KU LEUVEN

DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES
DPS16.06

MAY 2016

Second-order corrected likelihood for nonlinear models with fixed effects

Yutao SUN

Econometrics

Faculty of Economics
and Business

Second-order corrected likelihood for nonlinear models with fixed effects

Yutao Sun*
KU Leuven

Abstract

We introduce a second-order correction technique for nonlinear fixed-effect models exposed to the incidental parameter problem. This technique produces a bias-corrected log-likelihood function that possesses a bias only to the order (in expectation) of $O\left(T^{-3}\right)$ where T is the number of time periods. As a consequence, the maximizer of the corrected log-likelihood, the corrected estimator, is also only biased to the order of $O\left(T^{-3}\right)$. The technique applies to static nonlinear fixed-effect models in which N, the number of individuals, is allowed to grow rapidly and T is assumed to grow at a rate satisfying N/T^5 converging to 0. The proposed technique is general in the sense that it does not depend on a specific functional form of the log-likelihood function.

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

1 Introduction

Panel data are becoming ever more important in economic studies. In these studies, researchers often attempt to capture individual heterogeneity by introducing individual-specific parameters, or "fixed effects", in the model. When the number of time periods in the dataset is small, however, nonlinear models with a large number of fixed-effect parameters may produce maximum likelihood (ML) estimates that are severely biased. This is known as the incidental parameter problem (IPP) of Neyman and Scott (1948). To briefly introduce[1] the problem, let $i = 1, \cdots, N$ index the individuals and $t = 1, \cdots, T$ the time periods, and suppose that $N \to \infty$ while T remains fixed. Denote $\log f\left(Y_{it}; \theta, a_i\right)$ to be the log-likelihood associated with observation Y_{it} (possibly conditional on covariates X_{it}), where θ is the vector of parameters of interest that applies to all observations it and a_i is the fixed-effect nuisance parameter. Here $\widehat{\alpha}_i$, the ML estimator of a_i, only uses the data from the ith individual. Therefore, when T is fixed, $\widehat{\alpha}_i$ remains a random variable for every i even as $N \to \infty$. In many models, this introduces a bias in the log-likelihood function in the sense that the ML estimator $\widehat{\theta}$ of θ converges to a point $\theta_T \neq \theta_0$ where θ_0 is the true value of θ. When, however, $N, T \to \infty$ with T increasing only slowly, the random variation in $\widehat{\alpha}_i$ vanishes only slowly such that $\widehat{\theta}$ inherits the slow convergence and the limiting distribution of $\widehat{\theta}$ is not centered

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

[1]See also Lancaster (2000) for a comprehensive survey.

at θ_0. In the course of nearly seven decades since IPP was discovered, numerous researchers have attempted to obtain solutions, either exact or approximate, as well as either analytical or numerical, to the IPP. In the early years, solutions are usually model-specific or depending on conditions that may be considered strict. For example Cox and Reid (1987) propose a consistent estimator of θ when a certain type of orthogonality between θ and a_i can be found. Here the problem is that the existence of this type of orthogonality is generally not guaranteed.

On the other hand, many researchers look at general solutions that do not depend on the specific functional form of the underlying density. Such solutions are often approximate in the sense that they produce bias-corrected estimates that are unbiased to some specific order of magnitude. One way to obtain a bias correction is to adopt certain automated approaches such as the jackknife or the bootstrap. For example Dhaene and Jochmans (2015) propose the split-panel jackknife for both dynamic and static models. The approach splits the panels into two non-overlapping subpanels along the time dimension. The bias in $\widehat{\theta}$ is then estimated on each of the two subpanels and subtracted from $\widehat{\theta}$. While automated methods are usually easy to construct, the computational stress of these methods may be large depending on the structure of the dataset. Another way alternative to the automated approaches is to derive an analytical formula approximating the bias in $\widehat{\theta}$. The formula is estimated using the data and is subsequently subtracted from $\widehat{\theta}$, producing a bias-corrected estimate. This type of corrections includes, e.g., Hahn and Newey (2004) who develop a correction to estimators under independent observations, and Hahn and Kuersteiner (2011) who derive a formula for dynamic nonlinear panel models. These methods are usually less computationally intensive and possess finite-sample properties that are more desirable.

Alternatively, a correction performed on the objective function, i.e., the log-likelihood function, may also be of interest. Arellano and Hahn (2006) introduce an approximation to an infeasible log-likelihood function that is not exposed to the IPP. The proposed approximation is biased only to the order of magnitude of $O\left(T^{-2}\right)$ in expectation, and therefore, the maximizer in θ of the approximating function serves as a bias-corrected estimator that is biased only to the order of magnitude of $O\left(T^{-2}\right)$ in expectation. The proposed approximation is often called the first-order corrected likelihood and, when T is small, the corrected estimator obtained from maximizing the corrected log-likelihood may still be significantly biased, since the term that is of the order of magnitude of $O\left(T^{-2}\right)$ in expectation may still be large in magnitude. A possible way to overcome this situation is to seek a refined approximation that is biased to the order of magnitude of $O\left(T^{-3}\right)$ in expectation. The refined approximation is often called the second-order corrected likelihood. The corrected estimator obtained subsequently from maximizing the second-order corrected log-likelihood is then also biased to the order of magnitude of $O\left(T^{-3}\right)$ in expectation, which is of a higher order. The second-order corrected estimator is consistent and asymptotically unbiased under the asymptotic sequence $N/T^5 \to 0$ as $N, T \to \infty$. As a comparison, the required conditions are $N/T \to 0$ for the original estimator $\widehat{\theta}$ and $N/T^3 \to 0$ for the first-order corrected estimator. We develop the second-order corrected log-likelihood by extending the approach introduced by Arellano and Hahn (2006). Our second-order corrected likelihood can be applied to a general class of models provided that some mild assumptions are satisfied. The proposed corrected log-likelihood depends only on known quantities such as $\widehat{\alpha}_i$ and Y_{it} and hence, can be constructed in a straightforward way using the given data.

The rest of this paper is organized in the following way. In section 2, we introduce the settings and assumptions that are required for the derivation of the second-order corrected likelihood. In section 3, we provide the approach for obtaining the second-order corrected likelihood. In this section, we first revisit the derivation of Arellano and Hahn (2006) in order to introduce the differences and difficulties in the derivation of the second-order corrected log-likelihood. Next, we formally derive the second-order corrected likelihood. In section 4, we provide suggestive examples and simulations regarding the application of the second-order corrected likelihood under various models and designs. In section 5, we leave concluding remarks by briefly introducing possible routes for further studies.

2 Preliminary

Let Y_{it} denote the itth observation where $i = 1, \cdots, N$ and $t = 1, \cdots, T$ with $N/T^5 \to 0$ as $N, T \to \infty$. We assume that Y_{it} are independent across i and t, i.e., we restrict our study to the static models. Let $f(Y_{it}; \theta, a_i)$ be the conditional density of Y_{it} where θ is the parameter of interest that is the same for every Y_{it} and a_i is the fixed-effect parameter which can only be estimated from the ith individual. Let

$$\begin{aligned} \alpha_i(\theta) &\equiv \arg\max_{a_i} \frac{1}{T} \sum\nolimits_t \mathbb{E} \log f(Y_{it}; \theta, a_i), \\ \widehat{\alpha}_i(\theta) &\equiv \arg\max_{a_i} \frac{1}{T} \sum\nolimits_t \log f(Y_{it}; \theta, a_i). \end{aligned}$$

We make the following assumptions about the density and about $\alpha_i(\theta)$ and $\widehat{\alpha}_i(\theta)$.

Assumption 1. *Suppose* $(\theta, \alpha_1(\theta), \cdots, \alpha_N(\theta), \widehat{\alpha}_1(\theta), \cdots, \widehat{\alpha}_N(\theta)) \in \operatorname{int}(\Theta \times A^{2N})$ *where* $\Theta \times A^{2N}$ *is compact and* $\operatorname{int}(\cdot)$ *denotes the interior of a set.*

1. *For every* θ, $\alpha_i(\theta)$ *and* $\widehat{\alpha}_i(\theta)$ *are unique.*
2. *For every* θ *and every nonnegative integer* $r \leq 4$, $\nabla_{a_i}^r \log f(Y_{it}; \theta, a_i)$ *exists and satisfies*

$$\left| \nabla_{a_i}^r \log f(Y_{it}; \theta, a_i) \right| < \infty$$

for every $\alpha_i(\theta)$ *and* $\widehat{\alpha}_i(\theta)$ *where* $\nabla_{a_i}^r$ *denotes the* r*th derivative w.r.t.* a_i.

Assumption 2. *The second derivative of* $\log f(Y_{it}; \theta, a_i)$ *w.r.t.* a_i *satisfies*

$$\frac{1}{T} \sum_t \nabla_{a_i}^2 \log f(Y_{it}; \theta, a_i) < 0$$

for every

$$a_i \in \{\alpha_i(\theta), \widehat{\alpha}_i(\theta)\}.$$

Whereas assumption 1 is standard, assumption 2 - strict concavity of the likelihood - may deserve some extra words. In general, this is an acceptable assumption - see, e.g., Newey and McFadden (1994, chap. 35). However, there are cases where complications arise. Consider the probit model with $Y_{it} = 1$ for all t in some specific i. Under this situation, $\max_{a_i} 1/T \sum_t \log f(Y_{it}; \theta, a_i)$ is achieved at $a_i \to \infty$ irrespectively of θ such that $\lim_{a_i \to \infty} 1/T \sum_t \nabla^2_{a_i} \log f(Y_{it}; \theta, a_i) \to_p 0$. We regard this as somewhat nonstandard and hence exclude detailed discussions about this situation. We will, however, point out the exact places where this assumption is inevitable - see remark 1, 2, and 3 below.

Next, for every arbitrarily given i (hence the index i is omitted) and θ, let

$$
\begin{aligned}
l(\theta) &\equiv \frac{1}{T} \sum\nolimits_t \log f(Y_{it}; \theta, \alpha_i(\theta)), \\
\widehat{l}(\theta) &\equiv \frac{1}{T} \sum\nolimits_t \log f(Y_{it}; \theta, \widehat{\alpha}_i(\theta)).
\end{aligned}
$$

When T is fixed, $\widehat{\alpha}_i(\theta)$ remains a random variable for every i even as $N \to \infty$. In this setting, the random variation in $\widehat{\alpha}_i(\theta)$ does not vanish as $N \to \infty$. In many models, this induces a bias to the log-likelihood function in the sense that the ML estimator of θ, $\widehat{\theta} \equiv \arg\max_\theta 1/N \sum_i \widehat{l}(\theta)$, converges to a wrong value, say θ_T; i.e., $\text{plim}_{N\to\infty} \widehat{\theta} = \theta_T \neq \theta_0$ where $\theta_T \equiv \text{plim}_{N\to\infty} \arg\max_\theta 1/N \sum_i \widehat{l}(\theta)$ and $\theta_0 \equiv \text{plim}_{N\to\infty} \arg\max_\theta 1/N \sum_i l(\theta)$. When, however, $N, T \to \infty$ with T increasing much slower than N, the random variation in $\widehat{\alpha}_i(\theta)$ vanishes only slowly. In that case, $\widehat{\theta}$ inherits this slow convergence such that the limiting distribution of $\widehat{\theta}$ is not centered at θ_0.

In what follows, we assume that the expectation exists and that the stochastic order operator and the expectation can be interchanged. Here it is obvious that $l(\theta)$ is not exposed to IPP and hence, can be thought of as an infeasible target log-likelihood function to which an approximation - say $\widehat{l}^{(k)}(\theta)$ - can be constructed, where $\widehat{l}^{(k)}(\theta)$ satisfies

$$
\mathbb{E} l(\theta) = \mathbb{E} \widehat{l}^{(k)}(\theta) + o\left(T^{-k}\right)
$$

in which $\mathbb{E}(\cdot)$ denotes the expectation under the true density $f(Y_{it}; \theta_0, \alpha_i(\theta_0))$ and $\widehat{l}^{(k)}(\theta)$ depends on $\widehat{\alpha}_i(\theta)$ instead of $\alpha_i(\theta)$. This approximation may then serve as a corrected log-likelihood function such that a less biased estimator of θ_0 may be constructed simply as

$$
\widehat{\theta}^{(k)} \equiv \arg\max_\theta \frac{1}{N} \sum_i \widehat{l}^{(k)}(\theta).
$$

Arellano and Hahn (2006) provide the approximating function for $k = 1$, the first order, which takes the form (for a single i)

$$
\widehat{l}^{(1)}(\theta) \equiv \widehat{l}(\theta) + \frac{\widehat{b}_1}{T}
$$

in which $\widehat{b}_1$ denotes some function evaluated at $\widehat{\alpha}_i(\theta)$. The estimator of θ derived from $\widehat{l}^{(1)}(\theta)$ is then biased only to the order of $o\left(T^{-1}\right)$. When the higher-order bias term $o\left(T^{-1}\right)$ is not negligible, a refined approximation for, e.g., $k = 2$ must be constructed. The approximation should take the form

$$
\widehat{l}^{(2)}(\theta) \equiv \widehat{l}(\theta) + \frac{\widehat{b}_1}{T} + \frac{\widehat{b}_2}{T^2}
$$

where $\widehat{b}_2$, similar to $\widehat{b}_1$, is a function evaluated at $\widehat{\alpha}_i(\theta)$.

3 Bias Correction

3.1 Review of the First-order Correction

To derive $\widehat{b}_2$, it may be useful to study the derivation of $\widehat{b}_1$, i.e., to first replicate the result of Arellano and Hahn (2006). Let

$$
\begin{aligned}
l_r &\equiv l_r(\theta) \equiv \frac{1}{T}\sum\nolimits_t \nabla_a^r \log f(Y_{it};\theta,a_i)\big|_{a_i=\alpha_i(\theta)}, \\
\widehat{l}_r &\equiv \widehat{l}_r(\theta) \equiv \frac{1}{T}\sum\nolimits_t \nabla_a^r \log f(Y_{it};\theta,a_i)\big|_{a_i=\widehat{\alpha}_i(\theta)};
\end{aligned}
$$

and, for simplicity, let $l \equiv l(\theta)$, $\widehat{l} \equiv \widehat{l}(\theta)$, $\alpha = \alpha_i(\theta)$, and $\widehat{\alpha} = \widehat{\alpha}_i(\theta)$. Observing

$$
\widehat{\alpha}-\alpha = O_p\left(T^{-\frac{1}{2}}\right), \qquad l_1 = O_p\left(T^{-\frac{1}{2}}\right), \qquad l_2 = O_p(1);
$$

and, for a regular problem, $\widehat{l}$ can be Taylor-expanded around α,

$$
\begin{aligned}
\widehat{l} &= l + l_1(\widehat{\alpha}-\alpha) + \frac{1}{2}l_2(\widehat{\alpha}-\alpha)^2 + O_p\left(T^{-\frac{3}{2}}\right) \\
l &= \widehat{l} - l_1(\widehat{\alpha}-\alpha) - \frac{1}{2}l_2(\widehat{\alpha}-\alpha)^2 + O_p\left(T^{-\frac{3}{2}}\right).
\end{aligned} \tag{3.1}
$$

Here for the case where $k=2$, equation (3.1) needs to be extended to the order of $O_p\left(T^{-2}\right)$. This is fairly straightforward. Similarly[2], $\widehat{l}_1 = 0$ can be Taylor-expanded around α,

$$
0 = l_1 + l_2(\widehat{\alpha}-\alpha) + O_p\left(T^{-1}\right) \tag{3.2}
$$

where, as $l_2 < 0$,

$$
(\widehat{\alpha}-\alpha) = -\frac{l_1}{l_2} + O_p\left(T^{-1}\right). \tag{3.3}
$$

Note that assumption 1 itself does not guarantee that $\widehat{l}$ and $\widehat{l}_1$ are analytic and hence Taylor-expandable. That is, there are functions that are infinitely differentiable but are nowhere analytic. See Hille (2005, chap. 10) for exact conditions of analyticity and, e.g., Darst (1973) for proof that most infinitely differentiable functions are nowhere analytic. Nevertheless, in the likelihood context, $\widehat{\alpha}-\alpha = O_p\left(T^{-1/2}\right)$ ensures that equation (3.1) and (3.2) are convergent as $T \to \infty$.

Remark 1. *Assumption 2 plays a role here. If l_2 were not guaranteed to be nonzero, equation (3.3) would be undefined. It is possible to replace l_2 with $\mathbb{E}l_2$ to avoid this situation. This, however, only postpones the problem to a later stage - see remark 2 below.*

For the case where $k=2$, equation (3.2) must be extended to the order of $O_p\left(T^{-3/2}\right)$ such that a higher-order version of equation (3.3) can be constructed. This is only marginally difficult, since a technique similar to Pace and Salvan (1997, chap. 9.3) could be adopted to derive the expansion of $(\widehat{\alpha}-\alpha)$ in order to produce a higher-order version of equation (3.3).

[2] See also Cox and Snell (1968).

The technique will be introduced in section 3.2. Next, combine equation (3.1) and (3.3),

$$\begin{aligned} l &= \widehat{l} - l_1 \left(-\frac{l_1}{l_2}\right) - \frac{1}{2} l_2 \left(-\frac{l_1}{l_2}\right)^2 + O_p\left(T^{-\frac{3}{2}}\right) \\ &= \widehat{l} + \frac{1}{2}\frac{l_1^2}{l_2} + O_p\left(T^{-\frac{3}{2}}\right). \end{aligned}$$

For $k = 2$, a combination of the higher-order expansion of $(\widehat{\alpha} - \alpha)$ and the extended version of equation (3.1) must be computed. The requires rising a polynomial to the power of 4. This is slightly difficult, since the technique proposed by Provost and Ratemi (2011) may be invoked. Additionally, a simplification to this calculation will be introduced in section 3.3. Here note that, as $\widehat{l}_1 = 0$, if one replaces l_1^2 with $\widehat{l}_1^2$, the ratio l_1^2/l_2 disappears completely. Therefore, a refined version of l_1^2/l_2 needs to be constructed. By the definition of l_1,

$$\begin{aligned} l_1^2 &= \frac{1}{T^2}\sum_t \left(\nabla_a \log f\left(Y_{it};\theta,a\right)\right)^2\Big|_{a=\alpha} \\ &\quad + \frac{1}{T^2}\sum_{t_1\neq t_2} \nabla_a \log f\left(Y_{it_1};\theta,a\right)\nabla_a \log f\left(Y_{it_2};\theta,a\right)\Big|_{a=\alpha} \end{aligned}$$

where, as Y_{it} are independent,

$$\mathbb{E}\sum_{t_1\neq t_2} \nabla_a \log f\left(Y_{it_1};\theta,a\right)\nabla_a \log f\left(Y_{it_2};\theta,a\right)\Big|_{a=\alpha} = 0 \tag{3.4}$$

such that

$$\mathbb{E}l = \mathbb{E}\widehat{l} + \frac{\mathbb{E}b_1}{T} + O\left(T^{-2}\right)$$

with

$$b_1 = \frac{1/T\sum_t \left(\nabla_a \log f\left(Y_{it};\theta,a\right)\right)^2\big|_{a=\alpha}}{2l_2}.$$

For $k = 2$, the identification of terms having zero expectation, as similar to equation (3.4), is necessary. This is mathematically involved. We provide additional notations and propositions that may ease the difficulty. They are given in section 3.3. Here replacing b_1 with $\widehat{b}_1$, b_1 evaluated at $\widehat{\alpha}$, will introduce a bias since, typically,

$$\mathbb{E}\widehat{b}_1 = \mathbb{E}b_1 + O\left(T^{-1}\right).$$

However, for $k = 1$, this bias can be neglected, as

$$\begin{aligned} \mathbb{E}l &= \mathbb{E}\widehat{l} + \frac{\mathbb{E}b_1}{T} + O\left(T^{-2}\right) \\ &= \mathbb{E}\widehat{l} + \frac{\mathbb{E}b_1}{T} + \frac{1}{T}O\left(T^{-1}\right) + O\left(T^{-2}\right) \\ &= \mathbb{E}\widehat{l} + \frac{\mathbb{E}\widehat{b}_1}{T} + O\left(T^{-2}\right) \end{aligned}$$

from which the first-order corrected likelihood then follows as

$$\widehat{l}^{(1)}(\theta) \equiv \widehat{l} + \frac{\widehat{b}_1}{T}.$$

$\widehat{l}^{(1)}(\theta)$ can be constructed easily from the sample since $\widehat{l}$ and $\widehat{b}_1$ depend only on known quantities $\widehat{\alpha}_i$, θ, and Y_{it}.

Remark 2. *Assumption 2 is important for the first-order corrected likelihood to be defined; i.e., $\widehat{b}_1$ contains $\widehat{l}_2$ in the denominator such that $\widehat{l}_2 < 0$ for every θ must be guaranteed.*

As described above, since $\mathbb{E}\widehat{b}_1 = \mathbb{E}b_1 + O\left(T^{-1}\right)$, replacing $\mathbb{E}b_1$ with $\mathbb{E}\widehat{b}_1$ will introduce a bias to the order of $O\left(T^{-2}\right)$. This bias was negligible for the case where $k = 1$. For $k = 2$, this must be taken into account. To deal with this situation, a procedure similar to the one dealing with the pair $\left(l, \widehat{l}\right)$ must also be applied on the pair $\left(b_1, \widehat{b}_1\right)$. This is particularly involved. However, since the procedure dealing with the pair $\left(l, \widehat{l}\right)$ is identical to the one dealing with $\left(b_1, \widehat{b}_1\right)$, we omit, to a very large extent, the details in the derivation of the latter. The result will be given in section 3.4.

3.2 Stochastic Expansion of Fixed-effect Estimator

For the second-order correction, equation (3.2) must be continued to include the term with an order of $O_p\left(T^{-3/2}\right)$, i.e.,

$$\begin{aligned} 0 &= l_1 + l_2(\widehat{\alpha}-\alpha) + \frac{1}{2!}l_3(\widehat{\alpha}-\alpha)^2 + \frac{1}{3!}l_4(\widehat{\alpha}-\alpha)^3 + O_p\left(T^{-2}\right) \\ (\widehat{\alpha}-\alpha) &= -\frac{l_1}{l_2} - \frac{1}{2!}\frac{l_3}{l_2}(\widehat{\alpha}-\alpha)^2 - \frac{1}{3!}\frac{l_4}{l_2}(\widehat{\alpha}-\alpha)^3 + O_p\left(T^{-2}\right). \end{aligned} \tag{3.5}$$

Here, as $l_2 < 0$ is assumed, the ratios in equation (3.5) are well-defined. Next, consider

$$(\widehat{\alpha}-\alpha) = a_{1/2} + a_{2/2} + a_{3/2} + O_p\left(T^{-2}\right) \tag{3.6}$$

in which $a_{j/2}$ are unknown random variables satisfying $a_{j/2} = O_p\left(T^{-j/2}\right)$ and $a_{j/2} \neq o_p\left(T^{-j/2}\right)$. $a_{j/2}$ can be solved via a recursive procedure similar to the one described in Pace and Salvan (1997, chap. 9.3). Combining equation (3.5) and (3.6) leads to

$$\begin{aligned} & a_{1/2} + a_{2/2} + a_{3/2} \\ = & -\frac{l_1}{l_2} - \frac{1}{2!}\frac{l_3}{l_2}\left(a_{1/2}+a_{2/2}+a_{3/2}\right)^2 - \frac{1}{3!}\frac{l_4}{l_2}\left(a_{1/2}+a_{2/2}+a_{3/2}\right)^3 + O_p\left(T^{-2}\right) \\ = & -\frac{l_1}{l_2} - \frac{1}{2!}\frac{l_3}{l_2}\left(a_{1/2}^2 + 2a_{1/2}a_{2/2}\right) - \frac{1}{3!}\frac{l_4}{l_2}a_{1/2}^3 + O_p\left(T^{-2}\right). \end{aligned}$$

Collecting the terms on the left-hand side and the right-hand side by their stochastic order,

$$a_{1/2} = -\frac{l_1}{l_2}, \qquad a_{2/2} = -\frac{l_3}{2l_2}a_{1/2}^2, \qquad a_{3/2} = -\frac{l_3}{l_2}a_{1/2}a_{2/2} - \frac{1}{6}\frac{l_4}{l_2}a_{1/2}^3$$

and, after a recursive substitution,

$$a_{1/2} = -\frac{l_1}{l_2}, \qquad a_{2/2} = -\frac{l_1^2 l_3}{2l_2^3}, \qquad a_{3/2} = -\frac{l_3^2 l_1^3}{2l_2^5} + \frac{l_1^3 l_4}{6l_2^4}.$$

The order of each term can be identified easily. As $l_1 = O_p\left(T^{-1/2}\right)$ and $l_r = O_p(1)$ for $1 < r \leq 3$, it is apparent that $a_{1/2} = O_p\left(T^{-1/2}\right)$, $a_{2/2} = O_p\left(T^{-1}\right)$, and $a_{3/2} = O_p\left(T^{-3/2}\right)$ and, when $l_1 \neq o_p\left(T^{-1/2}\right)$ and $l_r \neq o_p(1)$ for $1 < r \leq 3$, $a_{1/2} \neq o_p\left(T^{-1/2}\right)$, $a_{2/2} \neq o_p\left(T^{-1}\right)$, and $a_{3/2} \neq o_p\left(T^{-3/2}\right)$. Combining equation (3.6) with $a_{1/2}$ to $a_{3/2}$,

$$(\widehat{\alpha} - \alpha) = -\frac{l_1}{l_2} - \frac{l_1^2 l_3}{2l_2^3} + \frac{l_1^3 l_4}{6l_2^4} - \frac{l_3^2 l_1^3}{2l_2^5} + O_p\left(T^{-2}\right). \tag{3.7}$$

A caution that must be observed is that, if $\widehat{\alpha}$ is plugged in, equation (3.7) is not yet accurate for the calculation of a second-order corrected ML estimate (viewing $\widehat{\alpha}$ and α as generic ML estimators). This is because of two complications. First, as $\mathbb{E}\widehat{a}_{1/2} = \mathbb{E}a_{1/2} + O\left(T^{-1}\right)$, the plug-in version of equation (3.7) possesses a bias to the order, in expectation, of $O\left(T^{-1}\right)$, which is larger than the targeted (in expectation) $O\left(T^{-2}\right)$. Second, in a regular problem, the term that is to the order of $O_p\left(T^{-2}\right)$ must also be included such that the second-order corrected ML estimate is unbiased, in expectation, to the order of $O\left(T^{-2}\right)$; i.e., equation (3.7) must be extended to an additional order such that the remainder term is of the order of $O_p\left(T^{-5/2}\right)$. For the computation of a bias-corrected ML estimate, see Ferrari et al. (1996), for instance. Whereas we do not deal with the first problem, the second point described above can easily be solved. The recursive substitution procedure can be continued to produce an arbitrary order expansion of $(\widehat{\alpha} - \alpha)$. In appendix A, we list the first 8 terms of the expansion. In addition, the expansion of $(\widehat{\alpha} - \alpha)$ is not unique. Other versions include, e.g., Bartlett (1953a and 1953b), Haldane and Smith (1956), and Rilstone et al. (1996).

Furthermore, by the definition of $a_{j/2}$, it is straightforward that $\mathbb{E}a_{j/2} \neq 0$ in general. This reflects the fact that the ML estimator of a nonlinear model is often biased - see Box (1971) for details. However, equation (3.7) can be transformed such that the term that is of the order of $O_p\left(T^{-1/2}\right)$ has a zero expected value; i.e., it is possible to replace $a_{1/2}$ with $-l_1/\mathbb{E}l_2 + b$ for some $b = O_p\left(T^{-1}\right)$. The transformed version is in line with the asymptotic theory of the ML estimator. That is, $\sqrt{T}(\widehat{\alpha} - \alpha) \to_d \mathcal{N}(0, \Sigma)$, where $\mathcal{N}(\cdot, \cdot)$ is the normal density and Σ is the asymptotic variance. As $T \to \infty$, $\sqrt{T}(\widehat{\alpha} - \alpha) \to_p -\sqrt{T} l_1/\mathbb{E}l_2$ where $\mathbb{E}(l_1/\mathbb{E}l_2) = 0$ and, by the central limit theorem, $\sqrt{T} l_1 \to_d \mathcal{N}(0, \Omega)$, where Ω is the asymptotic variance of $\sqrt{T} l_1$. This implies $\sqrt{T}(\widehat{\alpha} - \alpha) \to_d \mathcal{N}(0, \Sigma)$ with Σ determined by Ω and $\mathbb{E}l_2$.

3.3 Stochastic Expansion of Likelihood

In this section, we seek to obtain an expansion of the likelihood,

$$\mathbb{E}l = \mathbb{E}\widehat{l} + \frac{\mathbb{E}b_1}{T} + \frac{\mathbb{E}b_2'}{T^2} + O\left(T^{-3}\right)$$

in which b_1 and b_2' are $O_p(1)$. To do so, first Taylor-expand $\widehat{l}$ around α,

$$\begin{aligned}
\widehat{l} &= l + l_1(\widehat{\alpha} - \alpha) + \frac{1}{2!} l_2 (\widehat{\alpha} - \alpha)^2 + \frac{1}{3!} l_3 (\widehat{\alpha} - \alpha)^3 + \frac{1}{4!} l_4 (\widehat{\alpha} - \alpha)^4 + O_p\left(T^{-\frac{5}{2}}\right) \\
l &= \widehat{l} - l_1(\widehat{\alpha} - \alpha) - \frac{1}{2} l_2 (\widehat{\alpha} - \alpha)^2 - \frac{1}{6} l_3 (\widehat{\alpha} - \alpha)^3 - \frac{1}{24} l_4 (\widehat{\alpha} - \alpha)^4
\end{aligned}$$

$$+O_p\left(T^{-\frac{5}{2}}\right). \tag{3.8}$$

Here the combination of equation (3.7) and (3.8) can be done in a simplified way. Since

$$\begin{aligned}
(\widehat{\alpha}-\alpha)^2 &= a_{1/2}^2+2a_{1/2}a_{2/2}+2a_{1/2}a_{3/2}+a_{2/2}^2+O_p\left(T^{-\frac{5}{2}}\right),\\
(\widehat{\alpha}-\alpha)^3 &= a_{1/2}^3+3a_{1/2}^2a_{2/2}+O_p\left(T^{-\frac{5}{2}}\right),\\
(\widehat{\alpha}-\alpha)^4 &= a_{1/2}^4+O_p\left(T^{-\frac{5}{2}}\right);
\end{aligned}$$

the number of terms needed to construct $(\widehat{\alpha}-\alpha)^r$ from equation (3.7) decreases as r increases. With this observation, plug in equation (3.7) into (3.8),

$$l=\widehat{l}\ \underbrace{+\frac{l_1^2}{2l_2}}_{[A]}\ \underbrace{+\frac{l_3l_1^3}{6l_2^3}}_{[B]}\ \underbrace{+\frac{l_3^2l_1^4}{8l_2^5}}_{[C]}\ \underbrace{-\frac{l_4l_1^4}{24l_2^4}}_{[D]}\ +O_p\left(T^{-\frac{5}{2}}\right). \tag{3.9}$$

The next step would be to compute the expectation, term-by-term, of equation (3.9) in order to drop those terms whose expected value is 0. Note that we only need to discover those terms whose expected value is 0. Here term $[A]$ to $[D]$ are all ratios, so that the expectation of a ratio needs to be computed for each of them. We do this in two steps. For each ratio, we first identify the expectation of the numerator. Here an additional difficulty is that the numerator in each ratio is a product of several sums. Such a product must be expanded. This type of expansion is essentially a calculation of a product of sums, e.g., $(a+b)(c+d)=ac+ad+bc+bd$ where a to d are random variables. The expectation then follows as $\mathbb{E}(a+b)(c+d)=\mathbb{E}ac+\mathbb{E}ad+\mathbb{E}bc+\mathbb{E}bd$. To ease the representation of this expansion, let us introduce some additional notation.

Notation 1. *For given positive integers J and M, let power $p_{jm}\in\mathbb{N}$ and order $r_{jm}\in\mathbb{N}$ with $j=1,\cdots,J$ and $m=1,\cdots,M$. Let*

$$R\equiv\begin{pmatrix} r_{11} & \cdots & r_{1M}\\ \vdots & \ddots & \vdots\\ r_{J1} & \cdots & r_{JM}\end{pmatrix},\qquad P\equiv\begin{pmatrix} p_{11} & \cdots & p_{1M}\\ \vdots & \ddots & \vdots\\ p_{J1} & \cdots & p_{JM}\end{pmatrix},$$

$$\mathcal{T}\equiv\left\{(t_1,\cdots,t_J)\,|t_j=1,\cdots,T;t_j\neq t_{j'}\forall j\neq j';j,j'=1,\cdots,J\right\}$$

where $r_{jm}=0$ if and only if $p_{jm}=0$, if $p_{jm}=0$ then $p_{jm'}=0$ for $m'>m$, and $\sum_{m=1}^M p_{jm}>0$ and $\sum_{m=1}^M r_{jm}>0$. For $1(\cdot)$ being the indicator function, let

$$\begin{aligned}
\mathcal{P}(R,P) &\equiv J-\frac{1}{2}\sum\nolimits_{j=1}^{J}1\left(\sum\nolimits_{m=1}^{M}r_{jm}=1\wedge\sum\nolimits_{m=1}^{M}p_{jm}=1\right),\\
\mathcal{L}(R,P) &\equiv \frac{1}{T^{\mathcal{P}(R,P)}}\sum_{(t_1,\cdots,t_J)\in\mathcal{T}}\prod_{j=1}^{J}\prod_{m=1}^{M}\left(\nabla_a^{r_{jm}}\log f\left(Y_{it_j};\theta,a\right)\big|_{a=\alpha}\right)^{p_{jm}}.
\end{aligned}$$

Let further

$$\mathcal{L}_{\mathcal{P}}\binom{p_{11},\cdots,p_{1M};\cdots;p_{J1},\cdots p_{JM}}{r_{11},\cdots,r_{1M};\cdots;r_{J1},\cdots r_{JM}}\equiv\mathcal{L}(R,P).$$

Note that, for any R and P, $\mathcal{P}(R,P)$ is a half integer or an integer between $J/2$ and

J, and that $\mathcal{L}(R,P)$ is invariant if the rows of R and P are rearranged accordingly. Next, we introduce the following results to identify the expected value and the stochastic order of $\mathcal{L}(R,P)$.

Proposition 1. *Let*

$$\begin{aligned}\mathcal{J}_0(R,P) &\equiv \left\{j|1\le j\le J\wedge\left(\sum\nolimits_{m=1}^{M} r_{jm}=1\wedge\sum\nolimits_{m=1}^{M} p_{jm}=1\right)\right\},\\ \mathcal{J}_1(R,P) &\equiv \{1,\cdots,J\}\backslash\mathcal{J}_0(R,P).\end{aligned}$$

Suppose

$$\prod_{j\in\mathcal{J}_1(R,P)}\prod_{m=1}^{M}\left(\nabla_a^{r_{jm}}\log f\left(Y_{it_j};\theta,a\right)\big|_{a=\alpha}\right)^{p_{jm}} \tag{3.10}$$

is nonconstant (i.e., stochastic) for every $r_{jm}\le 4$ and

$$\mathbb{E}\left(\prod_{j\in\mathcal{J}_1(R,P)}\prod_{m=1}^{M}\left(\nabla_a^{r_{jm}}\log f\left(Y_{it_j};\theta,a\right)\big|_{a=\alpha}\right)^{p_{jm}}\right)\neq 0, \tag{3.11}$$

then $\mathcal{P}(R,P)$ is the smallest half integer or integer such that $\mathcal{L}(R,P)=O_p(1)$.

Proof. See Appendix B. ■

Lemma 1. *$\mathbb{E}\mathcal{L}(R,P)=0$ if $\mathcal{P}(R,P)<J$. Additionally, when condition (3.10) and (3.11) in proposition 1 are satisfied, $\mathbb{E}\mathcal{L}(R,P)=0$ if and only if $\mathcal{P}(R,P)<J$.*

Proof. See appendix C. ■

Note that, when condition (3.10) or (3.11) is not satisfied, $\mathcal{L}(R,P)$ is still $O_p(1)$. However, $\mathcal{P}(R,P)$ is no longer the smallest half integer or integer.

We then give some examples about the use of notation 1.

Example 1. *Letting*

$$R=\begin{pmatrix}1 & 2\end{pmatrix},\qquad P=\begin{pmatrix}2 & 1\end{pmatrix};$$

$$\mathcal{L}(R,P)=\frac{1}{T}\sum\nolimits_{t}\left(\left(\nabla_a\log f\left(Y_{it};\theta,a\right)\right)^2\nabla_a^2\log f\left(Y_{it};\theta,a\right)\right)\big|_{a=\alpha}.$$

Example 2. *Letting*

$$R=\begin{pmatrix}1 & 0\\ 2 & 3\end{pmatrix},\qquad P=\begin{pmatrix}1 & 0\\ 1 & 2\end{pmatrix};$$

$$\begin{aligned}&\mathcal{L}(R,P)\\ =\ &\frac{1}{T^{\mathcal{P}(R,P)}}\sum_{t_1\neq t_2}\left(\nabla_a\log f\left(Y_{it_1};\theta,a\right)\nabla_a^2\log f\left(Y_{it_2};\theta,a\right)\left(\nabla_a^3\log f\left(Y_{it_2};\theta,a\right)\right)^2\Big|_{a=\alpha}\right)\end{aligned}$$

with

$$\sum\nolimits_{j=1}^{J}1\left(\sum\nolimits_{m=1}^{M} r_{jm}=1\wedge\sum\nolimits_{m=1}^{M} p_{jm}=1\right)=1$$

such that

$$\mathcal{P}(R,P)=\frac{3}{2}.$$

Example 3. *Letting*

$$R = \left(1\right), \qquad P = \left(2\right), \qquad R' = \left(2\right), \qquad P' = \left(1\right);$$

the second Barttlet identity can be expressed as

$$\mathbb{E}\mathcal{L}\left(R,P\right) = -\mathbb{E}\mathcal{L}\left(R',P'\right)$$

in which, notice that,

$$\mathcal{L}\left(R',P'\right) \equiv l_2.$$

Example 4. *It can be computed that*

$$\begin{aligned} l_1^2 &= \frac{1}{T^2}\sum_t \left(\nabla_a \log f\left(Y_{it};\theta,a\right)\right)^2\Big|_{a=\alpha} \\ &\quad + \frac{1}{T^2}\sum_{t_1\neq t_2} \nabla_a \log f\left(Y_{it_1};\theta,a\right)\nabla_a \log f\left(Y_{it_2};\theta,a\right)\Big|_{a=\alpha} \\ &= \frac{1}{T}\mathcal{L}_1\tbinom{2}{1} + \frac{1}{T}\mathcal{L}_1\tbinom{1;1}{1;1}. \end{aligned}$$

In the $k=1$ case, the first step stops after the calculation in example 4, since the structure of the expansion of the log-likelihood is rather simple. However in the $k=2$ case, the products of sums involved have a more complicated structure. For instance,

$$l_3 l_1^3 = \left(\frac{1}{T}\sum_t \nabla_a^3 \log f\left(Y_{it};\theta,a\right)\Big|_{a=\alpha}\right)\left(\frac{1}{T}\sum_t \nabla_a \log f\left(Y_{it};\theta,a\right)\Big|_{a=\alpha}\right)^3,$$

which is a product of four sums. For this reason, we perform the above-mentioned expansion iteratively. For $[A]$ to $[D]$, we begin by expanding l_1^2 on the numerator. For instance in $[B]$, l_1^2 will be expanded while $l_1 l_3$ left intact. After this calculation, we substitute the expanded term $1/T\mathcal{L}_1\binom{2}{1} + 1/T\mathcal{L}_1\binom{1;1}{1;1}$ back into the numerator and rewrite the ratio into a sum of several ratios. Each of the rewritten ratios only contains products in the numerator.

Example 5. *After the first iteration,* $[B]$ *will become*

$$[B] = \frac{l_3 l_1 \mathcal{L}_1\binom{2}{1}}{6Tl_2^3} + \frac{l_3 l_1 \mathcal{L}_1\binom{1;1}{1;1}}{6Tl_2^3}.$$

When this is finished for all $[A]$ to $[D]$, we will obtain a new version of equation (3.9) containing, say, K ratios. In the next iteration, for each of the K ratios, we perform a similar calculation. We expand the product of two $\mathcal{L}\left(R,P\right)$, each satisfying $\mathbb{E}\mathcal{L}\left(R,P\right)=0$. If the numerator has only one such $\mathcal{L}\left(R,P\right)$, we expand the product of this particular $\mathcal{L}\left(R,P\right)$ and any other $\mathcal{L}\left(R,P\right)$. Here note that any l_r can be expressed as an $\mathcal{L}\left(R,P\right)$ and that the condition $\mathbb{E}\mathcal{L}\left(R,P\right)=0$ can be verified by lemma 1. We stop the iterative procedure if each of the harvested ratios satisfies exactly one of the following conditions.

1. The numerator is a product of a constant, one or several l_r, or one or several $\mathcal{L}\left(R,P\right)$ if none of them have a zero expectation.

2. The numerator is a product of a constant and an $\mathcal{L}\left(R,P\right)$ satisfying $\mathbb{E}\mathcal{L}\left(R,P\right)=0$.

Example 6. *Continuing with the above example, we compute*

$$\begin{aligned} l_1\mathcal{L}_1\binom{2}{1} &= \frac{1}{T}\mathcal{L}_1\binom{3}{1}+\frac{1}{T^{1/2}}\mathcal{L}_{1.5}\binom{2;1}{1;1}, \\ l_1\mathcal{L}_1\binom{1;1}{1;1} &= \frac{1}{T^{1/2}}\mathcal{L}_{1.5}\binom{1;1;1}{1;1;1}+\frac{2}{T^{1/2}}\mathcal{L}_{1.5}\binom{1;2}{1;1} \end{aligned}$$

such that

$$[B] = \frac{l_3\mathcal{L}_1\binom{3}{1}}{6T^2 l_2^3} + \frac{l_3\mathcal{L}_{1.5}\binom{2;1}{1;1}}{2T^{1.5} l_2^3} + \frac{l_3\mathcal{L}_{1.5}\binom{1;1;1}{1;1;1}}{6T^{1.5} l_2^3}$$

in which the first ratio satisfies the first condition and hence needs not be further processed, whereas the others need to be expanded further.

After this procedure, the identification of the stochastic order and the expectation of a ratio would become easier. When a ratio satisfies condition 1, it can be stochastically expanded into a term with a nonzero expectation plus a higher-order term. Such a ratio must be kept in the derivation. When, however, a ratio satisfies condition 2, a stochastic expansion of the ratio will contain a term with a zero expectation plus a higher-order term. In this case, the leading term can be dropped whereas the higher-order term may need to be investigated further (see below).

Example 7. *Continuing with the above example, it is clear that*

$$\begin{aligned} \frac{l_3\mathcal{L}_1\binom{3}{1}}{6T^2 l_2^3} &= O_p\left(T^{-2}\right), \\ \mathbb{E}\frac{l_3\mathcal{L}_1\binom{3}{1}}{6T^2 l_2^3} &= \frac{\mathbb{E}\left(l_3\mathcal{L}_1\binom{3}{1}\right)}{6T^2\mathbb{E}\left(l_2^3\right)} + \frac{o(1)}{T^2} = \frac{\mathbb{E}l_3\mathbb{E}\mathcal{L}_1\binom{3}{1}}{6T^2\mathbb{E}\left(l_2^3\right)} + o\left(T^{-2}\right) \end{aligned}$$

in which

$$\frac{\mathbb{E}l_3\mathbb{E}\mathcal{L}_1\binom{3}{1}}{6T^2\mathbb{E}\left(l_2^3\right)} \neq 0.$$

Also note that the harvested ratios depend on the calculation above; i.e., the expression obtained from the procedure would be different if, e.g., l_1^4 from $[D]$ were expanded at once. This implies that the expansion of the likelihood does not have a unique representation. However, all variants should be identical in the sense that they evaluate to the same value[3]. It is also equivalent if one expands each numerator into a sum of several $\mathcal{L}(R,P)$ regardless of condition 1 and 2. This technique can be favorable when T is small; however, when T is large, this will deliver too many terms, complicating the derivation. We present the exact approach of calculating the product of two $\mathcal{L}(R,P)$ in appendix D. In what follows, we use this procedure when computing the product in the numerator of any ratio.

Formally,

$$[A] = \frac{\mathcal{L}_1\binom{2}{1}}{2Tl_2} + \underbrace{\frac{\mathcal{L}_1\binom{1;1}{1;1}}{2Tl_2}}_{[E]},$$

[3]We implemented the above procedure as a computer symbolic algorithm and numerically verified that equation (3.9) delivers the same value, up to some numerical roundoff error of a typical magnitude of 10^{-16}, as the expression derived from the above procedure.

$$
\begin{aligned}
[B] &= \underbrace{\frac{\mathcal{L}_{2.5}\binom{1;2;1}{1;1;3}}{2T^{1.5}l_2^3}}_{[F]}+\underbrace{\frac{\mathcal{L}_{2.5}\binom{1;1;1;1}{1;1;1;3}}{6T^{1.5}l_2^3}}_{[G]}+\frac{\mathcal{L}_2\binom{2;1,1}{1;1,3}}{2T^2l_2^3}+\underbrace{\frac{\mathcal{L}_2\binom{1;1;1,1}{1;1;1,3}}{2T^2l_2^3}}_{[H]}+\frac{l_3\mathcal{L}_1\binom{3}{1}}{6T^2l_2^3}+O_p\left(T^{-\frac{5}{2}}\right),\\
[C] &= \underbrace{\frac{3\mathcal{L}_4\binom{1;1;2;1;1}{1;1;1;3;3}}{4T^2l_2^5}}_{[I]}+\underbrace{\frac{\mathcal{L}_4\binom{1;1;1;1;1;1}{1;1;1;1;3;3}}{8T^2l_2^5}}_{[J]}+\frac{l_3^2\mathcal{L}_1\binom{2}{1}^2}{8T^2l_2^5}+\frac{l_3^2\mathcal{L}_2\binom{2;2}{1;1}}{4T^2l_2^5}+O_p\left(T^{-\frac{5}{2}}\right),\\
[D] &= \underbrace{-\frac{\mathcal{L}_3\binom{1;1;2;1}{1;1;1;4}}{4T^2l_2^4}}_{[K]}\underbrace{-\frac{\mathcal{L}_3\binom{1;1;1;1;1}{1;1;1;1;4}}{24T^2l_2^4}}_{[L]}-\frac{l_4\mathcal{L}_2\binom{2;2}{1;1}}{12T^2l_2^4}-\frac{l_4\mathcal{L}_1\binom{2}{1}^2}{24T^2l_2^4}+O_p\left(T^{-\frac{5}{2}}\right).
\end{aligned}
$$

It can be verified that terms $[H]$ to $[L]$ have a zero expectation to the order of $O\left(T^{-2}\right)$. For example,

$$
\begin{aligned}
\mathbb{E}\frac{\mathcal{L}_2\binom{1;1;1,1}{1;1;1,3}}{2T^2l_2^3} &= \frac{1}{T^2}\frac{\mathbb{E}\mathcal{L}_2\binom{1;1;1,1}{1;1;1,3}}{2\mathbb{E}\left(l_2^3\right)}+o\left(T^{-2}\right)\\
&= o\left(T^{-2}\right).
\end{aligned}
$$

Terms $[E]$, $[F]$, and $[G]$ need to be investigated further, due to the following reasoning. Ratio $[E]$, $[F]$, and $[G]$ satisfy condition 2 such that the leading terms in their stochastic expansions have a zero expectation; that is,

$$
[E]=E'+O_p\left(T^{-3/2}\right),\qquad [F]=F'+O_p(T^{-2}),\qquad [G]=G'+O_p(T^{-2})
$$

where $\mathbb{E}E'=\mathbb{E}F'=\mathbb{E}G'=0$. However, $O_p\left(T^{-3/2}\right)$ and $O_p(T^{-2})$ are lower than $O_p\left(T^{-5/2}\right)$ such that they also need to be calculated. To deal with these terms, first observe that, as $l_2=\mathbb{E}l_2+O_p\left(T^{-1/2}\right)$, $1/l_2$ can be expanded, i.e.,

$$
\frac{1}{l_2}=\frac{1}{\mathbb{E}l_2}-\frac{1}{\left(\mathbb{E}l_2\right)^2}\left(l_2-\mathbb{E}l_2\right)+\frac{1}{\left(\mathbb{E}l_2\right)^3}\left(l_2-\mathbb{E}l_2\right)^2+O_p\left(T^{-\frac{3}{2}}\right). \tag{3.12}
$$

Properties of equation (3.12) are studied by, e.g., Rice (2008). It is also known that the Taylor series of a reciprocal function is only convergent in a specific region, which, in our setting, is $2\mathbb{E}l_2<l_2<0$. This, however, does not contradict the use of the above-mentioned Taylor series in our setting, since $l_2\to_p\mathbb{E}l_2$ when $T\to\infty$. The second step[4] is to replace $1/l_2$ in $[E]$, $[F]$, and $[G]$ by equation (3.12).

Formally,

$$
\begin{aligned}
[E] &= \frac{\mathcal{L}_1\binom{1;1}{1;1}}{2T}\left(\frac{1}{\mathbb{E}l_2}-\frac{1}{\left(\mathbb{E}l_2\right)^2}\left(l_2-\mathbb{E}l_2\right)+\frac{1}{\left(\mathbb{E}l_2\right)^3}\left(l_2-\mathbb{E}l_2\right)^2+O_p\left(T^{-\frac{3}{2}}\right)\right)\\
&= \frac{3\mathcal{L}_1\binom{1;1}{1;1}}{2T\mathbb{E}l_2}-\frac{3\mathcal{L}_2\binom{1;1;1}{1;1;2}}{2T\left(\mathbb{E}l_2\right)^2}+\frac{\mathcal{L}_3\binom{1;1;1;1}{1;1;2;2}}{2T\left(\mathbb{E}l_2\right)^3}-\frac{3\mathcal{L}_{1.5}\binom{1;1,1}{1;1,2}}{T^{1.5}\left(\mathbb{E}l_2\right)^2}+\frac{2\mathcal{L}_{2.5}\binom{1;1;1,1}{1;2;1,2}}{T^{1.5}\left(\mathbb{E}l_2\right)^3}\\
&\quad \underbrace{+\frac{\mathcal{L}_2\binom{1,1;1,1}{1,2;1,2}}{T^2\left(\mathbb{E}l_2\right)^3}}_{[E.1]}+\frac{\mathcal{L}_2\binom{1;1;2}{1;1;2}}{2T^2\left(\mathbb{E}l_2\right)^3}+O_p\left(T^{-\frac{5}{2}}\right)
\end{aligned}
$$

[4] Recall that the first step was to expand the numerator.

in which only $[E.1]$ has a nonzero expectation. Similarly,

$$\begin{aligned}
[F] &= \frac{\mathcal{L}_{2.5}\binom{1;2;1}{1;1;3}}{2T^{1.5}}\left(\frac{1}{\mathbb{E}l_2}-\frac{1}{(\mathbb{E}l_2)^2}(l_2-\mathbb{E}l_2)+\frac{1}{(\mathbb{E}l_2)^3}(l_2-\mathbb{E}l_2)^2+O_p\left(T^{-\frac{3}{2}}\right)\right)^3 \\
&= \frac{2\mathcal{L}_{2.5}\binom{1;2;1}{1;1;3}}{T^{1.5}(\mathbb{E}l_2)^3}-\frac{3\mathcal{L}_{3.5}\binom{1;2;1;1}{1;1;3;2}}{2T^{1.5}(\mathbb{E}l_2)^4}-\underbrace{\frac{3\mathcal{L}_3\binom{2;1;1,1}{1;3;1,2}}{2T^2(\mathbb{E}l_2)^4}}_{[F.1]}+O_p\left(T^{-\frac{5}{2}}\right)
\end{aligned}$$

in which only $[F.1]$ has a nonzero expectation and

$$\begin{aligned}
[G] &= \frac{\mathcal{L}_{2.5}\binom{1;1;1;1}{1;1;1;3}}{6T^{1.5}}\left(\frac{1}{\mathbb{E}l_2}-\frac{1}{(\mathbb{E}l_2)^2}(l_2-\mathbb{E}l_2)+\frac{1}{(\mathbb{E}l_2)^3}(l_2-\mathbb{E}l_2)^2+O_p\left(T^{-\frac{3}{2}}\right)\right)^3 \\
&= \frac{2\mathcal{L}_{2.5}\binom{1;1;1;1}{1;1;1;3}}{3T^{1.5}(\mathbb{E}l_2)^3}-\frac{\mathcal{L}_{3.5}\binom{1;1;1;1;1}{1;1;1;3;2}}{2T^{1.5}(\mathbb{E}l_2)^4}-\frac{3\mathcal{L}_3\binom{1;1;1;1,1}{1;1;3;1,2}}{2T^2(\mathbb{E}l_2)^4}+O_p\left(T^{-\frac{5}{2}}\right)
\end{aligned}$$

in which all ratios have a zero expectation.

Now, drop terms $[G]$ to $[L]$, and replace term $[E]$ and $[F]$ with $[E.1]$ and $[F.1]$ respectively to construct

$$\mathbb{E}l=\mathbb{E}\widehat{l}+\frac{\mathbb{E}b_1}{T}+\frac{\mathbb{E}b_2'}{T^2}+O\left(T^{-3}\right) \tag{3.13}$$

where

$$\begin{aligned}
b_1 &\equiv \frac{\mathcal{L}_1\binom{2}{1}}{2l_2} \\
b_2' &\equiv \frac{\mathcal{L}_2\binom{1,1;1,1}{1,2;1,2}}{(\mathbb{E}l_2)^3}-\frac{3\mathcal{L}_3\binom{2;1;1,1}{1;3;1,2}}{2(\mathbb{E}l_2)^4}+\frac{\mathcal{L}_2\binom{2;1,1}{1;1,3}}{2l_2^3}+\frac{l_3\mathcal{L}_1\binom{3}{1}}{6l_2^3} \\
&\quad -\frac{l_4\mathcal{L}_2\binom{2;2}{1;1}}{12l_2^4}-\frac{l_4\mathcal{L}_1\binom{2}{1}^2}{24l_2^4}+\frac{l_3^2\mathcal{L}_1\binom{2}{1}^2}{8l_2^5}+\frac{l_3^2\mathcal{L}_2\binom{2;2}{1;1}}{4l_2^5}.
\end{aligned}$$

Here it can be verified that each ratio in b_1 and b_2' is $O_p(1)$ such that, in general, $b_1=O_p(1)$ and $b_2'=O_p(1)$.

3.4 Corrected Likelihood

If α were available, equation (3.13) could be directly constructed to approximate $\mathbb{E}l$ to the order of $O\left(T^{-2}\right)$. When, however, only $\widehat{\alpha}$ is available, $\mathbb{E}\widehat{b}_1$, the plug-in estimate, is not sufficiently accurate for $\mathbb{E}b_1$. To deal with this problem, we apply the same procedure on b_1/T. Taylor-expanding $\widehat{b}_1/T$ around α,

$$\begin{aligned}
\frac{\widehat{b}_1}{T} &= \frac{b_1}{T}+\frac{1}{T}\nabla b_1(\widehat{\alpha}-\alpha)+\frac{1}{2}\frac{1}{T}\nabla^2 b_1(\widehat{\alpha}-\alpha)^2+O_p\left(T^{-\frac{5}{2}}\right) \\
\frac{b_1}{T} &= \frac{\widehat{b}_1}{T}-\frac{1}{T}\nabla b_1(\widehat{\alpha}-\alpha)-\frac{1}{2}\frac{1}{T}\nabla^2 b_1(\widehat{\alpha}-\alpha)^2+O_p\left(T^{-\frac{5}{2}}\right)
\end{aligned}$$

where ∇^r denotes the r-th derivative w.r.t. a. Plug in equation (3.7) for $(\widehat{\alpha}-\alpha)$ and rearrange to obtain

$$\begin{aligned}\frac{b_1}{T} &= \frac{\widehat{b}_1}{T}+\frac{l_1\mathcal{L}_1\binom{1,1}{1,2}}{Tl_2^2}-\frac{l_1^2\mathcal{L}_1\binom{2}{2}}{2Tl_2^3}-\frac{l_1^2\mathcal{L}_1\binom{1,1}{1,3}}{2Tl_2^3}+\frac{l_1^2l_4\mathcal{L}_1\binom{2}{1}}{4Tl_2^4}\\&\quad+\frac{3l_1^2l_3\mathcal{L}_1\binom{1,1}{1,2}}{2Tl_2^4}-\frac{3l_1^2l_3^2\mathcal{L}_1\binom{2}{1}}{4Tl_2^5}-\frac{l_1l_3\mathcal{L}_1\binom{2}{1}}{2Tl_2^3}+O_p\left(T^{-\frac{5}{2}}\right).\end{aligned}$$

Apply the same procedure in section 3.3,

$$\frac{\mathbb{E}b_1}{T}=\frac{\mathbb{E}\widehat{b}_1}{T}+\frac{\mathbb{E}b_{1,1}}{T^2}+O\left(T^{-3}\right) \tag{3.14}$$

where

$$\begin{aligned}b_{1,1} &\equiv \frac{3\mathcal{L}_3\binom{2;1;1,1}{1;3;1,2}}{2\left(\mathbb{E}l_2\right)^4}-\frac{2\mathcal{L}_2\binom{1,1;1,1}{1,2;1,2}}{\left(\mathbb{E}l_2\right)^3}+\frac{\mathcal{L}_1\binom{2,1}{1,2}}{l_2^2}-\frac{\mathcal{L}_2\binom{2;1,1}{1;1,3}}{2l_2^3}-\frac{3l_3^2\mathcal{L}_1\binom{2}{1}^2}{4l_2^5}\\&\quad-\frac{\mathcal{L}_1\binom{2}{1}\mathcal{L}_1\binom{2}{2}}{2l_2^3}-\frac{\mathcal{L}_1\binom{2}{1}\mathcal{L}_1\binom{1,1}{1,3}}{2l_2^3}-\frac{l_3\mathcal{L}_1\binom{3}{1}}{2l_2^3}+\frac{l_4\mathcal{L}_1\binom{2}{1}^2}{4l_2^4}\\&\quad+\frac{3l_3\mathcal{L}_1\binom{2}{1}\mathcal{L}_1\binom{1,1}{1,2}}{2l_2^4}.\end{aligned}$$

Combining equation (3.13) and (3.14),

$$\mathbb{E}l=\mathbb{E}\widehat{l}+\frac{\mathbb{E}\widehat{b}_1}{T}+\frac{\mathbb{E}b_2}{T^2}+O\left(T^{-3}\right) \tag{3.15}$$

with

$$\begin{aligned}b_1 &\equiv \frac{\mathcal{L}_1\binom{2}{1}}{2l_2},\\ b_2 &\equiv b_2'+b_{1,1}\\ &= -\frac{\mathcal{L}_2\binom{1,1;1,1}{1,2;1,2}}{\left(\mathbb{E}l_2\right)^3}-\frac{l_3\mathcal{L}_1\binom{3}{1}}{3l_2^3}-\frac{l_4\mathcal{L}_2\binom{2;2}{1;1}}{12l_2^4}+\frac{5l_4\mathcal{L}_1\binom{2}{1}^2}{24l_2^4}-\frac{5l_3^2\mathcal{L}_1\binom{2}{1}^2}{8l_2^5}+\frac{l_3^2\mathcal{L}_2\binom{2;2}{1;1}}{4l_2^5}\\&\quad+\frac{\mathcal{L}_1\binom{2,1}{1,2}}{l_2^2}-\frac{\mathcal{L}_1\binom{2}{1}\mathcal{L}_1\binom{2}{2}}{2l_2^3}-\frac{\mathcal{L}_1\binom{2}{1}\mathcal{L}_1\binom{1,1}{1,3}}{2l_2^3}+\frac{3l_3\mathcal{L}_1\binom{2}{1}\mathcal{L}_1\binom{1,1}{1,2}}{2l_2^4}.\end{aligned}$$

Here b_2 can be replaced with $\widehat{b}_2$ - b_2 evaluated at $\widehat{\alpha}$. This induces a bias, in expectation, to the order of $O\left(T^{-3}\right)$ which can be neglected. $\mathbb{E}l_2$ in b_2 can be replaced l_2, which also induces a negligible bias. The second-order corrected likelihood (for a single i) can then be constructed as

$$\widehat{l}^{(2)}\equiv\widehat{l}+\frac{\widehat{b}_1}{T}+\frac{\widehat{b}_2}{T^2}$$

where the right-hand side depends only on known quantities Y_{it}, $\widehat{\alpha}$, and θ and hence, can be constructed in a straightforward way.

Remark 3. *Assumption 2 is also significant here, i.e., equation (3.15) would be undefined if* $\widehat{l}_2=0$. *Here it is also possible to derive an alternative version of equation (3.15) where* *the denominators contain only* $\mathbb{E}l_2$. *This would avoid the problem. However, as*

$$\mathbb{E}\left(\frac{\widehat{\mathcal{L}}_1\binom{2}{1}}{\mathbb{E}\widehat{l}_2}\right) \neq \mathbb{E}\left(\frac{\widehat{\mathcal{L}}_1\binom{2}{1}}{\widehat{l}_2}\right)$$

in general, such an alternative formula would contain a bias to the order of $O\left(T^{-2}\right)$ *in expectation if no treatment were employed to deal with the inequality of the expectations above. Possible solutions, e.g., Hartley and Ross (1954), de Paschal (1961), and Ogliore et al. (2011), can be adopted for this. However, these solutions all depend on* $\widehat{l}_2 < 0$, *rendering the inevitability of assumption 2.*

Under the asymptotic sequence $N/T^5 \to 0$ as $N, T \to \infty$, $O_p\left(T^{-5/2}\right) = o_p\left(N^{-1/2}\right)$ such that $1/N \sum_i \widehat{l}^{(2)}(\theta)$ is consistent, i.e.,

$$\begin{aligned}
\frac{1}{N}\sum_i \widehat{l}^{(2)}(\theta) &= \frac{1}{N}\sum_i l(\theta) + O_p\left(N^{-\frac{1}{2}}\right) \\
&= \underset{N\to\infty}{\text{plim}} \frac{1}{N}\sum_i l(\theta) + O_p\left(N^{-\frac{1}{2}}\right)
\end{aligned}$$

implying that, under certain regularity conditions,

$$\sqrt{NT}\left(\widehat{\theta}^{(2)} - \theta\right) \to_d \mathcal{N}(0, \Sigma)$$

where $\mathcal{N}(\cdot,\cdot)$ is the normal density with mean zero and covariance matrix

$$\begin{aligned}
\Sigma &\equiv H^{-1}\mathbb{E}\left(s_{it}s_{it}'\right)H^{-1}, \\
s_{it} &\equiv \nabla_\theta \log f\left(Y_{it};\theta,\alpha(\theta)\right), \qquad H \equiv -\frac{1}{N}\sum_i \mathbb{E}\nabla_{\theta\theta'} l(\theta).
\end{aligned}$$

4 Example and Simulation

4.1 Analytical Correction of Many-normal-mean Model

Our first example is the many-normal-mean model introduced by Neyman and Scott (1948). Consider $Y_{it} \sim \mathcal{N}(\alpha_i, \theta_0)$ where α_i, the mean, is different across i, and θ_0, the variance, is the same for all it. In this setting, the variance is the parameter of interest whereas the means are nuisance. It can be shown[5] that, under fixed T, $\widehat{\theta} \to_p \theta_T \neq \theta_0$, while under increasing T, $\widehat{\theta} - \theta_0 = O_p\left(T^{-1}\right)$. To see this, observe that

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

such that the log-likelihood writes

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

[5]See, e.g., Neyman and Scott (1948) or Lancaster (2000).

Under this setting, $\alpha_i = \mathbb{E}_i Y_{it}$ and $\widehat{\alpha}_i = 1/N \sum_t Y_{it}$ for every θ where $\mathbb{E}_i(\cdot)$ denotes the expectation computed on a specific i.

There would be no IPP if α_i were plugged in for a_i, i.e.,

$$\overline{\theta} = \frac{1}{NT} \sum_{it} (Y_{it} - \alpha_i)^2$$

is fully unbiased and consistent under $N \to \infty$ even when T is fixed. However, as α_i is not always feasible, $\widehat{\alpha}_i$ must be plugged in, under which case

$$\widehat{\theta} = \frac{1}{NT} \sum_{it} (Y_{it} - \widehat{\alpha}_i)^2 \to_p \theta_0 - \frac{\theta_0}{T}$$

as $N \to \infty$. When T increases, $\widehat{\theta} - \theta_0 = O_p\left(T^{-1}\right)$. Next, we apply equation (3.15) to the model. Because of the fact that, for every given θ, $\nabla_a^2 \log f(Y_{it}; \theta, a)$ is a constant, we can anticipate

$$\nabla_a^r \log f(Y_{it}; \theta, a) = l_r = 0 \text{ for } r > 2, \qquad l_2 = -1/\theta, \qquad \mathbb{E}\mathcal{L}_2\binom{1,1;1,1}{1,2;1,2} = 0,$$
$$\mathcal{L}_1\binom{2,1}{1,2}/l_2^2 = \mathcal{L}_1\binom{2}{1}/l_2, \qquad \mathcal{L}_1\binom{2}{1}\mathcal{L}_1\binom{2}{2}/2l_2^3 = \mathcal{L}_1\binom{2}{1}/2l_2$$

such that

$$\widehat{l}^{(2)} = \widehat{l} + \frac{1}{T}\frac{\widehat{\mathcal{L}}_1\binom{2}{1}}{2\widehat{l}_2} + \frac{1}{T^2}\frac{\widehat{\mathcal{L}}_1\binom{2}{1}}{2\widehat{l}_2}$$

where $\widehat{\mathcal{L}}_1\binom{2}{1}$ is understood as evaluating $\mathcal{L}_1\binom{2}{1}$ at $\widehat{\alpha}_i$. It follows then that, as

$$\frac{\widehat{\mathcal{L}}_1\binom{2}{1}}{\widehat{l}_2} = -\frac{1}{T}\sum_t \frac{(Y_{it} - \widehat{\alpha}_i)^2}{2\theta},$$

the log-likelihood (with $\widehat{\alpha}_i$ plugged in) becomes

$$-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\theta - \left(\frac{1}{NT} + \frac{1}{NT^2} + \frac{1}{NT^3}\right)\sum_{it}\frac{(Y_{it} - \widehat{\alpha}_i)^2}{2\theta}.$$

Equating the first derivative (w.r.t. θ) to 0,

$$-\frac{1}{2}\frac{1}{\theta} + \left(\frac{1}{NT} + \frac{1}{NT^2} + \frac{1}{NT^3}\right)\sum_{it}\frac{(Y_{it} - \widehat{\alpha}_i)^2}{2\theta} = 0$$

in which, as $\theta \neq 0$,

$$\begin{aligned}\widehat{\theta}^{(2)} &= \left(\frac{1}{T} + \frac{1}{T} + \frac{1}{T^3}\right)\frac{1}{N}\sum_{it}(Y_{it} - \widehat{\alpha}_i)^2 \\ &= \left(1 - \frac{1}{T^3}\right)\frac{1}{N(T-1)}\sum_{it}(Y_{it} - \widehat{\alpha}_i)^2.\end{aligned}$$

Under $N \to \infty$ and fixed T,

$$\frac{1}{N(T-1)}\sum_{it}(Y_{it} - \widehat{\alpha}_i)^2 \to_p \theta_0$$

such that

$$\widehat{\theta}^{(2)} - \theta_0 \rightarrow_p \frac{\theta_0}{T^3}.$$

4.2 Correction of Logit Model

In this section we present a simulation study of the logit model. Note that, the static logit model has an alternative analytical correction approach, the conditional logit - see Andersen (1970), Heckman (1981), or Chamberlain (1985). Let $1(\cdot)$ be the indicator function and consider the model

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

where ε_{it} is standard-logistically distributed and X_{it} is a scalar covariate. For the logit model, the individual log-likelihood is

$$\log f\left(Y_{it}; \theta, a_i\right) \equiv \left(1 - Y_{it}\right)\left(-X_{it}\theta - a_i\right) - \log\left(1 + \exp\left(-X_{it}\theta - a_i\right)\right).$$

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

1. $X_{it} \sim \mathcal{N}(0,1)$ and $\alpha_i = 0$. This represents the case where the model could be consistently estimated by the pooled logit.
2. $X_{it} \sim \mathcal{N}(0,1)$ and $\alpha_i \sim \mathcal{N}(0,1/16)$. This represents the case where the model could be consistently estimated by the random-effect logit.
3. $X_{it} \sim \mathcal{N}(\alpha_i,1)$ with $\alpha_i \sim \mathcal{N}(0,1/16)$. In this design, X_{it} and α_i are correlated such that the model must be estimated by the fixed-effect logit.

The number of replications of the Monte Carlo study is $1,000$ with N set to be $10,000$ in each. θ_0 and T are varied according to the description in the tables. A comparison across the three designs may conclude that the estimates $\widehat{\theta}$, $\widehat{\theta}^{(1)}$, and $\widehat{\theta}^{(2)}$ are, respectively, very similar across designs. This reflects the fact that the IPP enters when one allows α_i to be estimated, instead of when the fixed-effect model is the true underlying model. That is, estimating the fixed-effect logit on a dataset where pooled logit or random-effect logit could be estimated consistently induces an IPP bias. Second, the proposed second-order bias correction is effective in the sense that the bias is reduced sufficiently even when T is only 5. When T is 20, the estimate $\widehat{\theta}^{(2)}$ is almost unbiased. As a comparison, under the same design and the same T, the bias in the original estimate $\widehat{\theta}$ is roughly 5% to 6% relative to θ_0. Furthermore, the bias is roughly symmetric around 0, since it can be seen that the magnitude of the relative bias are similar when θ_0 is flipped around 0. This phenomenon may indicate that the bias is 0 when $\theta_0 = 0$.

Another point that may be of interest is that the bias is roughly $O_p\left(T^{-3}\right)$. To see this, first note that, under the condition

$$\mathbb{E}\left(\widehat{\theta}^{(2)} - \theta_0\right) = O\left(T^{-3}\right),$$

the bias in $\widehat{\theta}^{(2)}$ should be reduced by a factor of roughly 1/8 when T is doubled. This is the case in the results. Focus on table 1 and let $B_T\left(\theta_0\right)$ denote the absolute value of the relative bias in $\widehat{\theta}^{(2)}$ under some θ_0 and T. It can be calculated, e.g., as

$$\frac{B_{10}(0.5)}{B_{20}(0.5)} = \frac{0.0015}{0.0002} = 7.5, \qquad \frac{B_5(-0.5)}{B_{10}(-0.5)} = \frac{0.0179}{0.0021} \approx 8.5238,$$
$$\frac{B_{10}(1)}{B_{20}(1)} = \frac{0.0035}{0.0005} = 7, \qquad \frac{B_5(-1)}{B_{10}(-1)} = \frac{0.0283}{0.0035} \approx 8.0857.$$

Figure 1 presents plots of the profiled log-likelihoods, for $T = 5$ and $T = 10$, of the logit model $Y_{it} = 1(X_{it}\theta_0 + \alpha_i + \varepsilon_{it} \geq 0)$ where ε_{it} is standard-logistically distributed, $X_{it} \sim \mathcal{N}(0, 4)$, $\alpha_i = 0$, and $\theta_0 = 0.5$. $N = 10,000$. Here $X_{it} \sim \mathcal{N}(0, 4)$ to introduce a sufficient variation such that the curves are steeper and visually distinguishable. The plotted quantities are $\sum_i \widehat{l}(\theta)$ (circle), $\sum_i \widehat{l}^{(1)}(\theta)$ (triangle), $\sum_i \widehat{l}^{(2)}(\theta)$ (square), and $\sum_i \mathbb{E}l(\theta)$ (asterisk). The profile log-likelihoods are computed for $\theta = 0.3, \cdots, 0.7$ with a step of 0.01, and the vertical lines indicate the maximizers. A comparison of the two graphs shows that when T increases from 5 to 10, $\widehat{l}^{(2)}(\theta)$ converges faster to $\mathbb{E}l(\theta)$ than $\widehat{l}^{(1)}(\theta)$, which itself converges to the expected likelihood faster than $\widehat{l}(\theta)$. Here a distinct feature is that, when T is as small as 5, $\widehat{l}^{(2)}(\theta)$ is already very accurate, compared to $\widehat{l}^{(1)}(\theta)$ and $\widehat{l}(\theta)$, as an approximation of $\mathbb{E}l(\theta)$. Under the above setting, the maximizer of $\widehat{l}^{(2)}(\theta)$ is already very close[6] to the maximizer of $\mathbb{E}l(\theta)$, which is θ_0.

[6] Because θ is chosen discretely, we would rather not use the phrase “exactly the same”.

Table 1: Second Order - Simulation Results for Logit Model - Design 1

Setting	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE
$T=5$		$\theta_0 = 0.5$			$\theta_0 = -0.5$			$\theta_0 = 1$			$\theta_0 = -1$	
$\widehat{\theta}$	0.6313	0.2626	0.1320	−0.6310	0.2621	0.1318	1.2940	0.2940	0.2946	−1.2928	0.2928	0.2934
$\widehat{\theta}^{(1)}$	0.5383	0.0766	0.0400	−0.5379	0.0758	0.0396	1.0893	0.0893	0.0906	−1.0883	0.0883	0.0895
$\widehat{\theta}^{(2)}$	0.4912	−0.0176	0.0137	−0.4910	**−0.0179**	0.0138	0.9725	−0.0275	0.0307	−0.9717	**−0.0283**	0.0309
$T=10$		$\theta_0 = 0.5$			$\theta_0 = -0.5$			$\theta_0 = 1$			$\theta_0 = -1$	
$\widehat{\theta}$	0.5573	0.1146	0.0579	−0.5570	0.1139	0.0576	1.1229	0.1229	0.1233	−1.1230	0.1230	0.1234
$\widehat{\theta}^{(1)}$	0.5095	0.0190	0.0122	−0.5092	0.0184	0.0120	1.0196	0.0196	0.0216	−1.0197	0.0197	0.0218
$\widehat{\theta}^{(2)}$	0.4993	**−0.0015**	0.0075	−0.4990	**−0.0021**	0.0076	0.9965	**−0.0035**	0.0094	−0.9965	**−0.0035**	0.0098
$T=20$		$\theta_0 = 0.5$			$\theta_0 = -0.5$			$\theta_0 = 1$			$\theta_0 = -1$	
$\widehat{\theta}$	0.5267	0.0534	0.0272	−0.5267	0.0534	0.0272	1.0564	0.0564	0.0567	−1.0567	0.0567	0.0571
$\widehat{\theta}^{(1)}$	0.5023	0.0045	0.0056	−0.5022	0.0045	0.0056	1.0044	0.0044	0.0073	−1.0047	0.0047	0.0078
$\widehat{\theta}^{(2)}$	0.4999	**−0.0002**	0.0050	−0.4999	−0.0002	0.0051	0.9995	**−0.0005**	0.0058	−0.9998	−0.0002	0.0061

Notes: Bias is presented relative to θ_0. The number of replications is $1,000$, $N = 10,000$. Model: $Y_{it} = 1\left(X_{it}\theta_0 + \alpha_i + \varepsilon_{it} \geq 0\right)$ where ε_{it} is standard-logistically distributed, $X_{it} \sim \mathcal{N}(0,1)$, and $\alpha_i = 0$. $\widehat{\theta}$ is the original estimate, $\widehat{\theta}^{(1)}$ is the first-order corrected, and $\widehat{\theta}^{(2)}$ is the second-order corrected.

Table 2: Second Order - Simulation Results for Logit Model - Design 2

Setting	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE
$T=5$		$\theta_0 = 0.5$			$\theta_0 = -0.5$			$\theta_0 = 1$			$\theta_0 = -1$	
$\widehat{\theta}$	0.6311	0.2621	0.1318	−0.6307	0.2615	0.1314	1.2941	0.2941	0.2947	−1.2938	0.2938	0.2944
$\widehat{\theta}^{(1)}$	0.5381	0.0763	0.0400	−0.5377	0.0753	0.0393	1.0895	0.0895	0.0907	−1.0892	0.0892	0.0904
$\widehat{\theta}^{(2)}$	0.4907	−0.0187	0.0143	−0.4904	−0.0192	0.0140	0.9721	−0.0279	0.0307	−0.9717	−0.0283	0.0310
$T=10$		$\theta_0 = 0.5$			$\theta_0 = -0.5$			$\theta_0 = 1$			$\theta_0 = -1$	
$\widehat{\theta}$	0.5574	0.1148	0.0580	−0.5570	0.1141	0.0576	1.1232	0.1232	0.1236	−1.1230	0.1230	0.1235
$\widehat{\theta}^{(1)}$	0.5095	0.0191	0.0120	−0.5092	0.0185	0.0119	1.0198	0.0198	0.0217	−1.0196	0.0196	0.0219
$\widehat{\theta}^{(2)}$	0.4991	−0.0018	0.0072	−0.4988	−0.0024	0.0075	0.9962	−0.0038	0.0095	−0.9960	−0.0040	0.0102
$T=20$		$\theta_0 = 0.5$			$\theta_0 = -0.5$			$\theta_0 = 1$			$\theta_0 = -1$	
$\widehat{\theta}$	0.5267	0.0535	0.0273	−0.5268	0.0536	0.0274	1.0570	0.0570	0.0574	−1.0568	0.0568	0.0572
$\widehat{\theta}^{(1)}$	0.5023	0.0045	0.0056	−0.5023	0.0047	0.0057	1.0049	0.0049	0.0080	−1.0047	0.0047	0.0078
$\widehat{\theta}^{(2)}$	0.4999	−0.0003	0.0051	−0.4999	−0.0001	0.0052	0.9998	−0.0002	0.0064	−0.9997	−0.0003	0.0062

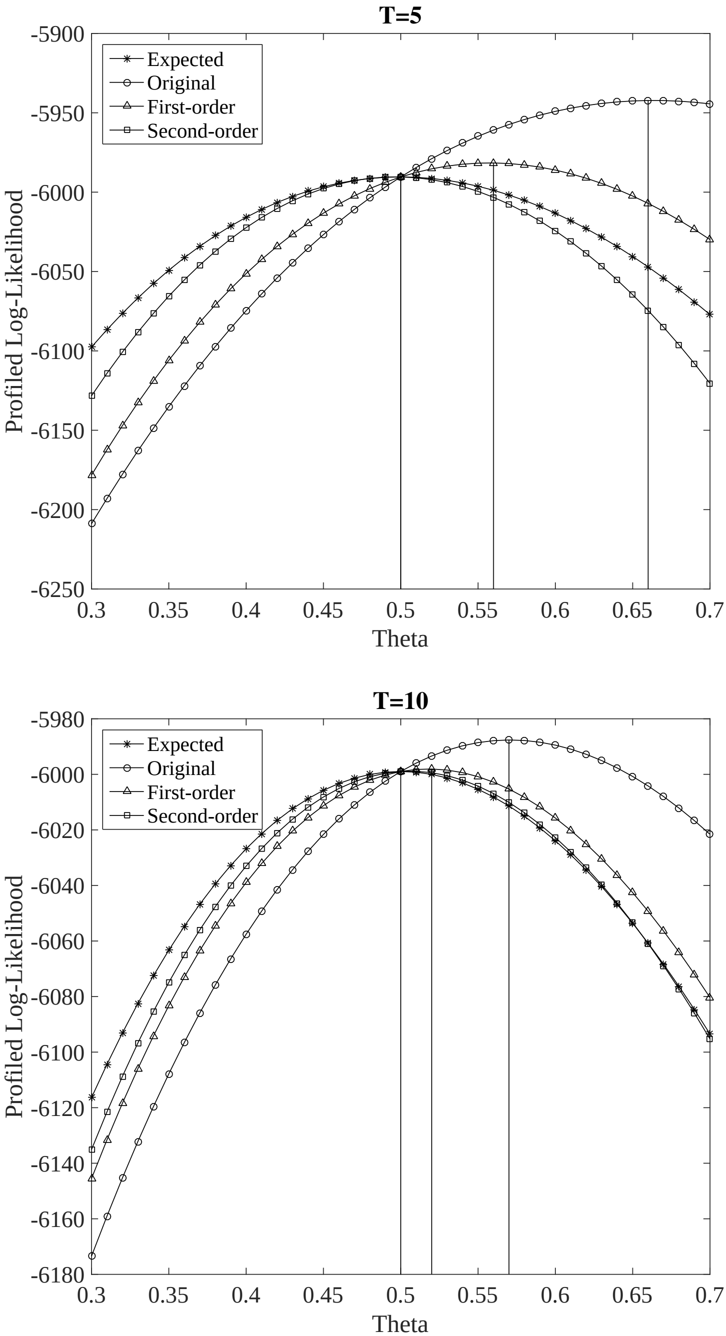
Notes: Bias is presented relative to θ_0. The number of replications is $1,000$, $N = 10,000$. Model: $Y_{it} = 1\left(X_{it}\theta_0 + \alpha_i + \varepsilon_{it} \geq 0\right)$ where ε_{it} is standard-logistically distributed, $X_{it} \sim \mathcal{N}(0,1)$, and $\alpha_i = \mathcal{N}(0, 1/16)$. $\widehat{\theta}$ is the original estimate, $\widehat{\theta}^{(1)}$ is the first-order corrected, and $\widehat{\theta}^{(2)}$ is the second-order corrected.

Table 3: Second Order - Simulation Results for Logit Model - Design 3

Setting	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE
$T=5$		$\theta_0=0.5$			$\theta_0=-0.5$			$\theta_0=1$			$\theta_0=-1$	
$\widehat{\theta}$	0.6311	0.2622	0.1318	-0.6309	0.2618	0.1316	1.2946	0.2946	0.2953	-1.2920	0.2920	0.2927
$\widehat{\theta}^{(1)}$	0.5382	0.0763	0.0398	-0.5378	0.0757	0.0395	1.0905	0.0905	0.0916	-1.0884	0.0884	0.0896
$\widehat{\theta}^{(2)}$	0.4904	-0.0193	0.0141	-0.4909	-0.0183	0.0137	0.9709	-0.0291	0.0316	-0.9718	-0.0282	0.0309
$T=10$		$\theta_0=0.5$			$\theta_0=-0.5$			$\theta_0=1$			$\theta_0=-1$	
$\widehat{\theta}$	0.5573	0.1146	0.0579	-0.5569	0.1138	0.0575	1.1244	0.1244	0.1248	-1.1230	0.1230	0.1234
$\widehat{\theta}^{(1)}$	0.5094	0.0188	0.0120	-0.5091	0.0182	0.0118	1.0205	0.0205	0.0224	-1.0197	0.0197	0.0219
$\widehat{\theta}^{(2)}$	0.4987	-0.0026	0.0074	-0.4988	-0.0024	0.0075	0.9956	-0.0044	0.0099	-0.9965	-0.0035	0.0099
$T=20$		$\theta_0=0.5$			$\theta_0=-0.5$			$\theta_0=1$			$\theta_0=-1$	
$\widehat{\theta}$	0.5266	0.0533	0.0272	-0.5266	0.0533	0.0272	1.0573	0.0573	0.0577	-1.0567	0.0567	0.0571
$\widehat{\theta}^{(1)}$	0.5021	0.0042	0.0055	-0.5022	0.0044	0.0055	1.0048	0.0048	0.0078	-1.0048	0.0048	0.0077
$\widehat{\theta}^{(2)}$	0.4997	-0.0006	0.0051	-0.4998	-0.0004	0.0050	0.9995	-0.0005	0.0062	-0.9998	-0.0002	0.0060

Notes: Bias is presented relative to θ_0. The number of replications is $1,000$, $N=10,000$. Model: $Y_{it}=1\left(X_{it}\theta_0+\alpha_i+\varepsilon_{it}\geq 0\right)$ where ε_{it} is standard-logistically distributed, $X_{it}\sim\mathcal{N}\left(\alpha_i,1\right)$, and $\alpha_i=\mathcal{N}\left(0,1/16\right)$. $\widehat{\theta}$ is the original estimate, $\widehat{\theta}^{(1)}$ is the first-order corrected, and $\widehat{\theta}^{(2)}$ is the second-order corrected.

Figure 1: Second Order - Plot of Profiled Log-likelihood for Logit



Notes: Computed on a single simulated dataset. $N = 10,000$. Model: $Y_{it} = 1\left(X_{it}\theta_0 + \alpha_i + \varepsilon_{it} \geq 0\right)$ where ε_{it} is standard-logistically distributed, $X_{it} \sim \mathcal{N}\left(0, 4\right)$, $\alpha_i = 0$, and $\theta_0 = 0.5$. θ chosen from 0.3 to 0.7 with a step of 0.01. Circle: $\sum_i \widehat{l}\left(\theta\right)$; triangle: $\sum_i \widehat{l}^{(1)}\left(\theta\right)$; square: $\sum_i \widehat{l}^{(2)}\left(\theta\right)$; asterisk: $\sum_i \mathbb{E}l\left(\theta\right)$. All curves are vertically shifted such that they coincide at θ_0. Vertical lines at maximizers.

4.3 Correction of Probit Model

The next example is the probit model. Researchers such as Greene et al. (2002) and Fernández-Val (2009) have also studied the probit model in detail. While the IPP in the logit model may be solved via the conditional logit, there seems to be little literature on the analytical solution of a probit model beyond the first order. Consider the model

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

where $\varepsilon_{it} \sim \mathcal{N}(0,1)$, and the individual log-likelihood function

$$\ln f\left(Y_{it};\theta,a_i\right) \equiv \ln\Phi\left[\left(2Y_{it}-1\right)\left(X_{it}\theta + a_i\right)\right].$$

where $\Phi(\cdot)$ is the standard normal distribution function.

Tables 4, 5, and 6 present the simulation results of the probit model under similar designs as explained in section 4.2. Compared to the logit model, we observe similar features in the probit model. The second-order correction is effective in reducing the bias. When $T = 5$ and $\theta_0 = 0.5$, $\widehat{\theta}$ is biased roughly 25% to 26% whereas $\widehat{\theta}^{(2)}$ is biased only 4% to 5% approximately. When $T = 20$, $\widehat{\theta}^{(2)}$ is almost unbiased under all designs. However, when comparing the RMSEs to the logit model, it seems that the variations in $\widehat{\theta}$, $\widehat{\theta}^{(1)}$, and $\widehat{\theta}^{(2)}$ are larger. For example in design 1, when $\theta_0 = 1$ and $T = 5$, the RMSEs of $\widehat{\theta}$ are larger than that of $\widehat{\theta}$ in the logit case and the RMSEs of $\widehat{\theta}^{(2)}$ are roughly 2 times that of $\widehat{\theta}^{(2)}$ in the logit case.

Another difference is that, in most presented cases, the bias in $\widehat{\theta}^{(2)}$ reduces by a factor smaller than 1/8 when T is doubled. For instance, in design 1, $B_5(-0.5)/B_{10}(-0.5) = 0.0356/0.0016 = 22.25$, which is higher than 8. This, however, does not contradict the assumption that $\mathbb{E}\left(\widehat{\theta}^{(2)} - \theta_0\right) = O\left(T^{-3}\right)$, since every quantity approaching 0 faster than T^{-3} as $T \to \infty$ is also $O\left(T^{-3}\right)$.

Figure 2 presents plots of the profiled log-likelihoods, for $T = 5$ and $T = 10$, of the probit model under the same setting as described in section 4.2 except that the model is replaced with $Y_{it} = 1\left(X_{it}\theta_0 + \alpha_i + \varepsilon_{it} \geq 0\right)$ where $\varepsilon_{it} \sim \mathcal{N}(0,1)$. Here the same pattern in the logit case follows. $\widehat{l}^{(2)}(\theta)$ serves as a better approximation than $\widehat{l}^{(1)}(\theta)$. The difference in the probit case is that, when T is 5, the maximizer of $\sum_i \widehat{l}^{(2)}(\theta)$ does not coincide with that of $\sum_i \mathbb{E}l(\theta)$. This is, in fact, in line with the simulation.

Table 4: Second Order - Simulation Results for Probit Model - Design 1

Setting	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE
$T=5$		$\theta_0 = 0.5$			$\theta_0 = -0.5$			$\theta_0 = 1$			$\theta_0 = -1$	
$\widehat{\theta}$	0.6239	0.2478	0.1242	−0.6237	0.2474	0.1240	1.3286	0.3286	0.3290	−1.3272	0.3272	0.3275
$\widehat{\theta}^{(1)}$	0.5581	0.1162	0.0587	−0.5579	0.1159	0.0585	1.1846	0.1846	0.1852	−1.1834	0.1834	0.1839
$\widehat{\theta}^{(2)}$	0.5180	0.0359	0.0195	−0.5178	**0.0356**	0.0193	1.0771	0.0771	0.0781	−1.0761	0.0761	0.0771
$T=10$		$\theta_0 = 0.5$			$\theta_0 = -0.5$			$\theta_0 = 1$			$\theta_0 = -1$	
$\widehat{\theta}$	0.5504	0.1008	0.0507	−0.5507	0.1013	0.0510	1.1297	0.1297	0.1299	−1.1292	0.1292	0.1295
$\widehat{\theta}^{(1)}$	0.5121	0.0242	0.0130	−0.5124	0.0247	0.0133	1.0378	0.0378	0.0384	−1.0374	0.0374	0.0381
$\widehat{\theta}^{(2)}$	0.5005	0.0011	0.0047	−0.5008	**0.0016**	0.0049	1.0040	0.0040	0.0077	−1.0036	0.0036	0.0077
$T=20$		$\theta_0 = 0.5$			$\theta_0 = -0.5$			$\theta_0 = 1$			$\theta_0 = -1$	
$\widehat{\theta}$	0.5229	0.0459	0.0232	−0.5231	0.0462	0.0233	1.0575	0.0575	0.0577	−1.0578	0.0578	0.0580
$\widehat{\theta}^{(1)}$	0.5027	0.0053	0.0043	−0.5028	0.0056	0.0044	1.0084	0.0084	0.0096	−1.0087	0.0087	0.0098
$\widehat{\theta}^{(2)}$	0.4998	−0.0004	0.0034	−0.4999	−0.0001	0.0033	0.9999	−0.0001	0.0046	−1.0002	0.0002	0.0044

Notes: Bias is presented relative to θ_0. The number of replications is $1,000$, $N = 10,000$. Model: $Y_{it} = 1\left(X_{it}\theta_0 + \alpha_i + \varepsilon_{it} \geq 0\right)$ where $\varepsilon_{it} \sim \mathcal{N}(0,1)$, $X_{it} \sim \mathcal{N}(0,1)$, and $\alpha_i = 0$. $\widehat{\theta}$ is the original estimate, $\widehat{\theta}^{(1)}$ is the first-order corrected, and $\widehat{\theta}^{(2)}$ is the second-order corrected.

Table 5: Second Order - Simulation Results for Probit Model - Design 2

Setting	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE
$T = 5$		$\theta_0 = 0.5$			$\theta_0 = -0.5$			$\theta_0 = 1$			$\theta_0 = -1$	
$\widehat{\theta}$	0.6266	0.2532	0.1270	-0.6271	0.2542	0.1274	1.3331	0.3331	0.3334	-1.3318	0.3318	0.3321
$\widehat{\theta}^{(1)}$	0.5604	0.1208	0.0609	-0.5608	0.1216	0.0613	1.1888	0.1888	0.1894	-1.1876	0.1876	0.1881
$\widehat{\theta}^{(2)}$	0.5200	0.0400	0.0213	-0.5203	0.0407	0.0216	1.0804	0.0804	0.0813	-1.0793	0.0793	0.0802
$T = 10$		$\theta_0 = 0.5$			$\theta_0 = -0.5$			$\theta_0 = 1$			$\theta_0 = -1$	
$\widehat{\theta}$	0.5518	0.1036	0.0521	-0.5518	0.1036	0.0521	1.1311	0.1311	0.1314	-1.1310	0.1310	0.1313
$\widehat{\theta}^{(1)}$	0.5130	0.0261	0.0139	-0.5130	0.0261	0.0140	1.0385	0.0385	0.0392	-1.0385	0.0385	0.0391
$\widehat{\theta}^{(2)}$	0.5012	0.0023	0.0050	-0.5012	0.0024	0.0050	1.0042	0.0042	0.0079	-1.0041	0.0041	0.0079
$T = 20$		$\theta_0 = 0.5$			$\theta_0 = -0.5$			$\theta_0 = 1$			$\theta_0 = -1$	
$\widehat{\theta}$	0.5235	0.0471	0.0238	-0.5236	0.0471	0.0238	1.0582	0.0582	0.0585	-1.0585	0.0585	0.0587
$\widehat{\theta}^{(1)}$	0.5029	0.0058	0.0044	-0.5029	0.0059	0.0044	1.0085	0.0085	0.0097	-1.0088	0.0088	0.0100
$\widehat{\theta}^{(2)}$	0.4999	-0.0001	0.0032	-0.5000	-0.0001	0.0033	0.9998	-0.0002	0.0045	-1.0000	0.0000	0.0047

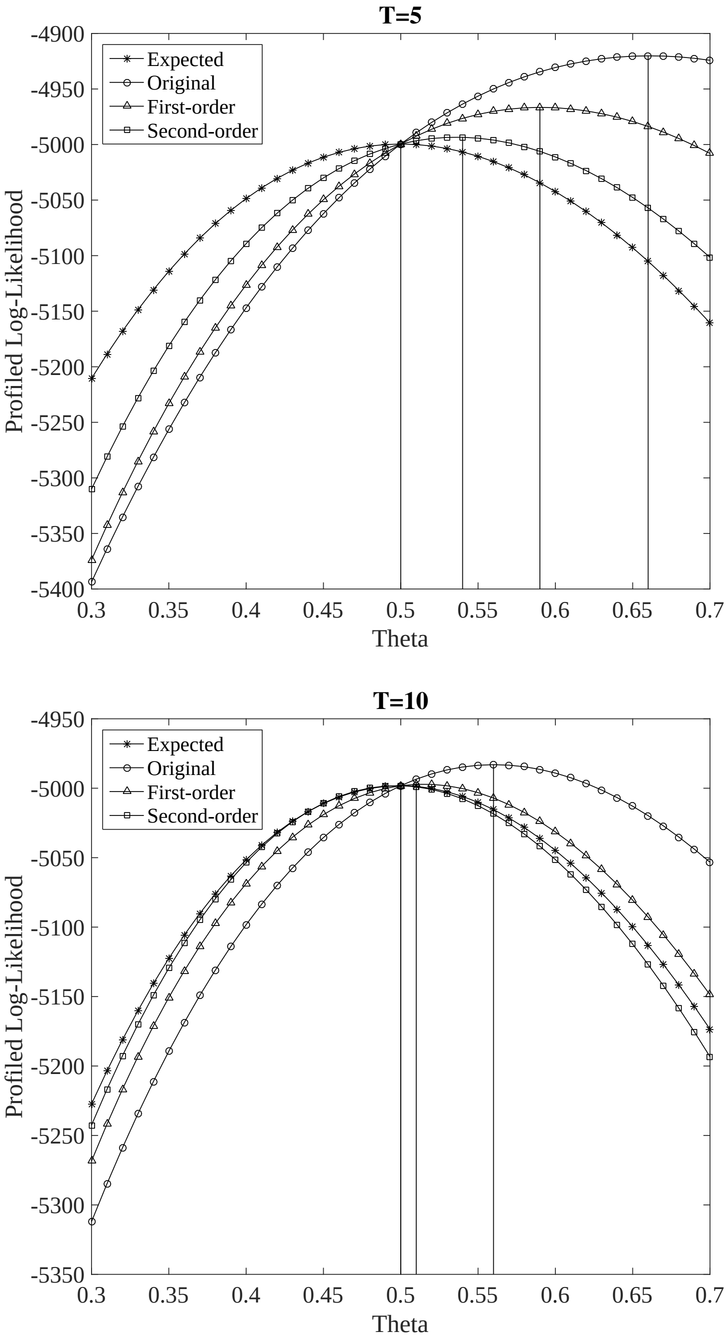
Notes: Bias is presented relative to θ_0. The number of replications is $1,000$, $N = 10,000$. Model: $Y_{it} = 1\left(X_{it}\theta_0 + \alpha_i + \varepsilon_{it} \geq 0\right)$ where $\varepsilon_{it} \sim \mathcal{N}\left(0,1\right)$, $X_{it} \sim \mathcal{N}\left(0,1\right)$, and $\alpha_i = \mathcal{N}\left(0,1/16\right)$. $\widehat{\theta}$ is the original estimate, $\widehat{\theta}^{(1)}$ is the first-order corrected, and $\widehat{\theta}^{(2)}$ is the second-order corrected.

Table 6: Second Order - Simulation Results for Probit Model - Design 3

Setting	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE
$T=5$	$\theta_0=0.5$			$\theta_0=-0.5$			$\theta_0=1$			$\theta_0=-1$		
$\widehat{\theta}$	0.6310	0.2620	0.1313	−0.6241	0.2481	0.1244	1.3445	0.3445	0.3449	−1.3275	0.3275	0.3278
$\widehat{\theta}^{(1)}$	0.5642	0.1283	0.0646	−0.5582	0.1164	0.0588	1.1994	0.1994	0.1999	−1.1837	0.1837	0.1842
$\widehat{\theta}^{(2)}$	0.5233	0.0466	0.0244	−0.5180	0.0361	0.0196	1.0886	0.0886	0.0894	−1.0764	0.0764	0.0773
$T=10$	$\theta_0=0.5$			$\theta_0=-0.5$			$\theta_0=1$			$\theta_0=-1$		
$\widehat{\theta}$	0.5536	0.1072	0.0539	−0.5507	0.1015	0.0510	1.1379	0.1379	0.1382	−1.1290	0.1290	0.1293
$\widehat{\theta}^{(1)}$	0.5142	0.0283	0.0151	−0.5123	0.0247	0.0133	1.0433	0.0433	0.0439	−1.0372	0.0372	0.0379
$\widehat{\theta}^{(2)}$	0.5020	0.0040	0.0054	−0.5007	0.0014	0.0050	1.0073	0.0073	0.0099	−1.0034	0.0034	0.0077
$T=20$	$\theta_0=0.5$			$\theta_0=-0.5$			$\theta_0=1$			$\theta_0=-1$		
$\widehat{\theta}$	0.5243	0.0486	0.0246	−0.5232	0.0463	0.0234	1.0612	0.0612	0.0614	−1.0577	0.0577	0.0579
$\widehat{\theta}^{(1)}$	0.5033	0.0065	0.0047	−0.5028	0.0056	0.0042	1.0098	0.0098	0.0109	−1.0086	0.0086	0.0097
$\widehat{\theta}^{(2)}$	0.5001	0.0003	0.0033	−0.4999	−0.0002	0.0032	1.0004	0.0004	0.0046	−1.0001	0.0001	0.0044

Notes: Bias is presented relative to θ_0. The number of replications is $1,000$, $N=10,000$. Model: $Y_{it}=1\left(X_{it}\theta_0+\alpha_i+\varepsilon_{it}\geq 0\right)$ where $\varepsilon_{it}\sim\mathcal{N}(0,1)$, $X_{it}\sim\mathcal{N}(\alpha_i,1)$, and $\alpha_i=\mathcal{N}(0,1/16)$. $\widehat{\theta}$ is the original estimate, $\widehat{\theta}^{(1)}$ is the first-order corrected, and $\widehat{\theta}^{(2)}$ is the second-order corrected.

Figure 2: Second Order - Plot of Profiled Log-likelihood for Probit



Notes: Computed on a single simulated dataset. $N = 10,000$. Model: $Y_{it} = 1\left(X_{it}\theta_0 + \alpha_i + \varepsilon_{it} \geq 0\right)$ where $\varepsilon_{it} \sim \mathcal{N}(0,1)$, $X_{it} \sim \mathcal{N}(0,4)$, $\alpha_i = 0$, and $\theta_0 = 0.5$. θ chosen from 0.3 to 0.7 with a step of 0.01. Circle: $\sum_i \widehat{l}(\theta)$; triangle: $\sum_i \widehat{l}^{(1)}(\theta)$; square: $\sum_i \widehat{l}^{(2)}(\theta)$; asterisk: $\sum_i \mathbb{E}l(\theta)$. All curves are vertically shifted such that they coincide at θ_0. Vertical lines at maximizers.

4.4 Application to Poisson

The last example we present is the static Poisson model, which is not exposed to IPP. This is noted by, e.g., Lancaster (2002). Therefore, we briefly investigate the consequence of an application of the correction technique to a model that is not exposed to IPP. Consider[7] Y_{it} being Poisson-distributed with mean $\exp(X_{it}\theta_0 + \alpha_i)$ where $X_{it} \sim \mathcal{U}(0,1)$, $\alpha_i = 0$, and $\theta_0 = 0.5$. Table 7 presents the simulation result for the Poisson model. It can be found that, when applied to the Poisson model, the correction technique would generate a bias that is relatively insignificant in percentage. Such bias increases with the order of correction, which may be a consequence of introducing extra variation into the model; i.e., the variation of $\widehat{b}_1$ and $\widehat{b}_2$ may affect $\widehat{\theta}$ in an undesired way. However, this bias approaches 0 fairly fast when T increases.

Table 7: Second Order - Simulation Results for Poisson Model

	Mean	Bias	RMSE	Mean	Bias	RMSE
	$T = 10$			$T = 20$		
$\widehat{\theta}$	0.4997	-0.0005	0.0099	0.5002	0.0004	0.0069
$\widehat{\theta}^{(1)}$	0.4984	-0.0033	0.0100	0.4998	-0.0004	0.0069
$\widehat{\theta}^{(2)}$	0.4969	-0.0061	0.0103	0.4996	-0.0009	0.0069

Notes: Bias is presented relative to θ_0. The number of replications is $1,000$, $N = 10,000$. Model: Y_{it} Poisson-distributed with mean $\exp(X_{it}\theta_0 + \alpha_i)$ with $X_{it} \sim \mathcal{U}(0,1)$, $\alpha_i = 0$, and $\theta_0 = 0.5$. $\widehat{\theta}$ is the original estimate, $\widehat{\theta}^{(1)}$ is the first-order corrected, and $\widehat{\theta}^{(2)}$ is the second-order corrected.

5 Concluding Remarks

We propose a second-order corrected log-likelihood function such that, when this log-likelihood function is maximized, the resulting estimator of θ possesses a bias that is less than the original estimator $\widehat{\theta}$. The corrected log-likelihood function serves as an approximation, to the order of $O_p\left(T^{-5/2}\right)$, to the infeasible log-likelihood l and, under the asymptotic sequence of $N/T^5 \to 0$ as $N, T \to \infty$ and certain regularity conditions, the proposed approximation is consistent for l (and hence $\mathbb{E}l$). The proposed technique applies, up to certain assumptions, to any density or mass functions that are smooth in the sense that the fourth derivative exists. In addition, our approach naturally extends to unbalanced panels, since the corrected log-likelihood function is derived for a single i independently of T.

Our research casts some light on several subjects that are worth studying. First, we have not studied the correction of the variance estimator. As noted by, e.g., Dhaene and Jochmans (2015), the ML asymptotic variance is too small that inferences based on such variance may produce, e.g., confidence intervals that are too narrow when T is small. Therefore, a correction of the variance may be beneficial for small-T samples. An alternative research direction concerning the variance may be to study how various variance estimators such as the “sandwich” variance estimator - see, e.g., Freedman (2012) - can incorporate the corrected log-likelihood.

Second, we have not looked at higher-order corrections under the presence of IPP. When T is small, a higher-order correction may produce more advantageous bias-corrected esti-

[7]See, e.g., Cameron and Trivedi (2013) for details about a fixed-effect Poisson model.

mates as compared to the one produced by the second-order correction and, therefore, seems to be worth pursuing.

Third, as the second-order correction introduces an approximation to the infeasible log-likelihood, it would be of interest to study how the approximation may benefit statistical tests and other inferences based on the likelihood. Examples of these inferences include the likelihood ratio test or the AIC. Alternatively, it would be useful to study how the proposed approach extends to other functions such as the average effects or the pseudo-likelihood.

Last but not least, as the IPP under dynamic models is usually much more severe than under static models, it would be of particular interest to investigate how our approach extends to dynamic models. For the linear autoregressive model, Nickell (1981) develops an analytical formula for the bias of the estimator of the autoregressive parameter. For nonlinear autoregressive models, there seems to be little higher-order development on this subject. Specifically to our approach, a modified version of the weights described in Arellano and Hahn (2006) may apply.

6 Appendix

A Higher-order Expansion of ML Estimator

$$(\widehat{\alpha}-\alpha)=\sum\nolimits_{j=1}^{8} a_{j/2}+O_p\left(T^{-\frac{9}{2}}\right)$$

where

$$\begin{aligned}
a_{1/2} &= -\frac{l_1}{l_2}, \qquad a_{2/2}=-\frac{l_1^2 l_3}{2l_2^3}, \qquad a_{3/2}=\frac{l_1^3 l_4}{6l_2^4}-\frac{l_1^3 l_3^2}{2l_2^5},\\
a_{4/2} &= \frac{5l_1^4 l_3 l_4}{12l_2^6}-\frac{5l_1^4 l_3^3}{8l_2^7}-\frac{l_1^4 l_5}{24l_2^5},\\
a_{5/2} &= \frac{l_1^5 l_6}{120l_2^6}-\frac{l_1^5 l_4^2}{12l_2^7}-\frac{7l_1^5 l_3^4}{8l_2^9}-\frac{l_1^5 l_3 l_5}{8l_2^7}+\frac{7l_1^5 l_3^2 l_4}{8l_2^8},\\
a_{6/2} &= \frac{7l_1^6 l_3 l_6}{240l_2^8}-\frac{21l_1^6 l_3^5}{16l_2^{11}}-\frac{l_1^6 l_7}{720l_2^7}+\frac{7l_1^6 l_4 l_5}{144l_2^8}-\frac{7l_1^6 l_3 l_4^2}{18l_2^9}-\frac{7l_1^6 l_3^2 l_5}{24l_2^9}+\frac{7l_1^6 l_3^3 l_4}{4l_2^{10}},\\
a_{7/2} &= \frac{l_1^7 l_8}{5040l_2^8}-\frac{l_1^7 l_5^2}{144l_2^9}+\frac{l_1^7 l_4^3}{18l_2^{10}}-\frac{33l_1^7 l_3^6}{16l_2^{13}}-\frac{5l_1^7 l_3^2 l_4^2}{4l_2^{11}}-\frac{l_1^7 l_3 l_7}{180l_2^9}-\frac{l_1^7 l_4 l_6}{90l_2^9}\\
&\quad +\frac{3l_1^7 l_3^2 l_6}{40l_2^{10}}-\frac{5l_1^7 l_3^3 l_5}{8l_2^{11}}+\frac{55l_1^7 l_3^4 l_4}{16l_2^{12}}+\frac{l_1^7 l_3 l_4 l_5}{4l_2^{10}},\\
a_{8/2} &= \frac{l_1^8 l_3 l_8}{1120l_2^{10}}-\frac{429l_1^8 l_3^7}{128l_2^{15}}-\frac{55l_1^8 l_3^3 l_4^2}{16l_2^{13}}-\frac{l_1^8 l_9}{40320l_2^9}+\frac{l_1^8 l_4 l_7}{480l_2^{10}}+\frac{l_1^8 l_5 l_6}{320l_2^{10}}\\
&\quad -\frac{5l_1^8 l_3 l_5^2}{128l_2^{11}}-\frac{5l_1^8 l_4^2 l_5}{96l_2^{11}}+\frac{55l_1^8 l_3 l_4^3}{144l_2^{12}}-\frac{l_1^8 l_3^2 l_7}{64l_2^{11}}+\frac{11l_1^8 l_3^3 l_6}{64l_2^{12}}-\frac{165l_1^8 l_3^4 l_5}{128l_2^{13}}\\
&\quad +\frac{429l_1^8 l_3^5 l_4}{64l_2^{14}}+\frac{55l_1^8 l_3^2 l_4 l_5}{64l_2^{12}}-\frac{l_1^8 l_3 l_4 l_6}{16l_2^{11}}.
\end{aligned}$$

B Proof of Proposition 1

We use the following extra notation in appendix B, C, and D. Let R_j and P_j be the j-th row of, respectively, R and P; and let

$$\begin{aligned} s_{it}\left(R_j, P_j\right) &\equiv \prod_{m=1}^{M} \left(\nabla_a^{r_{jm}} \ln f\left(Y_{it}; \theta, a\right)\big|_{a=\alpha}\right)^{p_{jm}}, \\ \mathcal{S}\left(R, P\right) &\equiv \sum_{(t_1,\cdots,t_J)\in\mathcal{T}} \prod_{j=1}^{J} s_{it_j}\left(R_j, P_j\right). \end{aligned}$$

Note also that

$$\begin{aligned} J &= \left|\mathcal{J}_0\left(R,P\right)\right| + \left|\mathcal{J}_1\left(R,P\right)\right|, \\ \left|\mathcal{J}_0\left(R,P\right)\right| &= \sum_{j=1}^{J} 1\left(\sum_{m=1}^{M} r_{jm} = 1 \wedge \sum_{m=1}^{M} p_{jm} = 1\right). \end{aligned}$$

As $T \to \infty$, the stochastic variable $s_{it}\left(R_j, P_j\right)$ satisfied

$$\frac{1}{T}\sum_t s_{it}\left(R_j, P_j\right) \xrightarrow{p} \mathbb{E}s_{it}\left(R_j, P_j\right),$$
$$\frac{1}{T}\sum_t s_{it}\left(R_j, P_j\right) = \mathbb{E}s_{it}\left(R_j, P_j\right) + O_p\left(T^{-\frac{1}{2}}\right).$$

It follows that, when $j \in \mathcal{J}_1\left(R,P\right)$ such that $\mathbb{E}s_{it}\left(R_j, P_j\right) \neq 0$,

$$\sum_t s_{it}\left(R_j, P_j\right) = O_p\left(T\right);$$

whereas, when $j \in \mathcal{J}_0\left(R,P\right)$ such that $\mathbb{E}s_{it}\left(R_j, P_j\right) = 0$,

$$\sum_t s_{it}\left(R_j, P_j\right) = O_p\left(T^{\frac{1}{2}}\right).$$

Now

$$\mathcal{S}\left(R,P\right) = \sum_{(t_1,\cdots,t_J)\in\mathcal{T}} \left(\prod_{j\in\mathcal{J}_0(R,P)} s_{it_j}\left(R_j, P_j\right)\right)\left(\prod_{j\in\mathcal{J}_1(R,P)} s_{it_j}\left(R_j, P_j\right)\right)$$

is a J-fold summation with each fold being $\sum_t s_{it}\left(R_j, P_j\right)$. Therefore,

$$\begin{aligned} \mathcal{S}\left(R,P\right) &= \left(\prod_{j\in\mathcal{J}_0(R,P)} O_p\left(T^{\frac{1}{2}}\right)\right)\left(\prod_{j\in\mathcal{J}_1(R,P)} O_p\left(T\right)\right) \\ &= O_p\left(T^{\frac{\left|\mathcal{J}_0(R,P)\right|}{2}}\right) O_p\left(T^{\left|\mathcal{J}_1(R,P)\right|}\right) \\ &= O_p\left(T^{\left|\mathcal{J}_1(R,P)\right| + \frac{1}{2}\left|\mathcal{J}_0(R,P)\right|}\right). \end{aligned}$$

Since $\mathcal{P}\left(R,P\right) = J - \frac{1}{2}\left|\mathcal{J}_0\left(R,P\right)\right| = \left|\mathcal{J}_1\left(R,P\right)\right| + \frac{1}{2}\left|\mathcal{J}_0\left(R,P\right)\right|$, it is obvious that

$$\mathcal{L}\left(R,P\right) = \frac{1}{T^{J-\frac{1}{2}\left|\mathcal{J}_0(R,P)\right|}}\mathcal{S}\left(R,P\right) = O_p\left(1\right).$$

C Proof of Lemma 1

$\mathcal{P}(R,P) < J$ is equivalent to $|\mathcal{J}_0(R,P)| > 0$, and hence, it follows that

$$\mathcal{S}(R,P) = \sum_{(t_1,\cdots,t_J)\in\mathcal{T}} \left(\prod_{j\in\mathcal{J}_1(R,P)} s_{it_j}(R_j,P_j)\right)\left(\prod_{j\in\mathcal{J}_0(R,P)} s_{it_j}(R_j,P_j)\right).$$

Now by independence,

$$\mathbb{E}\mathcal{S}(R,P) = \sum_{(t_1,\cdots,t_J)\in\mathcal{T}} \left(\prod_{j\in\mathcal{J}_1(R,P)} \mathbb{E}s_{it_j}(R_j,P_j)\right)\left(\prod_{j\in\mathcal{J}_0(R,P)} \mathbb{E}s_{it_j}(R_j,P_j)\right)$$

in which, since $\mathbb{E}s_{it}(R_j,P_j) = 0$ for $j \in \mathcal{J}_0(R,P)$, $\mathbb{E}\mathcal{S}(R,P) = 0$ such that $\mathbb{E}\mathcal{L}(R,P) = 0$.

D Product of $\mathcal{L}(R,P)$ and $\mathcal{L}(R',P')$

The product of several $\mathcal{L}(R,P)$ and/or several l_r can be computed iteratively by rewriting l_r into $\mathcal{L}(R,P)$. To calculate $\mathcal{L}(R,P)\mathcal{L}(R',P')$, we shall focus on the calculation of $\mathcal{S}(R,P)\mathcal{S}(R',P')$. Let

$$R \equiv \begin{pmatrix} r_{11} & \cdots & r_{1M} \\ \vdots & \ddots & \vdots \\ r_{J1} & \cdots & r_{JM} \end{pmatrix}, \qquad P \equiv \begin{pmatrix} p_{11} & \cdots & p_{1M} \\ \vdots & \ddots & \vdots \\ p_{J1} & \cdots & p_{JM} \end{pmatrix},$$
$$R' \equiv \begin{pmatrix} r'_{11} & \cdots & r'_{1M'} \\ \vdots & \ddots & \vdots \\ r'_{J'1} & \cdots & r'_{J'M'} \end{pmatrix}, \qquad P' \equiv \begin{pmatrix} p'_{11} & \cdots & p'_{1M'} \\ \vdots & \ddots & \vdots \\ p'_{J'1} & \cdots & p'_{J'M'} \end{pmatrix}$$

where we suppose w.l.o.g. that $J' \le J$; let

$$c_j \equiv \binom{P_j}{R_j}, \qquad c'_{j'} \equiv \binom{P'_{j'}}{R'_{j'}};$$

and let

$$\langle c_j, c'_{j'}\rangle \equiv \binom{P_j, P'_{j'}}{R_j, R'_{j'}}$$

in which $P_j, P'_{j'}$ and $R_j, R'_{j'}$ are simply row-joined respectively. Now let

$$\mathcal{S}(c_1,\cdots,c_J) \equiv \mathcal{S}(R,P),\ \mathcal{S}\left(c'_1,\cdots,c'_{J'}\right) \equiv \mathcal{S}\left(R',P'\right)$$

in which pairs p_{jm}, r_{jm} and/or $p'_{j'm'}, r'_{j'm'}$ are removed if, respectively, $p_{jm} = r_{jm} = 0$ and/or $p'_{j'm'} = r'_{j'm'} = 0$. For $z = 0,\cdots,J'$ and for given $j_1,\cdots,j_z,j'_1,\cdots,j'_z$, let

$$\mathcal{C}_{j\neq j_1,\cdots,j_z} \equiv \{c_j | j = 1,\cdots,J; j \neq j_1,\cdots,j_z\},$$
$$\mathcal{C}'_{j\neq j'_1,\cdots,j'_z} \equiv \left\{c'_j | j = 1,\cdots,J'; j \neq j'_1,\cdots,j'_z\right\}$$

such that, e.g.,

$$\mathcal{S}\left(c_1,\cdots,c_J\right)\equiv\mathcal{S}\left(c_1,\cdots,c_n,\mathcal{C}_{j\neq 1,\cdots,n}\right),$$
$$\mathcal{S}\left(c'_1,\cdots,c'_{J'}\right)\equiv\mathcal{S}\left(c'_1,\cdots,c'_n,\mathcal{C}'_{j\neq 1,\cdots,n}\right)$$

for every nonnegative integer $n\leq J'$. Here $\mathcal{C}_{j\neq j_1,\cdots,j_z}$ (or $\mathcal{C}'_{j\neq j'_1,\cdots,j'_z}$) serves as a collection of c_j (or $c'_{j'}$) for $j\neq j_1,\cdots,j_z$ (or $j'\neq j'_1,\cdots,j'_z$). It needs to be noted that $j_1,\cdots,j_z$ may not necessarily be identical to $j'_1,\cdots,j'_z$, and that, even when they are indeed identical, $\mathcal{C}_{j\neq j_1,\cdots,j_z}$ and $\mathcal{C}'_{j\neq j'_1,\cdots,j'_z}$ may still be different.

Suppose, for every positive integer $u,v\leq J'$, $j_u\neq j_v\forall u\neq v$ and $j'_u\neq j'_v\forall u\neq v$. It then can be calculated that

$$\begin{aligned}
&\mathcal{S}\left(c_1,\cdots,c_J\right)\mathcal{S}\left(c'_1,\cdots,c'_{J'}\right)\\
=\;&\mathcal{S}\left(c_1,\cdots,c_J,c'_1,\cdots,c'_{J'}\right)\\
&+\mathcal{S}\left(\left\langle c_1,c'_1\right\rangle,\cdots,c_J,c'_2,\cdots,c'_{J'}\right)+\cdots+\mathcal{S}\left(c_1,\cdots,\left\langle c_J,c'_1\right\rangle,c'_2,\cdots,c'_{J'}\right)\\
&+\cdots\\
&+\mathcal{S}\left(\left\langle c_1,c'_{J'}\right\rangle,\cdots,c_J,c'_1,\cdots,c'_{J'-1}\right)+\cdots+\mathcal{S}\left(c_1,\cdots,\left\langle c_J,c'_{J'}\right\rangle,c'_1,\cdots,c'_{J'-1}\right)\\
&+\sum_{\substack{j_1<j_2\in(1,\cdots,J)\\ j'_1,j'_2\in(1,\cdots,J')}}\mathcal{S}\left(c_1,\cdots,\left\langle c_{j_1},c'_{j'_1}\right\rangle,\cdots,\left\langle c_{j_2},c'_{j'_2}\right\rangle,\cdots,c_J,\mathcal{C}'_{j\neq j'_1,j'_2}\right)\\
&+\sum_{\substack{j_1<j_2<j_3\in(1,\cdots,J)\\ j'_1,j'_2,j'_3\in(1,\cdots,J')}}\mathcal{S}\left(c_1,\cdots,\left\langle c_{j_1},c'_{j'_1}\right\rangle,\cdots,\left\langle c_{j_3},c'_{j'_3}\right\rangle,\cdots,c_J,\mathcal{C}'_{j\neq j'_1,j'_2,j'_3}\right)\\
&+\sum_{z\in(4,\cdots,J')}\sum_{\substack{j_1<\cdots<j_z\in(1,\cdots,J)\\ j'_1,\cdots,j'_z\in(1,\cdots,J')}}\mathcal{S}\left(\left\langle c_{j_1},c'_{j'_1}\right\rangle,\cdots,\left\langle c_{j_z},c'_{j'_z}\right\rangle,\mathcal{C}_{j\neq j_1,\cdots,j_z},\mathcal{C}'_{j\neq j'_1,\cdots,j'_z}\right)
\end{aligned}$$

or, in a more compact form,

$$\begin{aligned}
&\mathcal{S}\left(c_1,\cdots,c_J\right)\mathcal{S}\left(c'_1,\cdots,c'_{J'}\right)\\
=\;&\sum_{z\in(0,\cdots,J')}\sum_{\substack{j_1<\cdots<j_z\in(1,\cdots,J)\\ j'_1,\cdots,j'_z\in(1,\cdots,J')}}\mathcal{S}\left(\left\langle c_{j_1},c'_{j'_1}\right\rangle,\cdots,\left\langle c_{j_z},c'_{j'_z}\right\rangle,\mathcal{C}_{j\neq j_1,\cdots,j_z},\mathcal{C}'_{j\neq j'_1,\cdots,j'_z}\right)
\end{aligned}$$

in which

$$\mathcal{S}\left(\left\langle c_{j_1},c'_{j'_1}\right\rangle,\cdots,\left\langle c_{j_z},c'_{j'_z}\right\rangle,\mathcal{C}_{j\neq j_1,\cdots,j_z},\mathcal{C}'_{j\neq j'_1,\cdots,j'_z}\right)\equiv 0$$

if

$$\begin{aligned}
T &< z+\left|\mathcal{C}_{j\neq j_1,\cdots,j_z}\right|+\left|\mathcal{C}'_{j\neq j'_1,\cdots,j'_z}\right|=J+J'-z\\
z &< J+J'-T.
\end{aligned}$$

Furthermore, when T is fixed, it can be simplified that

$$\mathcal{S}\left(c_1,\cdots,c_J\right)\mathcal{S}\left(c'_1,\cdots,c'_{J'}\right)$$

$$= \sum_{\max(J+J'-T,0)\leq z\leq J'} \sum_{\substack{j_1<\cdots<j_z\in(1,\cdots,J)\\ j'_1,\cdots,j'_z\in(1,\cdots,J')}} \mathcal{S}\left(\begin{array}{c} \left\langle c_{j_1}, c'_{j'_1}\right\rangle, \cdots, \left\langle c_{j_z}, c'_{j'_z}\right\rangle, \\ \mathcal{C}_{j\neq j_1,\cdots,j_z}, \mathcal{C}'_{j\neq j'_1,\cdots,j'_z} \end{array} \right).$$

References

Andersen, E. B. (1970). Asymptotic properties of conditional maximum-likelihood estimators. *Journal of the Royal Statistical Society. Series B (Methodological) 32*(2), 283–301.

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.

Bartlett, M. S. (1953a). Approximate confidence intervals. *Biometrika*, 12–19.

Bartlett, M. S. (1953b). Approximate confidence intervals. ii. more than one unknown parameter. *Biometrika*, 306–317.

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

Cameron, A. C. and P. K. Trivedi (2013). *Regression analysis of count data*, Volume 53. Cambridge university press.

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

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.

Darst, R. (1973). Most infinitely differentiable functions are nowhere analytic. *Canad. Math. Bull 16*, 597–598.

de Paschal, J. N. (1961). Unbiased ratio estimators in stratified sampling. *Journal of the American Statistical Association 56*(293), 70–87.

Dhaene, G. and K. Jochmans (2015). Split-panel jackknife estimation of fixed-effect models. *The Review of Economic Studies*, rdv007.

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

Ferrari, S. L., D. A. Botter, G. M. Cordeiro, and F. Cribari-Neto (1996). Second-and third-order bias reduction for one-parameter family models. *Statistics & probability letters 30*(4), 339–345.

Freedman, D. A. (2012). On the so-called huber sandwich estimator and robust standard errors. *The American Statistician.*

Greene, W., C. Han, and P. Schmidt (2002). The bias of the fixed effects estimator in nonlinear models. *Unpublished Manuscript, Stern School of Business, NYU*.

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

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

Haldane, J. and S. M. Smith (1956). The sampling distribution of a maximum-likelihood estimate. *Biometrika*, 96–103.

Hartley, H. and A. Ross (1954). Unbiased ratio estimators.

Heckman, J. (1981). The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. *Manski, C.F., McFadden, D. (Eds.), Structural Analysis of Discrete Data with Econometric Applications. MIT Press, Cambridge, MA.*.

Hille, E. (2005). *Analytic function theory*, Volume 2. American Mathematical Soc.

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.

Newey, W. K. and D. McFadden (1994). *Handbook of Econometrics*. Elsevier.

Neyman, J. and E. L. Scott (1948). Consistent estimates based on partially consistent observations. *Econometrica: Journal of the Econometric Society*, 1–32.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica: Journal of the Econometric Society*, 1417–1426.

Ogliore, R., G. Huss, and K. Nagashima (2011). Ratio estimation in sims analysis. *Nuclear instruments and methods in physics research section B: beam interactions with materials and atoms 269*(17), 1910–1918.

Pace, L. and A. Salvan (1997). *Principles of statistical inference: from a Neo-Fisherian perspective*, Volume 4. World Scientific.

Provost, S. B. and W. M. Ratemi (2011). Polynomial expansions via embedded pascals triangles. *Acta et Commentationes Universitatis Tartuensis de Mathematica 15-1*, 45–60.

Rice, S. H. (2008). A stochastic version of the price equation reveals the interplay of deterministic and stochastic processes in evolution. *BMC evolutionary biology 8*(1), 262.

Rilstone, P., V. Srivastava, and A. Ullah (1996). The second-order bias and mean squared error of nonlinear estimators. *Journal of Econometrics 75*(2), 369–395.