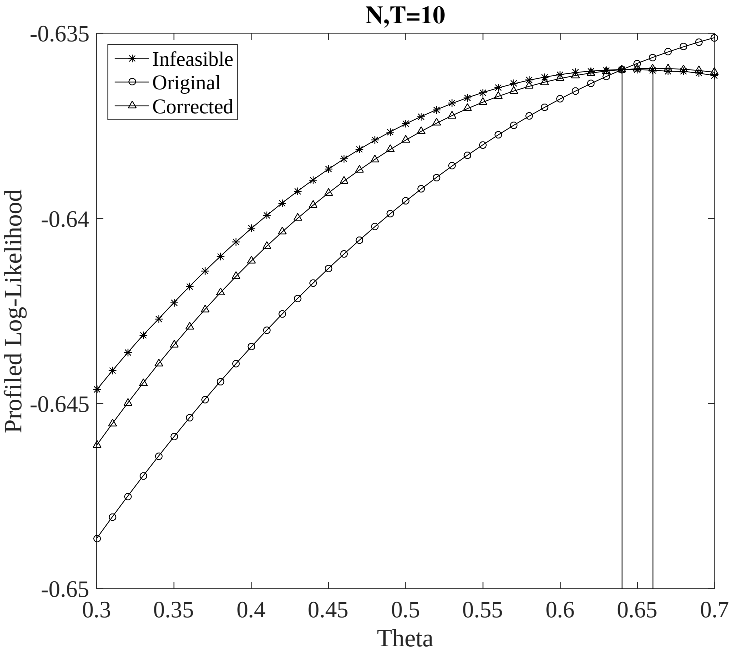
Notes: Bias is presented relative to θ_0. The number of replications is $1,000$. Model: $Y_{it} = 1\left(X_{it}\theta_0 + \alpha_i + \gamma_t + \varepsilon_{it} \geq 0\right)$ where ε_{it} is standard-logistically distributed, $X_{it} \sim \mathcal{N}(0,1)$, $\alpha_i \sim \mathcal{N}(0,1/16)$, and $\gamma_t \sim \mathcal{N}(0,1/16)$. $\widehat{\theta}$ is the original estimate, $\widetilde{\theta}$ is the bias-corrected estimate.

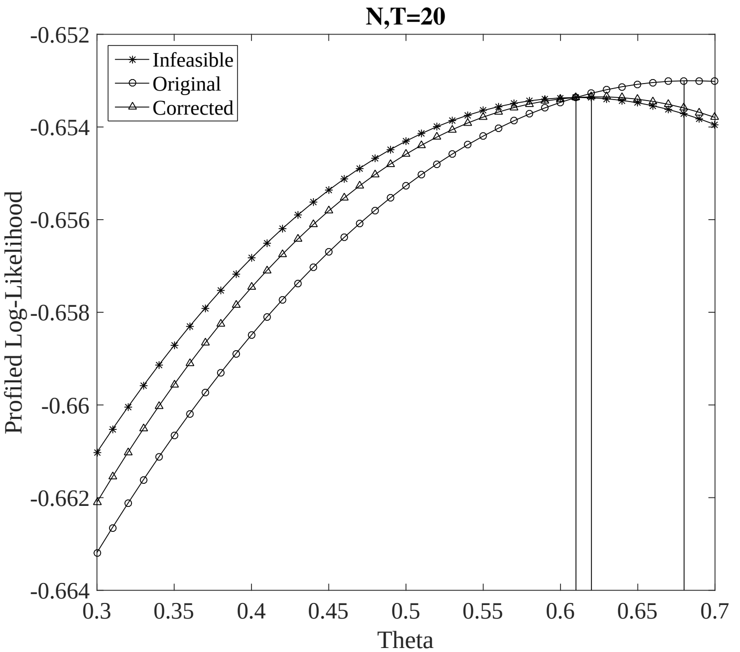
Table 3: Double IPP - Simulation Result for Logit Model - Design 3

Setting	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE
$N,T=10$	$\theta_0=0.5$			$\theta_0=-0.5$			$\theta_0=1$			$\theta_0=-1$		
$\widehat{\theta}$	0.6350	0.2701	0.3498	−0.6273	0.2546	0.3423	1.3176	0.3176	0.5619	−1.3092	0.3092	0.5541
$\widetilde{\theta}$	0.5215	0.0430	0.2654	−0.5286	0.0572	0.2647	1.0640	0.0640	0.3586	−1.0718	0.0718	0.3659
$N,T=20$	$\theta_0=0.5$			$\theta_0=-0.5$			$\theta_0=1$			$\theta_0=-1$		
$\widehat{\theta}$	0.5611	0.1222	0.1504	−0.5536	0.1071	0.1397	1.1370	0.1370	0.2164	−1.1399	0.1399	0.2139
$\widetilde{\theta}$	0.5086	0.0172	0.1251	−0.5060	0.0120	0.1171	1.0252	0.0252	0.1511	−1.0319	0.0319	0.1475
$N,T=40$	$\theta_0=0.5$			$\theta_0=-0.5$			$\theta_0=1$			$\theta_0=-1$		
$\widehat{\theta}$	0.5263	0.0527	0.0662	−0.5284	0.0567	0.0638	1.0574	0.0574	0.0963	−1.0598	0.0598	0.0945
$\widetilde{\theta}$	**0.4995**	−0.0010	0.0580	−0.5050	0.0101	0.0547	**1.0021**	0.0021	0.0730	−1.0070	0.0070	0.0697
$N,T=80$	$\theta_0=0.5$			$\theta_0=-0.5$			$\theta_0=1$			$\theta_0=-1$		
$\widehat{\theta}$	0.5151	0.0302	0.0314	−0.5137	0.0275	0.0310	1.0296	0.0296	0.0459	−1.0285	0.0285	0.0449
$\widetilde{\theta}$	**0.5019**	0.0038	0.0269	−0.5017	0.0034	0.0271	**1.0024**	0.0024	0.0343	−1.0021	0.0021	0.0337

Notes: Bias is presented relative to θ_0. The number of replications is $1,000$. Model: $Y_{it} = 1\left(X_{it}\theta_0 + \alpha_i + \gamma_t + \varepsilon_{it} \geq 0\right)$ where ε_{it} is standard-logistically distributed and $X_{it} \sim \mathcal{N}\left(\alpha_i + \gamma_t, 1\right)$ with $\alpha_i \sim \mathcal{N}\left(0, 1/16\right)$ and $\gamma_t \sim \mathcal{N}\left(0, 1/16\right)$. $\widehat{\theta}$ is the original estimate, $\widetilde{\theta}$ is the bias-corrected estimate.

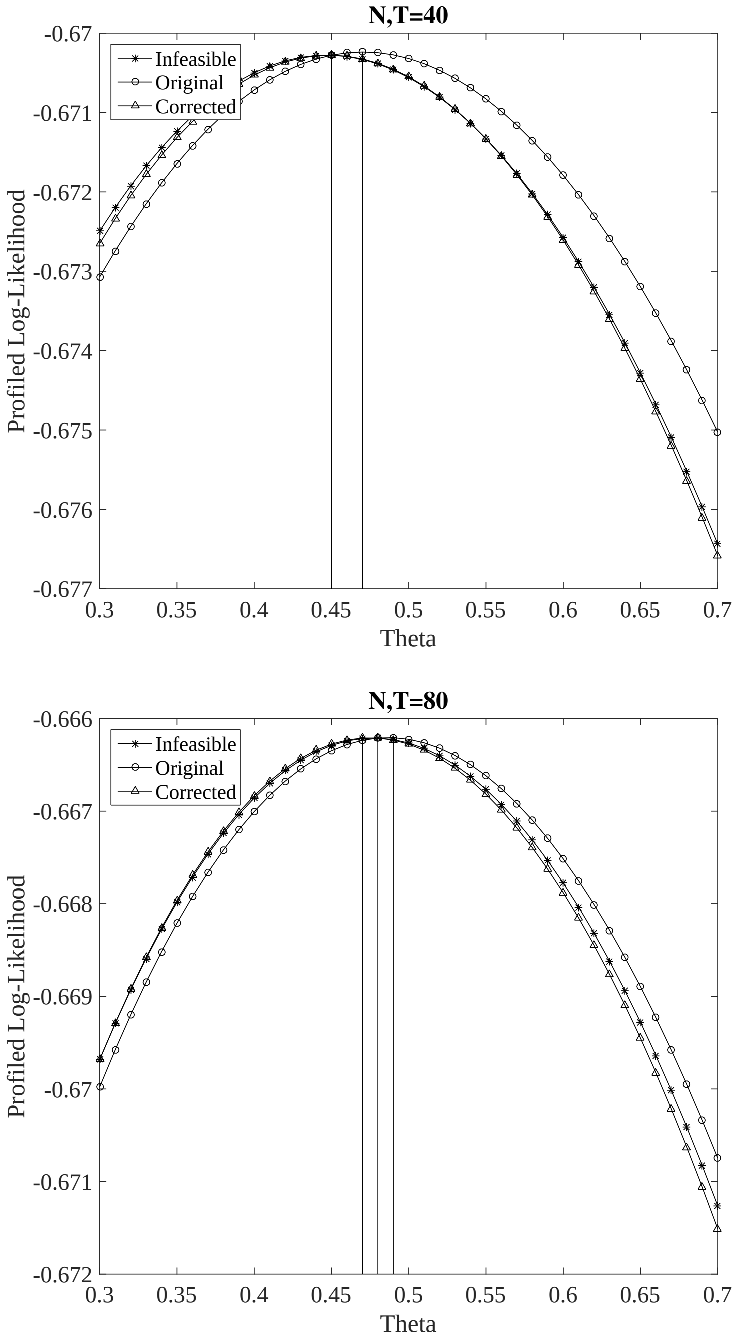
Figure 1: Double IPP - Plot of Profiled Log-likelihood for Logit - Part 1





Notes: Computed on a single simulated dataset. Model: $Y_{it} = 1\left(X_{it}\theta_0 + \alpha_i + \gamma_t + \varepsilon_{it} \geq 0\right)$ where ε_{it} is standard-logistically distributed, $X_{it} \sim \mathcal{N}(0,1)$, $\alpha_i = \gamma_t = 0$, and $\theta_0 = 0.5$. θ chosen from the region depicted on the horizontal axis with a step of 0.01. Circle: $\widehat{L}(\theta)$; triangle: $\widetilde{L}(\theta)$; and asterisk: $L(\theta)$. All curves are vertically shifted such that they coincide at $\overline{\theta}$ (maximizer of the infeasible log-likelihood). Vertical lines at maximizers.

Figure 2: Double IPP - Plot of Profiled Log-likelihood for Logit - Part 2



Notes: Computed on a single simulated dataset. Model: $Y_{it} = 1\left(X_{it}\theta_0 + \alpha_i + \gamma_t + \varepsilon_{it} \geq 0\right)$ where ε_{it} is standard-logistically distributed, $X_{it} \sim \mathcal{N}(0,1)$, $\alpha_i = \gamma_t = 0$, and $\theta_0 = 0.5$. θ chosen from the region depicted on the horizontal axis with a step of 0.01. Circle: $\widehat{L}(\theta)$; triangle: $\widetilde{L}(\theta)$; and asterisk: $L(\theta)$. All curves are vertically shifted such that they coincide at $\overline{\theta}$ (maximizer of the infeasible log-likelihood). Vertical lines at maximizers.

4.3 Correction of Static Probit

Next, we consider the probit model

$$Y_{it} = 1\left(X_{it}\theta_0 + \alpha_i + \gamma_t + \varepsilon_{it} \geq 0\right)$$

where $\varepsilon_{it} \sim \mathcal{N}(0,1)$ and X_{it} is a scalar covariate.

Tables 4, 5, and 6 present results of the simulation of the probit model under similar designs as in section 4.2. We find similar patterns as in the logit example. The correction is generally sufficient and does not induce large dispersion on the estimators. For example, when $N, T = 80$, $\widetilde{\theta}$ is only slightly biased (maximum 0.3% in all design) whereas $\widehat{\theta}$ is still roughly 3% biased. In addition, the variation of $\widehat{\theta}$ and $\widetilde{\theta}$ are smaller than those from the logit model when θ_0 is small, e.g., 0.5.

Figures 3 and 4 present plots of the profiled log-likelihood functions for $N, T = 10$, $N, T = 20$, $N, T = 40$, and $N, T = 80$. The model is $Y_{it} = 1\left(X_{it}\theta_0 + \alpha_i + \gamma_t + \varepsilon_{it} \geq 0\right)$ where $\varepsilon_{it} \sim \mathcal{N}(0,1)$, $X_{it} \sim \mathcal{N}(0,1)$, $\alpha_i = \gamma_t = 0$, and $\theta_0 = 0.5$. The plotted quantities are $\widehat{L}(\theta)$ (circle), $\widetilde{L}(\theta)$ (triangle), and $L(\theta)$ (asterisk), computed for $\theta = 0.3, \cdots, 0.7$ with a step of 0.01. The plotted quantities are evaluated on a single simulated dataset. We observe a similar pattern as in the logit case, i.e., $\widetilde{L}(\theta)$ serves as a better approximation of $L(\theta)$ than $\widehat{L}(\theta)$.

Table 4: Double IPP - Simulation Result for Probit Model - Design 1

Setting	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE
$N, T = 10$	$\theta_0 = 0.5$			$\theta_0 = -0.5$			$\theta_0 = 1$			$\theta_0 = -1$		
$\widehat{\theta}$	0.6408	0.2816	0.2654	-0.6428	0.2856	0.2582	1.3875	0.3875	0.5597	-1.3957	0.3957	0.6036
$\widetilde{\theta}$	0.5483	0.0966	0.1909	-0.5508	0.1017	0.1851	1.1352	0.1352	0.3190	-1.1375	0.1375	0.3409
$N, T = 20$	$\theta_0 = 0.5$			$\theta_0 = -0.5$			$\theta_0 = 1$			$\theta_0 = -1$		
$\widehat{\theta}$	0.5547	0.1093	0.1027	-0.5517	0.1034	0.0989	1.1380	0.1380	0.1915	-1.1372	0.1372	0.1914
$\widetilde{\theta}$	0.5125	0.0250	0.0807	-0.5099	0.0198	0.0775	1.0331	0.0331	0.1201	-1.0322	0.0322	0.1206
$N, T = 40$	$\theta_0 = 0.5$			$\theta_0 = -0.5$			$\theta_0 = 1$			$\theta_0 = -1$		
$\widehat{\theta}$	0.5241	0.0482	0.0479	-0.5243	0.0485	0.0466	1.0603	0.0603	0.0813	-1.0582	0.0582	0.0805
$\widetilde{\theta}$	0.5031	0.0062	0.0394	-0.5032	0.0064	0.0381	1.0090	0.0090	0.0515	-1.0068	0.0068	0.0523
$N, T = 80$	$\theta_0 = 0.5$			$\theta_0 = -0.5$			$\theta_0 = 1$			$\theta_0 = -1$		
$\widehat{\theta}$	0.5115	0.0230	0.0217	-0.5120	0.0241	0.0221	1.0283	0.0283	0.0383	-1.0284	0.0284	0.0385
$\widetilde{\theta}$	0.5009	0.0019	0.0180	-0.5015	0.0030	0.0181	1.0025	0.0025	0.0250	-1.0027	0.0027	0.0252

Notes: Bias is presented relative to θ_0. The number of replications is $1,000$. Model: $Y_{it} = 1\left(X_{it}\theta_0 + \alpha_i + \gamma_t + \varepsilon_{it} \geq 0\right)$ where $\varepsilon_{it} \sim \mathcal{N}(0,1)$, $X_{it} \sim \mathcal{N}(0,1)$, and $\alpha_i = \gamma_t = 0$. $\widehat{\theta}$ is the original estimate, $\widetilde{\theta}$ is the bias-corrected estimate.

Table 5: Double IPP - Simulation Result for Probit Model - Design 2

Setting	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE
$N,T=10$		$\theta_0=0.5$			$\theta_0=-0.5$			$\theta_0=1$			$\theta_0=-1$	
$\widehat{\theta}$	0.6453	0.2905	0.2708	-0.6484	0.2968	0.2742	1.3889	0.3889	0.5683	-1.3976	0.3976	0.5854
$\widetilde{\theta}$	0.5501	0.1002	0.1947	-0.5531	0.1061	0.1959	1.1305	0.1305	0.3261	-1.1383	0.1383	0.3399
$N,T=20$		$\theta_0=0.5$			$\theta_0=-0.5$			$\theta_0=1$			$\theta_0=-1$	
$\widehat{\theta}$	0.5592	0.1184	0.1069	-0.5572	0.1144	0.1044	1.1431	0.1431	0.1921	-1.1497	0.1497	0.2007
$\widetilde{\theta}$	0.5161	0.0321	0.0830	-0.5140	0.0280	0.0806	1.0354	0.0354	0.1167	-1.0411	0.0411	0.1231
$N,T=40$		$\theta_0=0.5$			$\theta_0=-0.5$			$\theta_0=1$			$\theta_0=-1$	
$\widehat{\theta}$	0.5239	0.0478	0.0465	-0.5242	0.0484	0.0474	1.0637	0.0637	0.0848	-1.0604	0.0604	0.0830
$\widetilde{\theta}$	0.5026	0.0053	0.0380	-0.5029	0.0058	0.0390	1.0116	0.0116	0.0534	-1.0083	0.0083	0.0535
$N,T=80$		$\theta_0=0.5$			$\theta_0=-0.5$			$\theta_0=1$			$\theta_0=-1$	
$\widehat{\theta}$	0.5112	0.0225	0.0219	-0.5117	0.0234	0.0224	1.0286	0.0286	0.0387	-1.0283	0.0283	0.0386
$\widetilde{\theta}$	0.5007	0.0013	0.0184	-0.5011	0.0023	0.0187	1.0025	0.0025	0.0252	-1.0022	0.0022	0.0255

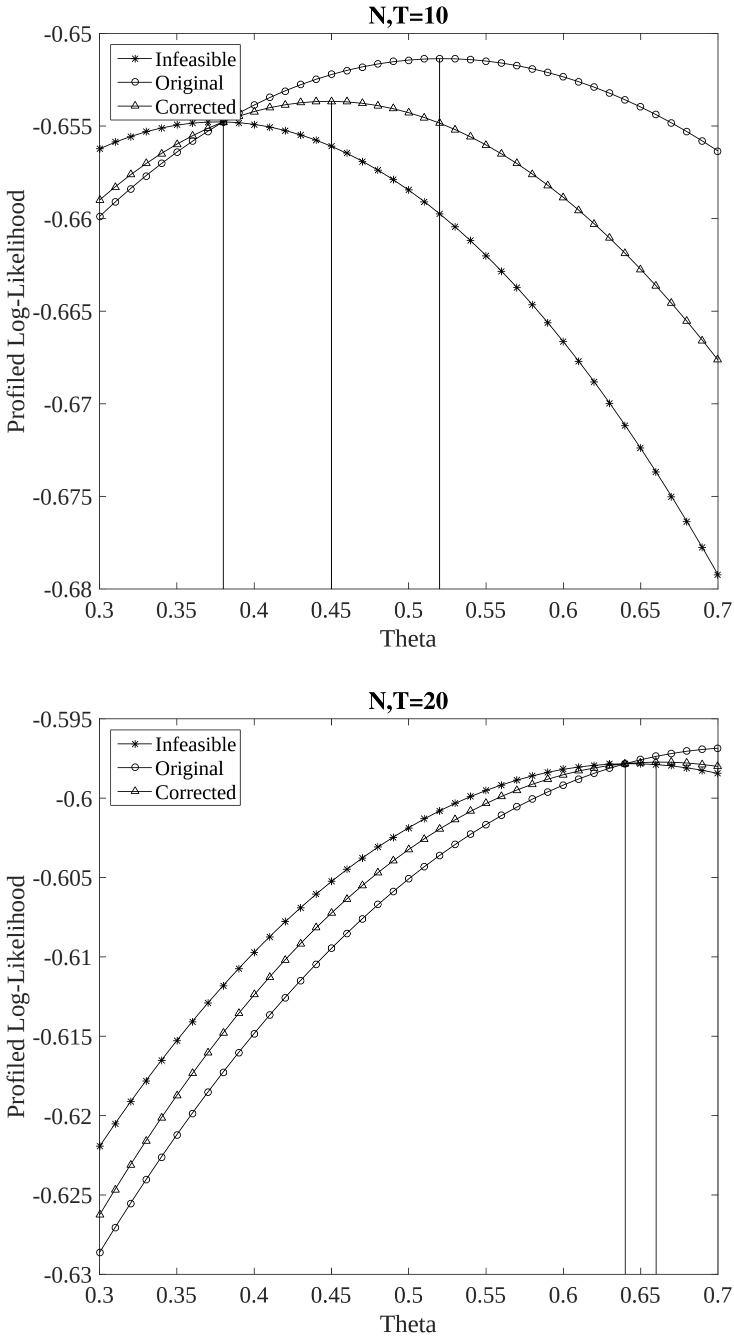
Notes: Bias is presented relative to θ_0. The number of replications is $1,000$. Model: $Y_{it} = 1\left(X_{it}\theta_0 + \alpha_i + \gamma_t + \varepsilon_{it} \geq 0\right)$ where $\varepsilon_{it} \sim \mathcal{N}(0,1)$, $X_{it} \sim \mathcal{N}(0,1)$, $\alpha_i \sim \mathcal{N}(0,1/16)$, and $\gamma_t \sim \mathcal{N}(0,1/16)$. $\widehat{\theta}$ is the original estimate, $\widetilde{\theta}$ is the bias-corrected estimate.

Table 6: Double IPP - Simulation Result for Probit Model - Design 3

Setting	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE
$N, T = 10$		$\theta_0 = 0.5$			$\theta_0 = -0.5$			$\theta_0 = 1$			$\theta_0 = -1$	
$\widehat{\theta}$	0.6601	0.3203	0.2875	-0.6240	0.2479	0.2512	1.4746	0.4746	0.6667	-1.3921	0.3921	0.5634
$\widetilde{\theta}$	0.5531	0.1062	0.1981	-0.5345	0.0691	0.1835	1.1777	0.1777	0.3744	-1.1374	0.1374	0.3233
$N, T = 20$		$\theta_0 = 0.5$			$\theta_0 = -0.5$			$\theta_0 = 1$			$\theta_0 = -1$	
$\widehat{\theta}$	0.5609	0.1217	0.1102	-0.5544	0.1088	0.1027	1.1510	0.1510	0.2027	-1.1336	0.1336	0.1864
$\widetilde{\theta}$	0.5143	0.0285	0.0849	-0.5117	0.0235	0.0807	1.0363	0.0363	0.1226	-1.0260	0.0260	0.1156
$N, T = 40$		$\theta_0 = 0.5$			$\theta_0 = -0.5$			$\theta_0 = 1$			$\theta_0 = -1$	
$\widehat{\theta}$	0.5253	0.0506	0.0502	-0.5256	0.0512	0.0474	1.0668	0.0668	0.0888	-1.0601	0.0601	0.0825
$\widetilde{\theta}$	0.5015	0.0031	0.0412	-0.5048	0.0095	0.0384	1.0096	0.0096	0.0550	-1.0076	0.0076	0.0525
$N, T = 80$		$\theta_0 = 0.5$			$\theta_0 = -0.5$			$\theta_0 = 1$			$\theta_0 = -1$	
$\widehat{\theta}$	0.5123	0.0246	0.0233	-0.5105	0.0211	0.0213	1.0302	0.0302	0.0410	-1.0280	0.0280	0.0377
$\widetilde{\theta}$	0.5003	0.0006	0.0194	-0.5002	0.0005	0.0181	1.0021	0.0021	0.0266	-1.0018	0.0018	0.0244

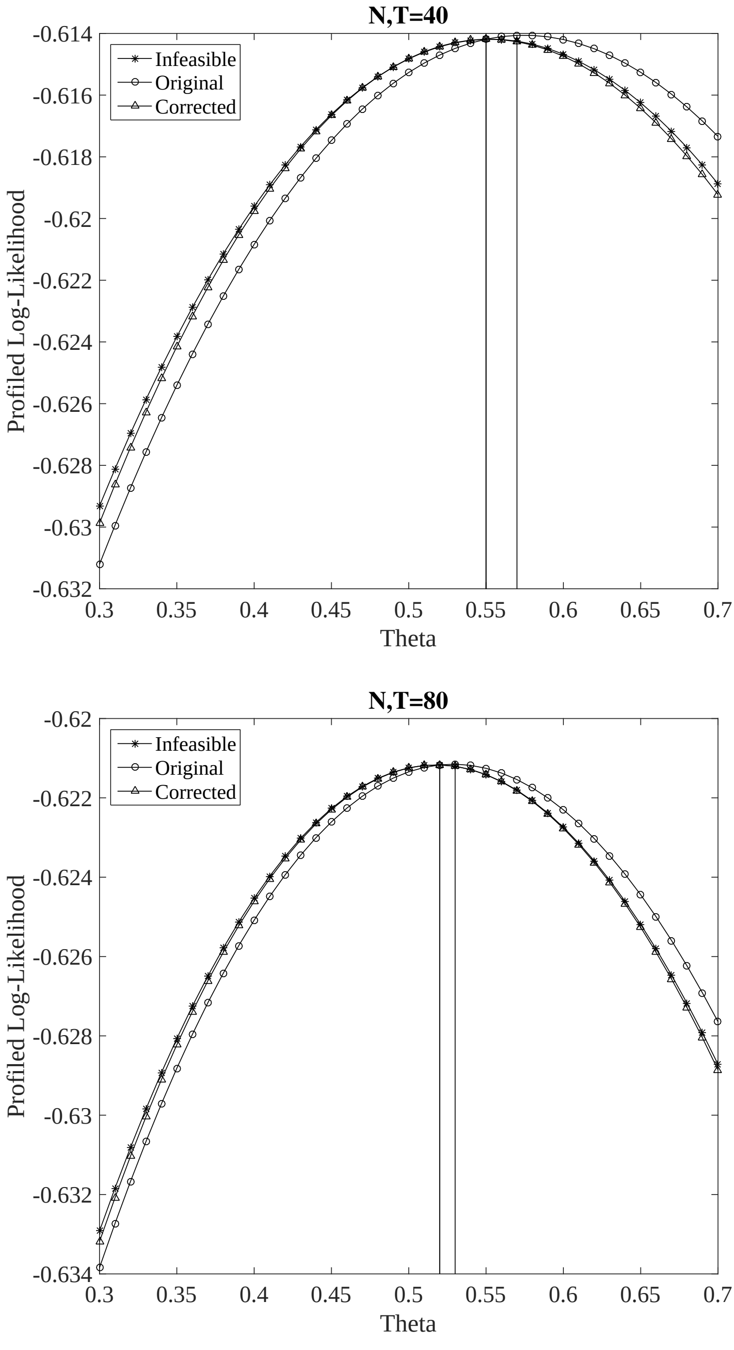
Notes: Bias is presented relative to θ_0. The number of replications is 1,000. Model: $Y_{it} = 1(X_{it}\theta_0 + \alpha_i + \gamma_t + \varepsilon_{it} \geq 0)$ where $\varepsilon_{it} \sim \mathcal{N}(0,1)$ and $X_{it} \sim \mathcal{N}(\alpha_i + \gamma_t, 1)$ with $\alpha_i \sim \mathcal{N}(0, 1/16)$ and $\gamma_t \sim \mathcal{N}(0, 1/16)$. $\widehat{\theta}$ is the original estimate, $\widetilde{\theta}$ is the bias-corrected estimate.

Figure 3: Double IPP - Plot of Profiled Log-likelihood for Probit - Part 1



Notes: Computed on a single simulated dataset. Model: $Y_{it} = 1\left(X_{it}\theta_0 + \alpha_i + \gamma_t + \varepsilon_{it} \geq 0\right)$ where $\varepsilon_{it} \sim \mathcal{N}(0,1)$, $X_{it} \sim \mathcal{N}(0,1)$, $\alpha_i = \gamma_t = 0$, and $\theta_0 = 0.5$. θ chosen from the region depicted on the horizontal axis with a step of 0.01. Circle: $\widehat{L}(\theta)$; triangle: $\widetilde{L}(\theta)$; and asterisk: $L(\theta)$. All curves are vertically shifted such that they coincide at $\overline{\theta}$ (maximizer of the infeasible log-likelihood). Vertical lines at maximizers.

Figure 4: Double IPP - Plot of Profiled Log-likelihood for Probit - Part 2



Notes: Computed on a single simulated dataset. Model: $Y_{it} = 1\left(X_{it}\theta_0 + \alpha_i + \gamma_t + \varepsilon_{it} \geq 0\right)$ where $\varepsilon_{it} \sim \mathcal{N}(0,1)$, $X_{it} \sim \mathcal{N}(0,1)$, $\alpha_i = \gamma_t = 0$, and $\theta_0 = 0.5$. θ chosen from the region depicted on the horizontal axis with a step of 0.01. Circle: $\widehat{L}(\theta)$; triangle: $\widetilde{L}(\theta)$; and asterisk: $L(\theta)$. All curves are vertically shifted such that they coincide at $\overline{\theta}$ (maximizer of the infeasible log-likelihood). Vertical lines at maximizers.

5 Conclusion

The estimator $\widehat{\theta}$ of the parameter that is common to all observations in a nonlinear fixed-effect model with both individual and time effect could contain a substantial bias. When N/T converges to a constant, the bias enters the asymptotic distribution of $\sqrt{NT}\left(\widehat{\theta}-\theta_0\right)$ such that the distribution is not centered at 0. We propose a likelihood-based bias correction technique that eliminates such bias to the first order. We focus on a simple setting where Y_{it} is static and the model contains only individual and time effects, and we show that our method is effective, given the large variation, in correcting the bias even when N and T are still small. Our method does not impose restrictions on how the effects enter the model, and therefore, covers a very general class of models in which the individual and time effects do not enter additively.

In addition, we briefly discuss the incorporation of dynamic models where Y_{it} are correlated across i and t and the accommodation of models with more than two sets of fixed effects. However, these discussions are very brief in the sense that further research may be necessary. For instance, we argue that dynamic models can be implemented with our correction technique provided that the observation-level scores are averaged with the Bartlett kernel weight. Such weight is optimal in the setting where only individual effects are present but is not guaranteed to remain optimal when both individual and time effects are included. In addition, we have not investigated the effect of different choices in the weights used to average the observation-level score.

Alternatively, one may also wonder if a higher-order approximation of $L(\theta)$ can be derived. While this may be worth studying, such a correction may be difficult to derive. To see this, suppose that $\widehat{L}$ follows an asymptotic expansion

$$\widehat{L}(\theta)=L(\theta)+\frac{B_1(\theta)}{T}+\frac{D_1(\theta)}{N}+\frac{W(\theta)}{NT}+\frac{B_2(\theta)}{T^2}+\frac{D_2(\theta)}{N^2}+\cdots$$

for some $B_j(\theta)$ and $D_j(\theta)$ defined in a similar way as $B(\theta)$ and $D(\theta)$ and some $W(\theta)$ depending only on θ. Here the existence of $B_j(\theta)$ and $D_j(\theta)$ is due to the inclusion of individual and time effects whereas the existence of $W(\theta)$ is due to the fact that $\widehat{L}(\theta)$ is, in general, nonlinear in θ, i.e., the log-likelihood function would still contain a bias (away from the expected value) even when the individual and time effects were not included. Our method essentially eliminates $B_1(\theta)$ and $D_1(\theta)$ while $W(\theta)$ is left untreated. For a higher-order bias correction technique, $B_1(\theta)$, $D_1(\theta)$, $B_2(\theta)$, $D_2(\theta)$, and as well as $W(\theta)$ must all be eliminated.

Beyond the proposed directions, further studies may be conducted for, e.g., a variance estimator of $\widetilde{\theta}$ that possesses more desired finite-sample properties than the standard ML variance; or, e.g., how $\widetilde{L}(\theta)$ would benefit inferences based on likelihood such as the likelihood ratio test.

References

Alvarez, J. and M. Arellano (2003). The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica 71*(4), 1121–1159.

Arellano, M. and S. Bonhomme (2009). Robust priors in nonlinear panel data models. *Econometrica*, 489–536.

Arellano, M. and J. Hahn (2005). Understanding bias in nonlinear panel models: Some recent developments. *Documentos de Trabajo (CEMFI)* (7), 1.

Arellano, M. and J. Hahn (2006). A likelihood-based approximate solution to the incidental parameter problem in dynamic nonlinear models with multiple effects. *Documentos de Trabajo (CEMFI) 13*(13), 1–49.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica 77*(4), 1229–1279.

Box, M. (1971). Bias in nonlinear estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 171–201.

Charbonneau, K. B. (2014). Multiple fixed effects in binary response panel data models. Technical report, Bank of Canada Working Paper.

Chen, M., I. Fernandez-Val, and M. Weidner (2014). Nonlinear panel models with interactive effects. *arXiv preprint arXiv:1412.5647*.

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

Cox, D. R. and E. J. Snell (1968). A general definition of residuals. *Journal of the Royal Statistical Society. Series B (Methodological)*, 248–275.

Dhaene, G. and K. Jochmans (2015). Split-panel jackknife estimation of fixed-effect models. *The Review of Economic Studies 82*(3), 991–1030.

Fernández-Val, I. (2009). Fixed effects estimation of structural parameters and marginal effects in panel probit models. *Journal of Econometrics 150*(1), 71–85.

Fernández-Val, I. and M. Weidner (2016). Individual and time effects in nonlinear panel models with large N,T. *Journal of Econometrics*.

Greene, W. (2004). Fixed effects and bias due to the incidental parameters problem in the tobit model. *Econometric Reviews 23*(2), 125–147.

Hahn, J. and G. Kuersteiner (2011). Bias reduction for dynamic nonlinear panel models with fixed effects. *Econometric Theory 27*(06), 1152–1191.

Hahn, J., G. Kuersteiner, and U. U. Davis (2007). Bandwidth choice for bias estimators in dynamic nonlinear panel models. Technical report, Citeseer.

Hahn, J. and W. Newey (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica 72*(4), 1295–1319.

Honoré, B. E. (1993). Orthogonality conditions for tobit models with fixed effects and lagged dependent variables. *Journal of Econometrics 59*(1), 35–61.

Hsiao, C., M. H. Pesaran, and A. K. Tahmiscioglu (2002). Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of econometrics 109*(1), 107–150.

Lancaster, T. (2000). The incidental parameter problem since 1948. *Journal of Econometrics 95*(2), 391–413.

Lancaster, T. (2002). Orthogonal parameters and panel data. *The Review of Economic Studies 69*(3), 647–666.

Moreira, M. (2008). A maximum likelihood method for the incidental parameter problem. Technical report, National Bureau of Economic Research.

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

Okui, R. (2010). Asymptotically unbiased estimation of autocovariances and autocorrelations with long panel data. *Econometric Theory 26*(05), 1263–1304.