Title
Essays in Econometrics

Author
Liu, Xueyuan

Publication Date
2022

UNIVERSITY OF CALIFORNIA

Los Angeles

Essays in Econometrics

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Economics

by

Xueyuan Liu

2022

ABSTRACT OF THE DISSERTATION

Essays in Econometrics

by

Xueyuan Liu
Doctor of Philosophy in Economics
University of California, Los Angeles, 2022
Professor Jinyong Hahn, Co-Chair
Professor Shuyang Sheng, Co-Chair

The dissertation consists of three chapters on different econometric topics.

The first chapter studies jackknife bias reduction for simulated maximum likelihood estimator of discrete choice models. We propose to reduce asymptotic biases of simulated maximum likelihood estimators (SMLE) by using a jackknife method similar to Dhaene and Jochmans (2015), which was originally proposed to reduce bias in nonlinear panel models. Lee (1995) investigates the asymptotic bias of the SMLE, and derives the analytical formula of higher order bias due to simulation. However, implementation of Lee (1995)'s method requires analytical characterization of the higher order bias, which may not be convenient for practice. Because the jackknife method does not require an explicit characterization of the bias, it may be a practically attractive alternative to Lee (1995)'s estimator.

The second chapter studies estimation of average treatment effects for massively unbalanced binary outcomes. The maximum likelihood estimator (MLE) of the average treatment effects (ATE) in the logit model for binary outcomes may have a significant second order bias if the event has a low probability. The analysis of rare events is relevant for economics because

some of the big data sets are collected from online sources where the number of events (such as " clicks" and " purchases") is much smaller than the number of nonevents. The literature about rare events (King and Zeng, 2001; Chen and Giles, 2012; Rilstone, 1996; Wang, 2020) does not shed light on the finite sample behavior of logit MLE and ATE if events are rare. In this chapter, we also derive the second order bias of the logit ATE estimator and propose bias-corrected estimators of the ATE. We also propose a variation on the logit model with parameters that are elasticities. Finally, we propose a computational trick that avoids numerical instability in the case of estimation for rare events.

The third chapter studies a Vuong test (Vuong, 1989) for panel data models with fixed effects. This chapter generalizes the Vuong test to nonlinear panel models where the dimension of incidental parameters grows with the sample size. The incidental parameters (Neyman and Scott, 1948) that affect the unbiasedness of the parameters of interest are also important for panel data models as they capture unobserved heterogeneity. The discrepancy in incidental parameters plays an important role in model selection; for example, as noted by MacKinnon et al. (2020), there is a vast literature on the cluster-robust inference that assumes the structure of the clusters is correctly specified, which is often violated. In the presence of incidental parameters, we cannot easily apply the classical Vuong test to select a panel data model. This chapter proposes a new model selection test for panel data models by extending the classical Vuong test, which selects from two parametric likelihood models based on their Kullback–Leibler information criterion (KLIC). This chapter proposes three different test statistics for researchers who need to deal with all possible relationships between candidate models: overlapping models, nested models, and strictly nonnested models. These three model relationships are classified according to the structure of low-dimensional parameter of interest and high-dimensional incidental parameters. We allow for disagreements about incidental parameters and obtain specification tests based on a modified likelihood function.

The dissertation of Xueyuan Liu is approved.

Elisabeth Honka

Zhipeng Liao

Rosa Liliana Matzkin

Shuyang Sheng, Committee Co-Chair

Jinyong Hahn, Committee Co-Chair

University of California, Los Angeles

2022

To my family, for the love, care and support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to express my sincere gratitude for the unending support and guidance that I have received from my advisors, Jinyong Hahn, Rosa Matzkin, Zhipeng Liao and Shuyang Sheng. I am grateful for the opportunity to work with them. I would also like to thank Andres Santos, Denis Chetverikov, and all participants of the proseminars, workshops in Econometrics.

Finally, I would like to thank my husband, Xinyu Huang, for his unconditional love, help, and support.

VITA

2017	M.A. in Economics, National School of Development/CCER, Peking University.
2019	M.A. in Economics, Department of Economics, UCLA.
2019–2022	Teaching Assistant, Department of Economics, UCLA.

CHAPTER 1

Jackknife Bias Reduction for Simulated Maximum Likelihood Estimator of Discrete Choice Models

1.1 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. Lee (1995) then goes on and constructs bias-adjusted SMLE's by estimating such a bias using the analytic formula.

Implementation of Lee (1995)'s 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 jackknife method due to Dhaene and Jochmans (2015), which was originally proposed to reduce bias in nonlinear panel models. Dhaene and Jochmans (2015)'s intuition may be attractive for practice because it only requires computation of the SMLE a few times.

Our main results are presented in Section 1.2. All the proofs are collected in Section 1.3.

1.2 Main Results

We start with a review of Lee (1995). Consider a standard model of discrete responses. Let $C = \{1, \ldots, L\}$ be a set of mutually exclusive and exhaustive alternatives. For each alternative $l \in C$, let $P\left(l \middle| \theta, x\right)$ denote the probability that such alternative is chosen, where x denotes the vector consisting of all distinct explanatory variable, 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

$$\mathcal{L}_c(\theta) \equiv \sum_{i=1}^{n} \sum_{l=1}^{L} d_{li} \ln P\left(l \middle| \theta, x_i\right). \tag{1.1}$$

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

$$\frac{1}{n} \sum_{i=1}^{n} \sum_{l=1}^{L} d_{li} \frac{\partial \ln P\left(l \middle| \theta, x_i\right)}{\partial \theta} = 0. \tag{1.2}$$

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

$$f_{r,l}(\theta, x_i) \equiv \frac{1}{r} \sum_{j=1}^{r} h_l\left(v_j^{(i)}, x_i, \theta\right), \tag{1.3}$$

where $v_j^{(i)}$, $j = 1, \ldots, r$ denote r Monte Carlo draw for observation i from $\gamma(v)$. Because $E\left[f_{r,l}(\theta, x_i) \middle| x_i\right] = P\left(l \middle| \theta, x_i\right)$, the $f_{r,l}(\theta, x_i)$ is a conditionally unbiased simulator. The SMLE $\hat{\theta}_I$ 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 $n \to \infty$,[1] we will replace it by a more specific rate $r = O\left(n^{\delta}\right)$, 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: $G(\theta, z) \equiv \sum_{l=1}^{L} d_l \left[\partial \ln f_l(\theta, x) / \partial\theta\right]$, $H(\theta, z) \equiv \sum_{l=1}^{L} d_l \left[\partial^2 \ln f_l(\theta, x) / \partial\theta\partial\theta'\right]$, $e_{r,l}(x) \equiv f_{r,l}(\theta_0, x_i) - P(l|\, \theta_0, x_i)$, and $e_l(v, x) \equiv h_l(v, x, \theta_0) - P(l|\, \theta_0, x_i)$.[2] All proofs are collected in Section 1.3.

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[3] 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 $\hat{\theta}_I$ to the simulation bias, modulo the $o_p(1)$ remainder term. If $1/2 < \delta \leq 1$, 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 split sample jackknife estimator for all these cases.

Proposition 1 *Suppose that Lee (1995)'s Assumptions 1 and 2 are satisfied. Further suppose*

[1]The SMLE $\hat{\theta}_I$ satisfies the first-order condition $\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L}\left(d_{li} - f_{r,l}\left(\hat{\theta}_I, x_i\right)\right)\left[\partial \ln f_{r,l}\left(\hat{\theta}_I, x_i\right) \big/ \partial\theta\right] = 0$. Because $\sum_{l=1}^{L}\left(d_{li} - f_{r,l}(\theta_0, x_i)\right)\left[\partial \ln f_{r,l}(\theta_0, x_i) / \partial\theta\right]$ 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)'s imposed Assumption 3.

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

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

that $r = O\left(n^{\delta}\right)$ *for some* $\delta > 0$*. We then have*

$$\sqrt{n}\left(\hat{\theta}_I - \theta_0\right) = \Omega\left\{S_n + L_n + n^{1/2} r^{-1}\bar{\mu} + B_{1,n} + B_{2,n} + O_p\left(n^{-\delta}\right)\right\} \quad \textit{for } 1/2 < \delta \leq 1,$$
$$\sqrt{n}\left(\hat{\theta}_I - \theta_0\right) = \Omega\left\{S_n + L_n + n^{1/2} r^{-1}\bar{\mu} + O_p\left(n^{-1/2}\right)\right\} \quad \textit{for } \delta = 1/2,$$
$$\sqrt{n}\left(\hat{\theta}_I - \theta_0\right) = \Omega\left\{S_n + B_{1,n} + B_{2,n} + o_p\left(n^{-1/2}\right)\right\} \quad \textit{for } \delta > 1,$$
$$r\left(\hat{\theta}_I - \theta_0\right) = \Omega\bar{\mu} + o_p(1) \quad \textit{for } 0 < \delta < 1/2,$$

where Ω *denotes the inverse of the information matrix, and*

$$S_n \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} G(z_i),$$
$$L_n \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{l=1}^{L} d_{li} \frac{1}{P_l(x_i)} \left\{\frac{\partial e_{r,l}(x_i)}{\partial \theta} - \frac{\partial P_l(x_i)}{\partial \theta} e_{r,l}(x_i)\right\},$$
$$\bar{\mu} \equiv \sum_{l=1}^{L} \left[\frac{1}{P_l(x)} \left\{- \mathrm{Cov}\left(h_l(v,x), \left.\frac{\partial h_l(v,x)}{\partial \theta}\right| x\right) + \frac{\partial P_l(x_i)}{\partial \theta} \mathrm{Var}\left(h_l(v,x)| x\right)_i\right\}\right],$$
$$B_{1,n} \equiv \left[n^{-1} \sum_{i=1}^{n} H(z_i) - E\left[H(z_i)\right]\right] \Omega S_n,$$
$$B_{1,n} \equiv \frac{n^{-1}}{2} \begin{bmatrix} S_n' \Omega E\left[\partial H(z_i)/\partial \theta_1\right] \Omega S_n \\ \vdots \\ S_n' \Omega E\left[\partial H(z_i)/\partial \theta_K\right] \Omega S_n \end{bmatrix}.$$

We now present the split sample 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)'s idea to the current situation, we define the split sample jackknife estimator as

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

Dhaene and Jochmans (2015) discuss the intuition and theory underlying the split sample 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

$$\begin{aligned}
\sqrt{n}\left(\hat{\theta}_I-\theta_0\right) &\approx \Omega\left\{S_n+L_n+n^{1/2}r^{-1}\bar{\mu}\right\} \quad \text{for } 1/2<\delta\leq 1, \\
\sqrt{n}\left(\hat{\theta}_I-\theta_0\right) &\approx \Omega\left\{S_n+L_n+n^{1/2}r^{-1}\bar{\mu}\right\} \quad \text{for } \delta=1/2, \\
\sqrt{n}\left(\hat{\theta}_I-\theta_0\right) &\approx \Omega\left\{S_n\right\} \quad \text{for } \delta>1, \\
r\left(\hat{\theta}_I-\theta_0\right) &\approx \Omega\bar{\mu} \quad \text{for } 0<\delta<1/2,
\end{aligned}$$

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 $E\left[\sqrt{n}\left(\hat{\theta}_I-\theta_0\right)\right]\approx\Omega n^{1/2}r^{-1}\bar{\mu}$ assuming that we can exchange expectations and approximations. In other words, we have

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

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

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

Using that $2\hat{\theta}_I-\bar{\theta}_{\mathcal{S}_1}$ is less biased than $\hat{\theta}_I$, 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 (1995)'s Assumptions 1 and 2 are satisfied. Further suppose* *that* $r = O\left(n^{\delta}\right)$ *for some* $\delta > 0$.. *We then have*

$$\sqrt{n}\left(\tilde{\theta}_{1/2} - \theta_0\right) = \Omega\left\{S_n + L_n + B_{1,n} + B_{2,n} + O_p\left(n^{-\delta}\right)\right\} \quad \textit{for } 1/2 < \delta \leq 1,$$
$$\sqrt{n}\left(\tilde{\theta}_{1/2} - \theta_0\right) = \Omega\left\{S_n + L_n + O_p\left(n^{-1/2}\right)\right\} \quad \textit{for } \delta = 1/2,$$
$$\sqrt{n}\left(\tilde{\theta}_{1/2} - \theta_0\right) = \Omega\left\{S_n + B_{1,n} + B_{2,n} + o_p\left(n^{-1/2}\right)\right\} \quad \textit{for } \delta > 1,$$
$$r\left(\tilde{\theta}_{1/2} - \theta_0\right) = o_p(1) \quad \textit{for } 0 < \delta < 1/2.$$

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 split sample jackknife estimator removes such bias. Because the split sample jackknife estimator does not require separate analytic characterization of the simulation bias, which is required for implementation of Lee (1995)'s procedure, it may have certain practical advantage. Implementation of the split sample jackknife requires computation of $\bar{\theta}_{\mathcal{S}_1}$ and $\bar{\theta}_{\mathcal{S}_2}$, so if the computational burden is not serious, the split sample jackknife estimator can be an attractive alternative to Lee (1995)'s procedure.

1.3 Proofs

1.3.1 Proof of Proposition 1

Lee (1995)'s Theorem 3 gives us

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

where

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

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

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

Note that $E[Q_n] = n^{1/2}r^{-1}\bar{\mu}$, where $\bar{\mu}$ does not depend on r, by Lee (1995)'s Equation (3.6). Therefore, $E[Q_n] = O\left(n^{1/2}r^{-1}\right) = O\left(n^{(1-2\delta)/2}\right)$, 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 (1995)'s Theorem 3 results in

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

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 $Q_n = E[Q_n] + O_p(r^{-1}) = E[Q_n] + O_p\left(n^{-1/2}\right)$, from which we further obtain

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

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

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

so Lee (1995)'s Theorem 3 results in

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

Finally, because $L_n = O_p\left(r^{-1/2}\right)$ by Lee (1995)'s 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).

1.3.2 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

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

where

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

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

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

where we note that

$$2n^{1/2}r^{-1}\bar{\mu} - \frac{n^{1/2}m^{-1}\bar{\mu} + n^{1/2}m^{-1}\bar{\mu}}{2} = n^{1/2}\left(\frac{r}{2}\right)^{-1}\bar{\mu} - n^{1/2}m^{-1}\bar{\mu} = 0 \tag{1.4}$$

because $r = 2m$. Finally, using Lemma 1 below, 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

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

where we use (1.4) and Lemma 1.

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

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

Lemma 1 $2L_n - \frac{L_{n,_{(1)}} + L_{n,_{(2)}}}{2} = L_n$.

Proof. We note that

$$\begin{aligned}
&2e_{r,l}\left(x_i\right)-\frac{e_{m,l,(1)}\left(x_i\right)+e_{m,l,(2)}\left(x_i\right)}{2}\\
&=2\left(\frac{1}{r}\sum_{j=1}^{r}h_l\left(v_j^{(i)},x_i,\theta_0\right)-P\left(l|\,\theta_0,x_i\right)\right)\\
&-\frac{\frac{1}{m}\sum_{j=1}^{m}h_l\left(v_j^{(i)},x_i,\theta_0\right)-P\left(l|\,\theta_0,x_i\right)+\frac{1}{m}\sum_{j=m+1}^{r}h_l\left(v_j^{(i)},x_i,\theta_0\right)-P\left(l|\,\theta_0,x_i\right)}{2}\\
&=\frac{2}{r}\sum_{j=1}^{r}h_l\left(v_j^{(i)},x_i,\theta_0\right)-\frac{1}{2m}\sum_{j=1}^{r}h_l\left(v_j^{(i)},x_i,\theta_0\right)-P\left(l|\,\theta_0,x_i\right)\\
&=\frac{1}{r}\sum_{j=1}^{r}h_l\left(v_j^{(i)},x_i,\theta_0\right)-P\left(l|\,\theta_0,x_i\right)=e_{r,l}\left(x_i\right),
\end{aligned}$$

which implies that

$$\begin{aligned}
&2L_n-\frac{L_{n,(1)}+L_{n,(2)}}{2}\\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L}d_{li}\frac{1}{P_l\left(x_i\right)}\left(\begin{array}{c}\frac{\partial\left(2e_{r,l}(x_i)-\frac{e_{m,l,(1)}(x_i)+e_{m,l,(2)}(x_i)}{2}\right)}{\partial\theta}\\ -\frac{\partial P_l(x_i)}{\partial\theta}\left(2e_{r,l}\left(x_i\right)-\frac{e_{m,l,(1)}(x_i)+e_{m,l,(2)}(x_i)}{2}\right)\end{array}\right)\\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L}d_{li}\frac{1}{P_l\left(x_i\right)}\left(\frac{\partial e_{r,l}\left(x_i\right)}{\partial\theta}-\frac{\partial P_l\left(x_i\right)}{\partial\theta}e_{r,l}\left(x_i\right)\right)\\
&=L_n.
\end{aligned}$$

■

CHAPTER 2

Estimation of Average Treatment Effects for Massively Unbalanced Binary Outcomes

2.1 Introduction

We examine the properties of the maximum likelihood estimator (MLE) for logit models when the outcome is the occurrence or not of a low probability event. King and Zeng (2001) considered logistic regression for rare events data and focused on correcting the bias of estimators of the regression coefficients and event probabilities. They were motivated by the fact that in political science data the binary dependent variable takes the value one (for " events", such as wars, coups, presidential vetoes, the decision of citizens to run for political office, or infection by an uncommon disease) much less frequently than the value zero (for " nonevents"). The analysis of rare events is relevant for economics because some of the big data sets are collected from online sources where the number of events (such as " clicks" and " purchases") is much smaller than the number of nonevents.

King and Zeng (2001) considered various statistical problems with rare event data, including sample selection problems and finite sample biases. They primarily discussed issues related to the predicted event probabilities. In this chapter, we focus on the finite sample bias in the estimator of the average treatment effects (ATE). We derive the higher order bias of the logit MLE and the implied bias of the estimator of the ATE, and analyze the finite sample properties of the bias corrected estimator by Monte Carlo simulations.

We note that the higher order properties of the logit MLE have already been analyzed by Chen and Giles (2012). Our focus is different from Chen and Giles (2012) in that we analyze the implied ATE, which can be interpreted to be a nonlinear transformation of the parameters of the logit model. We recognize that Rilstone (1996) analyzed the higher order properties of some fixed nonlinear transformations of parameter estimates. The ATE can be understood to be a data dependent nonlinear transformation of the logit parameters, and as such, Rilstone (1996) analysis does not apply to the ATE.

It may seem unnatural to use a higher order expansion to derive the finite sample bias in the case of rare events. Indeed, Wang (2020) proposed an intuitive asymptotic approximation where the intercept term of the logit model diverges to negative infinity as a function of the sample size, so that the implied probability of the event converges to zero. We show that his asymptotic approximation is equivalent to the usual first-order asymptotic approximation where the sample size grows to infinity and the parameters are fixed. Therefore, Wang (2020)'s results have implications for the efficiency of various sampling methods, but do not shed light on the finite sample behavior of logit MLE if events are rare.

In Section 2.2, we derive the higher order bias of the logit MLE as well as of the related estimator of the ATE. In Section 2.3, we provide intuition for the bias of the logit MLE exploiting the invariance property of the MLE. In Section 2.4, we argue why the higher order approach should be preferred over Wang (2020)'s intuitive asymptotics. In Section 2.5, we propose a new binary response model with constant elasticity, and in Section 2.6, we develop a trick to reduce the numerical instability in the calculation of the MLE for to rare events. In Section 2.7, we present the simulation evidence.

2.2 Higher Order Bias of the Logit MLE and Related ATE Estimator

In this section, we consider the logit model where the binary dependent variable $y_i = 1$ with probability equal to $\Lambda\left(x_i'\theta_0\right)$, where $\Lambda\left(t\right) \equiv \exp\left(t\right) / \left(1 + \exp\left(t\right)\right)$ denotes the cumulative distribution function (CDF) of the logistic distribution. We first review the second order bias of the generic MLE as discussed in Rilstone et al. (1996), and consider the second order bias of logit MLE, as was done in Chen and Giles (2012). We show that the second order bias thus calculated is larger if events are rare. We go on to characterize the second order bias of the average treatment effect (ATE).

2.2.1 Second Order Bias of MLE

The idea underlying the second order expansion of the generic MLE is straightforward. Suppose that the density of the random vector z_i is given by $f(z;\theta_0)$, and the MLE $\widehat{\theta}$ maximizes the joint log likelihood $\sum_{i=1}^n \log f(z_i;\theta)$. The first order condition is $\frac{1}{n}\sum_{i=1}^n v\left(z_i;\widehat{\theta}\right)=0$, where $v(z;\theta)=\partial \log f(z;\theta)/\partial\theta$. By manipulating this first order condition, we can derive the implication that $\widehat{\theta}-\theta_0$ can be decomposed into the sum of the first order term of order $O_p(n^{-1/2})$, the second order term of order $O_p(n^{-1})$, and the third order term of the smaller order. Specifically, it can be shown[1] that

$$\widehat{\theta}-\theta_0=\frac{1}{\sqrt{n}}\theta^{\epsilon}(0)+\frac{1}{2n}\theta^{\epsilon\epsilon}(0)+o_p\left(n^{-1}\right), \tag{2.1}$$

where

$$\theta^{\epsilon}(0)\equiv-\left(E\left[\frac{\partial v(z_i;\theta_0)}{\partial\theta'}\right]\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n v(z_i;\theta_0)\right),$$

$$\begin{aligned}\frac{1}{2}\theta^{\epsilon\epsilon}(0)\equiv&-\frac{1}{2}\left(E\left[\frac{\partial v(z_i;\theta_0)}{\partial\theta'}\right]\right)^{-1}\begin{bmatrix}\theta^{\epsilon}(0)'\left(E\left[\frac{\partial^2 v_1(z_i;\theta_0)}{\partial\theta\partial\theta'}\right]\right)\theta^{\epsilon}(0)\\ \vdots\\ \theta^{\epsilon}(0)'\left(E\left[\frac{\partial^2 v_K(z_i;\theta_0)}{\partial\theta\partial\theta'}\right]\right)\theta^{\epsilon}(0)\end{bmatrix}\\ &-\left(E\left[\frac{\partial v(z_i;\theta_0)}{\partial\theta'}\right]\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\frac{\partial v(z_i;\theta_0)}{\partial\theta'}-E\left[\frac{\partial v(z_i;\theta_0)}{\partial\theta'}\right]\right)\right)\theta^{\epsilon}(0).\end{aligned}$$

Letting B denote a consistent estimator of $E\left[\frac{1}{2}\theta^{\epsilon\epsilon}(0)\right]$, the bias corrected estimator $\widetilde{\theta}$ is calculated as $\widetilde{\theta}=\widehat{\theta}-B/n$. In Appendix 2.9.3, we review the characterization of the second

[1] See Appendix 2.9.4 for details.

order bias for generic MLE, and in Section 2.9.4, we discuss the second order bias for logit MLE, replicating Chen and Giles (2012).[2] The discussion in Section 2.9.4 indicates that the bias corrected estimator can be computed by the following algorithm:

1. Calculate the MLE $\widehat{\theta}$.

2. Let

$$\hat{\Lambda}_i \equiv \frac{\exp\left(x_i'\widehat{\theta}\right)}{1+\exp\left(x_i'\widehat{\theta}\right)}.$$

3. Let

$$A \equiv -\frac{1}{n}\sum_{i=1}^{n}\hat{\Lambda}_i\left(1-\hat{\Lambda}_i\right)x_i x_i'. \tag{2.2}$$

4. Let

$$C_k \equiv -\frac{1}{n}\sum_{i=1}^{n}\hat{\Lambda}_i\left(1-\hat{\Lambda}_i\right)\left(1-2\hat{\Lambda}_i\right)x_{i,k}x_i x_i'. \tag{2.3}$$

5. Let

$$T_k = \frac{1}{2}\operatorname{trace}\left(C_k A^{-1}\right). \tag{2.4}$$

6. Let

$$B = A^{-1}\begin{bmatrix} T_1 \\ \vdots \\ T_k \end{bmatrix}. \tag{2.5}$$

7. Let $\widetilde{\theta} = \widehat{\theta} - B/n$.

[2] Our bias formula looks somewhat different from the one presented in King and Zeng (2001), which in turn is based on McCullagh and Nelder (2019). It can be shown that the two bias formulae are in fact identical. See Appendix 2.9.5.

2.2.2 Second Order Bias of the ATE

We now present the main result of this section. We consider the treatment effect model where we have $y_i = 1$ with probability $\Lambda\left(x_i'\theta_{0,(1)}\right)$ under treatment ($D = 1$), and with probability $\Lambda\left(x_i'\theta_{0,(0)}\right)$ under control ($D = 0$). We can estimate the $\theta_{0,(1)}$ and $\theta_{0,(0)}$ for the two sub-samples with $D = 1$ and 0.[3] The average treatment effect is equal to $E\left[\Lambda\left(x_i'\theta_{0,(1)}\right)\right] - E\left[\Lambda\left(x_i'\theta_{0,(0)}\right)\right]$, which can be estimated by the natural estimator

$$\frac{1}{n}\sum_{i=1}^{n}\Lambda\left(x_i'\widehat{\theta}_{(1)}\right) - \frac{1}{n}\sum_{i=1}^{n}\Lambda\left(x_i'\widehat{\theta}_{(0)}\right),$$

where $\widehat{\theta}_{(1)}$ and $\widehat{\theta}_{(0)}$ denote the MLE of $\theta_{0,(1)}$ and $\theta_{0,(0)}$ using the treated and control subsamples. When the outcome 1 is a rare event under treatment and/or control, the bias of the natural estimator above can be corrected, where the bias correction has to reflect the nonlinearity of Λ as well as the finite sample bias of the MLE. In particular, the bias correction follows from the second order expansion

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\Lambda\left(x_i'\widehat{\theta}_{(1)}\right) - \frac{1}{n}\sum_{i=1}^{n}\Lambda\left(x_i'\theta_{0,(1)}\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\Lambda\left(x_i'\theta_{0,(1)}\right)\left(1-\Lambda\left(x_i'\theta_{0,(1)}\right)\right)x_i'\left(\widehat{\theta}_{(1)}-\theta_{0,(1)}\right) \\
&+ \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\Lambda\left(x_i'\theta_{0,(1)}\right)\left(1-\Lambda\left(x_i'\theta_{0,(1)}\right)\right)\left(1-2\Lambda\left(x_i'\theta_{0,(1)}\right)\right)\left(x_i'\left(\widehat{\theta}_{(1)}-\theta_{0,(1)}\right)\right)^2 \\
&+ o_p\left(n^{-1}\right).
\end{aligned} \tag{2.6}$$

In the expansion (2.6), the remainder term is noted to be of order $o_p(n^{-1})$. This reflects the assumption that the number of treated n_1 is the same order of magnitude as n, i.e., $n_1/n = O(1)$. If the number of treated (or the number of control n_0) is very small relative to the sample size, the above approximation may not be very accurate. The bias of the ATE estimator derives from the two terms on the right. From (2.1), we can see that the first term

[3] We assume that the treatment assignment is unconfounded

on the right of (2.6) has the expansion

$$\left(\frac{1}{n}\sum_{i=1}^{n}\Lambda\left(x_i'\theta_{0,(1)}\right)\left(1-\Lambda\left(x_i'\theta_{0,(1)}\right)\right)x_i'\right)\left(\frac{1}{\sqrt{n_1}}\theta_{(1)}^{\epsilon}(0)+\frac{1}{2n_1}\theta_{(1)}^{\epsilon\epsilon}(0)+o_p\left(n_1^{-1}\right)\right),$$

and likewise, the second term on the right of (2.6) has the expansion

$$\text{trace}\left\{\begin{array}{c}\left(\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\Lambda\left(x_i'\theta_{0,(1)}\right)\left(1-\Lambda\left(x_i'\theta_{0,(1)}\right)\right)\left(1-2\Lambda\left(x_i'\theta_{0,(1)}\right)\right)x_ix_i'\right)\\ \cdot\left(\frac{1}{\sqrt{n_1}}\theta_{(1)}^{\epsilon}(0)+\frac{1}{2n_1}\theta_{(1)}^{\epsilon\epsilon}(0)+o_p\left(n_1^{-1}\right)\right)\left(\frac{1}{\sqrt{n_1}}\theta_{(1)}^{\epsilon}(0)+\frac{1}{2n_1}\theta_{(1)}^{\epsilon\epsilon}(0)+o_p\left(n_1^{-1}\right)\right)'\end{array}\right\}.$$

Here, the $\theta_{(1)}^{\epsilon}(0)$ and $\theta_{(1)}^{\epsilon\epsilon}(0)$ are the counterparts of $\theta_{(1)}^{\epsilon}(0)$ and $\theta_{(1)}^{\epsilon\epsilon}(0)$ for $\widehat{\theta}_{(1)}$. Ignoring all the terms of order $o_p\left(n^{-1}\right)$ while using the assumption $n_1/n = O(1)$, we get that the second order bias through $\widehat{\theta}_{(1)}$ is equal to

$$\begin{aligned}&\frac{1}{2n_1}\left(\frac{1}{n}\sum_{i=1}^{n}\Lambda\left(x_i'\theta_{0,(1)}\right)\left(1-\Lambda\left(x_i'\theta_{0,(1)}\right)\right)x_i'\right)E_{(1)}\left[\theta_{(1)}^{\epsilon\epsilon}(0)\right]\\ &+\frac{1}{2n_1}\text{trace}\left\{\begin{array}{c}\left(\frac{1}{n}\sum_{i=1}^{n}\Lambda\left(x_i'\theta_{0,(1)}\right)\left(1-\Lambda\left(x_i'\theta_{0,(1)}\right)\right)\left(1-2\Lambda\left(x_i'\theta_{0,(1)}\right)\right)x_ix_i'\right)\\ \cdot E_{(1)}\left[\theta_{(1)}^{\epsilon}(0)\,\theta_{(1)}^{\epsilon}(0)'\right]\end{array}\right\},\end{aligned}\tag{2.7}$$

where $E_{(1)}$ denotes the expectation taken with respect to the distribution of the treated. The first term of (2.7) depends on $E_{(1)}\left[\theta_{(1)}^{\epsilon\epsilon}(0)\right]$, and hence, is directly related to the second order bias of $\widehat{\theta}_{(1)}$ itself. The second term of (2.7) is the expected value of the second term of (2.6). It comes from the second order Taylor series expansion of $\frac{1}{n}\sum_{i=1}^{n}\Lambda\left(x_i'\widehat{\theta}_{(1)}\right)$, and as such, it reflects the nonlinearity of Λ. See Section 2.2.3 below for details on implementing the bias correction, if this expansion is applied to both $\frac{1}{n}\sum_{i=1}^{n}\Lambda\left(x_i'\widehat{\theta}_{(1)}\right)$ and $\frac{1}{n}\sum_{i=1}^{n}\Lambda\left(x_i'\widehat{\theta}_{(0)}\right)$.

2.2.3 Details of Estimating the Bias of the ATE

Applying the same reasoning as (2.7) to $\widehat{\theta}_{(0)}$, and combining it with the information equality that leads to the simplification

$$E_{(1)}\left[\theta_{(1)}^{\epsilon}(0)\,\theta_{(1)}^{\epsilon}(0)'\right]=-\left(E_{(1)}\left[\frac{\partial v}{\partial\theta'}\right]\right)^{-1},$$
$$E_{(0)}\left[\theta_{(0)}^{\epsilon}(0)\,\theta_{(0)}^{\epsilon}(0)'\right]=-\left(E_{(0)}\left[\frac{\partial v}{\partial\theta'}\right]\right)^{-1},$$

we can conclude that the second order bias of the ATE is given by

$$\begin{aligned}
&\frac{1}{2n_1}\left(\frac{1}{n}\sum_{i=1}^{n}\Lambda\left(x_i'\theta_{0,(1)}\right)\left(1-\Lambda\left(x_i'\theta_{0,(1)}\right)\right)x_i'\right)E_{(1)}\left[\theta_{(1)}^{\epsilon\epsilon}\left(0\right)\right] \\
&-\frac{1}{2n_0}\left(\frac{1}{n}\sum_{i=1}^{n}\Lambda\left(x_i'\theta_{0,(0)}\right)\left(1-\Lambda\left(x_i'\theta_{0,(0)}\right)\right)x_i'\right)E_{(0)}\left[\theta_{(0)}^{\epsilon\epsilon}\left(0\right)\right] \\
&-\frac{1}{2n_1}\operatorname{trace}\left\{\left(\frac{1}{n}\sum_{i=1}^{n}\Lambda\left(x_i'\theta_{0,(1)}\right)\left(1-\Lambda\left(x_i'\theta_{0,(1)}\right)\right)\left(1-2\Lambda\left(x_i'\theta_{0,(1)}\right)\right)x_ix_i'\right)\left(E_{(1)}\left[v^{\theta}\right]\right)^{-1}\right\} \\
&+\frac{1}{2n_0}\operatorname{trace}\left\{\left(\frac{1}{n}\sum_{i=1}^{n}\Lambda\left(x_i'\theta_{0,(0)}\right)\left(1-\Lambda\left(x_i'\theta_{0,(0)}\right)\right)\left(1-2\Lambda\left(x_i'\theta_{0,(0)}\right)\right)x_ix_i'\right)\left(E_{(0)}\left[v^{\theta}\right]\right)^{-1}\right\}.
\end{aligned} \tag{2.8}$$

which can be estimated in the standard way.

1. Suppose that there are n_1 observations such that $D = 1$. We estimate $\theta_{0,(1)}$ by MLE $\widehat{\theta}_{(1)}$ from this sample. Our preceding discussion implies that the counterparts of $\widehat{E}\left[\theta^{\epsilon\epsilon}\left(0\right)\right]$ and $\widehat{E}\left[v^{\theta}\right]$, which we will denote as $\widehat{E}_{(1)}\left[\theta_{(1)}^{\epsilon\epsilon}\left(0\right)\right]$ and $\widehat{E}_{(1)}\left[v^{\theta}\right]$ can be characterized by the following steps:

   (a) Calculate the MLE $\widehat{\theta}_{(1)}$.

   (b) Let
   $$\hat{\Lambda}_{i,(1)} = \frac{\exp\left(x_i'\widehat{\theta}_{(1)}\right)}{1+\exp\left(x_i'\widehat{\theta}_{(1)}\right)}. \tag{2.9}$$

   (c) Let
   $$\widehat{E}_{(1)}\left[v^{\theta}\right] \equiv A_{(1)} = -\frac{1}{n_1}\sum_{D=1}\hat{\Lambda}_{i,(1)}\left(1-\hat{\Lambda}_{i,(1)}\right)x_ix_i'. \tag{2.10}$$

   (d) Let
   $$C_{k,(1)} = -\frac{1}{n_1}\sum_{D=1}\hat{\Lambda}_{i,(1)}\left(1-\hat{\Lambda}_{i,(1)}\right)\left(1-2\hat{\Lambda}_{i,(1)}\right)x_{i,k}x_ix_i'.$$

   (e) Let
   $$T_{k,(1)} = \frac{1}{2}\operatorname{trace}\left(C_{k,(1)}A_{(1)}^{-1}\right).$$

(f) Let

$$\widehat{E}_{(1)}\left[\theta^{\epsilon\epsilon}_{(1)}(0)\right] \equiv \left(A_{(1)}\right)^{-1}\begin{bmatrix} T_{1,(1)} \\ \vdots \\ T_{k,(1)} \end{bmatrix}. \tag{2.11}$$

2. Likewise, we calculate the $\widehat{E}_{(0)}\left[\theta^{\epsilon\epsilon}_{(0)}(0)\right]$ and $\widehat{E}_{(0)}\left[v^{\theta}\right]$:

(a) Calculate the MLE $\widehat{\theta}_{(0)}$.

(b) Let

$$\hat{\Lambda}_{i,(0)} \equiv \frac{\exp\left(x_i'\widehat{\theta}_{(0)}\right)}{1+\exp\left(x_i'\widehat{\theta}_{(0)}\right)}. \tag{2.12}$$

(c) Let

$$\widehat{E}_{(0)}\left[v^{\theta}\right] \equiv A_{(0)} = -\frac{1}{n_0}\sum_{D=0}\hat{\Lambda}_{i,(0)}\left(1-\hat{\Lambda}_{i,(0)}\right)x_i x_i'. \tag{2.13}$$

(d) Let

$$C_{k,(0)} = -\frac{1}{n_0}\sum_{D=0}\hat{\Lambda}_{i,(0)}\left(1-\hat{\Lambda}_{i,(0)}\right)\left(1-2\hat{\Lambda}_{i,(0)}\right)x_{i,k}x_i x_i'.$$

(e) Let

$$T_{k,(0)} = \frac{1}{2}\operatorname{trace}\left(C_{k,(0)}A^{-1}_{(0)}\right).$$

(f) Let

$$\widehat{E}_{(0)}\left[\theta^{\epsilon\epsilon}_{(0)}(0)\right] \equiv \left(A_{(0)}\right)^{-1}\begin{bmatrix} T_{1,(0)} \\ \vdots \\ T_{k,(0)} \end{bmatrix}. \tag{2.14}$$

3. The second order bias is computed to be

$$\begin{aligned}\frac{B_{ATE}}{n} \equiv & \frac{1}{2n_1}\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\Lambda}_{i,(1)}\left(1-\hat{\Lambda}_{i,(1)}\right)x_i'\right)\widehat{E}_{(1)}\left[\theta_{(1)}^{\epsilon\epsilon}(0)\right] \\ & -\frac{1}{2n_0}\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\Lambda}_{i,(0)}\left(1-\hat{\Lambda}_{i,(0)}\right)x_i'\right)\widehat{E}_{(0)}\left[\theta_{(0)}^{\epsilon\epsilon}(0)\right] \\ & -\frac{1}{2n_1}\text{trace}\left\{\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\Lambda}_{i,(1)}\left(1-\hat{\Lambda}_{i,(1)}\right)\left(1-2\hat{\Lambda}_{i,(1)}\right)x_i x_i'\right)\left(\widehat{E}_{(1)}\left[v^{\theta}\right]\right)^{-1}\right\} \\ & +\frac{1}{2n_0}\text{trace}\left\{\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\Lambda}_{i,(0)}\left(1-\hat{\Lambda}_{i,(0)}\right)\left(1-2\hat{\Lambda}_{i,(0)}\right)x_i x_i'\right)\left(\widehat{E}_{(0)}\left[v^{\theta}\right]\right)^{-1}\right\}\end{aligned} \tag{2.15}$$

using (2.9), (2.10), (2.11), (2.12), (2.13), and (2.14) computed in previous steps.

2.3 Intuition

In this section, we provide the intuition underlying the second order bias. Using a model without regressors, we explain that the bias is due to the inherent nonlinearity of the logit model. We show this using the invariance property of the MLE. The invariance is also used to explain the bias of the predicted probabilities in models without regressors.

We consider the binary response model where $y = 1$ with probability equal to $\Lambda(\theta) = \exp(\theta)/(1+\exp(\theta))$. The second order bias of the logit MLE, presented in Appendix 2.9.4, simplifies to

$$-\frac{1}{2n}\frac{(1-2\Lambda)}{\Lambda(1-\Lambda)} \tag{2.16}$$

with $\Lambda = \Lambda(\theta_0)$. Note that if $\Lambda \approx 0$, the bias is negative. Also note that

$$\lim_{\Lambda\to 0}\left(-\frac{(1-2\Lambda)}{\Lambda(1-\Lambda)}\right) = -\infty$$

so the bias is larger if we are dealing with rare events. In order to understand the bias, we note that the first order condition can be rewritten as $\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \Lambda\left(\widehat{\theta}\right)\right) = 0$. In other words, the logit MLE solves

$$\Lambda\left(\widehat{\theta}\right)=\overline{y} \tag{2.17}$$

and hence

$$\widehat{\theta}=\Lambda^{-1}\left(\overline{y}\right)=\ln\frac{\overline{y}}{1-\overline{y}}.$$

By the CLT, we have $\sqrt{n}\left(\overline{y}-\Lambda\right)\to N\left(0,\Lambda\left(1-\Lambda\right)\right)$, so we can write that

$$\overline{y}=\Lambda+\frac{1}{\sqrt{n}}\left(\sqrt{\Lambda\left(1-\Lambda\right)}Z+o_p\left(1\right)\right)$$

where $Z\sim N\left(0,1\right)$. It follows that

$$\begin{aligned}\widehat{\theta}&=\ln\left(\frac{\Lambda+\frac{1}{\sqrt{n}}\left(\sqrt{\Lambda\left(1-\Lambda\right)}Z+o_p\left(1\right)\right)}{1-\Lambda-\frac{1}{\sqrt{n}}\left(\sqrt{\Lambda\left(1-\Lambda\right)}Z+o_p\left(1\right)\right)}\right)\\&=\ln\left(\frac{\Lambda}{1-\Lambda}\right)+\frac{n^{-1/2}}{\sqrt{\Lambda\left(1-\Lambda\right)}}Z-n^{-1}\frac{1}{2}\frac{1-2\Lambda}{\Lambda\left(1-\Lambda\right)}Z^2+o_p\left(n^{-1}\right)\\&=\theta_0+\frac{n^{-1/2}}{\sqrt{\Lambda\left(1-\Lambda\right)}}Z-n^{-1}\frac{1}{2}\frac{1-2\Lambda}{\Lambda\left(1-\Lambda\right)}Z^2+o_p\left(n^{-1}\right)\end{aligned}$$

It follows that the second order bias of $\widehat{\theta}$ is

$$-n^{-1}\frac{1}{2}\frac{1-2\Lambda}{\Lambda\left(1-\Lambda\right)}E\left[Z^2\right]=-n^{-1}\frac{1}{2}\frac{1-2\Lambda}{\Lambda\left(1-\Lambda\right)} \tag{2.18}$$

confirming the second order bias calculation in (2.16).

Remark 1 *The first order condition of the MLE can be a convenient tool to understand the (lack of) bias of the average of the predicted probabilities in a logit model with regressors. For this purpose, we note that the probability $E\left[\Lambda\left(x'\theta_0\right)\right]$ of $y=1$ can be estimated by $\frac{1}{n}\sum_{i=1}^{n}\Lambda\left(x_i'\widehat{\theta}\right)$. We will assume that the first component of x_i is an intercept term. If so, we recall by the first order condition that $\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\Lambda\left(x_i'\widehat{\theta}\right)\right)=0$, and therefore, so*

$$E\left[\frac{1}{n}\sum_{i=1}^{n}\Lambda\left(x_i'\widehat{\theta}\right)\right]=E\left[\frac{1}{n}\sum_{i=1}^{n}y_i\right]=\frac{1}{n}\sum_{i=1}^{n}\Lambda\left(x'\theta_0\right).$$

It follows that the bias is exactly zero. Now note that the second order bias is merely an approximation of the actual bias, and one may want to assess whether the higher order *approximation does a good job by asking whether the second order bias for the predicted probability is close to the actual bias (i.e., zero). The same second-order expansion as in Section 2 gives*

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\Lambda\left(x_i'\widehat{\theta}\right)-\frac{1}{n}\sum_{i=1}^{n}\Lambda\left(x_i'\theta_0\right)=&\frac{1}{n}\sum_{i=1}^{n}\Lambda\left(x_i'\theta_0\right)\left(1-\Lambda\left(x_i'\theta_0\right)\right)x_i'\left(\widehat{\theta}-\theta_0\right)\\&+\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\Lambda\left(x_i'\theta_0\right)\left(1-\Lambda\left(x_i'\theta_0\right)\right)\left(1-2\Lambda\left(x'\theta_0\right)\right)\left(x_i'\left(\widehat{\theta}-\theta_0\right)\right)^2\\&+o_p\left(n^{-1}\right)\end{aligned}$$

Straightforward algebra shows that the second order bias of the $\frac{1}{n}\sum_{i=1}^{n}\Lambda\left(x_i'\widehat{\theta}\right)$ *is indeed zero, confirming the intuition. See Appendix 2.9.6.*

Remark 2 *The second order bias of the ATE estimator can be shown to be zero if the treatment is randomly assigned. See Appendix 2.9.7 for a proof. By random assignment, we mean that* D *is independent of* x *so that the propensity score* $\Pr\left(D_i=1|\, x_i\right)$ *is constant in* x*. Under random assignment, the distribution of* x *is identical across the two subsamples* $D=1$ *and* $D=0$*. Therefore, the intuition to be discussed in Remark 1 applies, and the second order bias is zero. Under unconfounded treatment assignment we should expect some amount of second order bias. Its magnitude is an empirical matter.*

Remark 3 *Note that the ATE estimator is the difference of the average predicted probabilities for the treated and the controls. This estimator is biased because the average is over the full sample and not over the subsamples of the treated and the controls. Under random assignment, though, the distribution of* X *is the same in the subsamples, so that the ATE estimator is unbiased, even if events are rare.*

2.4 Comparison with Wang (2020)

We now compare our higher order asymptotics with Wang (2020)'s asymptotics where the probability of $y=1$ is assumed to converge to zero as a function of the sample size. Using the same model without regressors that was discussed in the previous section, we examine Wang (2020)'s asymptotic analysis, and argue that his asymptotics is identical to the traditional fixed parameter asymptotics for all practical purposes.

The data are $y_1, \ldots, y_n$ which are IID Bernoulli variables such that $y_i = 1$ with probability p_n. We assume that $p_n \propto n^{-\delta}$ with $0 \leq \delta < 1$, and consider the normalized sum

$$\sum_{i=1}^{n} \frac{y_i - p_n}{\sqrt{np_n (1 - p_n)}}.$$

In Appendix 2.9.8 we show that the Lyapunov condition is satisfied[4] and

$$\frac{\sqrt{n} (\overline{y} - p_n)}{\sqrt{p_n (1 - p_n)}} = \sum_{i=1}^{n} \frac{y_i - p_n}{\sqrt{np_n (1 - p_n)}} \to N(0, 1), \tag{2.19}$$

so we can write

$$\overline{y} = p_n + \frac{\sqrt{p_n (1 - p_n)}}{\sqrt{n}} Z_n,$$

where $Z_n = O_p(1)$ is such that $E[Z_n] = 0$ and $\operatorname{Var}(Z_n) = 1$. In Appendix 2.9.9 we derive the expansion

$$\sqrt{np_n (1 - p_n)} \left(\widehat{\theta} - \theta_0\right) = Z_n - \frac{1}{2} \frac{1 - 2p_n}{\sqrt{np_n (1 - p_n)}} Z_n^2 + O_p\left(n^{-(1-\delta)}\right). \tag{2.20}$$

The expansion (2.20) implies that the higher order bias can be calculated as

$$E\left[-\frac{1}{2} \frac{1 - 2p_n}{np_n (1 - p_n)} Z_n^2\right] = -\frac{1}{2} \frac{1 - 2p_n}{np_n (1 - p_n)}.$$

If we compare this expression with the higher order bias (2.18) based on the fixed probability Λ, equating Λ and p_n, we see that the higher order bias under the asymptotics where $p_n \to 0$ is identical to the higher order bias under the asymptotics where p_n is fixed at Λ. In other words, Wang's asymptotics gives the same higher order bias as the fixed parameter asymptotics.

This analysis raises the question about the relevance of Wang (2020)'s asymptotic framework for our purpose. It is helpful to make an explicit link between p_n and the logit

[4] Appendix 2.9.8 also notes that $\delta = 1$ (so that $p_n \propto n^{-1}$) is not compatible with asymptotic normality, and that this rate is not appropriate for logit models.

model, which we will do by writing $p_n = \Lambda(\theta_n)$. The sum of y_i over the entire sample is binom$(n, \Lambda(\theta_n))$. Using Wang's notation, we have $n_1 \sim$binom$(n, \Lambda(\theta_n))$. His equation (2) means that he is considering $p_n \propto n^{-\delta}$ with $0 \leq \delta < 1$, as we do here.[5]

His Theorem 1 boils down to $\sqrt{n_1}\left(\widehat{\theta} - \theta\right) \to N(0,1)$ in the model with only an intercept. Using his equation (3), this is equivalent to the approximation $\sqrt{n\Lambda(\theta_n)}\left(\widehat{\theta} - \theta\right) \approx N(0,1)$ or

$$\widehat{\theta} \approx N\left(\theta, \frac{1}{n\Lambda(\theta_n)}\right) = N\left(\theta, \frac{1}{np_n}\right).$$

If we ignore the higher order term involving Z_n^2, our (2.20) along with (2.19) implies $\sqrt{np_n(1-p_n)}\left(\widehat{\theta} - \theta\right) \to N(0,1)$ or

$$\widehat{\theta} \approx N\left(\theta, \frac{1}{np_n(1-p_n)}\right),$$

which is the same approximation that we obtain when $p_n = \Lambda$ and does not vary as a function of the sample size. When $p_n \approx 0$, the difference between the two standard errors is small

$$\frac{1/(np_n)}{1/(np_n(1-p_n))} = 1 - p_n \approx 1.$$

Therefore, Wang (2020)'s asymptotic framework leads to the same first order asymptotic approximation as our (2.20) that was derived using the classical first order asymptotics with fixed parameters. A straightforward calculation suggests that even with regressors, Wang (2020)'s asymptotic framework does not offer a substantively different approximation than the classical first order asymptotic approximation. In particular Wang is silent on the second order bias.

[5]His equation (2) implies that $\Lambda(\theta_n) \to 0$, $n\Lambda(\theta_n) \to \infty$. Note that $\Lambda(\theta_n) \to 0$ means that $1/(1+\exp(\theta_n)) \to 1$, which in turn means that $\exp(\theta_n) \to 0$, or $\theta_n \to -\infty$. Also note that $n\Lambda(\theta_n) \to \infty$ rules out the Poisson approximation, because if $\Lambda(\theta_n) \propto n^{-1}$, we cannot have $n\Lambda(\theta_n) \to \infty$. On the other hand, $n\Lambda(\theta_n) \to \infty$ is satisfied as long as $\Lambda(\theta_n) \propto n^{-\delta}$ with $0 \leq \delta < 1$.

2.5 A New Binary Response Model for Rare Events

In this chapter, the primary object of interest is the ATE if the outcome is the occurrence of a rare event. On the other hand, in some applications it is more useful to have an estimate of the elasticity of the rare event probability with respect to the independent variables. Consider the logit model where $y = 1$ with probability $\Lambda(\alpha + x\beta)$, where x is a scalar that is measured on the log scale. The elasticity of the event probability with respect to x is equal to

$$\frac{\Lambda'(\alpha + x\beta)\beta}{\Lambda(\alpha + x\beta)} = (1 - \Lambda(\alpha + x\beta))\beta,$$

where we used $\partial\Lambda(t)/\partial t = \Lambda(t)(1 - \Lambda(t))$. Note that $1 - \Lambda(\alpha + x\beta) \approx 1$ if the event is rare, so the logit model exhibits near constant elasticity for rare events. If the elasticity is approximated by β, one may be interested in correcting the bias of the MLE of β. The simulation results in Tables 2.2, 2.3, and 2.4 show, that the bias in the MLE for β is modest but may be important depending on the application.[6] Therefore, one can use the bias corrected estimator of the slope coefficient as the bias corrected estimator of the elasticity in the rare event case.

Given that the β is only an approximate elasticity, the case for bias correction may not be so compelling for logit models. Instead one may choose to work with a model that has a constant elasticity. Let $P(x)$ denote the probability that $y = 1$ as a function of x. A model that has a constant elasticity should satisfy

$$\frac{P'(x)}{P(x)} = \text{constant}$$

assuming that x is measured in logs. This is equivalent to

$$\frac{d\ln P(x)}{dx} = \beta$$

[6] In Table 2.2, for the $\alpha_0 = -2.5$ and $n = 500$ combination, the bias of $\hat{\beta}$ is 0.0168, where the true value of β is 1. So, the MLE overestimates the elasticity by 1.68 %, which is reduced to 0.41 % by the bias corrected estimator.

for some β, so by integration we obtain $\ln P(x) = \alpha + \beta x$, or

$$P(x) = \exp(\alpha + \beta x)$$

as a model of constant elasticity.[7] Because $\exp(\alpha + \beta x) \approx \exp(\alpha + \beta x) / (1 + \exp(\alpha + \beta x))$ when $\exp(\alpha + \beta x) \approx 0$,[8] one can argue that this new model is an approximation of the logit model but with the convenient feature of a constant elasticity.

2.6 Artificial Censoring - Overcoming the Numerical Instability

In our Monte Carlo simulations, we encountered numerical stability problems with the computation of the MLE for extremely small values of α. The problem is that the optimization algorithm does not converge for such values of α that may be visited during the search for the MLE. As a consequence, we could not calculate the MLE for these data sets. Given that the log likelihood of the logit model is globally concave, in theory this sort of problem should not happen. The problem is that the Hessian of the log-likelihood is close to singular for rare events. We offer a simple solution, which seems to resolve the problem.

Consider the logit model where $y = 1$ with probability $\Lambda(\alpha + x\beta)$, where x is a scalar. We would normally maximize the log-likelihood

$$L(\alpha, \beta) = \sum_{i=1}^{n} \left(y_i \log \Lambda(\alpha + x_i \beta) + (1 - y_i) \log(1 - \Lambda(\alpha + x_i \beta)) \right).$$

---

[7]When $\alpha \approx -\infty$ and the support of x is bounded, we can guarantee that $\exp(\alpha + \beta x) < 1$. If the support of x is not bounded, but if we are sure that $\alpha + \beta x < 0$ for most values of x, we may want to adopt a parameterization

$$P(x) = \Psi(\alpha + \beta x)$$

where $\Psi(t) = \frac{1}{2}\exp(t)$ if $t < 0$, and $\Psi(t) = 1 - \frac{1}{2}\exp(-t)$ if $t > 0$.

[8]It is in the sense that

$$\frac{t}{1+t} = t + O(t^2).$$

It is straightforward to see that

$$\frac{\partial^2 L(\alpha,\beta)}{\partial\alpha^2} = -\sum_{i=1}^{n} \Lambda(\alpha + x_i\beta)(1 - \Lambda(\alpha + x_i\beta)),$$
$$\frac{\partial^2 L(\alpha,\beta)}{\partial\alpha\partial\beta} = -\sum_{i=1}^{n} \Lambda(\alpha + x_i\beta)(1 - \Lambda(\alpha + x_i\beta))\, x_i,$$
$$\frac{\partial^2 L(\alpha,\beta)}{\partial\beta^2} = -\sum_{i=1}^{n} \Lambda(\alpha + x_i\beta)(1 - \Lambda(\alpha + x_i\beta))\, x_i^2,$$

so the second derivative is in theory strictly negative definite. On the other hand, if $\alpha \approx -\infty$, we have $\Lambda(\alpha + x_i\beta) \approx 0$ and as a consequence, the second derivative matrix is close to zero, and therefore close to singular. Because the step length of the Newwton-Raphson algorithm depends on the inverse of the Hessian, the algorithm may not converge.

In order to overcome this problem, we consider artificial censoring of the $y = 0$ outcomes. The rare event logit model is unchanged and x is always observed. However we censor observations with $y = 0$ with probability π. So there are three possible outcomes, $y = 1$ (and observed), with probability $\Lambda(\alpha + x\beta)$, $y = 0$ and observed with probability $(1-\pi)(1 - \Lambda(\alpha + x\beta))$, and $y = 0$ and not observed with probability $\pi(1 - \Lambda(\alpha + x\beta))$.

Therefore, the probability that the econometrician observes outcome $y = 1$ conditional on x, is

$$\begin{aligned}
\frac{\Lambda(\alpha + x\beta)}{\Lambda(\alpha + x\beta) + (1-\pi)(1 - \Lambda(\alpha + x\beta))} &= \frac{\frac{\exp(\alpha + x\beta)}{1+\exp(\alpha+x\beta)}}{\frac{\exp(\alpha+x\beta)}{1+\exp(\alpha+x\beta)} + \exp(\delta)\frac{1}{1+\exp(\alpha+x\beta)}} \\
&= \frac{\exp(\alpha + x\beta)}{\exp(\alpha + x\beta) + \exp(\delta)} \\
&= \frac{\exp(\alpha^* + x\beta)}{\exp(\alpha^* + x\beta) + 1} \\
&= \Lambda(\alpha^* + x\beta),
\end{aligned}$$

and the probability the econometrician observes $y = 0$ given x is

$$\begin{aligned}
\frac{(1-\pi)\left(1-\Lambda\left(\alpha+x\beta\right)\right)}{\Lambda\left(\alpha+x\beta\right)+(1-\pi)\left(1-\Lambda\left(\alpha+x\beta\right)\right)} &= \frac{\exp\left(\delta\right)\frac{1}{1+\exp(\alpha+x\beta)}}{\frac{\exp(\alpha+x\beta)}{1+\exp(\alpha+x\beta)}+\exp\left(\delta\right)\frac{1}{1+\exp(\alpha+x\beta)}} \\
&= \frac{\exp\left(\delta\right)}{\exp\left(\alpha+x\beta\right)+\exp\left(\delta\right)} \\
&= \frac{1}{\exp\left(\alpha^{*}+x\beta\right)+1} \\
&= 1-\Lambda\left(\alpha^{*}+x\beta\right),
\end{aligned}$$

where $\Lambda\left(\alpha+x\beta\right)+(1-\pi)\left(1-\Lambda\left(\alpha+x\beta\right)\right)$ is the probability that the outcome is observed, and

$$\exp\left(\delta\right) \equiv 1-\pi, \quad \alpha^{*} \equiv \alpha-\delta.$$

Note that $\delta < 0$. If we choose δ (i.e., π) such that α^{*} is not close to $-\infty$, the Hessian for the censored sample is not (close to) singular. In general π can be chosen so that the expected number of observed 0 outcomes is about equal to the number of observed 1 outcomes.

Because δ is chosen by the econometrician, from an estimate α^{*} for the artificially censored sample we can back out α.

In order to see whether this trick is useful, we drew one sample of size $n = 50,000$ and another sample of size $n = 100,000$ such that $\alpha = -10$, $\beta_1 = \cdots = \beta_9 = 1$, and all the nine independent variables are independent and $N(0,1)$. We used $\beta_1 = \cdots = \beta_9 = 0$ as the starting value of the Newton-Raphson algorithm. As for the starting value for α, we chose the true value of α (i.e., -10) for the full sample, while we used $\alpha^{*} = -10-\delta$ for the subsample after the random censoring. In our sample of $n = 50,000$, the number of 1's was 116 and the missing probability is set to $\pi = 1-0.002$, implying that $\delta = \ln(0.002) = -6.21$.) Convergence results for Newton-Raphson method are shown in Table 2.1.

2.7 Monte Carlo Simulations

2.7.1 Bias-corrected MLE

We examined the performance of MLE estimators and bias-corrected MLE estimators in a sampling experiment. We consider a logit model in which $y_i = 1$ with probability $p_i = e^{\alpha + x_i \beta} \big/ \left(1 + e^{\alpha + x_i \beta}\right)$. For simplicity, we assume that x_i is a scalar random variable with the uniform distribution U[0,1]. We let $\hat{\alpha}$ and $\hat{\beta}$ denote the MLE estimators of α and β, and $\tilde{\alpha}$ and $\tilde{\beta}$ the bias corrected estimators, using the formula discussed in Section 2.2.

Tables 2.2, 2.3, and 2.4 present the mean biases of these estimators for various combinations of parameters, based on Monte Carlo simulations with 5000 runs. Note that the αs were chosen ~~of~~ so that events are rare. Consistent with the theory, the bias corrected estimators remove most of the bias.

The bias of the MLE increases if the probability of the event decreases, and the bias decreases with the sample size.The mean bias of the bias corrected MLE does not follow the same pattern which confirms that the bias correction removes the systematic bias of the MLE.

2.7.2 Bias-corrected ATE Estimator

In this section, we report properties of bias corrected estimators of the average treatment effect (ATE). The treatment assignment is assumed to be unconfounded. In our simulations, we generate the conditioning variable x separately for the treated and the controls. The distribution of x is allowed to be different for these two sub-samples. When they are equal, the treatment assignment is independent of x so that the propensity score is constant, and the treatment is randomly assigned. When the distribution of x is different for the treated and controls, the treatment is not randomly assigned.

Total sample size is n; n_1 is the number of $D_i = 1$, the treated, and n_0 is the number of

$D_i = 0$, the controls. The number of replications is 1000. We let $\widehat{ATE_1}$ and $\widehat{ATE_2}$ denote the estimator of the ATE based on the MLE $\left(\widehat{\theta}_{(1)}, \widehat{\theta}_{(0)}\right)$, and on the biased corrected MLE, $\left(\widetilde{\theta}_{(1)}, \widetilde{\theta}_{(0)}\right)$, respectively, i.e.,

$$\widehat{ATE_1} \equiv \frac{1}{n}\sum_{i=1}^{n} \Lambda\left(x_i'\widehat{\theta}_{(1)}\right) - \frac{1}{n}\sum_{i=1}^{n} \Lambda\left(x_i'\widehat{\theta}_{(0)}\right), \tag{2.21}$$
$$\widehat{ATE_2} \equiv \frac{1}{n}\sum_{i=1}^{n} \Lambda\left(x_i'\widetilde{\theta}_{(1)}\right) - \frac{1}{n}\sum_{i=1}^{n} \Lambda\left(x_i'\widetilde{\theta}_{(0)}\right).$$

Let $\widetilde{ATE}_1$ and $\widetilde{ATE}_2$ denote the bias corrected versions of $\widehat{ATE_1}$ and $\widehat{ATE_2}$. The higher order bias of $\widehat{ATE_1}$ can be removed by using (2.15) in Appendix 2.2.3. Likewise, the higher order bias of $\widehat{ATE_2}$ can be removed, noting that the first two terms in (2.15) can be ignored because $\left(\widetilde{\theta}_{(1)}, \widetilde{\theta}_{(0)}\right)$ already are bias-corrected.

In Tables 2.5 - 2.18, we report the mean bias of these estimators. The mean biases over the Monte Carlo replications are calculated as the averages of $\widehat{ATE_1} - ATE_0$, $\widehat{ATE_2} - ATE_0$, $\widetilde{ATE}_1 - ATE_0$, $\widetilde{ATE}_2 - ATE_0$, where

$$ATE_0 \equiv \frac{1}{n}\sum_{i=1}^{n} \Lambda\left(x_i'\theta_{0,(1)}\right) - \frac{1}{n}\sum_{i=1}^{n} \Lambda\left(x_i'\theta_{0,(0)}\right).$$

We will discuss the Monte Carlo results separately for the " random assignment" case, where the distribution of x is identical across the treatment and control subsamples, and for the " nonrandom assignment", where the distribution of x is different across the treatment and control subsamples.

We adopt a different notation for the parameters of the model:

$$\theta_{0,(0)} = (\alpha_0 \ \beta_0)' \quad \theta_{0,(1)} = (\alpha_1 \ \beta_1)'$$

with α_0, α_1 the intercepts and β_0, β_1 the slope coefficients in the logit outcome models for the treated and controls.

2.7.2.1 Bias-corrected ATE - Constant Propensity Score

As was discussed in Remark 2, the second order bias of the ATE is zero when the propensity score is constant. In order to verify this result, we will first consider the case that the distributions of x are identical over the $D = 1$ and $D = 0$ subsamples. In Tables 2.5 - 2.14, we evaluate the performance of various estimators of the ATE under random assignment. Overall, we see that the original ATE estimator $\widehat{ATE}_1$ is largely free of bias, consistent with Remark 2. All estimators are unbiased even in the rare event case and if the treatment assignment is unbalanced. The bias also does not depend on the sample size and the event probability.

In Table 2.5, $\beta_1 = 2$ and $\alpha_1 = \alpha_0 + 1$ for $D_i = 1$. For $D_i = 0$, we let $\beta_0 = 1$, and we consider the different values of α_0 listed in Table 2.5. In Table 2.13, the treatment effect is larger, $\beta_1 = 4$. The rest of the data generating process (DGP hereafter) is unchanged. The conclusions are the same as for Table 2.5. In Table 2.14, we let $\beta_1 = \beta_0 = 1$, and we consider different values of $\alpha_1 = \alpha_0$ listed in Table 2.14. For each parameter value the ATE is 0. The rest of the DGP is unchanged. The conclusions are the same as for Table 2.5.

2.7.2.2 Bias-corrected ATE - General Propensity Score

As was discussed in Remark 2, the second order bias of the ATE is in general not equal to zero when the propensity score is not constant, and therefore, the distribution of x is different for the treated and the controls. In Tables 2.15 - 2.18, we consider such " nonrandom assignment". The discussion in Remark 2 suggests that the bias is large if there is a stark difference between the distribution of x in the two subsamples. Table 2.15 is meant to represent such a situation: for $D_i = 1$, $X \sim N[4, 1]$, and for $D_i = 0$, $X \sim N[0, 1]$. Figure 2.1 shows the implied propensity scores for the DGPs considered in Table 2.15 and Table 2.16. We see a modest bias for the selected parameter values. In any case, the bias corrected estimator removes most of the bias. The estimator based on the bias-corrected MLE is as

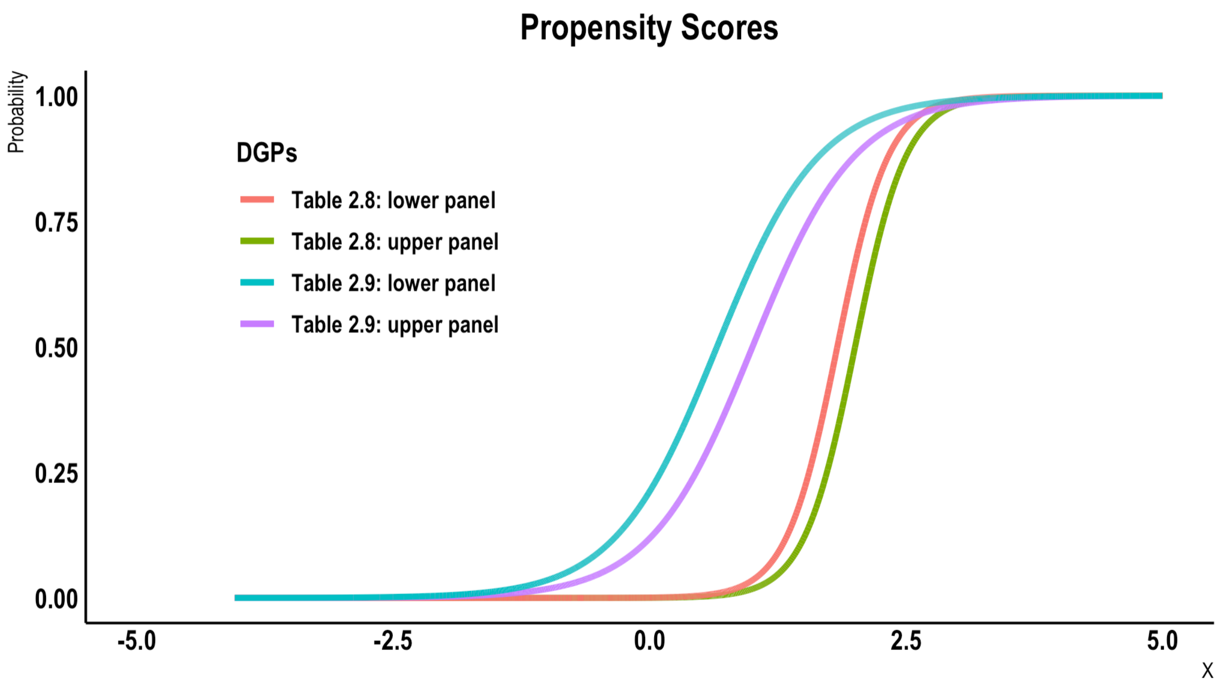

Figure 2.1: Distribution of Propensity Scores

biased as that based on the non-corrected MLE. The bias induced by the curvature of Λ is relatively large.

Based on Remark 2, we can speculate that if the difference of the distributions is not as stark, we expect the ATE to have a smaller bias. In order to verify this conjecture, we consider in Table 2.16 the case where the X is distributed as $N[2,1]$ and $N[0,1]$ in the treated and control subsamples, respectively. In general, the bias in $\widehat{ATE}_1$ is a lot smaller than in Table 2.15.

The bias formula (2.33) shows that it is an average of (2.31) and (2.32) weighted by the density of X. If the θs are identical across the two subsamples, then the counterparts of (2.31) and (2.32) are also identical across the two subsamples. Therefore, the bias of $\widehat{\theta}_{(1)}$ and $\widehat{\theta}_{(0)}$ may be similar, with the difference only arising from the possible difference of " weights". Therefore, one may think that the biases of the two terms on the right of (2.21) may almost cancel each other out when the θs are similar. In order to examine this conjecture, we considered cases where the θs are identical across the two subsamples in Tables 2.17 and 2.18. In Table 2.17, we consider nonrandom assignment, for $D_i = 1$: $X \sim N[4, 1]$; for $D_i = 0$: $X \sim N[0, 1]$. For both $D_i = 1$ and $D_i = 0$: $\beta_0 = \beta_1 = 1$, $\alpha_0 = \alpha_1$ is listed in the table. In Table 2.18, we consider nonrandom assignment and same parameters of interests, for $D_i = 1$: $X \sim N[2, 1]$; for $D_i = 0$: $X \sim N[0, 1]$. For both $D_i = 1$ and $D_i = 0$: $\beta_0 = \beta_1 = 1$, $\alpha_0 = \alpha_1$ is listed in the table. We conclude that the intuition that the biases cancel is not correct.

2.7.3 Tables

Table 2.1: Non-convergence and Artificial Censoring

Initial guess	α	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
n=50000										
	Full Sample									
$\alpha = -10$	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	Random Censoring									
$\alpha = -10 - (-6.21)$	-9.81	0.79	1.06	0.97	0.57	0.78	1.00	0.91	0.76	0.89
n=100000										
	Full Sample									
$\alpha = -10$	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	Random Censoring									
$\alpha = -10 - (-6.91)$	-10.4	0.95	1.44	1.19	1.02	1.21	0.99	1.24	1.41	0.77

Table 2.2: Mean Bias of MLE; $\beta_0 = 1$

	$\alpha_0 = -2.5$		$\alpha_0 = -3$		$\alpha_0 = -3.5$		$\alpha_0 = -4$	
	p[9] $= 12.25\%$		$p = 7.83\%$		$p = 4.91\%$		$p = 3.04\%$	
	Mean bias of α							
	$\hat{\alpha}$	$\tilde{\alpha}$	$\hat{\alpha}$	$\tilde{\alpha}$	$\hat{\alpha}$	$\tilde{\alpha}$	$\hat{\alpha}$	$\tilde{\alpha}$
n=500	-0.0249	-0.0033	-0.0357	-0.0005	-0.0703	-0.0109	-0.1164	-0.0130
n=750	-0.0179	-0.0036	-0.0243	-0.0013	-0.0435	-0.0053	-0.0784	-0.0135
n=1000	-0.0060	0.0046	-0.0143	0.0028	-0.0297	-0.0016	-0.0499	-0.0028
n=5000	-0.0021	0.0000	-0.0055	-0.0021	-0.0076	-0.0021	-0.0115	-0.0025
	Mean bias of β							
	$\hat{\beta}$	$\tilde{\beta}$	$\hat{\beta}$	$\tilde{\beta}$	$\hat{\beta}$	$\tilde{\beta}$	$\hat{\beta}$	$\tilde{\beta}$
n=500	0.0168	0.0041	0.0183	-0.0006	0.0391	0.0088	0.0630	0.0112
n=750	0.0152	0.0068	0.0165	0.0042	0.0319	0.0124	0.0564	0.0239
n=1000	-0.0007	-0.0068	0.0070	-0.0021	0.0189	0.0046	0.0285	0.0054
n=5000	-0.0002	-0.0014	0.0030	0.0012	0.0036	0.0008	0.0036	-0.0007

[9] p denotes the probability of y=1 for each parameter combination.

Table 2.3: Mean Bias of MLE; $\beta_0 = 1.5$

	$\alpha_0 = -2.5$		$\alpha_0 = -3$		$\alpha_0 = -3.5$		$\alpha_0 = -4$	
	$p = 15.62\%$		$p = 10.19\%$		$p = 6.48\%$		$p = 4.05\%$	
	Mean bias of α							
	$\hat{\alpha}$	$\tilde{\alpha}$	$\hat{\alpha}$	$\tilde{\alpha}$	$\hat{\alpha}$	$\tilde{\alpha}$	$\hat{\alpha}$	$\tilde{\alpha}$
n=500	-0.0243	-0.0043	-0.0345	-0.0028	-0.0617	-0.0098	-0.1049	-0.0169
n=750	-0.0155	-0.0023	-0.0224	-0.0017	-0.0374	-0.0037	-0.0674	-0.0116
n=1000	-0.0058	0.0018	-0.0128	0.0027	-0.0252	-0.0004	-0.0395	0.0013
n=5000	-0.0022	-0.0003	-0.0045	-0.0014	-0.0075	-0.0027	-0.0105	-0.0027
	Mean bias of β							
	$\hat{\beta}$	$\tilde{\beta}$	$\hat{\beta}$	$\tilde{\beta}$	$\hat{\beta}$	$\tilde{\beta}$	$\hat{\beta}$	$\tilde{\beta}$
n=500	0.0190	0.0031	0.0270	0.0042	0.0444	0.0092	0.0724	0.0146
n=750	0.0168	0.0063	0.0193	0.0043	0.0278	0.0052	0.0585	0.0219
n=1000	0.0040	-0.0060	0.0076	-0.0034	0.0190	0.0023	0.0274	0.0011
n=5000	0.0008	-0.0007	0.0025	0.0037	0.0042	0.0009	0.0060	0.0010

Table 2.4: Mean Bias of MLE; $\beta_0 = 2$

	$\alpha_0 = -2.5$		$\alpha_0 = -3$		$\alpha_0 = -3.5$		$\alpha_0 = -4$	
	$p = 19.76\%$		$p = 13.23\%$		$p = 8.58\%$		$p = 5.43\%$	
	Mean bias of α							
	$\hat{\alpha}$	$\tilde{\alpha}$	$\hat{\alpha}$	$\tilde{\alpha}$	$\hat{\alpha}$	$\tilde{\alpha}$	$\hat{\alpha}$	$\tilde{\alpha}$
n=500	-0.0240	-0.0054	-0.0309	-0.0023	-0.0532	-0.0079	-0.0888	-0.0142
n=750	-0.0136	-0.0013	-0.0195	-0.0007	-0.0322	-0.0026	-0.0535	-0.0056
n=1000	-0.0062	0.0030	-0.0120	0.0020	-0.0194	0.0025	-0.0342	0.0010
n=5000	-0.0027	-0.0010	-0.0044	-0.0016	-0.0064	-0.0021	-0.0087	-0.0019
	Mean bias of β							
	$\hat{\beta}$	$\tilde{\beta}$	$\hat{\beta}$	$\tilde{\beta}$	$\hat{\beta}$	$\tilde{\beta}$	$\hat{\beta}$	$\tilde{\beta}$
n=500	0.0232	0.0048	0.0270	0.0021	0.0487	0.0118	0.0720	0.0140
n=750	0.0170	0.0049	0.0192	0.0028	0.0306	0.0066	0.0447	0.0077
n=1000	0.0044	-0.0046	0.0084	-0.0038	0.0146	-0.0031	0.0280	0.0009
n=5000	0.0023	0.0006	0.0035	0.0010	0.0046	0.0011	0.0058	0.0005

Table 2.5: Mean Bias of ATE Estimators; Random Assignment

$\beta_0 = 1$	$\alpha_0 = -2.5$		$\alpha_0 = -3$		$\alpha_0 = -3.5$		$\alpha_0 = -4$	
$\beta_1 = 2$	$\alpha_1 = -1.5$		$\alpha_1 = -2$		$\alpha_1 = -2.5$		$\alpha_1 = -3$	
$x \mid D = 1 \sim N(0,1)$	$p_1 = 28.61\%$		$p_1 = 22.45\%$		$p_1 = 17.17\%$		$p_1 = 12.86\%$	
$x \mid D = 0 \sim N(0,1)$	$p_0 = 10.45\%$		$p_0 = 6.89\%$		$p_0 = 4.41\%$		$p_0 = 2.76\%$	
	$\frac{n_1}{n} = \frac{1}{2}$							
	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$
n=1000	0.0005	0.0005	0.0000	0.0000	0.0005	0.0005	0.0003	0.0003
n=1500	-0.0001	-0.0001	0.0004	0.0004	0.0001	0.0001	-0.0004	-0.0004
n=2000	0.0004	0.0004	0.0003	0.0003	0.0003	0.0003	0.0000	0.0000
	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$
n=1000	-0.0003	0.0006	-0.0009	0.0004	-0.0005	0.0012	-0.0007	0.0007
n=1500	-0.0007	0.0000	-0.0002	0.0004	-0.0005	0.0000	-0.0011	-0.0004
n=2000	0.0000	0.0004	-0.0001	0.0003	-0.0002	0.0002	-0.0004	0.0002
	$\frac{n_1}{n} = \frac{2}{3}$							
	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$
n=1500	0.0000	0.0000	-0.0007	-0.0007	-0.0006	-0.0006	-0.0008	-0.0008
n=2250	0.0000	0.0000	0.0000	0.0000	0.0008	0.0008	0.0005	0.0005
n=3000	-0.0006	-0.0006	-0.0002	-0.0002	0.0002	0.0002	0.0003	0.0003
	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$
n=1500	-0.0011	0.0000	-0.0019	-0.0008	-0.0019	-0.0003	-0.0021	-0.0006
n=2250	-0.0006	0.0002	-0.0007	0.0000	0.0000	0.0009	-0.0004	0.0008
n=3000	-0.0011	-0.0006	-0.0008	-0.0003	-0.0004	0.0002	-0.0004	0.0003

Table 2.6: Mean Bias of ATE Estimators; Random Assignment

$\beta_0 = 1$	$\alpha_0 = -2.5$		$\alpha_0 = -3$		$\alpha_0 = -3.5$		$\alpha_0 = -4$	
$\beta_1 = 4$	$\alpha_1 = -1.5$		$\alpha_1 = -2$		$\alpha_1 = -2.5$		$\alpha_1 = -3$	
$x \mid D = 1 \sim N(0,1)$	$p_1 = 36.57\%$		$p_1 = 32.39\%$		$p_1 = 28.40\%$		$p_1 = 24.68\%$	
$x \mid D = 0 \sim N(0,1)$	$p_0 = 10.45\%$		$p_0 = 6.89\%$		$p_0 = 4.41\%$		$p_0 = 2.76\%$	
	$\frac{n_1}{n} = \frac{1}{2}$							
	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$
n=1000	0.0004	0.0004	0.0002	0.0002	0.0012	0.0012	0.0008	0.0008
n=1500	-0.0003	-0.0003	0.0000	0.0000	0.0007	0.0007	0.0005	0.0005
n=2000	0.0008	0.0008	0.0004	0.0004	-0.0002	-0.0002	0.0003	0.0004
	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$
n=1000	-0.0007	0.0005	-0.0010	0.0006	-0.0002	0.0019	-0.0006	0.0013
n=1500	-0.0011	-0.0003	-0.0009	0.0000	-0.0002	0.0007	-0.0005	0.0005
n=2000	0.0002	0.0007	-0.0002	0.0004	-0.0008	-0.0002	-0.0003	0.0004
	$\frac{n_1}{n} = \frac{2}{3}$							
	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$
n=1500	0.0002	0.0002	-0.0005	-0.0005	-0.0004	-0.0004	-0.0002	-0.0002
n=2250	0.0002	0.0002	0.0000	0.0000	0.0004	0.0004	0.0005	0.0005
n=3000	-0.0003	-0.0003	-0.0005	-0.0005	-0.0002	-0.0002	-0.0002	-0.0002
	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$
n=1500	-0.0010	0.0004	-0.0018	-0.0006	-0.0019	-0.0001	-0.0019	0.0000
n=2250	-0.0005	0.0004	-0.0010	0.0000	-0.0006	0.0005	-0.0006	0.0008
n=3000	-0.0009	-0.0003	-0.0011	-0.0005	-0.0010	-0.0003	-0.0010	-0.0001

Table 2.7: Mean Bias of ATE Estimators; Random Assignment

$\beta_0 = 1$	$\alpha_0 = -2.5$		$\alpha_0 = -3$		$\alpha_0 = -3.5$		$\alpha_0 = -4$	
$\beta_1 = 1$	$\alpha_1 = -1.5$		$\alpha_1 = -2$		$\alpha_1 = -2.5$		$\alpha_1 = -3$	
$x\|D = 0, 1 \sim N(0, 1)$	$p_1 = p_0 = 10.45\%$		$p_1 = p_0 = 6.89\%$		$p_1 = p_0 = 4.41\%$		$p_1 = p_0 = 2.76\%$	
	$\frac{n_1}{n} = \frac{1}{2}$							
	$\widehat{ATE}_1$	$\widetilde{ATE}_1$	$\widehat{ATE}_1$	$\widetilde{ATE}_1$	$\widehat{ATE}_1$	$\widetilde{ATE}_1$	$\widehat{ATE}_1$	$\widetilde{ATE}_1$
n=1000	0.0006	0.0006	0.0007	0.0007	0.0006	0.0006	0.0004	0.0004
n=1500	-0.0009	-0.0009	-0.0003	-0.0003	0.0000	0.0000	0.0000	0.0000
n=2000	-0.0003	-0.0003	-0.0001	-0.0001	0.0000	0.0000	0.0000	0.0000
	$\widehat{ATE}_2$	$\widetilde{ATE}_2$	$\widehat{ATE}_2$	$\widetilde{ATE}_2$	$\widehat{ATE}_2$	$\widetilde{ATE}_2$	$\widehat{ATE}_2$	$\widetilde{ATE}_2$
n=1000	0.0006	0.0008	0.0007	0.0012	0.0006	0.0016	0.0004	0.0006
n=1500	-0.0009	-0.0009	-0.0003	-0.0002	0.0000	0.0000	0.0000	0.0001
n=2000	-0.0003	-0.0004	-0.0001	-0.0002	0.0000	0.0000	0.0000	0.0000
	$\frac{n_1}{n} = \frac{2}{3}$							
	$\widehat{ATE}_1$	$\widetilde{ATE}_1$	$\widehat{ATE}_1$	$\widetilde{ATE}_1$	$\widehat{ATE}_1$	$\widetilde{ATE}_1$	$\widehat{ATE}_1$	$\widetilde{ATE}_1$
n=1500	-0.0005	-0.0005	-0.0005	-0.0005	-0.0002	-0.0002	-0.0003	-0.0003
n=2250	0.0003	0.0003	0.0000	0.0000	0.0004	0.0004	0.0003	0.0003
n=3000	0.0000	0.0000	-0.0004	-0.0004	0.0000	0.0000	0.0001	0.0001
	$\widehat{ATE}_2$	$\widetilde{ATE}_2$	$\widehat{ATE}_2$	$\widetilde{ATE}_2$	$\widehat{ATE}_2$	$\widetilde{ATE}_2$	$\widehat{ATE}_2$	$\widetilde{ATE}_2$
n=1500	-0.0011	-0.0004	-0.0013	-0.0007	-0.0010	-0.0011	-0.0012	-0.0002
n=2250	0.0000	0.0005	-0.0004	0.0002	-0.0001	0.0005	-0.0003	0.0007
n=3000	-0.0003	0.0000	-0.0007	-0.0004	-0.0004	0.0000	-0.0003	0.0002

Table 2.8: Mean Bias of ATE Estimators; Non-random Assignment

$\beta_0 = 1$	$\alpha_0 = -2.5$		$\alpha_0 = -3$		$\alpha_0 = -3.5$		$\alpha_0 = -4$	
$\beta_1 = 1$	$\alpha_1 = -3$		$\alpha_1 = -3$		$\alpha_1 = -3$		$\alpha_1 = -3$	
$x\|D = 1 \sim N(4,1)$	$p_1 = 69.60\%$							
$x\|D = 0 \sim N(0,1)$	$p_0 = 10.45\%$		$p_0 = 6.89\%$		$p_0 = 4.41\%$		$p_0 = 2.76\%$	
	$\frac{n_1}{n} = \frac{1}{2}$							
	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$
n=1000	0.0049	-0.0010	0.0061	-0.0010	0.0068	-0.0004	0.0033	-0.0014
n=1500	0.0037	-0.0002	0.0064	0.0015	0.0079	0.0028	0.0085	0.0051
n=2000	0.0022	-0.0007	0.0027	-0.0010	0.0038	-0.0003	0.0009	-0.0024
	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$
n=1000	0.0061	-0.0012	0.0073	-0.0019	0.0073	-0.0019	0.0025	-0.0030
n=1500	0.0045	-0.0004	0.0072	0.0014	0.0083	0.0027	0.0079	0.0046
n=2000	0.0029	-0.0006	0.0033	-0.0010	0.0040	-0.0006	0.0005	-0.0032
	$\frac{n_1}{n} = \frac{2}{3}$							
	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$
n=1500	0.0056	-0.0008	0.0062	-0.0017	0.0072	-0.0006	0.0047	0.0003
n=2250	0.0053	0.0010	0.0090	0.0034	0.0102	0.0041	0.0091	0.0051
n=3000	0.0037	0.0005	0.0067	0.0024	0.0088	0.0041	0.0068	0.0033
	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$
n=1500	0.0071	-0.0011	0.0077	-0.0014	0.0080	-0.0018	0.0040	-0.0013
n=2250	0.0064	0.0004	0.0100	0.0033	0.0107	0.0038	0.0083	0.0040
n=3000	0.0046	0.0004	0.0075	0.0025	0.0091	0.0041	0.0062	0.0027

Table 2.9: Mean Bias of ATE Estimators; Non-random Assignment

$\beta_0 = 1$	$\alpha_0 = -2$		$\alpha_0 = -3$		$\alpha_0 = -3.5$		$\alpha_0 = -4$	
$\beta_1 = 1$	$\alpha_1 = -2$		$\alpha_1 = -2$		$\alpha_1 = -2$		$\alpha_1 = -2$	
$x\|D = 1 \sim N(2,1)$	$p_1 = 50.05\%$							
$x\|D = 0 \sim N(0,1)$	$p_0 = 10.45\%$		$p_0 = 6.89\%$		$p_0 = 4.41\%$		$p_0 = 2.76\%$	
	$\frac{n_1}{n} = \frac{1}{2}$							
	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$
n=1000	0.0002	-0.0002	0.0000	0.0000	-0.0003	0.0005	-0.0023	-0.0004
n=1500	0.0006	0.0003	0.0014	0.0014	0.0014	0.0020	0.0007	0.0021
n=2000	0.0000	-0.0002	-0.0003	-0.0003	-0.0004	0.0000	-0.0014	-0.0005
	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$
n=1000	0.0006	-0.0002	-0.0003	0.0002	-0.0013	0.0008	-0.0041	-0.0007
n=1500	0.0009	0.0003	0.0012	0.0014	0.0007	0.0019	-0.0005	0.0023
n=2000	0.0002	-0.0002	-0.0004	-0.0003	-0.0008	0.0002	-0.0023	0.0000
	$\frac{n_1}{n} = \frac{2}{3}$							
	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$
n=1500	-0.0008	-0.0010	-0.0023	-0.0018	-0.0030	-0.0016	-0.0045	-0.0017
n=2250	0.0010	0.0089	0.0017	0.0020	0.0013	0.0023	-0.0002	0.0018
n=3000	0.0000	0.0000	0.0006	0.0008	0.0005	0.0013	-0.0008	0.0007
	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$
n=1500	-0.0005	-0.0010	-0.0027	-0.0018	-0.0044	-0.0014	-0.0069	-0.0012
n=2250	0.0012	0.0009	0.0014	0.0020	0.0003	0.0022	-0.0020	0.0017
n=3000	0.0002	0.0000	0.0004	0.0008	-0.0002	0.0013	-0.0021	0.0013

Table 2.10: Mean Bias of ATE Estimators; Non-random Assignment

$\beta_0 = 1$	$\alpha_0 = -2.5$		$\alpha_0 = -3$		$\alpha_0 = -3.5$		$\alpha_0 = -4$	
$\beta_1 = 1$	$\alpha_1 = -2.5$		$\alpha_1 = -3$		$\alpha_1 = -3.5$		$\alpha_1 = -4$	
$x\|D = 1 \sim N(4,1)$	$p_1 = 77.81\%$		$p_1 = 69.60\%$		$p_1 = 60.19\%$		$p_1 = 50.05\%$	
$x\|D = 0 \sim N(0,1)$	$p_0 = 10.45\%$		$p_0 = 6.89\%$		$p_0 = 4.41\%$		$p_0 = 2.76\%$	
	$\frac{n_1}{n} = \frac{1}{2}$							
	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$
n=1000	0.0044	-0.0021	0.0061	-0.0010	0.0062	-0.0005	0.0027	-0.0013
n=1500	0.0041	-0.0003	0.0064	0.0015	0.0083	0.0034	0.0090	0.0061
n=2000	0.0025	-0.0008	0.0027	-0.0010	0.0040	0.0002	0.0010	-0.0020
	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$
n=1000	0.0058	-0.0023	0.0073	-0.0019	0.0066	-0.0019	0.0018	-0.0028
n=1500	0.0050	-0.0005	0.0072	0.0014	0.0086	0.0033	0.0083	0.0056
n=2000	0.0033	-0.0007	0.0033	-0.0010	0.0043	0.0000	0.0006	-0.0027
	$\frac{n_1}{n} = \frac{2}{3}$							
	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$
n=1500	0.0047	-0.0019	0.0062	-0.0017	0.0073	-0.0004	0.0042	0.0000
n=2250	0.0051	0.0006	0.0090	0.0034	0.0104	0.0045	0.0092	0.0054
n=3000	0.0037	0.0003	0.0067	0.0024	0.0085	0.0039	0.0066	0.0032
	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$
n=1500	0.0062	-0.0023	0.0077	-0.0014	0.0081	-0.0016	0.0036	-0.0015
n=2250	0.0062	0.0000	0.0100	0.0033	0.0109	0.0041	0.0084	0.0043
n=3000	0.0045	0.0002	0.0075	0.0025	0.0089	0.0038	0.0060	0.0026

Table 2.11: Mean Bias of ATE Estimators; Non-random Assignment

$\beta_0 = 1$	$\alpha_0 = -2.5$		$\alpha_0 = -3$		$\alpha_0 = -3.5$		$\alpha_0 = -4$	
$\beta_1 = 1$	$\alpha_1 = -2.5$		$\alpha_1 = -3$		$\alpha_1 = -3.5$		$\alpha_1 = -4$	
$x\|D = 1 \sim N(2,1)$	$p_1 = 39.93\%$		$p_1 = 30.38\%$		$p_1 = 22.13\%$		$p_1 = 15.50\%$	
$x\|D = 0 \sim N(0,1)$	$p_0 = 10.45\%$		$p_0 = 6.89\%$		$p_0 = 4.41\%$		$p_0 = 2.76\%$	
	$\frac{n_1}{n} = \frac{1}{2}$							
	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$
n=1000	-0.0001	-0.0005	-0.0006	-0.0005	-0.0009	0.0000	-0.0026	-0.0006
n=1500	0.0003	0.0000	0.0010	0.0010	0.0008	0.0014	-0.0002	0.0012
n=2000	0.0001	-0.0001	-0.0002	-0.0002	-0.0006	-0.0001	-0.0017	-0.0007
	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$
n=1000	0.0003	-0.0004	-0.0007	-0.0003	-0.0017	0.0003	-0.0042	-0.0010
n=1500	0.0006	0.0000	0.0009	0.0010	0.0002	0.0014	-0.0013	0.0014
n=2000	0.0003	0.0000	-0.0002	-0.0001	-0.0010	0.0000	-0.0025	-0.0003
	$\frac{n_1}{n} = \frac{2}{3}$							
	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$
n=1500	-0.0008	-0.0010	-0.0024	-0.0020	-0.0035	-0.0020	-0.0046	-0.0017
n=2250	0.0009	0.0008	0.0014	0.0017	0.0012	0.0022	-0.0004	0.0016
n=3000	-0.0002	-0.0003	0.0005	0.0007	0.0003	0.0010	-0.0007	0.0007
	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$
n=1500	-0.0005	-0.0010	-0.0027	-0.0019	-0.0046	-0.0019	-0.0068	-0.0013
n=2250	0.0012	0.0008	0.0011	0.0017	0.0003	0.0021	-0.0020	0.0015
n=3000	0.0000	-0.0003	0.0003	0.0007	-0.0003	0.0011	-0.0019	0.0014

Table 2.12: Mean Bias of ATE Estimators; Random Assignment

$\beta_0 = (1,1)$	$\alpha_0 = -2.5$		$\alpha_0 = -3$		$\alpha_0 = -3.5$		$\alpha_0 = -4$	
$\beta_1 = (2,2)$	$\alpha_1 = -1.5$		$\alpha_1 = -2$		$\alpha_1 = -2.5$		$\alpha_1 = -3$	
$x \mid D = 1 \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}\right)$				$x \mid D = 0 \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}\right)$				
				$\frac{n_1}{n} = \frac{1}{2}$				
	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$
n=1000	-0.0003	0.0006	-0.0003	0.0008	0.0006	0.0019	0.0012	0.0025
n=1500	-0.0003	0.0004	0.0006	0.0013	0.0000	0.0009	0.0004	0.0012
n=2000	-0.0021	-0.0016	-0.0012	-0.0006	-0.0004	0.0003	-0.0003	0.0004
	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$
n=1000	-0.0006	0.0006	-0.0005	0.0008	0.0005	0.0019	0.0011	0.0025
n=1500	-0.0005	0.0004	0.0004	0.0013	0.0000	0.0009	0.0003	0.0012
n=2000	-0.0022	-0.0016	-0.0013	-0.0006	-0.0004	0.0003	-0.0003	0.0004
				$\frac{n_1}{n} = \frac{2}{3}$				
	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$
n=1500	-0.0006	0.0005	-0.0001	0.0011	-0.0003	0.0011	0.0001	0.0016
n=2250	-0.0003	0.0004	0.0000	0.0009	0.0002	0.0011	0.0003	0.0012
n=3000	-0.0003	0.0002	0.0000	0.0006	0.0002	0.0009	0.0003	0.0011
	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$
n=1500	-0.0011	0.0005	-0.0007	0.0011	-0.0008	0.0010	-0.0004	0.0015
n=2250	-0.0007	0.0004	-0.0003	0.0009	-0.0002	0.0011	0.0000	0.0012
n=3000	-0.0006	0.0002	-0.0003	0.0006	0.0000	0.0009	0.0000	0.0011

Table 2.13: Mean Bias of ATE Estimators; Random Assignment

$\beta_0 = (1,1)$	$\alpha_0 = -2.5$		$\alpha_0 = -3$		$\alpha_0 = -3.5$		$\alpha_0 = -4$	
$\beta_1 = (4,4)$	$\alpha_1 = -1.5$		$\alpha_1 = -2$		$\alpha_1 = -2.5$		$\alpha_1 = -3$	
$x \mid D=1 \sim N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\frac{1}{2} & 0\\ 0 & \frac{1}{2}\end{bmatrix}\right)$				$x \mid D=0 \sim N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\frac{1}{2} & 0\\ 0 & \frac{1}{2}\end{bmatrix}\right)$				
	$\frac{n_1}{n} = \frac{1}{2}$							
	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$
n=1000	0.0000	0.0013	0.0009	0.0025	0.0006	0.0024	0.0000	0.0021
n=1500	-0.0001	0.0008	-0.0002	0.0009	0.0006	0.0019	0.0000	0.0014
n=2000	-0.0009	-0.0003	-0.0008	0.0000	-0.0012	-0.0003	-0.0007	0.0003
	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$
n=1000	-0.0003	0.0013	0.0006	0.0025	0.0004	0.0024	0.0000	0.0021
n=1500	-0.0003	0.0008	-0.0003	0.0009	0.0005	0.0019	0.0000	0.0014
n=2000	-0.0011	-0.0003	-0.0010	0.0000	-0.0013	-0.0003	-0.0007	0.0003
	$\frac{n_1}{n} = \frac{2}{3}$							
	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$
n=1500	0.0000	0.0013	0.0000	0.0014	-0.0005	0.0012	0.0000	0.0017
n=2250	0.0005	0.0013	0.0011	0.0021	0.0008	0.0019	0.0002	0.0014
n=3000	0.0002	0.0008	0.0003	0.0011	0.0002	0.0010	0.0001	0.0011
	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$
n=1500	-0.0005	0.0013	-0.0007	0.0013	-0.0010	0.0012	-0.0006	0.0017
n=2250	0.0001	0.0013	0.0007	0.0021	0.0004	0.0019	-0.0002	0.0014
n=3000	-0.0001	0.0008	0.0000	0.0011	-0.0001	0.0010	-0.0001	0.0011

Table 2.14: Mean Bias of ATE Estimators; Random Assignment

$\beta_0 = (1,1)$	$\alpha_0 = -2.5$		$\alpha_0 = -3$		$\alpha_0 = -3.5$		$\alpha_0 = -4$	
$\beta_1 = (1,1)$	$\alpha_1 = -1.5$		$\alpha_1 = -2$		$\alpha_1 = -2.5$		$\alpha_1 = -3$	
$x \mid D = 1 \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}\right)$					$x \mid D = 0 \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}\right)$			
	$\frac{n_1}{n} = \frac{1}{2}$							
	$\widehat{ATE}_1$	$\widetilde{ATE}_1$	$\widehat{ATE}_1$	$\widetilde{ATE}_1$	$\widehat{ATE}_1$	$\widetilde{ATE}_1$	$\widehat{ATE}_1$	$\widetilde{ATE}_1$
n=1000	-0.0002	0.0002	0.0003	0.0007	0.0012	0.0015	0.0007	0.0009
n=1500	0.0000	0.0003	0.0000	0.0003	0.0009	0.0011	0.0004	0.0006
n=2000	-0.0012	-0.0009	-0.0008	-0.0006	-0.0009	-0.0007	-0.0004	-0.0003
	$\widehat{ATE}_2$	$\widetilde{ATE}_2$	$\widehat{ATE}_2$	$\widetilde{ATE}_2$	$\widehat{ATE}_2$	$\widetilde{ATE}_2$	$\widehat{ATE}_2$	$\widetilde{ATE}_2$
n=1000	-0.0005	0.0002	0.0002	0.0007	0.0010	0.0015	0.0005	0.0009
n=1500	-0.0002	0.0003	0.0000	0.0003	0.0008	0.0011	0.0003	0.0006
n=2000	-0.0013	-0.0009	-0.0009	-0.0006	-0.0009	-0.0007	-0.0005	-0.0003
	$\frac{n_1}{n} = \frac{2}{3}$							
	$\widehat{ATE}_1$	$\widetilde{ATE}_1$	$\widehat{ATE}_1$	$\widetilde{ATE}_1$	$\widehat{ATE}_1$	$\widetilde{ATE}_1$	$\widehat{ATE}_1$	$\widetilde{ATE}_1$
n=1500	0.0000	0.0008	0.0006	0.0014	0.0003	0.0011	0.0004	0.0014
n=2250	0.0000	0.0004	0.0001	0.0007	0.0004	0.0010	0.0004	0.0011
n=3000	-0.0002	0.0002	-0.0003	0.0000	0.0003	0.0007	0.0002	0.0007
	$\widehat{ATE}_2$	$\widetilde{ATE}_2$	$\widehat{ATE}_2$	$\widetilde{ATE}_2$	$\widehat{ATE}_2$	$\widetilde{ATE}_2$	$\widehat{ATE}_2$	$\widetilde{ATE}_2$
n=1500	-0.0005	0.0008	0.0000	0.0014	-0.0003	0.0011	-0.0001	0.0013
n=2250	-0.0005	0.0004	-0.0002	0.0007	0.0000	0.0010	0.0000	0.0010
n=3000	-0.0004	0.0002	-0.0006	0.0000	0.0000	0.0007	0.0000	0.0007

Table 2.15: Mean Bias of ATE Estimators; Non-random Assignment

$\beta_0 = (1,1)$	$\alpha_0 = -2.5$		$\alpha_0 = -3$		$\alpha_0 = -3.5$		$\alpha_0 = -4$	
$\beta_1 = (1,1)$	$\alpha_1 = -3$		$\alpha_1 = -3$		$\alpha_1 = -3$		$\alpha_1 = -3$	

$$x \mid D = 1 \sim N\left(\begin{bmatrix} 2 \\ 2 \end{bmatrix}, \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}\right) \qquad x \mid D = 0 \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}\right)$$

				$\frac{n_1}{n} = \frac{1}{2}$				
	$\widehat{ATE}_1$	$\widetilde{ATE}_1$	$\widehat{ATE}_1$	$\widetilde{ATE}_1$	$\widehat{ATE}_1$	$\widetilde{ATE}_1$	$\widehat{ATE}_1$	$\widetilde{ATE}_1$
n=1000	0.0032	-0.0007	0.0035	-0.0003	0.0038	0.0002	0.0038	0.0003
n=1500	0.0013	-0.0013	0.0011	-0.0012	0.0014	-0.0009	0.0013	-0.0008
n=2000	0.0006	-0.0013	0.0009	-0.0009	0.0010	-0.0007	0.0012	-0.0005
	$\widehat{ATE}_2$	$\widetilde{ATE}_2$	$\widehat{ATE}_2$	$\widetilde{ATE}_2$	$\widehat{ATE}_2$	$\widetilde{ATE}_2$	$\widehat{ATE}_2$	$\widetilde{ATE}_2$
n=1000	0.0018	-0.0007	0.0020	-0.0003	0.0022	0.0002	0.0022	0.0003
n=1500	0.0004	-0.0012	0.0002	-0.0012	0.0004	-0.0009	0.0004	-0.0008
n=2000	0.0000	-0.0013	0.0002	-0.0009	0.0003	-0.0007	0.0004	-0.0005
				$\frac{n_1}{n} = \frac{2}{3}$				
	$\widehat{ATE}_1$	$\widetilde{ATE}_1$	$\widehat{ATE}_1$	$\widetilde{ATE}_1$	$\widehat{ATE}_1$	$\widetilde{ATE}_1$	$\widehat{ATE}_1$	$\widetilde{ATE}_1$
n=1500	0.0014	0.0000	0.0013	0.0001	0.0010	0.0000	0.0012	0.0004
n=2250	0.0003	-0.0006	0.0009	0.0002	0.0011	0.0005	0.0008	0.0003
n=3000	0.0002	-0.0005	0.0004	-0.0002	0.0005	0.0000	0.0004	0.0000
	$\widehat{ATE}_2$	$\widetilde{ATE}_2$	$\widehat{ATE}_2$	$\widetilde{ATE}_2$	$\widehat{ATE}_2$	$\widetilde{ATE}_2$	$\widehat{ATE}_2$	$\widetilde{ATE}_2$
n=1500	0.0003	0.0000	0.0001	0.0001	-0.0002	0.0000	0.0000	0.0003
n=2250	-0.0004	-0.0006	0.0002	0.0002	0.0003	0.0005	0.0000	0.0003
n=3000	-0.0004	-0.0005	-0.0002	-0.0002	-0.0001	0.0000	-0.0002	0.0000

Table 2.16: Mean Bias of ATE Estimators; Non-random Assignment

$\beta_0 = (1,1)$	$\alpha_0 = -2$		$\alpha_0 = -3$		$\alpha_0 = -3.5$		$\alpha_0 = -4$	
$\beta_1 = (1,1)$	$\alpha_1 = -2$		$\alpha_1 = -2$		$\alpha_1 = -2$		$\alpha_1 = -2$	
$x \mid D = 1 \sim N\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}\right)$				$x \mid D = 0 \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}\right)$				
	$\frac{n_1}{n} = \frac{1}{2}$							
	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$
n=1000	0.0017	0.0009	0.0006	0.0003	0.0009	0.0008	0.0009	0.0009
n=1500	-0.0017	-0.0021	-0.0015	-0.0017	-0.0011	-0.0013	-0.0012	-0.0012
n=2000	-0.0014	-0.0018	-0.0008	-0.0009	-0.0006	-0.0007	-0.0005	-0.0005
	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$
n=1000	0.0018	0.0009	0.0005	0.0003	0.0007	0.0008	0.0007	0.0009
n=1500	-0.0016	-0.0021	-0.0016	-0.0017	-0.0013	-0.0013	-0.0014	-0.0012
n=2000	-0.0014	-0.0018	-0.0008	-0.0009	-0.0007	-0.0007	-0.0006	-0.0005
	$\frac{n_1}{n} = \frac{2}{3}$							
	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$
n=1500	0.0009	0.0009	0.0008	0.0013	0.0006	0.0012	0.0008	0.0015
n=2250	0.0001	0.0001	0.0008	0.0011	0.0010	0.0014	0.0007	0.0012
n=3000	0.0000	0.0001	0.0004	0.0007	0.0006	0.0009	0.0004	0.0008
	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$
n=1500	0.0005	0.0009	0.0003	0.0013	0.0000	0.0012	0.0001	0.0015
n=2250	0.0000	0.0001	0.0005	0.0011	0.0006	0.0014	0.0003	0.0012
n=3000	0.0000	0.0001	0.0002	0.0007	0.0003	0.0009	0.0001	0.0008

Table 2.17: Mean Bias of ATE Estimators; Non-random Assignment

$\beta_0 = (1,1)$	$\alpha_0 = -2.5$		$\alpha_0 = -3$		$\alpha_0 = -3.5$		$\alpha_0 = -4$	
$\beta_1 = (1,1)$	$\alpha_1 = -2.5$		$\alpha_1 = -3$		$\alpha_1 = -3.5$		$\alpha_1 = -4$	
$x \mid D = 1 \sim N\left(\begin{bmatrix} 2 \\ 2 \end{bmatrix}, \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}\right)$				$x \mid D = 0 \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}\right)$				
	$\frac{n_1}{n} = \frac{1}{2}$							
	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$
n=1000	0.0052	-0.0024	0.0035	-0.0003	0.0014	0.0000	0.0013	0.0012
n=1500	0.0034	-0.0015	0.0011	-0.0013	0.0005	-0.0004	0.0003	0.0003
n=2000	0.0021	-0.0016	0.0009	-0.0009	0.0003	-0.0004	0.0002	0.0002
	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$
n=1000	0.0023	-0.0023	0.0020	-0.0003	0.0006	0.0000	0.0007	0.0011
n=1500	0.0015	-0.0015	0.0002	-0.0013	0.0000	-0.0004	0.0000	0.0003
n=2000	0.0007	-0.0015	0.0002	-0.0009	-0.0001	-0.0004	0.0000	0.0002
	$\frac{n_1}{n} = \frac{2}{3}$							
	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$
n=1500	0.0017	-0.0014	0.0013	0.0001	0.0010	0.0010	0.0011	0.0018
n=2250	0.0010	-0.0011	0.0009	0.0002	0.0012	0.0012	0.0009	0.0014
n=3000	0.0009	-0.0006	0.0004	-0.0002	0.0009	0.0009	0.0005	0.0008
	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$
n=1500	0.0000	-0.0014	0.0001	0.0001	0.0000	0.0010	0.0002	0.0018
n=2250	-0.0002	-0.0011	0.0002	0.0002	0.0006	0.0012	0.0004	0.0014
n=3000	0.0000	-0.0006	-0.0002	-0.0002	0.0004	0.0009	0.0000	0.0008

Table 2.18: Mean Bias of ATE Estimators; Non-random Assignment

$\beta_0=(1,1)$	$\alpha_0=-2.5$		$\alpha_0=-3$		$\alpha_0=-3.5$		$\alpha_0=-4$	
$\beta_1=(1,1)$	$\alpha_1=-2.5$		$\alpha_1=-3$		$\alpha_1=-3.5$		$\alpha_1=-4$	
$x \mid D=1 \sim N\left(\begin{bmatrix}1\\1\end{bmatrix},\begin{bmatrix}\frac{1}{2}&0\\0&\frac{1}{2}\end{bmatrix}\right)$				$x \mid D=0 \sim N\left(\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}\frac{1}{2}&0\\0&\frac{1}{2}\end{bmatrix}\right)$				
	$\frac{n_1}{n}=\frac{1}{2}$							
	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$
n=1000	0.0004	0.0003	0.0003	0.0008	0.0006	0.0014	0.0007	0.0018
n=1500	-0.0002	-0.0002	0.0000	0.0004	0.0005	0.0010	0.0002	0.0009
n=2000	-0.0002	-0.0002	-0.0005	-0.0003	-0.0005	-0.0001	-0.0003	0.0003
	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$
n=1000	0.0005	0.0003	0.0005	0.0007	0.0007	0.0014	0.0007	0.0018
n=1500	0.0000	-0.0002	0.0002	0.0004	0.0005	0.0010	0.0002	0.0009
n=2000	-0.0001	-0.0002	-0.0005	-0.0003	-0.0005	-0.0001	-0.0002	0.0003
	$\frac{n_1}{n}=\frac{2}{3}$							
	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$	$\widehat{ATE_1}$	$\widetilde{ATE_1}$
n=1500	0.0015	0.0020	0.0007	0.0015	0.0002	0.0013	0.0005	0.0018
n=2250	0.0006	0.0009	0.0011	0.0017	0.0011	0.0019	0.0008	0.0017
n=3000	0.0006	0.0009	0.0006	0.0010	0.0005	0.0011	0.0004	0.0011
	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$	$\widehat{ATE_2}$	$\widetilde{ATE_2}$
n=1500	0.0011	0.0020	0.0003	0.0015	-0.0003	0.0013	0.0000	0.0018
n=2250	0.0003	0.0009	0.0009	0.0017	0.0008	0.0019	0.0005	0.0017
n=3000	0.0005	0.0009	0.0004	0.0010	0.0003	0.0011	0.0002	0.0011

2.8 Conclusion

If the treatment assignment is unconfounded, then the MLE of the ATE in the logit model for binary outcomes is biased both because the MLE of the parameters of the logit model are biased and the curvature of the logit model contributes to the bias of the ATE estimator. The bias is larger if the event has a low probability, the case we focus on in this chapter.

The logit ATE estimator is unbiased if the treatment assignment is random. This is obvious if there are no covariates, but it is also true with covariates.

We derive the second order bias of the logit ATE estimator. We also propose bias-corrected estimators of the ATE.

Simulation experiments show that it is not sufficient to bias correct the MLE of the logit parameters, that the bias of the logit ATE estimator is moderately large, and that the bias-corrected estimators remove most of the bias.

Finally we propose a variation on the logit model with parameters that are elasticities. We also propose a computational trick that avoids numerical instability in the case of estimation for rare events.

2.9 Appendix

2.9.1 Second Order Bias of Logit MLE

We derive the second order bias of the logit MLE. For notational simplicity, we will omit the i subscript whenever obvious.

2.9.2 Second Order Expansion of Generic MLE

The MLE $\widehat{\theta}$ maximizes the joint log likelihood $\sum_{i=1}^{n} \log f(z_i; c)$, and satisfies the first order condition $0 = \frac{1}{n} \sum_{i=1}^{n} v\left(z_i; \widehat{\theta}\right)$. Let F denote the collection of (marginal) distribution functions of z. Let $\hat{F}$ denote the empirical distribution function. Define $F(\epsilon) \equiv F + \epsilon\sqrt{n}\left(\hat{F} - F\right)$ for $\epsilon \in \left[0, n^{-1/2}\right]$. For each fixed ϵ, let $\theta(\epsilon)$ be the solution to the estimating equation

$$0 = \int v\left[\cdot; \theta(\epsilon)\right] dF(\epsilon). \tag{2.22}$$

Note that (2.22) is equivalent to $0 = \frac{1}{n} \sum_{i=1}^{n} v\left(z_i; \theta\left(n^{-1/2}\right)\right)$ when evaluated at $\epsilon = n^{-1/2}$, so we can see that $\theta\left(n^{-1/2}\right) = \widehat{\theta}$. By a Taylor series expansion, we have

$$\widehat{\theta} - \theta_0 = \theta\left(\frac{1}{\sqrt{n}}\right) - \theta(0) = \frac{1}{\sqrt{n}}\theta^{\epsilon}(0) + \frac{1}{2}\left(\frac{1}{\sqrt{n}}\right)^2 \theta^{\epsilon\epsilon}(0) + \frac{1}{6}\left(\frac{1}{\sqrt{n}}\right)^3 \theta^{\epsilon\epsilon\epsilon}(\tilde{\epsilon}), \tag{2.23}$$

where $\theta^{\epsilon}(\epsilon) \equiv d\theta(\epsilon)/d\epsilon$, $\theta^{\epsilon\epsilon}(\epsilon) \equiv d^2\theta(\epsilon)/d\epsilon^2$, ..., and $\tilde{\epsilon}$ is somewhere in between 0 and $n^{-1/2}$.

Let

$$h(\cdot, \epsilon) \equiv v\left[\cdot; \theta(\epsilon)\right]. \tag{2.24}$$

The first order condition may be written as

$$0 = \int h(\cdot, \epsilon)\, dF(\epsilon). \tag{2.25}$$

Differentiating repeatedly with respect to ϵ, we obtain

$$0 = \int \frac{dh(\cdot, \epsilon)}{d\epsilon} dF(\epsilon) + \int h(\cdot, \epsilon)\, d\Delta_n, \tag{2.26}$$

$$0 = \int \frac{d^2 h(\cdot, \epsilon)}{d\epsilon^2} dF(\epsilon) + 2\int \frac{dh(\cdot, \epsilon)}{d\epsilon} d\Delta_n, \tag{2.27}$$

$$0 = \int \frac{d^3 h(\cdot, \epsilon)}{d\epsilon^3} dF(\epsilon) + 3\int \frac{d^2 h(\cdot, \epsilon)}{d\epsilon^2} d\Delta_n, \tag{2.28}$$

where $\Delta_n \equiv \sqrt{n}\left(\hat{F} - F\right)$. We will ignore the third order term, which can be justified under the type of regularity conditions discussed in Hahn and Newey (2004), and find the analytic expression for the second order bias.

Rewrite (2.26) as

$$0 = \left(\int v^{\theta}\left[\cdot;\theta\left(\epsilon\right)\right] dF\left(\epsilon\right)\right)\theta^{\epsilon}\left(\epsilon\right) + \int v\left[\cdot;\theta\left(\epsilon\right)\right] d\Delta_{n},$$

where

$$v^{\theta} \equiv \frac{\partial v\left[\cdot;\theta\left(\epsilon\right)\right]}{\partial\theta'}.$$

Evaluating it at $\epsilon = 0$, and noting that $E\left[v_i\right] = 0$, we obtain

$$0 = \left(\int v^{\theta}\left[\cdot;\theta\left(0\right)\right] dF\right)\theta^{\epsilon}\left(0\right) + \int v\left[\cdot;\theta\left(0\right)\right]\Delta_{n},$$

so

$$\theta^{\epsilon}\left(0\right) = -\left(E\left[v^{\theta}\right]\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} v_i\right). \tag{2.29}$$

Recall that $\dim\left(\theta\right) = K$, and write

$$v\left[\cdot;\theta\left(\epsilon\right)\right] = \begin{bmatrix} v_1\left[\cdot;\theta\left(\epsilon\right)\right] \\ \vdots \\ v_K\left[\cdot;\theta\left(\epsilon\right)\right] \end{bmatrix}.$$

We can then write

$$\frac{d^2 h\left(\cdot,\epsilon\right)}{d\epsilon^2} = \begin{bmatrix} d^2 h_1\left(\cdot,\epsilon\right)/\,d\epsilon^2 \\ \vdots \\ d^2 h_K\left(\cdot,\epsilon\right)/\,d\epsilon^2 \end{bmatrix} = \begin{bmatrix} \theta^{\epsilon}\left(\epsilon\right)' \frac{\partial^2 v_1\left[\cdot;\theta\left(\epsilon\right)\right]}{\partial\theta\partial\theta'}\theta^{\epsilon}\left(\epsilon\right) + \frac{\partial v_1\left[\cdot;\theta\left(\epsilon\right)\right]}{\partial\theta'}\theta^{\epsilon\epsilon}\left(\epsilon\right) \\ \vdots \\ \theta^{\epsilon}\left(\epsilon\right)' \frac{\partial^2 v_K\left[\cdot;\theta\left(\epsilon\right)\right]}{\partial\theta\partial\theta'}\theta^{\epsilon}\left(\epsilon\right) + \frac{\partial v_K\left[\cdot;\theta\left(\epsilon\right)\right]}{\partial\theta'}\theta^{\epsilon\epsilon}\left(\epsilon\right) \end{bmatrix},$$

so we can rewrite (2.27) as

$$0 = \begin{bmatrix} \theta^{\epsilon}\left(\epsilon\right)'\left(\int \frac{\partial^2 v_1\left[\cdot;\theta\left(\epsilon\right)\right]}{\partial\theta\partial\theta'} dF\left(\epsilon\right)\right)\theta^{\epsilon}\left(\epsilon\right) \\ \vdots \\ \theta^{\epsilon}\left(\epsilon\right)'\left(\int \frac{\partial^2 v_K\left[\cdot;\theta\left(\epsilon\right)\right]}{\partial\theta\partial\theta'} dF\left(\epsilon\right)\right)\theta^{\epsilon}\left(\epsilon\right) \end{bmatrix} + \begin{bmatrix} \int \frac{\partial v_1\left[\cdot;\theta\left(\epsilon\right)\right]}{\partial\theta'} dF\left(\epsilon\right) \\ \vdots \\ \int \frac{\partial v_K\left[\cdot;\theta\left(\epsilon\right)\right]}{\partial\theta'} dF\left(\epsilon\right) \end{bmatrix}\theta^{\epsilon\epsilon}\left(\epsilon\right) + 2\begin{bmatrix} \int \frac{\partial v_1\left[\cdot;\theta\left(\epsilon\right)\right]}{\partial\theta'} d\Delta_n \\ \vdots \\ \int \frac{\partial v_K\left[\cdot;\theta\left(\epsilon\right)\right]}{\partial\theta'} d\Delta_n \end{bmatrix}\theta^{\epsilon}\left(\epsilon\right).$$

Evaluating it at $\epsilon = 0$, we obtain

$$0 = \begin{bmatrix} \theta^{\epsilon}\left(0\right)'\left(E\left[\frac{\partial^2 v_1}{\partial\theta\partial\theta'}\right]\right)\theta^{\epsilon}\left(0\right) \\ \vdots \\ \theta^{\epsilon}\left(0\right)'\left(E\left[\frac{\partial^2 v_K}{\partial\theta\partial\theta'}\right]\right)\theta^{\epsilon}\left(0\right) \end{bmatrix} + \begin{bmatrix} E\left[\frac{\partial v_1}{\partial\theta'}\right] \\ \vdots \\ E\left[\frac{\partial v_K}{\partial\theta'}\right] \end{bmatrix}\theta^{\epsilon\epsilon}\left(0\right) + 2\begin{bmatrix} \int \frac{\partial v_1\left[\cdot;\theta\left(\epsilon\right)\right]}{\partial\theta'} d\Delta_n \\ \vdots \\ \int \frac{\partial v_K\left[\cdot;\theta\left(\epsilon\right)\right]}{\partial\theta'} d\Delta_n \end{bmatrix}\theta^{\epsilon}\left(0\right),$$

so

$$\frac{1}{2}\theta^{\epsilon\epsilon}(0) = -\frac{1}{2}\left(E\left[v^{\theta}\right]\right)^{-1}\begin{bmatrix} \theta^{\epsilon}(0)'\left(E\left[\frac{\partial^2 v_1}{\partial\theta\partial\theta'}\right]\right)\theta^{\epsilon}(0) \\ \vdots \\ \theta^{\epsilon}(0)'\left(E\left[\frac{\partial^2 v_K}{\partial\theta\partial\theta'}\right]\right)\theta^{\epsilon}(0) \end{bmatrix} - \left(E\left[v^{\theta}\right]\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(v_i^{\theta} - E\left[v_i^{\theta}\right]\right)\right)\theta^{\epsilon}(0).$$

2.9.3 Second order bias

We now calculate $E\left[\frac{1}{2}\theta^{\epsilon\epsilon}(0)\right]$. We first compute the expectation of

$$E\left[v^{\theta}\right]\left(\frac{1}{2}\theta^{\epsilon\epsilon}(0)\right) = -\frac{1}{2}\begin{bmatrix} \theta^{\epsilon}(0)'\left(E\left[\frac{\partial^2 v_1}{\partial\theta\partial\theta'}\right]\right)\theta^{\epsilon}(0) \\ \vdots \\ \theta^{\epsilon}(0)'\left(E\left[\frac{\partial^2 v_K}{\partial\theta\partial\theta'}\right]\right)\theta^{\epsilon}(0) \end{bmatrix} - \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(v_i^{\theta} - E\left[v_i^{\theta}\right]\right)\right)\theta^{\epsilon}(0).$$

The kth component of the first term has an expectation equal to

$$\begin{aligned} -\frac{1}{2}E\left[\theta^{\epsilon}(0)'\left(E\left[\frac{\partial^2 v_k}{\partial\theta\partial\theta'}\right]\right)\theta^{\epsilon}(0)\right] &= -\frac{1}{2}E\left[\operatorname{trace}\left(\theta^{\epsilon}(0)'\left(E\left[\frac{\partial^2 v_k}{\partial\theta\partial\theta'}\right]\right)\theta^{\epsilon}(0)\right)\right] \\ &= -\frac{1}{2}E\left[\operatorname{trace}\left(E\left[\frac{\partial^2 v_k}{\partial\theta\partial\theta'}\right]\theta^{\epsilon}(0)\,\theta^{\epsilon}(0)'\right)\right] \\ &= -\frac{1}{2}\operatorname{trace}\left(E\left[\frac{\partial^2 v_k}{\partial\theta\partial\theta'}\right]E\left[\theta^{\epsilon}(0)\,\theta^{\epsilon}(0)'\right]\right). \end{aligned}$$

Because

$$\begin{aligned} E\left[\theta^{\epsilon}(0)\,\theta^{\epsilon}(0)'\right] &= E\left[\left(-\left(E\left[v^{\theta}\right]\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}v_i\right)\right)\left(-\left(E\left[v^{\theta}\right]\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}v_i\right)\right)'\right] \\ &= \left(E\left[v^{\theta}\right]\right)^{-1}E\left[vv'\right]\left(E\left[v^{\theta}\right]\right)^{-1} \\ &= -\left(E\left[v^{\theta}\right]\right)^{-1}, \end{aligned}$$

where we used the information equality, we can write

$$-\frac{1}{2}E\left[\theta^{\epsilon}(0)'\left(E\left[\frac{\partial^2 v_k}{\partial\theta\partial\theta'}\right]\right)\theta^{\epsilon}(0)\right] = \frac{1}{2}\operatorname{trace}\left(E\left[\frac{\partial^2 v_k}{\partial\theta\partial\theta'}\right]\left(E\left[v^{\theta}\right]\right)^{-1}\right).$$

The kth component of the second term has an expectation equal to

$$
\begin{aligned}
&-E\left[\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(v_{k,i}^{\theta}-E\left[v_{k,i}^{\theta}\right]\right)\right)\theta^{\epsilon}(0)\right]\\
&=-E\left[\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(v_{k,i}^{\theta}-E\left[v_{k,i}^{\theta}\right]\right)\right)\left(-\left(E\left[v^{\theta}\right]\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}v_i\right)\right)\right]\\
&=E\left[\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{\partial v_{k,i}}{\partial\theta'}-E\left[\frac{\partial v_{k,i}}{\partial\theta'}\right]\right)\right)\left(\left(E\left[v^{\theta}\right]\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}v_i\right)\right)\right]\\
&=E\left[\operatorname{trace}\left\{\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{\partial v_{k,i}}{\partial\theta'}-E\left[\frac{\partial v_{k,i}}{\partial\theta'}\right]\right)\right)\left(\left(E\left[v^{\theta}\right]\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}v_i\right)\right)\right\}\right]\\
&=E\left[\operatorname{trace}\left\{\left(\left(E\left[v^{\theta}\right]\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}v_i\right)\right)\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{\partial v_{k,i}}{\partial\theta'}-E\left[\frac{\partial v_{k,i}}{\partial\theta'}\right]\right)\right)\right\}\right]\\
&=\operatorname{trace}\left(\left(E\left[v^{\theta}\right]\right)^{-1}E\left[v_i\left(\frac{\partial v_{k,i}}{\partial\theta'}-E\left[\frac{\partial v_{k,i}}{\partial\theta'}\right]\right)\right]\right)\\
&=\operatorname{trace}\left(E\left[v\left(\frac{\partial v_{k}}{\partial\theta'}-E\left[\frac{\partial v_{k}}{\partial\theta'}\right]\right)\right]\left(E\left[v^{\theta}\right]\right)^{-1}\right).
\end{aligned}
$$

Therefore, the kth component of $E\left[v^{\theta}\right]E\left[\left(\frac{1}{2}\theta^{\epsilon\epsilon}(0)\right)\right]$ is equal to

$$
\begin{aligned}
&\frac{1}{2}\operatorname{trace}\left(E\left[\frac{\partial^2 v_k}{\partial\theta\partial\theta'}\right]\left(E\left[v^{\theta}\right]\right)^{-1}\right)+\operatorname{trace}\left(E\left[v\left(\frac{\partial v_{k}}{\partial\theta'}-E\left[\frac{\partial v_{k}}{\partial\theta'}\right]\right)\right]\left(E\left[v^{\theta}\right]\right)^{-1}\right)\\
&=\operatorname{trace}\left(\left(\frac{1}{2}E\left[\frac{\partial^2 v_k}{\partial\theta\partial\theta'}\right]+E\left[v\left(\frac{\partial v_{k}}{\partial\theta'}-E\left[\frac{\partial v_{k}}{\partial\theta'}\right]\right)\right]\right)\left(E\left[v^{\theta}\right]\right)^{-1}\right). \qquad (2.30)
\end{aligned}
$$

2.9.4 Getting back to Logit model

We now consider the general logit model. We assume that y is equal to 1 with probability equal to

$$\Lambda(x'\theta)=\frac{\exp(x'\theta)}{1+\exp(x'\theta)},$$

where x is the vector of regressors that may include the intercept term. The log likelihood is then given by

$$\log f = y\log\Lambda(x'\theta)+(1-y)\log(1-\Lambda(x'\theta)).$$

Using that $\frac{d}{dt}\Lambda(t) = \Lambda(t)(1-\Lambda(t))$, we obtain

$$v = \frac{\partial \log f}{\partial \theta} = \left(\frac{y}{\Lambda} - \frac{1-y}{1-\Lambda}\right)\Lambda(1-\Lambda)x = (y-\Lambda)x,$$

$$v^{\theta} = \frac{\partial^2 \log f}{\partial\theta\partial\theta'} = -\Lambda(1-\Lambda)xx',$$

$$\frac{\partial v_k}{\partial \theta'} = \frac{\partial((y-\Lambda)x_k)}{\partial\theta'} = -\Lambda(1-\Lambda)x_k x', \tag{2.31}$$

$$\frac{\partial^2 v_k}{\partial\theta\partial\theta'} = \frac{\partial^2((y-\Lambda)x_k)}{\partial\theta\partial\theta'} = -\Lambda(1-\Lambda)(1-2\Lambda)x_k x x'. \tag{2.32}$$

Note that

$$E\left[v v_k^{\theta}\right] = E\left[((y-\Lambda)x)(-\Lambda(1-\Lambda)x_k x')\right] = 0,$$

because the conditional expectation of y given x is equal to Λ. It follows that $E\left[v\left(\frac{\partial v_k}{\partial\theta'} - E\left[\frac{\partial v_k}{\partial\theta'}\right]\right)\right] = 0$ and the second order bias (2.30) simplifies; for the logit model, the kth component of $E\left[v^{\theta}\right] E\left[\left(\frac{1}{2}\theta^{\epsilon\epsilon}(0)\right)\right]$ is equal to

$$\frac{1}{2}\operatorname{trace}\left(E\left[\frac{\partial^2 v_k}{\partial\theta\partial\theta'}\right]\left(E\left[v^{\theta}\right]\right)^{-1}\right). \tag{2.33}$$

Remark 4 *In (2.2) and (2.3), it can be seen that A is an estimator of $E\left[v^{\theta}\right]$ and C_k is an estimator of $E\left[\frac{\partial^2 v_k}{\partial\theta\partial\theta'}\right]$.*

2.9.5 Comparison with King and Zeng (2001)'s Bias Formula

In terms of comparison with King and Zeng (2001)'s (11), we note that their bias formula can be rewritten $\left(\frac{1}{n}X'WX\right)^{-1}\left(\frac{1}{n}X'W\xi\right)$, where W in our context is equal to a diagonal matrix with diagonal elements equal to $\hat{\Lambda}_i\left(1-\hat{\Lambda}_i\right)$, $\xi_i = \frac{Q_{ii}}{2}\left(2\hat{\Lambda}_i - 1\right)$, and $Q = X\left(X'WX\right)^{-1}X'$. Note first that

$$\frac{1}{n}X'WX = \frac{1}{n}\sum_{i=1}^{n} x_i\hat{\Lambda}_i\left(1-\hat{\Lambda}_i\right)x_i' = -A, \tag{2.34}$$

where A is from (2.2). Also note that $Q = \frac{1}{n}X\left(\frac{1}{n}X'WX\right)^{-1}X' = -\frac{1}{n}XA^{-1}X'$, so we have $Q_{ii} = -\frac{1}{n}x_i'Ax_i = -\frac{1}{n}\operatorname{trace}(x_i x_i' A)$. It follows that the k-th component of $\frac{1}{n}X'W\xi$ is equal to

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} x_{i,k}\hat{\Lambda}_i\left(1-\hat{\Lambda}_i\right)\xi_i \\
&= \frac{1}{n}\sum_{i=1}^{n} x_{i,k}\hat{\Lambda}_i\left(1-\hat{\Lambda}_i\right)\frac{1}{2}\left(-\frac{1}{n}\operatorname{trace}\left(Ax_ix_i'\right)\right)\left(2\hat{\Lambda}_i-1\right) \\
&= \frac{1}{2n}\frac{1}{n}\sum_{i=1}^{n}\hat{\Lambda}_i\left(1-\hat{\Lambda}_i\right)\left(1-2\hat{\Lambda}_i\right)\operatorname{trace}\left(x_ix_i'A\right) \\
&= \frac{1}{2n}\operatorname{trace}\left(\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\Lambda}_i\left(1-\hat{\Lambda}_i\right)\left(1-2\hat{\Lambda}_i\right)x_ix_i'\right)A\right) \\
&= -\frac{1}{2n}\operatorname{trace}\left(C_kA\right) = -\frac{1}{n}T_k,
\end{aligned}
$$

where C_k and T_k are from (2.3) and (2.4). Therefore, we can understand that

$$
\frac{1}{n}X'W\xi = -\frac{1}{n}\begin{bmatrix} T_1 \\ \vdots \\ T_k \end{bmatrix}. \tag{2.35}
$$

Combining (2.34) and (2.35), we obtain

$$
\left(\frac{1}{n}X'WX\right)^{-1}\left(\frac{1}{n}X'W\xi\right) = (-A)^{-1}\left(-\frac{1}{n}\begin{bmatrix} T_1 \\ \vdots \\ T_k \end{bmatrix}\right) = B
$$

for B in (2.5).

2.9.6 Zero Second Order Bias of the Average Predicted Probability

Similar calculation as in the previous section indicates that the second order bias of is $\frac{1}{n}\sum_{i=1}^{n}\Lambda\left(x_i'\widehat{\theta}\right)$ can be estimated by

$$
\begin{aligned}
&\frac{1}{2n}\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\Lambda}_i\left(1-\hat{\Lambda}_i\right)x_i'\right)\widehat{E}\left[\theta^{\epsilon\epsilon}(0)\right] \\
&-\frac{1}{2n}\operatorname{trace}\left\{\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\Lambda}_i\left(1-\hat{\Lambda}_i\right)\left(1-2\hat{\Lambda}_i\right)x_ix_i'\right)\left(\widehat{E}\left[v^{\theta}\right]\right)^{-1}\right\}.
\end{aligned}
$$

Recalling that the first component of x_i is an intercept term, i.e., 1, we may rewrite it as

$$\frac{1}{2n}\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\Lambda}_i\left(1-\hat{\Lambda}_i\right)x_{i,1}x_i'\right)\widehat{E}\left[\theta^{\epsilon\epsilon}\left(0\right)\right]$$
$$-\frac{1}{2n}\operatorname{trace}\left\{\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\Lambda}_i\left(1-\hat{\Lambda}_i\right)\left(1-2\hat{\Lambda}_i\right)x_{i,1}x_ix_i'\right)\left(\widehat{E}\left[v^{\theta}\right]\right)^{-1}\right\}. \tag{2.36}$$

Recalling that $\widehat{E}\left[v^{\theta}\right]=-\frac{1}{n}\sum_{i=1}^{n}\hat{\Lambda}_i\left(1-\hat{\Lambda}_i\right)x_{i,1}x_i'$, we can understand the first term above as

$$-\frac{1}{2n}\left(\text{1st row of }\widehat{E}\left[v^{\theta}\right]\right)\widehat{E}\left[\theta^{\epsilon\epsilon}\left(0\right)\right].$$

In Appendix 2.9.1, we saw that the kth component of $\widehat{E}\left[v^{\theta}\right]\widehat{E}\left[\theta^{\epsilon\epsilon}\left(0\right)\right]$ is

$$-\operatorname{trace}\left\{\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\Lambda}_i\left(1-\hat{\Lambda}_i\right)\left(1-2\hat{\Lambda}_i\right)x_{i,k}x_ix_i'\right)\left(\widehat{E}\left[v^{\theta}\right]\right)^{-1}\right\},$$

so we can see that (2.36) can be rewritten

$$-\frac{1}{2n}\left(\text{1st row of }\widehat{E}\left[v^{\theta}\right]\right)\widehat{E}\left[\theta^{\epsilon\epsilon}\left(0\right)\right]+\frac{1}{2n}\left(\text{1st component of }\widehat{E}\left[v^{\theta}\right]\widehat{E}\left[\theta^{\epsilon\epsilon}\left(0\right)\right]\right)=0.$$

2.9.7 Zero Second Order Bias of the ATE Under Random Assignment

In view of (2.8) in Appendix 2.2.3, it suffices to prove that

$$\frac{n}{2n_1}\left(\frac{1}{n}\sum_{i=1}^{n}\Lambda\left(x_i'\theta_{0,(1)}\right)\left(1-\Lambda\left(x_i'\theta_{0,(1)}\right)\right)x_i'\right)E_{(1)}\left[\theta^{\epsilon\epsilon}\left(0\right)\right]$$
$$-\frac{n}{2n_1}\operatorname{trace}\left\{\left(\frac{1}{n}\sum_{i=1}^{n}\Lambda\left(x_i'\theta_{0,(1)}\right)\left(1-\Lambda\left(x_i'\theta_{0,(1)}\right)\right)\left(1-2\Lambda\left(x_i'\theta_{0,(1)}\right)\right)x_ix_i'\right)\left(E_{(1)}\left[v^{\theta}\right]\right)^{-1}\right\}$$
$$=o_p\left(1\right), \tag{2.37}$$

and

$$\frac{1}{2n_0}\left(\frac{1}{n}\sum_{i=1}^{n}\Lambda\left(x_i'\theta_{0,(0)}\right)\left(1-\Lambda\left(x_i'\theta_{0,(0)}\right)\right)x_i'\right)E_{(0)}\left[\theta^{\epsilon\epsilon}\left(0\right)\right]$$
$$-\frac{1}{2n_0}\operatorname{trace}\left\{\left(\frac{1}{n}\sum_{i=1}^{n}\Lambda\left(x_i'\theta_{0,(0)}\right)\left(1-\Lambda\left(x_i'\theta_{0,(0)}\right)\right)\left(1-2\Lambda\left(x_i'\theta_{0,(0)}\right)\right)x_ix_i'\right)\left(E_{(0)}\left[v^{\theta}\right]\right)^{-1}\right\}$$
$$=o_p\left(1\right),$$

under random assignment. We will only prove the former, because the latter can be established similarly.

Write

$$\frac{1}{n}\sum_{i=1}^{n}\Lambda\left(x_i'\theta_{0,(1)}\right)\left(1-\Lambda\left(x_i'\theta_{0,(1)}\right)\right)x_i' = \frac{n_1}{n}\frac{1}{n_1}\sum_{D_i=1}\Lambda\left(x_i'\theta_{0,(1)}\right)\left(1-\Lambda\left(x_i'\theta_{0,(1)}\right)\right)x_i' + \frac{n_0}{n}\frac{1}{n_0}\sum_{D_i=0}\Lambda\left(x_i'\theta_{0,(1)}\right)\left(1-\Lambda\left(x_i'\theta_{0,(1)}\right)\right)x_i'.$$

Under the random assignment, the x_i has the identical distribution so

$$\frac{1}{n_1}\sum_{D_i=1}\Lambda\left(x_i'\theta_{0,(1)}\right)\left(1-\Lambda\left(x_i'\theta_{0,(1)}\right)\right)x_i' = \frac{1}{n_0}\sum_{D_i=0}\Lambda\left(x_i'\theta_{0,(1)}\right)\left(1-\Lambda\left(x_i'\theta_{0,(1)}\right)\right)x_i' + o_p\left(1\right),$$

and hence, we can write

$$\begin{aligned}
&\left(\frac{1}{n}\sum_{i=1}^{n}\Lambda\left(x_i'\theta_{0,(1)}\right)\left(1-\Lambda\left(x_i'\theta_{0,(1)}\right)\right)x_i'\right)E_{(1)}\left[\theta^{\epsilon\epsilon}\left(0\right)\right] \\
&= \left(\frac{1}{n_1}\sum_{D_i=1}\Lambda\left(x_i'\theta_{0,(1)}\right)\left(1-\Lambda\left(x_i'\theta_{0,(1)}\right)\right)x_i'\right)E_{(1)}\left[\theta^{\epsilon\epsilon}\left(0\right)\right] + o_p\left(1\right) \\
&= \left(\text{1st row of } -E_{(1)}\left[v^{\theta}\right]\right)E_{(1)}\left[\theta^{\epsilon\epsilon}\left(0\right)\right] + o_p\left(1\right). \qquad (2.38)
\end{aligned}$$

Likewise, we can write

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\Lambda\left(x_i'\theta_{0,(1)}\right)\left(1-\Lambda\left(x_i'\theta_{0,(1)}\right)\right)\left(1-2\Lambda\left(x_i'\theta_{0,(1)}\right)\right)x_ix_i' \\
&= \frac{1}{n_1}\sum_{D_i=1}\Lambda\left(x_i'\theta_{0,(1)}\right)\left(1-\Lambda\left(x_i'\theta_{0,(1)}\right)\right)\left(1-2\Lambda\left(x_i'\theta_{0,(1)}\right)\right)x_ix_i' + o_p\left(1\right) \\
&= -E_{(1)}\left[\frac{\partial^2 v_1}{\partial\theta\partial\theta'}\right] + o_p\left(1\right)
\end{aligned}$$

under random assignment, so

$$\begin{aligned}
&\text{trace}\left\{\left(\frac{1}{n}\sum_{i=1}^{n}\Lambda\left(x_i'\theta_{0,(1)}\right)\left(1-\Lambda\left(x_i'\theta_{0,(1)}\right)\right)\left(1-2\Lambda\left(x_i'\theta_{0,(1)}\right)\right)x_ix_i'\right)\left(E_{(1)}\left[v^{\theta}\right]\right)^{-1}\right\} \\
&= -E_{(1)}\left[\frac{\partial^2 v_1}{\partial\theta\partial\theta'}\right]\left(E_{(1)}\left[v^{\theta}\right]\right)^{-1} + o_p\left(1\right) \\
&= -\text{1st component of } \widehat{E}\left[v^{\theta}\right]\widehat{E}\left[\theta^{\epsilon\epsilon}\left(0\right)\right] + o_p\left(1\right). \qquad (2.39)
\end{aligned}$$

Combining (2.38) and (2.39), we conclude that (2.37) under random assignment as long as $n_1, n_0 \to \infty$ at the same rate.

2.9.8 Lyapunov Condition

Note that

$$
\begin{aligned}
E\left|\frac{y_i - p_n}{\sqrt{np_n(1-p_n)}}\right|^3 &= \left(\frac{1-p_n}{\sqrt{np_n(1-p_n)}}\right)^3 p_n + \left(\frac{p_n}{\sqrt{np_n(1-p_n)}}\right)^3 (1-p_n) \\
&= \frac{1}{n\sqrt{n}} \frac{2p_n^2 - 2p_n + 1}{\sqrt{p_n(1-p_n)}},
\end{aligned}
$$

so

$$
\sum_{i=1}^{n} E\left|\frac{y_i - p_n}{\sqrt{np_n(1-p_n)}}\right|^3 = \frac{1}{\sqrt{n}} \frac{2p_n^2 - 2p_n + 1}{\sqrt{p_n(1-p_n)}} \to 0
$$

if $p_n \propto n^{-\delta}$ with $0 \le \delta < 1$.

If $p_n \propto \frac{1}{n}$, we can see that

$$
\lim_{n\to\infty} \frac{1}{\sqrt{n}} \frac{2p_n^2 - 2p_n + 1}{\sqrt{p_n(1-p_n)}} = 1,
$$

so Lyapunov condition is violated. This is consistent with the fact that we should expect Poisson approximation not normal approximation when p_n is very small. For concreteness, we will assume that $p_n = \frac{\lambda}{n}$ and analyze

$$
\begin{aligned}
\widehat{\theta} - \theta &= \ln\frac{\overline{y}}{1-\overline{y}} - \ln\frac{p_n}{1-p_n} = \ln\frac{\overline{y}}{p_n} - \ln\frac{1-\overline{y}}{1-p_n} \\
&= \ln\frac{n\overline{y}}{\lambda} - \ln\frac{1-\overline{y}}{1-\frac{\lambda}{n}}.
\end{aligned}
$$

As for the second term, we can use $n\overline{y} = O_p(1)$ to conclude that

$$
\ln\frac{1-\overline{y}}{1-\frac{\lambda}{n}} = \ln\frac{1-O_p\left(\frac{1}{n}\right)}{1-\frac{\lambda}{n}} = o_p(1),
$$

while $\ln\frac{n\overline{y}}{\lambda}$ is asymptotically equivalent to log of Poisson(λ) divided by its own mean. Problem is that we need to confront the fact that a Poisson distribution has a positive probability of being equal to 0, at which point the ln is not well-defined.

2.9.9 Derivation of (2.20)

We analyze

$$\widehat{\theta}-\theta=\ln\frac{\overline{y}}{1-\overline{y}}-\ln\frac{p_n}{1-p_n}=\ln\frac{\overline{y}}{p_n}-\ln\frac{1-\overline{y}}{1-p_n}.$$

As for the second term, we have

$$\begin{aligned}
\ln\frac{1-\overline{y}}{1-p_n}&=\ln\frac{1-\left(p_n+\frac{\sqrt{p_n(1-p_n)}}{\sqrt{n}}Z_n\right)}{1-p_n}\\
&=\ln\left(1-\frac{\sqrt{p_n}}{\sqrt{n\left(1-p_n\right)}}Z_n\right)\\
&=-\frac{\sqrt{p_n}}{\sqrt{n\left(1-p_n\right)}}Z_n-\frac{1}{2}\left(\frac{\sqrt{p_n}}{\sqrt{n\left(1-p_n\right)}}Z_n\right)^2+O_p\left(n^{-3(1+\delta)/2}\right)\\
&=-\frac{\sqrt{p_n}}{\sqrt{n\left(1-p_n\right)}}Z-\frac{1}{2}\left(\frac{\sqrt{p_n}}{\sqrt{n\left(1-p_n\right)}}\right)^2Z_n^2+O_p\left(n^{-3(1+\delta)/2}\right)
\end{aligned}$$

noting that $\frac{\sqrt{p_n}}{\sqrt{n(1-p_n)}}=O\left(n^{-(1+\delta)/2}\right)$. As for the first term, we have

$$\begin{aligned}
\ln\frac{\overline{y}}{p_n}&=\ln\left(1+\frac{\sqrt{1-p_n}}{\sqrt{np_n}}Z_n\right)\\
&=\frac{\sqrt{1-p_n}}{\sqrt{np_n}}Z_n-\frac{1}{2}\left(\frac{\sqrt{1-p_n}}{\sqrt{np_n}}Z_n\right)^2+O\left(n^{-3(1-\delta)/2}\right)\\
&=\frac{\sqrt{1-p_n}}{\sqrt{np_n}}Z_n-\frac{1}{2}\left(\frac{\sqrt{1-p_n}}{\sqrt{np_n}}\right)^2Z_n^2+O\left(n^{-3(1-\delta)/2}\right)
\end{aligned}$$

noting that $\frac{\sqrt{1-p_n}}{\sqrt{np_n}}=O\left(n^{-(1-\delta)/2}\right)$. To conclude, we have

$$\begin{aligned}
\widehat{\theta}-\theta&=\ln\frac{\overline{y}}{1-\overline{y}}-\ln\frac{p_n}{1-p_n}=\ln\frac{\overline{y}}{p_n}-\ln\frac{1-\overline{y}}{1-p_n}\\
&=\frac{\sqrt{1-p_n}}{\sqrt{np_n}}Z_n-\frac{1}{2}\left(\frac{\sqrt{1-p_n}}{\sqrt{np_n}}\right)^2Z_n^2+O_p\left(n^{-3(1-\delta)/2}\right)\\
&+\frac{\sqrt{p_n}}{\sqrt{n\left(1-p_n\right)}}Z_n+\frac{1}{2}\left(\frac{\sqrt{p_n}}{\sqrt{n\left(1-p_n\right)}}\right)^2Z_n^2+O_p\left(n^{-3(1+\delta)/2}\right)\\
&=\frac{1}{\sqrt{np_n\left(1-p_n\right)}}Z_n-\frac{1}{2}\frac{1-2p_n}{np_n\left(1-p_n\right)}Z_n^2+O_p\left(n^{-3(1-\delta)/2}\right),
\end{aligned}$$

and using $\sqrt{np_n\left(1-p_n\right)} = O_p\left(n^{-(1-\delta)/2}\right)$, we conclude that

$$\sqrt{np_n\left(1-p_n\right)}\left(\widehat{\theta}-\theta\right) = Z_n - \frac{1}{2}\frac{1-2p_n}{\sqrt{np_n\left(1-p_n\right)}}Z_n^2 + O_p\left(n^{-(1-\delta)}\right).$$

CHAPTER 3

A Vuong Test for Panel Data Models with Fixed Effects

3.1 Introduction

When researchers working with panel data seek to select among models in terms of a few parameters of interest, they can turn to model selection tests. The classical test is the Vuong test. However, incidental parameters—high-dimensional parameters that affect the unbiasednesss of the parameters of interest—are also important for panel data models as they capture unobserved heterogeneity. In the presence of incidental parameters, we cannot easily apply the classical Vuong test to select a panel data model.

An advantage of panel models is to deal with unobserved heterogeneity, which can be modeled as time-invariant individual-specific effects. With the fixed effects approaches to panel data models, we do not need to impose distributional assumptions on the unobserved effects, thereby allowing the unobserved effects to be arbitrarily related with the observed covariates. The fixed effect estimators treat the unobserved effects as parameters to be estimated. It is well-known that maximum likelihood estimates of the panel data models suffer from the incidental parameters problem as noted by Neyman and Scott (1948); that is, the estimation of the fixed effects when the time dimension is short can be severely biased. Some approaches circumvent the inconsistency problem based on the intuition that there are clever ways to avoid estimating the fixed effects; for instance, in linear models, fixed effects are numerically equivalent to the within-group estimator that removes the individual effects by taking differences within each individual, so that first differencing yields models free of fixed effects. However, such estimators generally only apply to specific models, and the existence of such estimators seems to be quite rare when we want to analyze fixed effects estimation of average partial effects (APEs), which are averages of functions of the data, parameters, and unobserved effects. Moreover, fixed effects are unavoidable if we are interested in settings where individuals are clustered at different levels. For example, students may be clustered at class, school, and district level; or observations are clustered according to household, county, and state codes. Therefore, dealing with the incidental parameter problem coming from the noise of estimation of fixed effects is important for panel data models. From this perspective, it is natural to propose a test for model selection among various panel data models in the presence of incidental parameters.

The fixed-T approximation limitations can be overcome by an alternative asymptotic approximation that considers sequences of panels where both N and T increase to infinity, see Arellano and Hahn (2007), Hahn and Kuersteiner (2011) and Hahn and Newey (2004). This large-T approximation makes the incidental parameter problem of fixed effects estimation become an asymptotic bias problem that is easier to tackle. These bias-corrected estimators are designed to remove the O(1/T) bias and generally applicable. There are several ways to achieve this goal in the literature: Hahn and Newey (2004) and Hahn and Kuersteiner (2011) construct an analytical or numerical bias correction of a fixed-effects estimator for nonlinear panel data models. Another approach is to consider estimators from bias-corrected moment equations, see Woutersen (2002), Arellano (2003), Carro (2007) and Fernández-Val (2005). In addition, Pace and Salvan (2006) and Arellano and Hahn (2007) propose estimation from a bias-corrected objective function relative to some target criterion. Another strand of literature focuses on a modified objective function where the correction term is designed to remove the O(1/T) bias of the resulting estimator (Bester and Hansen, 2009; Arellano and Hahn, 2016). The correction term could be trace-based or determinant-based.

A vast literature assumes that incidental parameters are correctly specified. But often this may not be the case. For example, a researcher may incorrectly specify incidental parameters if she clusters her data at zip code, city, or country level when in fact the data are actually clustered at the individual level. In fact, the specification problem in panel data models is more severe than in cross sectional data because incidental parameters change over time, thereby generating a different set of potential model specifications for every period of time.

At present, we do not have a test for model selection for these situations where there are many incidental parameters. Without a test specifically designed for panel data with incidental parameters, the misspecifications of discrepancy in the structure of incidental parameters could have serious consequences for model selection. Because it affects the unbiasedness of the low-dimensional parameters of interests. If we ignore the specifications of incidental parameters and only select models based on low-dimensional parameters, we may choose the wrong model.

This chapter proposes a new model selection test for panel data models by extending the classical Vuong test, which selects from two parametric likelihood models based on their Kullback–Leibler information criterion (KLIC). Suppose there are two panel models, for example, panel probit and panel logit. We do not know whether both or one of the two models would be misspecified because the true model is unknown. But it is of interest to know whether one of the two models is superior to the other. Following Vuong (1989), we derive an LR based test of the null hypothesis that two models are equivalent in terms of their distances to the true model. Under the null hypothesis, both models are equally close to the true data distribution in terms of the Kullback–Leibler (KL) divergence. When the null does not hold, the tests direct the researcher to the model closer to the true distribution with probability approaching one. An important feature of panel data models is the specification of incidental parameters or fixed effects, which is different from the classical Vuong test. We exploit a modified objective function to deal with infinite-dimensional nuisance parameters. This chapter is different from Lee and Phillips (2015) in that they assume the parameter space of fixed effects is common across the candidate models. Therefore their tests choose the model that best fits the data generating process when only a subset of the parameters is of central interest, which is a special case in this chapter.

This chapter provides a test to select a better model from two competing nonlinear panel models with incidental parameters. It may seem that we can easily extend the criterion function in classical Vuong test to panel data models. Indeed, when there is no incidental parameter, the classical Vuong test allows the researcher to select between two parametric likelihood models based on their Kullback–Leibler information criterion (KLIC) . However, with incidental parameters, the estimators for panel data models are severely biased. This is called the incidental parameters problem, as noted by Neyman and Scott (1948).

In order to extend classical Vuong test to panel data with incidental parameters, we propose three new test statistics based on a new criterion function. In classical Vuong test, the object function is maximized at pseudo-true values. However, the expectation of the concentrated likelihood for panel data with incidental parameters is not maximized at the true value of the parameter. To be consistent with classical Vuong test, we use a modified likelihood function as the new criterion function. The purpose is to generate a closer approximation to the target likelihood function.

The discrepancy in incidental parameters could have serious consequences for model selection; for example, as noted by MacKinnon et al. (2020), there is a vast literature on cluster-robust inference that assumes the structure of the clusters is correctly specified, which is often violated. An interesting case investigated in their paper is a test for the appropriate clustering level in linear regression models. They show that clustering at either the classroom or school level is better than no clustering using data from the Tennessee Student Teacher Achievement Ratio (STAR) experiment. More generally, if we have observations taken from individuals in different geographical locations, there could possibly be clustering at the zip-code, city, county, state, or country level. Even in this simple cross-sectional setting, we need to choose one model among many different specifications of fixed effects. In panel data models, it highly possible that the fixed effects change over time, thereby generating different model specifications for every period of time. This chapter's main goal is to provide a way to select a better model from two competing nonlinear panel models with fixed effects, which allows for disagreements about both parameters of interests and incidental parameters.

We offer three different test statistics for researchers who need to deal with all possible relationships between candidate models: overlapping models,nested models, and strictly nonnested models. These three model relationships are classified according to the structure of low-dimensional parameter of interest and high-dimensional incidental parameters. It is shown that these three test statistics have different convergence rates. Users can choose one test according to the specific model relationship.

Since both finite-dimensional parameters of interest and infinite-dimensional incidental parameters present in the model, the non-nested hypotheses is different from the literature featuring Cox (1961), Atkinson (1970), Pesaran (1974), Pesaran and Deaton (1978), Mizon and Richard (1986), Gourieroux and Monfort (1995), Ramalho and Smith (2002), Bontemps et al. (2008), among others. There are three possible situations in which two models are strictly non-nested. First, they share the same structure of incidental parameters but have different parameters of interest, which are non-nested. For example, panel logit and panel probit with identical individual-level fixed effects. Second, they have the same parameters of interest but different specifications of incidental parameters, such as panel logit models clustered at different levels. Third, both the parameters of interest and incidental parameters are different in those two models. Continue with the example, panel logit and panel probit clustered at different levels.

The classical Vuong test suffers from a discontinuity problem in the asymptotic distribution of the test statistic, which means that the asymptotics depends on whether the models are nested, non-nested, or overlapping, as noted by Shi (2015), Hsu and Shi (2017), Liao and Shi (2020), Liu and Lee (2019) and so on. Shi (2015) shows that the classical Vuong tests either have severe size distortion or poor power due to this discontinuity problem and propose a one-step nondegenerate Vuong-type test for moment-based models. Liao and Shi (2020) then extend the test to semi/nonparametric models. Hsu and Shi (2017) introduce some additional randomness into the test statistic and derive a one-step test for model selection between conditional moment restriction models. Liu and Lee (2019) show that their intuition carries over to spatial models. This chapter follows the manner of Shi (2015) and Liao and Shi (2020), to construct bias-corrected test statistics for panel model selection. The test achieves uniformly asymptotic size control and is consistent regardless of the true DGPs for non-nested, nested, and overlapping models.

This chapter is organized as follows. Section 3.2 sets up the objective function for model selection and compares it with the classical Vuong test. Section 3.3 conducts the test for non-nested models, Section 3.4 is for nested models, and Section 3.5 is for overlapping models. Useful lemmas and some important proofs are in Section 3.6.

3.2 The classical Vuong test and extension to panel data models

3.2.1 The classical Vuong test

When there is no incidental parameter, the classical Vuong test allows the researcher to select between two parametric likelihood models based on their Kullback–Leibler information criterion (KLIC). In this section, we review the classical Vuong test and show that how it could be adopted to panel data models with fixed effects.

Assume that we have two parametric models $\mathcal{F}$ and $\mathcal{G}$ to choose from, their densities are $y_i \sim f(y;\theta) : \theta \in \Theta \subset R^{d_\theta}$ and $y_i \sim g(y;\gamma) : \gamma \in \Gamma \subset R^{d_\gamma}$ respectively. Since we do not know the true underlining data generating process (DGP), we are interested in the comparison between these two models: which one is closer to the truth. The classical Vuong test (Vuong, 1989) looks at their distances to the DGP in terms of the Kullback and Leibler Information Criterion (KLIC). Let $\psi_i(\phi)$ be the logarithm of the ratio of the two p.d.f.s:

$$\sum_{i=1}^{n} \psi_i(\phi) = \sum_{i=1}^{n} \log f(y_i;\theta) - \sum_{i=1}^{n} \log g(y_i;\gamma)$$

$\phi_0 = (\theta_0', \gamma_0')$ is the concatenated vector of the pseudo-true values that maximizes the expectation under the density functions $f(y_i;\theta)$ and $g(y_i;\gamma)$:

$$\theta_0 = \arg\max_{\theta\in\Theta} E\left[\sum_{i=1}^{n} \log f(y_i;\theta)\right]$$
$$\gamma_0 = \arg\max_{\gamma\in\Gamma} E\left[\sum_{i=1}^{n} \log g(y_i;\gamma)\right]$$

The distance between the DGP and these two models is measured by the minimum KLIC among distributions in the model. And we would like to select the best model among a collection of competing models is the one that is closest to the true model. In order to do that, Vuong (1989) adopts the following test hypotheses:

$$H_0 : E\left[\sum_{i=1}^{n} \psi_i(\phi_0)\right] = 0$$
$$H_f : E\left[\sum_{i=1}^{n} \psi_i(\phi_0)\right] > 0$$
$$H_g : E\left[\sum_{i=1}^{n} \psi_i(\phi_0)\right] < 0$$

Under H_0, $\mathcal{F}$ and $\mathcal{G}$ are equally good since they are equally distant from true distribution in the Kullback-Leibler sense. If $\mathcal{F}$ is a "better" model, $E[\sum_{i=1}^{n}\psi_i(\phi_0)]$ is expected to be "big", therefore under H_f, f is favored since it is closer to the true distribution. Under H_g, g is "better" as $E[\sum_{i=1}^{n}\psi_i(\phi_0)]$ is small. The test statistics $LR_n = \sum_{i=1}^{n}\psi_i(\hat{\phi}_n)$ is the sample analogue of $LR_0 = \sum_{i=1}^{n}\psi_i(\phi_0)$, where $\widehat{\phi}_n = \left(\widehat{\theta}'_n, \widehat{\gamma}'_n\right)$, and that

$$\widehat{\theta}_n = \arg\max_{\theta\in\Theta} \sum_{i=1}^{n} \log f(y_i;\theta)$$
$$\widehat{\gamma}_n = \arg\max_{\gamma\in\Gamma} \sum_{i=1}^{n} \log g(y_i;\gamma)$$

If these two models are strictly non-nested: $\mathcal{F}\cap\mathcal{G} = \varnothing$, which is the easiest case, we have

$$n^{-1/2}LR_n = n^{-1/2}\sum_{i=1}^{n}\psi_i\left(\hat{\phi}_n\right) \xrightarrow{d} N\left(0,\omega^2\right)$$

The estimator for the variance term is

$$\hat{\omega}_n^2 = n^{-1} \sum_{i=1}^{n} \left[\psi_i \left(\hat{\phi}_n \right) \right]^2 \xrightarrow{p} \omega^2$$

Thus we have the following asymptotics:

$$H_0 : n^{-1/2} LR_n / \hat{\omega}_n \xrightarrow{d} N(0,1)$$
$$H_f : n^{-1/2} LR_n / \hat{\omega}_n \xrightarrow{\text{a.s.}} +\infty$$
$$H_g : n^{-1/2} LR_n / \hat{\omega}_n \xrightarrow{\text{a.s.}} -\infty$$

If these two models are nested, for example, $\mathcal{F}$ nests $\mathcal{G}$: $\mathcal{G} \subset \mathcal{F}$. Assume that $\dim(\theta) \geq \dim(\gamma)$, under H_0:

$$2LR_n \xrightarrow{d} \chi^2(\dim(\theta) - \dim(\gamma))$$

H_0 is rejected if $2LR_n > c\left(\hat{Q}_n, 1-\alpha\right)$, $c\left(\hat{Q}_n, 1-\alpha\right)$ is $1-\alpha$ quantile of $Y'QY$. $\hat{Q}_n$ is a consistent estimator of Q.

3.2.2 Incidental parameter problem in panel data models

With incidental parameters, the estimators for panel data models are severely biased. This is called the incidental parameters problem, as noted by Neyman and Scott (1948). Since it violates the unbiasedness properties in classical Vuong test, we need to tackle with it for model selection tests.

In this part, we review the bias-corrected estimator and explain why the modified likelihood function behaves more like a genuine likelihood function. Consider panel observations $\{y_{it}\}$ for $i = 1, 2, \ldots, n$ and $t = 1, 2, \ldots, T$. The density function is $y_{it} \sim f(y|\theta_0, \alpha_{i0})$, where θ_0 is the common parameter of interest and α_{i0} denotes individual fixed effect, which are considered to be non-stochastic constants. A maximization estimator is defined by

$$\left(\hat{\theta}, \hat{\alpha}_1, \ldots, \hat{\alpha}_n\right) = \underset{\theta, \gamma_1, \ldots, \gamma_n}{\text{argmax}} \sum_{i=1}^{n} \sum_{t=1}^{T} \log f\left(x_{it}; \theta, \alpha_i\right)$$

We assume that if n is fixed and $T \to \infty$, the estimator $\left(\hat{\theta}, \hat{\alpha}_1, \ldots, \hat{\alpha}_n\right)$ is consistent for $(\theta_0, \alpha_{10}, \ldots, \alpha_{n0})$ .To simplify notation, we assume $\dim(\alpha_i) = 1$, concentrating out the α_i leads to the characterization

$$\hat{\theta} \equiv \underset{\theta}{\operatorname{argmax}} \sum_i \sum_t \log f\left(y_{it} \mid \theta, \widehat{\alpha}_i(\theta)\right)$$
$$\widehat{\alpha}_i(\theta) \equiv \underset{\alpha}{\operatorname{argmax}} \sum_t \log f\left(y_{it} \mid \theta, \alpha\right)$$

The $\widehat{\alpha}_i(\theta)$ on the data only through the i-th observation so there are only T observations to estimate α_i. Let

$$\theta_T \equiv \underset{\theta}{\operatorname{argmax}} \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} E\left[\sum_{t=1}^{T} \log f\left(y_{it} \mid \theta, \widehat{\alpha}_i(\theta)\right)\right]$$

is biased in general $\theta_T \neq \theta_0$ (incidental parameters problem). Before proceeding to the bias-corrected estimator, it will be useful to define some notation:

$$\begin{aligned}
u_{it}(\theta, \alpha) \equiv& \frac{\partial}{\partial\theta} \log f\left(y_{it} \mid \theta, \alpha\right) \\
v_{it}(\theta, \alpha) \equiv& \frac{\partial}{\partial\alpha} \log f\left(y_{it} \mid \theta, \alpha\right) \\
U_{it}(\theta, \alpha) \equiv& u_{it}(\theta, \alpha) - v_{it}(\theta, \alpha) E\left[v_{it}^{\alpha_i}\right]^{-1} E\left[u_{it}^{\alpha_i}\right] \\
b_i\left(\theta_0\right) =& -\left(\frac{E\left[v_{it} U_{it}^{\alpha_i}\right]}{E\left[v_{it}^{\alpha_i}\right]} - \frac{E\left[U_{it}^{\alpha_i\alpha_i}\right] E\left[v_{it}^2\right]}{2\left(E\left[v_{it}^{\alpha_i}\right]\right)^2}\right) \\
\mathcal{I}_i \equiv& -E\left[\frac{\partial U_{it}\left(\theta_0, \alpha_0\right)}{\partial\theta}\right] \\
B =& \left(\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathcal{I}_i\right)^{-1} \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} b_i\left(\theta_0\right)
\end{aligned}$$

We use the short-hand notation $u_{it} \equiv u_{it}\left(\theta_0, \alpha_0\right), v_{it} \equiv v_{it}\left(\theta_0, \alpha_0\right), U_{it} \equiv U_{it}\left(\theta_0, \alpha_0\right)$ and we denote by v_{it}^{α} and $v_{it}^{\alpha\alpha}$ the first and second derivatives of v_{it} with respect to α_i. We denote $\widetilde{\theta}$ as biased-corrected MLE estimator, which is

$$\tilde{\theta} \equiv \widehat{\theta} - \frac{\widehat{B}}{T} \tag{3.1}$$

$\widehat{B}$ is an estimator of the bias term, which can be a sample analogue of B.[1] The idea behind this method is to expand the incidental parameters bias of the estimator on the order of magnitude T, and to subtract an estimate of the leading term of the bias from the estimator. The bias stems from the fact that we use $\sum_i \sum_t \log f\left(y_{it} \mid \theta, \widehat{\alpha}_i(\theta)\right)$ rather than $\sum_i \sum_t \log f(y_{it}; \theta_0, \alpha_i(\theta_0))$ to estimate θ.

3.2.3 The new criterion function

It may seem that we can easily extend the criterion function in classical Vuong test to panel data models, but the incidental parameter problem induces a difficulty. Using a modified likelihood function as the new criterion function, we extend the classical Vuong test to panel data models with incidental parameters.

Intuitively, we can treat the likelihood function which is evaluated at bias-corrected estimator $\tilde{\theta}$ in equation (3.1) as criterion function $\psi_f\left(y_{it}, \tilde{\theta}, \widehat{\alpha}_i(\tilde{\theta})\right)$:

$$\sum_i \sum_t \psi_f\left(y_{it}, \tilde{\theta}, \widehat{\alpha}_i(\tilde{\theta})\right) \equiv \sum_i \sum_t \log f\left(y_{it}; \tilde{\theta}, \widehat{\alpha}_i(\tilde{\theta})\right)$$

In classical Vuong test, the object function is maximized at pseudo-true values. However, the expectation of the concentrated likelihood $\psi_f\left(y_{it}, \tilde{\theta}, \widehat{\alpha}_i(\tilde{\theta})\right)$ is not maximized at the true value of the parameter. If we use plug-in estimator of incidental parameters $\widehat{\alpha}_i(\theta)$ instead of $\alpha_i(\theta_0)$, the first-order derivative of $\sum_i \sum_t \log f\left(y_{it} \mid \theta, \widehat{\alpha}_i(\theta)\right)$ with respect to θ is not centered at zero when $\theta = \widetilde{\theta}$. In order to be consistent with classical Vuong test, we use a modified

[1] According to Arellano and Hahn (2007), we can estimate the bias term using $\widehat{B}(\widehat{\theta}) = \left(\frac{1}{n}\sum_{i=1}^n \widehat{\mathcal{I}}_i\right)^{-1} \frac{1}{n}\sum_{i=1}^n \widehat{b}_i(\widehat{\theta})$, $\widehat{\mathcal{I}}_i = -\left(\widehat{E}_T\left[\widehat{u}_{it}^{\theta}\right] - \widehat{E}_T\left[\widehat{u}_{it}^{\alpha_i}\right]\widehat{E}_T\left[\widehat{v}_{it}^{\alpha_i}\right]^{-1}\widehat{E}_T\left[\widehat{u}_{it}^{\alpha_i\prime}\right]\right)$, and

$$\widehat{b}_i(\widehat{\theta}) = \left(\frac{-\widehat{E}_T\left[\widehat{v}_{it}^2\right]}{\widehat{E}_T\left[\widehat{v}_{it}^{\alpha_i}\right]}\right)\left\{-\frac{1}{\left(-\widehat{E}_T\left[\widehat{v}_{it}^2\right]\right)}\left(\widehat{E}_T\left[\widehat{v}_{it}\widehat{u}_{it}^{\alpha_i}\right] - \widehat{E}_T\left[\widehat{v}_{it}\widehat{v}_{it}^{\alpha_i}\right]\frac{\widehat{E}_T\left[\widehat{u}_{it}^{\alpha_i}\right]}{\widehat{E}_T\left[\widehat{v}_{it}^{\alpha_i}\right]}\right)\frac{1}{2\widehat{E}_T\left[\widehat{v}_{it}^{\alpha_i}\right]}\left(\widehat{E}_T\left[\widehat{u}_{it}^{\alpha_i\alpha_i}\right] - \widehat{E}_T\left[\widehat{v}_{it}^{\alpha_i\alpha_i}\right]\frac{\widehat{E}_T\left[\widehat{u}_{it}^{\alpha_i}\right]}{\widehat{E}_T\left[\widehat{v}_{it}^{\alpha_i}\right]}\right)\right\}$$

where $\widehat{E}_T(.) = \sum_{t=1}^T (.)/T$, $\widehat{u}_{it}^{\theta} = u_{it}^{\theta}\left(\widehat{\theta}, \widehat{\alpha}_i(\widehat{\theta})\right)$, $\widehat{u}_{it}^{\alpha_i} = u_{it}^{\alpha_i}\left(\widehat{\theta}, \widehat{\alpha}_i(\widehat{\theta})\right)$, etc.

likelihood function instead:

$$LM_{nT}(\widetilde{\theta}) \equiv \sum_i \sum_t \psi_f \left(y_{it}, \tilde{\theta}, \widehat{\alpha}_i(\tilde{\theta}) \right) - \sum_i \widehat{R}^*_{fi}(\widetilde{\theta})$$

The purpose of including an extra modification term $\sum_i \widehat{R}^*_{fi}(\widetilde{\theta})$, which is defined in equation (3.3), is to generate a closer approximation to the target likelihood function $\sum_i \sum_t \psi_f \left(\theta_0, \alpha_i(\theta_0)\right)$. For example, Bester and Hansen (2009), Arellano and Hahn (2016) and Lee and Phillips (2015) consider trace-based correction term $\widehat{R}^*_{fi}(\widetilde{\theta})$, which depends exclusively on the Hessian and the outer product of the scores of the fixed effects. In the literature, it is shown that the modified likelihood function has zero first-order derivative at $\tilde{\theta}$. We will discuss the formula of $\widehat{R}^*_{fi}(\widetilde{\theta})$ in different models.

Based on the modified likelihood functions, we offer three different test statistics for researchers who need to deal with all possible relationships between candidate models.

3.3 The non-nested models

There are three possible situations in which two models are strictly non-nested. First, they share the same structure of incidental parameters but have different parameters of interest, which are non-nested. For example, panel logit and panel probit with identical individual-level fixed effects.

Second, they have the same parameters of interest but different specifications of incidental parameters, such as panel logit models clustered at different levels, e.g. the zip-code, city, county, state, or country level.

Third, both the parameters of interest and incidental parameters are different in those two models. For instance, panel logit clustered at individual level and panel probit clustered at zip-code level.

Suppose there are two panel models f and g which are strictly non-nested but have same structure of fixed effects, for example, the binary panel model with fixed effects is characterized by $y_{it} = 1(\gamma_{i0} + z_{it}'\theta_0 + \varepsilon_{it} \geq 0)$,where ε_{it} conditional on z_{it} either has a logistic or standard normal distribution. The former one is defined by model f and the latter one is model g, they are considered as non-nested cases. It is necessary to consider both finite-dimensional parameters of interest and infinite-dimensional incidental parameters when determining model relationships. Continue with the previous example in the introduction, we have observations taken from individuals in different geographical locations, there could possibly be clustering at the zip-code, city, county, state, or country level.

The main assumption we make about nonnested models is as follows:

Assumption 1 *Suppose that f and g have different parameters of interest and structure of incidental parameters but they are non-nested, i.e., there is no $(\theta_0, \gamma_0, \alpha_{i0}, \lambda_{i0}) \in \Theta \times \Upsilon \times A \times \Lambda$ such that $f(y_{it}, \theta_0, \alpha_i) = g(y_{it}, \gamma_0, \lambda_i)\, \forall y_{it} \in Y$.*

We consider the strictly non-nested models and denote the modified likelihood function evaluated at $\widetilde{\theta}$ in equation (3.1) as (take model f as an example):

$$LM_{nT}(\widetilde{\theta}) \equiv \sum_i \sum_t \psi_f(y_{it}, \widetilde{\theta}, \widehat{\alpha}_i(\widetilde{\theta})) - \sum_i \widehat{R}^*_{fi}(\widetilde{\theta}) \tag{3.2}$$

Without the modification term $\sum_i \widehat{R}^*_{fi}(\widetilde{\theta})$, $\frac{1}{nT}\sum_i \sum_t \psi_f\left(y_{it}, \widetilde{\theta}, \widehat{\alpha}_i(\widetilde{\theta})\right)$ does not have zero first-order derivatives at $\tilde{\theta}$, which violates the property of objective function in the classical Vuong test. Expanding equation (3.2) around $(\theta_0, \widehat{\alpha}_i(\theta_0))$, the Taylor's Theorem implies that for $\widetilde{\widetilde{\theta}}$ in between θ_0 and $\widetilde{\theta}$:

$$\begin{aligned}
&\sum_i \sum_t \psi_f\left(y_{it}, \widetilde{\theta}, \widehat{\alpha}_i(\widetilde{\theta})\right) - \sum_i \widehat{R}^*_{fi}(\widetilde{\theta}) \\
=&\sum_i \sum_t \psi_f(y_{i,t}; \theta_0, \widehat{\alpha}_i(\theta_0)) - \sum_i \widehat{R}^*_{fi}(\theta_0) && \text{(i)} \\
+&\frac{\partial\left[\sum_i \sum_t \psi_f(y_{i,t}; \theta_0, \widehat{\alpha}_i(\theta_0)) - \sum_i \widehat{R}^*_{fi}(\theta_0)\right]}{\partial\theta}(\tilde{\theta} - \theta_0) && \text{(ii)} \\
+&\frac{1}{2}(\tilde{\theta} - \theta_0)' \frac{\partial^2([\sum_i \sum_t \psi_f(y_{i,t}; \widetilde{\widetilde{\theta}}, \widehat{\alpha}_i(\widetilde{\widetilde{\theta}})) - \sum_i R_{fi}(\widetilde{\widetilde{\theta}})])}{\partial\theta\partial\theta'}(\tilde{\theta} - \theta_0) && \text{(iii)}
\end{aligned}$$

where

$$
\begin{aligned}
\widehat{R}^{*}_{fi}(\theta) &\equiv -\frac{1}{2}\frac{\frac{1}{T}\sum_t v^2_{fi,t}(\theta,\alpha_i(\theta))}{E\left[v^{\alpha}_{fi,t}(\theta,\alpha_i(\theta))\right]} \\
\widehat{R}^{*}_{fi}(\theta_0) &\equiv -\frac{1}{2}\frac{\frac{1}{T}\sum_t v^2_{fi,t}}{E\left[v^{\alpha}_{fi,t}\right]} \\
\widehat{R}^{*}_{fi}(\widetilde{\theta}) &\equiv -\frac{1}{2}\frac{\frac{1}{T}\sum_t \widehat{v}^2_{fi,t}}{\frac{1}{T}\sum_t \widehat{v}^{\alpha}_{fi,t}}
\end{aligned}
\tag{3.3}
$$

and $v \equiv v(\theta_0,\alpha_i(\theta_0))$, $v^{\alpha}_{fi,t} \equiv v^{\alpha}_{fi,t}(\theta_0,\alpha_i(\theta_0))$,$\widehat{v}_{fi,t} \equiv v_{fi,t}[\tilde{\theta},\widehat{\alpha}_i(\tilde{\theta})]$, $\widehat{v}^{\alpha}_{fi,t} \equiv v^{\alpha}_{i,t}[\tilde{\theta},\widehat{\alpha}_i(\tilde{\theta})]$. According to Proposition 4 in the appendix, we prove that in the first term (i):

$$
\sum_i\sum_t \psi_f(y_{i,t};\theta_0,\widehat{\alpha}_i(\theta_0)) - \sum_i R_{fi}(\theta_0) = \sum_i\sum_t \psi_f(y_{it},\theta_0,\alpha_i(\theta_0)) + o_p(1) \tag{3.4}
$$

where $R_{fi}(\theta_0) \equiv -\frac{1}{2}\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{fi,t}\right)^2}{E\left[v^{\alpha}_{fi,t}\right]}$. Therefore the first term (i) can be written as:

$$
\sum_i\sum_t \psi_f(y_{it},\theta_0,\alpha_i(\theta_0)) + \sum_i R_{fi}(\theta_0) - \sum_i \widehat{R}^{*}_{fi}(\theta_0) \tag{3.5}
$$

In Proposition 3, the second term (ii) is shown to be

$$
\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}\right)\mathcal{I}_f^{-1}\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}\right) + o_p(1) \tag{3.6}
$$

where $\mathcal{I}_f \equiv \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_{fi}$. We deduce that the third term (iii) becomes:

$$
-\frac{1}{2}\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}\right)\mathcal{I}_f^{-1}\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}\right) + o_p(1) \tag{3.7}
$$

Combining equations (3.5), (3.6) and (3.7):

$$
\begin{aligned}
&\sum_i\sum_t \psi_f\left(y_{it},\widetilde{\theta},\widehat{\alpha}_i(\widetilde{\theta})\right) - \sum_i \widehat{R}^{*}_{fi}(\widetilde{\theta}) && (3.8)\\
=&\sum_i\sum_t \psi_f(y_{it},\theta_0,\alpha_i(\theta_0)) + \sum_i R_{fi}(\theta_0) - \sum_i \widehat{R}^{*}_{fi}(\theta_0) && \text{(a)}\\
&+\frac{1}{2}\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}\right)\mathcal{I}_f^{-1}\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}\right) && \text{(b)}
\end{aligned}
$$

The first term (a) is $O_p\left(\sqrt{nT}\right)$ since we know that from Proposition 6 and 9 in the appendix:

$$\frac{1}{\sqrt{n}}\sum_i\left[R_{fi}\left(\theta_0\right)-\widehat{R}^*_{fi}(\theta_0)\right]\xrightarrow{d}N(0,\sigma^2_{U_f})$$

$$\frac{1}{\sqrt{nT}}\sum_i\left[\sum_t\psi_f\left(y_{it},\theta_0,\alpha_i(\theta_0)\right)-E\left(\sum_t\psi_f\left(y_{it},\theta_0,\alpha_i(\theta_0)\right)\right)\right]\xrightarrow{d}N(0,\omega^2_f)$$

where

$$\sigma^2_{U_f}\equiv\lim_{n\to\infty}\frac{1}{2n}\sum_i\frac{\left(E\left[v^2_{fi,t}\right]\right)^2}{\left(E\left[v^{\alpha}_{fi,t}\right]\right)^2} \tag{3.9}$$

$$\omega^2_f\equiv\lim_{n\to\infty}\frac{1}{n}\sum_i E\left[\psi_f\left(y_{it},\theta_0,\alpha_i(\theta_0)\right)-E\left(\psi_f\left(y_{it},\theta_0,\alpha_i(\theta_0)\right)\right)\right]^2 \tag{3.10}$$

The second term (b) is $O_p\left(1\right)$. We divide the equation (3.8) by $\sqrt{nT}$ at both sides, we have:

$$\begin{aligned}\frac{1}{\sqrt{nT}}\left[\sum_i\sum_t\psi_f\left(y_{it},\widetilde{\theta},\widehat{\alpha}_i(\widetilde{\theta})\right)-\sum_i\widehat{R}^*_{fi}(\widetilde{\theta})\right]=&\frac{1}{\sqrt{nT}}\sum_i\sum_t\psi_f\left(y_{it},\theta_0,\alpha_i(\theta_0)\right)\\&+\frac{1}{\sqrt{nT}}\sum_i\left[R_{fi}\left(\theta_0\right)-\widehat{R}^*_{fi}(\theta_0)\right]+o_p\left(1\right)\\=&\frac{1}{\sqrt{nT}}\sum_i\sum_t\psi_f\left(y_{it},\theta_0,\alpha_i(\theta_0)\right)+o_p\left(1\right)\end{aligned}$$

and hence

$$\frac{1}{\sqrt{nT}}\sum_i\sum_t\psi_f\left(y_{it},\widetilde{\theta},\widehat{\alpha}_i(\widetilde{\theta})\right)-\frac{1}{\sqrt{nT}}\sum_i\widehat{R}^*_{fi}(\widetilde{\theta})=\frac{1}{\sqrt{nT}}\sum_i\sum_t\psi_f\left(y_{it},\theta_0,\alpha_i(\theta_0)\right)+o_p(1) \tag{3.11}$$

we further denote:

$$LM_{nT}\left(\tilde{\theta},\tilde{\gamma}\right)\equiv\left(\sum_i\sum_t\psi_f\left(y_{it},\widetilde{\theta},\widehat{\alpha}_i(\widetilde{\theta})\right)-\sum_i\widehat{R}^*_{fi}(\widetilde{\theta})\right)-\left(\sum_i\sum_t\psi_g\left(y_{it},\widetilde{\gamma},\widehat{\alpha}_i(\widetilde{\gamma})\right)-\sum_i\widehat{R}^*_{gi}(\widetilde{\gamma})\right)$$

$$LM_{nT}\left(\theta_0,\gamma_0\right)\equiv\sum_i\sum_t\psi_f\left(y_{it},\theta_0,\alpha_i(\theta_0)\right)-\sum_i\sum_t\psi_g\left(y_{it},\theta_0,\alpha_i(\theta_0)\right)$$

Equation (3.11) becomes:

$$\frac{1}{\sqrt{nT}}LM_{nT}(\tilde{\theta},\tilde{\gamma})=\frac{1}{\sqrt{nT}}LM_{nT}\left(\theta_0,\gamma_0\right)+o_p(1) \tag{3.12}$$

Theorem 1 *If assumption 1 and condition 1 hold, then under* $H_0 : E\left[LM_{nT}\left(\theta_0, \gamma_0\right)\right] = 0 :$

$$\frac{\frac{1}{\sqrt{nT}} LM_{nT}(\tilde{\theta}, \tilde{\gamma})}{\widehat{\omega}_n} \xrightarrow{d} N(0,1)$$

$H_f : E\left[LM_{nT}\left(\theta_0, \gamma_0\right)\right] > 0 :$

$$\frac{\frac{1}{\sqrt{nT}} LM_{nT}(\tilde{\theta}, \tilde{\gamma})}{\widehat{\omega}_n} \xrightarrow{a.s.} +\infty$$

$H_g : E\left[LM_{nT}\left(\theta_0, \gamma_0\right)\right] < 0 :$

$$\frac{\frac{1}{\sqrt{nT}} LM_{nT}(\tilde{\theta}, \tilde{\gamma})}{\widehat{\omega}_n} \xrightarrow{a.s.} -\infty$$

where $\widehat{\omega}_n^2$ *is an estimator for* ω^2*, and* $\widehat{\omega}_n^2 = \frac{1}{nT} \sum_i \sum_t [\psi_f \left(\widetilde{\theta}, \widehat{\alpha}_i(\widetilde{\theta})\right) - \psi_g \left(\tilde{\gamma}, \widehat{\alpha}_i(\tilde{\gamma})\right)]^2$, $\omega^2 = \lim_{n \to \infty} \frac{1}{n} \sum_i E\left[\psi_f\left(\theta_0, \alpha_i(\theta_0)\right) - \psi_g\left(\gamma_0, \alpha_i(\gamma_0)\right)\right]^2$.

Proof. See Proposition 7, 9 and appendix 3.6.4. ■

3.4 The nested models

When we consider the nested cases, there are two possibilities: (i) there is no disagreement about low-dimensional parameters; (ii) the high-dimensional incidental parameters is unchanging over these specifications. For the first case, the test converges at root n, which is slower than the classical Vuong test. This is because the test is driven by the discrepancy in high-dimensional incidental parameters. For the second case, we have chi-sqaure distribution based on modified criterion function, which is identical to classical Vuong test.

Example 1 *(Different cluster levels) As noted by MacKinnon et al. (2020), there is a vast literature on cluster-robust inference that assumes the structure of the clusters is correctly specified, which is often violated. An interesting case investigated in their paper is a test for the appropriate clustering level in linear regression models. They show that clustering at either the classroom or school level is better than no clustering using data from the Tennessee Student*

Teacher Achievement Ratio (STAR) experiment. More generally, if we have observations taken from individuals in different geographical locations, there could possibly be clustering at the zip-code, city, county, state, or country level.

Consider the panel model with fixed effects with known variance of the error term, but the cluster levels change over time:

$$\psi_f(y_{i,t};\theta,\alpha_i) = -\frac{1}{2}\ln\sigma^2 - \frac{1}{2}\frac{(y_{it}-x_{it}\theta-\alpha_i)^2}{\sigma^2} \quad \text{for} \quad t=1,2,\ldots,2J$$

$$\psi_g(y_{i,t};\theta,\alpha_i) = \begin{cases} -\frac{1}{2}\ln\sigma^2 - \frac{1}{2}\frac{(y_{it}-x_{it}\theta-\alpha_{i1})^2}{\sigma^2} & \text{for} \quad t=1,2,\ldots,J \\ -\frac{1}{2}\ln\sigma^2 - \frac{1}{2}\frac{(y_{it}-x_{it}\theta-\alpha_{i2})^2}{\sigma^2} & \text{for} \quad t=J+1,2,\ldots,2J \end{cases}$$

For model f, the fixed effects α_i is constant across different time periods, while for model g, α_i changes with time. For instance, observations are clustered at individual level at first half period of time, but clustered at county level later on. We can treat α_i as different cluster levels of observations. Model f is nested in model g since f is equivalent to g when $\alpha_{i1}=\alpha_{i2}$. Assume that $\sigma^2=1$, and the initial value is taken from a stationary distribution, and we obtain

$$\psi_f(y_{i,t};\theta,\alpha_i) = -\frac{1}{2}(y_{it}-x_{it}\theta-\alpha_i)^2$$

We note that for model f:

$$\begin{aligned} \alpha_i(\theta) &= E(y_{it}-x_{it}\theta), & \widehat{\alpha}_i\left(\widetilde{\theta}\right) &= \textstyle\sum_{t=1}^{t=2J}\left(y_{it}-x_{it}\widetilde{\theta}\right). \\ R_{fi}(\theta_0) &= -\tfrac{1}{2}\left[\tfrac{1}{2J}\textstyle\sum_{t=1}^{t=2J}(y_{it}-x_{it}\theta_0-\alpha_i)\right]^2, & \widehat{R}_{fi}\left(\widetilde{\theta}\right) &= -\tfrac{1}{2}\left[\tfrac{1}{2J}\textstyle\sum_{t=1}^{t=2J}\left(y_{it}-x_{it}\widetilde{\theta}-\widehat{\alpha}_i\left(\widetilde{\theta}\right)\right)\right]^2. \\ R^*_{fi}(\theta_0) &= -\tfrac{1}{2}E\left[\tfrac{1}{2J}\textstyle\sum_{t=1}^{t=2J}(y_{it}-x_{it}\theta_0-\alpha_i)^2\right], & \widehat{R}^*_{fi}\left(\widetilde{\theta}\right) &= -\tfrac{1}{2}\left[\tfrac{1}{2J}\textstyle\sum_{t=1}^{t=2J}\left[y_{it}-x_{it}\widetilde{\theta}-\widehat{\alpha}_i\left(\widetilde{\theta}\right)\right]^2\right]. \end{aligned}$$

We can also compute the modification terms for model g for $t=1,2,\ldots,J$ and $t=J+1,2,\ldots,2J$, respectively.

3.4.1 Case (i) no disagreement about parameters of interest

We consider the first case and define the nested relationship as follows:

Assumption 2 *Suppose that f and g share the parameters of interest θ_0, but g nests f, i.e., there exists a function $\phi_\alpha(.)$ from A^F to A^G such that for any $\alpha_{fi}(\theta_0)$ in A^F: $f(y_{it}, \theta_0, \alpha_{fi}(\theta_0)) = g(y_{it}, \theta_0, \phi(\alpha_{gi}(\theta_0))) \, \forall y_{it} \in Y$.*

Divide equation (3.8) by $\sqrt{n}$ at both sides, since the term (b) is $O_p(1)$, we have the following expansion for model g:

$$\frac{1}{\sqrt{n}} \sum_i \sum_{t=1}^{t=2J} \psi_f (y_{i,t}; \theta_0, \alpha_i(\theta_0)) + \sum_i \frac{1}{\sqrt{n}} \left[R_{fi}(\theta_0) - \widehat{R}^*_{fi}(\theta_0) \right] \tag{3.13}$$

$$= \frac{1}{\sqrt{n}} \sum_i \sum_{t=1}^{t=2J} \psi_f \left(y_{it}, \widetilde{\theta}, \widehat{\alpha}_i(\widetilde{\theta}) \right) - \frac{1}{\sqrt{n}} \sum_i \widehat{R}^*_{fi}(\widetilde{\theta}) + o_p(1) \tag{3.14}$$

where $\widehat{R}^*_{fi}(\widetilde{\theta}) \equiv -\frac{1}{2} \frac{\frac{1}{2J} \sum_{t=1}^{t=2J} \widehat{v}^2_{fi,t}}{\frac{1}{2J} \sum_{t=1}^{t=2J} \widehat{v}^\alpha_{fi,t}}$. Similarly, for model model g, we compute the statistic for two different time periods separately, the corresponding modification terms are:

$$\widehat{R}^*_{1gi}(\widetilde{\theta}) \equiv -\frac{1}{2} \frac{\frac{1}{J} \sum_{t=1}^{t=J} \widehat{v}^2_{1gi,t}}{\frac{1}{J} \sum_{t=1}^{t=J} \widehat{v}^\alpha_{1gi,t}}, \quad \widehat{R}^*_{2gi}(\widetilde{\theta}) \equiv -\frac{1}{2} \frac{\frac{1}{J} \sum_{t=J+1}^{t=2J} \widehat{v}^2_{2gi,t}}{\frac{1}{J} \sum_{t=J+1}^{t=2J} \widehat{v}^\alpha_{2gi,t}}.$$

We define

$$\begin{aligned} LM_{nT}\left(\tilde{\theta}, \widehat{\alpha}_i, \widehat{\alpha}_{i1}, \widehat{\alpha}_{i2}\right) \equiv & \sum_i \left\{ \sum_{t=1}^{t=2J} \psi_f \left(y_{it}, \tilde{\theta}, \widehat{\alpha}_i(\tilde{\theta}) \right) - \sum_{t=1}^{t=J} \psi_g \left(y_{it}, \tilde{\theta}, \widehat{\alpha}_{i1}(\tilde{\theta}) \right) \right. \\ & \left. - \sum_{t=J+1}^{t=2J} \psi_g \left(y_{it}, \tilde{\theta}, \widehat{\alpha}_{i2}(\tilde{\theta}) \right) \right\} - \sum_i \left[\widehat{R}^*_{fi}(\tilde{\theta}) - \widehat{R}^*_{1gi}(\tilde{\theta}) - \widehat{R}^*_{2gi}(\tilde{\theta}) \right] \end{aligned}$$

Theorem 2 *If assumption 2 and condition 1 hold, then under $H_0 : E\left[LM_{nT}(\theta_0, \alpha_i, \alpha_{i1}, \alpha_{i2})\right] = 0$:*

$$\frac{\frac{1}{\sqrt{n}} LM_{nT}(\widetilde{\theta}, \widehat{\alpha}_i, \widehat{\alpha}_{i1}, \widehat{\alpha}_{i2})}{\widehat{\sigma}_{U^{nested}}} \xrightarrow{d} N(0,1)$$

Under $H_g : E\left[LM_{nT}(\theta_0, \alpha_i, \alpha_{i1}, \alpha_{i2})\right] < 0$:

$$\frac{\frac{1}{\sqrt{n}} LR_{nT}(\widetilde{\theta}, \widehat{\alpha}_i, \widehat{\alpha}_{i1}, \widehat{\alpha}_{i2})}{\widehat{\sigma}_{U^{nested}}} \xrightarrow{a.s.} -\infty$$

where $\widehat{\sigma}^2_{U nested}$ *is an estimator for* $\sigma^2_{U nested}$*:*

$$\widehat{\sigma}^2_{U nested} \equiv \frac{J-1}{2nJ} \sum_i \frac{\left(\frac{1}{J}\sum_{t=1}^{t=J} \widehat{v}^2_{1gi,t}\right)^2}{\left(\frac{1}{J}\sum_{t=1}^{t=J} \widehat{v}^{\alpha}_{1gi,t}\right)^2}$$

$$\sigma^2_{U nested} \equiv \lim_{n\to\infty} \frac{1}{2n} \sum_i \frac{\left(E\left[v^2_{1gi,t}\right]\right)^2}{\left(E\left[v^{\alpha}_{1gi,t}\right]\right)^2}$$

If the information matrix identity holds, $\sigma^2_{U nested} = \frac{1}{2}$.

Proof. See Proposition 8 and proof of Theorem 3.6.4 in the appendix. ■

3.4.2 Case (ii) no disagreement about incidental parameters

We assume that the parameter of interest θ is different across candidate models while the incidental parameters α_i is unchanging over these specifications. Take the following case as an example:

$$f: \quad y_{it} = \alpha_i + x_{it}\theta_1 + \varepsilon_{it}$$

$$g: \quad y_{it} = \alpha_i + x_{it1}\theta_1 + x_{it2}\theta_2 + \varepsilon_{it}$$

Similar to nested case (i), model f is a special case for model g when $\theta_2 = 0$. A formal definition is as follows:

Assumption 3 *Suppose that* f *and* g *have same structure of incidental parameters but* g *nests* f*, i.e., there exists a function* $\phi(.)$ *from* Θ *to* Γ *such that for any* θ *in* Θ*:* $f\left(y_{it}, \theta_0, \alpha_i(\theta_0)\right) = g\left(y_{it}, \phi(\theta_0), \alpha_i\left(\phi(\theta_0)\right)\right) \forall y_{it} \in Y$.

Under the null hypothesis, $R_{fi}\left(\theta_0\right) = R_{gi}(\gamma_0)$ and $\widehat{R}^*_{fi}(\theta_0) = \widehat{R}^*_{gi}(\gamma_0)$ since these two models have same structure of fixed effects. We obtain the following expansion from equation (3.8):

$$LM_{nT}\left(\tilde{\theta}, \tilde{\gamma}\right) = LM_{nT}\left(\theta_0, \gamma_0\right) + \frac{NT}{2}(\widetilde{\theta}-\theta_0)'\mathcal{I}_f(\widetilde{\theta}-\theta_0) - \frac{NT}{2}(\widetilde{\gamma}-\gamma_0)'\mathcal{I}_g(\widetilde{\gamma}-\gamma_0) + o_p(1)$$

therefore under null assumption, assume that $\dim(\gamma) \geq \dim(\theta)$ (see proof in the appendix 3.6.4),

$$-2LM_{nT}\left(\tilde{\theta},\tilde{\gamma}\right) \xrightarrow{d} \chi^2(\dim(\gamma)-\dim(\theta))$$

Theorem 3 *If assumption 3, the information matrix identity holds for model f, condition 1 holds, then under $H_0 : E\left[LM_{nT}(\theta_0,\gamma_0)\right]=0$, for any $x>0$,*

$$\Pr\left(-2LM_{nT}\left(\tilde{\theta},\tilde{\gamma}\right)<x\right)-c\left(\hat{Q}_n,1-\alpha\right) \xrightarrow{a.s.} 0$$

Under $H_g : E\left[LM_{nT}(\theta_0,\gamma_0)\right]<0$:

$$-2LM_{nT}\left(\tilde{\theta},\tilde{\gamma}\right) \xrightarrow{a.s.} +\infty$$

where $c(Q,1-\alpha)$ is the $1-\alpha$ quantile of $\chi^2(\dim(\gamma)-\dim(\theta))$.

Proof. *See the appendix 3.6.4.* ■

Lee and Phillips (2015) assume that the parameter space of fixed effects is common across the candidate models, their tests choose the model that best fits the data generating process when only a subset of the parameters is of central interest, which is equivalent to our results if the information matrix identity holds (see a proof of equivalence in the appendix 3.6.4).

3.5 The overlapping models

Here we consider two models that are overlapping and they are not non-nested. We consider the case when the variance term might be zero, and the converge rate of the test is root n. This case is different from classical Vuong test as there is no uniform formula for overlapping models in the literature.

Assumption 4 *Suppose that f and g are overlapping and they are not non-nested.*

The test statistic for overlapping models are similar to case (i) in nested models, since the asymptotic distribution is mostly driven by differences in incidental parameters. It is pointed out by Shi (2015) that the high-order bias may dominate the leading terms in LR_n and result in size distortion if we follow Vuong (1989)'s framework and construct a two-step test for overlapping nonnested models. Some recent papers propose one-step nondegenerate test for different data models: Shi (2015) shows that the classical Vuong tests either have severe size distortion or poor power due to this discontinuity problem and propose a one-step nondegenerate Vuong-type test for moment-based models, Liao and Shi (2020) then extend the test to semi/nonparametric models. Hsu and Shi (2017) introduce some additional randomness into the test statistic and derive a one-step test for model selection between conditional moment restriction models, Liu and Lee (2019) show that their intuition carries over to spatial models. This chapter follows the manner of Shi (2015) and Liao and Shi (2020), to construct bias-corrected test statistics for panel model selection. The test achieves uniformly asymptotic size control and is consistent regardless of the true DGPs for non-nested, nested, and overlapping models.

Consider a simple case when $LM_{nT}(\theta_0, \gamma_0) = 0$, and the degeneracy problem is considered under the assumption that f and g share the parameters of interest θ_0, while they have different structures of incidental parameters, which corresponds to previous analysis for nested case (i):

$$
\begin{aligned}
\frac{1}{\sqrt{n}} LM_{nT}\left(\tilde{\theta}, \widetilde{\gamma}\right) =& \frac{1}{\sqrt{n}} \sum_i \left[R_{fi}(\theta_0) - R_{gi}(\gamma_0) - \left(R_{fi}^* - R_{gi}^*\right)\right] + o_p(1) \\
\xrightarrow{d}& N\left(0, \sigma_U^2\right)
\end{aligned}
$$

where σ_U^2 is defined in equation (3.15), the proof is included in Proposition 7 in the appendix.

Theorem 4 *If assumption 4 and condition 1 hold, then under* $H_0 : E\left[LM_{nT}(\theta_0, \gamma_0)\right] = 0$:

$$
\frac{\frac{1}{\sqrt{n}} LM_{nT}(\tilde{\theta}, \widetilde{\gamma})}{\widehat{\sigma}_U} \xrightarrow{d} N(0,1)
$$

under $H_f : E\left[LM_{nT}\left(\theta_0, \gamma_0\right)\right] > 0$:

$$\frac{\frac{1}{\sqrt{n}} LM_{nT}(\tilde{\theta}, \widetilde{\gamma})}{\widehat{\sigma}_U} \overset{a.s.}{\to} +\infty$$

under $H_g : E\left[LM_{nT}\left(\theta_0, \gamma_0\right)\right] < 0$:

$$\frac{\frac{1}{\sqrt{n}} LM_{nT}(\tilde{\theta}, \widetilde{\gamma})}{\widehat{\sigma}_U} \overset{a.s.}{\to} -\infty$$

where $\widehat{\sigma}_U^2$ *is an estimator for* σ_U^2*, and that*

$$\widehat{\sigma}_U^2 \equiv \frac{T-1}{2nT} \sum_i \left[\frac{\frac{1}{T}\sum_{t=1}^{t=T} \widehat{v}_{fi,t}^2}{\frac{1}{T}\sum_{t=1}^{t=T} \widehat{v}_{fi,t}^{\alpha}} - \frac{\frac{1}{T}\sum_{t=1}^{t=T} \widehat{v}_{gi,t}^2}{\frac{1}{T}\sum_{t=1}^{t=T} \widehat{v}_{gi,t}^{\alpha}} \right]^2$$

$$\sigma_U^2 \equiv \lim_{n\to\infty} \frac{1}{2n} \sum_i E\left[\frac{E\left[v_{fi,t}^2\right]}{E\left[v_{fi,t}^{\alpha}\right]} - \frac{E\left[v_{gi,t}^2\right]}{E\left[v_{gi,t}^{\alpha}\right]} \right]^2 \tag{3.15}$$

3.6 Appendix

3.6.1 Regularity Conditions

Condition 1

1. $n, T \to \infty$ *such that* $(n/T) \to \rho$*,where* $0 < \rho < \infty$.

2. *(i) The function* $\log f\left(.;\theta,\alpha\right)$ *is continuous in* $(\theta,\alpha) \in \mathcal{Y}$; *(ii) the parameter space* $\mathcal{Y}$ *is compact; (iii) there exists a function such that* $\left|\partial \log f\left(y_{it};\theta,\alpha_i\right)\right| \leq M\left(y_{it}\right)$,

$$\left|\frac{\partial \log f\left(y_{it};\theta,\alpha_i\right)}{\partial\left(\theta,\alpha_i\right)}\right| \leq M\left(y_{it}\right)$$

and $\sup_i E\left[M\left(y_{it}\right)^{33}\right] < \infty$.

3. *For each* $\eta > 0$,

$$\inf_i \left[G_{(i)}\left(\theta_0, \alpha_{i0}\right) - \sup_{|(\theta,\alpha):|(\theta,\alpha)-(\theta_0,\alpha_0)|>\eta|} G_{(i)}(\theta,\alpha) \right] > 0$$

where

$$\widehat{G}_{(i)}(\theta, \alpha_i) \equiv T^{-1}\sum_{t=1}^{T}\log f(y_{it};\theta,\alpha_i) \equiv T^{-1}\sum_{t=1}^{T} g(y_{it};\theta,\alpha_i)$$

$$G_{(i)}(\theta,\alpha_i) \equiv E[\log f(y_{it};\theta,\alpha_i)]$$

Let $\mathcal{I}_i \equiv E[U_{it}U_{it}']$ *.*

4. *(i) There exists some* $M(y_{it})$ *such that*

$$\left|\frac{\partial^{m_1+m_2}\log f(z_{it};\theta,\alpha_i)}{\partial\theta^{m_1}\partial\alpha_i^{m_2}}\right| \leq M(z_{it}), \quad 0 \leq m_1+m_2 \leq 1,\ldots,6$$

and $\sup_i E\left[M(y_{it})^Q\right] < \infty$ *for some* $Q > 64$*; (ii)* $\bar{E}[\mathcal{I}_i] > 0$*; (iii)* $\min_i E[v_{it}^2] > 0$.

5. $\sup_i\left[\frac{1}{T}\left|\frac{\partial S_n(\theta)}{\partial\theta'}\right|\right] = o_p(1)$*, where*

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

and $\frac{1}{T}\sum_{t=1}^{T}\tilde{V}_{it} = -\left(E\left[\frac{\partial v_i}{\partial\alpha_i}\right]\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}v_{it}\right)$.

3.6.2 Lemmas

Lemma 2 *(Arellano and Hahn, 2016, Theorem 2) Under condition 1,*

$$\sqrt{nT}(\tilde{\theta}-\theta_0) = \mathcal{I}_f^{-1}\left\{\frac{1}{\sqrt{nT}}\sum_i\sum_t U_{fi,t}(\theta_0,\alpha_{i0})\right\} + o_p(1)$$

and that

$$\sqrt{nT}(\tilde{\theta}-\theta_0) \xrightarrow{d} N\left(0,\mathcal{I}_f^{-1}\right)$$

Lemma 3 *(Hahn and Kuersteiner, 2011, Lemma 7) Assume that* $\{Q_t, t = 1,2...\}$ *is a stationary, mixing sequence with* $E[Q_t] = 0$ *and* $E\left[|Q_t|^{2r+\delta}\right] < \infty$ *for any positive integer* r*, some* $\delta > 0$ *and all* t*. Let* $\mathcal{A}_t \equiv \sigma(Q_t, Q_{t-1}, Q_{t-2},\ldots), \mathcal{B}_t \equiv \sigma(Q_t, Q_{t+1}, Q_{t+2},\ldots)$*, and*

$$\alpha(m) \equiv \sup_t \sup_{A\in\mathcal{A}_t, B\in\mathcal{B}_{t+m}} |P(A\cap B) - P(A)P(B)| \tag{3.16}$$

Then, for any m *such that* $1 \leq m < C(r)\,n$,

$$E\left[\left(\sum_{i=1}^{n} Q_i\right)^{2r}\right] \leq C(r)\left|Q_t\right|^{2r+\delta}\left[n^r m^{2r}\right] \tag{3.17}$$

Lemma 4 *(Hahn and Kuersteiner, 2011, Lemma 1) Suppose that, for each* i, $\{\xi_{it}, t = 1, 2, ...\}$ *is a mixing sequence with* $E[\xi_{it}] = 0$ *for all* i, t. *Let* $\mathcal{A}_t^{\xi i} \equiv \sigma(\xi_{it}, \xi_{it-1}, \xi_{it-2}, \ldots), \mathcal{B}_t^{\xi i} \equiv \sigma(x_{it}, x_{it+1}, x_{it+2}, \ldots)$, *and*

$$\alpha_i(m) \equiv \sup_t \sup_{A \in \mathcal{A}_t^{\xi i}, B \in \mathcal{B}_{t+m}^{\xi i}} |P(A \cap B) - P(A)P(B)| \tag{3.18}$$

Assume that $\sup_i |\alpha_i(m)| \leq Ca^m$ *for some* a *such that* $0 < a < 1$ *and some* $0 < C < \infty$. *We assume that* $\{\xi_{it}, t = 1, 2, 3, ...\}$ *are independent across* i. *We also assume that* $n = O(T)$. *Finally, assume that* $E\left[|\xi_{it}|^{6+\delta}\right] < \infty$ *for some* $\delta > 0$. *We then have*

$$P\left[\max_{1 \leq i \leq n}\left|\frac{1}{T}\sum_{t=1}^{T}\xi_{it}\right| > \eta\right] = o\left(T^{-1}\right) \tag{3.19}$$

for every $\eta > 0$. *Now assume that* $E\left[|\xi_{it}|^{10q+12+\delta}\right] < \infty$ *for some* $\delta > 0$ *and some integer* $q \geq 1$. *Then,*

$$P\left[\max_{1 \leq i \leq n}\left|\frac{1}{T}\sum_{t=1}^{T}\xi_{it}\right| > \eta T^{\frac{1}{10}-v}\right] = o\left(T^{-q}\right) \tag{3.20}$$

for every $\eta > 0$ *and* $0 < v < (100q + 120)^{-1}$.

Lemma 5 *(Hahn and Kuersteiner, 2011, Lemma 3) Assume that* y_{it} *satisfies condition 1, and let* $\xi(y_{it}, \phi)$ *be a function indexed by the parameter* $\phi \in int\Phi$, *where* Φ *is a convex subset of* $\mathbb{R}^p$. *For any sequence* $\phi_i \in int\Phi$, *assume that* $E[\xi(y_{it}, \phi_i)] = 0$. *Further assume that* $\sup_\phi \|\xi(y_{it}, \phi)\| \leq \mathbf{M}(y_{it})$ *for some* $\mathbf{M}(y_{it})$ *such that* $E\left[\mathbf{M}(y_{it})^4\right] < \infty$. *Let* $\Sigma_{nT} = \sum_{i=1}^{n}\Sigma_{iT}^{\xi\xi}$ *with* $\Sigma_{iT}^{\xi\xi} = Var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi(y_{it}, \phi_i)\right)$. *Denote the smallest eigenvalue of* $\Sigma_{iT}^{\xi\xi}$ *by* λ_{iT}^{ξ}, *and assume that* $\inf_i \inf_T \lambda_{iT}^{\xi} > 0$. *Then,*

$$\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\xi(y_{it}, \phi_i) \xrightarrow{d} N\left(0, f^{\xi\xi}\right), \text{ and } \sup_i \left\|\Sigma_{iT}^{\xi\xi} - f_i^{\xi\xi}\right\| \to 0, \tag{3.21}$$

where $f^{\xi\xi} \equiv \lim \frac{1}{n}\sum_{i=1}^{n} f_i^{\xi\xi}$, *and* $f_i^{\xi\xi} \equiv \sum_{j=-\infty}^{\infty} E\left[\xi(y_{it}, \phi_i)\xi(y_{it-j}, \phi_i)'\right]$.

Lemma 6 *(Hahn and Kuersteiner, 2011, Theorem 4)* $P\left[\max_{1\leq i\leq n}\max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}}\left|\widehat{\alpha}_i(\epsilon)-\alpha_{i0}\right|\geq\eta\right]=o\left(T^{-1}\right)$ *for every* $\eta>0$.

Lemma 7 *(i)* $\alpha_i^\theta=-\frac{E\left[v_{it}^\theta\right]}{E\left[v_{it}^\alpha\right]}$; *(ii)* $\frac{\partial\widehat{\alpha}_i}{\partial\theta}=-\frac{\sum_{t=1}^T v_{it}^\theta}{\sum_{t=1}^T v_{it}^\alpha}$.

Proof. *Consider* $\alpha_i\left(\theta,F_i(\varepsilon)\right)$ *solves the estimating equation*

$$\int v_i\left[\theta,\alpha_i\left(\theta,F_i(\varepsilon)\right)\right]dF_i(\varepsilon)=0 \tag{3.22}$$

Differentiating the LHS with respect to θ *and* ε, *we obtain*

$$0=\int_i^\theta dF_i(\varepsilon)+\alpha_i^\theta\int v_i^\alpha dF_i(\varepsilon) \tag{3.23}$$

$$0=\alpha_i^\varepsilon\int_i^{\alpha_i}dF_i(\varepsilon)+\int_i d\Delta_{iT} \tag{3.24}$$

where

$$\Delta_{iT}=\frac{dF_i(\varepsilon)}{d\varepsilon}=\sqrt{T}\left(\widehat{F}_i-F_i\right) \tag{3.25}$$

We solve for these equations and evaluate them at $\varepsilon=0$ *gives (i):*

$$\alpha_i^\theta=-E\left[v_{it}^\alpha\right]^{-1}E\left[v_{it}^\theta\right]=O_p(1) \tag{3.26}$$

For (ii), by the definition of $\widehat{\alpha}_i(\theta)$,

$$\sum_{t=1}^T v_{it}\left(\theta,\widehat{\alpha}_i(\theta)\right)=0 \tag{3.27}$$

Differentiating the LHS with respect to θ, *we obtain*

$$\sum_{t=1}^T v_{it}^\theta+\sum_{t=1}^T v_{it}^\alpha\left[\frac{\partial\widehat{\alpha}_i}{\partial\theta}\right]=0 \tag{3.28}$$

it follows that

$$\frac{\partial\widehat{\alpha}_i}{\partial\theta}=-\frac{\sum_{t=1}^T v_{it}^\theta}{\sum_{t=1}^T v_{it}^\alpha} \tag{3.29}$$

■

Lemma 8

$$\frac{1}{nT}\sum_i \sqrt{T}\left(\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right)=-\frac{1}{nT}\sum_i \frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)}{E\left[v_{i,t}^{\alpha}\right]}+O_p\left(\frac{1}{T\sqrt{T}}\right) \tag{3.30}$$

$$\frac{1}{nT}\sum_i\left(\sqrt{T}\left(\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right)\right)^2=\frac{1}{nT}\sum_i \frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{(E\left[v_{i,t}^{\alpha}\right])^2}+O_p\left(\frac{1}{T\sqrt{T}}\right) \tag{3.31}$$

$$\frac{1}{nT}\sum_i\left(\sqrt{T}\left(\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right)\right)^3=-\frac{1}{nT}\sum_i \frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^3}{(E\left[v_{i,t}^{\alpha}\right])^3}+o_p\left(\frac{1}{T\sqrt{T}}\right) \tag{3.32}$$

Proof.

Since we have

$$\frac{1}{T}\sum_{t=1}^{T} v\left(y_{i,t};\theta_0,\widehat{\alpha}_i\left(\theta_0\right)\right)=0$$

Let $F \equiv (F_1,\ldots,F_n)$ *denote the collection of marginal distribution functions of* y_{it}*.Let* $\hat{F}_i$ *denote the empirical distribution function for the observation* i*. Define* $F_i(\epsilon) \equiv F_i + \epsilon\sqrt{T}\left(\hat{F}_i - F_i\right)$ *for* $\epsilon \in \left[0, T^{-1/2}\right]$*. For each fixed* θ *and* ϵ*, let* $\alpha_i(\epsilon)$ *be the solution to the estimating equation*

$$0=\int v\left[\cdot;\theta_0,\alpha_i(\epsilon)\right]dF_i(\epsilon)$$

By a Taylor series expansion, we have

$$\widehat{\alpha}_i\left(\theta_0\right)-\alpha_{i0}=\alpha_i\left(\frac{1}{\sqrt{T}}\right)-\alpha_i(0)=\frac{1}{\sqrt{T}}\alpha_i^{\epsilon}(0)+\frac{1}{2}\left(\frac{1}{\sqrt{T}}\right)^2\alpha_i^{\epsilon\epsilon}(0)+\frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^3\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})$$

where $\alpha_i^{\epsilon}(\epsilon) \equiv d\alpha_i(\epsilon)/d\epsilon, \alpha_i^{\epsilon\epsilon}(\epsilon) \equiv d^2\alpha_i(\epsilon)/d\epsilon^2, \ldots,$ *and* $\widetilde{\epsilon}$ *is somewhere in between 0 and* $\frac{1}{\sqrt{T}}$*.*

Let $h_i(\cdot,\epsilon) \equiv v\left[\cdot;\theta_0,\alpha_i(\epsilon)\right]$*, the first order condition could be written as*

$$0=\int h_i(\cdot,\epsilon)dF_i(\epsilon)$$

Differentiating repeatedly with respect to ϵ*,we obtain*

$$0=\int \frac{dh_i(\cdot,\epsilon)}{d\epsilon}dF_i(\epsilon)+\int h_i(\cdot,\epsilon)d\Delta_{iT} \tag{A.1}$$

$$0=\int \frac{d^2h_i(\cdot,\epsilon)}{d\epsilon^2}dF_i(\epsilon)+2\int \frac{dh_i(\cdot,\epsilon)}{d\epsilon}d\Delta_{iT} \tag{A.2}$$

$$0=\int \frac{d^3h_i(\cdot,\epsilon)}{d\epsilon^3}dF_i(\epsilon)+3\int \frac{d^2h_i(\cdot,\epsilon)}{d\epsilon^2}d\Delta_{iT} \tag{A.3}$$

where $\Delta_{iT} \equiv \sqrt{T}\left(\hat{F}_i - F_i\right)$.

(A.1) $\alpha_i^{\epsilon}(0)$ *Evaluating (A.1)*

$$0 = \left(\int v^{\alpha}\left[\cdot;\theta_0,\alpha_i(\epsilon)\right] dF_i(\epsilon)\right)\alpha_i^{\epsilon}(\epsilon) + \int v\left[\cdot;\theta_0,\alpha_i(\epsilon)\right] d\Delta_{iT}$$

at $\epsilon = 0$*, and noting that* $E\left[v_{i,t}\right] = 0$*, we obtain*

$$0 = \left(\int v^{\alpha}\left[\cdot;\theta_0,\alpha_i(0)\right] dF_i\right)\alpha_i^{\epsilon}(0) + \int v\left[\cdot;\theta_0,\alpha_i(0)\right] d\Delta_{iT}$$

so $\alpha_i^{\epsilon}(0) = -\left(E\left[v_{i,t}^{\alpha}\right]\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} v_{i,t}\right)$.

(A.2) $\alpha_i^{\epsilon\epsilon}(0)$ *Evaluating (A.2)*

$$\begin{aligned}0 = &\left(\int v^{\alpha\alpha}\left[\cdot;\theta_0,\alpha_i(\epsilon)\right] dF_i(\epsilon)\right)\left(\alpha_i^{\epsilon}(\epsilon)\right)^2 + \left(\int v^{\alpha}\left[\cdot;\theta_0,\alpha_i(\epsilon)\right] dF_i(\epsilon)\right)\alpha_i^{\epsilon\epsilon}(\epsilon)\\ &+ 2\left(\int v^{\alpha}\left[\cdot;\theta_0,\alpha_i(\epsilon)\right] d\Delta_{iT}\right)\alpha_i^{\epsilon}(\epsilon)\end{aligned}$$

at $\epsilon = 0$*, we obtain*

$$0 = E\left[v_{i,t}^{\alpha\alpha}\right]\left(\alpha_i^{\epsilon}(0)\right)^2 + E\left[v_{i,t}^{\alpha}\right]\alpha_i^{\epsilon\epsilon}(0) + 2\left(\int v^{\alpha}\left[\cdot;\theta_0,\alpha_i(0)\right] d\Delta_{iT}\right)\alpha_i^{\epsilon}(0)$$

so that

$$\begin{aligned}\alpha_i^{\epsilon\epsilon}(0) =& -\left(E\left[v_{i,t}^{\alpha}\right]\right)^{-1} E\left[v_{i,t}^{\alpha\alpha}\right]\left(\alpha_i^{\epsilon}(0)\right)^2 - 2\left(E\left[v_{i,t}^{\alpha}\right]\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(v_{i,t}^{\alpha} - E\left[v_{i,t}^{\alpha}\right]\right)\right)\alpha_i^{\epsilon}(0)\\ =& -\left(E\left[v_{i,t}^{\alpha}\right]\right)^{-1} E\left[v_{i,t}^{\alpha\alpha}\right]\left(E\left[v_{i,t}^{\alpha}\right]\right)^{-2}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} v_{i,t}\right)^2\\ &+ 2\left(E\left[v_{i,t}^{\alpha}\right]\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(v_{i,t}^{\alpha} - E\left[v_{i,t}^{\alpha}\right]\right)\right)\left(E\left[v_{i,t}^{\alpha}\right]\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} v_{i,t}\right)\\ =& \left(E\left[v_{i,t}^{\alpha}\right]\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} v_{i,t}\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[2 * \frac{v_{i,t}^{\alpha} - E\left[v_{i,t}^{\alpha}\right]}{E\left[v_{i,t}^{\alpha}\right]} - \frac{E\left[v_{i,t}^{\alpha\alpha}\right] v_{i,t}}{\left(E\left[v_{i,t}^{\alpha}\right]\right)^2}\right]\\ =& 2\left(E\left[v_{i,t}^{\alpha}\right]\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} v_{i,t}\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} w_{i,t}\right)\end{aligned}$$

where $w_{i,t} \equiv \frac{v_{i,t}^{\alpha} - E\left[v_{i,t}^{\alpha}\right]}{E\left[v_{i,t}^{\alpha}\right]} - \frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{2\left(E\left[v_{i,t}^{\alpha}\right]\right)^2} v_{i,t}$.

(A.3) $\alpha_i^{\epsilon\epsilon\epsilon}(\epsilon)$ *Evaluating (A.3)*

$$0 = \left(\int v^{\alpha\alpha\alpha}\left[\cdot;\theta_0,\alpha_i(\epsilon)\right]dF_i(\epsilon)\right)\left(\alpha_i^{\epsilon}(\epsilon)\right)^3 + \left(\int v^{\alpha}\left[\cdot;\theta_0,\alpha_i(\epsilon)\right]dF_i(\epsilon)\right)\alpha_i^{\epsilon\epsilon\epsilon}(\epsilon)$$
$$+ 3\left(\int v^{\alpha\alpha}\left[\cdot;\theta_0,\alpha_i(\epsilon)\right]d\Delta_{iT}\right)\left(\alpha_i^{\epsilon}(\epsilon)\right)^2 + 2\left(\int v^{\alpha}\left[\cdot;\theta_0,\alpha_i(\epsilon)\right]d\Delta_{iT}\right)\alpha_i^{\epsilon\epsilon}(\epsilon)$$

we see that

$$P\left[\max_i |\alpha_i^{\epsilon\epsilon\epsilon}(\epsilon)| > C\right] = o\left(\frac{1}{T}\right)$$

so that

$$P\left[\frac{1}{n}\sum_i |\alpha_i^{\epsilon\epsilon\epsilon}(\epsilon)| > C\right] \le P\left[\max_i |\alpha_i^{\epsilon\epsilon\epsilon}(\epsilon)| > C\right] = o\left(\frac{1}{T}\right)$$

We therefore have:

$$\begin{aligned}
\frac{1}{nT}\sum_i \sqrt{T}\left[\widehat{\alpha}_i(\theta_0) - \alpha_i(\theta_0)\right] =& \frac{1}{nT}\sum_i \left[\alpha_i^{\epsilon}(0) + \frac{1}{2}\left(\frac{1}{\sqrt{T}}\right)\alpha_i^{\epsilon\epsilon}(0) + \frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^2 \alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\right] \\
=& \frac{1}{nT}\sum_i \left\{-\left(E\left[v_{i,t}^{\alpha}\right]\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} v_{i,t}\right)\right\} \\
&+ \frac{1}{nT}\sum_i \left(\frac{1}{\sqrt{T}}\right)\left(E\left[v_{i,t}^{\alpha}\right]\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} v_{i,t}\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} w_{i,t}\right) \\
&+ \frac{1}{nT}\sum_i \left[\frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^2 \alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\right] \\
=& -\frac{1}{nT}\sum_i \frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)}{E\left[v_{i,t}^{\alpha}\right]} + O_p\left(\frac{1}{T\sqrt{T}}\right)
\end{aligned}$$

and that

$$\begin{aligned}\frac{1}{nT}\sum_i\left[\sqrt{T}\left(\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right)\right]^2 =&\frac{1}{nT}\sum_i\left[\alpha_i^{\epsilon}(0)+\frac{1}{2}\left(\frac{1}{\sqrt{T}}\right)\alpha_i^{\epsilon\epsilon}(0)+\frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^2\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\right]^2\\
=&\frac{1}{nT}\sum_i[\alpha_i^{\epsilon}(0)]^2+\frac{1}{4}\left(\frac{1}{T}\right)\frac{1}{nT}\sum_i[\alpha_i^{\epsilon\epsilon}(0)]^2+\frac{1}{36}\left(\frac{1}{T}\right)^2\frac{1}{nT}\sum_i[\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})]^2\\
&+\frac{1}{nT\sqrt{T}}\sum_i\alpha_i^{\epsilon}(0)\alpha_i^{\epsilon\epsilon}(0)+\frac{1}{3}\left(\frac{1}{\sqrt{T}}\right)^2\frac{1}{nT}\sum_i\alpha_i^{\epsilon}(0)\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\\
&+\frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^3\frac{1}{nT}\sum_i\alpha_i^{\epsilon\epsilon}(0)\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\\
=&\frac{1}{nT}\sum_i\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{(E\left[v_{i,t}^{\alpha}\right])^2}+O_p\left(\frac{1}{T\sqrt{T}}\right)\end{aligned}$$

also,

$$\begin{aligned}\frac{1}{nT}\sum_i\left[\sqrt{T}\left(\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right)\right]^3 =&\frac{1}{nT}\sum_i\left[\alpha_i^{\epsilon}(0)+\frac{1}{2}\left(\frac{1}{\sqrt{T}}\right)\alpha_i^{\epsilon\epsilon}(0)+\frac{1}{6}\left(\frac{1}{T}\right)\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\right]^3\\
=&\frac{1}{nT}\sum_i[\alpha_i^{\epsilon}(0)]^3+\frac{1}{nT^2\sqrt{T}}\sum_i[\frac{1}{2}\alpha_i^{\epsilon\epsilon}(0)]^3+\frac{1}{nT^4}\sum_i\left[\frac{1}{6}\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\right]^3\\
&+3\frac{1}{nT\sqrt{T}}\sum_i[\alpha_i^{\epsilon}(0)]^2\frac{1}{2}\alpha_i^{\epsilon\epsilon}(0)+3\frac{1}{nT^2}\sum_i\alpha_i^{\epsilon}(0)\frac{1}{4}\alpha_i^{\epsilon\epsilon}(0)\\
&+3\frac{1}{nT^2}\sum_i[\alpha_i^{\epsilon}(0)]^2\frac{1}{6}\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})+3\frac{1}{nT^3}\sum_i\alpha_i^{\epsilon}(0)\frac{1}{36}[\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})]^2\\
&+3\frac{1}{nT^3\sqrt{T}}\sum_i\frac{1}{2}\alpha_i^{\epsilon\epsilon}(0)\frac{1}{36}[\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})]^2+3\frac{1}{nT^3}\sum_i\frac{1}{4}[\alpha_i^{\epsilon\epsilon}(0)]^2\frac{1}{6}\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\\
&+6\frac{1}{nT^2\sqrt{T}}\sum_i\alpha_i^{\epsilon}(0)\frac{1}{2}\alpha_i^{\epsilon\epsilon}(0)\frac{1}{6}\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\\
=&-\frac{1}{nT}\sum_i\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^3}{(E\left[v_{i,t}^{\alpha}\right])^3}+o_p\left(\frac{1}{T\sqrt{T}}\right)\end{aligned}$$

■

Lemma 9 *(Arellano and Hahn, 2016, Theorem 5):*

$$\frac{\partial\left[\frac{1}{n}\sum_i\widehat{R}_{fi}^{*}(\widetilde{\theta})\right]}{\partial\theta}=\frac{1}{n}\sum_{i=1}^{n}E\left[U_{it}^{\alpha_i}\tilde{V}_{it}\right]+\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}E\left[U_{it}^{\alpha_i\alpha_i}\right]E\left[\left(\tilde{V}_{it}\right)^2\right]+o_p\left(1\right)$$

where $\frac{1}{T}\sum_{t=1}^{T}\tilde{V}_{it}=-\left(E\left[\frac{\partial v_i}{\partial \alpha_i}\right]\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}v_{it}\right)$.

Lemma 10 *(Rao and Mitra, 1971, Theorem 9.2.1) Let* $Y\sim N_p(\mu,\Sigma)$*, where* Σ *may be singular, then* $Y'AY\xrightarrow{d}\chi^2(k)$ *iff*

$$\Sigma A\Sigma A\Sigma=\Sigma A\Sigma \tag{3.33}$$

where $k=tr\,(A\Sigma)$.

Lemma 11 *Under condition 1 in the appendix,*

$$\widetilde{\widetilde{A}}_{f,n}=\frac{\partial^2(\frac{1}{nT}[\sum_i\sum_t \log f(y_{i,t};\widetilde{\widetilde{\theta}},\widehat{\alpha}_i(\widetilde{\widetilde{\theta}}))-\sum_i R_{fi}(\widetilde{\widetilde{\theta}})])}{\partial\theta\partial\theta'}=-\frac{1}{n}\sum_i\mathcal{I}_i$$

Proof. To prove $\widetilde{\widetilde{A}}_{f,n}=-\frac{1}{n}\sum_i\mathcal{I}_i$, it suffices to prove $-\frac{\partial^2(\frac{1}{nT}[\sum_i\sum_t \log f(y_{i,t};\widetilde{\widetilde{\theta}},\widehat{\alpha}_i(\widetilde{\widetilde{\theta}}))])}{\partial\theta\partial\theta'}=\frac{1}{n}\sum_i\mathcal{I}_i$ and $\frac{\partial^2(\frac{1}{nT}[\sum_i R_i(\widetilde{\widetilde{\theta}})])}{\partial\theta\partial\theta'}=o_p(1)$. Since

$$\begin{aligned}\frac{\partial^2(\frac{1}{nT}[\sum_i\sum_t \log f(y_{i,t};\widetilde{\widetilde{\theta}},\widehat{\alpha}_i(\widetilde{\widetilde{\theta}}))])}{\partial\theta\partial\theta'}&=\frac{1}{nT}\sum_i\sum_t\frac{\partial u(y_{i,t};\widetilde{\widetilde{\theta}},\widehat{\alpha}_i(\widetilde{\widetilde{\theta}}))}{\partial\theta}\\&=\frac{1}{nT}\sum_i\sum_t\left[u^\theta(y_{i,t};\widetilde{\widetilde{\theta}},\widehat{\alpha}_i(\widetilde{\widetilde{\theta}}))-v^\theta(y_{i,t};\widetilde{\widetilde{\theta}},\widehat{\alpha}_i(\widetilde{\widetilde{\theta}}))\left(\frac{\partial\widehat{\alpha}_i(\widetilde{\widetilde{\theta}})}{\partial\theta}\right)\right]\\&=\frac{1}{nT}\sum_i\sum_t\left[u^\theta(y_{i,t};\widetilde{\widetilde{\theta}},\widehat{\alpha}_i(\widetilde{\widetilde{\theta}}))-v^\theta(y_{i,t};\widetilde{\widetilde{\theta}},\widehat{\alpha}_i(\widetilde{\widetilde{\theta}}))\left(\frac{\frac{1}{T}\sum_t v_{it}^\theta}{\frac{1}{T}\sum_t v_{it}^{\alpha_i}}\right)\right]\\&=\frac{1}{nT}\sum_i\sum_t U_{it}^\theta\left(\widetilde{\widetilde{\theta}},\widehat{\alpha}_i(\widetilde{\widetilde{\theta}})\right)+o_p(1)\end{aligned}$$

the first and second equalities hold by the definition of $u\left(y_{it},\widetilde{\theta},\widehat{\alpha}_i(\widetilde{\theta})\right)$, the third equality follows from Lemma 7, the fourth equality hold by the definition of $U(y_{it};\theta,\alpha\,(\theta))$, and we have

$$\left|\frac{1}{nT}\sum_i\sum_t u_{it}^\theta\left(\widetilde{\widetilde{\theta}},\widehat{\alpha}_i(\widetilde{\widetilde{\theta}})\right)-\frac{1}{nT}\sum_i\sum_t u_{it}^\theta\left(\theta_0,\alpha_i\left(\theta_0\right)\right)\right| \tag{3.34}$$

$$\leq\left(\max_i\frac{1}{T}\sum_t M\left(y_{i,t}\right)\right)\times\left(|\widetilde{\widetilde{\theta}}-\theta_0|+\max_i\left|\widehat{\alpha}_i(\widetilde{\widetilde{\theta}})-\alpha_i\left(\theta_0\right)\right|\right) \tag{3.35}$$

and that

$$\left|\frac{1}{nT}\sum_i\sum_t u_{it}^{\alpha}\left(\tilde{\tilde{\theta}},\widehat{\alpha}_i(\tilde{\tilde{\theta}})\right)-\frac{1}{nT}\sum_i\sum_t u_{it}^{\alpha}\left(\theta_0,\alpha_i\left(\theta_0\right)\right)\right| \tag{3.36}$$

$$\leq\left(\max_i\frac{1}{T}\sum_t M\left(y_{i,t}\right)\right)\times\left(|\tilde{\tilde{\theta}}-\theta_0|+\max_i\left|\widehat{\alpha}_i(\tilde{\tilde{\theta}})-\alpha_i\left(\theta_0\right)\right|\right) \tag{3.37}$$

follows from 4 in Condition 1, we have $\max_i \frac{1}{T}\sum_t M\left(y_{i,t}\right)=O_p(1)$. Because of Lemma 2,

$$P\left(|\tilde{\tilde{\theta}}-\theta_0|\geq\eta\right)=o\left(T^{-1}\right)$$

for $\tilde{\tilde{\theta}}$ lies in between $\tilde{\theta}$ and θ_0. $P\left[\max_{1\leq i\leq n}|\widehat{\alpha}_i(\tilde{\tilde{\theta}})-\alpha_i\left(\theta_0\right)|\geq\eta\right]=o\left(T^{-1}\right)$ is proved in Lemma 8. Therefore we conclude that

$$\frac{1}{nT}\sum_i\sum_t U_{it}^{\theta}\left(\tilde{\tilde{\theta}},\widehat{\alpha}_i(\tilde{\tilde{\theta}})\right)=\frac{1}{nT}\sum_i\sum_t U_{it}^{\theta}\left(\theta_0,\alpha_i\left(\theta_0\right)\right)+o_p(1) \tag{3.38}$$

$$=\frac{1}{n}\sum_i E\left[U_{it}^{\theta}\right]+o_p(1)=-\frac{1}{n}\sum_i\mathcal{I}_i+o_p(1) \tag{3.39}$$

The second equality follows from Lemma 5. As it is shown in Lemma 9, $\frac{1}{n}\sum_i\frac{\partial R_{fi}(\theta_0)}{\partial\theta}=S_{fn}\left(\theta_0\right)$. According to Condition 1, we have

$$\frac{\partial^2\frac{1}{nT}\sum_i R_i(\tilde{\tilde{\theta}})}{\partial\theta\partial\theta'}=o_p(1)$$

given $(n/T)\rightarrow\rho$. ■

Lemma 12 *Suppose that*

$$K_i\left(\cdot;\theta_0,\alpha_i(\theta_0,\epsilon)\right)=\frac{\partial^{m_1+m_2}\psi\left(y_{it};\theta_0,\alpha_i(\theta_0,\epsilon)\right)}{\partial\alpha_i^m} \tag{3.40}$$

for some $m\leq 1,...,5$. *Then for any* $\eta>0$, *we have*

$$P\left[\max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}}\left|\frac{1}{n}\sum_{i=1}^n\int K_i\left(\cdot;\theta_0,\alpha_i(\theta_0,\epsilon)\right)dF_i(\epsilon)-\frac{1}{n}\sum_{i=1}^n E\left[K_i\left(y_{it};\theta_0,\alpha_{i0}\right)\right]\right|>\eta\right]=o\left(T^{-1}\right) \tag{3.41}$$

and

$$P\left[\max_i \max_{0\le\epsilon\le\frac{1}{\sqrt{T}}} \left|\int K_i\left(\cdot;\theta_0,\alpha_i(\theta_0,\epsilon)\right)dF_i(\epsilon) - E\left[K_i\left(y_{it};\theta_0,\alpha_{i0}\right)\right]\right| > \eta\right] = o\left(T^{-1}\right) \quad (3.42)$$

Also,

$$P\left[\max_i \max_{0\le\epsilon\le\frac{1}{\sqrt{T}}} \left|\int K_i\left(\cdot;\theta_0,\alpha_i(\theta_0,\epsilon)\right)d\Delta_{iT}\right| > CT^{\frac{1}{10}-v}\right] = o\left(T^{-1}\right) \quad (3.43)$$

for some constant $C>0$ and $0<v<(100q+120)^{-1}$.

Proof. Note that

$$\left\|\int K_i\left(\cdot;\theta_0,\alpha_i(\theta_0,F_i(\epsilon))\right)dF_i(\epsilon) - \int K_i\left(y_{it};\theta_0,\alpha_{i0}\right)dF_i\right\| \quad (3.44)$$

$$\le \left\|\int K_i\left(\cdot;\theta_0,\alpha_i(\theta_0,F_i(\epsilon))\right)dF_i(\epsilon) - \int K_i\left(y_{it};\theta_0,\alpha_{i0}\right)dF_i(\epsilon)\right\| \quad (3.45)$$

$$+ \left\|\int K_i\left(y_{it};\theta_0,\alpha_{i0}\right)dF_i(\epsilon) - \int K_i\left(y_{it};\theta_0,\alpha_{i0}\right)dF_i\right\| \quad (3.46)$$

$$\le \int M\left(y_{it}\right)\left(\left|\alpha_i(\theta_0,F_i(\epsilon)) - \alpha_{i0}\right|\right)d\left|F_i(\epsilon)\right| \quad (3.47)$$

$$+ \epsilon\sqrt{T}\left\|\int K_i\left(y_{it};\theta_0,\alpha_{i0}\right)d\left(\hat{F}_i - F_i\right)\right\| \quad (3.48)$$

Therefore, we have

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\int K_i\left(\cdot;\theta_0,\alpha_i(\theta_0,F_i(\epsilon))\right)dF_i(\epsilon) - \int K_i\left(y_{it};\theta_0,\alpha_{i0}\right)dF_i\right\| \quad (3.49)$$

$$\le \left(\frac{1}{n}\sum_{i=1}^{n}\left(\alpha_i(\theta_0,F_i(\epsilon)) - \alpha_{i0}\right)^2\right)^{\frac{1}{2}} \left(\frac{1}{n}\sum_{i=1}^{n}\left(E\left[M\left(y_{it}\right)\right] + \frac{1}{T}\sum_{t=1}^{T}M\left(y_{it}\right)\right)^2\right)^{\frac{1}{2}} \quad (3.50)$$

$$+ \left\|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}K_i\left(y_{it};\theta_0,\alpha_{i0}\right) - E\left[K_i\left(y_{it};\theta_0,\alpha_{i0}\right)\right]\right)\right\| \quad (3.51)$$

the RHS of which can be bounded by using Lemmas 4 and 6 in absolute value by some $\eta > 0$

with probability $1 - o(T^{-1})$. Because

$$\left| \int K_i\left(\cdot; \theta_0, \alpha_i(\theta_0, F_i(\epsilon))\right) dF_i(\epsilon) - E\left[K_i\left(y_{it}; \theta_0, \alpha_{i0}\right)\right] \right| \tag{3.52}$$

$$\leq \left|\alpha_i(\theta_0, F_i(\epsilon)) - \alpha_i\right| \cdot \left(E\left[M\left(y_{it}\right)\right] + \frac{1}{T}\sum_{t=1}^{T} M\left(y_{it}\right) \right) \tag{3.53}$$

$$+ \left| \frac{1}{T}\sum_{t=1}^{T} M\left(y_{it}\right) - E\left[M\left(y_{it}\right)\right] \right| \tag{3.54}$$

we can bound

$$\max_i \max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}} \left| \int K_i\left(\cdot; \theta_0, \alpha_i(\theta_0, F_i(\epsilon))\right) dF_i(\epsilon) - E\left[K_i\left(y_{it}; \theta_0, \alpha_{i0}\right)\right] \right| \tag{3.55}$$

in absolute value by some $\eta > 0$ with probability $1 - o(T^{-1})$. Using 4 in Condition 1 and Lemma 4, we also deduce that $\max_i \left|\int K_i\left(\cdot; \theta_0, \alpha_i(\theta_0, F_i(\epsilon))\right) d\Delta_{iT}\right|$ can be bounded by in absolute value by $CT^{\frac{1}{10}-v}$ for some constant $C > 0$ and $0 < v < \frac{1}{160}$ with probability $1 - o(T^{-1})$. ■

Lemma 13 *(Arellano and Hahn, 2016, Lemma 14)*

$$P\left[\max_i \max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}} \left|\alpha_i^{\epsilon}\left(\epsilon\right)\right| > CT^{\frac{1}{10}-v}\right] = o\left(T^{-1}\right) \tag{3.56}$$

$$P\left[\max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}} \left|\mu^{\epsilon}(\epsilon)\right| > CT^{\frac{1}{10}-v}\right] = o\left(T^{-1}\right) \tag{3.57}$$

$$P\left[\max_i \max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}} \left|\alpha_i^{\epsilon\epsilon}\left(\epsilon\right)\right| > C\left(T^{\frac{1}{10}-v}\right)^2\right] = o\left(T^{-1}\right) \tag{3.58}$$

$$P\left[\max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}} \left|\mu^{\epsilon\epsilon}(\epsilon)\right| > C\left(T^{\frac{1}{10}-v}\right)^2\right] = o\left(T^{-1}\right) \tag{3.59}$$

$$P\left[\max_i \max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}} \left|\alpha_i^{\epsilon\epsilon\epsilon}\left(\epsilon\right)\right| > C\left(T^{\frac{1}{10}-v}\right)^3\right] = o\left(T^{-1}\right) \tag{3.60}$$

$$P\left[\max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}} \left|\mu^{\epsilon\epsilon\epsilon}(\epsilon)\right| > C\left(T^{\frac{1}{10}-v}\right)^3\right] = o\left(T^{-1}\right) \tag{3.61}$$

for some constant $C > 0$ and $0 < v < (100q + 120)^{-1}$.

Lemma 14

$$\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U\left(y_{it};\theta_0,\widehat{\alpha}_i\left(\theta_0\right)\right)=\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}+\sqrt{\frac{n}{T}}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}U_{it}^{\alpha_i}\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widetilde{V}_{it}\right)$$
$$+\sqrt{\frac{n}{T}}\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}E\left[U_{it}^{\alpha_i\alpha_i}\right]\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{V}_{it}\right)^2+o_p\left(1\right)$$

where $\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)=-\left(E\left[\frac{\partial v_i}{\partial \alpha_i}\right]\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}v_{it}\right)+o_p\left(\frac{1}{\sqrt{T}}\right)=\frac{1}{T}\sum_{t=1}^{T}\tilde{V}_{it}+o_p\left(\frac{1}{\sqrt{T}}\right)$.

Proof. Let $F\equiv\left(F_1,\ldots,F_n\right)$ denote the collection of marginal distribution functions of y_{it}.Let $\hat{F}_i$ denote the empirical distribution function for the observation i. Define $F_i(\epsilon)\equiv F_i+\epsilon\sqrt{T}\left(\hat{F}_i-F_i\right)$ for $\epsilon\in\left[0,T^{-1/2}\right]$. For each fixed θ and ϵ, let $\alpha_i(\epsilon)$ be the solution to the estimating equation

$$0=\int v_i\left[\cdot;\theta_0,\alpha_i(\theta_0,F_i(\epsilon))\right]dF_i(\epsilon)$$

and let $\mu\left(F\left(\epsilon\right)\right)$ be the solution to the estimating equation

$$0=\sum_{i=1}^{n}\int\left[U_i\left(y_{it};\theta_0,\alpha_i\left(\theta_0,F_i(\epsilon)\right)\right)-\mu\left(F\left(\epsilon\right)\right)\right]dF_i(\epsilon)\tag{3.62}$$

Note that $\mu\left(F\left(0\right)\right)=0$, and

$$\mu\left(\widehat{F}\right)\equiv\mu\left(F\left(\frac{1}{\sqrt{T}}\right)\right)=\frac{1}{n}\sum_{i=1}^{n}U_i\left(y_{it};\theta_0,\alpha_i\left(\theta_0,F_i\left(\frac{1}{\sqrt{T}}\right)\right)\right)\tag{3.63}$$
$$=\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}U\left(y_{it};\theta_0,\widehat{\alpha}_i\left(\theta_0\right)\right)\tag{3.64}$$

By a Taylor series expansion, we have

$$\mu\left(\widehat{F}\right)-\mu\left(F\right)=\frac{1}{\sqrt{T}}\mu^{\epsilon}(0)+\frac{1}{2}\left(\frac{1}{\sqrt{T}}\right)^2\mu^{\epsilon\epsilon}(0)+\frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^3\mu^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\tag{3.65}$$

where $\mu^{\epsilon}(\epsilon)\equiv d\mu(F(\epsilon))/d\epsilon,\mu^{\epsilon\epsilon}(\epsilon)\equiv d^2\mu(F(\epsilon))/d\epsilon^2,\ldots$, and $\widetilde{\epsilon}$ is somewhere in between 0 and $\frac{1}{\sqrt{T}}$.

(C.1) $\mu^{\epsilon}(0)$ Let

$$h_i^{\mu}(\cdot,\epsilon)\equiv U_i\left(y_{it};\theta_0,\alpha_i\left(\theta_0,F_i(\epsilon)\right)\right)-\mu(F(\epsilon))\tag{3.66}$$

The first order condition could be written as

$$0 = \frac{1}{n}\sum_{i=1}^{n}\int h_i^{\mu}(\cdot,\epsilon)dF_i(\epsilon)$$

Differentiating repeatedly with respect to ϵ, we obtain

$$0 = \frac{1}{n}\sum_{i=1}^{n}\int \frac{dh_i^{\mu}(\cdot,\epsilon)}{d\epsilon}dF_i(\epsilon) + \frac{1}{n}\sum_{i=1}^{n}\int h_i^{\mu}(\cdot,\epsilon)d\Delta_{iT} \tag{C.1}$$

$$0 = \frac{1}{n}\sum_{i=1}^{n}\int \frac{d^2h_i^{\mu}(\cdot,\epsilon)}{d\epsilon^2}dF_i(\epsilon) + 2\frac{1}{n}\sum_{i=1}^{n}\int \frac{dh_i^{\mu}(\cdot,\epsilon)}{d\epsilon}d\Delta_{iT} \tag{C.2}$$

$$0 = \frac{1}{n}\sum_{i=1}^{n}\int \frac{d^3h_i^{\mu}(\cdot,\epsilon)}{d\epsilon^3}dF_i(\epsilon) + 3\frac{1}{n}\sum_{i=1}^{n}\int \frac{d^2h_i^{\mu}(\cdot,\epsilon)}{d\epsilon^2}d\Delta_{iT} \tag{C.3}$$

where $\Delta_{iT} \equiv \sqrt{T}\left(\hat{F}_i - F_i\right)$. Evaluating (C.1)

$$0 = \frac{1}{n}\sum_{i=1}^{n}\int \left[U_i^{\alpha_i}\left(y_{it};\theta_0,\alpha_i\left(\theta_0,F_i(\epsilon)\right)\right)\alpha_i^{\epsilon}\left(\theta_0,F_i(\epsilon)\right) - \mu^{\epsilon}(F(\epsilon))\right]dF_i(\epsilon) \tag{3.67}$$

$$+ \frac{1}{n}\sum_{i=1}^{n}\int \left[U_i\left(y_{it};\theta_0,\alpha_i\left(\theta_0,F_i(\epsilon)\right)\right) - \mu(F(\epsilon))\right]d\Delta_{iT} \tag{3.68}$$

at $\epsilon = 0$, and noting that $E\left[U_i^{\alpha_i}\right] = 0$, combining Lemma 8, we obtain

$$\mu^{\epsilon}(0) = \frac{1}{n}\sum_{i=1}^{n}\int U_i d\Delta_{iT} \tag{3.69}$$

(C.2) $\mu^{\epsilon\epsilon}(0)$ Evaluating (C.2)

$$\begin{aligned}
0 = &-\frac{1}{n}\sum_{i=1}^{n}\int \mu^{\epsilon\epsilon}(F(\epsilon))dF_i(\epsilon)\\
&+\frac{1}{n}\sum_{i=1}^{n}\int U_i^{\alpha_i\alpha_i}\left(y_{it};\theta_0,\alpha_i\left(\theta_0,F_i(\epsilon)\right)\right)\left(\alpha_i^{\epsilon}(\theta_0,F_i(\epsilon))\right)^2 dF_i(\epsilon)\\
&+\frac{1}{n}\sum_{i=1}^{n}\int U_i^{\alpha_i}\left(y_{it};\theta_0,\alpha_i\left(\theta_0,F_i(\epsilon)\right)\right)\alpha_i^{\epsilon\epsilon}(\theta_0,F_i(\epsilon))dF_i(\epsilon)\\
&+\frac{2}{n}\sum_{i=1}^{n}\int \left[U_i^{\alpha_i}\left(y_{it};\theta_0,\alpha_i\left(\theta_0,F_i(\epsilon)\right)\right)\alpha_i^{\epsilon}(\theta_0,F_i(\epsilon)) - \mu^{\epsilon}(F(\epsilon))\right]d\Delta_{iT}
\end{aligned}$$

at $\epsilon = 0$, and noting that $E\left[U_i^{\alpha_i}\right] = 0$, combining Lemma 8, we obtain

$$\begin{aligned}
\mu^{\epsilon\epsilon}(0) =& \frac{1}{n}\sum_{i=1}^{n} E\left[U_i^{\alpha_i\alpha_i}\right](\alpha_i^{\epsilon})^2 + \frac{2}{n}\sum_{i=1}^{n}(\int U_i^{\alpha_i}\left(y_{it};\theta_0,\alpha_i\left(\theta_0\right)\right)d\Delta_{iT})\alpha_i^{\epsilon}(\theta_0, F_i(0)) \\
=& \frac{1}{n}\sum_{i=1}^{n} E\left[U_i^{\alpha_i\alpha_i}\right]\left[\left(E\left[\frac{\partial v_i}{\partial \alpha_i}\right]\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} v_{it}\right)\right]^2 \\
& - \frac{2}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} U_{it}^{\alpha_i}\right)\left(E\left[\frac{\partial v_i}{\partial \alpha_i}\right]\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} v_{it}\right)
\end{aligned}$$

(C.3) We can ignore $\frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^3\mu^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})$ according to Lemma 13.

■

Lemma 15 $\frac{1}{nT}\sum_i\left(\sum_t v_{i,t}\right)\left(\widehat{\alpha}_i\left(\theta_0\right) - \alpha_i\left(\theta_0\right)\right) = -\frac{1}{nT}\sum_i \frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{E\left[v_{i,t}^{\alpha}\right]} + O_p\left(\frac{1}{T\sqrt{T}}\right)$

Proof. From Lemma 8, we deduce that

$$\begin{aligned}
\frac{1}{nT}\sum_i\left(\sum_t v_{i,t}\right)\left[\widehat{\alpha}_i\left(\theta_0\right) - \alpha_i\left(\theta_0\right)\right] =& \frac{1}{nT}\sum_i\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\sqrt{T}\left[\widehat{\alpha}_i\left(\theta_0\right) - \alpha_i\left(\theta_0\right)\right] \\
=& \frac{1}{nT}\sum_i\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\left[-\frac{\frac{1}{\sqrt{T}}\sum_t v_{i,t}}{E\left[v_{i,t}^{\alpha}\right]}\right. \\
& \left. + \frac{1}{\sqrt{T}}\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\left(\frac{1}{\sqrt{T}}\sum_t w_{i,t}\right)}{E\left[v_{i,t}^{\alpha}\right]} + \frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^2\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\right] \\
=& \frac{1}{nT}\sum_i\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\left[-\frac{\frac{1}{\sqrt{T}}\sum_t v_{i,t}}{E\left[v_{i,t}^{\alpha}\right]}\right] \qquad \text{(B1)} \\
& + \frac{1}{nT}\sum_i\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\left[\frac{1}{\sqrt{T}}\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\left(\frac{1}{\sqrt{T}}\sum_t w_{i,t}\right)}{E\left[v_{i,t}^{\alpha}\right]}\right] \\
& \qquad \text{(B2)} \\
& + \frac{1}{nT}\sum_i\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\left[\frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^2\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\right] \qquad \text{(B3)}
\end{aligned}$$

It can be shown that

$$(\text{B1}) = -\frac{1}{nT}\sum_i \frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{E\left[v_{i,t}^{\alpha}\right]}$$

$$(\text{B2}) = O_p\left(\frac{1}{T\sqrt{T}}\right)$$

$$(\text{B3}) = o_p\left(\frac{1}{T\sqrt{T}}\right)$$

Since $\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right) = O_p(1)$, and $\frac{1}{n}\sum_i \alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon}) = o_p\left(\frac{1}{T}\right)$, $\frac{1}{nT}\sum_i \alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon}) = o_p\left(\frac{1}{T^2}\right)$,

$$\frac{1}{nT}\sum_i \left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\left[\frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^2 \alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\right] = O_p(1) * \frac{1}{T^2}\frac{1}{n}\sum_i \alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon}) = o_p\left(\frac{1}{T\sqrt{T}}\right)$$

Therefore we have

$$\frac{1}{nT}\sum_i \left(\sum_t v_{i,t}\right)\left(\widehat{\alpha}_i(\theta_0) - \alpha_i(\theta_0)\right) = -\frac{1}{nT}\sum_i \frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{E\left[v_{i,t}^{\alpha}\right]} + O_p\left(\frac{1}{T\sqrt{T}}\right)$$

■

Lemma 16 $\frac{1}{nT}\sum_i \left(\sum_t v_{i,t}\right)\left(\widehat{\alpha}_i(\theta_0) - \alpha_i(\theta_0)\right) = -\frac{1}{nT}\sum_i \frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{E\left[v_{i,t}^{\alpha}\right]} + O_p\left(\frac{1}{T\sqrt{T}}\right)$

Proof. From Lemma 8, we deduce that

$$\begin{aligned}
\frac{1}{nT}\sum_i \left(\sum_t v_{i,t}\right)\left[\widehat{\alpha}_i(\theta_0) - \alpha_i(\theta_0)\right] =& \frac{1}{nT}\sum_i \left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\sqrt{T}\left[\widehat{\alpha}_i(\theta_0) - \alpha_i(\theta_0)\right] \\
=& \frac{1}{nT}\sum_i \left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\left[-\frac{\frac{1}{\sqrt{T}}\sum_t v_{i,t}}{E\left[v_{i,t}^{\alpha}\right]}\right. \\
&+ \frac{1}{\sqrt{T}}\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\left(\frac{1}{\sqrt{T}}\sum_t w_{i,t}\right)}{E\left[v_{i,t}^{\alpha}\right]} + \left.\frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^2 \alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\right] \\
=& \frac{1}{nT}\sum_i \left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\left[-\frac{\frac{1}{\sqrt{T}}\sum_t v_{i,t}}{E\left[v_{i,t}^{\alpha}\right]}\right] \qquad (\text{B1}) \\
&+ \frac{1}{nT}\sum_i \left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\left[\frac{1}{\sqrt{T}}\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\left(\frac{1}{\sqrt{T}}\sum_t w_{i,t}\right)}{E\left[v_{i,t}^{\alpha}\right]}\right] \qquad (\text{B2}) \\
&+ \frac{1}{nT}\sum_i \left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\left[\frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^2 \alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\right] \qquad (\text{B3})
\end{aligned}$$

It can be shown that

$$
\begin{aligned}
(\text{B1}) &= -\frac{1}{nT}\sum_i \frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{E\left[v_{i,t}^{\alpha}\right]} \\
(\text{B2}) &= O_p\left(\frac{1}{T\sqrt{T}}\right) \\
(\text{B3}) &= o_p\left(\frac{1}{T\sqrt{T}}\right)
\end{aligned}
$$

Since $\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right) = O_p(1)$, and $\frac{1}{n}\sum_i \alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon}) = o_p\left(\frac{1}{T}\right)$, $\frac{1}{nT}\sum_i \alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon}) = o_p\left(\frac{1}{T^2}\right)$,

$$
\frac{1}{nT}\sum_i \left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\left[\frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^2 \alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\right] = O_p(1) * \frac{1}{T^2}\frac{1}{n}\sum_i \alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon}) = o_p\left(\frac{1}{T\sqrt{T}}\right)
$$

Therefore we have

$$
\frac{1}{nT}\sum_i \left(\sum_t v_{i,t}\right)\left(\widehat{\alpha}_i\left(\theta_0\right) - \alpha_i\left(\theta_0\right)\right) = -\frac{1}{nT}\sum_i \frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{E\left[v_{i,t}^{\alpha}\right]} + O_p\left(\frac{1}{T\sqrt{T}}\right)
$$

■

Lemma 17 $\frac{1}{2}\frac{1}{nT}\sum_i \left(\sum_t v_{i,t}^{\alpha}\right)\left(\left(\widehat{\alpha}_i\left(\theta_0\right) - \alpha_i\left(\theta_0\right)\right)\right)^2 = \frac{1}{2}\frac{1}{nT}\sum_i \frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{E\left[v_{i,t}^{\alpha}\right]} + O_p\left(\frac{1}{T\sqrt{T}}\right)$

Proof.

$$
\frac{1}{T}\sum_t v_{i,t}^{\alpha} = E\left[v_{i,t}^{\alpha}\right] + \frac{1}{\sqrt{T}}\frac{1}{\sqrt{T}}\sum_t \left(v_{i,t}^{\alpha} - E\left[v_{i,t}^{\alpha}\right]\right)
$$

where

$$
\frac{1}{\sqrt{T}}\sum_t \left(v_{i,t}^{\alpha} - E\left[v_{i,t}^{\alpha}\right]\right) = O_p(1)
$$

according to condition 1. Then we have

$$
\frac{1}{nT}\sum_i \left(\frac{1}{T}\sum_t v_{i,t}^{\alpha}\right)\left(\sqrt{T}\left(\widehat{\alpha}_i\left(\theta_0\right) - \alpha_i\left(\theta_0\right)\right)\right)^2 \tag{3.70}
$$

$$
= \frac{1}{nT}\sum_i \left(E\left[v_{i,t}^{\alpha}\right] + \frac{1}{\sqrt{T}}\left(\frac{1}{\sqrt{T}}\sum_t \left(v_{i,t}^{\alpha} - E\left[v_{i,t}^{\alpha}\right]\right)\right)\right) \tag{3.71}
$$

$$
* \left[-\frac{\frac{1}{\sqrt{T}}\sum_t v_{i,t}}{E\left[v_{i,t}^{\alpha}\right]} + \frac{1}{\sqrt{T}}\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\left(\frac{1}{\sqrt{T}}\sum_t w_{i,t}\right)}{E\left[v_{i,t}^{\alpha}\right]} + \frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^2 \alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\right]^2 \tag{3.72}
$$

Writing

$$\left[-\frac{\frac{1}{\sqrt{T}}\sum_t v_{i,t}}{E\left[v_{i,t}^{\alpha}\right]}+\frac{1}{\sqrt{T}}\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\frac{1}{\sqrt{T}}\sum_t w_{i,t})}{E\left[v_{i,t}^{\alpha}\right]}+\frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^2\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\right]^2 \tag{3.73}$$

$$=\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{\left(E\left[v_{i,t}^{\alpha}\right]\right)^2}-\frac{2}{\sqrt{T}}\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2\left(\frac{1}{\sqrt{T}}\sum_t w_{i,t}\right)}{\left(E\left[v_{i,t}^{\alpha}\right]\right)^2}+\frac{1}{T}\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2\left(\frac{1}{\sqrt{T}}\sum_t w_{i,t}\right)^2}{\left(E\left[v_{i,t}^{\alpha}\right]\right)^2} \tag{3.74}$$

$$+\frac{1}{36}\left(\frac{1}{T}\right)^2[\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})]^2-\frac{1}{3}\left(\frac{1}{\sqrt{T}}\right)^2\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\frac{\frac{1}{\sqrt{T}}\sum_t v_{i,t}}{E\left[v_{i,t}^{\alpha}\right]} \tag{3.75}$$

$$+\frac{1}{3}\left(\frac{1}{\sqrt{T}}\right)^2\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\frac{1}{\sqrt{T}}\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\left(\frac{1}{\sqrt{T}}\sum_t w_{i,t}\right)}{E\left[v_{i,t}^{\alpha}\right]} \tag{3.76}$$

We obtain

$$\frac{1}{nT}\sum_i\left(\frac{1}{T}\sum_t v_{i,t}^{\alpha}\right)\left(\sqrt{T}\left(\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right)\right)^2 \tag{3.77}$$

$$=\frac{1}{nT}\sum_i\left[E\left[v_{i,t}^{\alpha}\right]+O_p(\frac{1}{\sqrt{T}})\right] \tag{3.78}$$

$$*\left[\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{\left(E\left[v_{i,t}^{\alpha}\right]\right)^2}-\frac{2}{\sqrt{T}}\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2\left(\frac{1}{\sqrt{T}}\sum_t w_{i,t}\right)}{\left(E\left[v_{i,t}^{\alpha}\right]\right)^2}+\frac{1}{T}\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2\left(\frac{1}{\sqrt{T}}\sum_t w_{i,t}\right)^2}{\left(E\left[v_{i,t}^{\alpha}\right]\right)^2}\right. \tag{3.79}$$

$$+\frac{1}{36}\left(\frac{1}{T}\right)^2[\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})]^2-\frac{1}{3}\left(\frac{1}{\sqrt{T}}\right)^2\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\frac{\frac{1}{\sqrt{T}}\sum_t v_{i,t}}{E\left[v_{i,t}^{\alpha}\right]} \tag{3.80}$$

$$\left.+\frac{1}{3}\left(\frac{1}{\sqrt{T}}\right)^2\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\frac{1}{\sqrt{T}}\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\left(\frac{1}{\sqrt{T}}\sum_t w_{i,t}\right)}{E\left[v_{i,t}^{\alpha}\right]}\right] \tag{3.81}$$

According to Lemma 8, we have

$$\frac{1}{nT}\sum_i\left(\frac{1}{T}\sum_t v_{i,t}^{\alpha}\right)\left(\sqrt{T}\left(\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right)\right)^2 \tag{3.82}$$

$$=\frac{1}{nT}\sum_i\left[\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{E\left[v_{i,t}^{\alpha}\right]}-\frac{2}{\sqrt{T}}\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2\left(\frac{1}{\sqrt{T}}\sum_t w_{i,t}\right)}{E\left[v_{i,t}^{\alpha}\right]}+\frac{1}{T}\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2\left(\frac{1}{\sqrt{T}}\sum_t w_{i,t}\right)^2}{E\left[v_{i,t}^{\alpha}\right]}\right. \tag{3.83}$$

$$+\frac{1}{36}\left(\frac{1}{T}\right)^2[\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})]^2E\left[v_{i,t}^{\alpha}\right]-\frac{1}{3}\left(\frac{1}{\sqrt{T}}\right)^2\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\left[\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right] \tag{3.84}$$

$$+\frac{1}{3}\left(\frac{1}{\sqrt{T}}\right)^2\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\frac{1}{\sqrt{T}}\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)\left(\frac{1}{\sqrt{T}}\sum_t w_{i,t}\right) \tag{3.85}$$

$$\left.+\frac{1}{nT}\sum_i O_p\left(\frac{1}{\sqrt{T}}\right)\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{\left(E\left[v_{i,t}^{\alpha}\right]\right)^2}+o_p\left(\frac{1}{T\sqrt{T}}\right)\right] \tag{3.86}$$

Therefore,

$$\frac{1}{nT}\sum_i\left(\frac{1}{T}\sum_t v_{i,t}^{\alpha}\right)\left(\sqrt{T}\left(\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right)\right)^2 \tag{3.87}$$

$$=\frac{1}{nT}\sum_i\left[\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{E\left[v_{i,t}^{\alpha}\right]}+\frac{1}{\sqrt{T}}\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2\left(\frac{1}{\sqrt{T}}\sum_t\left(\left(v_{i,t}^{\alpha}-E\left[v_{i,t}^{\alpha}\right]\right)-2E\left[v_{i,t}^{\alpha}\right]w_{i,t}\right)\right)}{\left(E\left[v_{i,t}^{\alpha}\right]\right)^2}\right] \tag{3.88}$$

$$+o_p\left(\frac{1}{T\sqrt{T}}\right) \tag{3.89}$$

Because

$$\begin{aligned}\left(v_{i,t}^{\alpha}-E\left[v_{i,t}^{\alpha}\right]\right)-2E\left[v_{i,t}^{\alpha}\right]w_{i,t}&=\left(v_{i,t}^{\alpha}-E\left[v_{i,t}^{\alpha}\right]\right)-2E\left[v_{i,t}^{\alpha}\right]\left(\frac{v_{i,t}^{\alpha}-E\left[v_{i,t}^{\alpha}\right]}{E\left[v_{i,t}^{\alpha}\right]}-\frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{2\left(E\left[v_{i,t}^{\alpha}\right]\right)^2}v_{i,t}\right)\\&=\left(v_{i,t}^{\alpha}-E\left[v_{i,t}^{\alpha}\right]\right)-2\left(v_{i,t}^{\alpha}-E\left[v_{i,t}^{\alpha}\right]\right)+\frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{E\left[v_{i,t}^{\alpha}\right]}v_{i,t}\\&=-\left(v_{i,t}^{\alpha}-E\left[v_{i,t}^{\alpha}\right]\right)+\frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{E\left[v_{i,t}^{\alpha}\right]}v_{i,t}\end{aligned}$$

We further write

$$\frac{1}{nT}\sum_i\left(\frac{1}{T}\sum_t v_{i,t}^{\alpha}\right)\left(\sqrt{T}\left(\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right)\right)^2 \tag{3.90}$$

$$=\frac{1}{nT}\sum_i\left[\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{E\left[v_{i,t}^{\alpha}\right]}+\frac{1}{\sqrt{T}}\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2\left(\frac{1}{\sqrt{T}}\sum_t\left(-\left(v_{i,t}^{\alpha}-E\left[v_{i,t}^{\alpha}\right]\right)+\frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{E\left[v_{i,t}^{\alpha}\right]}v_{i,t}\right)\right)}{\left(E\left[v_{i,t}^{\alpha}\right]\right)^2}\right] \tag{3.91}$$

$$+o_p\left(\frac{1}{T\sqrt{T}}\right) \tag{3.92}$$

It follows that

$$\frac{1}{2}\frac{1}{nT}\sum_i\left(\sum_t v_{i,t}^{\alpha}\right)\left(\left(\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right)\right)^2=\frac{1}{2}\frac{1}{nT}\sum_i\left(\frac{1}{T}\sum_t v_{i,t}^{\alpha}\right)\left(\sqrt{T}\left(\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right)\right)^2 \tag{3.93}$$

$$=\frac{1}{2}\frac{1}{nT}\sum_i\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{E\left[v_{i,t}^{\alpha}\right]}+O_p\left(\frac{1}{T\sqrt{T}}\right) \tag{3.94}$$

■

Lemma 18 $\frac{1}{6}\frac{1}{nT}\sum_i\left(\sum_t v_{i,t}^{\alpha\alpha}\left(y_{i,t};\theta_0,\widehat{\alpha}_i\left(\theta_0\right)\right)\right)\left(\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right)^3=-\frac{1}{6}\frac{1}{nT\sqrt{T}}\sum_i E\left[v_{i,t}^{\alpha\alpha}\right]\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^3}{\left(E\left[v_{i,t}^{\alpha}\right]\right)^3}+o_p\left(\frac{1}{T\sqrt{T}}\right)$

Proof.

$$\frac{1}{T}\sum_t v_{i,t}^{\alpha\alpha}(y_{i,t};\theta_0,\widehat{\widehat{\alpha}}_i\left(\theta_0\right))=E\left[v_{i,t}^{\alpha\alpha}\right]+O_p\left(\frac{1}{\sqrt{T}}\right)$$

Follows from Lemma 8, we write

$$\frac{1}{6}\frac{1}{nT}\sum_i\left(\sum_t v_{i,t}^{\alpha\alpha}\left(y_{i,t};\theta_0,\widehat{\alpha}_i\left(\theta_0\right)\right)\right)\left(\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right)^3 \tag{3.95}$$

$$=\frac{1}{6nT\sqrt{T}}\sum_i(\frac{1}{T}\sum v_{i,t}^{\alpha\alpha}(y_{i,t};\theta_0,\widehat{\widehat{\alpha}}_i\left(\theta_0\right)))\left(\sqrt{T}\left(\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right)\right)^3 \tag{3.96}$$

$$=\frac{1}{6nT\sqrt{T}}\sum_i\left(E\left[v_{i,t}^{\alpha\alpha}\right]+O_p\left(\frac{1}{\sqrt{T}}\right)\right)\times\left[-\frac{\frac{1}{\sqrt{T}}\sum_t v_{i,t}}{E\left[v_{i,t}^{\alpha}\right]}+O_p\left(\frac{1}{\sqrt{T}}\right)+\frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^2\alpha_i^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})\right]^3 \tag{3.97}$$

$$=-\frac{1}{6}\frac{1}{nT\sqrt{T}}\sum_i E\left[v_{i,t}^{\alpha\alpha}\right]\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^3}{(E\left[v_{i,t}^{\alpha}\right])^3}+o_p\left(\frac{1}{T\sqrt{T}}\right) \tag{3.98}$$

■

Lemma 19 *(Hahn and Newey, 2004, Lemma 22, Lemma 24, Lemma 25)*

$$E\left[\left(\frac{1}{T}\sum_{t\neq t'}^{T}v_{fi,t}v_{fi,t'}\right)\right]^2=O_p\left(1\right) \tag{3.99}$$

$$E\left[\left(\frac{1}{T\sqrt{T}}\sum_{t\neq t'}^{T}v_{fi,t}^2v_{fi,t'}\right)\right]^2=O_p\left(1\right) \tag{3.100}$$

$$E\left[\left(\frac{1}{T\sqrt{T}}\sum_{t\neq t'\neq t''\neq t}^{T}v_{fi,t}v_{fi,t'}v_{fi,t''}\right)\right]^2=O_p\left(1\right) \tag{3.101}$$

3.6.3 Propositions

Proposition 3

$$\frac{\partial\left[\sum_i\sum_t\psi_f(y_{i,t};\theta_0,\widehat{\alpha}_i(\theta_0))-\sum_i\widehat{R}_{fi}^*(\theta_0)\right]}{\partial\theta}(\tilde{\theta}-\theta_0)=\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}\right)\mathcal{I}_f^{-1}\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}\right)+o_p\left(1\right)$$

Proof. We can The first order derivative can be written as:

$$\frac{\partial\left[\frac{1}{\sqrt{nT}}\sum_i\sum_t\psi_f(y_{i,t};\theta_0,\widehat{\alpha}_i(\theta_0))-\frac{1}{\sqrt{nT}}\sum_i\widehat{R}^*_{fi}(\theta_0)\right]}{\partial\theta}$$

$$=\frac{1}{\sqrt{nT}}\sum_i\sum_t u(y_{i,t};\theta_0,\widehat{\alpha}_i(\theta_0))-\sqrt{\frac{n}{T}}S_{fn}(\theta_0)$$

$$=\frac{1}{\sqrt{nT}}\sum_i\sum_t\left[u(y_{i,t};\theta_0,\widehat{\alpha}_i(\theta_0))-v(y_{i,t};\theta_0,\widehat{\alpha}_i(\theta_0))\frac{E\left[u_{it}^{\alpha_i}\right]}{E\left[v_{it}^{\alpha_i}\right]}\right]-\sqrt{\frac{n}{T}}S_{fn}(\theta_0)$$

$$=\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U(y_{i,t};\theta_0,\widehat{\alpha}_i(\theta_0))-\sqrt{\frac{n}{T}}S_{fn}(\theta_0)$$

The first equality holds by the definition of $u(y_{i,t};\theta_0,\widehat{\alpha}_i(\theta_0))$ and $S_{fn}(\theta_0)$, the second equality follows from $v(y_{i,t};\theta_0,\widehat{\alpha}_i(\theta_0))=0$, the third equality hold by the definition of $U(y_{i,t};\theta_0,\widehat{\alpha}_i(\theta_0))$. According to Lemma 14, we deduce that

$$\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U\left(y_{it};\theta_0,\widehat{\alpha}_i(\theta_0)\right)=\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}+\sqrt{\frac{n}{T}}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}U_{it}^{\alpha_i}\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widetilde{V}_{it}\right)$$
$$+\sqrt{\frac{n}{T}}\frac{1}{2n}\sum_{i=1}^{n}E\left[U_{it}^{\alpha_i\alpha_i}\right]\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{V}_{it}\right)^2+o_p(1)$$

where

$$\widehat{\alpha}_i(\theta_0)-\alpha_i(\theta_0)=-\left(E\left[\frac{\partial v_i}{\partial\alpha_i}\right]\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}v_{it}\right)+o_p\left(\frac{1}{\sqrt{T}}\right)=\frac{1}{T}\sum_{t=1}^{T}\tilde{V}_{it}+o_p\left(\frac{1}{\sqrt{T}}\right)$$

according to Lemma 9,

$$\frac{\partial\left[\frac{1}{\sqrt{nT}}\sum_i\widehat{R}^*_{fi}(\widetilde{\theta})\right]}{\partial\theta}=\sqrt{\frac{n}{T}}S_{fn}(\theta_0)$$
$$=\sqrt{\frac{n}{T}}\frac{1}{n}\sum_{i=1}^{n}E\left[U_{it}^{\alpha_i}\tilde{V}_{it}\right]+\sqrt{\frac{n}{T}}\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}E\left[U_{it}^{\alpha_i\alpha_i}\right]E\left[\left(\tilde{V}_{it}\right)^2\right]+o_p(1)$$

we therefore have

$$\frac{\partial\left[\frac{1}{\sqrt{nT}}\sum_i\sum_t\psi_f(y_{i,t};\theta_0,\widehat{\alpha}_i(\theta_0))-\frac{1}{\sqrt{nT}}\sum_i\widehat{R}^*_{fi}(\theta_0)\right]}{\partial\theta}=\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}+o_p(1)$$

according to Lemma 2, we conclude that

$$\frac{\partial\left[\sum_i\sum_t\psi_f(y_{i,t};\theta_0,\widehat{\alpha}_i(\theta_0))-\sum_i\widehat{R}^*_{fi}(\theta_0)\right]}{\partial\theta}(\tilde{\theta}-\theta_0)=\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}\right)\mathcal{I}_f^{-1}\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}\right)+o_p(1)$$

■

Proposition 4 $\frac{1}{nT}\sum_i\sum_t[\psi_f\left(y_{it},\theta_0,\widehat{\alpha}(\theta_0)\right)-\psi_f\left(y_{it},\theta_0,\alpha_i(\theta_0)\right)]=\frac{1}{nT}\sum_i R_{fi}\left(\theta_0\right)+O_p\left(\frac{1}{T\sqrt{T}}\right)$

Proof. We decompose the left hand side of the above equation into a sum of three terms. For some $\overline{\alpha}_i\left(\theta_0\right)$ on the line segment adjoining $\widehat{\alpha}_i\left(\theta_0\right)$ and $\alpha_i\left(\theta_0\right)$,

$$
\begin{aligned}
&\frac{1}{nT}\sum_i\sum_t\left(\psi\left(y_{i,t};\theta_0,\widehat{\alpha}_i\left(\theta_0\right)\right)-\psi\left(y_{i,t};\theta_0,\alpha_i\left(\theta_0\right)\right)\right)\\
=&\frac{1}{nT}\sum_i\left(\sum_t v_{i,t}\right)\left(\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right)\\
&+\frac{1}{2}\frac{1}{nT}\sum_i\left(\sum_t v_{i,t}^{\alpha}\right)\left(\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right)^2\\
&+\frac{1}{6}\frac{1}{nT}\sum_i\left(\sum_t v_{i,t}^{\alpha\alpha}\left(y_{i,t};\theta_0,\overline{\alpha}_i\left(\theta_0\right)\right)\right)\left(\widehat{\alpha}_i\left(\theta_0\right)-\alpha_i\left(\theta_0\right)\right)^3
\end{aligned}
$$

By Lemma 15, Lemma 17 and Lemma 18, we conclude that

$$
\begin{aligned}
&\frac{1}{\sqrt{nT}}\sum_i\sum_t[\psi_f\left(y_{it},\theta_0,\widehat{\alpha}(\theta_0)\right)-\psi_f\left(y_{it},\theta_0,\alpha_i(\theta_0)\right)]\\
=&-\frac{1}{nT}\sum_i\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{E\left[v_{i,t}^{\alpha}\right]}+\frac{1}{\sqrt{T}}\frac{1}{nT}\sum_i\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2\left(\frac{1}{\sqrt{T}}\sum_t w_{i,t}\right)}{E\left[v_{i,t}^{\alpha}\right]}+o_p\left(\frac{1}{T\sqrt{T}}\right)\\
&+\frac{1}{2}\frac{1}{nT}\sum_i\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{E\left[v_{i,t}^{\alpha}\right]}+\frac{1}{\sqrt{T}}\frac{1}{2}\frac{1}{nT}\sum_i\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2\left(\frac{1}{\sqrt{T}}\sum_t\left(-\left(v_{i,t}^{\alpha}-E\left[v_{i,t}^{\alpha}\right]\right)+\frac{E\left[v_{i,t}^{\alpha,a}\right]}{E\left[v_{i,t}^{\alpha}\right]}v_{i,t}\right)\right)}{\left(E\left[v_{i,t}^{\alpha}\right]\right)^2}\\
&-\frac{1}{6}\frac{1}{nT\sqrt{T}}\sum_i E\left[v_{i,t}^{\alpha\alpha}\right]\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^3}{\left(E\left[v_{i,t}^{\alpha}\right]\right)^3}+o_p\left(\frac{1}{T\sqrt{T}}\right)
\end{aligned}
$$

and that

$$\frac{1}{\sqrt{nT}}\sum_i\sum_t[\psi_f\left(y_{it},\theta_0,\widehat{\alpha}(\theta_0)\right)-\psi_f\left(y_{it},\theta_0,\alpha_i(\theta_0)\right)]$$

$$=-\frac{1}{2}\frac{1}{nT}\sum_i\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{E\left[v_{i,t}^{\alpha}\right]}$$

$$+\frac{1}{\sqrt{T}}\frac{1}{nT}\sum_i\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2\left\{\frac{1}{\sqrt{T}}\sum_t\left[w_{i,t}+\frac{1}{2E\left[v_{i,t}^{\alpha}\right]}\left(-\left(v_{i,t}^{\alpha}-E\left[v_{i,t}^{\alpha}\right]\right)+\frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{E\left[v_{i,t}^{\alpha}\right]}v_{i,t}\right)\right]\right\}}{E\left[v_{i,t}^{\alpha}\right]}$$

$$-\frac{1}{6}\frac{1}{nT\sqrt{T}}\sum_i E\left[v_{i,t}^{\alpha\alpha}\right]\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^3}{\left(E\left[v_{i,t}^{\alpha}\right]\right)^3}+o_p\left(\frac{1}{T\sqrt{T}}\right)$$

We write

$$w_{i,t}+\frac{1}{2E\left[v_{i,t}^{\alpha}\right]}\left(-\left(v_{i,t}^{\alpha}-E\left[v_{i,t}^{\alpha}\right]\right)+\frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{E\left[v_{i,t}^{\alpha}\right]}v_{i,t}\right)-\frac{1}{6}\frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{\left(E\left[v_{i,t}^{\alpha}\right]\right)^2}v_{i,t}$$

$$=\frac{v_{i,t}^{\alpha}-E\left[v_{i,t}^{\alpha}\right]}{E\left[v_{i,t}^{\alpha}\right]}-\frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{2\left(E\left[v_{i,t}^{\alpha}\right]\right)^2}v_{i,t}-\frac{v_{i,t}^{\alpha}-E\left[v_{i,t}^{\alpha}\right]}{2E\left[v_{i,t}^{\alpha}\right]}+\frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{2\left(E\left[v_{i,t}^{\alpha}\right]\right)^2}v_{i,t}-\frac{1}{6}\frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{\left(E\left[v_{i,t}^{\alpha}\right]\right)^2}v_{i,t}$$

$$=\frac{1}{2}\frac{v_{i,t}^{\alpha}-E\left[v_{i,t}^{\alpha}\right]}{E\left[v_{i,t}^{\alpha}\right]}-\frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{6\left(E\left[v_{i,t}^{\alpha}\right]\right)^2}v_{i,t}$$

We obtain

$$\frac{1}{\sqrt{nT}}\sum_i\sum_t[\psi_f\left(y_{it},\theta_0,\widehat{\alpha}(\theta_0)\right)-\psi_f\left(y_{it},\theta_0,\alpha_i(\theta_0)\right)]$$

$$=-\frac{1}{2}\frac{1}{nT}\sum_i\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{E\left[v_{i,t}^{\alpha}\right]}$$

$$+\frac{1}{\sqrt{T}}\frac{1}{nT}\sum_i\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2\left(\frac{1}{\sqrt{T}}\sum_t\left(\frac{1}{2}\frac{v_{i,t}^{\alpha}-E\left[v_{i,t}^{\alpha}\right]}{E\left[v_{i,t}^{\alpha}\right]}-\frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{6\left(E\left[v_{i,t}^{\alpha}\right]\right)^2}v_{i,t}\right)\right)}{E\left[v_{i,t}^{\alpha}\right]}+o_p\left(\frac{1}{T\sqrt{T}}\right)$$

$$=\frac{1}{nT}\sum_i R_{fi}\left(\theta_0\right)+O_p\left(\frac{1}{T\sqrt{T}}\right)$$

■

Proposition 5 $\frac{1}{nT}\sum_i\sum_t[\psi_f\left(y_{it},\theta_0,\widehat{\alpha}(\theta_0)\right)-\psi_f\left(y_{it},\theta_0,\alpha_i(\theta_0)\right)]=\frac{1}{nT}\sum_i R_{fi}\left(\theta_0\right)+O_p\left(\frac{1}{T\sqrt{T}}\right)$

Proof. We decompose the left hand side of the above equation into a sum of three terms. For some $\overline{\alpha}_i(\theta_0)$ on the line segment adjoining $\widehat{\alpha}_i(\theta_0)$ and $\alpha_i(\theta_0)$,

$$\begin{aligned}
&\frac{1}{nT}\sum_i\sum_t\left(\psi\left(y_{i,t};\theta_0,\widehat{\alpha}_i(\theta_0)\right)-\psi\left(y_{i,t};\theta_0,\alpha_i(\theta_0)\right)\right)\\
=&\frac{1}{nT}\sum_i\left(\sum_t v_{i,t}\right)\left(\widehat{\alpha}_i(\theta_0)-\alpha_i(\theta_0)\right)\\
&+\frac{1}{2}\frac{1}{nT}\sum_i\left(\sum_t v_{i,t}^{\alpha}\right)\left(\widehat{\alpha}_i(\theta_0)-\alpha_i(\theta_0)\right)^2\\
&+\frac{1}{6}\frac{1}{nT}\sum_i\left(\sum_t v_{i,t}^{\alpha\alpha}\left(y_{i,t};\theta_0,\overline{\alpha}_i(\theta_0)\right)\right)\left(\widehat{\alpha}_i(\theta_0)-\alpha_i(\theta_0)\right)^3
\end{aligned}$$

By Lemma 15, Lemma 17 and Lemma 18, we conclude that

$$\begin{aligned}
&\frac{1}{\sqrt{nT}}\sum_i\sum_t[\psi_f\left(y_{it},\theta_0,\widehat{\alpha}(\theta_0)\right)-\psi_f\left(y_{it},\theta_0,\alpha_i(\theta_0)\right)]\\
=&-\frac{1}{nT}\sum_i\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{E\left[v_{i,t}^{\alpha}\right]}+\frac{1}{\sqrt{T}}\frac{1}{nT}\sum_i\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2\left(\frac{1}{\sqrt{T}}\sum_t w_{i,t}\right)}{E\left[v_{i,t}^{\alpha}\right]}+o_p\left(\frac{1}{T\sqrt{T}}\right)\\
&+\frac{1}{2}\frac{1}{nT}\sum_i\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{E\left[v_{i,t}^{\alpha}\right]}+\frac{1}{\sqrt{T}}\frac{1}{2}\frac{1}{nT}\sum_i\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2\left(\frac{1}{\sqrt{T}}\sum_t\left(-\left(v_{i,t}^{\alpha}-E\left[v_{i,t}^{\alpha}\right]\right)+\frac{E\left[v_{i,t}^{\alpha,a}\right]}{E\left[v_{i,t}^{\alpha}\right]}v_{i,t}\right)\right)}{\left(E\left[v_{i,t}^{\alpha}\right]\right)^2}\\
&-\frac{1}{6}\frac{1}{nT\sqrt{T}}\sum_i E\left[v_{i,t}^{\alpha\alpha}\right]\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^3}{\left(E\left[v_{i,t}^{\alpha}\right]\right)^3}+o_p\left(\frac{1}{T\sqrt{T}}\right)
\end{aligned}$$

and that

$$\begin{aligned}
&\frac{1}{\sqrt{nT}}\sum_i\sum_t[\psi_f\left(y_{it},\theta_0,\widehat{\alpha}(\theta_0)\right)-\psi_f\left(y_{it},\theta_0,\alpha_i(\theta_0)\right)]\\
=&-\frac{1}{2}\frac{1}{nT}\sum_i\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{E\left[v_{i,t}^{\alpha}\right]}\\
&+\frac{1}{\sqrt{T}}\frac{1}{nT}\sum_i\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2\left\{\frac{1}{\sqrt{T}}\sum_t\left[w_{i,t}+\frac{1}{2E\left[v_{i,t}^{\alpha}\right]}\left(-\left(v_{i,t}^{\alpha}-E\left[v_{i,t}^{\alpha}\right]\right)+\frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{E\left[v_{i,t}^{\alpha}\right]}v_{i,t}\right)\right]\right\}}{E\left[v_{i,t}^{\alpha}\right]}\\
&-\frac{1}{6}\frac{1}{nT\sqrt{T}}\sum_i E\left[v_{i,t}^{\alpha\alpha}\right]\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^3}{\left(E\left[v_{i,t}^{\alpha}\right]\right)^3}+o_p\left(\frac{1}{T\sqrt{T}}\right)
\end{aligned}$$

We write

$$
\begin{aligned}
& w_{i,t} + \frac{1}{2E\left[v_{i,t}^{\alpha}\right]}\left(-\left(v_{i,t}^{\alpha} - E\left[v_{i,t}^{\alpha}\right]\right) + \frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{E\left[v_{i,t}^{\alpha}\right]} v_{i,t}\right) - \frac{1}{6}\frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{\left(E\left[v_{i,t}^{\alpha}\right]\right)^2} v_{i,t} \\
= & \frac{v_{i,t}^{\alpha} - E\left[v_{i,t}^{\alpha}\right]}{E\left[v_{i,t}^{\alpha}\right]} - \frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{2\left(E\left[v_{i,t}^{\alpha}\right]\right)^2} v_{i,t} - \frac{v_{i,t}^{\alpha} - E\left[v_{i,t}^{\alpha}\right]}{2E\left[v_{i,t}^{\alpha}\right]} + \frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{2\left(E\left[v_{i,t}^{\alpha}\right]\right)^2} v_{i,t} - \frac{1}{6}\frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{\left(E\left[v_{i,t}^{\alpha}\right]\right)^2} v_{i,t} \\
= & \frac{1}{2}\frac{v_{i,t}^{\alpha} - E\left[v_{i,t}^{\alpha}\right]}{E\left[v_{i,t}^{\alpha}\right]} - \frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{6\left(E\left[v_{i,t}^{\alpha}\right]\right)^2} v_{i,t}
\end{aligned}
$$

We obtain

$$
\begin{aligned}
& \frac{1}{\sqrt{nT}} \sum_i \sum_t \left[\psi_f\left(y_{it}, \theta_0, \widehat{\alpha}(\theta_0)\right) - \psi_f\left(y_{it}, \theta_0, \alpha_i(\theta_0)\right)\right] \\
= & -\frac{1}{2}\frac{1}{nT}\sum_i \frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2}{E\left[v_{i,t}^{\alpha}\right]} \\
& + \frac{1}{\sqrt{T}}\frac{1}{nT}\sum_i \frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{i,t}\right)^2 \left(\frac{1}{\sqrt{T}}\sum_t \left(\frac{1}{2}\frac{v_{i,t}^{\alpha} - E\left[v_{i,t}^{\alpha}\right]}{E\left[v_{i,t}^{\alpha}\right]} - \frac{E\left[v_{i,t}^{\alpha\alpha}\right]}{6\left(E\left[v_{i,t}^{\alpha}\right]\right)^2} v_{i,t}\right)\right)}{E\left[v_{i,t}^{\alpha}\right]} + o_p\left(\frac{1}{T\sqrt{T}}\right) \\
= & \frac{1}{nT}\sum_i R_{fi}\left(\theta_0\right) + O_p\left(\frac{1}{T\sqrt{T}}\right)
\end{aligned}
$$

■

Proposition 6 $\frac{1}{\sqrt{n}}\sum_i \left[R_{fi}(\theta_0) - \widehat{R}_{fi}^{*}(\theta_0)\right] \xrightarrow{d} N(0, \sigma_{U_f}^2)$, *where* $\sigma_{U_f}^2 = \lim_{n\to\infty} \frac{1}{2n}\sum_i \frac{\left(E\left[v_{fi,t}^2\right]\right)^2}{\left(E\left[v_{fi,t}^{\alpha}\right]\right)^2}$.

Proof. We denote

$$
V_{fi,T} = -\frac{1}{2\sqrt{n}}\frac{\left(\frac{1}{\sqrt{T}}\sum_t v_{fi,t}\right)^2}{E\left[v_{fi,t}^{\alpha}\right]} + \frac{1}{2}\frac{\frac{1}{T}\sum_t v_{fi,t}^2}{E\left[v_{fi,t}^{\alpha}\right]} = -\frac{1}{T\sqrt{n}}\frac{\sum_{t=2}^{T} v_{fi,t}\left(\sum_{t'=1}^{t-1} v_{fi,t'}\right)}{E\left[v_{fi,t}^{\alpha}\right]}
$$

from which we have

$$
\frac{1}{\sqrt{n}}\sum_i \left[R_{fi}\left(\theta_0\right) - \widehat{R}_{fi}^{*}(\theta_0)\right] = \sum_i V_{fi,T}
$$

Observe that, for $t \geq 2$, let $N_{ft,t',i} = \frac{v_{fi,t}v_{fi,t'}}{E\left[v_{fi,t}^{\alpha}\right]}$. Then,

$$\begin{aligned}
\frac{1}{\sigma_{U_f}^2}\sum_i E\left[V_{fi,T}^2\right] =& \frac{1}{nT^2\sigma_{U_f}^2}\sum_i E\left[\left(\frac{\sum_{t=2}^{T} v_{fi,t}\left(\sum_{t'=1}^{t-1} v_{fi,t'}\right)}{E\left[v_{fi,t}^{\alpha}\right]}\right)^2\right] \\
=& \frac{1}{nT^2\sigma_{U_f}^2}\sum_i E\left[\sum_{t=2}^{T}\sum_{t'=1}^{t-1} N_{ft,t',i}^2\right] \\
&+ \frac{1}{nT^2\sigma_{U_f}^2}\sum_i E\left[\sum_{t\neq t'\neq t''\neq t}^{T} N_{ft,t',i}N_{ft,t'',i}\right] \\
&+ \frac{1}{nT^2\sigma_{U_f}^2}C\sum_i E\left[\sum_{t\neq t'\neq t''\neq t'''\neq t}^{T} N_{ft,t',i}N_{ft'',t''',i}\right]
\end{aligned}$$

where C is some positive integers which is not important for the analysis. The first equality holds by the definition of $V_{fi,T}$. The second equality holds because for any $t \neq t' \neq t'' \neq t$, $E\left[N_{ft,t',i}.N_{ft,t'',i}\right] = 0$ and for any $t \neq t' \neq t'' \neq t''' \neq t$, $E\left[N_{ft,t',i}N_{ft'',t''',i}\right] = 0$. Consider

$$E\left[\frac{1}{nT^2\sigma_{U_f}^2}\sum_i\sum_{t=2}^{T}\sum_{t'=1}^{t-1} N_{ft,t',i}^2\right] = \frac{1}{nT^2\sigma_{U_f}^2}\sum_{t=1}^{T-1}(T-t)\sum_i \frac{\left(E\left[v_{fi,t}^2\right]\right)^2}{\left(E\left[v_{fi,t}^{\alpha}\right]\right)^2} = 1$$

Also consider:

$$\begin{aligned}
&E\left[\frac{1}{nT^2\sigma_{U_f}^2}\sum_i\sum_{t=1}^{T-1}(T-t)\left(N_{ft,t',i}^2 - E\left(N_{ft,t',i}^2\right)\right)\right]^2 \\
=&\frac{1}{n^2T^4\sigma_{U_f}^4}\sum_i\sum_{t=1}^{T-1}(T-t)^2\left(E\left[N_{ft,t',i}^4\right] - \left[E\left(N_{ft,t',i}^2\right)\right]^2\right) \\
\leq&\frac{1}{n^2T\sigma_{U_f}^4}E\left[\sum_i\left[\frac{\left(\frac{1}{T}\sum_t v_{fi,t}^2\right)^4}{\left(E\left[v_{fi,t}^{\alpha}\right]\right)^4}\right]\right] \\
=&o_p(1)
\end{aligned}$$

the first equality holds under condition 1, the first inequality is because that $\left[E\left(N_{ft,t',i}^2\right)\right]^2 \geq 0$, and $\sum_{t=1}^{T-1}(T-t)^2 \leq T^3$. Then, we use Lyapunov Central Limit Theorem to show that

$$\frac{1}{\sqrt{n}}\sum_i\left[R_{fi}(\theta_0) - \widehat{R}_{fi}^{*}(\theta_0)\right] \xrightarrow{d} N(0,\sigma_{U_f}^2) \tag{3.102}$$

for this purpose, it is sufficient to verify the following Lyapunov condition: as $n \to \infty$,

$$\sigma_{U_f}^{-3} \sum_i E\left(|V_{fi,T}|^3\right) \to 0 \tag{3.103}$$

We deduce that

$$\begin{aligned}
\sigma_{U_f}^{-3} \sum_i E\left(|V_{fi,T}|^3\right) =& n^{-\frac{3}{2}} T^{-3} \sigma_{U_f}^{-3} \sum_i E\left[\left|\frac{\sum_{t=2}^{T} v_{fi,t}\left(\sum_{t'=1}^{t-1} v_{fi,t'}\right)}{E\left[v_{fi,t}^{\alpha}\right]}\right|^3\right] \\
\leq& \frac{1}{n^{\frac{3}{2}} T^3 \sigma_{U_f}^3} \sum_i E\left[\sum_{t=2}^{T} \sum_{t'=1}^{t-1} |N_{ft,t',i}|^3\right] \\
&+ \frac{1}{n^{\frac{3}{2}} T^3 \sigma_{U_f}^3} C_1 \sum_i E\left[\sum_{t\neq t'\neq t''\neq t}^{T} |N_{ft,t',i}|\, N_{ft,t'',i}^2\right] \\
&+ \frac{1}{n^{\frac{3}{2}} T^3 \sigma_{U_f}^3} C_2 \sum_i E\left[\sum_{t\neq t'\neq t''\neq t'''\neq t}^{T} |N_{ft,t',i} N_{ft,t'',i} N_{ft,t''',i}|\right] \\
&+ \frac{1}{n^{\frac{3}{2}} T^3 \sigma_{U_f}^3} C_3 \sum_i E\left[\sum_{t\neq t'\neq t''\neq t}^{T} |N_{ft,t',i} N_{ft,t'',i} N_{ft',t'',i}|\right] \\
&+ \frac{1}{n^{\frac{3}{2}} T^3 \sigma_{U_f}^3} C_4 \sum_i E\left[\sum_{t\neq t'\neq t''\neq t'''\neq t}^{T} \left|N_{ft,t',i}^2 N_{ft'',t''',i}\right|\right] \\
&+ \frac{1}{n^{\frac{3}{2}} T^3 \sigma_{U_f}^3} C_5 \sum_i E\left[\sum_{t\neq t'\neq t''\neq t'''\neq t''''\neq t}^{T} |N_{ft,t',i} N_{ft,t'',i} N_{ft''',t'''',i}|\right] \\
&+ \frac{1}{n^{\frac{3}{2}} T^3 \sigma_{U_f}^3} C_6 \sum_i E\left[\sum_{t\neq t'\neq t''\neq t'''\neq t''''\neq t'''''\neq t}^{T} |N_{ft,t',i} N_{ft'',t''',i} N_{ft'''',t''''',i}|\right]
\end{aligned}$$

The inequality follows from the property of absolute value. By the definition of $N_{ft,t',i}$,

$$\frac{1}{n^{\frac{3}{2}} T^3 \sigma_{U_f}^3} \sum_i E\left[\sum_{t=2}^{T} \sum_{t'=1}^{t-1} |N_{ft,t',i}|^3\right] \leq \frac{1}{n^{\frac{3}{2}} T^3 \sigma_{U_f}^3} \sum_i \sum_{t=2}^{T} \sum_{t'=1}^{t-1} \sup_i \left(E\left[M\left(y_{i,t}\right)\right]\right)^3 \tag{3.104}$$

$$\leq \frac{1}{n^{\frac{3}{2}} T \sigma_{U_f}^3} \sum_i E\left[M^3\left(y_{i,t}\right)\right] = o_p(1) \tag{3.105}$$

The first inequality holds under condition 1. The second inequality follows from $\sum_{t=1}^{T-1}(T-t) \leq$

T^2, we also have

$$\frac{1}{n^{\frac{3}{2}}T^3\sigma_{U_f}^3}\sum_i E\left[\sum_{t\neq t'\neq t''\neq t}^{T}\left|N_{ft,t',i}\right|N_{ft,t'',i}^2\right] \leq \frac{1}{n^{\frac{3}{2}}T^3\sigma_{U_f}^3}\sum_i\sum_{t\neq t'\neq t''\neq t}^{T}\sup_i\left(E\left[M\left(y_{i,t}\right)\right]\right)^3$$
$$= \frac{T\left(T-1\right)\left(T-2\right)}{n^{\frac{3}{2}}T^3\sigma_{U_f}^3}\sum_i\sup_i\left(E\left[M\left(y_{i,t}\right)\right]\right)^3 = o_p\left(1\right)$$

$$\frac{1}{n^{\frac{3}{2}}T^3\sigma_{U_f}^3}\sum_i E[\sum_{t\neq t'\neq t''\neq t}^{T}\left|N_{ft,t',i}N_{ft,t'',i}N_{ft',t'',i}\right|] \leq \frac{1}{n^{\frac{3}{2}}T^3\sigma_{U_f}^3}\sum_i\sum_{t\neq t'\neq t''\neq t}^{T}\sup_i\left(E\left[M\left(y_{i,t}\right)\right]\right)^3$$
$$= \frac{T\left(T-1\right)\left(T-2\right)}{n^{\frac{3}{2}}T^3\sigma_{U_f}^3}\sum_i\sup_i\left(E\left[M\left(y_{i,t}\right)\right]\right)^3 = o_p\left(1\right)$$

According to Lemma 19, we know that

$$E\left[\left(\frac{1}{T\sqrt{T}}\sum_{t\neq t'}^{T}v_{fi,t}^2 v_{fi,t'}\right)\right]^2 = O_p\left(1\right) \tag{3.106}$$

$$E\left[\left(\frac{1}{T\sqrt{T}}\sum_{t\neq t'\neq t''\neq t}^{T}v_{fi,t}v_{fi,t'}v_{fi,t''}\right)\right]^2 = O_p\left(1\right) \tag{3.107}$$

we deduce that

$$\frac{1}{n^{\frac{3}{2}}T^3\sigma_{U_f}^3}\sum_i E\left[\sum_{t\neq t'\neq t''\neq t'''\neq t}^{T}\left|N_{ft,t',i}N_{ft,t'',i}N_{ft,t''',i}\right|\right]$$
$$= \frac{1}{n^{\frac{3}{2}}T^3\sigma_{U_f}^3}\sum_i\sum_{t\neq t'\neq t''\neq t'''\neq t}^{T}E\left[\left|N_{ft,t',i}N_{ft,t'',i}N_{ft,t''',i}\right|\right]$$
$$\leq \frac{1}{n^{\frac{3}{2}}\sigma_{U_f}^3}\sum_i\frac{1}{\left[E\left[v_{fi,t}^{\alpha}\right]\right]^6}E\left[\left|\left(\frac{1}{T\sqrt{T}}\sum_{t\neq t'}^{T}v_{fi,t}^2 v_{fi,t'}\right)\left(\frac{1}{T\sqrt{T}}\sum_{t\neq t''\neq t'''\neq t}^{T}v_{fi,t}v_{fi,t''}v_{fi,t'''}\right)\right|\right]$$
$$\leq \frac{1}{n^{\frac{3}{2}}\sigma_{U_f}^3}\sum_i\frac{1}{\left[E\left[v_{fi,t}^{\alpha}\right]\right]^6}\left(E\left[\left(\frac{1}{T^{\frac{3}{2}}}\sum_{t\neq t'}^{T}v_{fi,t}^2 v_{fi,t'}\right)\right]^2\right)^{\frac{1}{2}}\left(E\left[\left(\frac{1}{T^{\frac{3}{2}}}\sum_{t\neq t''\neq t'''\neq t}^{T}v_{fi,t}v_{fi,t''}v_{fi,t'''}\right)\right]^2\right)^{\frac{1}{2}}$$
$$= o_p\left(1\right)$$

The first inequality holds by the definition of $N_{ft,t',i}$. The second inequality follows from

Holder's inequality. Likewise, we have

$$
\begin{aligned}
&\frac{1}{n^{\frac{3}{2}}T^{3}\sigma_{U_f}^{3}}\sum_{i}E\left[\sum_{t\neq t'\neq t''\neq t'''\neq t}^{T}\left|N_{ft,t',i}^{2}N_{ft'',t''',i}\right|\right]\\
\leq&\frac{1}{n^{\frac{3}{2}}\sigma_{U_f}^{3}}\sum_{i}\frac{1}{\left[E\left[v_{fi,t}^{\alpha}\right]\right]^{6}}E\left[\left|\left(\frac{1}{T^{\frac{3}{2}}}\sum_{t'\neq t''\neq t''\neq t'}^{T}v_{fi,t'}v_{fi,t''}v_{fi,t'''}\right)\left(\frac{1}{T^{\frac{3}{2}}}\sum_{t\neq t'}^{T}v_{fi,t}^{2}v_{fi,t'}\right)\right|\right]\\
\leq&\frac{1}{n^{\frac{3}{2}}\sigma_{U_f}^{3}}\sum_{i}\frac{1}{\left[E\left[v_{fi,t}^{\alpha}\right]\right]^{6}}\left(E\left[\left(\frac{1}{T^{\frac{3}{2}}}\sum_{t\neq t'}^{T}v_{fi,t}^{2}v_{fi,t'}\right)\right]^{2}\right)^{\frac{1}{2}}\left(E\left[\left(\frac{1}{T^{\frac{3}{2}}}\sum_{t'\neq t''\neq t''\neq t'}^{T}v_{fi,t'}v_{fi,t''}v_{fi,t'''}\right)\right]^{2}\right)^{\frac{1}{2}}\\
=&o_{p}\left(1\right)
\end{aligned}
$$

and that

$$
\begin{aligned}
&\frac{1}{n^{\frac{3}{2}}T^{3}\sigma_{U_f}^{3}}\sum_{i}E\left[\sum_{t\neq t'\neq t''\neq t'''\neq t''''\neq t}^{T}\left|N_{ft,t',i}N_{ft,t'',i}N_{ft''',t'''',i}\right|\right]\\
\leq&\frac{1}{n^{\frac{3}{2}}\sigma_{U_f}^{3}}\sum_{i}\frac{1}{\left[E\left[v_{fi,t}^{\alpha}\right]\right]^{6}}E\left[\left|\left(\frac{1}{T^{\frac{3}{2}}}\sum_{t'\neq t''\neq t'''\neq t'}^{T}v_{fi,t'}v_{fi,t''}v_{fi,t'''}\right)\left(\frac{1}{T^{\frac{3}{2}}}\sum_{t\neq t''''}^{T}v_{fi,t}^{2}v_{fi,t''''}\right)\right|\right]\\
\leq&\frac{1}{n^{\frac{3}{2}}\sigma_{U_f}^{3}}\sum_{i}\frac{1}{\left[E\left[v_{fi,t}^{\alpha}\right]\right]^{6}}\left(E\left[\left(\frac{1}{T^{\frac{3}{2}}}\sum_{t\neq t''''}^{T}v_{fi,t}^{2}v_{fi,t''''}\right)\right]^{2}\right)^{\frac{1}{2}}\left(E\left[\left(\frac{1}{T^{\frac{3}{2}}}\sum_{t'\neq t''\neq t'''\neq t'}^{T}v_{fi,t'}v_{fi,t''}v_{fi,t'''}\right)\right]^{2}\right)^{\frac{1}{2}}\\
=&o_{p}\left(1\right)
\end{aligned}
$$

we also have

$$
\begin{aligned}
&\frac{1}{n^{\frac{3}{2}}T^{3}\sigma_{U_f}^{3}}\sum_{i}E\left[\sum_{t\neq t'\neq t''\neq t'''\neq t''''\neq t'''''\neq t}^{T}\left|N_{ft,t',i}N_{ft'',t''',i}N_{ft'''',t''''',i}\right|\right] \\
\leq &\frac{1}{n^{\frac{3}{2}}\sigma_{U_f}^{3}}\sum_{i}\frac{1}{\left[E\left[v_{fi,t}^{\alpha}\right]\right]^{6}}E\left[\left|\left(\frac{1}{T^{\frac{3}{2}}}\sum_{t\neq t'\neq t''\neq t}^{T}v_{fi,t}v_{fi,t'}v_{fi,t''}\right)\right.\right. \\
&*\left.\left.\left(\frac{1}{T^{\frac{3}{2}}}\sum_{t'''\neq t''''\neq t'''''\neq t'''}^{T}v_{fi,t'''}v_{fi,t''''}v_{fi,t'''''}\right)\right|\right] \\
\leq &\frac{1}{n^{\frac{3}{2}}\sigma_{U_f}^{3}}\sum_{i}\frac{1}{\left[E\left[v_{fi,t}^{\alpha}\right]\right]^{6}}\left(E\left[\left(\frac{1}{T^{\frac{3}{2}}}\sum_{t'''\neq t''''\neq t'''''\neq t'''}^{T}v_{fi,t'''}v_{fi,t''''}v_{fi,t'''''}\right)\right]^{2}\right)^{\frac{1}{2}} \\
&*\left(E\left[\left(\frac{1}{T^{\frac{3}{2}}}\sum_{t\neq t'\neq t''\neq t}^{T}v_{fi,t}v_{fi,t'}v_{fi,t''}\right)\right]^{2}\right)^{\frac{1}{2}} \\
= &o_p(1)
\end{aligned}
$$

It follows that

$$
\sigma_{U_f}^{-3}\sum_{i}E\left(\left|V_{fi,T}\right|^{3}\right)=o_p(1) \tag{3.108}
$$

Therefore we conclude that

$$
\frac{1}{\sqrt{n}}\sum_{i}\left[R_{fi}(\theta_0)-\widehat{R}_{fi}^{*}(\theta_0)\right]=\sum_{i}V_{fi,T}\xrightarrow{d}N(0,\sigma_{U_f}^{2})
$$

■

Proposition 7 *$\frac{1}{\sqrt{n}}\sum_i\left[R_{fi}(\theta_0)-R_{gi}(\gamma_0)-\left(\widehat{R}_{fi}^{*}(\theta_0)-\widehat{R}_{gi}^{*}(\gamma_0)\right)\right]\xrightarrow{d}N(0,\sigma_U^2)$, where $\sigma_U^2=\lim_{n\to\infty}\frac{1}{2n}\sum_i\left[\frac{E\left[v_{fi,t}^2\right]}{E\left[v_{fi,t}^{\alpha}\right]}-\frac{E\left[v_{gi,t}^2\right]}{E\left[v_{gi,t}^{\alpha}\right]}\right]^2$.*

Proof. From Proposition 6, we deduce that

$$
\frac{1}{\sqrt{n}}\sum_{i}\left[R_{fi}(\theta_0)-\widehat{R}_{fi}^{*}(\theta_0)\right]\xrightarrow{d}N(0,\sigma_{U_f}^{2})
$$
$$
\frac{1}{\sqrt{n}}\sum_{i}\left[R_{gi}(\gamma_0)-\widehat{R}_{gi}^{*}(\gamma_0)\right]\xrightarrow{d}N(0,\sigma_{U_g}^{2})
$$

We denote

$$\frac{1}{\sqrt{n}}\sum_i\left[R_{fi}\left(\theta_0\right)-\widehat{R}^*_{fi}(\theta_0)\right]=\sum_i V_{fi,T} \tag{3.109}$$

$$\frac{1}{\sqrt{n}}\sum_i\left[R_{gi}\left(\gamma_0\right)-\widehat{R}^*_{gi}(\gamma_0)\right]=\sum_i V_{gi,T} \tag{3.110}$$

$$\sigma_U^2=\sigma_{U_f}^2+\sigma_{U_g}^2-2Cov(\sum_i V_{fi,T},\sum_i V_{gi,T}) \tag{3.111}$$

where

$$Cov(\sum_i V_{fi,T},\sum_i V_{gi,T})=\frac{1}{nT^2}\sum_i E\left[\frac{\sum_{t=2}^T v_{fi,t}\left(\sum_{t'=1}^{t-1} v_{fi,t'}\right)\sum_{t=2}^T v_{gi,t}\left(\sum_{t'=1}^{t-1} v_{gi,t'}\right)}{E\left[v_{fi,t}^\alpha\right]E\left[v_{gi,t}^\alpha\right]}\right] \tag{3.112}$$

$$=\frac{T-1}{2nT}\sum_i\frac{\left(E\left[v_{fi,t}^2\right]E\left[v_{gi,t}^2\right]\right)}{E\left[v_{fi,t}^\alpha\right]E\left[v_{gi,t}^\alpha\right]} \tag{3.113}$$

the first equality holds because $E\left[v_{fi,t}v_{gi,t'}v_{fi,t'}v_{gi,t''}\right]=0$ and $E\left[v_{fi,t}v_{gi,t'}v_{fi,t''}v_{gi,t'''}\right]=0$ for any $t\neq t'\neq t''\neq t'''\neq t$, the second equality follows from $\sum_{t=1}^{T-1}(T-t)=\frac{T(T-1)}{2}$. We conclude that

$$\frac{1}{\sqrt{n}}\sum_i\left[R_{fi}\left(\theta_0\right)-R_{gi}\left(\gamma_0\right)-\left(\widehat{R}^*_{fi}(\theta_0)-\widehat{R}^*_{gi}(\gamma_0)\right)\right]\xrightarrow{d}N(0,\sigma_U^2)$$

■

Proposition 8

$$\frac{1}{\sqrt{n}}\sum_i\left\{\left[R_{fi}(\theta_0)-\sum_i\widehat{R}^*_{fi}(\theta_0)\right]-\left[R_{1gi}(\theta_0)-\sum_i\widehat{R}^*_{1gi}(\theta_0)\right]-\left[R_{2gi}(\theta_0)-\sum_i\widehat{R}^*_{2gi}(\theta_0)\right]\right\}$$
$$\xrightarrow{d}N(0,\sigma^2_{U^{nested}_{t,n}})$$

where

$$\sigma^2_{U^{nested}_{t,n}}\equiv\sigma^2_{U_{1g}}=\lim_{n\to\infty}\frac{1}{2n}\sum_i\frac{\left(E\left[v_{1gi,t}^2\right]\right)^2}{\left(E\left[v_{1gi,t}^\alpha\right]\right)^2}$$

Proof. We denote $P_{fi}=\frac{1}{\sqrt{2J}}\sum_{t=1}^{t=2J}v_{fi,t}$, $P_{1gi}=\frac{1}{\sqrt{J}}\sum_{t=1}^{t=J}v_{1gi,t}$, $P_{2gi}=\frac{1}{\sqrt{J}}\sum_{t=J+1}^{t=2J}v_{2gi,t}$, $G_{fi}=E\left(\frac{1}{\sqrt{2J}}\sum_{t=1}^{t=2J}v_{fi,t}\right)^2$, $G_{1gi}=E(\frac{1}{\sqrt{J}}\sum_{t=1}^{t=J}v_{1gi,t})^2$, $G_{2gi}=E\left(\frac{1}{\sqrt{J}}\sum_{t=J+1}^{t=2J}v_{2gi,t}\right)^2$, under null hypothesis, $\alpha_{i1}=\alpha_{i2}=\alpha_i$, $E\left[v_{fi,t}^{\alpha}\right]=E\left[v_{1gi,t}^{\alpha}\right]=E\left[v_{2gi,t}^{\alpha}\right]$, $E\left(v_{2gi,t}^2\right)=E\left(v_{1gi,t}^2\right)=E\left(v_{fi,t}^2\right)$, $P_{fi}=\frac{1}{\sqrt{2}}\left(P_{1gi}+P_{2gi}\right)$, and that

$$
\begin{aligned}
&R_{fi}\left(\theta_0\right)-R_{1gi}\left(\theta_0\right)-R_{2gi}\left(\theta_0\right)-\left[\widehat{R}_{fi}^*(\theta_0)-\widehat{R}_{1gi}^*(\theta_0)-\widehat{R}_{2gi}^*(\theta_0)\right]\\
=&-\frac{1}{2}\frac{\left[\frac{1}{\sqrt{2}}\left(P_{1gi}+P_{2gi}\right)\right]^2-P_{1gi}^2-P_{2gi}^2}{E\left[v_{fi,t}^{\alpha}\right]}-\left[-\frac{1}{2}\frac{G_{fi}-G_{1gi}-G_{2gi}}{E\left[v_{fi,t}^{\alpha}\right]}\right]\\
=&\frac{\frac{1}{2}P_{1gi}^2+\frac{1}{2}P_{2gi}^2-P_{1gi}P_{2gi}-G_{fi}}{2E\left[v_{fi,t}^{\alpha}\right]}\\
=&\frac{\frac{1}{2}\left(P_{1gi}^2-G_{1gi}\right)}{2E\left[v_{fi,t}^{\alpha}\right]}+\frac{\frac{1}{2}\left(P_{2gi}^2-G_{2gi}\right)}{2E\left[v_{fi,t}^{\alpha}\right]}-\frac{P_{1gi}P_{2gi}}{2E\left[v_{fi,t}^{\alpha}\right]}
\end{aligned}
$$

From Proposition 6, we deduce that

$$
\begin{aligned}
\frac{1}{\sqrt{n}}\sum_i\left[R_{1gi}\left(\theta_0\right)-\widehat{R}_{1gi}^*(\theta_0)\right]&=\frac{1}{\sqrt{n}}\sum_i\left[-\frac{1}{2}\frac{\left(P_{1gi}^2-G_{1gi}\right)}{E\left[v_{1gi,t}^{\alpha}\right]}\right]\xrightarrow{d}N(0,\sigma_{U_{1g}}^2)\\
\frac{1}{\sqrt{n}}\sum_i\left[R_{2gi}\left(\theta_0\right)-\widehat{R}_{2gi}^*(\theta_0)\right]&=\frac{1}{\sqrt{n}}\sum_i\left[-\frac{1}{2}\frac{\left(P_{2gi}^2-G_{2gi}\right)}{E\left[v_{2gi,t}^{\alpha}\right]}\right]\xrightarrow{d}N(0,\sigma_{U_{2g}}^2)
\end{aligned}
$$

Since $E\left[P_{1gi}P_{2gi}\right]=0$,the mean becomes:

$$
E[\frac{1}{\sqrt{n}}\sum_i\left\{\left[R_{fi}(\theta_0)-\sum_i\widehat{R}_{fi}^*(\theta_0)\right]-\left[R_{1gi}(\theta_0)-\sum_i\widehat{R}_{1gi}^*(\theta_0)\right]-\left[R_{2gi}(\theta_0)-\sum_i\widehat{R}_{2gi}^*(\theta_0)\right]\right\}]=0
$$

and that

$$
\begin{aligned}
&Var\left[\frac{1}{\sqrt{n}}\sum_i(R_{fi}\left(\theta_0\right)-R_{1gi}\left(\theta_0\right)-R_{2gi}\left(\theta_0\right))\right]\\
&=\frac{1}{2}\sigma_{U_{1g}}^2+E\left[\frac{1}{\sqrt{n}}\sum_i\frac{P_{1gi}P_{2gi}}{2E\left[v_{fi,t}^{\alpha}\right]}\right]^2=\sigma_{U_{1g}}^2
\end{aligned}
\tag{3.114}
$$

where the first equality holds by $\sigma_{U_{1g}}^2=\sigma_{U_{2g}}^2$ under the null hypothesis and $E\left[P_{1gi}^3P_{2gi}\right]=0$, $E\left[P_{1gi}P_{2gi}^3\right]=0$. The second equality holds by the definition of P_{1gi},P_{1gi}, and $E\left[P_{1gi}^2\right]=E\left[P_{2gi}^2\right]$ under the null hypothesis. ■

Proposition 9 *under $H_0 : E\left[LR_{nT}\left(\theta_0, \gamma_0\right)\right] = 0$:*

$$\frac{1}{\sqrt{nT}} \sum_i \sum_t \left\{\left(\psi_f\left(y_{it}, \theta_0, \alpha_i\right) - \psi_g\left(y_{it}, \gamma_0, \lambda_i\right)\right)\right\} \xrightarrow{d} N(0, \omega^2) \tag{3.115}$$

where $\omega^2 = \frac{1}{nT} \sum_i \cdot \sum_t \left[\left(\psi_f\left(y_{it}, \theta_0, \alpha_i\right) - \psi_g\left(y_{it}, \gamma_0, \lambda_i\right)\right)\right]^2$.

Proof. *Similar to Proposition 7, we denote*

$$L_{i,T} = \frac{1}{\sqrt{nT}} \sum_t \left(\psi_f\left(y_{it}, \theta_0, \alpha_i\right) - \psi_g\left(y_{it}, \gamma_0, \lambda_i\right)\right) \tag{3.116}$$

it is sufficient to verify the following Lyapunov condition: as $n \to \infty$,

$$\omega^{-3} \sum_i E\left(\left|L_{i,T}\right|^3\right) \to 0 \tag{3.117}$$

which which is satisfied since as $n \to \infty$, let $P_{t,i} = \psi_f\left(y_{it}, \theta_0, \alpha_i\right) - \psi_g\left(y_{it}, \gamma_0, \lambda_i\right)$. Then,

$$\begin{aligned}
&\omega^{-3} \sum_i E\left(\left|L_{i,T}\right|^3\right) \\
=&n^{-\frac{3}{2}} T^{-\frac{3}{2}} \omega^{-3} \sum_i E\left[\left|\sum_t \left(\psi_f\left(y_{it}, \theta_0, \alpha_i\right) - \psi_g\left(y_{it}, \gamma_0, \lambda_i\right)\right)\right|^3\right] \\
\leq&n^{-\frac{3}{2}} T^{-\frac{3}{2}} \omega^{-3} \sum_i E\left[\sum_t \left|P_{t,i}\right|^3\right] + n^{\frac{3}{2}} T^{\frac{3}{2}} \omega^{3} \sum_i E\left[\sum_{t \neq t'} \left|P_{t,i}\right| P_{t',i}^2\right] \\
&+ n^{-\frac{3}{2}} T^{-\frac{3}{2}} \omega^{-3} \sum_i E\left[\sum_{t \neq t' \neq t'' \neq t} \left|P_{t,i} P_{t',i} P_{t'',i}\right|\right]
\end{aligned}$$

Since

$$n^{-\frac{3}{2}} T^{-\frac{3}{2}} \omega^{-3} \sum_i E\left[\sum_t \left|P_{t,i}\right|^3\right] \leq n^{-\frac{3}{2}} T^{-\frac{1}{2}} \omega^{-3} \sum_i \sup_i \left(E\left[M\left(y_{i,t}\right)\right]\right)^3 = o_p(1) \tag{3.118}$$

According to Lemma 19, we deduce that

$$\begin{aligned}
&n^{-\frac{3}{2}} \omega^{-3} \sum_i E\left[T^{-\frac{3}{2}} \sum_{t \neq t'} \left|P_{t,i}\right| P_{t',i}^2\right] \\
&\leq n^{-\frac{3}{2}} \omega^{-3} \sum_i \left(E\left[T^{-\frac{3}{2}} \sum_{t \neq t'} \left|P_{t,i}\right| P_{t',i}^2\right]^2\right)^{\frac{1}{2}} = o_p(1)
\end{aligned}$$

and that

$$n^{-\frac{3}{2}}\omega^{-3}\sum_i E\left[T^{-\frac{3}{2}}\sum_{t\neq t'\neq t''\neq t}\left|P_{t,i}P_{t',i}P_{t'',i}\right|\right]$$
$$\leq n^{-\frac{3}{2}}\omega^{-3}\sum_i\left(E\left[T^{-\frac{3}{2}}\sum_{t\neq t'\neq t''\neq t}\left|P_{t,i}P_{t',i}P_{t'',i}\right|\right]^2\right)^{\frac{1}{2}}$$
$$=o_p(1)$$

Therefore, we conclude that

$$\frac{1}{\sqrt{nT}}\sum_i\sum_t\left\{\left(\psi_f\left(y_{it},\theta_0,\alpha_i\right)-\psi_g\left(y_{it},\gamma_0,\lambda_i\right)\right)\right\}=\sum_i L_{i,T}\xrightarrow{d}N(0,\omega^2) \tag{3.119}$$

■

Proposition 10 $\frac{1}{\sqrt{n}}LR_{nT}\left(\theta_0,\gamma_0\right)+\frac{1}{\sqrt{n}}\sum_i\left[R_{fi}\left(\theta_0\right)-R_{gi}\left(\gamma_0\right)-\left(R^*_{fi}-R^*_{gi}\right)\right]\xrightarrow{d}N(0,\sigma^2_W)$, *where* $\sigma^2_W=T\omega^2+\sigma^2_U$.

Proof. We denote

$$W_{i,T}=V_{i,T}+L_{i,T} \tag{3.120}$$
$$V_{i,T}=-\frac{1}{\sqrt{n}}\frac{1}{T}\frac{\sum_{t=2}^T v_{fi,t}\left(\sum_{t'=1}^{t-1}v_{fi,t'}\right)}{E\left[v^\alpha_{fi,t}\right]}+\frac{1}{\sqrt{n}}\frac{1}{T}\frac{\sum_{t=2}^T v_{gi,t}\left(\sum_{t'=1}^{t-1}v_{gi,t'}\right)}{E\left[v^\alpha_{gi,t}\right]} \tag{3.121}$$
$$L_{i,T}=\frac{1}{\sqrt{n}}\left[\sum_t\left(\psi_f\left(y_{it},\theta_0,\alpha_i\right)-\psi_g\left(y_{it},\gamma_0,\lambda_i\right)\right)-E\left[\sum_t\left(\psi_f\left(y_{it},\theta_0,\alpha_i\right)-\psi_g\left(y_{it},\gamma_0,\lambda_i\right)\right)\right]\right] \tag{3.122}$$

Therefore we can write

$$\frac{1}{\sqrt{n}}LR_{nT}\left(\theta_0,\gamma_0\right)+\frac{1}{\sqrt{n}}\sum_i\left[R_{fi}\left(\theta_0\right)-R_{gi}\left(\gamma_0\right)-\left(R^*_{fi}-R^*_{gi}\right)\right] \tag{3.123}$$
$$=\sum_i V_{i,T}+\sum_i L_{i,T}=\sum_i W_{i,T} \tag{3.124}$$

From Proposition 7 and 9, we know that

$$\frac{1}{\sqrt{n}} LR_{nT}\left(\theta_0, \gamma_0\right) = \sum_i L_{i,T} \xrightarrow{d} N(0, T\omega^2)$$

$$\frac{1}{\sqrt{n}} \sum_i \left[R_{fi}\left(\theta_0\right) - R_{gi}\left(\gamma_0\right) - \left(\widehat{R}^*_{fi}(\theta_0) - \widehat{R}^*_{gi}(\gamma_0)\right)\right] = \sum_i V_{i,T} \xrightarrow{d} N(0, \sigma^2)$$

it is sufficient to verify the following Lyapunov condition: as $n \to \infty$,

$$\sigma_W^{-3} \sum_i E\left(\left|W_{i,T}\right|^3\right) \to 0 \tag{3.125}$$

which is satisfied since as $n \to \infty$,

$$\begin{aligned}
\sigma_W^{-3} \sum_i E\left(\left|W_{i,T}\right|^3\right) \leq & \sigma_W^{-3} \sum_i E\left(\left|V_{i,T}\right|^3 + \left|L_{i,T}\right|^3\right) \\
= & \frac{\sigma_U^3}{\sigma_W^3} \sigma_U^{-3} \sum_i E\left(\left|V_{i,T}\right|^3\right) + \frac{\left(\sqrt{T}\omega\right)^3}{\sigma_W^3} \left(\sqrt{T}\omega\right)^{-3} \sum_i E\left(\left|L_{i,T}\right|^3\right) \\
\to & 0
\end{aligned}$$

The first inequality holds by the convexity of the function $f\left(x\right) = \left|x\right|^3$. The first equality follows from Proposition 7 and 9. We denote

$$\begin{aligned}
V_{n,t} &= \left(v_{f1,t}, v_{f2,t}, \ldots v_{fn,t}, -v_{g1,t}, -v_{g2,t}, \ldots - v_{gn,t}\right)' \\
V_{n,t}^{\alpha} &= diag\left(v_{f1,t}^{\alpha}, v_{f2,t}^{\alpha}, \ldots v_{fn,t}^{\alpha}, -v_{g1,t}^{\alpha}, -v_{g2,t}^{\alpha}, \ldots - v_{gn,t}^{\alpha}\right) \\
H_n &= E\left[V_{n,t}^{\alpha}\right] \\
D_n &= E\left[V_{n,t} V_{n,t}'\right]
\end{aligned}$$

and further consider

$$
\begin{aligned}
&\sum_i E\left[L_{i,T}V_{i,T}\right] \\
=&-\frac{1}{nT}\sum_i E\left[\left[\left[\sum_t \left(\psi_f\left(y_{it},\theta_0,\alpha_i\right)-\psi_g\left(y_{it},\gamma_0,\lambda_i\right)\right)\right]\right.\right. \\
&*\left.\left.\left[\frac{\sum_{t=2}^T v_{fi,t}\left(\sum_{t'=1}^{t-1} v_{fi,t'}\right)}{E\left[v_{fi,t}^{\alpha}\right]}-\frac{1}{\sqrt{n}}\frac{1}{T}\frac{\sum_{t=2}^T v_{gi,t}\left(\sum_{t'=1}^{t-1} v_{gi,t'}\right)}{E\left[v_{gi,t}^{\alpha}\right]}\right]\right]\right] \\
=&-\frac{1}{nT}E\left[\sum_{t=2}^T\sum_{t'=1}^{t-1} LP\left(\theta_0,\gamma_0\right)V_{n,t}'\left[H_n\right]^{-1}V_{n,t'}\right] \\
=&-\frac{1}{nT}E\left[\sum_{t=2}^T\sum_{t'=1}^{t-1} \varrho_{n,t}' D_n^{1/2}\left[H_n\right]^{-1}V_{n,t'}\right]
\end{aligned}
$$

where $LP\left(\theta_0,\gamma_0\right)=\sum_i\left[\left(\psi_f\left(y_{it},\theta_0,\alpha_i\right)-\psi_g\left(y_{it},\gamma_0,\lambda_i\right)\right)\right]$, $\varrho_{n,t}=\left(D_n^{1/2}\right)^{+}Cov\left[LP\left(\theta_0,\gamma_0\right),V_{n,t}\right]$, and $\left(D_n^{1/2}\right)^{+}$ denotes the Moore-Penrose inverse of $D_n^{1/2}$. Note that

$$
\left(\left(D_n^{1/2}\right)^{+}Cov\left[LP\left(\theta_0,\gamma_0\right),V_{n,t}\right]\right)' D_n^{1/2}=Cov\left[LP\left(\theta_0,\gamma_0\right),V_{n,t}\right]'
$$

Then,

$$
\sum_i E\left[L_{i,T}V_{i,T}\right]=-\frac{1}{n}\sum_{t=1}^{T-1}\left(1-\frac{t}{T}\right)\varrho_{n,t}' D_n^{1/2}\left[H_n\right]^{-1}V_{n,t'}=o_p\left(1\right)
$$

follows from the fact that:

$$
\begin{aligned}
&E\left[\sum_i E\left[L_{i,T}V_{i,T}\right]\right]^2 \\
=&\frac{1}{n^2}\sum_{t=1}^{T-1}\left(1-\frac{t}{T}\right)^2\varrho_{n,t}' D_n^{1/2}\left[H_n\right]^{-1}D_n\left[H_n\right]^{-1}D_n^{1/2}\varrho_{n,t} \\
\leq&\frac{\kappa_{\max}^2}{n}=o_p\left(1\right)
\end{aligned}
$$

where $\kappa_{\max}^2$ is the maximum eigenvalue of matrix $\left(D_n\left[H_n\right]^{-1}\right)^2$. According to Lemma E.1 in Liao and Shi (2020), we have $\varrho_{n,t}' A\varrho_{n,t}\leq\kappa_{\max}\left(A\right)$ for any positive semi-definite matrix A. $\kappa_{\max}\left(A\right)$ is the maximum eigenvalue of matrix A. The inequality holds by $\left(1-\frac{t}{T}\right)^2\leq 1$.

Therefore, we conclude that

$$\frac{1}{\sqrt{n}}LR_{nT}\left(\theta_0,\gamma_0\right)+\frac{1}{\sqrt{n}}\sum_i\left[R_{fi}\left(\theta_0\right)-R_{gi}\left(\gamma_0\right)-\left(R^*_{fi}-R^*_{gi}\right)\right]\xrightarrow{d}N(0,\sigma^2_W)$$

■

3.6.4 Theorems

Proof of Theorem 1. From equation 3.12, we have:

$$\begin{aligned}&\frac{1}{\sqrt{nT}}[LR_{nT}\left(\tilde{\theta},\tilde{\gamma}\right)-E\left[LR_{nT}\left(\theta_0,\gamma_0\right)\right]]-\frac{1}{\sqrt{nT}}\sum_i\left[\widehat{R}_{fi}(\tilde{\theta})-\widehat{R}_{gi}(\tilde{\gamma})\right]\\ =&\frac{1}{\sqrt{nT}}[LR_{nT}\left(\theta_0,\gamma_0\right)-E\left[LR_{nT}\left(\theta_0,\gamma_0\right)\right]]+o_p(1)\end{aligned}\tag{T.1.1}$$

under $H_0 : E\left[LR_{nT}\left(\theta_0,\gamma_0\right)\right]=0$:

$$\frac{1}{\sqrt{nT}}LR_{nT}\left(\tilde{\theta},\tilde{\gamma}\right)-\frac{1}{\sqrt{nT}}\sum_i\left[\widehat{R}_{fi}(\tilde{\theta})-\widehat{R}_{gi}(\tilde{\gamma})\right]=\frac{1}{\sqrt{nT}}LR_{nT}\left(\theta_0,\gamma_0\right)+o_p(1)\tag{3.126}$$

from Proposition 9, we know that

$$\frac{1}{\sqrt{nT}}LR_{nT}(\theta_0,\gamma_0)\xrightarrow{d}N(0,\omega^2)\tag{3.127}$$

we denote $\widehat{\omega}^2_n$ as an estimator for ω^2, and $\widehat{\omega}^2_n=\frac{1}{nT}\sum_i\sum_t[\psi_f\left(\widetilde{\theta},\widehat{\alpha}_i(\widetilde{\theta})\right)-\psi_g\left(\tilde{\gamma},\widehat{\alpha}_i(\tilde{\gamma})\right)]^2$, $\omega^2=\lim_{n,T\to\infty}\frac{1}{nT}\sum_i\sum_t\left[\psi_f\left(\theta_0,\alpha_i(\theta_0)\right)-\psi_g\left(\gamma_0,\alpha_i(\gamma_0)\right)\right]^2$. The consistency results for $\widehat{\omega}^2_n$ can be similarly proved as previous part, more specifically, let

$$\omega^2_n=\frac{1}{nT}\sum_i\sum_t\left[\psi_f\left(\theta_0,\alpha_i(\theta_0)\right)-\psi_g\left(\gamma_0,\alpha_i(\gamma_0)\right)\right]^2$$

let $\frac{\partial u\left(y_{i,t};\tilde{\tilde{\tilde{\theta}}},\widehat{\alpha}_i(\tilde{\tilde{\tilde{\theta}}})\right)}{\partial\theta} = u^{\tilde{\tilde{\tilde{\theta}}}}_{fi,t}$, then we can prove that:

$$
\begin{aligned}
\widehat{\omega}_n^2 - \omega^2 =& \frac{1}{nT}\sum_i\sum_t 2\left[\psi_f\left(\theta_0,\alpha_i(\theta_0)\right) - \psi_g\left(\gamma_0,\alpha_i(\gamma_0)\right)\right] \\
& * \left\{\left[\psi_f\left(\widetilde{\theta},\widehat{\alpha}_i(\widetilde{\theta})\right) - \psi_g\left(\tilde{\gamma},\widehat{\alpha}_i(\tilde{\gamma})\right)\right] - \left[\psi_f\left(\theta_0,\alpha_i(\theta_0)\right) - \psi_g\left(\gamma_0,\alpha_i(\gamma_0)\right)\right]\right\} \\
& + \frac{1}{nT}\sum_i\sum_t 2\left\{\left[\psi_f\left(\widetilde{\theta},\widehat{\alpha}_i(\widetilde{\theta})\right) - \psi_g\left(\tilde{\gamma},\widehat{\alpha}_i(\tilde{\gamma})\right)\right] - \left[\psi_f\left(\theta_0,\alpha_i(\theta_0)\right) - \psi_g\left(\gamma_0,\alpha_i(\gamma_0)\right)\right]\right\}^2 \\
=& \frac{1}{nT}\sum_i\sum_t 2[\psi_f\left(\theta_0,\alpha_i(\theta_0)\right) - \psi_g\left(\gamma_0,\alpha_i(\gamma_0)\right)] \\
& * \{u^{\tilde{\tilde{\tilde{\theta}}}}_{fi,t}(\tilde{\tilde{\theta}} - \widehat{\theta})\left(\tilde{\theta} - \theta_0\right) + v_{fi,t}\left[\widehat{\alpha}_i(\theta_0) - \alpha_i(\theta_0)\right] + \frac{1}{2}v^{\alpha}_{fi,t}\left[\widehat{\alpha}_i(\theta_0) - \alpha_i(\theta_0)\right]^2 \\
& + \frac{1}{6}\widetilde{v}^{\alpha\alpha}_{fi,t}\left[\widehat{\alpha}_i(\theta_0) - \alpha_i(\theta_0)\right]^3 - u^{\tilde{\tilde{\tilde{\gamma}}}}_{gi,t}(\tilde{\tilde{\gamma}} - \widehat{\gamma})\left(\tilde{\gamma} - \gamma_0\right) - v_{gi,t}\left[\widehat{\alpha}_i(\gamma_0) - \alpha_i(\gamma_0)\right] \\
& - \frac{1}{2}v^{\alpha}_{gi,t}\left[\widehat{\alpha}_i(\gamma_0) - \alpha_i(\gamma_0)\right]^2 - \frac{1}{6}v^{\alpha\alpha}_{gi,t}\left[\widehat{\alpha}_i(\gamma_0) - \alpha_i(\gamma_0)\right]^3\} \\
& + \frac{1}{nT}\sum_i\sum_t 2 * \left\{u^{\tilde{\tilde{\tilde{\theta}}}}_{fi,t}(\tilde{\tilde{\theta}} - \widehat{\theta})\left(\tilde{\theta} - \theta_0\right) + v_{fi,t}\left[\widehat{\alpha}_i(\theta_0) - \alpha_i(\theta_0)\right] + \frac{1}{2}v^{\alpha}_{fi,t}\left[\widehat{\alpha}_i(\theta_0) - \alpha_i(\theta_0)\right]^2\right. \\
& + \frac{1}{6}\widetilde{v}^{\alpha\alpha}_{fi,t}\left[\widehat{\alpha}_i(\theta_0) - \alpha_i(\theta_0)\right]^3 - u^{\tilde{\tilde{\tilde{\gamma}}}}_{fi,t}(\tilde{\tilde{\gamma}} - \widehat{\gamma})\left(\tilde{\gamma} - \gamma_0\right) - v_{gi,t}\left[\widehat{\alpha}_i(\gamma_0) - \alpha_i(\gamma_0)\right] \\
& - \frac{1}{2}v^{\alpha}_{gi,t}\left[\widehat{\alpha}_i(\gamma_0) - \alpha_i(\gamma_0)\right]^2 - \frac{1}{6}v^{\alpha\alpha}_{gi,t}\left[\widehat{\alpha}_i(\gamma_0) - \alpha_i(\gamma_0)\right]^3\}^2 \\
=& o_p(1)
\end{aligned}
$$

■

Proof of Theorem 2. Since the asymptotic result is established, we only need to prove the consistency of $\widehat{\sigma}^2_{U^{nested}}$. As

$$
\widehat{\sigma}^2_{U^{nested}} - \sigma^2_{U^{nested}} = \frac{J-1}{2nJ}\sum_i\left[\frac{\left(\frac{1}{J}\sum_{t=1}^{t=J}\widehat{v}^2_{1gi,t}\right)^2}{\left(\frac{1}{J}\sum_{t=1}^{t=J}\widehat{v}^{\alpha}_{1gi,t}\right)^2} - \frac{\left(E\left[v^2_{1gi,t}\right]\right)^2}{\left(E\left[v^{\alpha}_{1gi,t}\right]\right)^2}\right]
$$

, it suffices to show that

$$\frac{1}{n}\sum_i\left[\frac{\left(\frac{1}{J}\sum_{t=1}^{t=J}\widehat{v}_{1gi,t}^2\right)^2}{\left(\frac{1}{J}\sum_{t=1}^{t=J}\widehat{v}_{1gi,t}^{\alpha}\right)^2}-\frac{\left(E\left[v_{1gi,t}^2\right]\right)^2}{\left(E\left[v_{1gi,t}^{\alpha}\right]\right)^2}\right] \tag{3.128}$$

$$=\frac{1}{n}\sum_i\left\{\left[\left(\frac{1}{J}\sum_{t=1}^{t=J}\widehat{v}_{1gi,t}^2\right)^2-\left(E\left[v_{1gi,t}^2\right]\right)^2\right]\left(\frac{1}{J}\sum_{t=1}^{t=J}\widehat{v}_{1gi,t}^{\alpha}\right)^{-2}\right\} \tag{3.129}$$

$$+\frac{1}{n}\sum_i\left\{\left(E\left[v_{1gi,t}^2\right]\right)^2\left[\left(\frac{1}{J}\sum_{t=1}^{t=J}\widehat{v}_{1gi,t}^{\alpha}\right)^{-2}-\left(E\left[v_{1gi,t}^{\alpha}\right]\right)^{-2}\right]\right\} \tag{3.130}$$

$$=o_p(1) \tag{3.131}$$

Since we can prove that

$$\begin{aligned}\widehat{v}_{1gi,t}\left(\widetilde{\theta},\widehat{\alpha}_i(\widetilde{\theta})\right)-v_{1gi,t}\left(\theta_0,\alpha_i(\theta_0)\right)=&\widehat{v}_{1gi,t}\left(\widetilde{\theta},\widehat{\alpha}_i(\widetilde{\theta})\right)-v_{1gi,t}\left(\theta_0,\widehat{\alpha}_i(\theta_0)\right)\\&+v_{1gi,t}\left(\theta_0,\widehat{\alpha}_i(\theta_0)\right)-v_{1gi,t}\left(\theta_0,\alpha_i(\theta_0)\right)\\=&\underbrace{\frac{\partial v_{1gi,t}(\widetilde{\widetilde{\widetilde{\theta}}},\widehat{\alpha}_i(\widetilde{\widetilde{\widetilde{\theta}}}))}{\partial\theta}\left(\widetilde{\widetilde{\theta}}-\widehat{\theta}\right)\left(\widetilde{\theta}-\theta_0\right)}_{O_p\left(\frac{1}{nT}\right)}+\underbrace{v_{1gi,t}^{\alpha}\left[\widehat{\alpha}_i(\theta_0)-\alpha_i(\theta_0)\right]}_{O_p\left(\frac{1}{\sqrt{T}}\right)}\\&+\underbrace{\frac{1}{2}v_{1gi,t}^{\alpha\alpha}\left[\widehat{\alpha}_i(\theta_0)-\alpha_i(\theta_0)\right]^2}_{O_p\left(\frac{1}{T}\right)}+\underbrace{\frac{1}{6}\widetilde{v}_{1gi,t}^{\alpha\alpha\alpha}\left[\widehat{\alpha}_i(\theta_0)-\alpha_i(\theta_0)\right]^3}_{O_p\left(\frac{1}{T\sqrt{T}}\right)}\end{aligned}$$

where $\widetilde{v}^{\alpha\alpha\alpha}_{1gi,t} = v^{\alpha\alpha\alpha}_{1gi,t}\left(\theta_0, \widetilde{\alpha}_i(\theta_0)\right)$, $\widetilde{\alpha}_i(\theta_0)$ is between $\widehat{\alpha}_i(\theta_0)$ and $\alpha_i(\theta_0)$. Therefore it is shown that

$$\begin{aligned}
&\frac{1}{J}\sum_t \widehat{v}^2_{1gi,t}\left(\widetilde{\theta}, \widehat{\alpha}_i(\widetilde{\theta})\right) - \frac{1}{J}\sum_t v^2_{1gi,t}(\theta_0, \alpha_i(\theta_0)) \\
&= \frac{1}{J}\sum_t 2v_{1gi,t}\left(\widehat{v}_{1gi,t} - v_{1gi,t}\right) + \frac{1}{J}\sum_t 2\left(\widehat{v}_{1gi,t} - v_{1gi,t}\right)^2 \\
&= \underbrace{\left[\frac{1}{J}\sum_t 2v_{1gi,t}\frac{\partial v_{1gi,t}(\widetilde{\widetilde{\widetilde{\theta}}}, \widehat{\alpha}_i(\widetilde{\widetilde{\widetilde{\theta}}}))}{\partial\theta}\right]\left(\widetilde{\widetilde{\theta}} - \widehat{\theta}\right)\left(\widetilde{\theta} - \theta_0\right)}_{O_p\left(\frac{1}{nT}\right)} + \underbrace{[\frac{1}{J}\sum_t 2v_{1gi,t}v^{\alpha}_{1gi,t}]\left[\widehat{\alpha}_i(\theta_0) - \alpha_i(\theta_0)\right]}_{O_p\left(\frac{1}{\sqrt{T}}\right)} \\
&+ \underbrace{[\frac{1}{J}\sum_t v_{1gi,t}v^{\alpha\alpha}_{1gi,t}]\left[\widehat{\alpha}_i(\theta_0) - \alpha_i(\theta_0)\right]^2}_{O_p\left(\frac{1}{T}\right)} + \underbrace{[\frac{1}{J}\sum_t \frac{1}{3}v_{1gi,t}\widetilde{v}^{\alpha\alpha\alpha}_{1gi,t}]\left[\widehat{\alpha}_i(\theta_0) - \alpha_i(\theta_0)\right]^3}_{O_p\left(\frac{1}{T\sqrt{T}}\right)} + O_p\left(\frac{1}{T}\right) \\
&= o_p(1)
\end{aligned}$$

and

$$\left(E\left[v^{\alpha}_{1gi,t}\right]\right)^{-2}\left[\left(\frac{1}{J}\sum_{t=1}^{t=J}\widehat{v}^2_{1gi,t}\right)^2 - \left(E\left[v^2_{1gi,t}\right]\right)^2\right] = o_p(1) \qquad \text{(T.1.4)}$$

Applying similar analysis to $\frac{1}{J}\sum_t \widehat{v}^{\alpha}_{1gi,t}$, we have

$$\begin{aligned}
\frac{1}{J}\sum_t \widehat{v}^{\alpha}_{1gi,t} - \frac{1}{J}\sum_t v^{\alpha}_{1gi,t} =& \frac{1}{J}\sum_t\left[\widehat{v}^{\alpha}_{1gi,t}\left(\widetilde{\theta}, \widehat{\alpha}_i(\widetilde{\theta})\right) - v^{\alpha}_{1gi,t}\left(\theta_0, \widehat{\alpha}_i(\theta_0)\right)\right] \\
&+ \frac{1}{J}\sum_t\left[v^{\alpha}_{1gi,t}\left(\theta_0, \widehat{\alpha}_i(\theta_0)\right) - v^{\alpha}_{1gi,t}(\theta_0, \alpha_i(\theta_0))\right] \\
=& \frac{1}{J}\sum_t \frac{\partial v^{\alpha}_{1gi,t}(\widetilde{\widetilde{\widetilde{\theta}}}, \widehat{\alpha}_i(\widetilde{\widetilde{\widetilde{\theta}}}))}{\partial\theta}\left(\widetilde{\widetilde{\theta}} - \widehat{\theta}\right)\left(\widetilde{\theta} - \theta_0\right) + \left(\frac{1}{J}\sum_t v^{\alpha\alpha}_{1gi,t}\right)\left[\widehat{\alpha}_i(\theta_0) - \alpha_i(\theta_0)\right] \\
&+ \frac{1}{2}\left(\frac{1}{J}\sum_t v^{\alpha\alpha\alpha}_{1gi,t}\right)\left[\widehat{\alpha}_i(\theta_0) - \alpha_i(\theta_0)\right]^2 + \frac{1}{6}\left(\frac{1}{J}\sum_t v^{\alpha\alpha\alpha\alpha}_{1gi,t}\right)\left[\widehat{\alpha}_i(\theta_0) - \alpha_i(\theta_0)\right]^3 \\
=& O_p(1)O_p\left(\frac{1}{nT}\right) + O_p(1)O_p\left(\frac{1}{\sqrt{T}}\right) = o_p(1) \qquad \text{(T.1.5)}
\end{aligned}$$

Based on law of large numbers,

$$\frac{1}{J}\sum_{t} v_{1gi,t}^{\alpha} - E\left[v_{1gi,t}^{\alpha}\right] = o_p(1) \tag{T.1.6}$$

$$\frac{1}{J}\sum_{t} v_{1gi,t}^{2} - E\left[v_{1gi,t}^{2}\right] = o_p(1) \tag{T.1.7}$$

Combining equation (T.1.4), (T.1.5), (T.1.6) and (T.1.7), we can prove that (T.1.2) is $o_p(1)$.

■

Proof of Theorem 3. Since

$$\begin{bmatrix} \sqrt{nT}(\widetilde{\theta}-\theta_0) \\ \sqrt{nT}(\widetilde{\gamma}-\gamma_0) \end{bmatrix} \xrightarrow{d} N(0, \Sigma_{\theta\gamma})$$

$$-LM_{nT}\left(\tilde{\theta},\tilde{\gamma}\right) = -\frac{NT}{2}(\widetilde{\theta}-\theta_0)'\mathcal{I}_f(\widetilde{\theta}-\theta_0) + \frac{NT}{2}(\widetilde{\gamma}-\gamma_0)'\mathcal{I}_g(\widetilde{\gamma}-\gamma_0) + o_p(1)$$

$-2LM_{nT}\left(\tilde{\theta},\tilde{\gamma}\right)$ can be considered as a quadratic form of a vector of normally distributed random variables:

$$\begin{bmatrix} \sqrt{nT}(\widetilde{\theta}-\theta_0) \\ \sqrt{nT}(\widetilde{\gamma}-\gamma_0) \end{bmatrix}' Q_{\theta\gamma} \begin{bmatrix} \sqrt{nT}(\widetilde{\theta}-\theta_0) \\ \sqrt{nT}\ (\widetilde{\gamma}-\gamma_0) \end{bmatrix}$$

where

$$\Sigma_{\theta\gamma} = \begin{bmatrix} \Sigma_{\theta}; & Cov_{\theta\gamma} \\ Cov_{\theta\gamma}; & \Sigma_{\gamma} \end{bmatrix}; Q_{\theta\gamma} = \begin{bmatrix} -\mathcal{I}_f & 0 \\ 0 & \mathcal{I}_g \end{bmatrix}$$

According to Lemma 10, $-2LM_{nT}\left(\tilde{\theta},\tilde{\gamma}\right)$ is asymptotically distributed as a central chi-square if and only if

$$\Sigma_{\theta\gamma}Q_{\theta\gamma}\Sigma_{\theta\gamma}Q_{\theta\gamma}\Sigma_{\theta\gamma} = \Sigma_{\theta\gamma}Q_{\theta\gamma}\Sigma_{\theta\gamma} \tag{3.132}$$

If the information matrix identity holds:

$$\mathcal{I}_f = E\left[U_{fit}\left(\theta_0,\alpha_0\right)U'_{fit}\left(\theta_0,\alpha_0\right)\right]$$

$$\mathcal{I}_g = E\left[U_{git}\left(\theta_0,\alpha_0\right)U'_{git}\left(\theta_0,\alpha_0\right)\right]$$

then equation (3.132) implies that

$$\mathcal{I}_f - \left(E\left[U_{fit}\left(\theta_0, \alpha_0\right) U'_{git}\left(\theta_0, \alpha_0\right)\right]\right) \mathcal{I}_g^{-1} \left(E\left[U_{git}\left(\theta_0, \alpha_0\right) U'_{fit}\left(\theta_0, \alpha_0\right)\right]\right) = 0 \tag{3.133}$$

which holds under the null hypothesis, it is also known that the degree of freedom is:

$$tr\left(Q_{\theta\gamma}\Sigma_{\theta\gamma}\right) = \dim(\gamma) - \dim(\theta)$$

Lee and Phillips (2015) consider the following profile likelihood information criterion:

$$\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\log f\left(z_{i,t};\tilde{\theta},\widehat{\alpha}_i(\widetilde{\theta})\right) - \frac{1}{nT}\sum_{i=1}^{n}\widehat{R}^*_{fi}(\widetilde{\theta}) + \frac{1}{nT}\operatorname{tr}\left\{J_f(\widehat{G})^{-1}I_f(\widehat{G})\right\}$$

where

$$J_f(\widehat{G}) = -\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{\partial^2\log f\left(z_{i,t};\widetilde{\theta},\alpha_i(\widetilde{\theta})\right)}{\partial\theta\partial\theta'}$$

$$I_f(\widehat{G}) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{\partial\log f\left(z_{i,t};\widetilde{\theta},\alpha_i(\widetilde{\theta})\right)}{\partial\theta}\frac{\partial\log f\left(z_{i,t};\widetilde{\theta},\widehat{\alpha}_i(\widetilde{\theta})\right)}{\partial\theta'}$$

Under H_0, the information matrix identity holds and gives

$$\operatorname{tr}\left\{J_f(\widehat{G})^{-1}I_f(\widehat{G})\right\} = \frac{\dim(\theta)}{nT}$$

it means that chi-square distribution is also achieved, which is equivalent to our results. ∎

Proof of Theorem 4. Since under H_0, the estimator of variance of $LR_{nT}(\theta_0, \gamma_0)$ is:

$$
\begin{aligned}
&\frac{1}{nT}\sum_i\sum_t[\psi_f\left(\widetilde{\theta}, \widehat{\alpha}_i(\widetilde{\theta})\right) - \psi_g\left(\tilde{\gamma}, \widehat{\lambda}_i(\tilde{\gamma})\right)]^2 \\
=&\frac{1}{nT}\sum_i\sum_t[\psi_f(\theta_0, \alpha_i(\theta_0)) - \psi_g(\gamma_0, \lambda_i(\gamma_0)) + \psi_f\left(\widetilde{\theta}, \widehat{\alpha}_i(\widetilde{\theta})\right) - \psi_f(\theta_0, \widehat{\alpha}_i(\theta_0)) + \psi_f(\theta_0, \widehat{\alpha}_i(\theta_0)) \\
&- \psi_f(\theta_0, \alpha_i(\theta_0)) - \psi_g\left(\widetilde{\gamma}, \widehat{\lambda}_i(\widetilde{\gamma})\right) + \psi_g\left(\gamma_0, \widehat{\lambda}_i(\gamma_0)\right) - \psi_g\left(\gamma_0, \widehat{\lambda}_i(\gamma_0)\right) + \psi_g(\gamma_0, \lambda_i(\gamma_0))]^2 \\
=&\frac{1}{nT}\sum_i\sum_t[\psi_f(\theta_0, \alpha_i(\theta_0)) - \psi_g(\gamma_0, \lambda_i(\gamma_0)) + \psi_f(\theta_0, \widehat{\alpha}_i(\theta_0)) - \psi_f(\theta_0, \alpha_i(\theta_0)) \\
&+ \psi_g\left(\gamma_0, \widehat{\lambda}_i(\gamma_0)\right) - \psi_g(\gamma_0, \lambda_i(\gamma_0))]^2 + o_p(1) \\
=&\frac{1}{nT}\sum_i\sum_t\left[\psi_f(\theta_0, \alpha_i(\theta_0)) - \psi_g(\gamma_0, \lambda_i(\gamma_0)) + v_{f,it}\left\{\widehat{\alpha}_i(\theta_0) - \alpha_i(\theta_0)\right\} - v_{g,it}\left\{\widehat{\lambda}_i(\gamma_0) - \lambda_i(\gamma_0)\right\}\right]^2 + o_p(1) \\
=&\frac{1}{nT}\sum_i\sum_t\left[\psi_f(\theta_0, \alpha_i(\theta_0)) - \psi_g(\gamma_0, \lambda_i(\gamma_0)) + v_{f,it}\left\{\widehat{\alpha}_i(\theta_0) - \alpha_i(\theta_0)\right\} - v_{g,it}\left\{\widehat{\lambda}_i(\gamma_0) - \lambda_i(\gamma_0)\right\}\right]^2 + o_p(1) \\
=&\frac{1}{nT}\sum_i\sum_t\left[\psi_f(\theta_0, \alpha_i(\theta_0)) - \psi_g(\gamma_0, \lambda_i(\gamma_0)) - v_{f,it}\frac{\frac{1}{T}\sum_t v_{fi,t}}{E\left[v^{\alpha}_{fi,t}\right]} + v_{g,it}\frac{\frac{1}{T}\sum_t v_{gi,t}}{E\left[v^{\alpha}_{gi,t}\right]}\right]^2 + o_p(1) \\
=&\frac{1}{nT}\sum_i\sum_t\left[\psi_f(\theta_0, \alpha_i(\theta_0)) - \psi_g(\gamma_0, \lambda_i(\gamma_0))\right]^2 + \frac{1}{nT}\sum_i\sum_t\left\{v_{f,it}\frac{\frac{1}{T}\sum_t v_{fi,t}}{E\left[v^{\alpha}_{fi,t}\right]} - v_{g,it}\frac{\frac{1}{T}\sum_t v_{gi,t}}{E\left[v^{\alpha}_{gi,t}\right]}\right\}^2 \\
&- \frac{2}{nT}\sum_i\sum_t\left[\psi_f(\theta_0, \alpha_i(\theta_0)) - \psi_g(\gamma_0, \lambda_i(\gamma_0))\right]\left\{v_{f,it}\frac{\frac{1}{T}\sum_t v_{fi,t}}{E\left[v^{\alpha}_{fi,t}\right]} - v_{g,it}\frac{\frac{1}{T}\sum_t v_{gi,t}}{E\left[v^{\alpha}_{gi,t}\right]}\right\} + o_p(1) \\
=&\omega^2 + \frac{1}{nT}\sum_i\frac{\left(E\left[v^2_{fi,t}\right]\right)^2}{\left(E\left[v^{\alpha}_{fi,t}\right]\right)^2} + \frac{1}{nT}\sum_i\frac{\left(E\left[v^2_{gi,t}\right]\right)^2}{\left(E\left[v^{\alpha}_{gi,t}\right]\right)^2} - 2\frac{1}{nT}\sum_i\frac{\left(E\left[v^2_{fi,t}v^2_{gi,t}\right]\right)}{E\left[v^{\alpha}_{fi,t}\right]E\left[v^{\alpha}_{gi,t}\right]} + o_p(1) \\
=&\frac{1}{T}\sigma^2_W - \frac{1}{T}\sigma^2_U + \frac{2}{T}\sigma^2_U + o_p(1) = \frac{1}{T}\sigma^2_W + \frac{1}{T}\sigma^2_U + o_p(1)
\end{aligned}
$$

Thus we have the estimator of variance of $\frac{1}{\sqrt{n}}LR_{nT}(\theta_0, \gamma_0)$ is:

$$\frac{1}{n}\sum_i\sum_t[\psi_f\left(\widetilde{\theta}, \widehat{\alpha}_i(\widetilde{\theta})\right) - \psi_g\left(\tilde{\gamma}, \widehat{\lambda}_i(\tilde{\gamma})\right)]^2 = \sigma^2_W + \sigma^2_U + o_p(1)$$

It means that $\frac{1}{n}\sum_i\sum_t\left[\psi_f\left(\widetilde{\theta}, \widehat{\alpha}_i(\widetilde{\theta})\right) - \psi_g\left(\tilde{\gamma}, \widehat{\lambda}_i(\tilde{\gamma})\right)\right]^2$ is a biased estimator of σ^2_W, we propose the bias-corrected variance term as follows:

$$\widehat{\sigma}^2_W = \frac{1}{n}\sum_i\sum_t[\psi_f\left(\widetilde{\theta}, \widehat{\alpha}_i(\widetilde{\theta})\right) - \psi_g\left(\tilde{\gamma}, \widehat{\lambda}_i(\tilde{\gamma})\right)]^2 - \widehat{\sigma}^2_U$$

■

Bibliography

Arellano, M. (2003). Discrete choices with panel data. *Investigaciones Económicas*, *27*(3), 423–458.

Arellano, M., & Hahn, J. (2007). Understanding bias in nonlinear panel models: Some recent developments. *Econometric Society Monographs*, *43*, 381.

Arellano, M., & Hahn, J. (2016). A likelihood-based approximate solution to the incidental parameter problem in dynamic nonlinear models with multiple effects. *Global Economic Review*, *45*(3), 251–274.

Atkinson, A. C. (1970). A method for discriminating between models. *Journal of the Royal Statistical Society: Series B (Methodological)*, *32*(3), 323–345.

Bester, C. A., & Hansen, C. (2009). A penalty function approach to bias reduction in nonlinear panel models with fixed effects. *Journal of Business & Economic Statistics*, *27*(2), 131–148.

Bontemps, C., Florens, J.-P., & Richard, J.-F. (2008). Parametric and non-parametric encompassing procedures. *Oxford Bulletin of Economics and Statistics*, *70*, 751–780.

Carro, J. M. (2007). Estimating dynamic panel data discrete choice models with fixed effects. *Journal of Econometrics*, *140*(2), 503–528.

Chen, Q., & Giles, D. E. (2012). Finite-sample properties of the maximum likelihood estimator for the binary logit model with random covariates. *Statistical Papers*, *53*(2), 409–426.

Cox, D. R. Tests of separate families of hypotheses. In: In *Proceedings of the fourth berkeley symposium on mathematical statistics and probability. 1.* (0). 1961, 96.

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

Fernández-Val, I. (2005). Estimation of structural parameters and marginal effects in binary choice panel data models with fixed effects. *Available at SSRN 869085.*

Gourieroux, C., & Monfort, A. (1995). Testing, encompassing, and simulating dynamic econometric models. *Econometric Theory*, 195–228.

Hahn, J., & Kuersteiner, G. (2011). Bias reduction for dynamic nonlinear panel models with fixed effects. *Econometric Theory*, *27*(6), 1152–1191.

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

Hajivassiliou, V. A., & Ruud, P. A. (1994). Classical estimation methods for ldv models using simulation. *Handbook of econometrics*, *4*, 2383–2441.

Hsu, Y.-C., & Shi, X. (2017). Model-selection tests for conditional moment restriction models. *Econometrics Journal*.

King, G., & Zeng, L. (2001). Logistic regression in rare events data. *Political analysis*, *9*(2), 137–163.

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

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

Lee, Y., & Phillips, P. C. (2015). Model selection in the presence of incidental parameters. *Journal of Econometrics*, *188*(2), 474–489.

Lerman, S., & Manski, C. (1981). On the use of simulated frequencies to approximate choice probabilities. *Structural analysis of discrete data with econometric applications*, *10*, 305–319.

Liao, Z., & Shi, X. (2020). A nondegenerate vuong test and post selection confidence intervals for semi/nonparametric models. *Quantitative Economics*, *11*(3), 983–1017.

Liu, T., & Lee, L.-F. (2019). A likelihood ratio test for spatial model selection. *Journal of Econometrics*, *213*(2), 434–458.

MacKinnon, J. G., Nielsen, M. Ø., & Webb, M. (2020). *Testing for the appropriate level of clustering in linear regression models* (tech. rep.). Queen's Economics Department Working Paper.

McCullagh, P., & Nelder, J. A. (2019). *Generalized linear models*. Routledge.

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica: Journal of the Econometric Society*, 995–1026.

Mizon, G. E., & Richard, J.-F. (1986). The encompassing principle and its application to testing non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, 657–678.

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

Pace, L., & Salvan, A. (2006). Adjustments of the profile likelihood from a new perspective. *Journal of Statistical Planning and Inference*, *136*(10), 3554–3564.

Pakes, A., & Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica: Journal of the Econometric Society*, 1027–1057.

Pesaran, M. H. (1974). On the general problem of model selection. *The Review of Economic Studies*, *41*(2), 153–171.

Pesaran, M. H., & Deaton, A. S. (1978). Testing non-nested nonlinear regression models. *Econometrica: Journal of the Econometric Society*, 677–694.

Ramalho, J. J., & Smith, R. J. (2002). Generalized empirical likelihood non-nested tests. *Journal of Econometrics*, *107*(1-2), 99–125.

Rao, C. R., & Mitra, S. K. (1971). Further contributions to the theory of generalized inverse of matrices and its applications. *Sankhyā: The Indian Journal of Statistics, Series A*, 289–300.

Rilstone, P. (1996). On the finite sample effects of nonlinear reparameterizations. *Econometric Reviews*, *15*(1), 33–50.

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

Shi, X. (2015). A nondegenerate vuong test. *Quantitative Economics*, *6*(1), 85–121.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, 307–333.

Wang, H. (2020). Logistic regression for massive data with rare events. *International Conference on Machine Learning*, 9829–9836.

Woutersen, T. (2002). *Robustness against incidental parameters* (tech. rep.). Research Report.