Economics Letters 219 (2022) 110784

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Jackknife bias reduction for simulated maximum likelihood estimator of discrete choice models

Jinyong Hahn *, Xueyuan Liu

UCLA Economics, BOX 951477, 8283 Bunche Hall, Los Angeles, CA 90095-1477, USA

ARTICLE INFO

Article history:
Received 5 June 2022
Received in revised form 23 July 2022
Accepted 2 August 2022
Available online 6 August 2022

JEL classification:
C15

Keywords:
Simulated maximum likelihood estimator
Split sample jackknife

ABSTRACT

We propose to reduce asymptotic biases of simulated maximum likelihood estimators (SMLE) by using a jackknife method similar to Dhaene and Jochmans (2015). Because the jackknife method does not require an explicit characterization of the bias, it may be a practically attractive alternative to Lee's (1995) estimator.

1. Introduction

We address the small sample bias of the simulated maximum likelihood estimator (SMLE). The SMLE was introduced primarily because in many models of discrete choice, the maximum likelihood estimation (MLE) is computationally impossible for all practical purpose. See Lerman and Manski (1981), McFadden (1989), Pakes and Pollard (1989), Lee (1992), or Hajivassiliou and Ruud (1994) for review of early literature. In order to derive the asymptotic normality of the SMLE, the number of simulation draws is often assumed to go to infinity sufficiently fast as a function of the sample size. Lee (1995) investigates the asymptotic bias of the SMLE, and derives the analytical formula of higher order bias due to simulation.[1] Lee (1995) then goes on and constructs bias-adjusted SMLE's by estimating such a bias using the analytic formula.[2]

Implementation of Lee's (1995) method requires analytical characterization of the higher order bias (due to simulation), which may not be convenient for practice. We propose to bypass the analytical characterization of the bias by modifying the split-sample (SS) jackknife method due to Dhaene and Jochmans (2015), which was originally proposed to reduce bias in nonlinear panel models. Dhaene and Jochmans (2015) intuition may be attractive for practice because it only requires computation of the SMLE a few times.

Our main results are presented in Section 2. All the proofs are collected in Section 3.

2. Main results

We start with a review of Lee (1995). Consider a standard model of discrete responses. Let $C = \{1, \dots, L\}$ be a set of mutually exclusive and exhaustive alternatives. For each alternative $l \in C$, let $P(l|\theta, x)$ denote the probability that such alternative is chosen, where x denotes the vector consisting of all distinct explanatory variables, and θ denotes the K-dimensional parameter. Let d_{li} denote a response indicator for individual i, equal to one when the observed response is the alternative l and zero otherwise. With a sample of size n of independent observations, the log likelihood function for the discrete choice model is

$$\mathcal{L}_c(\theta) \equiv \sum_{i=1}^{n} \sum_{l=1}^{L} d_{li} \ln P(l|\theta, x_i). \tag{1}$$

The classical MLE for θ is derived from the maximization of $\mathcal{L}_c(\theta)$. It is a solution of the score equation

$$\frac{1}{n} \sum_{i=1}^{n} \sum_{l=1}^{L} d_{li} \frac{\partial \ln P(l|\theta, x_i)}{\partial \theta} = 0. \tag{2}$$

* Corresponding author.
E-mail addresses: hahn@econ.ucla.edu (J. Hahn), liuxueyuan@ucla.edu (X. Liu).

[1] This bias is conceptually different from the generic second order bias that is due to the nonlinearity of the potentially infeasible MLE. See Lee (1995) for more discussion.

[2] Lee (1995) considers two versions of SMLE, one based on independently simulated moments and another one based on dependently simulated moments. Lee (1995, p. 449) notes that the latter has a larger asymptotic variance than the former, so we will focus our attention to the case of independently simulated moments.

https://doi.org/10.1016/j.econlet.2022.110784
0165-1765/

If the choice probabilities $P(l|\theta, x_i)$ are difficult to compute, we may want to use an unbiased simulator. Let $\gamma(v)$ be a density chosen for simulation. Assuming that $h_l(v, x, \theta)$ satisfies $P(l|\theta, x) = \int h_l(v, x, \theta)\gamma(v)\,dv$, we can work with the unbiased simulator

$$f_{r,l}(\theta, x_i) \equiv \frac{1}{r}\sum_{j=1}^{r} h_l\left(v_j^{(i)}, x_i, \theta\right), \tag{3}$$

where $v_j^{(i)}$, $j = 1, \ldots, r$ denote r Monte Carlo draws for observation i from $\gamma(v)$. Because $E\left[f_{r,l}(\theta, x_i)\middle| x_i\right] = P(l|\theta, x_i)$, the $f_{r,l}(\theta, x_i)$ is a conditionally unbiased simulator. The SMLE $\hat{\theta}_l$ is obtained by maximizing the simulated likelihood function

$$\mathcal{L}(\theta) \equiv \sum_{i=1}^{n}\sum_{l=1}^{L} d_{li} \ln f_{r,l}(\theta, x_i).$$

Lee (1995) imposes three assumptions. His Assumptions 1 and 2 are largely technical regularity conditions, and we impose them without explicitly spelling them out. As for his Assumption 3, which requires that $r \to \infty$ as $n \to \infty$,[3] we will replace it by a more specific rate $r = O(n^{\delta})$, where $\delta > 0$. We now present the implication of his Theorem 3 reflecting the assumption $r = O\left(n^{\delta}\right)$. We need to introduce a few symbols for this purpose: $G(\theta, z) \equiv \sum_{l=1}^{L} d_l\left[\partial \ln f_l(\theta, x)/\partial\theta\right]$, $H(\theta, z) \equiv \sum_{l=1}^{L} d_l\left[\partial^2 \ln f_l(\theta, x)\big/\partial\theta\partial\theta'\right]$, $P_l(x) \equiv P(l|\theta_0, x)$, $e_{r,l}(x) \equiv f_{r,l}(\theta_0, x) - P(l|\theta_0, x)$, and $e_l(v, x) \equiv h_l(v, x, \theta_0) - P(l|\theta_0, x)$.[4] All proofs are collected in Section 3.

Proposition 1. *Suppose that Lee's (1995) Assumptions 1 and 2 are satisfied. Further suppose that $r = O\left(n^{\delta}\right)$ for some $\delta > 0$. We then have*

$$\sqrt{n}\left(\hat{\theta}_l - \theta_0\right) = \Omega\left\{S_n + L_n + n^{1/2}r^{-1}\bar{\mu} + B_{1,n} + B_{2,n} + O_p\left(n^{-\delta}\right)\right\} \quad \text{for } 1/2 < \delta \le 1,$$

$$\sqrt{n}\left(\hat{\theta}_l - \theta_0\right) = \Omega\left\{S_n + L_n + n^{1/2}r^{-1}\bar{\mu} + O_p\left(n^{-1/2}\right)\right\} \quad \text{for } \delta = 1/2,$$

$$\sqrt{n}\left(\hat{\theta}_l - \theta_0\right) = \Omega\left\{S_n + B_{1,n} + B_{2,n} + o_p\left(n^{-1/2}\right)\right\} \quad \text{for } \delta > 1,$$

$$r\left(\hat{\theta}_l - \theta_0\right) = \Omega\bar{\mu} + o_p(1) \quad \text{for } 0 < \delta < 1/2,$$

where Ω denotes the inverse of the information matrix, and

$$S_n \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^{n} G(z_i),$$

$$L_n \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{1}{P_l(x_i)}\left\{\frac{\partial e_{r,l}(x_i)}{\partial\theta} - \frac{\partial \ln P_l(x_i)}{\partial\theta}e_{r,l}(x_i)\right\},$$

$$\bar{\mu} \equiv \sum_{l=1}^{L} E\left[\frac{1}{P_l(x)}\left\{-\operatorname{Cov}\left(h_l(v,x), \frac{\partial h_l(v,x)}{\partial\theta}\middle| x\right) + \frac{\partial \ln P_l(x)}{\partial\theta}\operatorname{Var}\left(h_l(v,x)|\, x\right)\right\}\right],$$

$$B_{1,n} \equiv \left[n^{-1}\sum_{i=1}^{n} H(z_i) - E\left[H(z_i)\right]\right]\Omega S_n,$$

$$B_{2,n} \equiv \frac{n^{-1/2}}{2}\begin{bmatrix} S_n'\Omega E\left[\partial H(z_i)/\partial\theta_1\right]\Omega S_n \\ \vdots \\ S_n'\Omega E\left[\partial H(z_i)/\partial\theta_K\right]\Omega S_n \end{bmatrix}.$$

From a pragmatic interpretation point of view, the case $\delta > 1$ is the least interesting in terms of understanding the bias due to the simulation. The terms S_n, $B_{1,n}$, and $B_{2,n}$ do not depend on the simulation draws $v_j^{(i)}$, while the term $\bar{\mu}$ reflects the bias due to the simulation. Because the $\bar{\mu}$ is not present when $\delta > 1$, it implies that the impact of simulation is ignored along with the $o_p\left(n^{-1/2}\right)$ remainder term, and the only higher order terms are the generic[5] higher order terms of the (computationally infeasible) MLE $B_{1,n} + B_{2,n}$. The case $0 < \delta < 1/2$ is the other extreme case, because we end up attributing the whole statistical properties of $\hat{\theta}_l$ to the simulation bias, modulo the $o_p(1)$ remainder term. If $1/2 < \delta \le 1$, the bias $n^{1/2}r^{-1}\bar{\mu}$ due to the simulation as well as the generic higher order terms of the MLE $B_{1,n} + B_{2,n}$ converge to zero in probability, while the bias $n^{1/2}r^{-1}\bar{\mu}$ due to the simulation becomes the first order asymptotic bias if $\delta = 1/2$. Our main result below discusses properties of the SS jackknife estimator for all these cases.

We now present the SS jackknife estimator. For this purpose, suppose that $r = 2m$, and let $\bar{\theta}_{\mathcal{S}_1}$ and $\bar{\theta}_{\mathcal{S}_2}$ denote maximizers of

$$\mathcal{L}_1(\theta) \equiv \sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\ln f_{m,l,(1)}(\theta, x_i) \equiv \sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\ln\left(\frac{1}{m}\sum_{j=1}^{m} h_l\left(v_j^{(i)}, x_i, \theta\right)\right),$$

$$\mathcal{L}_2(\theta) \equiv \sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\ln f_{m,l,(2)}(\theta, x_i) \equiv \sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\ln\left(\frac{1}{m}\sum_{j=m+1}^{2m} h_l\left(v_j^{(i)}, x_i, \theta\right)\right).$$

Applying Dhaene and Jochmans (2015) idea to the current situation, we define the SS jackknife estimator as

$$\tilde{\theta}_{1/2} \equiv 2\hat{\theta}_l - \frac{1}{2}\left(\bar{\theta}_{\mathcal{S}_1} + \bar{\theta}_{\mathcal{S}_2}\right).$$

Dhaene and Jochmans (2015) discuss the intuition and theory underlying the SS jackknife estimator for panel models, which we adapt to provide the intuition underlying our estimator here. Based on Proposition 1, we consider the intuitive approximation

$$\sqrt{n}\left(\hat{\theta}_l - \theta_0\right) \approx \Omega\left\{S_n + L_n + n^{1/2}r^{-1}\bar{\mu}\right\} \quad \text{for } 1/2 < \delta \le 1,$$

$$\sqrt{n}\left(\hat{\theta}_l - \theta_0\right) \approx \Omega\left\{S_n + L_n + n^{1/2}r^{-1}\bar{\mu}\right\} \quad \text{for } \delta = 1/2,$$

$$\sqrt{n}\left(\hat{\theta}_l - \theta_0\right) \approx \Omega\left\{S_n\right\} \quad \text{for } \delta > 1,$$

$$r\left(\hat{\theta}_l - \theta_0\right) \approx \Omega\bar{\mu} \quad \text{for } 0 < \delta < 1/2,$$

where generic higher order terms of the (computationally infeasible) MLE $B_{1,n} + B_{2,n}$ are ignored along with the remainder terms. This can be justified by recognizing that these terms are all smaller than the remaining terms in the order of magnitudes. Recognizing that the expectation of $S_n + L_n$ is zero, we can conclude that $E\left[\sqrt{n}\left(\hat{\theta}_l - \theta_0\right)\right] \approx \Omega n^{1/2}r^{-1}\bar{\mu}$ assuming that we can exchange expectations and approximations. In other words, we have

$$E\left[\hat{\theta}_l\right] \approx \theta_0 + \Omega r^{-1}\bar{\mu},$$

[3] The SMLE $\hat{\theta}_l$ satisfies the first-order condition $\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\left[\partial \ln f_{r,l}\left(\hat{\theta}_l, x_i\right)\big/\partial\theta\right] = 0$. Because $\sum_{l=1}^{L} d_{li}\left[\partial \ln f_{r,l}(\theta_0, x_i)\big/\partial\theta\right]$ does not necessarily have zero expectation, the SMLE is in general inconsistent unless $r \to \infty$ as a function of the sample size n, which explains why Lee (1995) imposed Assumption 3.

[4] Note that the last two symbols reflect Lee's (1995) convention of suppressing θ_0 when there is no confusion.

[5] See Lee (1995, p. 447) for example.

By the same token, we have

$$E\left[\bar{\theta}_{\mathcal{S}_1}\right] \approx \theta_0 + \Omega m^{-1}\bar{\mu},$$

so we expect

$$E\left[2\hat{\theta}_l - \bar{\theta}_{\mathcal{S}_1}\right] \approx \theta_0 + 2\Omega r^{-1}\bar{\mu} - \Omega m^{-1}\bar{\mu} = \theta_0.$$

Using that $2\hat{\theta}_l - \bar{\theta}_{\mathcal{S}_1}$ is less biased than $\hat{\theta}_l$, we can naturally think of $\tilde{\theta}_{1/2}$ as a more symmetric estimator using $\bar{\theta}_{\mathcal{S}_2}$ as well.

Below, we present the formal asymptotic expansion of $\sqrt{n}\left(\tilde{\theta}_{1/2} - \theta_0\right)$:

Proposition 2. *Suppose that Lee's (1995) Assumptions 1 and 2 are satisfied. Further suppose that $r = O\left(n^{\delta}\right)$ for some $\delta > 0$. We then have*

$$\sqrt{n}\left(\tilde{\theta}_{1/2} - \theta_0\right) = \Omega\left\{S_n + L_n + B_{1,n} + B_{2,n} + O_p\left(n^{-\delta}\right)\right\} \quad \text{for } 1/2 < \delta \leq 1,$$

$$\sqrt{n}\left(\tilde{\theta}_{1/2} - \theta_0\right) = \Omega\left\{S_n + L_n + O_p\left(n^{-1/2}\right)\right\} \quad \text{for } \delta = 1/2,$$

$$\sqrt{n}\left(\tilde{\theta}_{1/2} - \theta_0\right) = \Omega\left\{S_n + B_{1,n} + B_{2,n} + o_p\left(n^{-1/2}\right)\right\} \quad \text{for } \delta > 1,$$

$$r\left(\tilde{\theta}_{1/2} - \theta_0\right) = o_p(1) \quad \text{for } 0 < \delta < 1/2.$$

Comparing Propositions 1 and 2, we can see that the expansion of $\tilde{\theta}_{1/2}$ does not include the bias $n^{1/2}r^{-1}\bar{\mu}$ due to the simulation. In other words, the SS jackknife estimator removes such bias. As long as r increases at a faster rate than $n^{1/4}$, $\tilde{\theta}_{1/2}$ is efficient in the sense that the simulation-induced bias is small enough not to show up in the asymptotic distribution. The same properties hold for Lee's (1995) bias corrected estimator. Because the SS jackknife estimator does not require separate analytic characterization of the simulation bias, which is required for implementation of Lee's (1995) procedure, it may have certain practical advantage. Implementation of the SS jackknife requires computation of $\bar{\theta}_{\mathcal{S}_1}$ and $\bar{\theta}_{\mathcal{S}_2}$, so if the computational burden is not serious, the SS jackknife estimator can be an attractive alternative to Lee's (1995) procedure.

3. Proofs

3.1. Proof of Proposition 1

Lee's (1995) Theorem 3 gives us

$$\sqrt{n}\left(\hat{\theta}_l - \theta_0\right) = \Omega\left\{S_n + L_n + Q_n + B_{1,n} + B_{2,n} + O_p\left(\max\left[n^{-1/2}r^{-1/2}, n^{-1}, r^{-1}, n^{1/2}r^{-2}\right]\right)\right\}$$

where

$$Q_n \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{1}{P_l^2(x_i)}\left\{-\frac{\partial e_{r,l}(x_i)}{\partial\theta}e_{r,l}(x_i) + \frac{\partial \ln P_l(x_i)}{\partial\theta}e_{r,l}^2(x_i)\right\},$$

and $B_{1,n} = O_p\left(n^{-1/2}\right)$, and $B_{2,n} = O_p\left(n^{-1/2}\right)$. If r is chosen such that $r = O\left(n^{\delta}\right)$ with $1/2 < \delta < 1$, we have

$$O_p\left(\max\left[n^{-1/2}r^{-1/2}, n^{-1}, r^{-1}, n^{1/2}r^{-2}\right]\right) = O_p\left(\max\left[n^{-(\delta+1)/2}, n^{-1}, n^{-\delta}, n^{-(2\delta-1/2)}\right]\right) = O_p\left(n^{-\delta}\right).$$

Using $Q_n - E\left[Q_n\right] = O_p\left(r^{-1}\right) = O_p\left(n^{-\delta}\right)$, which is implied by Lee's (1995) Theorem 1, we obtain

$$\sqrt{n}\left(\hat{\theta}_l - \theta_0\right) = \Omega\left\{S_n + L_n + E\left[Q_n\right] + B_{1,n} + B_{2,n} + O_p\left(n^{-\delta}\right)\right\},$$

Note that $E\left[Q_n\right] = n^{1/2}r^{-1}\bar{\mu}$, where $\bar{\mu}$ does not depend on r, by Lee's (1995) Equation (3.6). Therefore, $E\left[Q_n\right] = O\left(n^{1/2}r^{-1}\right) = O\left(n^{(1-2\delta)/2}\right)$, which is larger than $O_p\left(n^{-1/2}\right)$ or $O_p\left(n^{-\delta}\right)$.

If $r = O\left(n^{1/2}\right)$, we have

$$O_p\left(\max\left[n^{-1/2}r^{-1/2}, n^{-1}, r^{-1}, n^{1/2}r^{-2}\right]\right) = O_p\left(\max\left[n^{-1/2}n^{-1/4}, n^{-1}, n^{-1/2}, n^{1/2}n^{-1}\right]\right) = O_p\left(n^{-1/2}\right),$$

and therefore, Lee's (1995) Theorem 3 results in

$$\sqrt{n}\left(\hat{\theta}_l - \theta_0\right) = \Omega\left\{S_n + L_n + Q_n + O_p\left(n^{-1/2}\right)\right\},$$

where we used $B_{1,n} = O_p\left(n^{-1/2}\right)$, and $B_{2,n} = O_p\left(n^{-1/2}\right)$. Note that the assumption $r = O\left(n^{1/2}\right)$ leads to the loss of our ability to tell $B_{1,n}$ and $B_{2,n}$ apart from the remainder term in Theorem 3, because they are of the same order of magnitude when $r = O\left(n^{1/2}\right)$. We have $Q_n = E\left[Q_n\right] + O_p\left(r^{-1}\right) = E\left[Q_n\right] + O_p\left(n^{-1/2}\right)$, from which we further obtain

$$\sqrt{n}\left(\hat{\theta}_l - \theta_0\right) = \Omega\left\{S_n + L_n + n^{1/2}r^{-1}\bar{\mu} + O_p\left(n^{-1/2}\right)\right\}.$$

If r is chosen such that $r = O\left(n^{\delta}\right)$ with $\delta \geq 1$, we have

$$\begin{aligned}O_p\left(\max\left[n^{-1/2}r^{-1/2}, n^{-1}, r^{-1}, n^{1/2}r^{-2}\right]\right) &= O_p\left(n^{-1}\right),\\ Q_n - E\left[Q_n\right] &= O_p\left(r^{-1}\right)\\ &= O_p\left(n^{-\delta}\right) = O_p\left(n^{-1}\right),\\ E\left[Q_n\right] &= O\left(n^{1/2}r^{-1}\right)\\ &= o\left(n^{-1/2}\right),\end{aligned}$$

so Lee's (1995) Theorem 3 results in

$$\begin{aligned}\sqrt{n}\left(\hat{\theta}_l - \theta_0\right) &= \Omega\left\{S_n + L_n + Q_n + B_{1,n} + B_{2,n} + O_p\left(n^{-1}\right)\right\}\\ &= \Omega\left\{S_n + L_n + B_{1,n} + B_{2,n} + o_p\left(n^{-1/2}\right)\right\}.\end{aligned}$$

Finally, because $L_n = O_p\left(r^{-1/2}\right)$ by Lee's (1995) Theorem 1, we have $L_n = o\left(n^{-1/2}\right)$, from which we get the third result.

If $0 < \delta < 1/2$, the result follows from his Corollary 1 (iii).

3.2. Proof of Proposition 2

Suppose that $1/2 < \delta \leq 1$. From Proposition 1, we get

$$\sqrt{n}\left(\bar{\theta}_{\mathcal{S}_1} - \theta_0\right) = \Omega\left\{S_n + L_{n,(1)} + n^{1/2}m^{-1}\bar{\mu} + B_{1,n} + B_{2,n} + O_p\left(n^{-\delta}\right)\right\},$$

where we note that the S_n, $B_{1,n}$, and $B_{2,n}$ in Proposition 1 do not depend on the simulation draws $v_j^{(i)}$, and the counterpart of $L_{n,(1)}$ is

$$L_{n,(1)} \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{1}{P_l(x_i)} \times \left\{\frac{\partial e_{m,l,(1)}(x_i)}{\partial\theta} - \frac{\partial \ln P_l(x_i)}{\partial\theta}e_{m,l,(1)}(x_i)\right\},$$

where

$$e_{m,l,(1)} \equiv f_{m,l,(1)}(\theta_0, x_i) - P(l|\,\theta_0, x_i) \equiv \frac{1}{m}\sum_{j=1}^{m} h_l\left(v_j^{(i)}, x_i, \theta\right) - P(l|\,\theta_0, x_i).$$

With similar expansion for $\sqrt{n}\left(\bar{\theta}_{\mathcal{S}_1} - \theta_0\right)$, we obtain

$$\sqrt{n}\left(\tilde{\theta}_{1/2} - \theta_0\right) = \sqrt{n}\left(2\left(\hat{\theta}_l - \theta_0\right) - \frac{1}{2}\left(\left(\bar{\theta}_{\mathcal{S}_1} - \theta_0\right) + \left(\bar{\theta}_{\mathcal{S}_2} - \theta_0\right)\right)\right)$$

$$
\begin{aligned}
&= \Omega \left\{2S_n + 2L_n + 2n^{1/2}r^{-1}\bar{\mu} + 2B_{1,n} + 2B_{2,n} + O_p\left(n^{-\delta}\right)\right\} \\
&\quad - \frac{1}{2}\Omega\left\{S_n + L_{n,(1)} + n^{1/2}m^{-1}\bar{\mu} + B_{1,n} + B_{2,n} + O_p\left(n^{-\delta}\right)\right\} \\
&\quad - \frac{1}{2}\Omega\left\{S_n + L_{n,(2)} + n^{1/2}m^{-1}\bar{\mu} + B_{1,n} + B_{2,n} + O_p\left(n^{-\delta}\right)\right\} \\
&= \Omega\left\{S_n + \left(2L_n - \frac{L_{n,(1)} + L_{n,(2)}}{2}\right) + B_{1,n} + B_{2,n} + O_p\left(n^{-\delta}\right)\right\},
\end{aligned}
$$

where we note that

$$
2n^{1/2}r^{-1}\bar{\mu} - \frac{n^{1/2}m^{-1}\bar{\mu} + n^{1/2}m^{-1}\bar{\mu}}{2} = n^{1/2}\left(\frac{r}{2}\right)^{-1}\bar{\mu} - n^{1/2}m^{-1}\bar{\mu} = 0 \tag{4}
$$

because $r = 2m$. Finally, we have $L_n = \left(L_{n,(1)} + L_{n,(2)}\right)/2$ from the definition, so we obtain

$$
\sqrt{n}\left(\tilde{\theta}_{1/2} - \theta_0\right) = \Omega\left\{S_n + L_n + B_{1,n} + B_{2,n} + O_p\left(n^{-\delta}\right)\right\}.
$$

Suppose now that $\delta = 1/2$. From Proposition 1, we get

$$
\sqrt{n}\left(\bar{\theta}_{\mathcal{S}_1} - \theta_0\right) = \Omega\left\{S_n + L_{n,(1)} + n^{1/2}m^{-1}\bar{\mu} + O_p\left(n^{-1/2}\right)\right\}.
$$

With similar expansion for $\sqrt{n}\left(\bar{\theta}_{\mathcal{S}_1} - \theta_0\right)$, we obtain

$$
\begin{aligned}
\sqrt{n}\left(\tilde{\theta}_{1/2} - \theta_0\right) &= \sqrt{n}\left(2\left(\hat{\theta}_I - \theta_0\right) - \frac{1}{2}\left(\left(\bar{\theta}_{\mathcal{S}_1} - \theta_0\right) + \left(\bar{\theta}_{\mathcal{S}_2} - \theta_0\right)\right)\right) \\
&= \Omega\left\{2S_n + 2L_n + 2n^{1/2}r^{-1}\bar{\mu} + O_p\left(n^{-1/2}\right)\right\} \\
&\quad - \frac{1}{2}\Omega\left\{S_n + L_{n,(1)} + n^{1/2}m^{-1}\bar{\mu} + O_p\left(n^{-1/2}\right)\right\} \\
&\quad - \frac{1}{2}\Omega\left\{S_n + L_{n,(2)} + n^{1/2}m^{-1}\bar{\mu} + O_p\left(n^{-1/2}\right)\right\} \\
&= \Omega\left\{S_n + L_n + O_p\left(n^{-1/2}\right)\right\},
\end{aligned}
$$

where we use (4) and $L_n = \left(L_{n,(1)} + L_{n,(2)}\right)/2$.

For the $\delta > 1$ case, the result follows from the fact that the three estimators $\hat{\theta}_I$, $\bar{\theta}_{\mathcal{S}_1}$, and $\bar{\theta}_{\mathcal{S}_2}$ all have the identical expansion presented in Proposition 1.

Finally, if $0 < \delta < 1/2$, we have by Proposition 1

$$
\begin{aligned}
r\left(\tilde{\theta}_{1/2} - \theta_0\right) &= r\left(2\left(\hat{\theta}_I - \theta_0\right) - \frac{1}{2}\left(\left(\bar{\theta}_{\mathcal{S}_1} - \theta_0\right) + \left(\bar{\theta}_{\mathcal{S}_2} - \theta_0\right)\right)\right) \\
&= 2r\left(\hat{\theta}_I - \theta_0\right) - \frac{2m}{2}\left(\bar{\theta}_{\mathcal{S}_1} - \theta_0\right) - \frac{2m}{2}\left(\bar{\theta}_{\mathcal{S}_2} - \theta_0\right) \\
&= 2\left(\Omega\bar{\mu} + o_p(1)\right) - \left(\Omega\bar{\mu} + o_p(1)\right) - \left(\Omega\bar{\mu} + o_p(1)\right) \\
&= o_p(1).
\end{aligned}
$$

Data availability

No data was used for the research described in the article.

References

Dhaene, G., Jochmans, K., 2015. Split-panel jackknife estimation of fixed-effect models. Rev. Econom. Stud. 82 (3), 991–1030.

Hajivassiliou, V.A., Ruud, P.A., 1994. Classical estimation methods for LDV models using simulation. Handb. Econom. 4, 2383–2441.

Lee, L.F., 1992. On efficiency of methods of simulated moments and maximum simulated likelihood estimation of discrete response models. Econom. Theory 8 (4), 518–552.

Lee, L.F., 1995. Asymptotic bias in simulated maximum likelihood estimation of discrete choice models. Econom. Theory 11 (3), 437–483.

Lerman, S., Manski, C., 1981. On the use of simulated frequencies to approximate choice probabilities. Struct. Anal. Discrete Data Econom. Appl. 10, 305–319.

McFadden, D., 1989. A method of simulated moments for estimation of discrete response models without numerical integration. Econometrica 995–1026.

Pakes, A., Pollard, D., 1989. Simulation and the asymptotics of optimization estimators. Econometrica 1027–1057.