THREE ESSAYS ON UNBALANCED PANEL DATA MODELS

By

Do Won Kwak

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Economics

2011

Abstract

THREE ESSAYS ON UNBALANCED PANEL DATA MODELS

By

Do Won Kwak

This dissertation consists of three chapters on panel data models. The first chapter examines unconditional logit and conditional logit, and pooled correlated random effect (CRE) logit approaches in binary panel data models with highly dependent data. Simulation results show that, first, the conditional logit method is not robust to violation of the conditional independence (CI) assumption. The magnitude of bias for the model coefficient is greater for the data with smaller time dimensions and higher serial correlations. Second, we find no significant finite sample biases for average partial effects (APEs) and associated rejection frequency of CRE logit in the presence of high serial correlation under correct specification of unobserved heterogeneity. Finally, we quantify two sources of bias into the part due to unobserved heterogeneity and the part due to serial correlation by using unconditional logit and conditional logit methods. Finite sample biases by both sources are important in binary panel models with highly persistent outcome and correlated individual fixed effects. As an empirical example, we apply conditional logit and pooled CRE logit to the Survey of Income and Program Participation (SIPP) data where welfare participation exhibits high persistence. The results imply that it is important to account for both state dependence and unobserved heterogeneity simultaneously.

The second chapter introduces two formal tests of the missing completely at random (MCAR) assumption in unbalanced panel data. In a Monte Carlo (MC) simulation, we provide the evidence of the substantial powers of tests. Inverse probability weighted (IPW) estimator and multiple imputation-based (MI) estimator under the missing at random (MAR) assumption are studied as methods for missing data when the MCAR assumption is violated. We suggest combining MI and IPW methods such that the MI method is applied to non-monotone missing data and the IPW method is applied to monotone missing data sequentially. Proposed tests and the MI-IPW method for missing data are applied to estimate the effect of class size reduction (CSR) on student scores for grades in K-3 using Project STAR. The result of empirical application shows the violation of the MCAR assumption and the MI-IPW estimates for the effects of CSR on student achievement scores in grades K-3 are about 4.5 to 6 percent while unweighted estimates are about 6 to 7.5 percent.

In the last chapter, we extend the robust inference method with heterogenous variance and autocorrelation in Hansen (2007) to unbalanced panel data with large time dimension. With the homogeneity assumption and random attrition, we derive the robust inference result using weighted least squares (WLS) for unbalanced panel data. We show that, our inference method with WLS in unbalanced panel data provides conservative t-test results in Bakirov and Szekely (2006) in the presence of attrition or heteroskedastic variance. We study the size and the power of a proposed inference method in unbalanced panel data using the Monte Carlo simulation. A MC simulation reveals that a proposed WLS method reduces the size distortion and improves power over the methods of Ibragimov and Muller (2009) and Hansen (2007).

ACKNOWLEDGMENTS

I would never been able to finish my dissertation without guidance, suggestions and support from several people. First, my deepest appreciation goes to my advisor, Jeffrey M. Wooldridge. He was readily available at every stage of the dissertation process and provided me his perspective and helpful comments all the time. I could never have reached the depths of this dissertation without his mentorship and insightful advice. Besides my advisor, I would like to sincerely thank the faculty members in Michigan State University, Todd Elder, Peter Schmidt, Timothy Vogelsang, and Tapabrata Maiti. Whenever I needed advice about my research, they helped make this dissertation a better work with their insights and suggestions. I also have been very fortunate to have conducted some part of my dissertation research at StataCorp as an intern. Simulations in my dissertation would not have been developed as it is now without computing and data expertise I had learnt at StataCorp. I would like to extend a special thanks to David Drukker and Rafal Raciborski.

I would like to thanks to my fellow graduate students. Jeff Brown, Cheol-Keun Cho, Dooyeon Cho, Junghwan Hyun, Myoung-Jin Keay, Jongduk Kim, Jinyoung Lee, Cuicui Lu, Suhyeon Nam, Iraj Rahmani, Seunghwa Rho, Shengwu Shang, Wei-Siang Wang, and Yali Wang had numerous discussions with me and I appreciate for their thoughtful comments and enjoyable conversation.

I would like to express my gratitude to my parents and younger brother for their unwavering love and support. Last but not least, I would like to thank my wife who was always there cheering me up and stood by me through the good and bad times.

Contents

List of Tables

List of Figures

Chapter 1

The robustness of conditional logit estimator in a binary panel data model

1.1 Introduction

Standard maximum likelihood estimators are usually inconsistent in binary panel data models with individual fixed effects when individual fixed effects are unspecified with a large number in the cross-section (N) and a fixed number of time (T). The problem arises since there are so many nuisance parameters (individual fixed effects) to estimate while the time dimension is fixed.[1] Fixed-T identification difficulty for micro panel can be overcome by using

[1] This situation is called incidental parameters problem. The references on the incidental parameters problem include Neyman and Scott [1948], Lancaster [2000] and Arellano and Hahn [2007]. Individual fixed effects is also called as unobserved heterogeneity hereafter.

the conditional logit approach if the logistic distributional assumption is true.[2] Chamberlain [2010] showed that conditional logit delivers $\sqrt{n}$ consistent estimation of the parameter of interest if individual-and-period specific errors are independent over time and logistic, and covariates are unbounded. However, these conditions for $\sqrt{n}$ consistency of conditional logit can be very restrictive in micro panel data applications. In particular, the independence assumption is restrictive in the applications with highly persistent response variables such as married women's labor force participation (Chay and Hyslop [2001]), married women's welfare participation (Chay and Hyslop [2001]), the incidence of external debt crises in developing countries (Hajivassiliou and McFadden [1998]) and habitual smoking behavior (Collado and Browning [2007]). Moreover, as fixed-T consistency results for model coefficients can not be extended to APEs, it becomes a serious limitation in the applications where researchers are interested in the effect of a policy on outcome. As we do not make any distributional assumption on heterogeneity (i.e. individual fixed effects), it is difficult to figure out what to plug in for unobserved heterogeneity when calculating the APEs. However, the conditional logit approach is useful for the estimations of log odd ratio and relative effects between covariates.

In this paper, we study the robustness of fixed-T consistency results for the conditional logit estimator to the violation of conditional independence assumption using a Monte Carlo (MC) experiment. Specifically, we introduce data generating processes (DGPs) which are

[2] These types of conditional methods are not available in general non-linear models and can not be extended to APEs estimation since they do not make any distributional assumption on unobserved heterogeneity. For general non-linear models, a recent literature focuses rather on approximately unbiased estimator than on the estimators with no bias at all. Arellano and Hahn [2007], Hahn and Newey [2004], Arellano [2003], Fernandez-Val [2009] and others provide bias correction methods for non-linear fixed effects panel data model. An extensive review is provided in Arellano and Hahn [2007].

designed to satisfy ideal conditions for the conditional logit estimator except the conditional independence (CI) assumption to focus on the effects by the violation of CI assumption. For simplicity, the violation of CI in DGPs is induced by correlated logistic errors of AR(1) using a copula. We introduce a copula in simulation since it allows the separation of the dependence structure over time from the marginal distribution for each time period. Thus, by specifying marginal distribution of outcome for each time period and using a copula to specify the dependence structure of marginal distributions of outcomes over time, we specify joint non-normal (logistic) distribution for outcomes over time. In a MC experiment, using DGPs with correlated logistic errors, we examine the performance of conditional logit, CRE logit with pooled MLE, and unconditional logit methods.[3] Finally, we quantify finite sample bias due to unobserved heterogeneity and due to true state dependence using unconditional logit and conditional logit estimators. Simulation with highly persistent binary panel data is important for the conditional logit estimator as well as for approximately unbiased estimators with bias correction since these methods require the conditional independence assumption but the effect of violation has not been studied much.

When interpreting the simulations results, we focus on finite sample bias for coefficients β and APEs in (1.1) while in DGPs, the focus is on the generation of correlated errors $u_{it}{}_{t=1}^{T}$ which are drawn from T-multivariate logistic distribution with serial correlation of AR(1).

[3] Correlated random effect logit approach is based on parametric specification of the distribution for unobserved heterogeneity as in Chamberlain [1980]. In the simulation and application, we use parametric specification for unobserved heterogeneity as follows. $D(c_i|\mathbf{x}_i) \sim \text{logistic}(\psi + \mathbf{x}_i\xi, \sigma_a^2)$. Unconditional logit approach is based on logit model with MLE where it estimate unobserved heterogeneity(c_i) as parameters. For the properties of unconditional logit see Arellano and Hahn [2007] and Andersen [1973]. For the case of T=2, Abrevaya [1997] shows the bias of unconditional MLE is 100 %.

$$y_{it} = 1[\mathbf{x}_{it}\beta + c_i + u_{it} \geq 0],\ i = 1, 2, ..., n,\ \text{and}\ t = 1, 2, ..., T \tag{1.1}$$

where n is large, T is fixed and $T \geq 2$, $\mathbf{x}_{it}$ is continuous or binary, c_i is correlated unobserved heterogeneity (with $\mathbf{x}_{it}$) and u_{it} is idiosyncratic error.

In simulation, the conditional logit estimator for β in (1.1) shows finite sample bias when the CI assumption is violated by serial correlation among errors. Finite sample bias is greater for the high ρ(serial correlation coefficient) and small T (the size of time dimension). The size distortion of inference (the bias of rejection frequency) with conditional logit is very severe for the data with $\rho \geq .4$ and it increases with n (the size of the cross-section). CRE logit with pooled MLE for APEs and the associated rejection frequency shows no finite sample bias regardless of the magnitude of ρ. Unconditional logit exhibits the bias due to both (inconsistent estimation of) unobserved heterogeneity and serial correlation. For unconditional logit estimator β with DGPs of $\rho \leq .4$, more bias is due to unobserved heterogeneity than due to serial correlation. For the unconditional logit estimator for APEs, more bias is due to unobserved heterogeneity than to serial correlation for DGPs with $\rho \leq .6$.

We analyze married women's welfare participation using the conditional logit model with MLE and CRE logit model with pooled MLE for β and APEs with the data of 8 waves for 1,934 married women, taken from Chay and Hyslop [2001]. Married women's welfare participation shows substantial serial persistence. This persistence could be due either to certain individual characteristics or due to true state dependence.[4] The estimation results

[4] In our example of welfare participation, true state dependence implies "narcotic" effect in Chay and Hyslop [2001]. Once a person starts to depend on welfare, participation itself change the person's mental attitude and behavior so that leads the person to rely more on welfare.

and a comparison of the estimates imply that much of the serial persistence of welfare participation is explained by unobserved heterogeneity (54%) but true state dependence (46%) is also an important source for serial persistence. Therefore, the empirical example shows that proper control of true state dependence as well as unobserved heterogeneity is very important in the model specifications of married women's welfare participation.

The rest of the chapter is organized as follows. In section 2, we explicitly explain the model and ideal assumptions for the conditional logit method and section 3 describes the procedure of DGPs using copula to satisfy ideal conditions for the conditional logit model except CI. Section 4 shows finite sample simulation results for conditional logit, CRE logit, and unconditional logit methods. Section 5 provides an empirical application pertaining to married women's welfare participation and section 6 contains a brief summary of the results and the conclusion.

1.2 Model

The conditional logit estimator allows to estimate β in (1.1) without making any assumption on unobserved heterogeneity since the logit transformation allows to separate the estimation of β from unobserved heterogeneity. (Andersen [1973], Chamberlain [1984] and Wooldridge [2010]) However, this comes with imposing CI assumption, which is not necessary for other competing estimators such as CRE logit with pooled MLE. In this section, we introduce conditional logit model and conditions for its ideal performance.

1.2.1 Ideal conditions for conditional logit estimator

Assumption 1 Our baseline model has latent variable representation.

$$y_{it} = 1[\mathbf{x}_{it}\beta + c_i + u_{it} \geq 0]$$

$$y_{it}^* = \mathbf{x}_{it}\beta + c_i + u_{it}$$

where $i = 1, 2, ...n$, $t = 1, 2, ..., T$

$$c_i \sim \text{ logistic } (\psi + \mathbf{x}_i\xi, \frac{\pi^2}{6})$$

$\mathbf{x}_{it}$ is continuous and $u_{it} \sim$ logistic$(0, \frac{\pi^2}{6})$

Assumption 2 Conditional strict exogeneity (SE)

$$E(y_{it}|\mathbf{x}_i, c_i) = E(y_{it}|\mathbf{x}_{it}, c_i) \text{ for each } t$$

$$P(y_{it} = 1|\mathbf{x}_i, c_i) = P(y_{it} = 1|\mathbf{x}_{it}, c_i) \text{ for each } t$$

Assumption 3 (Functional form marginal distribution) Conditional mean of binary response y_{it} is specified as

$$E(y_{it}|\mathbf{x}_{it}, c_i) = P(y_{it} = 1|\mathbf{x}_{it}, c_i) = \Lambda(\mathbf{x}_{it}\beta + c_i) \text{ for each } t = 1, 2, 3, ...T$$

$$\text{where } \Lambda(\mathbf{x}_{it}\beta + c_i) = \frac{\exp(\mathbf{x}_{it}\beta + c_i)}{1 + \exp(\mathbf{x}_{it}\beta + c_i)}$$

(i.e.) conditional marginal distribution of each y_{it} is logistic distribution.

Assumption 4 Conditional Independence (CI)

$$D(y_{i1,}y_{i2}, ..., y_{iT}|\mathbf{x}_i, c_i) = \prod_{t=1}^{T} D(y_{it}|\mathbf{x}_{it}, c_i)$$

Assumption 5 Other regularity conditions for the consistency of β such as convex support of β, and continuity and differentiability of objective function. We also assume random sampling over cross-section.

Theorem We assume, under assumptions 1-5, conditional logit estimator for β is a consistent estimator.

Proof See appendix.

The proof in appendix shows that the transformation using the sum of response variables as sufficient statistics eliminates unobserved heterogeneity. Obtained $\widehat{\beta}_{c-logit}$ estimator is a consistent and $\sqrt{n}$-asymptotically normal. Chamberlain [2010] also shows that the only binary panel model that provides fixed-T $\sqrt{n}$ consistent estimators for β is logit model.

Remark 1 *Proof explicitly shows that CI assumption is required to apply conditional MLE for β. CI can be easily violated when response variables show state dependence.*

1.3 DGP for correlated logistic errors

This section introduces the details of data generating processes(DGPs) in simulation for studying the robustness of conditional logit to the violation of CI assumption. The key is to

generate variables that satisfy ideal conditions for conditional logit except CI.

1.3.1 Correlated errors from multivariate logistic distribution

We introduce dependence among errors using a copula and logistic distribution. In general, specifying variables from multivariate logistic distribution can be very complicated as we have to define joint distribution. However, using copula, we can draw variables from multivariate logistic distribution by only specifying marginal distribution for each variable and copula which specify intended dependence among marginal distribution of each variable separately.

Copula

Copula is a function which links univariate marginal CDFs to their multivariate CDF with intended dependence among marginal distributions. Copula is useful for our purpose since any T-dimensional joint distribution function can be decomposed into its T marginal distributions and a copula function which completely describes the dependence structure among marginal distributions. For instance, in DGP, we use a copula to generate errors from a multivariate logistic distribution for which each marginal error has standard logistic distribution with intended dependence among marginal distribution of errors. We specify only conditional marginal distribution of z_{it} for each time and use a Gaussian copula to specify dependence among CDFs of $z_{i1}, z_{i2}, ..., z_{iT}$. We choose a Gaussian copula since most statistical packages allow users to randomly draw a set of variables from joint normal distribution with user-specified pair-wise covariance. Therefore, we specify the marginal distribution for each z_{it} and Gaussian copula for dependence among CDFs of $z_{i1}, z_{i2}, ..., z_{iT}$ as in (1.2).

$$\Phi(z_1, z_2, ..., z_T) = C(\Phi_1(z_1), ..., \Phi_t(z_t), ..., \Phi_T(z_T); \rho) \tag{1.2}$$

where $\Phi(\cdot)$ is joint normal, $\Phi_t(\cdot)$ is univariate normal of marginal distribution of each t, and $C(\cdot, \rho)$ is copula function where ρ governs dependence among marginal distributions.

Then, we transform normal marginal to logistic marginal at each period as in (1.3) and the dependence of multivariate normal by a Gaussian copula is inherited to the dependence among logistic marginal CDFs since copula is invariant to increasing transformation. Thus, after transforming of marginal distribution, we obtain (1.4).

$$\Phi_t(z_t) \underbrace{\rightarrow}_{\text{transformation of marginal CDF}} F_t(u_t) \tag{1.3}$$

$$F(u_1, u_2, ..., u_T) = C(F_1(u_1), ..., F_t(u_t), ..., F_T(u_T); \rho) \tag{1.4}$$

where $F(\cdot)$ is joint logistic distribution and $F_t(\cdot)$ is univariate logistic marginal distribution. Joint logistic density is derived using multivariate gaussian copula in the appendix.

Because of this invariance property of copula, dependence structure on $\mathbf{z}_{it}{}_{t=1}^{T}$ is maintained in $\mathbf{u}_{i}{}_{t=1}^{T}$ after increasing transformation. Generated $\mathbf{u}_{i}{}_{t=1}^{T}$ follows joint logistic distribution with correlation parameters ρ as in (1.2). Joint logistic density of $f(u_1, u_2, ..., u_T)$ and its derivation is provided in appendix.

Extension: Generate correlated variables from non-normal distribution Dependent variables generation using Gaussian copula can be extended to other non-normal multivariate distribution. For the method to work, the only requirement is to obtain the intended

marginal distribution from increasing transformation of the normal distribution. For instance, in our DGPs, each normal marginal CDF is transformed to logistic marginal CDF by increasing transformation and the Gaussian copula specifies the dependence structure among the marginal distributions. Dependence structure is maintained after increasing transforming by invariance property of copula.

Generation of a random vector from T-joint logistic distribution with dependence structure of AR(1) correlation

Consider a random vector $\mathbf{z} = (z_1, z_2, ..., z_T)'$ which is from T-variate multivariate normal distribution with AR(1) covariance matrix $Cov(\mathbf{z})$ of (1.5) and $z_t \sim N(0,1)\ \forall$ t. [5]

$$Cov(\mathbf{z}) = \begin{pmatrix} 1 & \rho & \cdots & \cdots & \rho^{T-1} \\ \rho & 1 & \cdots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \cdots & \rho^2 \\ \rho^{T-2} & \rho^{T-3} & \cdots & 1 & \rho \\ \rho^{T-1} & \rho^{T-2} & \cdots & \cdots & 1 \end{pmatrix} \tag{1.5}$$

Using gaussian copula in (1.6), we can express T-variate joint normal distribution as follows. Let Φ be a joint normal CDF for $\mathbf{z} = (z_1, z_2, ..., z_T)$ with marginal CDF Φ_t for each z_t, then, by Sklar's theorem in Sklar [1959], there exists unique copula C that connect each marginal CDF $\Phi_1(z_1), \Phi_2(z_2), ..., \Phi_T(z_T) \in \mathbb{R}^T$ with dependence structure of (1.5)

[5] We choose pairwise correlation of in (1.5) for simplicity. We have corr(Z)=cov(Z) since z_t is standard normal $\forall t$. The main advantage of using copula is the flexibility of model marginal distribution and dependence among marginal distribution separately. Moreover, the specification of dependence relationship among marginal CDFs is not limited to 2nd moment or linear relationship but is allowed to nonlinear and tail dependence. However, in the simulation, our focus is serial correlation among errors so that we use copula with only linear relationship specification for simplicity.

such that equation (1.6) is satisfied.

$$\Phi(z_1, z_2, ..., z_T) = C(\Phi_1(z_1), ..., \Phi_T(z_T); Cov(\mathbf{z})) \tag{1.6}$$

where Φ_t is normal CDF $\forall t = 1, ..., T$.

The transformation from a normal marginal distribution to a logistic marginal distribution is performed by $u_t = F_t^{-1}(\Phi(z_t)) = ln(\frac{\Phi(z_t)}{1-\Phi(z_t)})$ where $F_t(\cdot)$ is logistic CDF. We can rewrite (1.6) as the equation (1.7) for a copula C and $\mathbf{u} = (u_1, u_2, ..., u_T) \in \mathbb{R}^T$, such that

$$F(u_1, u_2, ..., u_T) = C(F_t(u_1), ..., F_t(u_T); \rho), \text{ where } F \text{ is joint logistic CDF.} \tag{1.7}$$

Because of the invariance dependence property of copula for increasing transformation of marginal distribution, $\{u_{it}\}, t = 1, 2, ..., T$, the dependent structure of $\mathbf{z}$ in (1.6) is maintained for $\mathbf{u}$ in (1.7) with the same C. [6] Lee [1979] also showed similar relationship between normal and logistic distribution for bivariate case.

1.4 Monte Carlo experiment

We generate data based on two model specifications. First one is where the model has only one covariate with unbounded support and individual fixed effects are correlated with the covariate. Second model has two covariates of unbounded continuous covariate and binary covariate and has individual fixed effects which are correlated with the binary covariate. In both models, we generated dependent errors from joint logistic distribution using Gaussian

[6] Linear dependence, (for instance, pairwise correlation), that we use here can not be maintained after transformation of marginal distribution from normal to logistic but general dependence measures such as Kendal's rank correlation are maintained under monotonic transformation such as from normal to logistic. (Schweizer and Sklar [1983] and Trivedi and Zimmer [2007])

copula. We perform simulation with binary covariate since conditional logit models with MLE requires covariates to have unbounded supports for $\sqrt{n}$ consistency result in Chamberlain [2010]. Thus, simulation with binary covariate provides information about how bounded support of covariate affects the property of conditional logit estimator.

1.4.1 The procedures of Data Generation Process: a continuous covariate

We generate data $\{y_{it}, x_{it}, c_i, u_{it}\}$ for $t = 1, 2, .., T$ and $i = 1, 2, ..., n$ based on model (1.1). We follow Heckman [1981], Green [2004] and Fernandez-Val [2009] for the design of the DGPs in this section. [7] We draw four items with 2,000 replications. The correlation across time-dimension is induced through error components **u** using a copula as illustrated in previous section 1.3.1. We generate four items as follow:

1. $x_{it} \sim iid\ N(0, 1)$. We generate an univariate covariate from identical and independent standard normal distribution.

2. Unobserved heterogeneity:

$$D(c_i|x_i) \sim \frac{x_{i1} + x_{i2} + ... + x_{ik}}{\sqrt{2k}} + \frac{a_i}{\sqrt{2}} \quad \text{where } a_i \sim \text{logistic}(0, \frac{\pi^2}{3}) \text{ and } k = [\frac{T}{2}]. \quad (1.8)$$

3. We generate each u_{it} such that it has standard logistic marginal distribution (i.e. $u_{it} \sim \log istic(0, \frac{\pi^2}{3})$), $Cov(u_{it}, u_{is}) = \rho^{|t-s|}$ for $t \neq s$, and $(u_1, u_2, ..., u_T)$ is drawn from joint logistic distribution by using the procedure in section 1.3.1.

[7] Devroye [1986] and Johnson [1987] contain extensive methods of generating multivariate random variables and vectors from non-normal distribution.

4. We generate y_{it} as follow.

$$y_{it} = 1[x_{it}\beta + c_i + u_{it} \geq 0] \text{ where } \beta = 1 \text{ and using } x_{it}, c_i, u_{it} \text{ as in} 1, 2, 3. \quad (1.9)$$

Generation of $\{x_{it}\}_{t=1}^{T}$ and c_i is straightforward and generation of $\{y_{it}\}_{t=1}^{T}$ is also straightforward once we have $\{u_{it}\}_{t=1}^{T}$. We use a copula in generating errors, $(u_{i1}, u_{i2}, ..., u_{iT})$ from T-variate logistic distribution while each u_{it} has a univariate logistic distribution and dependence structure among $(u_{i1}, u_{i2}, ..., u_{iT})$ is inherited from dependence structure of joint normal distribution for $(z_1, z_2, ..., z_T)$.

Generation of correlated logistic errors, u

Let $\mathbf{z} = (z_1, z_2, ..., z_T)'$ is variables drawn from T-variate multivariate normal distribution with dependent structure of AR(1) serial correlation. We obtain $(u_1, u_2, ..., u_T)$ by the increasing transformation of marginal distributions, $u_t = F_t^{-1}(\Phi_t(z_t))$ where F_t is logistic CDF and Φ_t is standard normal CDF.

$\mathbf{u} = (u_1, u_2, ..., u_T)$ is generated from the following algorithm:

(Step 1) Draw $z_1, z_2, ..., z_T$ from T-variate normal distribution with dependence structure of $cov(\mathbf{z})$ in (1.5).

(Step 2) Obtain each marginal CDF $e_t = \Phi_t(z_t)$ where $\Phi_t(\cdot)$ is standard normal CDF, and $\forall t = 1, ..., T$.

(Step 3) Transform z_t to u_t based on following formula. $u_t = F_t^{-1}(e_t) = F_t^{-1}(\Phi_t(z_t))$ where $F_t(\cdot) \sim logistic(0, \frac{\pi^2}{3})$, (i.e. $u_t = \ln(\frac{e_t}{1-e_t}) = \ln(\frac{\Phi_t(z_t)}{1-\Phi_t(z_t)})$)

(Step 4) Repeat steps of (1), (2), and (3) for all $i = 1, 2, ..., n$.

Table 1.1. Basic Descriptive Statistics for the Generated Errors

	u_1	u_2	u_3	u_4	u_5	u_6	u_7	u_8
$\rho = 0.1$	-.07(1.9)	.07(1.7)	.12(1.8)	.02(1.9)	-.04(1.8)	-0.05(1.9)	.02(1.9)	.08(1.9)
$\rho = 0.25$	-.03(1.8)	.03(1.8)	.08(1.9)	.03(1.9)	-.02(1.8)	-0.03(1.8)	.06(1.9)	.11(1.8)
$\rho = 0.5$	-.03(1.8)	.01(1.8)	.06(1.9)	.04(1.9)	.01(1.9)	-0.02(1.8)	.04(1.9)	.11(1.9)
$\rho = 0.75$	-.02(1.8)	-.01(1.8)	.02(1.9)	.02(1.8)	.01(1.9)	-0.01(1.9)	.03(1.9)	.07(1.9)
$\rho = 0.9$	-.02(1.8)	-.01(1.8)	.02(1.9)	.02(1.8)	.01(1.9)	-0.01(1.9)	.03(1.9)	.07(1.9)

Table 1.2. Pairwise correlation of the generated errors for $\rho = 0.1$

	u_1	u_2	u_3	u_4	u_5	u_6	u_7	u_8
u_1	1							
u_2	0.09	1						
u_3	0.06	0.13	1					
u_4	0.02	0.01	0.11	1				
u_5	0.03	0.01	-0.02	0.11	1			
u_6	0.03	0.06	0.01	0.02	0.11	1		
u_7	0.02	-0.05	0.01	0.05	0.01	0.09	1	
u_8	0.02	-0.002	0.02	0.07	0.04	0.02	0.08	1

Basic Statistics for correlated logistic errors Table 1.1 reports the MC point estimates of 2,000 replications for the mean and the standard deviation of generated errors from multivariate logistic distribution with n=100 and T=8. The estimates for standard deviations are reported in parenthesis. Table 1.2, 1.3, and 1.4 report pairwise correlations among generated errors for n=100, T=8 and ρ for 0.1, 0.5 and 0.9, respectively. Tables of basic statistics verify that pairwise correlations of $\mathbf{z}$ are maintained to the pairwise correlations among $\mathbf{u}$. Statistics for u_{it} such as mean and standard deviation is equal to mean and standard deviation of a variable from standard logistic distribution.

Table 1.3. Pairwise correlation of the generated errors for $\rho = 0.5$

	u_1	u_2	u_3	u_4	u_5	u_6	u_7	u_8
u_1	1							
u_2	0.51	1						
u_3	0.30	0.54	1					
u_4	0.14	0.28	0.52	1				
u_5	0.07	0.09	0.26	0.48	1			
u_6	-0.01	0.04	0.13	0.26	0.50	1		
u_7	0.02	0.03	0.08	0.14	0.23	0.52	1	
u_8	0.05	0.06	0.12	0.10	0.12	0.22	0.50	1

Table 1.4. Pairwise correlation of the generated errors for $\rho = 0.9$

	u_1	u_2	u_3	u_4	u_5	u_6	u_7	u_8
u_1	1							
u_2	0.90	1						
u_3	0.80	0.90	1					
u_4	0.72	0.80	0.90	1				
u_5	0.67	0.74	0.81	0.90	1			
u_6	0.59	0.66	0.73	0.82	0.91	1		
u_7	0.53	0.59	0.65	0.73	0.81	0.89	1	
u_8	0.48	0.54	0.59	0.67	0.74	0.81	0.91	1

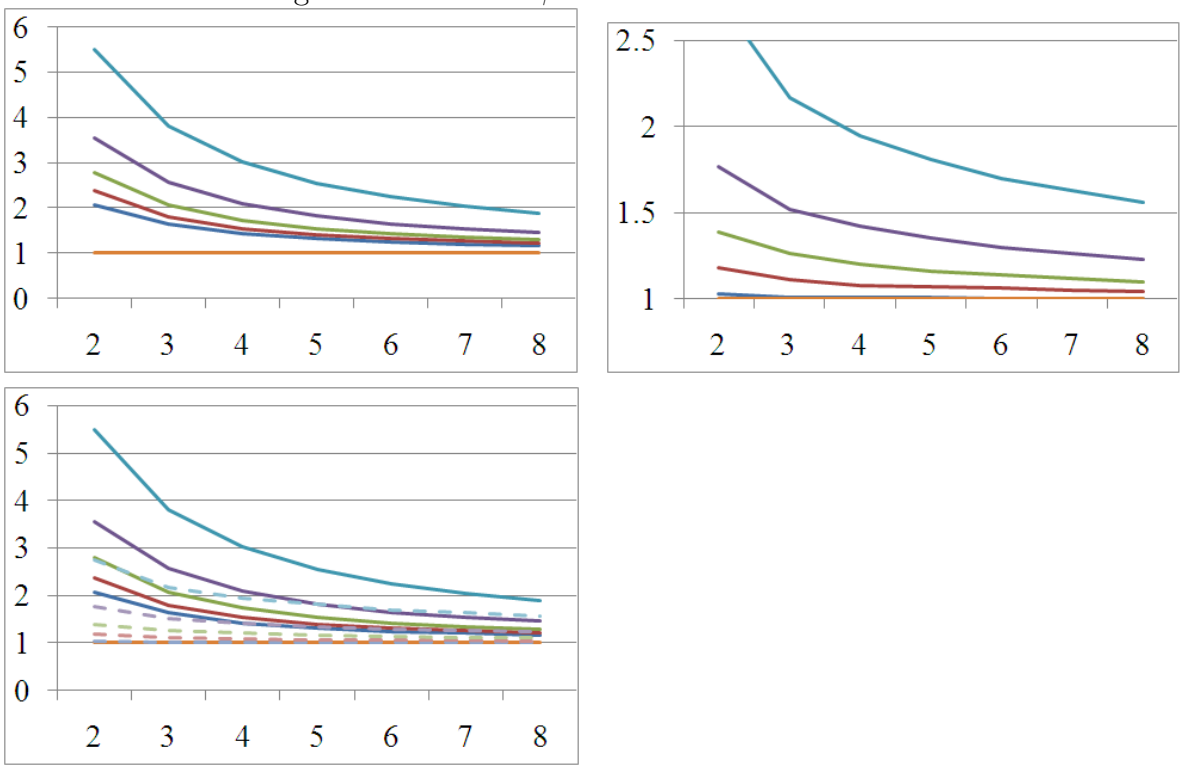
1.4.2 The bias and size distortion for coefficient: a continuous covariate

All the results presented in this section are based on 2,000 replications and the robust standard error of sandwich form that allows weakly dependent serial correlation. We vary three components T, n, and ρ in the DGPs since it is important to examine how conditional logit estimator for β behaves in finite sample as T, n, and ρ change. We are particularly interested in typical micro panel such as small T, large n with high persistence($\rho > 0.4$). Throughout the tables in this section, we report the mean of the MC point estimate for β (Mean), the standard deviation of the MC point estimate for β (SD), the mean of the MC point estimate of the robust standard error for β (SE), the rejection frequency ($1(p_{val} < 0.05)$), and 95-percent coverage rate for the rejection frequency.

Conditional logit for β

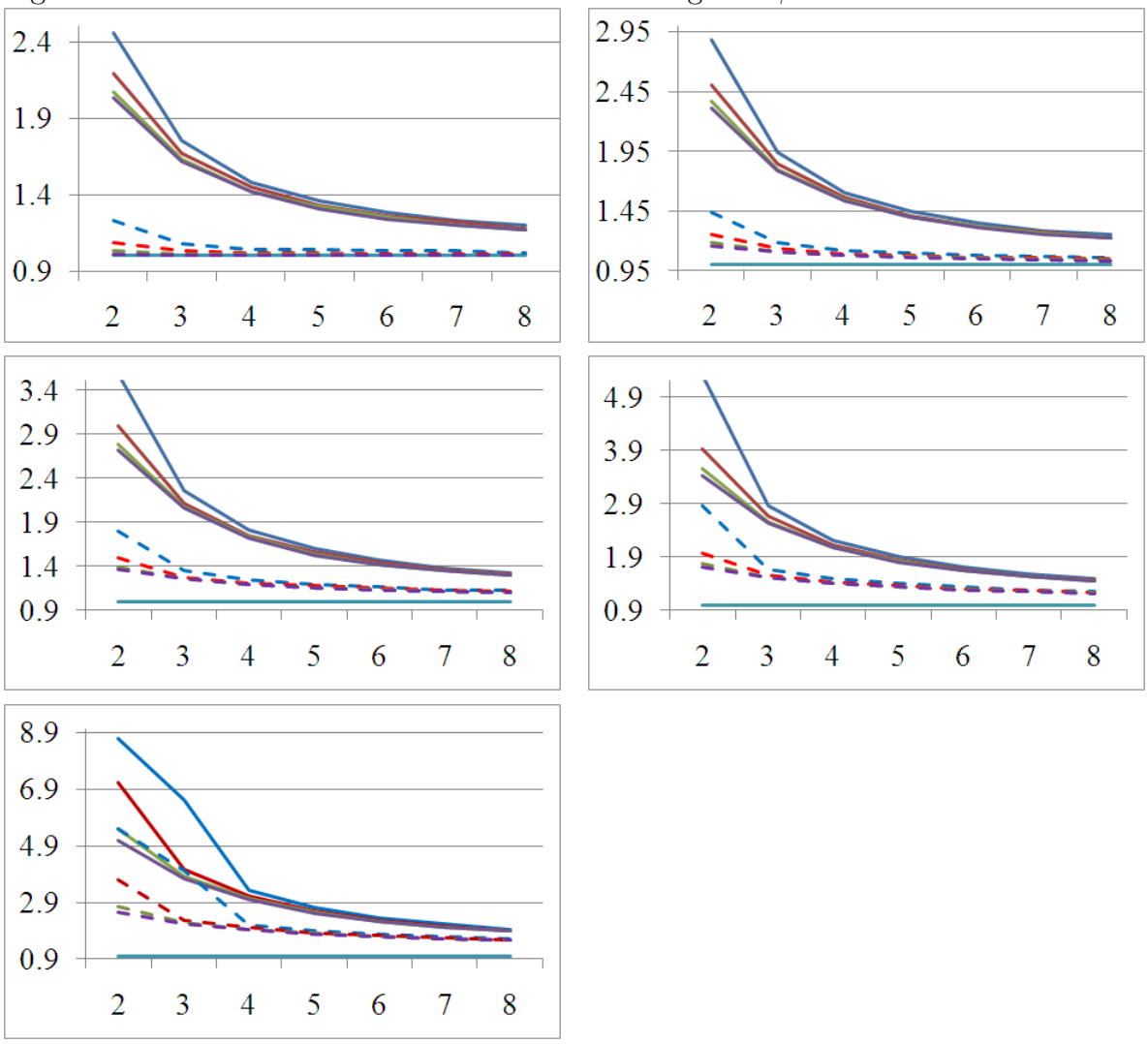
Figure 1.2 summarizes the estimation results for unconditional logit and conditional logit estimators for coefficient β. As the second panel of figure 1.2 shows, for ρ=0, conditional logit estimator for β shows no bias and rejection frequency in hypothesis test with conditional logit estimator for β shows no size distortion regardless of n and T. For ρ=0.2, the bias of conditional logit estimator of β is less than 10% for $T \geq$4. The rejection frequency reveals that a simulation with ρ=0.2 leads to over-rejection by 100% for $n = 400$. The size distortion increases with n. For ρ=0.4, the bias of conditional logit estimator for β is greater than 10% for $\forall T \leq 8$. The rejection frequency reveals that a simulation with

Figure 1.1. Bias for β with continuous covariate



(Note) x-axis: time dimension, y-axis: estimated β. True β is 1. Each color of solid line represents serial correlation coefficient, ρ - blue (0), brown (.2), green (.4), purple (.6), and teal-blue (.8). We use $n = 200$ as benchmark since the bias does not change much as we increases n over 200. The left panel of first row shows the estimates for unconditional logit estimator and the right panel of first row shows the estimates for conditional logit estimator. The left panel of second row shows the estimates from both of conditional (dashed line) and unconditional (solid line) estimators together. More results on simulation with various n, T, ρ are reported in Appendix. "For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation."

Figure 1.2. Bias of unconditional and conditional logit for β with continuous covariate



(Note) x-axis: time dimension, y-axis: estimated β and legend represent number of cross-section units - true (teal), blue (100), red (200), green (400) and purple (600) - by solid line for unconditional logit estimates and by dashed line for conditional logit estimates. ρ represents serial correlation coefficient. Each panel of figure represents estimated β for each ρ. Figures in the first row report estimated β for ρ =0 and ρ=.2. Figures in the second row report estimated β for ρ =.4 and ρ=.6.

ρ=0.4 leads to over-rejection by 700% for $n = 400$. For ρ=0.6, the bias of conditional logit estimator of β is greater than 20% for $\forall T \leq 8$. The rejection frequency with nominal level 0.05 approaches 0.95 and beyond as n increases over 400 with $\forall T \geq 4$. Finally, for ρ=0.8, the bias of conditional logit estimator for β is greater than 50% for $\forall T \leq 8$. The rejection frequency with nominal level 0.05 approaches 1 as n increases over 200 with $\forall T \geq 3$. Table 1.7 shows the magnitude of bias in % as ρ changes using unconditional logit estimates in MC experiment.

Figure 1.1 shows that the increase in T leads to the smaller bias for coefficient estimates but the increase in n does not change the magnitude of finite sample bias for coefficient β especially for $T \geq 3$. Overall, the magnitude of finite sample bias for coefficient β is greater with the higher value of ρ and the smaller T but is invariant to n. The size distortion becomes more severe as n increases and ρ increases while the change in T ($\forall T \geq 3$) does not change the rejection frequency much. Regardless of the lengths of T and n, the size distortion in inference is quite severe for ρ >0.2. In sum, simulation reveals that conditional logit estimator with correlated logistic errors provides biased coefficient estimates and rejection frequency in inference and, therefore, is not robust to the violation of CI assumption by serial correlation especially for ρ >0.2.

Unconditional logit for β: Decomposition of Bias

Left panel in the first row of figure 1.1 shows finite sample bias of unconditional logit for β. The bias for β is greater for smaller T and higher value of ρ while finite sample bias does not change much as n changes in figure 1.2. Furthermore, in table 1.5 and 1.6, using

unconditional logit estimator for β, simulation shows the decomposition of bias into the part due to unobserved heterogeneity(c_i) and the part due to serial correlation(ρ) for $n = 100$ and $n = 200$, respectively.

Bias for β Finite sample bias for unconditional logit is greater with smaller T and larger ρ. For instance, when there is no serial correlation, $\rho = 0$, the magnitude of bias for unconditional logit is greater than 100 % with T=2 and $\forall n \leq 600$ while the magnitude of bias for unconditional logit is about 20 % with T=8 and $\forall n \leq 600$. Even for T as large as 8, finite sample bias of unconditional logit estimator is about 17 % with no serial correlation and finite sample bias for unconditional logit increases up to about 20 %, 30 %, 45 % and 90 % as ρ increases up to .2, .4, .6, and .8 respectively.

Decomposition of bias For $\rho \leq$0.4, finite sample bias due to c_i(unobserved heterogeneity) dominates the bias due to ρ (serial correlation among errors) $\forall n, T$. For $T \geq$5, the bias and the decomposition of bias does not change much as n changes. For instance, in table 1.6, for the case of n=200 and T=8, finite sample bias due to unobserved heterogeneity is about 18 % while finite sample bias due to serial correlation is about 4 %, 13 %, 29 % and 73 % for ρ=.2, .4, .6 and .8 respectively. As ρ increases finite sample biases by both sources are substantial so that it is important to control both unobserved heterogeneity and serial correlation properly. Sometimes, we can control only one source of bias and, in this case, the choice of allowing serial dependence on response variable($D(\mathbf{y}_i|\mathbf{x}_i, c_i)$) or of allowing no restriction on $D(c_i|\mathbf{x}_i)$ should be depending on whether researchers have more information

Table 1.5. Decomposition of bias for β for $n = 100$

	c_i	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$
t=2	146	42	113	286	624
t=3	75	19	51	112	478
t=4	48	12	33	73	186
t=5	36	9	24	54	136
t=6	28	7	19	43	108
t=7	23	5	15	35	90
t=8	20	5	13	29	76

(Note) % of bias are reported in the cells. c_i represents the proportion of bias due to unobserved heterogeneity. ρ represents serial correlation coefficient for errors. Abrevaya [1997] shows that unconditional logit MLE for β has probability limit 2β. Thus, 46 % of bias is due to finite sample bias which has nothing to do with unobserved heterogeneity. With no serial correlation of error, we obtain $\hat{\beta}_{c-logit}$=1.23 and $\hat{\beta}_{unc-logit}$=2.46.

on $D(\mathbf{y}_i|\mathbf{x}_i, c_i)$ or $D(c_i|\mathbf{x}_i)$.[8] For instance, for the case where there exists high persistence of response variables even after conditioning on covariates, pooled methods that allowing serial dependence with parametric assumption on $D(c_i|\mathbf{x}_i)$ probably more appropriate than those methods that impose no restriction on $D(c_i|\mathbf{x}_i)$ and require conditional independence assumption on outcome $D(\mathbf{y}_i|\mathbf{x}_i, c_i)$.

1.4.3 The bias for APEs and rejection frequency: a continuous covariate

Since conditional logit approach avoids specifying individual fixed effects, APEs cannot be obtained for conditional logit estimator. In the estimation of APE for β in (1.1), we study the APEs for CRE logit with pooled MLE under correct specification for $D(c_i|\mathbf{x}_i)$ and we also compare these estimates with APEs for fixed effects(FE), LPM, and unconditional logit with

[8] It is like determining which functional forms assumption either $D(\mathbf{y}_i|\mathbf{x}_i, c_i)$ or $D(c_i|\mathbf{x}_i)$ is causing less distortion in statistical analysis.

Table 1.6. Decomposition of bias for β for $n = 200$

	c_i	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$
t=2	119	31	80	175	493
t=3	67	17	45	99	238
t=4	45	11	30	68	168
t=5	33	8	23	51	128
t=6	26	6	18	41	100
t=7	22	5	14	33	84
t=8	18	4	13	29	73

(Note) % of bias are reported in the cells. c_i represents the proportion of bias due to unobserved heterogeneity. ρ represents serial correlation coefficient for errors. 19 % of bias is due to finite sample bias which has nothing to do with unobserved heterogeneity. With no serial correlation of error, we obtain $\hat{\beta}_{c-logit}$=1.095 and $\hat{\beta}_{unc-logit}$=2.19.

Table 1.7. The bias for β due to serial correlation using conditional logit for $n = 200$

	c_i	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$
t=2	0	25	50	97	269
t=3	0	14	28	57	129
t=4	0	9	21	44	101
t=5	0	8	18	36	84
t=6	0	6	15	31	72
t=7	0	6	13	27	64
t=8	0	5	11	24	57

(Note) % of bias are reported in the cells. c_i represents the proportion of bias due to finite sample estimation error for unobserved heterogeneity. ρ represents serial correlation coefficient for errors.

MLE.[9] The APEs of CRE pooled-MLE and LPM with FE does not impose independence over time while unconditional logit does not impose any restriction on unobserved heterogeneity. Thus, unconditional logit is inconsistent estimator due to both unobserved heterogeneity and serial correlation. Figure 1.3 summarizes MC estimates of APEs for CRE logit with pooled MLE, unconditional logit with MLE and FE LPM. With correct specification of unobserved heterogeneity for CRE logit, only unconditional logit for APEs shows substantial bias. Finite sample bias for APEs of unconditional logit is positively correlated with the magnitude of serial correlation.

CRE logit with pooled MLE: APEs

Left panel of the second row in figure 1.3 summarizes the APEs for CRE logit with pooled MLE. The results with n=200 are reported as benchmark.[10]LPM and CRE logit with pooled MLE show no finite sample bias for APEs and associated rejection frequency. The correct coverage rate of rejection frequency of hypothesis test on APEs of β is obtained regardless of serial correlation (ρ) and the size of time dimension(T).[11]Finite sample bias is less than $6\%(= \frac{.165-.156}{.165})$ $\forall T$ and $\forall \rho$ and is less than 2% for most T and ρ. In short, CRE logit with pooled MLE provides reliable estimates for APEs of β and shows no size distortion in inference under correct specification of unobserved heterogeneity.[12]

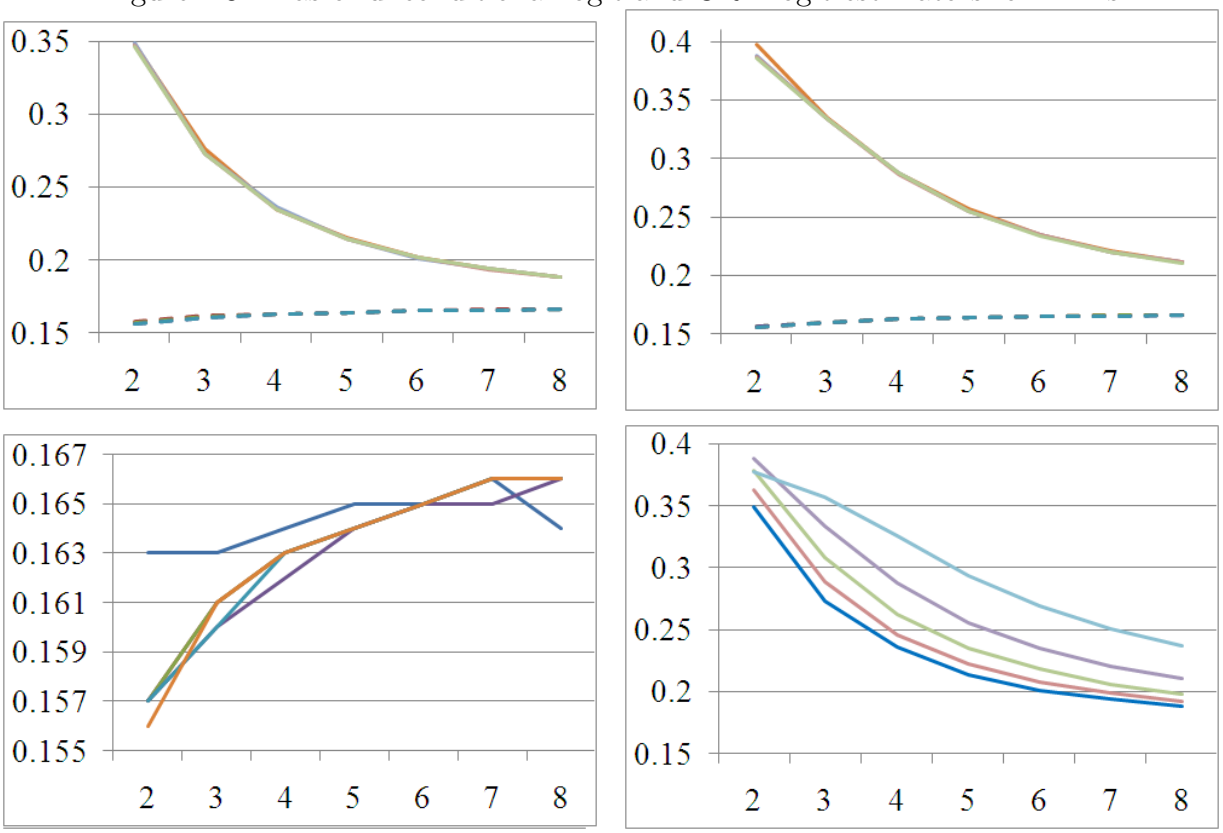
[9] Unconditional logit is logit model where individual fixed effects are estimated as parameters. CRE logit with pooled MLE is logit model with parametric assumption for unobserved heterogeneity as in DGPs of previous section. Pooled estimation is based on the assumption of wrong model.

[10] More results with various n and T are reported in appendix.

[11] The results do not change much as n changes. See appendix for details.

[12] The correct specification of unobserved heterogeneity is implicitly assumed in DGPs since c_i is generated to satisfy CRE logit model.

Figure 1.3. Bias of unconditional logit and CRE-logit estimators for APEs



(Note) x-axis: time dimension, y-axis: estimated β. In figures of first row, solid lines report APEs for unconditional logit and dashed lines report APEs for conditional logit. Color represent estimates with different number of cross-section units from n=100 to n=600. Figures in first row, left panel reports APEs with ρ=0 and right panel reports APEs with ρ=.6. For figures in second row, left panel report APEs for CRE logit with n=200 and various ρ - blue (true), red (0), greed (.2), purple (.4), teal (.6), and orange (.8) and right panel report APEs for unconditional logit with n=200 and various ρ - blue (0), red (.2), greed (.4), purple (.6), and teal (.8).

Unconditional logit with pooled MLE: APEs

Unconditional logit with MLE shows substantial finite sample bias for APEs and shows substantial size distortion in inference. Table 1.9 reports the estimates of β for unconditional logit. Finite sample bias for APEs is greater for smaller T and greater value of ρ. Even for ρ=0, finite sample bias is greater than 15 % $\forall T$. The bias for APEs is greater than 17 %, 42 %, 96 % and 204 % with ρ=.2, .4, .6 and .8, respectively, $\forall T$. The bias in rejection frequency is greater for smaller T and greater value of ρ but does not depend on n.[13]

The bias decomposition for APEs using unconditional logit Table 1.8 reveals that the decomposition of the bias for APEs into the part due to unobserved heterogeneity (c_i) and the part due to serial correlation (ρ) for n=200. The bias by c_i dominates the bias by serial correlation with $\rho \leq$0.6. This implies that, compared to the bias for β, the relative influence to the bias from c_i is greater for APEs. % of bias due to serial correlation decreases as T increases for $T \geq 4$.

In sum, first, serial correlation does not affect APEs for LPM and CRE logit with pooled MLE with correct specification of individual effects. Second, both estimates of LPM and CRE logit with pooled MLE under correct specification of individual effects show no significant finite sample bias for APEs and rejection frequency regardless of n, T and ρ. Third, the estimates of APEs for unconditional logit show substantial bias and its magnitude is positively correlated to serial correlation and negatively correlated to the magnitude of time dimension (T).

[13] See appendix for the results of the bias for rejection frequency.

Table 1.8. Decomposition of bias into two parts using unconditional logit for $n = 200$

	$c_i, \rho = 0$	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$
t=2	114	9	17	24	17
t=3	67	10	21	37	52
t=4	44	6	16	32	55
t=5	30	5	13	25	48
t=6	22	4	10	21	41
t=7	17	3	7	16	34
t=8	15	2	6	14	30

(Note) % of bias of unconditional logit APEs are reported in the cells. c_i represents the proportion of bias due to finite sample estimation error for unobserved heterogeneity. ρ represents serial correlation coefficient for errors. $n = 200$ case was reported as benchmark since very small difference of bias decomposition occurs as we change n. % of bias due to c_i is obtained from $\frac{APE_{unc,\rho=0} - APE_{True}}{APE_{True}}$ and % of bias due to serial correlation (ρ) is obtained from $\frac{APE_{unc,\rho>0} - APE_{unc,\rho=0}}{APE_{unc,\rho>0}}$.

Table 1.9. The bias of unconditional logit $n = 200$

	true	$c_i, \rho = 0$	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$
t=2	.163	.349	.363	.378	.388	.277
t=3	.163	.273	.289	.308	.334	.258
t=4	.164	.236	.246	.262	.288	.326
t=5	.165	.214	.222	.235	.255	.294
t=6	.165	.201	.208	.218	.235	.269
t=7	.166	.194	.199	.206	.220	.251
t=8	.164	.188	.192	.198	.211	.237

(Note) The bias are reported in the cells. $n = 200$ case was reported as benchmark since very small difference of bias decomposition occurs as we change n.

1.4.4 DGP: A model with both binary and continuous covariates

We generate data $\{y_{it}, x_{it}, D_{it}, c_i, u_{it}\}$ for $t = 1, 2, .., T$ and $i = 1, 2, ..., n$ based on model (1.10) with 1,000 replications.

$$y_{it} = 1[x_{it}\beta + D_{it}\delta + c_i + u_{it} \geq 0] \tag{1.10}$$

We draw the following five items as below. The correlation across time is induced through error components $\mathbf{u}$ as in section 1.3.1.

1. $x_{it} \sim iid\ N(0, 1)$. We generate an univariate covariate from identical and independent standard normal distribution.

2. $D_{it} = 1(d_{it} - .5 \geq 0)$ where $d_{it} \sim iidUnif(0, 1)$

3. Unobserved heterogeneity:

$$D(c_i|x_i) \sim \frac{D_{i1} + D_{i2} + ... + D_{ik}}{\sqrt{2k}} + \frac{a_i}{\sqrt{\frac{\pi^2}{3}}} \tag{1.11}$$

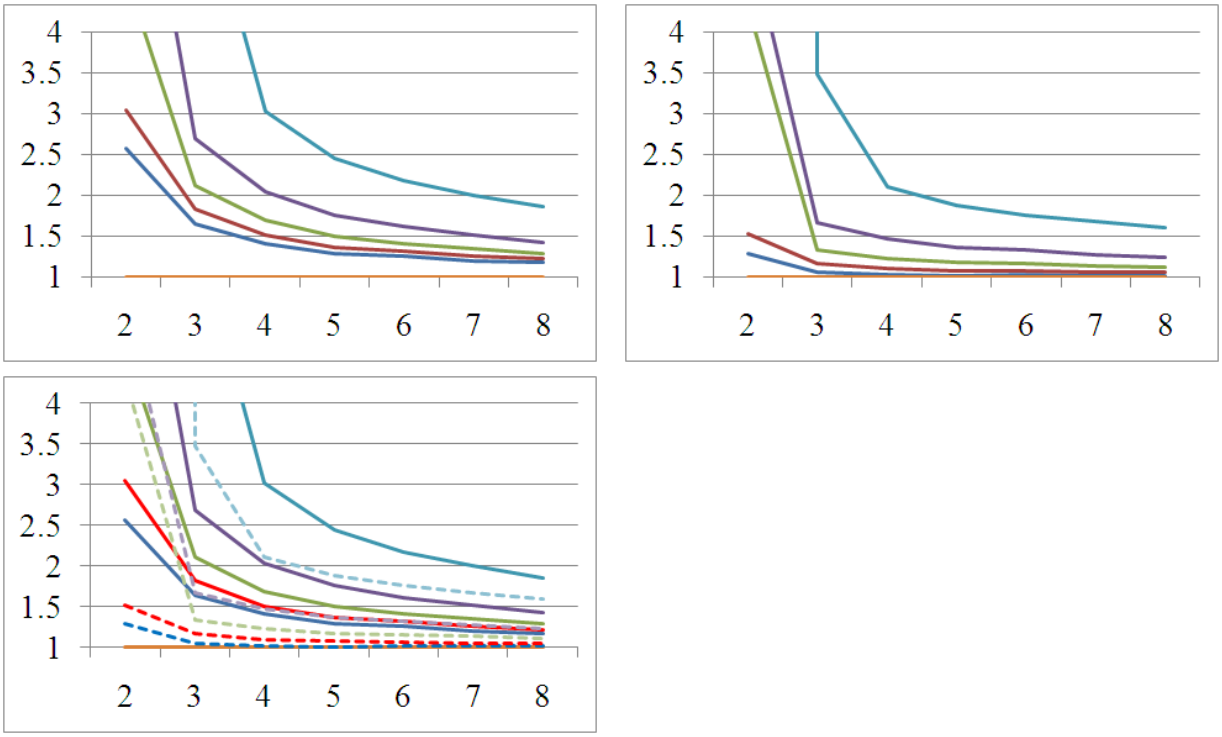
$$\text{where } a_i \sim \text{logistic}(0, \frac{\pi^2}{3}) \text{ and } [k = \frac{T}{2}].$$

4. We generate each u_{it} such that it has standard logistic marginal distribution (i.e. $u_{it} \sim \log istic(0, \frac{\pi^2}{3})$), $Cov(u_{it}, u_{is}) = \rho^{|t-s|}$ for $t \neq s$, and joint distribution of $(u_1, u_2, ..., u_T)$ is logistic.

5. We generate y_{it}

$$y_{it} = 1[x_{it}\beta + D_{it}\delta + c_i + u_{it} \geq 0] \tag{1.12}$$

Figure 1.4. Bias of unconditional and conditional logit estimators for β with binary covariate



(Note) x-axis: time dimension, y-axis: The estimates for β are reported for $n = 100$. The color of line represents serial correlation coefficient - orange (true), blue (0), red (.2), green (.4), purple (.6), teal-blue (.8). The left panel of first row shows the estimates for unconditional logit estimator and the right panel of first row shows the estimates for conditional logit estimator. The left panel of second row shows the estimates for both conditional (dashed line) and unconditional (solid line) estimators together.

using $x_{it}, D_{it}, c_i, u_{it}$ as in 1,2,3,4, β=.2 and δ =1.

1.4.5 The bias and size distortion for δ: binary covariate

Conditional logit for δ

As figure 1.4 shows, in the MC experiment with binary covariate, we reach the same conclusion as in the MC experiment with a continuous covariate. Smaller T and greater value of ρ

Table 1.10. Decomposition of bias into two parts $n = 100$

	c_i	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$
t=2	157	48	212	486	14343
t=3	65	18	47	104	450
t=4	41	10	28	63	162
t=5	29	8	21	47	116
t=6	26	6	15	36	92
t=7	20	6	15	32	80
t=8	18	4	11	25	68

(Note) % of bias are reported in the cells.

leads to greater magnitude of finite sample bias for coefficient estimate, δ, but the change in n does not affect the bias much. The size distortion becomes more severe as n and ρ increase while the change in $\forall T \geq 3$ does not have much effect on the rejection frequency. Regardless of the values of T and n, the size distortion is quite severe for ρ >0.2. Conditional logit estimator for δ and rejection frequency are biased and, therefore, are not robust to violation of CI assumption for ρ >0.2, $\forall n, T$.

Unconditional logit for δ: Decomposition of Bias

We obtain the same conclusion as DGP with a continuous covariate although the magnitude of bias is greater with a binary covariate especially for $T \leq 4$. The bias of unconditional logit for δ is greater for data with smaller T and higher value of ρ while the bias does not change much as n changes as shown in figure 1.4. Table 1.10 shows that, for $\rho \leq$0.4, the bias due to c_i(unobserved heterogeneity) dominates the bias due to ρ (serial correlation among errors) $\forall T$.

1.4.6 The bias for APEs and rejection frequency: binary covariate

In the estimation of APEs for β in (1.12), we study the APEs for CRE logit with pooled MLE and compare these estimates with APEs for LPM and unconditional logit with MLE. Figure 1.5 summarizes MC estimates for APEs of CRE logit with pooled MLE, unconditional logit with MLE and LPM.

As the left panel in the second row of figure 1.5 shows, LPM and CRE logit with pooled MLE exhibit no substantial finite sample bias of APEs for δ regardless of serial correlation (the magnitude of ρ) and size of T. The rejection frequency also shows no significant size distortion regardless of the magnitude of serial correlation (ρ) and size of T.[14] In short, CRE logit with pooled MLE provides reliable finite sample estimates for APEs of δ and a valid inference under correct specification of unobserved heterogeneity.

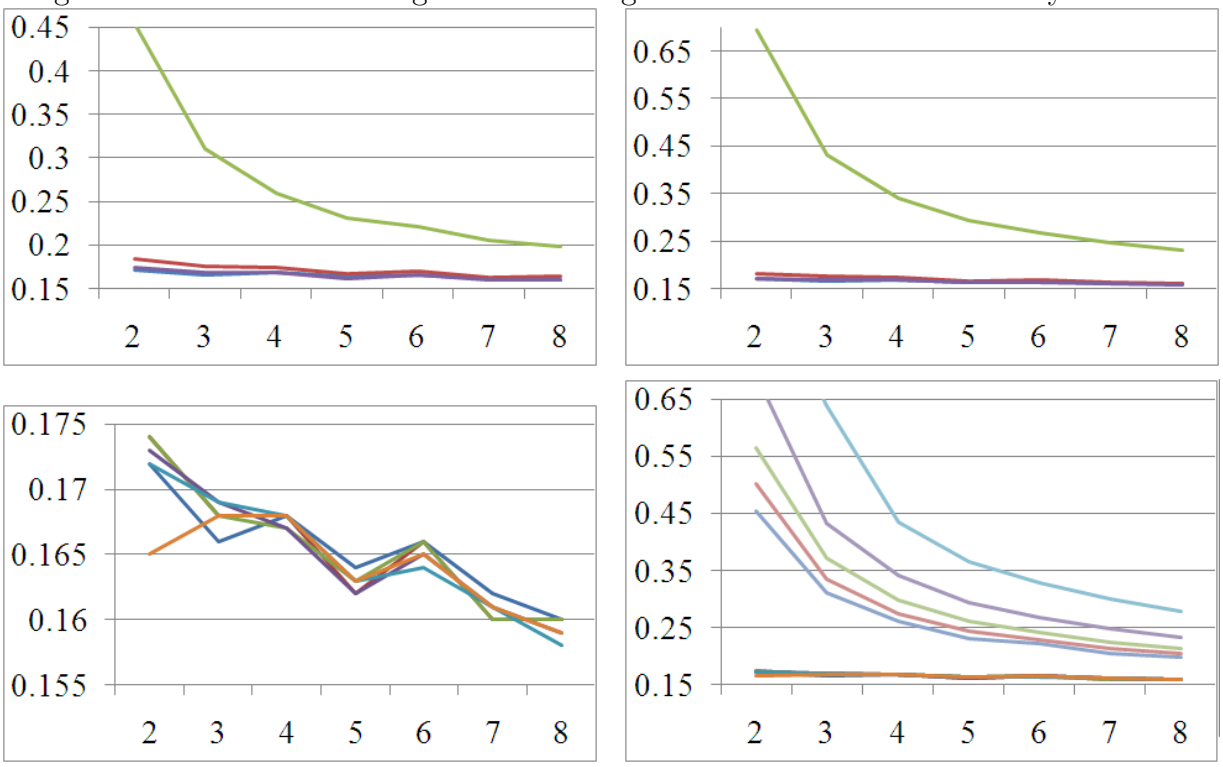
The bias decomposition for APE: binary covariate

Unconditional logit with MLE shows substantial bias for APEs and rejection frequency. The right panel in the second row of figure 1.5 shows the bias of APE of δ for unconditional logit. The bias for APE is greater for smaller T and larger value of ρ. Even for ρ=0, the magnitude of bias for APE of δ is greater than 24 % $\forall T$. The bias increases with the magnitude of ρ for all T. The bias for rejection frequency is greater for smaller T and larger value of ρ.[15]

[14] The results are reported in appendix.

[15] See appendix for further results of the bias for rejection frequency.

Figure 1.5. Unconditional logit and CRE-logit estimates of APEs for binary covariate



(Note) x-axis: time dimension, y-axis: APEs for β with $n = 100$ are reported. ρ represents serial correlation coefficient. Figures in first row report APEs with ρ=0 (left) and ρ=.6 (right) respectively. The colors of solid line represent various APE estimators; blue (true), red (LPM), green (unconditional logit), and purple (CRE logit). As we see, purple completely cover blue. The left panel of second row shows the estimates for CRE logit estimator and the right panel of second row shows the estimates for unconditional logit estimators with various serial correlation coefficient; blue (0), red (.2), green (.4), purple (.6) and teal-blue (.8).

Table 1.11. Decomposition of bias using unconditional logit for $n = 100$

	c_i	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$
t=2	164	27	65	140	255
t=3	87	14	37	73	197
t=4	55	7	23	48	104
t=5	41	7	18	38	82
t=6	33	4	13	28	65
t=7	27	5	12	26	59
t=8	25	4	10	22	51

(Note) % of bias of unconditional logit APEs are reported in the cells. c_i represents the proportion of bias due to finite sample estimation error for unobserved heterogeneity. ρ represents serial correlation coefficient for errors. $n = 100$ case was reported as benchmark since very small difference of bias decomposition occurs as we change n. % of bias due to c_i is obtained from $\frac{APE_{unc,\rho=0} - APE_{True}}{APE_{True}}$ and % of bias due to serial correlation (ρ) is obtained from $\frac{APE_{unc,\rho>0} - APE_{unc,\rho=0}}{APE_{unc,\rho>0}}$.

1.5 Panel data models for married women's welfare participation

In this section, we illustrate the application of conditional logit for β and CRE logit with pooled MLE for APEs using the Survey of Income and Program Participation (SIPP) data on the welfare participation of married women from Chay and Hyslop [2001].[16] As welfare participation of married women exhibits substantial serial persistence over time, we apply CRE logit with pooled MLE to obtain APEs of marital status and the number of children on welfare participation.

Welfare participation data are from 1990 panel of the SIPP which contains eight waves at four month intervals covering a 32-month period. The data set contains 1,934 married

[16] Pooled MLE with CRE logit model does not assume that the full conditional density of y give $\mathbf{x}, c$ is correctly specified but it is very useful since full conditional density of y give $\mathbf{x}, c$ is likely to be very complicated and pooled MLE does not restrict on serial dependence.

women over 8 periods. The sample contains women who are 18-65 years old and either receive AFDC payments during or before the sample period or whose average total family income during the sample period is below the family-specific average poverty level.[17]Welfare participation status shows clear serial correlation over time. For instance, out of 1,934 married women, 1,076 married women never participate in welfare program and 364 married women participate in welfare program for all sample periods. Basic statistics for variables which is used in estimation are provided in table 1.12.

We start estimation with static model of (1.13) for simplicity. We control unobserved heterogeneity with chamberlain device while we do not control for true state dependence explicitly.

$$participation_{it} = 1[\mathbf{z}_{it}\beta + f_t + c_i + u_{it} \geq 0],\ i = 1, 2, ..., n,\ \text{and}\ t = 1, 2, ..., T \qquad (1.13)$$

where $\mathbf{z}_{it}$ includes black race dummy, years of education, family poverty level status, age/10, $age^2/100$, marital status and number of children age less than 18. And f_t is period fixed effects.

[17]In summer 1993, the Nation had 36 million mothers 15 to 44 years old; 3.8 million of them (10 percent) were receiving AFDC (Aid to Families with Dependent Children) payments to help with the rearing of 9.7 million children. An additional 0.5 million women over 45 years old and 0.3 million fathers living with their dependent children also received AFDC.

Table 1.12. Basic statistics

	Obs	Mean	S.D	Min	Max
AFDC	15,472	.302	.459	0	1
married	15,472	.291	.454	0	1
kids	15,472	1.849	1.479	0	10
Black	15,472	.310	.463	0	1
Educ	15,472	11.260	2.700	0	18
Poverty	15,472	1026.426	352.849	519.755	2377.113
age/10	15,472	3.405	1.149	1.6	6.3
$age^2/100$	15,472	12.917	8.864	2.56	39.69

Note: There is no missing data. 15,472 = 8×1,934

1.5.1 Static panel data model for married women's welfare participation

Table 1.13 shows that allowing unobserved heterogeneity to be correlated with marital status and number of children has important effects on estimated APEs. LPM provides estimated coefficients of -0.27 and 0.05 on marital status and number of children, respectively, with both coefficients being statistically significant. Each child is estimated to increase the welfare participation probability by about 0.05 while being married is estimated to reduce the welfare participation probability by about 0.27. If we use logit and assume unobserved heterogeneity is independent of explanatory variables, the estimated APEs are a little bit higher for both marital status and number of children. When we use CRE logit model with pooled MLE which controls for unobserved heterogeneity, we obtain APEs that are very similar to the estimates by LPM. Moreover, the coefficients on the time averages of time-varying covariates are statistically significant and practically large. All these results imply that accounting for unobserved heterogeneity correctly is very important. CRE logit model with full MLE reveals the same conclusion as with partial MLE. Using GEE with logit link and correlated random

effect provides very similar estimates to the estimates for CRE logit model with full MLE.

Finally, column (6) contains the estimates from conditional logit. Unfortunately, the coefficient magnitudes are difficult to interpret because of scaling factor. The ratio of coefficients using conditional logit is $-\frac{2.80}{0.76} = (-3.69)$, which is quite different from the one, $-\frac{1.76}{0.34} = (-5.18)$, from CRE logit model with Pooled MLE. Moreover, if we do not control for unobserved heterogeneity, the ratio of coefficient for RE logit with Pooled MLE is $-\frac{2.28}{0.54} = (-4.22)$. The discrepancy for the ratio of coefficients between CRE logit and RE logit which does not control unobserved heterogeneity is possibly caused by either misspecification for the distribution of unobserved heterogeneity or uncontrolled unobserved heterogeneity. Moreover, the discrepancy for the ratio of coefficients between CRE logit and conditional logit is possibly caused by either misspecification for the distribution of unobserved heterogeneity or serial correlation. Therefore, for instance, if we assume distribution of unobserved heterogeneity is correct with Chamberlain's device, then we can infer that the discrepancy is originated from the bias by serial correlation for conditional logit and the bias by unobserved heterogeneity for RE logit. Moreover, in this case, because CRE logit with pooled MLE in column (3) allows arbitrary weak dependence, it seems sensible to rely on the estimates of CRE logit with pooled MLE under the assumptions that marital status and number of children variables are strictly exogenous conditional on unobserved heterogeneity and the assumption which unobserved heterogeneity can be specified by the time averages of time-varying covariates. However, without correct specification on $D(c_i|\mathbf{x}_i)$, no procedure strictly dominates the other. In particular, the choice among these procedures should be based on whether we are more confident for the correct specification about $D(\mathbf{y}_i|\mathbf{x}_i, c_i)$ or $D(c_i|\mathbf{x}_i)$.

Table 1.14 shows results for probit model using the same estimation methods as in table 1.13. Although standard errors are smaller for probit model, we can draw the identical conclusion as using logit model.

Table 1.13. Logit estimation for married women's welfare participation

	(1)	(2)		(3)		(4)		(5)		(6)
Model	Linear	logit		CRE logit		CRE logit		CRE logit		FE logit
	FE	pooled	MLE	pooled	MLE	GEE		Full	MLE	MLE
Estimation	coef	coef	APE	coef	APE	coef	APE	coef	APE	coef
married	-.2707	-2.276	-.3704	-1.757	-.2851	-1.591	-.2613	-3.305	-.2293	-2.802
	(.0278)	(.152)	(.0211)	(.196)	(.0308)	(.185)	(.0298)	(.248)	(.0122)	(.242)
kids	.0524	.539	.0877	.343	.0557	.294	.0483	.724	.0671	.761
	(.0098)	(.056)	(.0086)	(.060)	(.0096)	(.054)	(.0088)	(.091)	(.0084)	(.104)
$\overline{married}$				-.613		-.723		-1.762		
				(.231)		(.230)		(.349)		
$\overline{kids}$				.231		.242		.547		
				(.062)		(.061)		(.111)		
$\frac{1}{\sqrt{1+\hat{\sigma}^2}}$								.242		
log likelihood		-7571.91		-7552.25				-4028.83		-1412.62
number of women	1934	1934		1934		1934		1934		494

*Note: Cluster robust standard errors are reported in parentheses below all coefficients and APEs. All model include a full set of period dummy variables, time constant variables years of education, black race dummy variable, age/10, and $age^2/100$ which are not reported.

Table 1.14. Probit estimation for married women's welfare participation

	(1)	(2)		(3)		(4)		(5)		(6)
Model	Linear	probit		CRE probit		CRE probit		CRE probit		FE logit
	FE	pooled	MLE	pooled	MLE	GEE		Full	MLE	MLE
Estimation	coef	coef	APE	coef	APE	coef	APE	coef	APE	coef
married	-.2707	-1.251	-.3490	-.951	-.2644	-.906	-.2525	-1.870	-.2078	-2.802
	(.0278)	(.078)	(.0189)	(.103)	(.0281)	(.105)	(.0289)	(.134)	(.0124)	(.242)
kids	.0524	.309	.0861	.198	.0550	.177	.0492	.400	.0590	.761
	(.0098)	(.033)	(.0086)	(.035)	(.0097)	(.032)	(.0091)	(.052)	(.0078)	(.104)
$\overline{married}$				-.354		-1.304		-.726		
				(.126)		(.134)		(.217)		
$\overline{kids}$				.132		.404		.290		
				(.036)		(.037)		(.071)		
$\frac{1}{\sqrt{1+\hat{\sigma}^2}}$								.349		
log likelihood		-7600.63		-7580.68				-3976.44		-1412.62
sample	1934	1934		1934		1934		1934		494

*Note: Cluster robust standard errors are reported in parentheses below all coefficients and APEs. All model include a full set of period dummy variables, time constant variables years of education, black race dummy variable, age/10, and $age^2/100$ which are not reported.

1.5.2 Dynamic Unobserved effects logit models under strict exogeneity

Instead of controlling serial correlation by the models with pooled MLE methods, we probably control serial dependence among $\mathbf{y}_i$ by using a dynamic model with full MLE. Chay and Hyslop [2001] emphasizes the importance of introducing dynamic models to seriously consider both unobserved heterogeneity and true state dependence in welfare analysis. Explicitly distinguishing spurious sate dependence by individual permanent characteristics and true state dependence is possible only with dynamic models. Our treatment for dynamic model follows Wooldridge [2005] and Wooldridge [2010].

Suppose we date our observations starting at $t = 0$, so that y_{i0} is the first observation on outcome, y. For t = 1,...,T, we are interested in dynamic model of (1.14).

$$E(y_{it}|y_{i,t-1}, ..., y_{i0}, \mathbf{z}_i, c_i) = G[\rho y_{it-1} + \mathbf{z}_{it}\beta + c_i + u_{it} \geq 0] \tag{1.14}$$

where $\mathbf{z}_{it}$ is a vector of contemporaneous explanatory variables, and G can be the probit or logit function. There are several important points about this model. First, $\mathbf{z}_{it}$ is assumed to satisfy a strict exogeneity assumption (conditional on c_i). Second, the probability of success at time t is allowed to depend on the outcome in $t-1$ as well as unobserved heterogeneity, c_i.

Following Chay and Hyslop [2001], we focus the specification of dynamic model to investigate following two points. First, we are interested in hypothesis test with the null, H_0 : ρ=0. $\rho \neq 0$ implies the existence of state dependence after controlling for the unobserved heterogeneity, c_i. This provides information about the decomposition of welfare

participation persistence to state true dependence and unobserved heterogeneity. Second, the estimation for ρ and the marginal effects of marital status and number of children on welfare participation. We start with the specification of dynamic model of (1.15).

$$f(y_1, y_2, ...y_T|y_0, \mathbf{z}, c, \theta) = \prod_{t=1}^{T} f(y_t|y_{t-1}, ..., y_0, \mathbf{z}_t, c, \theta) \tag{1.15}$$

We condition on y_{i0} as well to solve the initial conditions problem. Furthermore, we propose a $D(c_i|y_0, \mathbf{z}_i)$ as in (1.17) to obtain $f(y_1, y_2, ..., y_T|y_0, \mathbf{z}_i)$ using (1.16). Given a density $h(c|y_0, \mathbf{z}, \delta)$, we have

$$f(y_1, y_2, ...y_T|y_0, \mathbf{z}, c, \theta) = \int_{\infty}^{\infty} f(y_1, y_2, ...y_T|y_0, \mathbf{z}, c, \theta) h(c|y_0, \mathbf{z}, \delta) dc \tag{1.16}$$

where G is standard normal CDF for probit model and

$$h(c|y_0, \mathbf{z}, \delta) \text{ is Normal}(\psi + \varsigma_0 y_{i0} + \mathbf{z}_i\xi, \sigma_a^2) \tag{1.17}$$

where $\mathbf{z}$ include $AFDC_{t-1}$, number of children, marital status, and family poverty level.

We can rewrite our estimation equation as in (1.18) using chamberlain device.

$$y_{it} = 1[\psi + \rho y_{it-1} + \mathbf{z}_{it}\beta + \varsigma_0 y_{i0} + \mathbf{z}_i\xi + a_i + u_{it} \geq 0] \tag{1.18}$$

Pooled MLE does not consistently estimate the scaled coefficients or the APEs in estimating dynamic model. This is because while $P(y_{it} = 1|\mathbf{z}_i, y_{it-1}, ..., y_{i0}, a_i)$=$\Phi(\psi + \rho y_{it-1} + \mathbf{z}_{it}\beta + \varsigma_0 y_{i0} + \mathbf{z}_i\xi + a_i)$ is true, $P(y_{it} = 1|\mathbf{z}_i, y_{it-1}, ..., y_{i0})$=$\Phi(\psi_a + \rho_a y_{it-1} + \mathbf{z}_{it}\beta_a + \varsigma_{0a} y_{i0} + \mathbf{z}_i\xi_a)$ does not hold true unless a_i is identically zero. Therefore, only random effect probit consistently estimate the scaled coefficients and the APEs. For comparison purpose we

also report CRE probit with pooled MLE and RE probit with MLE which does not control unobserved heterogeneity in table 1.15.

The column (4) in table 1.15 for dynamic model provides that CRE probit with MLE for $\hat{\rho}$ which is 1.290(s.e=.072) and is statistically different from zero even after controlling for unobserved heterogeneity. This is strong evidence for the existence of true state dependence. The computation of APE for $AFDC_{t-1}$ is based on averaging $\Phi(\hat{\psi_a}+\hat{\rho_a}+\mathbf{z}_{it}\hat{\beta_a}+\hat{\varsigma_{0a}}y_{i0}+\mathbf{z}_i\hat{\xi_a})$ - $\Phi(\hat{\psi_a} + \mathbf{z}_{it}\hat{\beta_a} + \hat{\varsigma_{0a}}y_{i0} + \mathbf{z}_i\hat{\xi_a})$ across all i and t.[18] CRE probit with full MLE for the APE of $AFDC_{t-1}$ is provided in the model (4) of table 1.15 and is about .175. In other words, averaged across all women and all time periods, the probability of participating in welfare at time t is about .175 higher if the women was in welfare at time $t-1$ than if she was not. This estimate controls for unobserved heterogeneity, number of children, marital status, family poverty level and women's education, race, age, age^2. The column (2) of table 1.15 reports the APEs for random effect probit with MLE which does not control unobserved heterogeneity. Estimated APE for state dependence is .384 which is higher than random effects estimates for which heterogeneity is controlled for. This implies that much of observed serial dependence in welfare participation could be attributed to unobserved heterogeneity. ($\frac{.384-.175}{.384} = .544$) About 54 % of observed serial dependence of welfare participation is due to heterogeneity while about 46 % of observed serial dependence is due to state dependence. In other words, simple dynamic RE probit model that ignore unobserved heterogeneity exaggerate true state dependence by 54 %. In sum, the analysis with dynamic model implies that both sources of persistent welfare participation are important since the

[18] Original coefficients are scaled by $\frac{1}{\sqrt{1+\hat{\sigma}^2}}$ =1, where $\hat{\sigma}^2$=.00000045 in the calculation of all APEs.

difference between the estimates clearly shows that welfare participation is depending on individual permanent characteristics as well as "narcotic" effects of welfare program.

Table 1.15. Probit estimation of dynamic model for married women's welfare participation

	(1)	(2)		(3)		(4)		(5)
Model	Linear	RE probit		CRE probit		CRE probit		FE logit
	FE	Full	MLE	pooled	MLE	Full	MLE	MLE
Estimation Method	coef	coef	APE	coef	APE	coef	APE	coef
$AFDC_1$	.3703	2.028	.3840	1.290	.1343	1.290	.1747	1.643
	(.0163)	(.073)	(.0058)	(.063)	(.0085)	(.062)	*	(.097)
married	-.1922	-.988	-.1076	-1.595	-.1545	-1.595	-.1545	-2.506
	(.0230)	(.094)	(.0057)	(.143)	(.0115)	(.136)	*	(.303)
kids	.0350	.211	.0220	.280	.0180	.280	.0178	.670
	(.0075)	(.037)	(.0024)	(.059)	(.0034)	(.051)	*	(.129)
$\frac{1}{\sqrt{1+\hat{\sigma}^2}}$		.734				.999		
unobs heterogeneity	controlled	not controlled		controlled		controlled		controlled
log likelihood		-2624.13		-1658.99		-1658.99		-936.45
sample	1934	1934		1934		1934		441

Note: Cluster robust standard errors are reported in parentheses below all coefficients and APEs. All models include a full set of period dummy variables, time constant variables years of education, black race dummy variable, age/10, $age^2/100$, initial period value for AFDC and married and kids variables for all periods which are not reported.

1.6 Conclusion

This paper examines the robustness of the conditional logit estimator to the violation of the CI assumption using correlated logistic errors in a binary panel data model with correlated individual fixed effects. Simulation results show that the conditional logit method is not robust to violation of the CI assumption. The magnitude of bias for coefficient is greater for the data with a smaller time dimension (T) and a higher serial correlation coefficient. (ρ) Under correct specification of unobserved heterogeneity, CRE logit with pooled MLE provides reliable estimates for APEs with no significant finite sample bias. Simulation for the rejection frequency of the hypothesis test shows evidence that CRE logit provides a reliable test in the presence of high serial correlation. Finite sample bias due to both unobserved heterogeneity and serial correlation is substantial. An empirical example of welfare participation which exhibits high persistence shows that both sources of bias are important. The decomposition shows that bias due to uncontrolled unobserved heterogeneity is about 54% and that due to serial correlation is about 46%. This implies the importance of accounting for both state dependence and unobserved heterogeneity.

Chapter 2

The MI-IPW method in an unbalanced panel data model: An empirical application on the effect of class size reduction on SAT scores for students in grades K-3

2.1 Introduction

Randomized experiments have become more popular in evaluating the impact of social programs recently. Angrist and Pischke [2009] reflects this trend by arguing that the causal inference for the effect of a program can be best answered by a randomized experiment.

However, perfect implementation of a randomized trial is extremely rare. In most experiments, experimental violations occur typically through noncompliance and attrition and can invalidate any statistical conclusions drawn from them. The possible bias from noncompliance can be addressed by the IV method using initial treatments as IV for actual treatments. However, standard methods for missing data, which include complete-case analysis, simple imputation and last-observation-carried-forward analysis, are not appropriate for a valid inference without assuming missing completely at random (MCAR). [1] Unfortunately, in most empirical applications, the MCAR assumption is unrealistic and the violation inevitably leads to biased estimates, incorrect standard errors and invalid inferences for unweighted complete-case analysis.[2]

In this chapter, we focus on the marginal treatment effect (MTE) of a program in a random experiment with ideal design at the beginning but subsequent missing data at the implementation stage. The purpose of this article is twofold. First, we provide robust Hausman tests of the missing completely at random (MCAR) assumption in an unbalanced linear panel data model for least-square (LS), fixed-effects (FE) and first-differencing (FD) estimators for which the rejections of tests imply invalid statistical analysis. Proposed Wald statistics for robust Hausman tests are based on comparing two consistent estimators under the null and it is robust in the sense that estimators used in the construction of the Wald statistics do not assume conditions for the estimators to be efficient.[3] The idea of the test

[1] MCAR and missing at random (MAR) assumption is in the sense defined by Rubin [1976] and Little and Rubin [2002]. The MCAR implies that the probability of missing is independent of both observed and unobserved variables. MAR implies that, conditional on observed data, the probability of missing does not depend on the unobserved data.

[2] Data is complete in the panel data for a particular period if a cross-section unit has observations for outcome and all covariates at that period.

[3] Original Hausman tests(Hausman [1978]) which compare random effect estimator and

is that, under the null, the difference in the estimates of two consistent estimators should be entirely due to sampling errors. This is an extension of a robust Hausman test of a balanced panel in [Wooldridge, 2010, p.321-334] to unbalanced panel data. We also conduct a Monte Carlo experiment of regression based variable addition tests for fixed-effects and first-differencing estimators to test the MCAR assumption.

Second, we introduce practical methods with valid statistical analyses for handling missing data under the missing at random (MAR) assumption which include the inverse probability weighted (IPW) method (Horvitz and Thompson [1952] and Robins et al. [1994]), the likelihood-based maximum likelihood (ML) method (Rubin [1987] and Little and Rubin [2002]) and the multiple imputation (MI) method (Buuren et al. [1999] and Schafer and Graham [2002]). Many microeconomic applications, that typically use the linear models involve various types of high-dimensional covariates which include continuous and categorical variables of various types (binary, count, fraction and so forth). Moreover, data can be missing for all covariates and outcome. Therefore, the likelihood based ML and the MI methods are not appropriate for unbalanced panel data with these types of missing variables since the ML and the MI methods require modeling the joint distribution of missing variables and the correct specification of a model for joint distribution would be extremely difficult, if not impossible. In many applications, although joint normality is assumed, we have very little information on the joint distribution of missing variables. On the other hand, the IPW method with missing outcome and covariates does not need to specify a fully parametric model for joint distribution of missing variables since probability weights can be modeled by a univariate normal (probit) or a logistic (logit) distribution.

fixed effect estimator assume efficiency for random effect estimator.

Meanwhile, the IPW method under MAR is criticized mainly due to the following three concerns. First, the IPW method is generally less efficient than the MI method with correct specification for missing variables since the IPW method only uses complete-case data while the MI method uses information in covariates and outcome even from incomplete case samples. (Clayton et al. [1999]) Therefore, for instance, if a linear regression model has high-dimensional covariates and only one variable in the covariates has missing data, then the loss of efficiency for the IPW method should be significant. However, for monotone missing data (i.e. attrition), there is no loss of information due to this factor since data missing occurs in outcome and all covariates simultaneously for a cross-section unit.[4] The second concern is that IPW estimates can be very unstable if the estimated probability weights are close to zero for certain subpopulation units(Little and Rubin [2002]). However, we can verify the relevance of this criticism by probing the estimates with actual data in the application. Finally, the IPW estimates can be sensitive to the model specification for the probability of selection (or missing). We provide an example of using an economic theory in the specification of the selection process and the MAR assumption. Particularly, in the empirical example, we extend Becker [1993]'s model of parent's school choice between public and private schools to specify the selection process that satisfies the MAR assumption. For simplicity, we use

[4] Recently, Robins and Rotnitzky [1995] and Scharfstein et al. [1999] propose an AIPW (i.e. doubly robust) estimator which allows to construct a theoretically efficient version of IPW so that the extension of IPW to AIPW estimator can overcome inefficiency problem at least asymptotically. An estimator is called doubly robust if it is consistent when either a model for conditional expectation(imputation model for missing variables) or a model for probability of selection is correctly specified. An AIPW estimator can attain efficiency in the sense that its variance can reach semi-parametric efficiency bound among the class of estimators that use $\sum_{i=1}^{n}\sum_{t=1}^{T} s_{it} \cdot moment_{it}(Observed_{it}, \theta) = 0$. More discussion for efficiency properties of AIPW estimators is provided in Bickel et al. [1998], Tsiatis [2006], Robins et al. [1994], and section in 2.4.1.

the logit or the probit models for the distribution of probability of selection.[5] Unfortunately, it is very hard to test how inaccurate the approximation of the specification for predictors and the logit or the probit models for probability of selection are in practice so we conduct a Monte Carlo simulation to provide evidence on the sensitivity of the IPW estimator to selection model misspecification.

We explicitly categorize unbalanced panel data into four types based on missing patterns. The four data types are the following: (i) complete case data, (ii) individual units leave the sample but re-enter the sample later (type-I missing data), (iii) individual units are in the sample but some variables are missing (type-II missing data), and (iv) attrition (type-III missing data). For type-I missing data, we drop the observations and do not use in the estimations while we assume these types of data missing are completely random or negligibly small. For type-II missing data, we apply multiple imputations based on the missing variable types. Finally, we propose a method which applies MI and IPW sequentially. Specifically, after multiple imputations of type-II missing variables, we treat type-II missing data as if they are complete case data. We then combine imputed type-II missing data with complete case data and hen apply the IPW method.

Tests of MCAR and the estimations with unweighted, IPW and MI-IPW methods under the MAR are conducted to gauge the effect of class size reduction (CSR) on SAT scores for students in K-3 grades using Tennessee's Project STAR. We employ Tennessee's Project STAR to evaluate the MI-IPW method in the application since it is a randomized experi-

[5] Nonparametric estimation of the probability of selection is also available from Hirano et al. [2003] but it suffers from the curse of dimensionality in practical applications.(Hall and Presnell [1999]).

ment with significant numbers of noncompliance and attrition.[6] For instance, 13% of 11,601 students switched from treatment groups to the control group or vice versa (i.e. 13% of the sample are defiers) and about 45% of the students either entered the program after kindergarten or left before the program ended and about 5% of the participating students had missing variables during in the STAR program. Moreover, potential differential attrition across treatment and control groups appears in basic statistics which show that a larger fraction of students left the sample in regular and regular-with-aide classes than from small classes for all grades from K to 3. As reported in Table 2.27, the difference in the average scores between students who stayed and drop-out students is greater in small classes than the difference in regular and regular-with-aide classes for both first and second grades although the differences are small. This implies that a relatively greater proportion of better performing students left the program from regular and regular-with-aide classes than from small classes. This is informal evidence of differential drop-out rates of students across different types of classes and suggests possible overestimation in the effect of CSR on SAT scores. Therefore, the MCAR assumption is possibly violated and a standard estimation with complete-case is probably inappropriate.[7]

Our application of proposed tests and estimations with Project STAR data proceed as follows. First, we apply formal tests of differential attrition across class types which lead to violation of the MCAR assumption. We test the MCAR assumption using Hausman-type tests and regression based variable addition tests using FD and FE estimators. In

[6] Tennessee's Project STAR is a longitudinal data with a single cohort of 11,601 students for grades in K-3. Missing data information for STAR is reported in table 2.26.

[7] Further details of findings with unweighted estimation with complete-case are provided in Word et al. [1990] and Krueger [1999].

the application with Project STAR, both Hausman-type tests and regression-based variable addition tests reject the MCAR and IPW and MI-IPW methods under MAR are proposed to overcome the endogeneity problem by violation of MCAR. Meanwhile, we cannot directly test with the null of weak exogeneity for a pooled LS estimator. This is because, for $s_{it} = 0$ (see (2.2) for definition), we do not have observations for either outcome or for some covariates so we cannot estimate $\mathbf{E}(u_{it}|s_{it}, \mathbf{x}_{it})$ directly. (see (2.1) for the definitions of variables).

Second, we use LS, reduced-form(IV)[8] and first-differencing estimators with IPW and MI-IPW methods in a linear unbalanced panel data model to estimate the effect of CSR on SAT score for grades K-3 using Tennessee's project STAR data. Specifically, we apply the IPW method using the logit (or probit) model for which all observed past covariates and outcome are adopted as predictors for probability of selection which is based on Becker [1993]'s school choice model. In the application, the estimated probability weights are strictly greater than 0.2 so that we can avoid a criticism of instability for the IPW method from the probability weights being too close to zero. Multiple imputations of chained equation estimates are also obtained and compared to IPW estimates. The estimates for the effect of CSR on SAT scores are about 4 to 6.5 percent for IPW and MI-IPW while unweighted estimates for the return of CSR are 5 to 7.5 percent. Unweighted estimators of complete-case overestimate the effect by about 1 to 2 percentage points. Direction of bias analysis for pooled estimators also suggests that unweighted estimators with complete-case reveal upward bias under an assumption of positive correlation between $\mathbf{E}(s_{it}|\mathbf{x}_{it})$ and $\mathbf{E}(s_{it+1}|\mathbf{x}_{it})$.

The rest of chapter is organized as follows. In section 2, we provide pooled LS, FE and FD

[8] Reduced-form estimation (IV hearafter) implies LS estimation where initial assignment of treatments is used as covariates instead of actual treatments as in Krueger [1999].

estimators in an unbalanced linear panel data model. Section 3 introduces robust Hausman-type tests and variable addition tests to test the MCAR assumption and to show the power of the tests. In section 4, we describe IPW and MI-IPW methods under MAR to deal with missing data and to perform robust analysis for the IPW method to the misspecification of selection model using Monte Carlo(MC) simulation. Section 5 illustrates the tests of the MCAR and the estimations of IPW and MI-IPW methods under MAR using project STAR in the analysis of the effect of CSR on SAT score for students in grades K-3. Section 6 provides a method with which to calculate the direction of bias when the MAR assumption is violated. Section 7 contains concluding remarks.

2.2 Linear panel data model with missing and noncompliance

We are mainly interested in the estimation of marginal treatment effect (MTE hearafter) of a program in linear unbalanced panel data model. In particular, we focus on three popular estimators, in linear panel data models, which include pooled LS, FE and FD estimators.

2.2.1 Model

We introduce selection indictor (2.2) to express explicitly the conditions of consistency and MCAR for LS, FE and FD estimators.

$$y_{it} = \mathbf{d}_{it} \cdot \beta + \mathbf{x}_{1it}\delta + \mathbf{x}_{2i}\gamma + f_t + c_i + u_{it} \text{ for } , i = 1, 2, .., N \text{ and } t = 1, 2, .., T \qquad (2.1)$$

$$s_{it} = \left\{ \begin{array}{ll} 1 & \text{if all of } y_{it}, \mathbf{x}_{it} \text{ are observed} \\ 0 & otherwise \end{array} \right\} \tag{2.2}$$

where y_{it} is an outcome, $\mathbf{d}_{it}$ is a vector of indicators for treatments, $\mathbf{x}_{1it}$ is a vector of time varying covariates, $\mathbf{x}_{2i}$ is a vector of time constant covariates, f_t is time effect, c_i is unobserved heterogeneity and u_{it} is idiosyncratic error. Let $\mathbf{x}_{it} = (\mathbf{d}_{it}, \mathbf{x}_{1it}, \mathbf{x}_{2i}, f_t)$, $\theta = (\beta', \delta', \gamma', 1)'$ and $v_{it} = c_i + u_{it}$. Then we can rewrite (2.1) as (2.3).

$$y_{it} = \mathbf{x}_{it} \cdot \theta + v_{it} \text{ for }, i = 1, 2, .., n \text{ and } t = 1, 2, .., T \tag{2.3}$$

We can express pooled LS, FE and FD estimators with selection indicator as (2.4), (2.5) and (2.6) respectively and explicitly state conditions for MCAR.

$$\widehat{\theta}_{pls} = (\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}\mathbf{x}'_{it}\mathbf{x}_{it})^{-1} \sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}\mathbf{x}'_{it}y_{it} \tag{2.4}$$

$$\widehat{\theta}_{FE} = (\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}\ddot{\mathbf{x}}'_{it}\ddot{\mathbf{x}}_{it})^{-1} \sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}\ddot{\mathbf{x}}'_{it}\ddot{y}_{it} \tag{2.5}$$

where $\ddot{\mathbf{x}}_{it} = \mathbf{x}_{it} - \overline{\mathbf{x}_i}$, $\overline{\mathbf{x}_i} = \frac{1}{T_i}\sum_{t=1}^{T} s_{it}\mathbf{x}_{it}$ and $T_i = \sum_{t=1}^{T} s_{it}$.

$$\widehat{\theta}_{FD} = (\sum_{i=1}^{N}\sum_{t=2}^{T} s^{FD}_{it}\Delta\mathbf{x}'_{it}\Delta\mathbf{x}_{it})^{-1} \sum_{i=1}^{N}\sum_{t=2}^{T} s^{FD}_{it}\Delta\mathbf{x}'_{it}\Delta y_{it} \tag{2.6}$$

where $\Delta\mathbf{x}_{it} = \mathbf{x}_{it} - \mathbf{x}_{it-1}$ for $t = 2, 3, .., T$ and $s^{FD}_{it} = \min\{s_{it}, s_{it-1}\}$.

Conditions for consistency with missing data

Pooled Least Squares For a pooled least squares(PLS) estimator, we essentially need rank condition (2.7) and exogeneity condition (2.8) to obtain a consistent estimator.

$$rank\mathbf{E}(s_{it}\mathbf{x}'_{it}\mathbf{x}_{it}) = dim(\theta) \text{ , } t = 1, 2, ..., T \tag{2.7}$$

$$\mathbf{E}(s_{it}\mathbf{x}'_{it}v_{it}) = 0 \text{ , } t = 1, 2, ..., T \tag{2.8}$$

Fixed effects Rank condition and exogeneity condition for fixed-effects estimator are (2.9) and (2.10), respectively.

$$rank\mathbf{E}(s_{it}\ddot{\mathbf{x}}_{it}'\ddot{\mathbf{x}}_{it}) = dim(\theta) \ , \ t = 1, 2, ..., T \tag{2.9}$$

$$\mathbf{E}(s_{it}\ddot{\mathbf{x}}_{it}'\ddot{u}_{it}) = 0 \ , \ t = 1, 2, ..., T \tag{2.10}$$

First differencing Rank condition and exogeneity condition for first differencing estimator are (2.11) and (2.12), respectively.

$$rank\mathbf{E}(s_{it}^{FD}\Delta\mathbf{x}_{it}'\Delta\mathbf{x}_{it}) = dim(\theta) \ , \ t = 2, ..., T \tag{2.11}$$

$$\mathbf{E}(s_{it}^{FD}\Delta\mathbf{x}_{it}'\Delta u_{it}) = 0 \ , \ t = 2, ..., T \tag{2.12}$$

Exogeneity assumption Strict exogeneity of (2.13) is sufficient for the exogeneity conditions of FE and FD estimators in (2.10) and (2.12) while weak exogeneity of (2.15) is sufficient for pooled estimator.

$$\mathbf{E}(u_{it}|\mathbf{s}_i, \mathbf{x}_i) = 0, \text{ for } t = 1, 2, ..., T \tag{2.13}$$

$$\text{where } \mathbf{s}_i = (s_{i1}, s_{i2}, ..., s_{iT}) \text{ and } \mathbf{x}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{iT}) \tag{2.14}$$

$$\mathbf{E}(u_{it}|s_{it}, \mathbf{x}_{it}) = 0 \text{ and } \mathbf{E}(c_i|s_{it}, \mathbf{x}_{it}) = 0, \text{ for } t = 1, 2, ..., T \tag{2.15}$$

Pooled estimator does not need strict exogeneity while it requires no correlation between selection and unobserved heterogeneity.

2.2.2 Tests of missing completely at random

If missing completely at random (MCAR) is satisfied, all three estimators should be consistent with random experiment.[9] We suggest a Hausman-type test using null hypothesis of strict exogeneity and $\mathbf{E}(c_i|s_{it}, \mathbf{x}_{it}) = 0$ in section 2.2.1.

MCAR MCAR in random experiment implies both weak and strict exogeneity while violation of either weak or strict exogeneity leads to violation of MCAR. Moreover, with proper rank conditions, MCAR assumption is sufficient for all of pooled LS, FE and FD estimators to be consistent while violation of either weak or strict exogeneity leads unweighted pooled LS or FD and FE estimators to be inconsistent respectively.

Null I: strict exogeneity Consider a case in which the violation of strict exogeneity is the only source of bias. Under strict exogeneity, both FE and FD estimators are consistent and the difference of FE and FD estimates should be entirely due to sampling error. We extend a robust Hausman test which compares FE and FD estimators in balanced panel data as in [Wooldridge, 2010, p.324-325] to unbalanced panel data model by constructing a Wald statistic based on $\sqrt{n}(\widehat{\beta}_{FE} - \widehat{\beta}_{FD})$. The rejection of null hypothesis implies the violation of strict exogeneity ($\text{Cov}(u_{it}, \mathbf{s}_i\mathbf{d}_i) \neq 0$) and both FE and FD estimators are inconsistent. In short, MCAR is violated and wnweighted FE and FD estimators are inconsistent.

Singularity Since we can not obtain the estimates for FE and FD coefficient of time constant covariates, the set of parameters that we can use to construct a Wald statistic

[9] See Wooldridge [2009] for more discussion of equivalence of three estimators under the null.

should exclude the coefficients from time constant variables. Thus, in typical applications, the indicators for the treatments of a program should vary over time to apply our method of comparing two consistent estimators under the null.

Null II: no correlation between selection and unobserved heterogeneity We propose a subsequent test for the case when we can not reject the null of strict exogeneity. We assume strict exogeneity hold both under the null and the alternative. Therefore, under the null, all conditions for the consistency of pooled, FE and FD estimators are satisfied while, under the alternative, $\mathbf{E}(c_i|s_{it}, \mathbf{x}_{it}) = 0$ is not satisfied. Still both fixed effect and first differencing estimators provide consistent estimators both under the null and the alternative while pooled estimator is consistent only under the null.[10]Thus, we can test the null hypothesis of MCAR by using differential correlation between selection and unobserved heterogeneity across treatments and control as alternative. We construct a Wald statistic based on $\sqrt{n}(\widehat{\beta}_{pls} - \widehat{\beta}_{FE})$ to perform a test of MCAR. The rejection of the null implies the violation of $\mathbf{E}(c_i|s_{it}, \mathbf{x}_{it}) = 0$ (conditional on $\mathbf{x}_{it}$ $Cov(c_i, s_{it}\mathbf{x}_{it}) \neq 0$) and this leads pooled estimator to be inconsistent because of $Cov(c_i, s_{it}\mathbf{x}_{it}) \neq 0$. Under ideal conditions and $\mathbf{E}(c_i|s_{it}, \mathbf{x}_{it}) \neq 0$, MCAR is violated and unweighted pooled estimator is inconsistent while, with FE and FD transformed linear models, MCAR assumption holds and unweighted FE and FD estimators are consistent.

[10]Strict exogeneity assumption is maintained both under the null and under the alternative. We also assume that rank conditions are satisfied for all estimators we consider in this section.

2.2.3 Implications from the rejection of tests

The rejection of strict exogeneity implies that both unweighted FE and FD estimators with complete-case are inconsistent. The failure of rejection of strict exogeneity and the success of rejection for $Cov(c_i,s_{it}|\mathbf{x}_{it})= 0$ leads unweighted pooled estimator with complete-case to be inconsistent while we can not reject the null that unweighted FE and FD estimators with complete case are consistent.[11]

The rejection of both tests implies missing indicator is probably correlated with both time-varying and time-constant unobserved variables even conditional on observed variables(c_i and $\mathbf{u}_i$ are correlated with $s_{it}x_{it}$) so that MCAR is violated and unweighted FE and FD estimators with complete case are invalid. Therefore, we need to control for missing data to produce a valid statistical analysis. Either FD or FE transformations are not enough to obtain a valid inference since the correlation between $\mathbf{u}_i$ and $\mathbf{s}_i\mathbf{x}_i$ is not zero.

2.3 Formal tests of MCAR

In this section, we introduce a test statistic used in performing robust Hausman tests for MCAR assumption and investigate the power of the test via Monte Carlo experiment.

[11] Under the null of these two Hausman-type tests, all other conditions for consistency including rank condition and other regularity conditions are assumed to be satisfied.

2.3.1 Hausman-type test I - Comparison of $\widehat{\theta}_{pls}$ and $\widehat{\theta}_{FE}$ using $\mathbf{E}(c_i, s_{it}\mathbf{x}_{it}) = \mathbf{0}$ as null

A test statistic is constructed to perform a Hausman-type test for the null of no correlation between unobserved heterogeneity and missing data. First, we stack two estimators $\widehat{\theta}_{pls}$ and $\widehat{\theta}_{FE}$. We denote stacked estimator as $\widehat{\theta}_a = (\widehat{\theta}'_{pols}, \widehat{\theta}'_{FE})'$ and denote restriction matrix as R of $k \times 2k$.

$$\sqrt{n}\widehat{\theta}_a = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\mathbf{x}'_{it}\mathbf{x}_{it} & 0 \\ 0 & \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\ddot{\mathbf{x}}'_{it}\ddot{\mathbf{x}}_{it} \end{bmatrix}^{-1}_{2k\times 2k} \begin{bmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\mathbf{x}'_{it}y_{it} \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\ddot{\mathbf{x}}'_{it}\ddot{y}_{it} \end{bmatrix}_{2k\times 1}$$

and $R = [I_k | -I_k]$.

We write the null hypothesis as (2.16) and this implies that under the null both pooled LS and FE estimators are consistent.

$$H_0 : R\theta_a = 0$$

$$R\theta_a = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 & -1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 & -1 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots & \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots & \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 & 0 & 0 & \cdots & 0 & -1 \end{bmatrix}_{k\times 2k} \begin{bmatrix} \theta_{1,ls} \\ \vdots \\ \theta_{k,ls} \\ \theta_{1,FE} \\ \vdots \\ \theta_{k,FE} \end{bmatrix}_{2k\times 1} \tag{2.16}$$

$$= \begin{bmatrix} \theta_{1,ls} - \theta_{1,FE} \\ \theta_{2,ls} - \theta_{2,FE} \\ \vdots \\ \vdots \\ \vdots \\ \theta_{k,ls} - \theta_{k,FE} \end{bmatrix}_{k \times 1} = \mathbf{0}$$

Using the null $H_0 : R\theta_a = 0$ [12]and consistency conditions in (2.2.1), we construct a Wald statistic which converges asymptotically to a chi-squared distribution with degree of freedom equal to number of restrictions, k in (2.16).

Assumptions

1. Conditional strict exogeneity : $E(u_{it}|\mathbf{x}_i, \mathbf{s}_i) = 0$, $\forall t$

2. no correlation between unobserved heterogeneity and missing data: $E(c_i|\mathbf{x}_{it}, s_{it}) = 0$.

3. Rank conditions: $E(\sum_{t=1}^{T} s_{it}^{FE} \ddot{\mathbf{x}}'_{it} \ddot{\mathbf{x}}_{it})$ and $E(\sum_{t=1}^{T} s_{it} \mathbf{x}'_{it} \mathbf{x}_{it})$ have full rank

$$\begin{aligned} p\lim \frac{1}{n} \sum_{i=1}^{n} \sum_{t=1}^{T} s_{it} \mathbf{x}'_{it} \mathbf{x}_{it} &= E(\sum_{t=1}^{T} s_{it} \mathbf{x}'_{it} \mathbf{x}_{it}) \\ p\lim \frac{1}{n} \sum_{i=1}^{n} \sum_{t=1}^{T} s_{it} \ddot{\mathbf{x}}'_{it} \ddot{\mathbf{x}}_{it} &= E(\sum_{t=1}^{T} s_{it} \ddot{\mathbf{x}}'_{it} \ddot{\mathbf{x}}_{it}) \end{aligned}$$

4. Conditions for applying CLT to following two objects, $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{t=1}^{T} s_{it} \mathbf{x}'_{it} u_{it}$ and $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{t=1}^{T} s_{it} \ddot{\mathbf{x}}'_{it} \ddot{u}_{it}$.

[12] This is equivalent to $H_0 : \theta_{j,PLS} = \theta_{j,FE}$ $\forall j = 1, 2, ...k$

A Wald Statistic for a robust Hausman test

Proposition 2 *Under assumption 1,2,3,and 4, for the test with null hypothesis of* H_0 : $R\theta_a = 0$, *we can define a Wald type statistic* W_a *which converges to* χ^2_k *distribution.*

$$W_a = [R\sqrt{n}(\widehat{\theta}_a)]'[R\widehat{var}(\sqrt{n}\widehat{\theta}_a)R']^{-1}R\sqrt{n}(\widehat{\theta}_a) \sim \chi^2_k$$

where

$$R\widehat{\theta}_a = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 & -1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 & -1 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots & \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots & \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 & 0 & 0 & \cdots & 0 & -1 \end{bmatrix}_{k\times 2k} \begin{bmatrix} \widehat{\theta}_{1,PLS} \\ \vdots \\ \widehat{\theta}_{k,PLS} \\ \widehat{\theta}_{1,FE} \\ \vdots \\ \widehat{\theta}_{k,FE} \end{bmatrix}_{2k\times 1} =$$

$$\begin{bmatrix} \widehat{\theta}_{1,PLS} - \widehat{\theta}_{1,FE} \\ \widehat{\theta}_{2,PLS} - \widehat{\theta}_{2,FE} \\ \vdots \\ \vdots \\ \vdots \\ \widehat{\theta}_{k,PLS} - \widehat{\theta}_{k,FE} \end{bmatrix}_{k\times 1}$$

Proof. See appendix. ■

2.3.2 Monte Carlo experiment I

We investigate the power of the test using Monte Carlo simulation. We generate a selection process which reduces full sample to either 75 or 50 percent of full sample. Generated

selection process induces a differential correlation between unobserved heterogeneity and selection across treatment and control. Thus, this particular selection process preserves asymptotic consistent results for FE and FD estimators. We focus on how the power of test changes as the fraction of complete-case data to full data changes to 75 and 50 percents.

Data generating process

Data generation process is based on following outcome and selection process.

The model for DGP

$$y_{it} = d_{it} \cdot \alpha + x_{it} \cdot \beta + c_i + u_{it}, \text{ for } i = 1, 2, .., n \text{ and } t = 1, 2, 3, 4$$

where d_{it} a binary variable and x_{it} is continuous variable.

$$s_{i1} = 1; \; s_{it} = 1(d_{it} \cdot \gamma + \frac{w_i}{\sqrt{2}} + \frac{e_{it}}{2} > a), \text{ where } \gamma = 1 \text{ for } i = 1, 2, .., n \text{ and } t = 2, 3, 4$$

Variables generation

1. $d_{it}^{\star} = \Phi(\frac{h_{it}+v_i}{\sqrt{2}})$ where $\Phi(\cdot)$ is standard normal CDF. $d_{it} = 1$ if $d_{it}^{\star} > 0.7$. $\alpha = 1$, $h_{it} \sim iidN(0, 1)$ and $v_i \sim iidN(0, 1)$.

2. $x_{it} \sim iidN(0, 1)$, $\beta = 0.2$

3. $u_{it} \sim iidN(0, 1)$

4. $c_i \sim \frac{\sqrt{3}w_i}{2} + \frac{a_i}{2}$, $w_i \sim iidN(0, 1)$ and $a_i \sim iidN(0, 1)$

5. $s_{it}^{\star} = \Phi(d_{it}\gamma + \frac{w_i}{\sqrt{2}} + \frac{e_{it}}{2})$ where γ=1 . $s_{it} = 1$ if $s_{it}^{\star} > a$, where a determines the proportion of missing data.

Table 2.1. Hausman-type test I: Null and Alternative ($\text{cov}(s_{it}d_{it},\, c_i) \neq 0$)

	Null	Alternative
Pooled LS	Consistent	Inconsistent
FE estimator	Consistent	Consistent

Table 2.2. Hausman-type test I: Null and Alternative ($\text{cov}(s_{it}d_{it},\, c_i) \neq 0$)

	Null	Alternative
RE estimator	Consistent and Efficient	Inconsistent
FE estimator	Consistent	Consistent

With this data generating process, the bias of pooled LS estimator for α is originated from w_i which causes a differential correlation between s_{it} and c_i across d_{it} while, for FE estimator, FE transformation eliminate c_i and FE estimator for α is consistent. This particular selection process leads to general missing.

The size and power of the Hausman-type I test is based on the estimators with characteristics in 2.1 under the null and alternative.

As DGP satisfy the conditions for Random effects (RE) estimator to be ideal under the null we can also carry out Hausman type-I test by stacking RE and FE estimators. For this case, the size and power of the Hausman-type I test is based on the estimators with characteristics in 2.2 under the null and alternative.

Simulation results

MC simulations are performed with 1,000 replications with T=4 and various n from 100 to 5,000 and cluster robust standard error is used in all simulations. We choose the size of T=4 to mimic the actual sample size of Tennessee's Project STAR in the application.

In DGP, a determines the missing proportion of data. Therefore, a=0 implies no missing

data so it is a balanced panel while a=0.5 implies 50 percent of data is missing and only 25 percent of data is complete case. Third column and fourth column in tables 2.3 and 2.4 represent the means of pooled LS and FE estimates for $\widehat{\alpha}$ respectively. P-value column shows that the mean of p-value for Wald statistic, W_a for Hausman-type test I. $1(P < .05)$ column shows the mean of the rejection of null at the nominal level 0.05. Finally, 95-CI column represents, the coverage for $1(P < .05)$.

Both in table 2.3 and table 2.4, the power of test increases with cross-sectional size, n, for all $a > 0$. For example, for the data of same size as Project STAR data for which n= 6,800 in each grade and a=0.5(missing fraction of data), the power of test is 1 in both 2.3 and 2.4. Overall, if n is greater than 5,000, the power of test is 1 regardless of the value for a(fraction of non-complete data to full data) in the simulation. In short, the power of robust Hausman-type test using the difference between pooled LS and FE is quite reliable to test MCAR for $n > 5,000$. As we see in tables 2.3 and 2.4, under alternative DGPs, both bias and standard error is smaller for RE estimator than those for pooled estimator.

In both 2.3 and 2.4, the panel with a=0 shows the results for balanced panel data and it reports no size distortion for all n and the coverage rate of rejection frequency contains true nominal level, 0.05, for all n.

Table 2.3. Pooled LS and FE estimates for α and p-value for W_a

$a=0$	True α	Pooled	FE	p-value	$r \equiv 1$(P<.05)	95-CI
$n=100$	1	1.003	1.002	0.502	0.054	(0.040,0.068)
$n=500$	1	1.003	1.002	0.510	0.062	(0.047,0.076)
$n=1,000$	1	0.996	0.998	0.496	0.047	(0.034,0.060)
$n=2,000$	1	0.999	1.001	0.490	0.047	(0.034,0.060)
$n=5,000$	1	0.998	1.000	0.510	0.048	(0.035,0.061)
$a=0.25$	True α	Pooled	FE	p-value	$r \equiv 1$(P<.05)	95-CI
$n=100$	1	0.810	1.000	0.374	0.155	(0.132,0.177)
$n=500$	1	0.812	1.000	0.132	0.533	(0.502,0.564)
$n=1,000$	1	0.804	0.997	0.033	0.846	(0.823,0.868)
$n=2,000$	1	0.810	1.001	0.002	0.990	(0.984,0.996)
$n=5,000$	1	0.807	0.999	0.000	1	(1,1)
$a=0.5$	True α	Pooled	FE	p-value	$r \equiv 1$(P<.05)	95-CI
$n=100$	1	0.758	1.002	0.348	0.182	(0.158,0.206)
$n=500$	1	0.754	1.000	0.089	0.664	(0.635,0.693)
$n=1,000$	1	0.747	0.999	0.014	0.936	(0.920,0.951)
$n=2,000$	1	0.751	0.999	0.001	0.995	(0.991,0.999)
$n=5,000$	1	0.750	1.000	0.000	1	(1,1)

*Note: 1(P<.05) is the rejection rate for the null of H_0:α=1 with nominal value 0.05.; a is the ratio of missing sample to full sample.; Cluster robust standard error has been used in all estimations.; p-value is the mean of p-value for W_a, r is rejection rate, and CI-95 is coverage for r.

Table 2.4. RE and FE estimates for α and p-value for W_a

$a=0$	True α	Pooled	FE	p-value	$r \equiv 1$(P<.05)	95-CI
$n=100$	1	.999	.999	0.512	0.051	(0.037,0.065)
$n=500$	1	1.000	1.000	0.495	0.055	(0.040,0.069)
$n=1,000$	1	1.000	1.001	0.497	0.050	(0.036,0.064)
$n=2,000$	1	1.001	1.002	0.507	0.054	(0.040,0.068)
$n=5,000$	1	1.000	1.000	0.496	0.045	(0.032,0.058)
$a=0.25$	True α	Pooled	FE	p-value	$r \equiv 1$(P<.05)	95-CI
$n=100$	1	0.960	1.000	0.397	0.141	(0.119,0.163)
$n=500$	1	0.962	1.001	0.166	0.489	(0.458,0.520)
$n=1,000$	1	0.958	0.998	0.055	0.768	(0.742,0.794)
$n=2,000$	1	0.962	1.001	0.007	0.965	(0.954,0.976)
$n=5,000$	1	0.960	1.000	0.000	1	(1,1)
$a=0.5$	True α	Pooled	FE	p-value	$r \equiv 1$(P<.05)	95-CI
$n=100$	1	0.973	1.005	0.433	0.119	(0.099,0.139)
$n=500$	1	0.968	1.000	0.225	0.325	(0.296,0.354)
$n=1,000$	1	0.967	1.001	0.105	0.616	(0.586,0.646)
$n=2,000$	1	0.965	0.999	0.026	0.879	(0.859,0.899)
$n=5,000$	1	0.967	1.001	0.000	1	(1,1)

*Note: 1(P<.05) is the rejection rate for the null of H_0:α=1 with nominal value 0.05.; a is the ratio of missing sample to full sample.; Cluster robust standard error has been used in all estimations.; p-value is the mean of p-value for W_a, r is rejection rate, and CI-95 is coverage for r.

2.3.3 Hausman-type test II - Comparison of $\widehat{\theta}_{FE}$ and $\widehat{\theta}_{FD}$, Null: strict exogeneity

A test statistic for a Hausman-type test of strict exogeneity is derived. We stack two unbiased estimators $\widehat{\theta}_{FE}, \widehat{\theta}_{FD}$ under strict exogeneity and denote stacked estimator as $\widehat{\theta}_b = (\widehat{\theta}'_{FE}, \widehat{\theta}'_{FD})'$ and restriction matrix as R of $k \times 2k$.

$$\sqrt{n}\widehat{\theta}_b = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\ddot{\mathbf{x}}'_{it}\ddot{\mathbf{x}}_{it} & 0 \\ 0 & \frac{1}{n}\sum_{i=1}^{n}\sum_{t=2}^{T} s_{it}^{FD}\Delta\mathbf{x}'_{it}\Delta\mathbf{x}_{it} \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\ddot{\mathbf{x}}'_{it}\ddot{y}_{it} \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{t=2}^{T} s_{it}^{FD}\Delta\mathbf{x}'_{it}\Delta y_{it} \end{bmatrix}$$

Using the null $H_0 : R\theta_b = 0$ and the assumptions for consistency in (2.2.1), we can construct a Wald statistic which converges asymptotically to a chi-squared distribution with following assumptions.

Assumptions

1. Conditional strict exogeneity : $E(u_{it}|\mathbf{x}_i, \mathbf{s}_i) = 0$, $\forall t$

2. Rank conditions: $E(\sum_{t=1}^{T} s_{it}\ddot{\mathbf{x}}'_{it}\ddot{\mathbf{x}}_{it})$ and $E(\sum_{t=2}^{T} s_{it}^{FD}\Delta\mathbf{x}'_{it}\Delta\mathbf{x}_{it})$ have full rank

$$p\lim \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\ddot{\mathbf{x}}'_{it}\ddot{\mathbf{x}}_{it} \;=\; E(\sum_{t=1}^{T} s_{it}\ddot{\mathbf{x}}'_{it}\ddot{\mathbf{x}}_{it})$$

$$p\lim \frac{1}{n}\sum_{i=1}^{n}\sum_{t=2}^{T} s_{it}^{FD}\Delta\mathbf{x}'_{it}\Delta\mathbf{x}_{it} = E(\sum_{t=2}^{T} s_{it}^{FD}\Delta\mathbf{x}'_{it}\Delta\mathbf{x}_{it})$$

3. Conditions for applying CLT to following two objects, $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\ddot{\mathbf{x}}'_{it}\ddot{u}_{it}$ and $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{t=2}^{T} s_{it}^{FD}\Delta\mathbf{x}'_{it}\Delta u_{it}$.

Proposition 3 *Under assumption 1,2,and 3, for the test with null hypothesis of* $H_0 : R\theta_b = 0$*, we can derive a Wald type statistic* W_b *that converges to* χ^2_k *distribution.*

$$W_b = [R\sqrt{n}(\widehat{\theta}_b)]'[R\widehat{var}(\sqrt{n}\widehat{\theta}_b)R']^{-1}R\sqrt{n}(\widehat{\theta}_b) \sim \chi^2_k$$

Proof. See appendix. It can be obtained in similar manner as Hausman type I test. ■

Singularity A test using FE and FD estimators can only include the coefficients for time varying covariates.[13]

2.3.4 Monte Carlo experiment: Test of MCAR using strict exogeneity assumption

We investigate the power of test for DGPs with selection processes which reduce full sample to either 75, 50 or 25 percent of full sample and induce a differential correlation between selection and idiosyncratic errors across d_{it}(treatment and control). Thus, these selection processes violate strict exogeneity assumption. We focus on how the power of test changes as we increase the missing proportion of data from 25 percent to 75 percent.

Data generating process

Data generation is based on following outcome and selection process.

The model for DGP

$$y_{it} = d_{it} \cdot \alpha + x_{it} \cdot \beta + f_t + c_i + u_{it} \tag{2.17}$$

[13] See Wooldridge [2010] for more discussion.

where $i = 1, 2, .., n$ and $t = 1, 2, 3, 4$

$$s_{i1} = 1;\ s_{it} = 1 \text{ if } \Phi(\frac{v_{it} + w_i + e_{it}}{\sqrt{3}} > cutoff) \text{ and } s_{it} = 0 \text{ otherwise} \tag{2.18}$$

Variables generation

1. $d^{\star}_{it} = \Phi(\frac{h_{it}+v_i}{\sqrt{2}})$ where $\Phi(\cdot)$ is standard normal CDF. $d_{it} = 1$ if $d^{\star}_{it} > 0.5$. and $d_{it} = 0$ otherwise. $\alpha = 1$, $h_{it} \sim iidN(0, 1)$ and $v_i \sim iidN(0, 1)$.

2. $$\begin{bmatrix} u_{it} \\ e_{it} \end{bmatrix} \sim BN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$$
   where BN is bivariate normal and ρ=0.7 in the simulation.

3. $x_{it} \sim iidN(0, 1)$, $\beta = -0.2$

4. $f_t \sim uniform(-1, 1)$

5. $c_i \sim \frac{\sqrt{3}w_i}{2} + \frac{v_i}{2}$, $w_i \sim iidN(0, 1)$ and $v_i \sim iidN(0, 1)$

6. $s^{\star}_{it} = \Phi(\frac{h_{it}+wi+e_{it}}{\sqrt{3}})$. $s_{it} = 1$ if $s^{\star}_{it} > a$ and $s_{it} = 0$ otherwise ,

   where a determines proportion of missing data since $s^{\star}_{it}$ is normal CDF.

Both FE and FD estimators for α are not valid by a differential correlation between s_{it} and u_{it} across d_{it} since $s_{it}d_{it}$ is correlated with u_{it} and the source of differential correlation h_{it} is time-varying. The differential correlation remains even after FE and FD transformations so that both FE and FD estimators are not reliable for unbalanced panel data with a >0.

Simulation results

Table 2.5 reports the size and power of Hausman-type II tests. Simulation performed with 1,000 replications with data set size of n ranges from 100 to 5,000 and T=4.

First, from the a=0 panel, we see in column $r \equiv 1(\text{P}<.05)$ that there is no size distortion and the coverage rate of rejection frequency contains true nominal level 0.05 for all n from 100 to 5,000. Second, the power of test increases with the proportion of missing data, a, and cross-sectional sample size, n. Third, for Project STAR data of a=0.5 and n=6,000, the power of test is about 0.6. Therefore, it is quite possible that we may not be able to reject strict exogeneity assumption even if selection and time-varying unobserved variables are correlated since finite sample power is strictly less than 1.

2.3.5 Variable addition test of strict exogeneity

Wooldridge [2009] proposed a simple regression based test of strict exogeneity for FE and FD estimators in an unbalanced panel data. This test simply adds selection indicator of s_{it+1} as an additional regressor in the linear regression model for FD and FE estimations. Then the t-statistics for the coefficient, λ on s_{it+1} in (2.19) and (2.20) can be used in t-tests of null hypothesis of $H_0 : \lambda = 0$. This is because the strict exogeneity implies any function of $\mathbf{s}_i = \{s_{i1}, s_{i2}, ...s_{iT}\}$ should not be correlated with $\mathbf{u}_i = (u_{i1}, u_{i2}, ...u_{iT})$ conditional on $\mathbf{x}_i$. The rejection of test implies the violation of strict exogeneity and violation of strict exogeneity is sufficient for violation of MCAR assumption so that both unweighted FE and FD estimators are invalid.

Table 2.5. FD and FE estimates for α and p-value for W_b

$a = 0$	True α	FD	FE	p-value	$r \equiv 1(\text{P}<.05)$	95-CI
$n = 100$	1	0.994	1.002	0.492	0.061	(0.046,0.076)
$n = 500$	1	1.001	1.002	0.504	0.045	(0.032,0.058)
$n = 1,000$	1	0.998	0.998	0.513	0.047	(0.034,0.060)
$n = 2,000$	1	0.999	0.999	0.499	0.044	(0.031,0.057)
$n = 5,000$	1	1.001	1.000	0.507	0.054	(0.040,0.068)
$a = 0.25$	True α	FD	FE	p-value	$r \equiv 1(\text{P}<.05)$	95-CI
$n = 100$	1	0.831	0.849	0.487	0.078	(0.061,0.095)
$n = 500$	1	0.834	0.854	0.469	0.061	(0.046,0.076)
$n = 1,000$	1	0.829	0.849	0.456	0.082	(0.065,0.099)
$n = 2,000$	1	0.831	0.850	0.418	0.121	(0.101,0.141)
$n = 5,000$	1	0.830	0.850	0.311	0.263	(0.236,0.290)
$a = 0.5$	True α	FD	FE	p-value	$r \equiv 1(\text{P}<.05)$	95-CI
$n = 100$	1	0.749	0.802	0.457	0.087	(0.070,0.104)
$n = 500$	1	0.743	0.803	0.388	0.167	(0.144,0.190)
$n = 1,000$	1	0.752	0.802	0.361	0.205	(0.180,0.230)
$n = 2,000$	1	0.753	0.808	0.259	0.352	(0.322,0.382)
$n = 5,000$	1	0.751	0.807	0.162	0.599	(0.569,0.629)
$a = 0.75$	True α	FD	FE	p-value	$r \equiv 1(\text{P}<.05)$	95-CI
$n = 100$	1	0.725	0.879	0.409	0.126	(0.105,0.147)
$n = 500$	1	0.717	0.869	0.296	0.320	(0.291,0.349)
$n = 1,000$	1	0.708	0.869	0.218	0.495	(0.464,0.526)
$n = 2,000$	1	0.711	0.869	0.159	0.611	(0.581,0.641)
$n = 5,000$	1	0.712	0.866	0.106	0.752	(0.725,0.779)

*Note: $1(P < .05)$ column shows the mean of the rejection of null at the nominal level 0.05.; a is the ratio of missing sample to full sample.; Cluster robust standard error has been used in all estimations.; p-value is the mean of p-value for W_b. FD column and FE column represents the mean of FD and FE estimates for $\widehat{\alpha}$ respectively. Finally, 95-CI column represents, the coverage for $1(P < .05)$.

$$\ddot{y}_{it} = \lambda \cdot \ddot{s}_{it+1} + \ddot{\mathbf{x}}'_{it}\theta + f_t + \ddot{u}_{it} \text{ , for } s_{it} = 1 \tag{2.19}$$

$$\Delta y_{it} = \lambda \cdot \Delta s_{it+1} + \Delta \mathbf{x}'_{it}\theta + \Delta f_t + \Delta u_{it} \text{ , for } s^{FD}_{it} = 1 \tag{2.20}$$

Although we can not directly test weak exogeneity by variable addition test for pooled estimator in (2.21) since we can not observed all $\mathbf{x}_{it}, y_{it}$ for individual i when s_{it} =0, we can obtain indirect evidence from variable addition test in the pooled LS based on the assumption of non-zero serial correlation of selection indicator. If we further assume that $\mathbf{E}(s_{it}|\mathbf{x}_{it})$ and $\mathbf{E}(s_{it+1}|\mathbf{x}_{it})$ is correlated, we can t-statistic on λ in (2.21) to test weak exogeneity and MCAR.

$$y_{it} = \lambda \cdot s_{it+1} + \mathbf{x}'_{it}\theta + f_t + c_i + u_{it} \text{ , for } s_{it} = 1 \tag{2.21}$$

Data generating process

We use the same DGP processes as in section 2.3.4 where we study the power of Hausman-type test for strict exogeneity. We only change the model of (2.17) from 2.3.4 to (2.19) and (2.20) for FE and FD estimators. In these DGPs, the selection and unobserved heterogeneity or time-varying idiosyncratic errors are differentially correlated across d_{it} so that both FD and FE transformations cannot eliminate the correlation and both FD and FE estimators are not valid.

The power of variable addition test for strict exogeneity assumption

Table 2.6 reports the estimates for $\widehat{\lambda}$ in (2.19) and (2.20) for FE and FD estimators. In all simulations, the number of replication is 1,000 with the data set of T=4 and n from 100

to 5,000 and the cluster standard error are used in all estimations. Selection indicators are generated to produce the fraction of missing data to full data to be ranged from a=0.1 to a=0.75.

For pooled LS estimator, the power of the test for the null, $H_0 : \lambda = 0$ in (2.21) is 1 for all a >0 and n ≥500 but this is not a direct test of MCAR for the pooled LS estimator. However, if we assume further non-zero correlation between $\mathbf{E}(s_{it}|\mathbf{x}_{it})$ and $\mathbf{E}(s_{it+1}|\mathbf{x}_{it})$, this test become valid test of weak exogeneity and MCAR.

For both FE and FD estimators, the power of test increases with n and a. For instance, for the case of a=0.5, the power of test increases from 0.07 to 0.26 for FE estimation while the power of test increases from 0.07 to 0.67 for FD estimation. Still, the power of test is significantly lower than 1 for the data with the actual size of the Project STAR. The power of variable addition tests is greater for FD estimator than FE estimator with the same n and a.

Compared to a robust Hausman-type test at the last column in table (2.6)which uses strict exogeneity, the power of variable addition test is greater for when $a \leq 0.5$ while the power of variable addition test is lower when a=0.75.

Table 2.6. Test of strict exogeneity variable addition test and Hausman test for W_b

	LS			FD			FE			Hausman
$a = 0.1$	t-stat	1(P<.05)	95-CI	t-stat	1(P<.05)	95-CI	t-stat	1(P<.05)	95-CI	1(P<.05)
$n = 100$	3.40	0.92	(0.90,0.94)	-0.13	0.08	(0.06,0.09)	-0.07	0.07	(0.05, 0.08)	
$n = 500$	7.30	1	(1,1)	-0.33	0.05	(0.04,0.06)	-0.18	0.05	(0.04,0.06)	
$n = 1,000$	10.35	1	(1,1)	-0.48	0.07	(0.05,0.09)	-0.23	0.06	(0.04,0.07)	
$n = 2,000$	14.54	1	(1,1)	-0.63	0.10	(0.08,0.11)	-0.39	0.07	(0.06, 0.09)	
$n = 5,000$	23.04	1	(1,1)	-1.06	0.18	(0.16,0.21)	-0.59	0.07	(0.06, 0.09)	
$a = 0.25$	t-stat	1(P<.05)	95-CI	t-stat	1(P<.05)	95-CI	t-stat	1(P<.05)	95-CI	1(P<.05)
$n = 100$	3.66	0.96	(0.95,0.97)	-0.23	0.07	(0.05,0.08)	-0.12	0.06	(0.04,0.07)	0.08
$n = 500$	8.14	1	(1,1)	-0.66	0.10	(0.08,0.12)	-0.44	0.07	(0.05,0.09)	0.06
$n = 1,000$	11.58	1	(1,1)	-0.84	0.15	(0.13,0.17)	-0.47	0.08	(0.07,0.10)	0.08
$n = 2,000$	16.35	1	(1,1)	-1.18	0.22	(0.20,0.25)	-0.67	0.10	(0.08,0.11)	0.12
$n = 5,000$	25.91	1	(1,1)	-1.85	0.45	(0.42,0.48)	-1.08	0.19	(0.16,0.21)	0.26
$a = 0.5$	t-stat	1(P<.05)	95-CI	t-stat	1(P<.05)	95-CI	t-stat	1(P<.05)	95-CI	1(P<.05)
$n = 100$	3.52	0.93	(0.92,0.95)	-0.32	0.07	(0.06,0.09)	-0.20	0.07	(0.06,0.09)	0.09
$n = 500$	7.94	1	(1,1)	-0.79	0.12	(0.10,0.14)	-0.46	0.07	(0.06,0.09)	0.17
$n = 1,000$	11.21	1	(1,1)	-1.05	0.19	(0.16,0.21)	-0.60	0.09	(0.07,0.11)	0.21
$n = 2,000$	15.97	1	(1,1)	-1.50	0.31	(0.28,0.34)	-0.86	0.15	(0.13,0.18)	0.35
$n = 5,000$	25.19	1	(1,1)	-2.37	0.67	(0.64,0.70)	-1.31	0.26	(0.23,0.29)	0.60
$a = 0.75$	t-stat	1(P<.05)	95-CI	t-stat	1(P<.05)	95-CI	t-stat	1(P<.05)	95-CI	1(P<.05)
$n = 100$	3.04	0.85	(0.83,0.87)	-0.31	0.10	(0.08,0.12)	-0.18	0.09	(0.07,0.11)	0.13
$n = 500$	6.84	1	(1,1)	-0.64	0.11	(0.09,0.12)	-0.36	0.06	(0.04,0.07)	0.32
$n = 1,000$	9.62	1	(1,1)	-0.87	0.14	(0.12,0.16)	-0.46	0.09	(0.07,0.11)	0.50
$n = 2,000$	13.59	1	(1,1)	-1.27	0.25	(0.23,0.28)	-0.70	0.11	(0.09,0.13)	0.61
$n = 5,000$	21.43	1	(1,1)	-2.04	0.52	(0.49,0.55)	-1.12	0.19	(0.16,0.21)	0.75

*Note: 1(P<.05) is the rejection rate of the null of H_0:α=1. a is the ratio of missing sample to full sample.

2.4 Estimations with MAR assumption: IPW and MI-IPW methods

We introduce the methods under MAR to deal with missing data when MCAR is violated. MAR assumes that the variables that explain selection (or missing) process are either included in outcome model, as covariates or past outcomes, or they are observed as auxiliary variables.

$$p(s_{it} = 1|y_{it}, \mathbf{z}_{it}) = p(s_{it} = 1|\mathbf{z}_{it}) = p(\mathbf{z}_{it}) \equiv G_{it} \tag{2.22}$$

where $t = 1, 2, 3$ and $\mathbf{z}_{it}$ is all observed variables at t.

Therefore, under MAR assumption of (2.22), individuals who have the same past outcome and covariates have the same conditional probability of selection (or missing) status. In particular, we focus on two of most popular methods to deal with missing data under MAR assumption which are IPW and MI methods.[14]

In this section, we first introduce the idea of the justification for IPW and MI-IPW methods. Then, we examine finite sample properties for IPW and MI-IPW methods via Monte Carlo simulation. We consider two DGPs: (i) outcome and selection models are correctly specified in first DGP and (ii) outcome model is correctly specified but selection model is misspecified. Simulation with misspecified selection model provides information on the robustness property of IPW estimators for MTE to the violation of MAR where the violation

[14] Unfortunately, it is not possible to test MAR assumption since it is never possible to know that all appropriate observed variables have been included as predictors in probability of selection model. Therefore, we suggest a method based on an economic theory for the specification of selection process where the justification of valid MAR is economic theory for the selection process.

of MAR is induced by the selection model misspecification.

2.4.1 IPW method

We first consider IPW method (or efficient version of AIPW method) in Scharfstein et al. [1999], Wooldridge [2007] and Tsiatis [2006]. These methods are based on the idea that, complete-case data which are weighted by valid probability of selection can create pseudo-data that mimic a random sample from the population of interest. Therefore standard estimation method applied to inverse probability weighted complete-case provides a consistent estimator. Correctly specified probability of selection model leads to the satisfaction of MAR assumption so that IPW estimator is consistent.

Consistency for IPW estimator

We consider a linear model with selection process as in (2.23) and (2.24) and we are primarily interested in MTE for $\mathbf{d}_{it}$, α.

$$y_{it} = \mathbf{d}_{it} \cdot \alpha + \mathbf{x}_{1it} \cdot \beta + f_{1t} + c_{1i} + u_{it} \tag{2.23}$$

where for $i = 1, 2, .., n$, $t = 1, 2, 3, 4$, $\mathbf{d}_{it}$ is treatment status indicator and $\mathbf{x}_{1it}$ is time-varying continuous covariates.

We exclusively focus on the missing pattern of monotone drop-out (i.e. attrition) in the modeling of selection process for IPW estimator.[15]

$$p_{i1} = 1;\ p_{it} = 1 \text{ if } \Phi(\mathbf{z}_{it} \cdot \gamma + f_{2t} + c_{2i} + e_{it} > a) \tag{2.24}$$

[15] We consider both monotone and non-monotone missing in the modeling of selection process for MI-IPW estimator.

and $p_{it} = 0$ otherwise for $i = 1, 2, .., n$ and $t = 2, 3, 4$. Then the objective function is provided in (2.25) for the case of balanced panel data model.

$$\sum_{i=1}^{n}\sum_{t=1}^{4} m_{it}(W_{it}, \theta) = 0 \text{ where } m_{it} = \mathbf{x}'_{it} v_{it} \tag{2.25}$$

where $v_{it} = c_i + u_{it}$, $\mathbf{x}_{it} = (\mathbf{d}_{it}, \mathbf{x}_{1it}, f_t)$ and $W_{it} = (y_{it}, \mathbf{x}_{it})$.

As we introduce missing data, we can rewrite the objective function (2.25) for unbalanced panel data with complete-case as the following estimating equation of (2.26).

$$\sum_{i=1}^{n}\sum_{t=1}^{4} s_{it} \cdot m_{it}(W_{it}, \theta) = 0 \tag{2.26}$$

The estimating equation for IPW estimator with complete-case in unbalanced panel data is provided in (2.27).

$$\sum_{i=1}^{n}\sum_{t=1}^{4} \frac{s_{it}}{p_{it}} \cdot m_{it}(\mathbf{W}_{it}, \theta) = 0 \tag{2.27}$$

We suppose $\widehat{p}_{it}$ is obtained and converges to $E(s_{it}|\mathbf{W}_{it})$ in (2.28). Then, under MAR assumption, we can extend the consistency result of unweighted pooled LS estimator in balanced panel data to IPW-pooled IPW estimator of (2.28) in unbalanced panel data.

$$\widehat{\beta}_{PLS} = (\sum_{i=1}^{n}\sum_{t=1}^{T} \frac{s_{it}\mathbf{x}'_{it}\mathbf{x}_{it}}{\widehat{p}(\mathbf{z}_{it})})^{-1}(\sum_{i=1}^{n}\sum_{t=1}^{T} \frac{s_{it}\mathbf{x}'_{it}y_{it}}{\widehat{p}(\mathbf{z}_{it})}) \tag{2.28}$$

where $\widehat{p}_{it}$ is estimated with either probit or logit model in the application.

Recall first that the crucial exogeneity condition for Pooled LS estimator to be consistent in balanced panel is given in (2.29).

$$E(\sum_{t=1}^{T} \mathbf{x}'_{it}(c_i + u_{it})) = 0 \tag{2.29}$$

We show the exogeneity condition of (2.29) for IPW-pooled LS estimator as below. We start from (2.30).

$$E(\sum_{t=1}^{T} s_{it}\mathbf{x}'_{it}(c_i + u_{it})) \tag{2.30}$$

Using law of iterated expectation and MAR, we can rewrite (2.30) as the following.

$$\begin{aligned} E(\sum_{t=1}^{T} s_{it}\mathbf{x}'_{it}(c_i + u_{it})) &= E(\sum_{t=1}^{T} E(s_{it}|y_{it}, \mathbf{z}_{it})\mathbf{x}'_{it}(c_i + u_{it})) \\ = E(\sum_{t=1}^{T} P(s_{it} = 1|y_{it}, \mathbf{z}_{it})\mathbf{x}'_{it}(c_i + u_{it})) &= E(\sum_{t=1}^{T} p(s_{it} = 1|\mathbf{z}_{it})\mathbf{x}'_{it}(c_i + u_{it})) \\ = E(\sum_{t=1}^{T} p(\mathbf{z}_{it})\mathbf{x}'_{it}(c_i + u_{it})) & \end{aligned}$$

Thus, provided that $\widehat{p}(\mathbf{z}_{it})$ is the consistent estimate for the correctly specified $p(\mathbf{z}_{it})$, we obtain consistent IPW-Pooled LS estimator if conditional strict exogeneity of (2.31) is satisfied.

$$\begin{aligned} E(\sum_{t=1}^{T} \frac{s_{it}\mathbf{x}'_{it}(c_i + u_{it})}{p(\mathbf{z}_{it})}) &= E(\sum_{t=1}^{T} E(s_{it}|y_{it}, \mathbf{z}_{it})\frac{\mathbf{x}'_{it}(c_i + u_{it})}{p(\mathbf{z}_{it})}) = E(\sum_{t=1}^{T} p(\mathbf{z}_{it})\frac{\mathbf{x}'_{it}(c_i + u_{it})}{p(\mathbf{z}_{it})}) \\ &= E(\sum_{t=1}^{T} \mathbf{x}'_{it}(c_i + u_{it})) = 0 \end{aligned} \tag{2.31}$$

As shown in (2.31), MAR assumption for selection process is critical for the validity of IPW method.

An example of using MAR for IPW Estimator

We consider the estimation of MTE for CSR on student achievement with project STAR data in illustrating the use of MAR assumption for the probability of selection.

We consider IPW-Pooled LS estimator for missing data.

$$\widehat{\beta}_{PLS} = (\sum_{i=1}^{n}\sum_{t=1}^{T} \frac{s_{it}\mathbf{x}'_{it}\mathbf{x}_{it}}{\widehat{p}(\mathbf{z}_{it})})^{-1}(\sum_{i=1}^{n}\sum_{t=1}^{T} \frac{s_{it}\mathbf{x}'_{it}y_{it}}{\widehat{p}(\mathbf{z}_{it})})$$

Specification for the probability of selection We estimate $p(s_{it}=1|\mathbf{z}_{it})$ using binary logit and probit with maximum likelihood model for $G_{it}(\mathbf{z}_{it},\gamma_0)$.

$$\begin{aligned} G_{it}(\mathbf{z}_{it},\gamma) &\equiv p(s_{it}=1|\mathbf{z}_{it})=p(\mathbf{z}_{it}),\ t=2,3,4 \\ \mathbf{z}_{it} &= (\mathbf{x}_{it-1},\ y_{it-1}),\ \text{or } \mathbf{z}_{it}=(\mathbf{x}_{it-1,...,}\mathbf{x}_1,\ y_{it-1},..,y_{i1}) \end{aligned}$$

where $\mathbf{z}_{it}$ is observed at t

(i) $\mathbf{x}_{it-1}$: class types (initial small, initial aide, cumulative years in small class, cumulative years in small class) , student characteristics (gender, race, free lunch status), teacher characteristics (race, experiences, highest degree dummy), class characteristics (fraction of classmates with kindergarten attendance, fraction of classmates with free lunch, fraction of female classmates , fraction of minority classmates) (ii) y_{it-1} : previous year SAT percentile score

We use logit binary response models for probability of selection

$$\pi_{it}=p(s_{it}=1|\mathbf{z}_{it},s_{it-1}=1)=P(\mathbf{z}_{it}\gamma+\epsilon_{it}>0)=\Lambda(\mathbf{z}_{it}\gamma),s_{it-1}=1.$$

$$\text{where } \Lambda(\mathbf{z}_{it}\gamma)=\frac{exp(\mathbf{z}_{it}\gamma)}{1+exp(\mathbf{z}_{it}\gamma)}$$

For the case of using logit model for selection probability, we can use conditional expectation argument sequentially to estimate selection probability as follows.

$$G_{i1} = 1,\ G_{i2}=\pi_{i2},\ G_{i3}=\pi_{i3}\cdot\pi_{i2},\ G_{i4}=\pi_{i4}\cdot\pi_{i3}\cdot\pi_{i2}$$

$$\widehat{G_{i1}} = 1,\widehat{G_{i2}}=\Lambda(s_{i1}\cdot z_{i2}\widehat{\gamma}),\widehat{G_{i3}}=\Lambda(s_{i1}\cdot z_{i2}\widehat{\gamma})\cdot\Lambda(s_{i2}\cdot z_{i3}\widehat{\gamma})$$

$$\widehat{G_{i4}}=\Lambda(s_{i1}\cdot z_{i2}\widehat{\gamma})\cdot\Lambda(s_{i2}\cdot z_{i3}\widehat{\gamma})\cdot\Lambda(s_{i3}\cdot z_{i4}\widehat{\gamma})$$

Once we obtain $\widehat{\gamma}$ from logit estimation, we can construct weights for inverse probability weighted (IPW) pooled LS estimator as in (2.32).

$$\widehat{\beta}_{PLS} = (\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{s_{it}\mathbf{x}_{it}'\mathbf{x}_{it}}{\widehat{G_{it}}})^{-1}(\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{s_{it}\mathbf{x}_{it}'y_{it}}{\widehat{G_{it}}}) \tag{2.32}$$

Idea behind the consistency of IPW estimator

Under MAR, those cross-section units that have the same predictors values (past outcome, covariates and auxiliary variables) should have the same probability of selection (missing). Thus, remaining cross-section units with predictors of low(high) probability of selection get high weights and play the role of as many cross-section units who leave sample and had the same predictors values. Using valid probability weights eliminates bias due to differential attrition across treatment and control. For instance, in the study of the MTE for CSR on SAT score, if those students who left sample and might have perform well if they remained in sample without treatment were exclusively in regular class, ignoring missing distorts MTE of small class since students in regular class misrepresent population. IPW method recovers random sample property of balanced panel since remaining students with the same predictors in probability of selection model as students who left sample play the role of themselves and missing students by receiving more weights.

2.4.2 AIPW estimator

The objective function of (2.33) provide an AIPW(Augmented IPW) estimator which is consistent and optimally efficient if data are MAR and conditional expectation is correctly

specified.

$$\sum_{i=1}^{n}\sum_{t=1}^{4}[\frac{s_{it}}{p_{it}}\cdot m_{it}(W_{it},\theta)+(1-\frac{s_{it}}{p_{it}})\mathbf{E}(m_{it}|\mathbf{I}_{t-1})]=0 \tag{2.33}$$

where $\mathbf{I}_{t-1}$ is all information available at $t-1$.

IPW estimator is consistent if probability of selection is correctly specified but is generally inefficient. While an AIPW estimator can be efficient and is consistent if either a model specification for probability of selection or a model specification for conditional expectation. (i.e. joint distribution of missing variable) An efficient AIPW estimator can be constructed with appropriate model for conditional expectation in (2.33).[16]

Unfortunately, we do not have much information for parametric specification of the model on conditional expectation, $\mathbf{E}[s_{it}\cdot m_{it}(W_{it},\theta)]=0$, to use in a practical application since it basically requires a correct specification for joint distribution of missing variables and, therefore, in the implementation, we focus on either the consistency of IPW estimator which only requires correct specification for probability of selection or MI-IPW estimator for which MI is based on MICE for $\mathbf{E}[s_{it}\cdot m_{it}(W_{it},\theta)]=0$. In the application of MI, imputation is

[16]AIPW is doubly robust and an estimator is called doubly robust if it is consistent when either a model for conditional expectation or a model for selection is correctly specified. Moreover, AIPW estimator is efficient as it delivers semi-parametric efficiency bound among the class of estimators that use $\sum_{i=1}^{n}\sum_{t=1}^{4}s_{it}\cdot m_{it}(W_{it},\theta)=0$. For more discussion on AIPW estimators is provided in Tsiatis [2006] and Robins et al. [1994]. Recently, we observed theoretical development of AIPW estimator(doubly robust(DR) estimator) which combines IPW method with imputation method as in equation (2.33). For instance, the first-term in objective function (2.33) is the same as IPW use for identification of α with complete-case and second-term in (2.33) is the same objective function as in imputation model.[17] Theoretical attractiveness of DR estimator such as robustness and efficiency has been studied in great details in Carpenter et al. [2006], Tsiatis [2006], Robins and Rotnitzky [1995], and Wooldridge [2007]. However, up to our knowledge, there is no simulation study to evaluate the usefulness of theoretical properties of IPW or DR estimators for the regression model with high-dimensional missing covariates of various forms including binary, count, ordinal and fraction. Most of simulation studies has only one missing continuous variable in the model and uses joint multivariate normal distribution for conditional expectation term in the simulations.(Bang and Robins [2005], Carpenter et al. [2006] and Graham et al. [2008]).

applied only to cross-section units who are remained in the sample with missing variables but not to cross-section units who leave sample once but come back to sample later.

2.4.3 Multiple imputation

We consider imputation based method to deal with missing data as in Rubin [1987], Buuren et al. [1999] and Ragnuthan et al. [2001]. Although multiple imputation has become increasingly popular in epidemiology, psychology, statistics and other social sciences, up to our knowledge, there has been no MI method appropriately designed for multiple treatments and missing occurs for outcome and all of covariates with various types. However, high-dimensional covariates with general missing or attrition is quite common in economic applications and the necessity of correct specification for the joint distribution of all missing covariates and outcome makes MI method unattractive in practice. For instance, in the application with Project STAR data, missing occurs in outcome and covariates including class size, free lunch status dummy, teacher's years of experience, teacher's master degree dummy, teacher's race dummy, fraction of female classmates, fraction of minority classmates and so forth. Therefore, we restrict the use of imputation based method to within sample individuals who have missing variables so that we do not apply imputation for the units who are not in the sample so outcome and all covariates are missing. In most application, for the cross-section units in the sample with missing variables, it is very rare that the number of missing variables is more than three so the burden of approximation is reduced significantly compared to imputing all variables for cross-section units who leave sample. In implementation, we apply a MI method using specification of a logistic distribution for binary variables

and normal distribution for all other continuous variables.

Among many MI methods used in practice, we implement fully conditional approach of multiple imputation in Royston [2004] and Royston [2005], using Stata *ice* command. [18]

Set-up for MICE Estimator

Conditional predictive distribution is estimated from the observed data under MAR. Missing variable are multiply imputed for within sample units. In implementation, we adopt a univariate conditional models in a Gibbs sampler type approach in Buuren et al. [1999]. We model each variable separately conditional on others observed variables. Sequences of missing variables imputation follow from most sample to least sample in terms of number of sample.

$$y_{it} = (y_O, y_M),\ y_O:\ \text{observed}\ , y_M:\ \text{missing} \tag{2.34}$$

$$\mathbf{x}_{it} = (\mathbf{x}_O, \mathbf{x}_M),\ \mathbf{x}_O:\ \text{observed}\ , \mathbf{x}_M:\ \text{missing} \tag{2.35}$$

We need to model all variables with missing data as in (2.36) and (2.37).

$$f(y_{M,it}|\mathbf{W}_{O,it-1})\ ,\ \forall t \tag{2.36}$$

$$f(\mathbf{x}_{M,it}|\mathbf{W}_{O,it-1})\ ,\ \forall t \tag{2.37}$$

where $\mathbf{W}_{O,it-1}$ is all observed variables at $t-1$.

[18] There are currently two widely used methods of model-based imputation: multiple imputation based on multivariate normal distribution for missing variables and multiple imputation by chained equation(MICE) in Buuren et al. [1999]. We implement a MI method as comparison purpose to IPW method since it is readily available to most of statistical packages such as Stata, SAS, and R.

We conduct multiple imputation by chained equations method in Stata implementation of Royston [2005] and type in below command.

```
ice y d x 'auxiliary' using temp.dta, m(10) cmd(d:logit)
```

where m(10) implies that *ice* produces 10 imputed data sets.

2.4.4 MI-IPW method

In unbalanced panel data, we suggest MI-IPW method for which implementation is based on the following. First, we sort data for each cross-section unit at each time into four category: complete case, monotone(attrition), non-monotone-I, and non-monotone-II. Non-monotone-I missing is for cross-section unit which is in sample but has missing variables while non-monotone-II missing is cross-section unit who leave sample but re-enter sample later. We assume MCAR for cross-section unit with missing patterns of non-monotone-II so that we ignore these types of data and still obtain valid statistical analysis. Second, we impute missing variables for non-monotone-I unit using the imputation methods in 2.4.3. Third, by treating imputed non-monotone-I units as complete case, we combine complete case and imputed non-monotone-I data and apply IPW. We obtain MTE $\widehat{\alpha_j}$ and its covariance matrix $\widehat{V_j}$. Finally, we iterate the second and third procedures multiple(M) times. By extending Rubin's formula(Rubin [1987]) for MI method to MI-IPW method, we obtain MI-IPW estimator for coefficient and variance matrix as in (2.38).

$$\widehat{\alpha} = \frac{1}{M}\sum_{j=1}^{M}\alpha_j,\ \widehat{V} = \frac{1}{M}\sum_{j=1}^{M}\widehat{V_j} + \frac{M+1}{M}B \tag{2.38}$$

where $B = \frac{1}{M-1}\sum_{j=1}^{M}(\alpha_j - \alpha)(\alpha_j - \alpha)'$

As probability weight in IPW method eliminates bias due to differential missing while MI for missing variables enhances the efficiency. Both IPW and MI method use MAR assumption in the specification of probability of selection (missing) and missing variables.

2.5 Monte Carlo experiment

We adopt data generating processes (DGPs) with monotone missing data(i.e. missing by attrition) so that data missing occurs simultaneously for outcome and all covariates for a cross-section unit. The covariates in outcome model include both discrete and continuous variables. We are primarily interested in obtaining consistent MTE using IPW method.

2.5.1 DGP I: Correct specification of both outcome and probability of selection models

We start with a DGP I which satisfies the correct specifications for both outcome and selection models. Conditional on auxiliary variables c_{2it}, selection is not correlated with unobserved errors in outcome equation so that MAR is satisfied for DGP I.

Model for data generating process We use a linear model with single treatment and single covariate for DGP. We generate data with n from 100 to 5,000 and t is 4 to mimic the actual size of Project STAR data.

$$y_{it} = d_{it} \cdot \alpha + x_{it} \cdot \beta + c_{1i} + u_{it}$$

where $i = 1, 2, ..., n$, d_{it} is a binary variable and x_{it} is continuous variable.

$$s_{i1} = 1;\ s_{it} = 1 \text{ if } f_i + c_{2it} + \epsilon_{it} > a \text{ and } s_{it-1} = 1,\ \text{for } i = 1, 2, .., n \text{ and } t = 2, 3, 4$$

where a is the cut-off value that determines the fraction of missing sample to full sample and a=0, -1, -2 are chosen and corresponding fraction of missing data to full data is 75, 50, and 25 percent, respectively.

Variables generation for DGP I

1. $d_{it} = 1$ if $f_i + q_{it} > 0$ and $d_{it} = 0$ otherwise, where $f_i \sim iidU(0,1)$, $q_{it} \sim iidN(0,1)$ and α =1

2. $x_{it} \sim iidU(0,1)$, $\beta = -0.2$

3. $c_{1i} = 0$

4. $u_{it} = r_{it} + v_{it}$ where $r_{it} \sim iidU(0,1)$ and

$$\begin{bmatrix} v_{i1} \\ v_{i2} \\ v_{i3} \\ v_{i4} \end{bmatrix} \sim MVN \left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \left(\begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix} \right) \right)$$

   where MVN is multivariate normal and ρ=0.7 in the simulation.

5. $s_{i1} = 1$ and $s_{it} = 1$ if $s_{it-1} = 1$ and $f_i + c_{2it} + \epsilon_{it} > a$ where $\epsilon_{it} \sim iidN(0,1)$, and $c_{2it} = r_{it} + q_{it}$.

Selection process and the potential source of bias The correlation among selection, treatment and unobserved errors are induced to be strictly positive in DGP. First, for a balanced data, d_{it} is not correlated with unobserved errors so that balanced panel unweighted estimators are consistent. The correlation between selection and treatment is induced by

f_i and q_{it} while the correlation between selection and unobserved errors are induced by r_{it}. Therefore, there is strictly positive differential correlation between selection and errors across d_{it}. Therefore, all pooled LS, FE and FD estimators should show finite sample bias. However, under MAR(i.e. conditional on $c_{2it} = r_{it}+q_{it}$ and f_i), the selection is not correlated with unobserved errors so both pooled LS and FD estimators with correct specification of probability of selection should not show bias more than sampling error.

Estimation of probability weight for IPW method Estimation of probability weight is based on (2.39) and (2.40). In Stata, we estimate $\widehat{p}_{it}$ for $t = 2, 3, 4$ using a command (probit s_{it} f_i c_{2it} if s_{it-1} =1 or logit s_{it} f_i c_{2it} if s_{it-1} =1)

$$\pi_{i1} = 1; p_{it}(s_{it} = 1|\mathbf{z}_{it}, s_{it-1} = 1) = \pi_{it} = \Phi(\mathbf{z}_{it} \cdot \gamma + e_{it} > 0) \tag{2.39}$$

where e_{it} is standard normal for $i = 1, 2, .., n$ and $t = 2, 3, 4$.

$$p_{it} = \pi_{it} \cdot \pi_{it-1} \ldots \pi_{i2}, \forall i \text{ and for } t = 2, 3, 4 \tag{2.40}$$

$\widehat{p}_{it}$ for $t = 2, 3, 4$ using a command in Stata as follows:

```
probit s f c2 if s_t-1=1
```

In this DGP, probit model with predictors, f_i and c_{2it}, provides the correct specification for probability of selection.

Imputation of missing variables using MICE In Stata, we impute missing variables for y, d, x using following specifications.

Table 2.7. Specification of models for missing variables

variable	distributional assumption	command	prediction equation
y_{Mit}	Normal	regress	$d_{Oit}\ x_{Oit}\ c_{O2it}\ f_{Oi}$
d_{Mit}	Logistic	logit	$y_{Oit}\ x_{Oit}\ c_{O2it}\ f_{Oi}$
x_{Mit}	Normal	regress	$y_{Oit}\ d_{Oit}\ c_{O2it}\ f_{Oi}$

M represents missing variable and O represent observed variable. In implementation, we type in a Stata command as follows (Royston [2005]): ice y d x c2 f using temp.dta, m(10) cmd(d: logit).

Simulation result: Correctly specified outcome and selection models

Table 2.8 reports unweighted estimates for pooled LS, FE and FD with DGP I. It shows that unweighted estimates are all biased since MCAR is violated. The biases of unweighted estimators do not depend on n but the biases for unweighted estimators decrease as a (missing fraction from full data) falls. Interestingly, all unweighted estimators underestimate the marginal treatment effect (MTE). The magnitude of biases among unweighted pooled LS , FE and FD estimators are not much different from each other.

Table 2.9 shows that IPW estimates for pooled LS and FD obtained using data from DGP I. As theory predicts, both IPW-pooled LS and IPW-FD estimators show no significant finite sample bias as the mean of MC point estimates converges to true value 1 as n increases for all a. For the inference with IPW LS estimator, the rejection frequency converges to true nominal level 0.05 as n increases and the coverage rate of rejection frequency contains 0.05 for all n and $a \leq$0.5. On the other hand, IPW-FD estimates are getting close to true MTE as n increases while the rejection frequency does not converges to 0.05 and the coverage rate of rejection frequency does not contain 0.05 for all n and a. There is over-rejection in the inference of IPW-FD estimator even for n=5,000 and all a. For example, the mean of MC

rejection frequency is .09 at the nominal level of .05.

Table 2.10 reports MI estimates for pooled LS, FE and FD with DGP I. The mean of MC point estimate for MTE of pooled LS is quite close to true value 1 while rejection rate coverage fails to contain 0.05 for all a and n. The means of MC point estimates for MTE and associated rejection frequency of FD are all biased. MI estimates for LS, FE and FD provide incorrect standard errors in finite sample.

Table 2.8. DGP I: MTE estimates for unweighted estimators

Unweighted		LS		FE		FD	
number of n	missing fraction	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)
200	0.75	0.968	.152	.951	.142	.952	.155
200	0.5	0.974	.111	.968	.082	.967	.093
200	0.25	0.987	.090	.986	.065	.987	.073
500	0.75	0.963	.099	.951	.086	.951	.095
500	0.5	0.974	.071	.969	.053	.970	.059
500	0.25	0.985	.058	.984	.040	.984	.046
1,000	0.75	0.963	.071	.951	.062	.951	.068
1,000	0.5	0.974	.048	.968	.037	.967	.042
1,000	0.25	0.988	.041	.985	.028	.985	.033
5,000	0.75	0.961	.031	.951	.026	.951	.029
5,000	0.5	0.974	.022	.969	.016	.969	.018
5,000	0.25	0.987	.018	.985	.013	.985	.015

*Note: Missing fraction is the fraction of missing sample to full sample.

Table 2.9. DGP I: MTE estimates for IPW estimators

IPW		LS				FD			
n	% missing	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)	r	coverage of r	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)	r	coverage of r
200	0.75	1.001	.246	.086	(.074,.099)	.986	.238	.143	(.128,.159)
200	0.5	.997	.142	.060	(.050,.070)	.993	.125	.090	(.077,.103)
200	0.25	1.001	.096	.050	(.040,.060)	1.000	.081	.090	(.077,.103)
500	0.75	.996	.179	.080	(.067,.091)	.985	.169	.122	(.107,.136)
500	0.5	1.001	.091	.047	(.037,.056)	.998	.079	.087	(.075,.099)
500	0.25	.998	.061	.049	(.040,.058)	.997	.050	.086	(.074,.098)
1,000	0.75	.997	.137	.078	(.066,.089)	.986	.127	.106	(.092,.119)
1,000	0.5	1.001	.065	.047	(.037,.056)	.996	.057	.096	(.083,.109)
1,000	0.25	1.000	.043	.049	(.039,.058)	.998	.036	.086	(.074,.098)
5,000	0.75	.994	.070	.055	(.045,.064)	.992	.065	.087	(.074,.099)
5,000	0.5	1.001	.030	.048	(.038,.057)	.997	.026	.082	(.070,.094)
5,000	0.25	1.001	.019	.046	(.036,.055)	.999	.016	.092	(.079,.105)

*Note: $r \equiv 1(p < .05)$ is the rejection rate for the null of H_0:α=1 with nominal value 0.05.; Cluster robust standard error has been used in all estimations.

Table 2.10. DGP I: MTE for estimators with multiple imputations

MI		LS		FE				FD			
n	% missing	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)	r	coverage of r	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)	r	coverage of r
200	0.75	1.006	.168	.980	.161	.551	(.529,.573)	.947	.157	.089	(.077,.102)
200	0.5	1.007	.112	.994	.093	.261	(.242,.280)	.973	.091	.065	(.054,.076)
200	0.25	1.002	.094	.996	.068	.128	(.113,.142)	.984	.074	.064	(.053,.074)
500	0.75	1.013	.106	.988	.101	.563	(.541,.585)	.951	.097	.093	(.080,.106)
500	0.5	1.003	.071	.991	.059	.277	(.257,.297)	.968	.058	.093	(.080,.106)
500	0.25	1.004	.059	.999	.042	.108	(.095,.122)	.986	.046	.065	(.054,.076)
1,000	0.75	1.011	.074	.987	.071	.559	(.537,.581)	.952	.066	.110	(.096,.124)
1,000	0.5	1.006	.051	.994	.042	.274	(.255,.294)	.970	.041	.107	(.094,.121)
1,000	0.25	1.000	.040	.998	.029	.096	(.083,.108)	.985	.033	.082	(.070,.094)

*Note: $r \equiv 1(p < .05)$ is the rejection rate for the null of H_0:α=1 with nominal value 0.05.; Cluster robust standard error has been used in all estimations.

2.5.2 DGP II: The correct probability of selection model with endogenous treatment effect

Consider the same data generating process as DGP I except $c_{i1} = f_i + a_i$ so that, even for a balanced panel data, the treatment, d_{it}, is correlated with unobserved heterogeneity. However, FE or FD transformation eliminates the source of bias in balanced panel data. In unbalanced panel data, unweighted pooled LS, FE and FD estimators are biased due to the correlations between selection and the treatment and between selection and time-varying errors. However, under MAR for probability of selection, IPW-FD estimator can eliminate the source of differential correlation between selection and unobserved heterogeneity across d_{it}. Once MAR assumption holds and the probability of selection is correctly specified, IPW-FD estimator should not show any bias more than finite sample error in simulation.

Simulation results

In table 2.11, unweighted pooled LS, FE and FD estimators show finite sample bias for the data with all n and $a > 0$. Unweighted Pooled LS estimator overestimates MTE while FE and FD estimators underestimate MTE for all n and a. The biases do not depend on n for all unweighted pooled LS, FE and FD estimators while the biases decrease as a (the fraction of missing sample to full sample) decreases.

Table 2.12 reports that IPW LS is biased and overestimates MTE. Compared to unweighted LS, IPW LS estimates has much larger variance but somewhat smaller bias. The bias for IPW LS estimator does not depend on n and a and the size distortion is prohibitively large for all n and a. On the other hand, IPW FD estimates are very close to true marginal

effect 1 while rejection frequency shows some evidence of over-rejection. Although the size distortion decreases as n increases, over-rejection problem remain even for n=5,000 and all a. The size distortion for IPW FD estimator does not disappear even for n=20,000.

Table 2.13 shows pooled LS FE and FD with MI . MC estimates show finite sample bias for all n and a. Pooled LS estimator with MI overestimates the marginal treatment effect while MI FD estimators underestimate the marginal treatment effect for all n and a. The biases do not depend on n for both MI LS and MI FD estimators while the biases decrease as a (fraction of missing sample to full sample) decreases. The mean of MC point estimates for MI FE estimator shows that the bias is quite small for all n and $a \leq 0.5$ while the size distortion is large and is negatively correlated to a.

Table 2.11. DGP II: MTE for unweighted estimators

Unweighted		LS		FE		FD	
number of n	missing fraction	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)
200	0.75	1.144	.171	.949	.140	.949	.154
200	0.5	1.142	.122	.968	.083	.968	.094
200	0.25	1.136	.098	.985	.064	.985	.064
500	0.75	1.150	.110	.952	.088	.952	.098
500	0.5	1.145	.077	.969	.051	.969	.058
500	0.25	1.138	.060	.985	.040	.984	.046
1,000	0.75	1.148	.075	.951	.059	.952	.065
1,000	0.5	1.143	.054	.970	.037	.970	.041
1,000	0.25	1.137	.043	.985	.028	.985	.031
5,000	0.75	1.148	.034	.952	.027	.952	.030
5,000	0.5	1.144	.024	.969	.016	.969	.018
5,000	0.25	1.138	.020	.985	.013	.985	.014

*Note: Missing fraction is the fraction of missing sample to full sample. Sensitivity analysis is conducted for unweighted LS, FE and FD estimators.

Table 2.12. DGP II: MTE estimates for IPW estimators

IPW		LS				FD			
n	% missing	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)	r	coverage of r	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)	r	coverage of r
200	0.75	1.129	.272	.151	(.136,.167)	.983	.235	.154	(.139,.171)
200	0.5	1.131	.150	.181	(.164,.198)	.993	.118	.098	(.084,.110)
200	0.25	1.132	.104	.261	(.242,.280)	.997	.080	.094	(.081,.106)
500	0.75	1.131	.196	.175	(.158,.191)	.987	.170	.129	(.114,.143)
500	0.5	1.134	.077	.344	(.323,.365)	.995	.079	.090	(.077,.103)
500	0.25	1.133	.063	.542	(.520,.564)	.997	.050	.091	(.078,.104)
1,000	0.75	1.136	.147	.253	(.233,.272)	.992	.127	.085	(.072,.097)
1,000	0.5	1.132	.072	.538	(.516,.560)	.998	.058	.084	(.071,.096)
1,000	0.25	1.133	.046	.831	(.815,.847)	.998	.035	.083	(.070,.095)
5,000	0.75	1.136	.074	.620	(.598,.641)	.991	.064	.085	(.073,.097)
5,000	0.5	1.133	.033	.972	(.965,.979)	.997	.026	.073	(.062,.084)
5,000	0.25	1.134	.021	.999	(.997,1)	.999	.015	.076	(.064,.088)

*Note: $r \equiv 1(p < .05)$ is the rejection rate for the null of H_0:α=1 with nominal value 0.05.; Cluster robust standard error has been used in all estimations.

Table 2.13. DGP II: MTE estimates for estimators with multiple imputations

MI		LS		FE				FD			
n	% missing	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)	r	coverage of r	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)	r	coverage of r
200	0.75	1.141	.176	.982	.163	.547	(.525,.569)	.950	.154	.073	(.062,.085)
200	0.5	1.142	.121	.993	.096	.270	(.251,.290)	.968	.094	.075	(.064,.087)
200	0.25	1.139	.100	.998	.067	.107	(.094,.121)	.985	.072	.056	(.046,.066)
500	0.75	1.145	.111	.989	.107	.578	(.557,.600)	.953	.097	.087	(.075,.100)
500	0.5	1.141	.076	.994	.060	.259	(.239,.278)	.970	.059	.084	(.072,.096)
500	0.25	1.136	.060	.997	.041	.100	(.087,.113)	.985	.046	.057	(.047,.067)
1,000	0.75	1.143	.078	.986	.073	.572	(.550,.594)	.949	.066	.117	(.102,.131)
1,000	0.5	1.137	.053	.991	.043	.289	(.269,.309)	.968	.040	.114	(.100,.128)
1,000	0.25	1.136	.045	.997	.030	.099	(.087,.113)	.984	.032	.080	(.068,.092)

*Note: $r \equiv 1(p < .05)$ is the rejection rate for the null of H_0:α=1 with nominal value 0.05.; Cluster robust standard error has been used in all estimations.

2.5.3 The robustness of IPW estimator to misspecification of the probability of selection

This section provides the simulation results for the sensitivity of probability of selection model misspecification. One of main criticism for IPW method is its instability of estimates to the misspecification of the probability of selection. We test the sensitivity of IPW estimator to the misspecification of probability of selection model using a DGP in which an omitted variable induces misspecification of probability of selection. For instance, in STAR empirical analysis, it is possible that we possibly omit some important family characteristics in the selection model.

DGP with selection model misspecification We use a linear model with a treatment and a covariate. We generate data with n from 100 to 5,000 and t=4 to mimic the actual size of Project STAR data.

$$y_{it} = d_{it} \cdot \alpha + x_{it} \cdot \beta + c_{1i} + u_{it}, \text{ for } i = 1, 2, .., n \text{ and } t = 1, 2, 3, 4$$

where d_{it} a binary variable and x_{it} is continuous variable.

Variable generation conducted as follows.

1. $d_{it} = 1$ if $f_i + q_{it} > 0$ and $d_{it} = 0$ otherwise, where $f_i \sim iidU(0,1)$, $q_{it} \sim iidN(0,1)$ and α =1

2. $x_{it} \sim iidU(0,1)$, $\beta = -0.2$

3. $c_{1i} \sim iidU(0,1)$

4. $u_{it} = r_{it} + v_{it}$ where $r_{it} = f_i + iidU(0,1)$ and

$$\begin{bmatrix} v_{i1} \\ v_{i2} \\ v_{i3} \\ v_{i4} \end{bmatrix} \sim MVN \left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \left(\begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix} \right) \right)$$

where MVN is multivariate normal and ρ=0.7 in the simulation.

The selection is complicated by including an interaction term which is omitted in the logit model estimation.

$$s_{i1} = 1 \text{ and } s_{it} = 1 \text{ if } s_{it-1} = 1 \text{ and } f_i + \delta y_{it-1} * c_{2it} + c_{2it} + \epsilon_{it} > a$$

where $\epsilon_{it} \sim iidN(0,1)$ and $c_{2it} = r_{it} + q_{it}$ and where a is the cut-off value that determines the fraction of missing sample to full sample and a=0, -1, -2 are chosen and corresponding fraction of missing data are about 75, 50, and 25 percents, respectively. For correctly specified selection model, IPW FD estimator is consistent since the unique source of differential correlation, f_i, can be eliminated by FD transformation.

The probability weights for IPW are obtained using following model specification for selection process, $\widehat{\pi_{it}}$, which is given by:

$$\pi_{it} = \frac{exp(\mathbf{z}_{it}\gamma)}{1 + exp(\mathbf{z}_{it}\gamma)}, \ t = 2,3,4 \text{ for } s_{it-1} = 1$$

where $\mathbf{z}_{it}$=(f_i, c_{2it}) and the omission of interaction term and logit estimation instead of probit lead to misspecification of selection model.

Simulation Result: Sensitivity of IPW estimators to probability of selection model misspecification We simulate with four values for the parameter of misspecifi-

cation, δ=0, 0.2, 0.5, and 1 where δ=0 implies correct specification. As the values of δ increases, the degree of misspecification increases. For instance, in DGP with δ=0.2, the variation by the omitted variable $(y_{it-1} * c_{2it})$ represents about 3 percent of total variation. Moreover, as δ increases to .5 and 1, the fraction of variation due to the variation of the omitted variable represents 10 and 25 percent of total variation respectively. This simulation intends to clarify the impact of misspecification of selection since we possibly do not observe some important variables in the selection model. Tables from 2.14 to 2.21 show simulation results. Even with the misspecification for the probability of selection model, the bias of IPW estimators is no greater than the bias of unweighted estimators with $\delta \leq$0.5. Figure 2.1 compares biases between unweighted and IPW methods for LS and FD estimators. Especially if missing fraction is less than or equal to 0.5, the biases for IPW estimators are less than those of unweighted estimators in the presence of misspecification of selection model for $\delta \leq$0.5. In sum, simulation results show that the estimates from IPW FD estimator does not exacerbate the bias even if selection model is misspecified when it is compared to the estimates from unweighted FD estimator with $\delta \leq$0.5.

Table 2.14. Selection model misspecification: MTE for unweighted estimators, δ=0

Unweighted		LS		FE		FD	
number of n	missing fraction	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)
100	0.75	1.265	.246	.955	.197	.953	.219
100	0.5	1.267	.169	.973	.102	.975	.118
100	0.25	1.269	.154	.981	.092	.981	.104
200	0.75	1.257	.179	.951	.136	.951	.147
200	0.5	1.258	.115	.976	.073	.977	.082
200	0.25	1.261	.106	.983	.064	.984	.073
500	0.75	1.251	.111	.949	.086	.948	.096
500	0.5	1.253	.072	.973	.044	.974	.052
500	0.25	1.257	.066	.982	.040	.982	.046
1,000	0.75	1.259	.081	.951	.063	.951	.070
1,000	0.5	1.259	.049	.975	.032	.974	.036
1,000	0.25	1.260	.045	.982	.028	.982	.032

*Note: Missing fraction is the fraction of missing sample to full sample. δ represents the degree of misspecification of the selection model.

Table 2.15. Selection model misspecification: MTE for IPW estimators, δ=0

IPW		LS				FD			
n	% missing	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)	r	coverage of r	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)	r	coverage of r
100	0.75	1.271	.354	.209	(.181,.236)	.989	.299	.171	(.146,.196)
100	0.5	1.271	.193	.342	(.313,.371)	.994	.139	.090	(.072,.108)
100	0.25	1.275	.165	.413	(.383,.444)	.994	.113	.081	(.064,.098)
200	0.75	1.250	.275	.233	(.207,.260)	.982	.216	.120	(.100,.140)
200	0.5	1.261	.134	.552	(.521,.583)	.998	.097	.087	(.070,.104)
200	0.25	1.265	.115	.663	(.634,.692)	1.000	.083	.079	(.062,.096)
500	0.75	1.243	.199	.384	(.354,.414)	.984	.160	.116	(.096,.136)
500	0.5	1.263	.086	.876	(.856,.896)	.998	.064	.088	(.070,.106)
500	0.25	1.263	.073	.955	(.942,.968)	.998	.051	.077	(.060,.094)
1,000	0.75	1.255	.153	.553	(.522,.584)	.990	.116	.097	(.079,.115)
1,000	0.5	1.266	.058	.990	(.984,.996)	.998	.043	.084	(.067,.101)
1,000	0.25	1.266	.049	.998	(.995,1)	.998	.035	.074	(.058,.090)

*Note: $r \equiv 1(p < .05)$ is the rejection rate for the null of H_0:α=1 with nominal value 0.05.; Cluster robust standard error has been used in all estimations.

Table 2.16. Selection model misspecification: MTE for unweighted estimators, δ=.2

Unweighted		LS		FE		FD	
number of n	missing fraction	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)
100	0.75	1.401	.240	.938	.165	.926	.183
100	0.5	1.303	.157	.973	.095	.966	.110
100	0.25	1.288	.146	.984	.089	.978	.102
200	0.75	1.417	.169	.941	.115	.929	.126
200	0.5	1.317	.112	.973	.069	.966	.078
200	0.25	1.295	.104	.982	.064	.977	.073
500	0.75	1.418	.107	.941	.075	.932	.084
500	0.5	1.316	.072	.974	.043	.968	.049
500	0.25	1.293	.067	.982	.040	.977	.045
1,000	0.75	1.415	.075	.939	.050	.929	.057
1,000	0.5	1.314	.050	.975	.030	.969	.035
1,000	0.25	1.292	.047	.983	.028	.979	.032

*Note: Missing fraction is the fraction of missing sample to full sample.

Table 2.17. Selection model misspecification: MTE for IPW estimators, δ=.2

IPW		LS				FD			
n	% missing	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)	$r \equiv 1(p < .05)$	coverage of r	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)	$r \equiv 1(p < .05)$	coverage of r
100	0.75	1.381	.300	.318	(.289,.347)	.930	.231	.141	(.119,.163)
100	0.5	1.288	.158	.458	(.426,.489)	.970	.115	.114	(.094,.134)
100	0.25	1.281	.144	.488	(.455,.521)	.981	.103	.106	(.086,.126)
200	0.75	1.405	.216	.533	(.502,.564)	.937	.163	.129	(.108,.150)
200	0.5	1.305	.116	.780	(.754,.806)	.970	.081	.116	(.096,.136)
200	0.25	1.289	.105	.809	(.784,.883)	.981	.074	.107	(.088,.126)
500	0.75	1.401	.140	.855	(.833,.877)	.942	.109	.141	(.119,.163)
500	0.5	1.304	.074	.992	(.986,.998)	.973	.051	.137	(.116,.158)
500	0.25	1.286	.068	.989	(.983,.995)	.981	.046	.129	(.108,.150)
1,000	0.75	1.395	.095	.980	(.971,.989)	.940	.073	.185	(.161,.209)
1,000	0.5	1.302	.050	1	(1,1)	.974	.036	.181	(.157,.205)
1,000	0.25	1.286	.047	1	(1,1)	.982	.033	.134	(.113,.155)

*Note: 1(P<.05) is the rejection rate for the null of H_0:α=1 with nominal value 0.05.; Cluster robust standard error has been used in all estimations.

Table 2.18. Selection model misspecification: MTE for unweighted estimators, δ=.5

Unweighted		LS		FE		FD	
number of n	missing fraction	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)
100	0.75	1.416	.215	.936	.142	.916	.157
100	0.5	1.303	.147	.971	.094	.962	.110
100	0.25	1.292	.143	.980	.089	.976	.105
200	0.75	1.424	.146	.936	.097	.918	.110
200	0.5	1.307	.101	.974	.064	.967	.074
200	0.25	1.288	.097	.982	.060	.977	.070
500	0.75	1.424	.098	.939	.062	.922	.070
500	0.5	1.306	.071	.974	.042	.965	.047
500	0.25	1.289	.068	.982	.039	.976	.045
1,000	0.75	1.423	.067	.941	.042	.924	.048
1,000	0.5	1.305	.048	.976	.028	.968	.033
1,000	0.25	1.287	.046	.983	.027	.978	.031

*Note: Missing fraction is the fraction of missing sample to full sample.

Table 2.19. Selection model misspecification: MTE for IPW estimators, δ=.5

IPW		LS				FD			
n	% missing	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)	$r \equiv 1(p < .05)$	coverage of r	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)	$r \equiv 1(p < .05)$	coverage of r
100	0.75	1.379	.259	.400	(.370,.431)	.914	.178	.142	(.120,.164)
100	0.5	1.297	.154	.505	(.472,.539)	.962	.112	.120	(.098,.141)
100	0.25	1.290	.145	.495	(.454,.535)	.977	.104	.114	(.088,.140)
200	0.75	1.389	.177	.656	(.627,.685)	.910	.124	.174	(.150,.198)
200	0.5	1.300	.105	.798	(.773,.823)	.968	.075	.109	(.090,.129)
200	0.25	1.285	.099	.799	(.773,.825)	.978	.071	.099	(.080,.119)
500	0.75	1.385	.113	.942	(.927,.957)	.918	.077	.257	(.230,.284)
500	0.5	1.299	.073	.991	(.985,.997)	.967	.048	.169	(.146,.192)
500	0.25	1.286	.069	.992	(.986,.998)	.977	.045	.138	(.117,.160)
1,000	0.75	1.384	.079	.995	(.991,.999)	.920	.053	.405	(.375,.435)
1,000	0.5	1.297	.049	1	(1,1)	.969	.033	.215	(.189,.241)
1,000	0.25	1.284	.046	1	(1,1)	.979	.031	.161	(.138,.184)

*Note: 1(P<.05) is the rejection rate for the null of H_0:α=1 with nominal value 0.05.; Cluster robust standard error has been used in all estimations.

Table 2.20. Selection model misspecification: MTE for unweighted estimators, δ=1.0

Unweighted		LS		FE		FD	
number of n	missing fraction	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)
100	0.75	1.378	.207	.929	.129	.905	.150
100	0.5	1.314	.156	.961	.093	.944	.107
100	0.25	1.298	.151	.966	.091	.955	.106
200	0.75	1.382	.150	.936	.093	.910	.107
200	0.5	1.313	.114	.966	.067	.952	.077
200	0.25	1.297	.110	.973	.064	.963	.072
500	0.75	1.374	.090	.931	.057	.905	.064
500	0.5	1.313	.067	.964	.042	.948	.047
500	0.25	1.299	.065	.972	.040	.960	.045
1,000	0.75	1.376	.064	.933	.038	.907	.045
1,000	0.5	1.311	.049	.964	.029	.949	.034
1,000	0.25	1.298	.047	.972	.027	.962	.032

*Note: Missing fraction is the fraction of missing sample to full sample.

Table 2.21. Selection model misspecification: MTE for IPW estimators, δ=1.0

IPW		LS				FD			
n	% missing	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)	$r \equiv 1(p < .05)$	coverage of r	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)	$r \equiv 1(p < .05)$	coverage of r
100	0.75	1.402	.248	.442	(.410,.473)	.902	.167	.170	(.146,.193)
100	0.5	1.323	.160	.529	(.496,.563)	.943	.108	.133	(.110,.156)
100	0.25	1.305	.154	.527	(.488,.565)	.955	.107	.127	(.102,.153)
200	0.75	1.412	.176	.730	(.702,.758)	.903	.115	.208	(.183,.233)
200	0.5	1.323	.118	.807	(.783,.832)	.952	.077	.149	(.127,.171)
200	0.25	1.305	.112	.817	(.792,.841)	.964	.072	.129	(.108,.150)
500	0.75	1.402	.105	.981	(.973,.989)	.896	.069	.416	(.385,.447)
500	0.5	1.322	.069	.997	(.994,1)	.947	.048	.257	(.230,.284)
500	0.25	1.307	.066	.996	(.992,1)	.960	.045	.204	(.179,.229)
1,000	0.75	1.406	.074	1	(1,1)	.897	.049	.614	(.584,.644)
1,000	0.5	1.321	.050	1	(1,1)	.948	.034	.423	(.392,.454)
1,000	0.25	1.305	.047	1	(1,1)	.962	.032	.297	(.269,.325)

*Note: 1(P<.05) is the rejection rate for the null of H_0:α=1 with nominal value 0.05.; Cluster robust standard error has been used in all estimations.

2.5.4 The robustness of first-differencing estimator with small within-variation

In most experiments, program participants ideally maintain their treatment or control groups status from the start to the end of a program. This inevitably induces very small within-variation for treatment variable and this is one of main criticisms for fixed effects and first differencing estimators in the program evaluation studies. Thus, we examine the sensitivity of IPW first-differencing estimator to small within-variation of treatment variable. The magnitude of within-variation is induced randomly and we assume correct specification for the probability of selection.

Motivation of sensitivity test for IPW FD estimator to small within variation

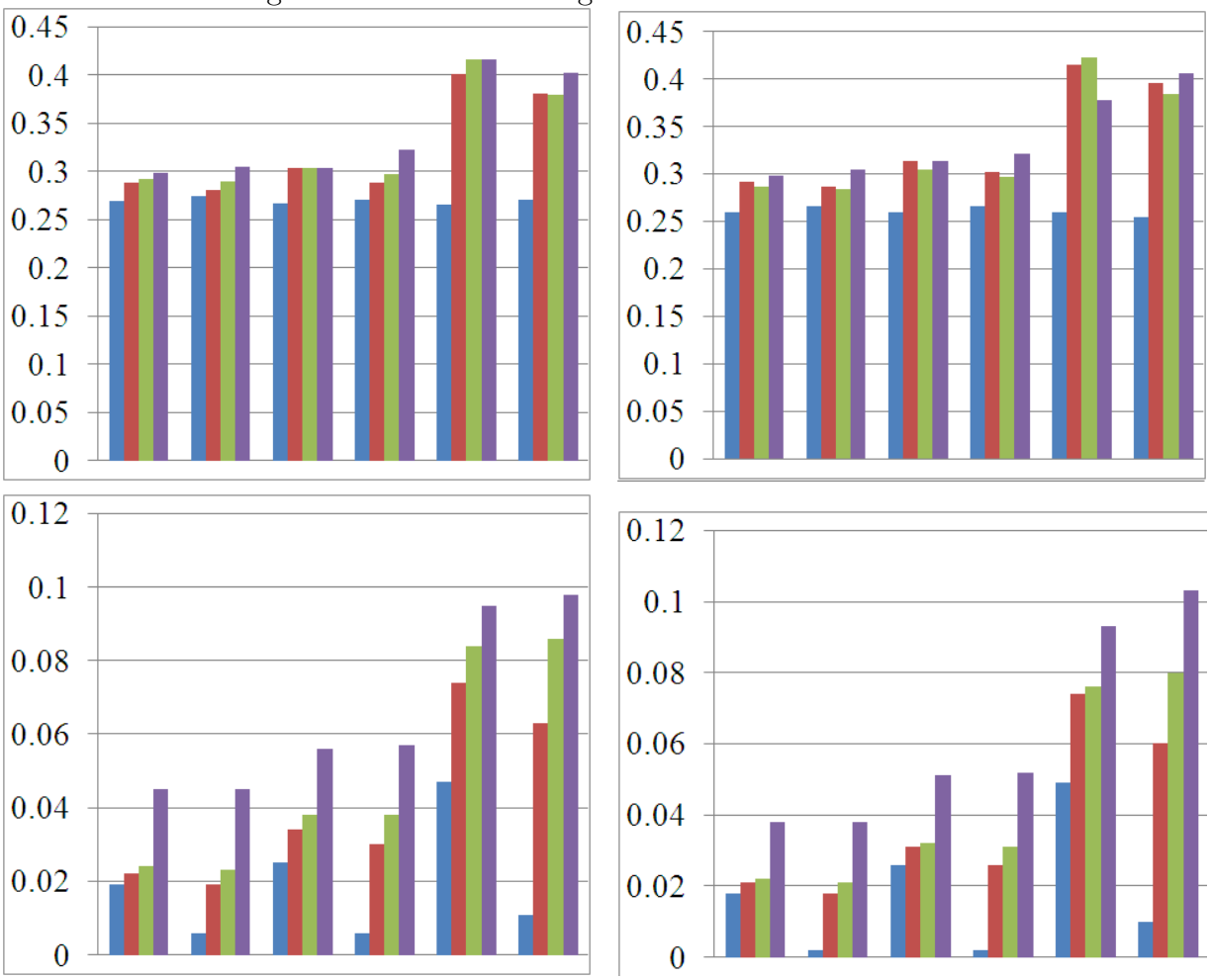
Fixed effects and first differencing estimators are very effective methods to eliminate the source of bias when across treatment and control groups, response variables are systematically different for missing data individual units and its difference is due to unobserved heterogeneity. However, identification of first differencing estimator is based only on within variation.

DGP for d_{it} with small within variation

Three different magnitudes of within-variation for treatment status d_{it} are induced randomly in the simulation.

No restriction on treatment status $L = 0$ First type use the generation of treatment variable without any restriction for within-variation imposed on for treatment variable. No

Figure 2.1. Bias of unweighted and IPW LS estimator



(Note) For figures in first row, six sets of bar represent LS (a=.25), IPW-LS (a=.25), LS (a=.5), IPW-LS (a=.5), LS(a=.75), and IPW-LS (a=.75) estimates from the left to the right where true value is 0. The colors of bar represent degree of mis-specification - blue (correctly specified), Red (δ=.2), green (δ=.5), and Purple (δ=1.0)- where δ is the magnitude of misspecification coefficient in the selection model and a is fraction of missing data. For figures in second row, six sets of bar represent FD (a=.25), IPW-FD (a=.25), FD (a=.5), IPW-FD (a=.5), FD(a=.75), and IPW-FD (a=.75) estimates from the left to the right. In both columns, the left panel reports simulation results for n=100 and right panel for n=1000.

restriction induces d_{it} for which about 68 percent of sample change status during four time periods.

$d_{it} = 1$ if $f_i + q_{it} > 0$ and $d_{it} = 0$ otherwise, where $f_i \sim iidU(0,1)$, $q_{it} \sim iidN(0,1)$

Small within variation treatment variable Smaller within-variations for the treatment status were induced with the following DGP. $d_{i1} = 1$ if $f_i + q_{i1} > 0$ and $d_{i1} = 0$ otherwise. $d_{it} = d_{it-1}$ if $-L < f_i + q_{it} < L$ and $d_{it} = 0$ if $f_i + q_{it} < -L$ and $d_{it} = 1$ if $f_i + q_{it} > L$. Therefore, the value L determines the magnitude of fraction of within-variation for a treatment variable.

$s_{i1} = 1$ and $s_{it} = 1$ if $s_{it-1} = 1$ and $f_i + c_{2it} + \epsilon_{it} > a$ where $\epsilon_{it} \sim iidN(0,1)$, and $c_{2it} = r_{it} + q_{it}$.

where a is cutoff value which is adjusted from DGP I.

The model for selection process is assumed to be specified correctly so that $\widehat{\pi_{it}}$ is obtained from the following:

$$\pi_{it} = \frac{exp(\mathbf{z}_{it}\gamma)}{1 + exp(\mathbf{z}_{it}\gamma)}, \ t = 2,3,4 \text{ for } s_{it-1} = 1$$

where $\mathbf{z}_{it}$=(f_i, c_{2it}).

Sensitivity of IPW FD estimator to small within variation for a treatment variable

We test the sensitivity of IPW FD estimator to small within variation of a treatment variable. DGPs are a treatment variable with full variation, with 50 % variation of units and with 15 % variation of units. The simulation intends to clarify the impact of small within variation on the consistency of IPW FD estimator.

Simulation results Tables from 2.22 to 2.24 show that the bias of FD and IPW-FD estimators to small within variation treatment. The key is that for missing fraction less than 50 % and within variation more than 50 %, IPW-FD estimate quite consistently estimate MTE. Figure 2.2 compares the biases between FD and IPW-FD methods. Especially if missing fraction is less than 0.5, IPW-FD method works quite well. Unless variation is less than 25 percent, the bias by random small within variation is not substantial.

Table 2.22. Sensitivity analysis to small within-variation: MTE with full within-variation

		FD		IPW-FD			
number of n	missing fraction	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)	$r \equiv 1(p < .05)$	coverage of r
100	0.75	.946	.219	.974	.290	.147	(.123,.171)
100	0.5	.967	.130	.992	.155	.104	(.085,.122)
100	0.25	.984	.101	.994	.107	.083	(.066,.101)
200	0.75	.949	.150	.974	.224	.126	(.105,.146)
200	0.5	.970	.090	.998	.111	.076	(.060,.092)
200	0.25	.987	.072	1.000	.078	.083	(.066,.100)
500	0.75	.948	.096	.979	.169	.122	(.101,.142)
500	0.5	.967	.057	.993	.079	.089	(.071,.107)
500	0.25	.987	.046	.998	.049	.095	(.077,.113)
1,000	0.75	.947	.068	.985	.124	.093	(.080,.106)
1,000	0.5	.968	.040	.994	.055	.084	(.067,.101)
1,000	0.25	.986	.032	.999	.036	.100	(.081,.119)

*Note: Missing fraction is the fraction of missing sample to full sample.

Table 2.23. Sensitivity analysis: MTE with within-variation of 50% sample only

		FD		IPW-FD			
number of n	missing fraction	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)	$r \equiv 1(p < .05)$	coverage of r
100	0.75	.941	.302	.964	.392	.190	(.163,.216)
100	0.5	.962	.177	.993	.218	.109	(.090,.129)
100	0.25	.987	.134	1.001	.145	.087	(.069,.105)
200	0.75	.952	.210	.987	.294	.145	(.123,.167)
200	0.5	.964	.120	.995	.156	.100	(.081,.119)
200	0.25	.983	.095	.998	.106	.081	(.064,.098)
500	0.75	.941	.137	.979	.223	.142	(.120,.164)
500	0.5	.961	.077	.998	.100	.071	(.055,.087)
500	0.25	.984	.058	.999	.067	.063	(.048,.078)
1,000	0.75	.933	.091	.983	.164	.115	(.095,.135)
1,000	0.5	.965	.053	1.001	.072	.063	(.048,.078)
1,000	0.25	.983	.041	1.000	.045	.071	(.055,.087)

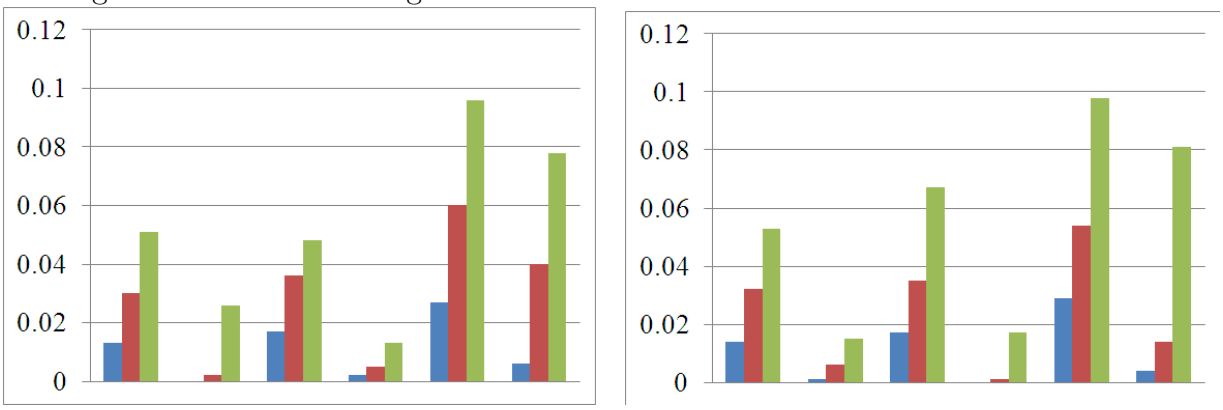
*Note: Missing fraction is the fraction of missing sample to full sample.

Table 2.24. Sensitivity analysis: MTE with within-variation of 15% sample only

		FD		IPW-FD			
number of n	missing fraction	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)	Mean($\widehat{\alpha}$)	SD($\widehat{\alpha}$)	$r \equiv 1(p < .05)$	coverage of r
100	0.75	.733	.735	.731	.793	.566	(.532,.599)
100	0.5	.935	.418	.968	.479	.221	(.195,.247)
100	0.25	.959	.280	.981	.299	.117	(.097,.137)
200	0.75	.904	.571	.922	.669	.412	(.381,.442)
200	0.5	.940	.273	.960	.343	.155	(.132,.177)
200	0.25	.973	.186	.994	.212	.090	(.077,.102)
500	0.75	.927	.314	.973	.485	.251	(.224,.278)
500	0.5	.942	.164	.974	.233	.116	(.096,.136)
500	0.25	.976	.116	1.003	.135	.060	(.045,.075)
1,000	0.75	.902	.208	.919	.385	.191	(.167,.215)
1,000	0.5	.946	.115	.986	.170	.097	(.079,.115)
1,000	0.25	.971	.082	1.004	.096	.055	(.041,.070)

*Note: Missing fraction is the fraction of missing sample to full sample.

Figure 2.2. Bias of unweighted and IPW FD estimator to small within variation



(Note) Six sets of bar represent FD and IPW-FD estimator according to covariate variation. From left to right, sets of bar represent FD (v=1), IPW-FD (v=1), FD (v=.5), IPW-FD (v=.5), FD (v=.15), and IPW-FD (v=.15) where v represents fraction of covariates which vary over time. Missing fraction is represent by the color of bar - Blue:25 % , Red: 50 %, and Green: 75 %. Left panel reports the estimates for n=200 and right panel reports the estimates for n=1,000.

2.6 Empirical application: The return to class size reduction (CSR) on SAT score

In this section, we estimate the effect of CSR on SAT score for grades in K-3 using Tennessee's STAR experiment data which is accumulated for a single cohort of students from 1985 to 1989. Data for class size, student characteristics, teacher characteristics, classroom peer characteristics, and student achievement scores are collected. However, substantial proportion of data is missing so that only about 50 percent of sample is complete-case. Out of 11,601 students, for each grade from K to 3, about 6,800 students are in the program in each grade.

We start by testing MCAR using Hausman-type test and variable addition test to deter-

mine whether we can use unweighted estimator with complete-case. As we reject the null of MCAR, we apply IPW and MI-IPW methods using Becker [1993]'s parent school choice model for the specification of the probability of selection.

This section proceeds as follows. We first briefly describe Tennessee's STAR experiment and provide detailed information on noncompliance and missing patterns of data.[19]Second, we provide Hausman-type test and variable addition test of MCAR. Third, we report IPW estimates and compare this to unweighted estimates.

2.6.1 Background information on Project STAR experiment of CSR

Origination of experiment Observational data has not been able to find consistent class size reduction effect on student performance (Hanushek [1986] and Hanushek [1997]) while experimental studies of class size reduction show strong and positive effect.[20] Therefore, the Tennessee legislature funded a randomized experiment, Tennessee's Project STAR (Student/Teacher Achievement Ratio), to provide the definitive answer to the effectiveness of CSR on student achievement in 1985.[21]

Experiment design It was a randomized experiment that assigned students and teachers to three different class types, a small-size class (target of 13-17 students), a regular-size

[19] More detailed description of STAR experiment is provided in Word et al. [1990].

[20] For instance, experimental studies such as Tennessee's Project STAR (Student/Teacher Achievement Ratio), Wisconsin's Student Achievement Guarantee in Education (SAGE) Program and Israel's Maimonides' Rule reported a strong link between class size reduction and improvement in student achievement.

[21] The experiment funded by the Tennessee State Legislature under Lama Alexander at a total cost of $12 million with $3 million for the first year of program.

class (target of 22-25 students), or a regular-size class with a full-time teacher's aide. The randomization was done within school for both students and teachers.[22] Each year about 6,800 students, 330 classes, and 79 schools in 46 districts participated in the program from 1985 to 1989. A single cohort of students who entered a participating school in the 1985-1986 school year participated in the experiment for grades in K-3. Thus, Project STAR experiment possess essential features of a controlled experiment designed to produce reliable evidence about the effects of reducing class size.[23]

Major findings in previous studies The evidence showed that the students in the smaller classes outperformed the students in the regular classes, whether or not the regular class teachers had a full-time aide helping them. In particular, the positive effect of smaller classes are greater for black and inner-city poor students. Moreover, students who had been in small class performed better even after treatment is over (Krueger and Whitmore [2001]).

2.6.2 Evidence of experimental violation

The validity of an experimental analysis depends on initial randomization and no differential attrition and refreshment across treatment and control groups. It is quite well supported by some evidences that initial class types were randomly assigned among students and teachers within school, but the Project STAR experiment suffers from implementation problems of

[22] All participating schools implemented at least one of each of the three types of classes in order to cancel out the possible influences coming from variations in the quality of the participating schools that might affect the quality of the classroom activity.

[23] See previous studies such as Word et al. [1990] and Krueger [1999] for more details of description on experiment design and implementation.

Table 2.25. Students in each class types for grades in K-3

	small	regular/aide	Total
Kindergarten			
	1900	4425	6325
Randomly assigned	1900	4425	6325
First grade			
Previously randomly assigned	1293	2867	4160
New entrant	384	1929	2313
Switchers	248	108	356
Switchers from previous years	248	108	356
Total	1925	4904	6829
Second grade			
Previously randomly assigned	1273	3402	4675
New entrant	366	1313	1679
Switchers	377	109	486
Switchers from previous years	192	47	239
Total	2016	4824	6840
Third grade			
Previously randomly assigned	1276	3567	4843
New entrant	373	908	1281
Switchers	525	153	678
Switchers from previous years	207	72	279
Total	2174	4628	6802

source: author's own calculation from the sample of Project STAR data.

noncompliance and differential missing across treatment and control groups.[24]

Noncompliance

Size of noncompliance sample There were sizable non-compliance of initial class type assignment of treatments and control. Among the complete samples of 24,275, about 2,200 samples were reassigned to different class types from initial assignment. For instance, table 2.25 shows the number of reassignments from small to other types of class or vice versa. For instance, first grade, out of 4,515 remaining students, 356 (7.9%) students switched class types from small to aide/regular or vice versa. For second and third grades, out of 5,049 and 5,413, remaining students 239 (4.9%) and 279 (5.2%) students switched class types, respectively. In sum, about 10 % of students were moved from one class type to another in a non-random manner. [25]

The solution to noncompliance If initial assignments of treatments and control were completely random, noncompliance problem can be overcome quite easily in linear panel data model by using Instrumental Variable(IV) method in which one uses initial treatments as IV for actual treatments.

Missing Data

Differential missing Ignoring missing causes problem in the estimation of MTE if missing is systematically different across treatment and control groups. Typically, attrition can

[24] [Krueger, 1999, p.502] for details of initial random assignment of class types for kindergartners.

[25] Krueger [1999] reported that most of these moves were due to student misbehavior, and were not typically the result of parental request for small class reassignment.

depend on some observed and unobserved characteristics and different characteristics across treatment and control groups can lead to differential attrition. Table 2.26 and 2.27 provide an indirect evidence for possible differential attrition across class types by comparing simple mean of SAT score across groups. Table 2.26 summarizes the frequency and fraction of missing for each class types and table 2.27 compares average SAT score for attritors and stayers of the program across class types. The difference of score between attritors and remainders is greater for students in regular classes and that in small classes.

Size of missing data There was sizable missing from the prior year's treatment and control groups. Table 2.27 shows that about 6,000 to 7,000 students for each grade, 1,809 students left the program in first grade, 1,608 students left the program in second grade, and 1,319 students left the program in third grade. For instance, among those students in the program at the beginning, about 45 percent of students left Project STAR program before finishing third grade. Overall, out of 11,601 students, about 48 % of all participant students either left the program before it ends or enter the program after kinder grade.

Table 2.26. Attrition and reassignments of class types (noncompliance)

A. Kindergarten to first grade					
Kindergarten	First grade				
	small	regular	regular with aide	leaving sample (All-K)	fraction of attrition
small	1293	60	48	380 (1,900)	0.20
regular	126	737	663	668 (2,194)	0.31
aide	122	761	706	642 (2,231)	0.29
new sample	384	1,026	903	2,313(new)/1,809(leaving)	
B. First grade to second grade					
First grade	Second grade				
	small	regular	regular with aide	leaving sample (All-G1)	fraction of attrition
small	1,435	23	24	443 (1,925)	0.23
regular	152	1,498	202	732 (2,584)	0.28
aide	40	115	1,560	650 (2,320)	0.28
new sample	389	693	709	1,791(new)/1,668(leaving)	
C. Second grade to third grade					
Second grade	Third grade				
	small	regular	regular with aide	leaving sample (All-G2)	fraction of attrition
small	1,564	37	35	380 (2,016)	0.188
regular	167	1,485	152	525 (2,329)	0.225
aide	40	76	1,857	522 (2,495)	0.209
new sample	403	487	481	1,371(new)/1,319(leaving)	

source: Project STAR and beyond: Database User's Guide(2007)

Table 2.27. SAT percentile scores at $t-1$ for Leavers and stayers

1st grader	Frequency	Mean	SE	Diff(small-reg/aide)
Stayer of small class at first grade	1,319	59.19	25.40	13.64
Leave small sample at first grade	453(25.6%)	45.55	28.22	
Stayer of regular or regular with aide class at first grade	2949	54.56	25.26	12.30
Leave regular or regular with aide class at first grade	1182(28.6%)	42.26	26.54	
Stayer of all class types at first grade	4267	55.99	25.40	12.81
Leaver of all class types at first grade	1635	43.18	27.05	
2nd grader	Frequency	Mean	SE	
Stayer of small class at second grade	1173	61.35	24.69	14.41
Leave small sample at second grade	276(19.0%)	46.94	27.13	
Stayer of regular or regular with aide class at second grade	2187	57.63	24.26	13.35
Leave regular or regular with aide class at second grade	631(22.4%)	44.28	25.17	
Stayer of all class types at second grade	3360	58.93	24.27	13.84
Leaver of all class types at second grade	907	45.09	25.80	

Note: Average score for the group are reported in mean column. Diff(small-reg/aide) column report the average score difference between stayers and leavers for each class type.

Basic statistics of missing data

Table 2.28 shows data used in the regression analysis for grades in K-3. Overall, 48 percent of data is missing and there are 16 different types of missing patterns appear in the data as in table 2.29. Among those students in the STAR program, about 5 participating students have missing variables as shown in table 2.28, but only about 52 percent of students have complete-case data. Therefore, there is severe attrition problem. Missing occurs in Project STAR due to students' attrition and missing variables of participating students.

Table 2.28. Data availability for covariates

	K	G1	G2	G3
Small class	6325	6829	6840	6802
Aide class	6325	6829	6840	6802
Regular class	6325	6829	6840	6802
White/Asian	11467	11467	11467	11467
Female	11581	11581	11581	11581
Free Lunch	6300	6650	6496	6520
Teacher white	6286	6810	6780	6737
Teacher experience	6304	6810	6739	6751
Teacher master degree and above	6304	6788	6744	6736
Fraction of free lunch in class	6325	6781	6645	6652
Fraction of kinder attendee in class	6325	6829	6840	6801
outcome (y_{it}) available	5907	6684	6559	6464
all covariates (x_{it}) available	6256	6584	6288	6428
Both y_{it} and all covariates (x_{it}) available ($\equiv A$)	5840	6430	5919	6086
Number of students in the program ($\equiv B$)	6325	6829	6840	6802
$\frac{A}{B}$	0.923	0.942	0.865	0.895

Table 2.29. Data availability by selection indicator

	K	G1	G2	G3	# of sample
case1	◯	◯	◯	◯	2418
case2	◯	◯	◯	×	513
case3	×	◯	◯	◯	1079
case4	◯	◯	×	×	898
case5	×	◯	◯	×	328
case6	×	×	◯	◯	935
case7	◯	×	×	×	1582
case8	×	◯	×	×	847
case9	×	×	◯	×	492
case10	×	×	×	◯	1154
case11	◯	×	◯	◯	101
case12	◯	×	◯	×	53
case13	◯	×	×	◯	52
case14	◯	◯	×	◯	223
case15	×	◯	×	◯	124
case16	×	×	×	×	802
Total					11601

(Note) ◯ implies $s_{it} = 1$ and × implies $s_{it} = 0$.

Table 2.30. Basic statistics: The mean and standard deviation

	K	G1	G2	G3
SAT score	52.44	52.82	52.40	51.85
	(26.49)	(27.46)	(26.99)	(27.17)
Small class	0.30	0.28	0.29	0.32
	(0.46)	(0.45)	(0.46)	(0.47)
Aide class	0.35	0.34	0.37	0.37
	(0.48)	(0.47)	(0.48)	(0.48)
Regular class	0.35	0.38	0.34	0.31
	(0.48)	(0.49)	(0.47)	(0.46)
White/Asian	0.67	0.67	0.65	0.67
	(0.47)	(0.47)	(0.48)	(0.47)
Female	0.49	0.48	0.48	0.48
	(0.50)	(0.50)	(0.50)	(0.50)
Free Lunch	0.48	0.52	0.51	0.51
	(0.50)	(0.50)	(0.50)	(0.50)
Teacher white	0.84	0.83	0.80	0.79
	(0.37)	(0.38)	(0.40)	(0.41)
Teacher experience	9.26	11.63	13.14	13.93
	(5.81)	(8.94)	(8.65)	(8.61)
Teacher master degree and above	0.34	0.34	0.36	0.43
	(0.47)	(0.47)	(0.48)	(0.49)
Fraction of free lunch in class	0.48	0.52	0.51	0.51
	(0.29)	(0.28)	(0.29)	(0.28)
Fraction of kinder attendee in class		0.66	0.54	0.47)
		(0.18)	(0.20)	(0.19)

(Note) Standard deviation is in parenthesis.

Selection indicator, s_{it} : Selection indicator s_{it} has value 1 only if both response variables and all covariates [small class (dummy), class with aide (dummy), student (race, gender, free lunch), teacher (race, experience, degree), class (fraction of free lunch, fraction of kindergarten attendees)] are observed, otherwise we assign s_{it} to be 0. And we also denote $T_i = \sum_{t=K}^{3} s_{it}$. Table 2.29 shows the number of students for each type of data based on using selection indicator s_{it}. Cases of data from 11 to 15 in table 2.29 are non-monotone II type missing so we assume MCAR of these missing patterns in the application of MI-IPW method.

2.6.3 The impact of CSR on SAT score for grades in K-3

We start by replicating complete-case estimations in previous studies under MCAR.

Model for the effect of CSR on SAT score

Using the model specification of Krueger [1999], we adopt linear panel data model as in (2.41).

Model Operationally ignoring missing in estimation, we can apply pooled LS or pooled reduced-form (IV) estimators. [26]

$$y_{it} = \mathbf{D}_{it} \cdot \beta + \mathbf{x}_{1it}\delta + \mathbf{x}_{2i}\gamma + f_t + c_i + u_{it} \tag{2.41}$$

$$\begin{aligned} i &= \text{student, } t = K, G1, G2, G3, \text{ where } \mathbf{D}_{it} \text{ : two treatments} \\ y_{it} &= \mathbf{x}_{it}\theta + c_i + u_{it}, \text{ where } \mathbf{x}_{it} = (\mathbf{D}_{it}\ \mathbf{x}_{1it}\ \mathbf{x}_{2i}\ f_t),\ \theta = (\beta'\ \delta'\ \gamma', 1)' \end{aligned}$$

[26] A reduced-form estimation uses initial treatments and control as explanatory variables in model specification.

where y_{it} is student academic achievement ,

$\mathbf{D}_{it}$ is vector of class type assignments- small and regular with aide,

Observed covariates include student, teacher, peer characteristics

$\mathbf{x}_{1it}$ time varying observable covariates including school dummy,

$\mathbf{x}_{2i}$ time-fixed observable covariates,

c_i unobserved heterogeneity, f_t time effect, and u_{it} unobserved errors

y (Achievement Score) Achievement score we use is Standard Achievement Test (SAT). It measures student achievement in reading, math, and word skills in grade K-3. Following Krueger [1999], we scaled the test score into percentile ranks. Percentile ranking was calculated using the scores for those students who are assigned to regular and regular with full-time aide class types only. And three separate percentile rankings for each subject were generated. Response variable y is obtained as average of these three percentile rankings. For those students who are not available all three percentile rankings, we use the average percentile ranking of available percentile rankings. If one of three subject scores is missing, we use only two subject test scores to obtain y and if two subject scores are missing, then we use only available one subject's test score for y. [27]

Covariates x include: (i) student characteristics: race, gender, free lunch status (ii) teacher characteristics: years of experience, highest degree dummy, race (iii) peer characteristics: fraction of classmates with free lunch, fraction of female classmates, fraction of

[27] Project STAR Public Access Data is obtained from www.heros-inc.org/data.html. All the following descriptions of variables used in this study are based on Project STAR and Beyond: Database User's Guide (2007).

classmates of kindergarten attendees, and fraction of minority classmates, and (iv) school variables: school indicator, school location indicator.

Unweighted estimations Results

Cross-section estimation Table 2.31 reports unweighted cross-section estimates for both LS and IV. The estimated cross-section return of small class is about 5 to 7 percentile points for grades in K-3. It is about $\frac{1}{5}$ of standard deviation of SAT score. The estimated cross-section return of the class with full-time aide is not statistically different from 0 for grades in K-3.

Table 2.31. Cross-section unweighted LS and reduced-form estimates

	OLS				reduced-form			
	K	G1	G2	G3	K	G1	G2	G3
Small	5.205	7.683	5.821	4.964	5.205	6.621	5.279	5.247
	(.749)	(.740)	(.776)	(0.822)	(.749)	(.758)	(.786)	(.814)
Aide	.259	1.866	1.642	-.492	.259	1.509	1.175	-.072
	(.698)	(.710)	(.733)	(.770)	(.698)	(.679)	(.700)	(.726)
R-squared	.309	.299	.278	.221	.309	.295	.276	.220
# of sample	5840	6430	5919	6086	5840	6430	5919	6086
Other covariates	School dummy, student and teacher characteristics				School dummy, student and teacher characteristics			
	OLS				reduced-form			
	K	G1	G2	G3	K	G1	G2	G3
Small	5.224	7.599	5.440	4.644	5.224	6.490	5.064	4.979
	(.749)	(.747)	(.782)	(0.836)	(.749)	(.767)	(.786)	(.817)
Aide	.434	2.012	1.445	-.791	.434	1.560	1.156	-.078
	(.698)	(.716)	(.738)	(.777)	(.698)	(.680)	(.699)	(.728)
R-squared	.311	.300	.280	.222	.311	.296	.279	.221
# of sample	5840	6430	5919	6086	5840	6430	5919	6086
Other covariates	School dummy, student, teacher peer(classmates) characteristics				School dummy, student, teacher peer(classmates) characteristics			

(Note) Students cluster-robust standard errors are provided in parentheses. Student characteristics include race, gender, and free lunch status and teacher characteristics include race, years of experience and highest degree dummy. Peer characteristics include the fraction of classmates with free lunch, fraction of minority classmates and fraction of female classmates. Reduced-form estimation use initial assignments of treatments as independent variables.

Pooled estimation of LS and FD estimation Table 2.33 reports unweighted FD estimates for the return of CSR on SAT score. First, the FD estimates for the return of CSR are about 6.4 percent ($1.6 \cdot 4 = 6.4$) while the return of CSR for unweighted pooled LS estimate are about 7.5 percent in table 2.32. Second, unweighted FD estimates imply that the class with full-time aide has statistically significant impact on the improvement of SAT score when compared to regular class without full-time aide while pooled estimates for the return of class with full-time aide are not statistically different from zero. Overall, with unweighted methods, the effect of CSR on SAT score is between 5(pooled IV) and 7.5(pooled LS) percents.

Table 2.32. Unweighted pooled LS and reduced-form

	LS: full sample		IV: full sample		LS: attrition sample		IV: attrition sample	
Initial small	3.243	3.201	4.991	4.903	3.474	3.441	4.856	4.806
	(.806)	(.806)	(.507)	(.507)	(.980)	(.980)	(.698)	(.699)
Initial aide	.495	.518	.782	.796	-.022	-.004	.205	.244
	(.671)	(.670)	(.507)	(.507)	(.797)	(.797)	(.663)	(.663)
Cumulative years in small	1.003	.978			.746	.743		
	(.388)	(.388)			(.436)	(.436)		
Cumulative years in aide	.262	.254			.221	.241		
	(.376)	(.377)			(.426)	(.427)		
R-squared	.228	.229	.226	.227	.227	.229	.226	.227
# of sample	24,275	24,275	24,275	24275	16,222	16,222	16,222	16,222
Student characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
teacher characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
peer characteristics	No	Yes	No	Yes	No	Yes	No	Yes
School, grade,PP-year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

(Note) Students cluster-robust standard errors are provided in parentheses. PP-year implies program participation year dummy. Grade implies dummies for grades in K-3. IV represents the estimation method for which initial assignments of class type were used as treatment variables.

Table 2.33. Unweighted FD estimates

	FD with full sample			FD with attrition sample		
Cumulative years in small	1.738	1.680	1.651	1.827	1.641	1.613
	(.362)	(.364)	(.364)	(.416)	(.413)	(.414)
Cumulative years in aide	1.492	1.370	1.359	1.738	1.692	1.689
	(.358)	(.363)	(.364)	(.426)	(.427)	(.428)
R-squared	.016	.020	.020	.015	.019	.020
# of sample	13214	12924	12924	9883	9793	9793
Student characteristics	Yes	Yes	Yes	Yes	Yes	Yes
teacher characteristics	No	Yes	Yes	No	Yes	Yes
peer characteristics	No	No	Yes	No	No	Yes
School, grade,PP-year	Yes	Yes	Yes	Yes	Yes	Yes

(Note) Students cluster-robust standard errors are provided in parentheses. Grade implies dummies for grades in K-3.

2.6.4 Test of strict exogeneity assumption

We formally test strict exogeneity using Hausman-type test which is based on unweighted FD and FE estimators. We perform Hausman-type II test and regression based variable addition test of Wooldridge [2009] for unbalanced panel data.

Test Results

Table 2.34 shows the rejection of the null hypothesis for strict exogeneity at nominal level of 0.05. This implies that the violation of strict exogeneity and unweighted FE and FD estimators with complete-case would produce bias and incorrect standard error. Furthermore, table 2.35 shows the results for a regression based variable addition tests and $H_0 : \lambda = 0$ in (2.19)-(2.21) are rejected for FD and FE estimation models. Tests imply MCAR is violated and, therefore, both unweighted FD and FE estimators are not valid.

Indirect evidence of weak exogeneity from LS and reduced-form(IV) Unfortunately, we cannot directly test conditional contemporaneous exogeneity(i.e. weak exogeneity conditional on selection $\mathbf{E}(s_{it} \cdot d_{it} \cdot error_{it})$ =0) for pooled LS and reduced-form(IV) estimators. However, if we assume that $E(s_{it}|\cdot)$ and $E(s_{it+1}|\cdot)$ are serially correlated, we can conjecture the sign of $\mathbf{E}(s_{it} \cdot d_{it} \cdot error_{it})$ from the sign of $\mathbf{E}(s_{it+1} \cdot d_{it} \cdot error_{it})$. LS and IV columns in table 2.35 report that $\mathbf{E}(s_{it+1} \cdot d_{it} \cdot error_{it})$ >0 and, thus for instance, if both $\mathbf{E}(s_{it} \cdot d_{it} \cdot u_{it})$ and $\mathbf{E}(s_{it+1} \cdot d_{it} \cdot error_{it})$ have the same sign, both LS and IV estimators are biased upward.

Table 2.34. Hausman-type test II: strict exogeneity

	FD	FE	FD	FE	FD	FE	FD	FE
Cumulative year in	1.569	.544	1.741	.477	1.682	.458	1.648	.442
small class	(.391)	(.321)	(.403)	(.324)	(.363)	(.326)	(.363)	(.327)
Cumulative year in	1.514	.809	1.501	.774	1.377	.705	1.360	.692
aide class	(.386)	(.319)	(.400)	(.322)	(.361)	(.323)	(.362)	(.324)
# of sample	31228		30309		29960		29960	
$H_0 : \widehat{\theta}_{pols} = \widehat{\theta}_{FE}$, $k = 2$	$\widehat{W}_2 = 4.82$		$\widehat{W}_2 = 6.70$		$\widehat{W}_2 = 6.37$		$\widehat{W}_2 = 6.16$	
P-values	{.089}		{.035}		{.041}		{.045}	
Student characteristics	No		Yes		Yes		Yes	
teacher characteristics	No		No		Yes		Yes	
peer characteristics	No		No		No		Yes	
School, grade,PP-year	Yes		Yes		Yes		Yes	

Table 2.35. Variable addition test: strict exogeneity

	PLS	PIV	FE	FD	PLS	PIV	FE	FD
Selection (S_{it+1})	10.439	10.474	4.571	3.889	10.975	10.287	4.185	3.604
	(.420)	(.420)	(.514)	(.544)	(.418)	(.419)	(.510)	(.540)
Initial small	2.702	5.047			2.848	4.113		
	(.919)	(.575)			(.923)	(.598)		
Initial aide	.076	.922			.185	.886		
	(.749)	(.513)			(.753)	(.516)		
Cumulative year in	1.602		1.276	2.054	.951		1.249	2.084
small class	(.522)		(.474)	(.491)	(.545)		(.478)	(0.494)
Cumulative year in	.802		1.866	2.085	.668		1.756	2.110
aide class	(.510)		(.468)	(.492)	(.516)		(.471)	(.498)
R-squared	.269	.269		.037	.272	.272		.037
# of sample	18386	18386	18386	8538	18189	18189	18189	8390
Student characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
teacher characteristics	No	No	No	No	Yes	Yes	Yes	Yes
peer characteristics	No	No	No	No	Yes	Yes	Yes	Yes
School, grade,PP-year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

2.6.5 Estimation with inverse probability weighted(IPW) and multiple imputation(MI)

The two formal tests in previous section show the evidence of differential correlation among unobserved heterogeneity, c_i, and idiosyncratic error, $\mathbf{u}_i = (u_{i1}, u_{i2}, ..., u_{iT})$ and selection indicator, $\mathbf{s}_i$ across d_{it} and this implies that MCAR is violated. Therefore, we apply IPW and MI-IPW methods using the probability of selection which satisfies MAR assumption as weights.

IPW estimation

In the estimation of the probability of selection (i.e. $\mathbf{E}[s_{it}|\mathbf{W}_{it-1}]$), we adopt the method in Robins et al. [1994] and Wooldridge [2010] for which the probability of selection is estimated in sequential manner for unbalanced panel data. We base our choice of the predictors for the probability of selection on parents's school choice between private and public. At the beginning of period, all students start in public school.

Parents's school choice and attrition

We use parents's school choice model in Becker [1993] to describe students attrition in STAR Program. Using this model, we specify the model for the probability of selection which satisfies MAR assumption. Estimated probability of selection is used as inverse probability weights in IPW estimation.

Assumptions

1. Parents maximize utility directly for two periods and indirectly infinite periods

2. heterogeneity takes two forms which are parent resource (h_{1i}) and student individual ability (h_{2i})

3. Each period choice is made for the resource allocation among (i) consumption (ii) spending on human capital of children (E_t)

4. Initially, as all students start in public school, we assume that the marginal cost of moving to private school is greater than the marginal benefit.

Model: Parent's optimization problem Following Becker [1993], at time t, parent's utility depends both on the utility of themselves and their children as in (2.42).

$$U_t = u(c_t) + \rho \cdot u(c_{t+1}) \tag{2.42}$$

where c_t is the consumption of parent, ρ is a constant which measures the relative weight for child's utility, c_{t+1} is children's consumption when they become adult.

If the preference (2.42) is the same for all generations and consumption during childhood can be ignored, then the utility of parent can be written as in (2.43).

$$U_t = \sum_{i=1}^{\infty} \rho^i \cdot u(c_{t+i}). \tag{2.43}$$

The utility of parent depends directly only on the utility of their own children and indirectly on all descendants.

Resource constraint for parent and children at period t is given by

$$y_t = c_t + E_t;\ y_t = h_{1i};\ y_{t+1} = f(h_{2i}, E_t); \text{with linearity } y_{t+1} = a(h_{2i}) \cdot E_t \tag{2.44}$$

where h_{1i} is parent's income, E_t parents' human capital investment for their children at t and $a > 0$ is the marginal return for human capital investment which varies across students as

their ability varies. For the sake of simplicity of illustration, we assume a linear production function for children in (2.44). [28]

At the optimum, with linearity assumption on children's production function, we have the following optimal condition.

$$u'(h_{1i} - E_t) = \rho \cdot \mathbf{E}_t(a(h_{2i}) \cdot u'(c_{t+1})|I_t) \tag{2.45}$$

where $\mathbf{E}_t(\cdot|I_t)$ is conditional expectation at period t using all available information, I_t, at the beginning of period t. LHS of (2.45) is marginal cost of moving to private school and RHS of (2.45) is expected marginal benefit of moving to private school. Parent does not have complete information on their children's ability so they update their children's ability as they obtain new information. At the beginning of program, for all participating students, LHS is greater than RHS in (2.45). At the end of each grade, parents update their expected marginal benefit of moving to private school as new information on student's score, class, teacher, and school quality become available. Therefore, attrition by school choice occurs as RHS becomes larger than LHS in (2.45) after new information arrives at the end of period.

There are other reasons for students move and leave STAR program such as parent job reallocation, the move due to sibling concerns, and others. We assume that the attrition of students by these reasons should not be causing differential attrition across class types since it is very hard to think that these reasons to move only occur to one type of class. Therefore, we assume that differential attrition across class types only occurs through parent's decision of the moving.

[28] We assume no transfer from parents to children other than human capital investment and bequest transfer. Bequest transfer is determined at the level which expected marginal returns from bequest transfer is equal to marginal return from human capital investment.

The probability of selection model The attrition of student is determined by updated information $\mathbf{E}_t(\cdot|I_t)$ between at the end of $t-1$ and at the beginning of t. I_t includes student's score (y_{it-1}) and updated class, teacher, and school characteristics($\mathbf{z}_{1it-1}$) where $\mathbf{z}_{1it-1}$ include class average on SAT score, teacher's experience, teacher's highest degree, the fraction of free lunch students and so on. In the application, we use y_{it-1} and ($\mathbf{z}_{1it-1}$) as predictors for the probability of selection model. In this setting, MAR is satisfied as in (2.46) and MAR implies that, once we know parent information I_t, we can figure out parent's school choice and attrition status. We use logit model in the application.

$$P(s_{it}=1|\mathbf{W}_{it}) = P(s_{it}=1|\mathbf{z}_{it})P(s_{it}=1|\mathbf{W}_{it}, s_{it-1}=1) = P(s_{it}=1|\mathbf{z}_{it}, s_{it-1}) = \pi_{it} \tag{2.46}$$

where $\mathbf{z}_{it} = (\mathbf{z}_{1it-1}, y_{it-1})$.

$$\pi_{it} = G(\mathbf{z}_{it};\gamma)\Lambda(\mathbf{z}_{it}\cdot\gamma) \tag{2.47}$$

where $\mathbf{w}_{it}$ =($\mathbf{x}_{it-1}, y_{it-1}$) and $\Lambda(\cdot) = \frac{exp(\cdot)}{1+exp(\cdot)}$.

2.6.6 IPW estimation

We illustrate IPW estimation with complete-case. MI-IPW estimation is conducted with pseudo-complete case data.[29]

We construct p_{it} in sequential manner using conditional expectation.

$$p_{it} = p(s_{it}=1|\mathbf{z}_{it}, s_{it-1}=1)\cdot p(s_{it-1}=1|\mathbf{z}_{it-1}, s_{it-2}=1)\cdots p(s_{i2}=1|\mathbf{z}_{i2}, s_{i1}=1),\ t=1,2,3 \tag{2.48}$$

[29] Pseudo-complete case is the case that combines complete case and non-monotone II which become complete case after imputations.

$$p_{it} = \pi_{it} \cdot \pi_{it-1} \cdots \pi_{i1}$$

Using the sequential approach as in (2.48), and assuming the sequence of binary response models is correctly specified as in (B.1), we can consistently $\widehat{\gamma}$. Once we obtain $\widehat{\gamma}$ from MLE, we use $G(\mathbf{z}_{it}, \widehat{\gamma})$ to construct $\widehat{p_{it}}$ for all i and t with $s_{it} = 1$.

IPW estimates: The return of CSR on SAT score

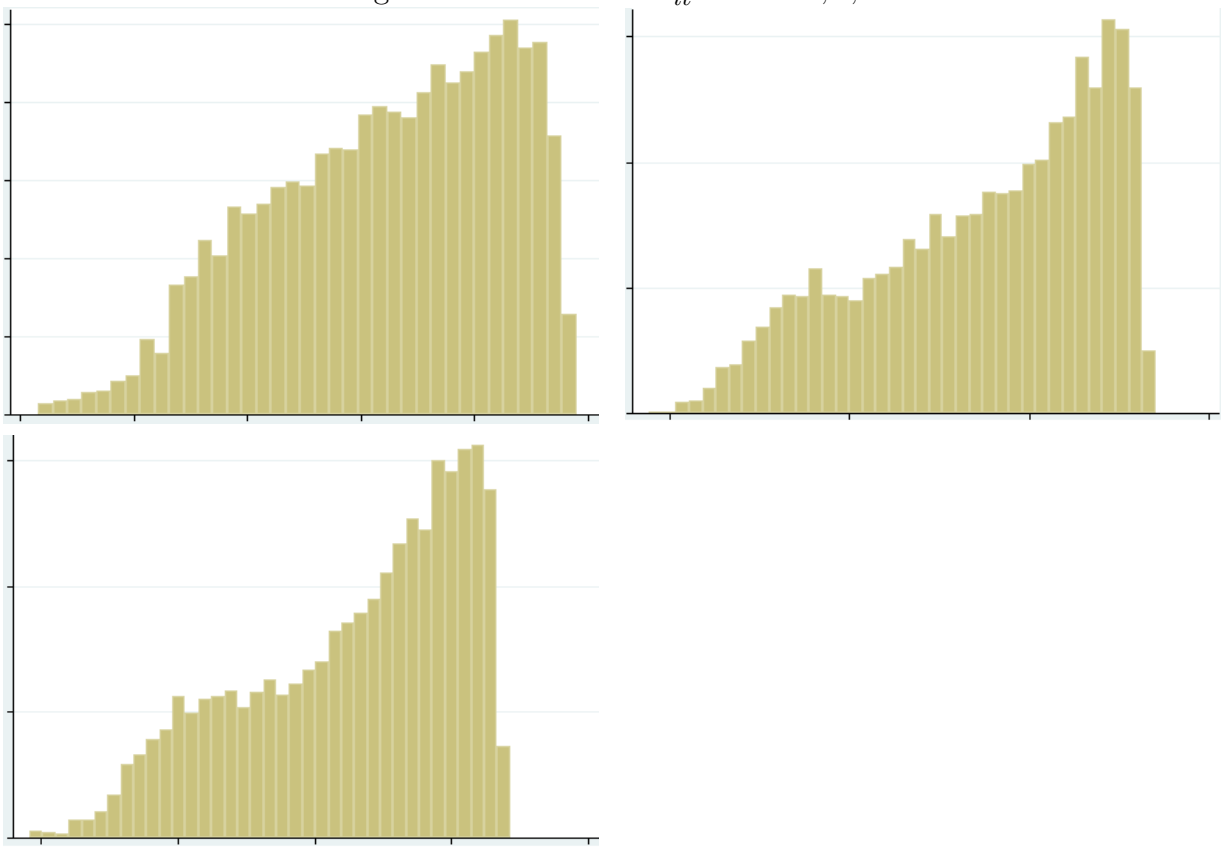
For cross-section estimation, we estimate $p_{it} = \pi_{it} = p(s_{it} = 1|\mathbf{z}_{it}, s_{it-1} = 1)$ using logit model. $\mathbf{z}_{it-1}$ include class types, student characteristics (gender, race, free lunch), teacher characteristics (race, experiences, highest degree), school fixed effects, and SAT percentile score of period $t - 1$.

Cross-section estimation of probability weight Figure 2.3 show estimated probability weight for cross-section estimation in grades K-3. One of practical criticism for IPW estimation is that estimates can be very unstable if certain subsets of population have very low non-attrition probabilities.(Little and Rubin [2002]) However, as wee see in figure 2.3, all of estimated $\widehat{p}_{it}$ is strictly greater than 0.4 so that our assumption of $p_{it} > 0$ is well satisfied. The overlap assumption of $\widehat{p}_{it}$ for treatment and control groups is also well satisfied as we see in figure 2.4.

IPW estimates: cross-section Table 2.36 reports cross-section IPW estimates of both LS and reduced-form for grades in 1-3. The estimated cross-section return of small class is about 5 to 7 percentile points. Students in small classes obtain about 5 to 7 percentile

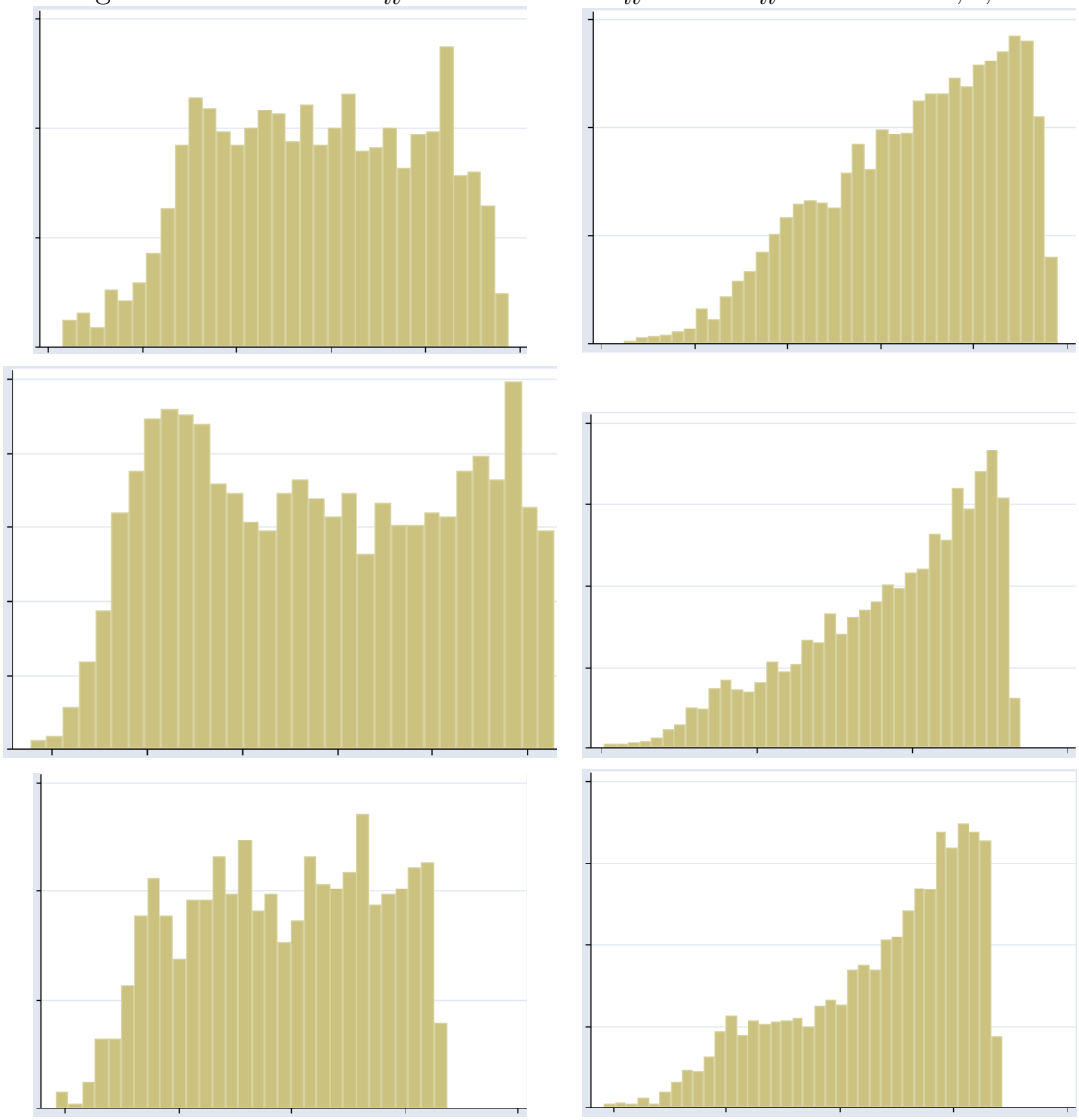
point higher test scores than students in regular class for grades in K-G3. It is about $\frac{1}{5}$ of standard deviation of SAT score. The estimated cross-section return of full-time aide class is not statistically different from 0. The implication of IPW estimates is no different from the one of unweighted estimates. In all grades, the difference between unweighted and IPW estimates is within 1 percent percentile.

Figure 2.3. Estimated π_{it} for $t = 1, 2, 3$



(Note) For all figures, x-axis represents $\widehat{\pi_{it}}$ and y-axis represents fraction. For the figure in first row and first column, $\widehat{\pi_{i1}}$ in x-axis is ranged from 0.4, to 0.9 and its unit interval is .1. y-axis is ranged from 0 to 0.05 and its unit interval is 0.01. For the figure in first row and second column, $\widehat{\pi_{i2}}$ in x-axis is ranged from 0.4, to 1 and its unit interval is 0.2. y-axis is ranged from 0 to 0.06 and its unit interval is 0.02. For the figure in second row and first column, $\widehat{\pi_{i3}}$ in x-axis is ranged from 0.6, to 1 and its unit interval is 0.1. y-axis is ranged from 0 to 0.06 and its unit interval is 0.02.

Figure 2.4. Estimated π_{it} conditional on $s_{it} = 0$ or $s_{it} = 1$ for $t = 1, 2, 3$



(Note) x-axis represents $\widehat{\pi_{it}}$ and y-axis represents fraction. We denote figure in ith row and jth column as $F_{(i,j)}$. $F_{(i,1)}$ shows $\widehat{\pi}_{it}$ conditional on $s_{it} = 0$ and $F_{(i,2)}$ shows $\widehat{\pi}_{it}$ conditional on $s_{it} = 1$. $F_{(1,1)}$ shows $\widehat{\pi}_{i1}$ conditional on $s_{i1} = 0$, x-asix is ranged from 0.4 to 0.9 and y-asix is ranged from 0 to 0.06 with interval of 0.02. $F_{(2,1)}$ shows $\widehat{\pi}_{i2}$ conditional on $s_{i2} = 0$, x-asix is ranged from 0.4 to 0.9 and y-asix is ranged from 0 to 0.05 with interval of 0.01. $F_{(3,1)}$ shows $\widehat{\pi}_{i3}$ conditional on $s_{i3} = 0$, x-asix is ranged from 0.6 to 1 and y-asix is ranged from 0 to 0.06 with interval of 0.02. $F_{(1,2)}$ shows $\widehat{\pi}_{i1}$ conditional on $s_{i1} = 1$, x-asix is ranged from 0.4 to 0.9 and y-asix is ranged from 0 to 0.06 with interval of 0.02. $F_{(2,2)}$ shows $\widehat{\pi}_{i2}$ conditional on $s_{i2} = 1$, x-asix is ranged from 0.4 to 1 and y-asix is ranged from 0 to 0.08 with interval of 0.02. $F_{(3,1)}$ shows $\widehat{\pi}_{i3}$ conditional on $s_{i3} = 1$, x-asix is ranged from 0.6 to 1 and y-asix is ranged from 0 to 0.08 with interval of 0.02.

Table 2.36. Cross-section IPW LS and IPW reduced-form estimation

	LS				reduced-form			
	K	G1	G2	G3	K	G1	G2	G3
Small	5.205	7.309	6.025	5.491	5.205	5.613	5.536	6.224
	(.749)	(.926)	(.928)	(0.953)	(.749)	(.960)	(.953)	(.945)
Aide	.259	1.090	2.375	.289	.259	.013	1.480	.820
	(.698)	(.950)	(.913)	(.904)	(.698)	(.898)	(.848)	(.833)
R-squared	.309	.318	.294	.234	.309	.313	.293	.235
# of sample	5840	4052	4338	4533	5840	4052	4338	4533
Other covariates	School dummy, student and teacher characteristics				School dummy, student and teacher characteristics			
	OLS				reduced-form			
	K	G1	G2	G3	K	G1	G2	G3
Small	5.224	7.282	5.595	5.355	5.224	5.508	5.312	6.072
	(.749)	(.935)	(.940)	(0.969)	(.749)	(.973)	(.955)	(.949)
Aide	.434	1.442	2.235	.092	.434	.076	1.544	.728
	(.698)	(.955)	(.921)	(.910)	(.698)	(.897)	(.847)	(.834)
R-squared	.311	.320	.296	.235	.311	.316	.293	.235
# of sample	5840	4052	4338	4533	5840	4052	4338	4533
Other covariates	School dummy, student, teacher peer(classmates) characteristics				School dummy, student, teacher peer(classmates) characteristics			

(Note) Students cluster-robust standard errors are reported in parentheses. Student characteristics include race, gender, and free lunch status and teacher characteristics include race, years of experience and highest degree dummy. Peer characteristics include the fraction of classmates with free lunch, fraction of minority classmates and fraction of female classmates. Reduced form estimation use initial assignments of treatments as independent variables.

Panel probability weights estimation $\widehat{\pi_{it}}$ is estimated using equation (2.49) by conditional MLE.

$$\pi_{it} = \Lambda(z_{it} \cdot \gamma) \text{ for } s_{it-1} = 1. \tag{2.49}$$

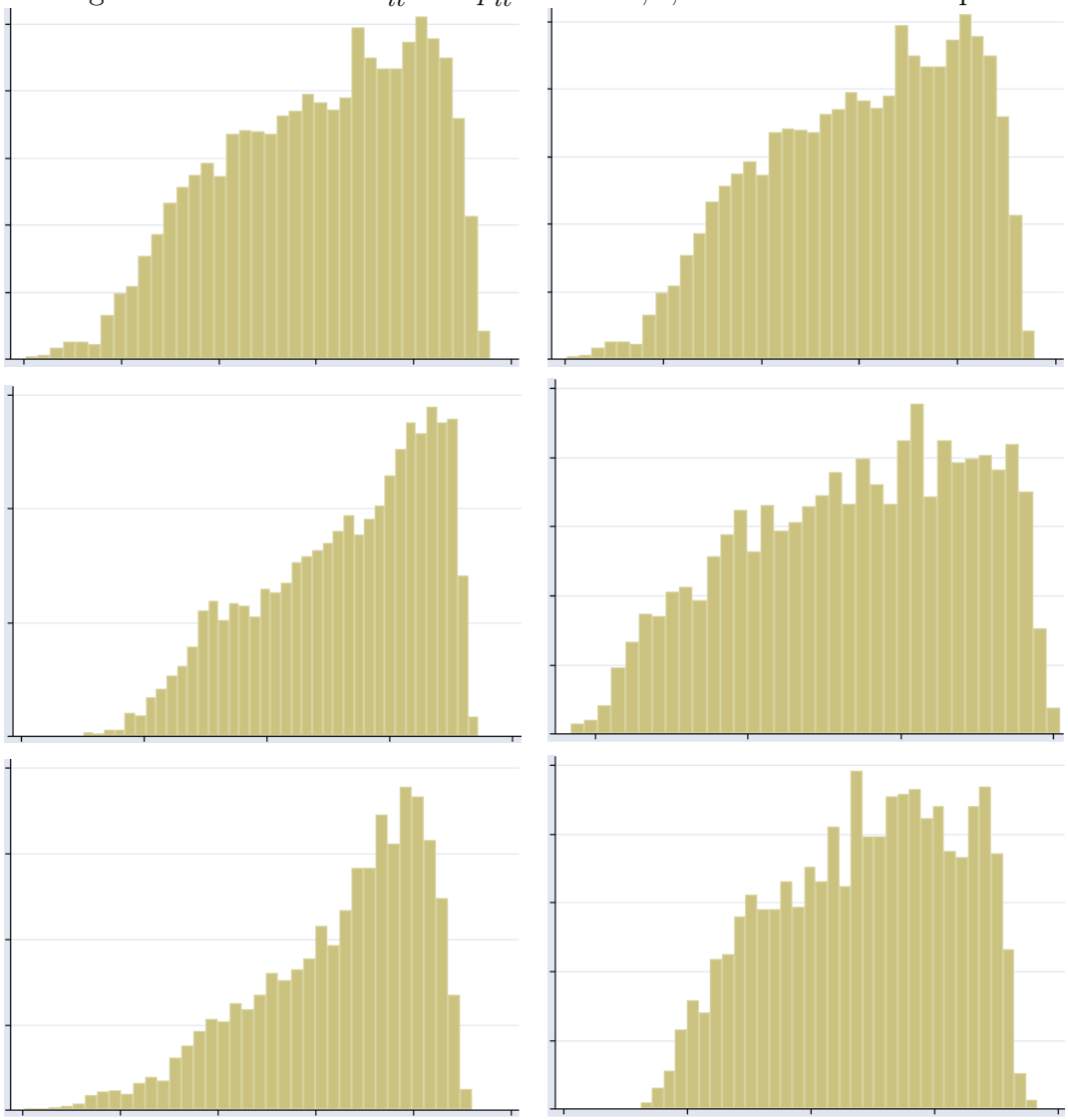
where $j = 1, 2, 3$.

The probability weight estimate, $\widehat{p_{it}}$ is sequentially constructed as in (2.50).

$$\widehat{p_{it}} = \widehat{\pi_{it}} \cdot \widehat{\pi_{it-1}} \cdot \cdots \cdot \widehat{\pi_{i1}}. \tag{2.50}$$

Figure 2.5 shows the histogram of estimated probability weight, $\widehat{\pi_{it}}$ and $\widehat{p_{it}}$ $\widehat{\pi_{it}}$ and $\widehat{p_{it}}$ are greater than .15 and .35 for all i and t respectively.

Figure 2.5. Estimated π_{it} and p_{it} for $t = 1, 2, 3$ with attrition sample



(Note) $F_{(i,j)}$ identifies the location of figure. i identifies the row and j identifies column. Figures in left panel of rows represent $\widehat{\pi_{it}}$. First, second, and third row show figures for $\widehat{\pi_{i1}}$, $\widehat{\pi_{i2}}$, and $\widehat{\pi_{i3}}$ respectively. Figures in right panel of rows represent $\widehat{\pi_{it}}$. First, second, and third row show figures for $\widehat{p_{i1}}$, $\widehat{p_{i2}}$, and $\widehat{p_{i3}}$ respectively. For $\widehat{\pi_{i1}}$ in $F_{(1,1)}$, x-axis is ranged from 0.4 to 0.9 and y-axis is ranged from 0 to 0.05 with interval of 0.01. For $\widehat{\pi_{i2}}$ in $F_{(2,1)}$, x-axis is ranged from 0.2 to 1 and y-axis is ranged from 0 to 0.06 with interval of 0.02. For $\widehat{\pi_{i3}}$ in $F_{(3,1)}$, x-axis is ranged from 0.5 to 1 and y-axis is ranged from 0 to 0.08 with interval of 0.02. For $\widehat{p_{i1}}$ in $F_{(1,2)}$, x-axis is ranged from 0.4 to 0.9 and y-axis is ranged from 0 to 0.05 with interval of 0.01. For $\widehat{p_{i2}}$ in $F_{(2,2)}$, x-axis is ranged from 0.2 to .8 and y-axis is ranged from 0 to 0.05 with interval of 0.01. For $\widehat{\pi_{i3}}$ in $F_{(3,2)}$, x-axis is ranged from 0 to .8 and y-axis is ranged from 0 to 0.05 with interval of 0.01.

Panel IPW estimation results Table 2.37 provides IPW pooled estimates. The estimated effects of IPW LS for the effect of CSR on SAT score is about 6.4 % while the estimated effect of unweighted LS is 7.3 % in table 2.32. Moreover, the estimated effects of IPW IV for the effect of CSR on SAT score is about 4.5 % while the estimated effect of unweighted IV is 5.5 %. Table 2.38 reports IPW FD estimates. The estimated effects of IPW FD estimator is 5.5 % while the effect for unweighted FD estimates is about 7 %. Again, the effect is large for the unweighted estimator. Thus, for all three estimates, unweighted estimates for the return of CSR on score are greater than IPW estimates.

Summary of estimation results: Unweighted and IPW estimates in unbalanced panel data

Unweighted estimates are greater than IPW estimates for pooled LS, pooled reduced-form(IV), FD estimators. However, the difference of estimates is very small for pooled LS, pooled reduced-form, and FD estimator and is within 2%. If IPW estimates are consistent, unweighted LS, reduced-form, and FD estimators overestimate the return of CSR about 1 to 2 percent. In sum, we conclude that the return of CSR on SAT score for grades in K-3 is within the range from 4 to 6.5 percent.

Table 2.37. IPW pooled LS, pooled IV

	Pooled LS			Pooled IV		
Initial small	2.152	2.374	2.328	4.500	4.497	4.448
	(1.163)	(1.163)	(1.163)	(.813)	(.815)	(.815)
Initial aide	-.601	-.414	-.390	-.221	-.123	-.074
	(.897)	(.905)	(.903)	(.762)	(.768)	(.768)
Cumulative years in small	1.154	1.033	1.038			
	(.506)	(.508)	(.508)			
Cumulative years in aide	.360	.275	.297			
	(.471)	(.476)	(.476)			
R-squared	.228	.229	.231	.228	.229	.230
# of sample	15395	15241	15241	15395	15241	15241
Student characteristics	Yes	Yes	Yes	Yes	Yes	Yes
teacher characteristics	No	Yes	Yes	No	Yes	Yes
peer characteristics	No	No	Yes	No	No	Yes
School, grade,PP-year	Yes	Yes	Yes	Yes	Yes	Yes

Table 2.38. Weighted FD estimates

	IPW FD			unweighted FD		
Cumulative years in small	1.317	1.380	1.361	1.827	1.641	1.613
	(.415)	(.416)	(.416)	(.416)	(.413)	(.414)
Cumulative years in aide	1.697	1.572	1.572	1.738	1.692	1.689
	(.433)	(.435)	(.435)	(.426)	(.427)	(.428)
R-squared	.014	.018	.019	.015	.019	.020
# of sample	9512	9401	9401	9883	9793	9793
Student characteristics	Yes	Yes	Yes	Yes	Yes	Yes
teacher characteristics	No	Yes	Yes	No	Yes	Yes
peer characteristics	No	No	Yes	No	No	Yes
School, grade,PP-year	Yes	Yes	Yes	Yes	Yes	Yes

2.6.7 MI-IPW method

The major disadvantage of imputation method in our application is the necessity of modeling the joint distribution of missing variables which include both categorical and continuous variables. For instance, our application of project STAR data includes score (continuous), free lunch (binary), teacher experience (continuous), teacher race (binary), many other class peer (fractional variables with having values between zero and one). However, we mitigate this problem of high dimensional covariates by focusing on non-monotone II type of missing in imputation. It is easy to implement MI method since built-in command for MI method exists with Stata, SAS and R.[30]

Models of missing variables with multiple imputation

Participating students with missing observations are about 5 percent at each grade and missing observations mostly occur to SAT score. Therefore, most of missing data comes from attrition of students for Project STAR data.

Imputation model for missing variables using MICE In Stata, we imputed missing variables which include outcome, treatments and covariates using following predictors and distributional assumption in table 2.39.

[30] We implement multiple imputation for the estimation of the return of CSR on score using "ice" Stata command with the simplest assumption on the distribution of missing variables. Further details on imputation methods under MAR have been studied extensively in theory and practice(Rubin [1987] and Little and Rubin [2002])

Table 2.39. Covariates and distributional assumption used for MICE

variables	distribution	command	prediction equation
SAT score	Normal	regress	treatments, student, teacher, peer characteristics
small class	Logistic	logit	SAT score, student, teacher, peer characteristics
aide class	Logistic	logit	SAT score, student, teacher, peer characteristics
free lunch	Logistic	logit	SAT score, treatments, student, teacher, peer characteristics
teacher's race	Logistic	logit	SAT score, treatments, student, teacher, peer characteristics
teacher's highest degree	Logistic	logit	SAT score, treatments, student, teacher, peer characteristics
teacher's years of experience	Logistic	logit	SAT score, treatments, student, teacher, peer characteristics
fraction of students with freelunch	Normal	regress	SAT score, treatments, student, teacher, peer characteristics
fraction of female students	Normal	regress	SAT score, treatments, student, teacher, peer characteristics
fraction of black students	Normal	regress	SAT score, treatments, student, teacher, peer characteristics

(Note) (i) Treatments are class-type indicators. (ii) Students characteristics include race, sex and free lunch status. (iii) Teacher characteristics include race, degree, and experience. (iv) Peer characteristics include the fraction of student with free lunch, fraction of female and fraction of black classmates. For panel data MI estimation, grade, program participation grade and school dummies are additionally included in predictions of all missing variables.

Estimates with MI-IPW

Using the set of imputed variables for STAR participating students as complete case, we apply IPW estimation.

Tables 2.40 and 2.41 reports pooled LS, pooled IV and pooled FD estimates for the return of CSR on score using MI-IPW method. There are about 4.5 percent more sample observations are used in MI-IPW method than IPW method. For instance, in the case of pooled LS estimator, MI-IPW method uses 17,468 observations while IPW method 15,395 observations in the estimation. The rest of missing data which is more than 40 percent of full-sample is due to attrition. The difference of estimates between IPW and MI-IPW is very small while standard errors are small for MI-IPW method so that MI-IPW estimates provide the same implications as those of IPW estimates. In sum, unweighted estimator overestimates CSR on SAT score about 1 to 2 percentage points.

Summary of results

The return of CSR on SAT score is about 4 to 6.5 percent for MI-IPW estimator while it is about 5 to 7.5 percent for unweighted estimator. Unweighted estimator overestimates the effect of CSR by about 1 to 2 percentage point. The return of aid class on SAT score is not statistically significant for pooled LS and IV estimates while it is statistically significant for FD estimator. The statistical significance of the return for both small class and aide class does not change whether we use either unweighted or IPW (MI-IPW) methods.

Table 2.40. MI-IPW pooled LS, pooled IV

	Pooled LS			Pooled IV		
Initial small	2.584	2.202	2.267	4.788	4.511	4.416
	(1.084)	(1.087)	(1.080)	(.747)	(.756)	(.760)
Initial aide	.170	-.127	-.094	.275	.066	.080
	(.840)	(.839)	(.843)	(.716)	(.716)	(.716)
Cumulative years in small	.997	1.065	.905			
	(.453)	(.455)	(.488)			
Cumulative years in aide	.090	.172	.164			
	(.421)	(.423)	(.435)			
R-squared	.201	.200	.206	.201	.200	.206
# of sample	17468	17468	17468	17468	17468	17468
Student characteristics	Yes	Yes	Yes	Yes	Yes	Yes
teacher characteristics	No	Yes	Yes	No	Yes	Yes
peer characteristics	No	No	Yes	No	No	Yes
School, grade,PP-year	Yes	Yes	Yes	Yes	Yes	Yes

Table 2.41. MI-IPW FD estimates

	MI-IPW FD			unweighted FD		
Cumulative years in small	.982	.894	.901	1.827	1.641	1.613
	(.411)	(.413)	(.413)	(.416)	(.413)	(.414)
Cumulative years in aide	1.241	1.358	1.460	1.738	1.692	1.689
	(.425)	(.430)	(.430)	(.426)	(.427)	(.428)
R-squared	.016	.020	.020	.015	.019	.020
# of sample	11,146	10,761	10,761	9883	9793	9793
Student characteristics	Yes	Yes	Yes	Yes	Yes	Yes
teacher characteristics	No	Yes	Yes	No	Yes	Yes
peer characteristics	No	No	Yes	No	No	Yes
School, grade,PP-year	Yes	Yes	Yes	Yes	Yes	Yes

2.7 Direction (sign) of the bias for unweighted estimators

In this section, we try to obtain information about the direction of the bias for unweighted estimators by making assumption that $E(v_{it}s_{it}|\mathbf{z}_{it})$ and $E(v_{it}s_{it+j}|\mathbf{z}_{it})$ has the same sign. This is because the source of bias for unweighted estimators is $E(v_{it}s_{it}|\mathbf{z}_{it})$ and we cannot directly estimate it. Although we can not obtain an estimate for equation (2.51), we can obtain an estimate for (2.52) and we infer $E(v_{it}s_{it}|\mathbf{z}_{it})$ from $E(v_{it}s_{it+j}|\mathbf{z}_{it})$.

$$E(v_{it}s_{it}|\mathbf{x}_{it},\mathbf{z}_{it}) = E(v_{it}s_{it}|\mathbf{z}_{it}) \text{ where } \mathbf{x}_{it} \subseteq \mathbf{z}_{it} \text{ and } \mathbf{z}_{it} \text{ is instrument} \tag{2.51}$$

where $v_{it} = c_i + u_{it}$

$$E(v_{it}s_{it+j}|\mathbf{z}_{it}) \text{ where } |j| \geq 1 \tag{2.52}$$

Under the assumption that $E(v_{it}s_{it}|\mathbf{z}_{it})$ and $E(v_{it}s_{it+j}|\mathbf{z}_{it})$ has the same sign, we can calculate the direction of the bias of unweighted IV estimates and thereby we can obtain the bound for true return of class size reduction.

2.7.1 Bound for the effects of small class

First, we consider an estimator for $E(v_{it}s_{it+j}|\mathbf{z}_{it})$.

Cross-section IV regression

$$\widehat{E(u_{it}s_{it+j}|\mathbf{z}_{it})} = \frac{1}{n}\sum_{i=1}^{n}\widehat{v}_{it}s_{it+j} \tag{2.53}$$

where $\widehat{v}_{it}$ be residual from unweighted IV estimator [31]We can rewrite IV estimator as follows.

$$\begin{aligned}\widehat{\theta}_{IV} &= (\sum_{i=1}^{n} s_{it}\mathbf{z}'_{it}\mathbf{z}_{it})^{-1}(\sum_{i=1}^{n} s_{it}\mathbf{z}'_{it}y_{it}) = (\sum_{i=1}^{n} s_{it}\mathbf{z}'_{it}\mathbf{z}_{it})^{-1}(\sum_{i=1}^{n} s_{it}\mathbf{z}'_{it}(\mathbf{z}_{it}\theta + v_{it})) \\ &= \theta + (\sum_{i=1}^{n} s_{it}\mathbf{z}'_{it}\mathbf{z}_{it})^{-1}(\sum_{i=1}^{n} \mathbf{z}'_{it}s_{it}v_{it})\end{aligned}$$

$$\begin{aligned}(\sum_{i=1}^{n} s_{it}\mathbf{z}'_{it}\mathbf{z}_{it})^{-1}(\sum_{i=1}^{n} \mathbf{z}'_{it}s_{it}v_{it}) &= (\frac{1}{n}\sum_{i=1}^{n} s_{it}\mathbf{z}'_{it}\mathbf{z}_{it})^{-1}(\frac{1}{n}\sum_{i=1}^{n} \mathbf{z}'_{it}s_{it}v_{it}) \\ &\xrightarrow{p} E(s_{it}\mathbf{z}'_{it}\mathbf{z}_{it})^{-1}E(\mathbf{z}'_{it}s_{it}v_{it}) \\ &\underbrace{=}_{\textbf{Since Z exogenous}} E(s_{it}\mathbf{z}'_{it}\mathbf{z}_{it})^{-1}E(\mathbf{z}'_{it})E(s_{it}v_{it})\end{aligned}$$

Thus, finally, we obtain cross-section IV estimator as follows.

$$\begin{aligned}\widehat{\theta}_{IV} &= \theta + E(s_{it}\mathbf{z}'_{it}\mathbf{z}_{it})^{-1}E(\mathbf{z}'_{it})E(s_{it}v_{it}) \\ \widehat{\theta}_{IV} &> \theta \text{ if } \text{sign}(E(s_{it}v_{it})) \text{ is positive} \\ &< \theta \text{ if } \text{sign}(E(s_{it}v_{it})) \text{ is positive}\end{aligned}$$

Therefore, $\widehat{\theta}_{IV}$ becomes the upper bound for θ_{true} if $\text{sign}(E(s_{it}v_{it}))$ is positive and lower bound if $\text{sign}(E(s_{it}v_{it}))$ is negative.

pooled IV regression

$$\widehat{E(\sum_{t=K}^{G3} u_{it}s_{it+j}|\mathbf{z}_{it})} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=K}^{G3} \widehat{v}_{it}s_{it+j} \tag{2.54}$$

$$\widehat{v}_{it} = y_{it} - \mathbf{z}_{it}\widehat{\theta}_{pooledIV} \text{ , where } \widehat{\theta}_{pooledIV} = (\sum_{i=1}^{n}\sum_{t=K}^{G3} s_{it}\mathbf{z}'_{it}\mathbf{z}_{it})^{-1}(\sum_{i=1}^{n}\sum_{t=K}^{G3} s_{it}\mathbf{z}'_{it}y_{it})$$

[31] This is not conventional IV estimator. We define IV estimator here to be equivalent estimator as in Krueger [1999] where he called this estimator as reduced form estimator.

Similarly, as in cross-section IV case, we obtain an pooled IV estimator as

$$\begin{aligned}
\widehat{\theta}_{pooledIV} &= \theta + E(\sum_{t=K}^{G3} s_{it}\mathbf{z}'_{it}\mathbf{z}_{it})^{-1}E(\sum_{t=K}^{G3}\mathbf{z}'_{it}s_{it}v_{it}) \\
\widehat{\theta}_{IV} &> \theta \text{ if sign}(E(s_{it}v_{it})) \text{ is positive} \\
&< \theta \text{ if sign}(E(s_{it}v_{it})) \text{ is positive}
\end{aligned}$$

This is because $E(\mathbf{z}_{it}) > 0$ and $\mathbf{z}_{it}$ is exogenous in our data.

Estimation results

Pairwise correlation coefficient estimates for (2.53) is reported in table 2.42. The estimates for $E(\widehat{v_{it}s_{it+j}}|\mathbf{z}_{it}) = \frac{1}{n}\sum_{i=1}^{n}\widehat{v}_{it}s_{it+j}$ is positive and statistically significantly different from zero $\forall\ j \geq 1$ in all grades. (i.e. residuals from cross-section IV estimators and lead of selection is positively correlated in statistically significant way.) However, pairwise correlation coefficients for residuals and lag of selection are not statistically different from zero. In other word, we simply cannot reject the null hypothesis of $E(v_{it}s_{it-j}|\mathbf{z}_{it}) = 0$ for $j \geq 1$.

Thus, given that sample estimates for $E(\widehat{v_{it}s_{it+j}}|\mathbf{z}_{it}) = \frac{1}{n}\sum_{i=1}^{n}\widehat{v}_{it}s_{it+j}(\geq 0)$ is non-negative $\forall$ j and j $\neq$=0. It is positive for leads of selection and zero for lags of selection. Therefore, if we assume the of sign$(E(v_{it}s_{it}|\mathbf{z}_{it}))$=sign$(E(v_{it}s_{it+j}|\mathbf{z}_{it}))$ $\forall$ j$\geq$ 1, we can infer that $\widehat{\theta}_{IV}$ is biased upward. Thus, under this assumption, $\widehat{\theta}_{IV}$ overestimate true effects of small class. Table 2.43 shows the upper bound cross-section IV estimates.

Table 2.42. $corr(\widehat{v}_{it},\, s_{it+j})\ \forall$ j≥ 0

	covariates(1)				covariates(2)			
	$\widehat{u}_{iK}$	$\widehat{u}_{iG1}$	$\widehat{u}_{iG2}$	$\widehat{u}_{iG3}$	$\widehat{u}_{iK}$	$\widehat{u}_{iG1}$	$\widehat{u}_{iG2}$	$\widehat{u}_{iG3}$
s_{iK}		.0011	-.0019	-.0020		.0018	.0003	-.0007
s_{iG1}	.1954*		.0115	.0075	.1733*		.0205	.0097
s_{iG2}	.2539*	.2583*		.0074	.2196*	.2261*		-.0007
s_{iG3}	.2708*	.2810*	.1409*		.2282*	.2561*	.1157*	
	covariates(3)				covariates(4)			
	$\widehat{u}_{iK}$	$\widehat{u}_{iG1}$	$\widehat{u}_{iG2}$	$\widehat{u}_{iG3}$	$\widehat{u}_{iK}$	$\widehat{u}_{iG1}$	$\widehat{u}_{iG2}$	$\widehat{u}_{iG3}$
s_{iK}		.0049	.0014	.0022		.0048	.0012	.0022
s_{iG1}	.1692*		.0220	.0099	.1696*		.0224	.0099
s_{iG2}	.2156*	.2235*		.0038	.2156*	.2241*		.0041
s_{iG3}	.2215*	.2481*	.1073*		.2217*	.2485*	.1066*	

(Note) *: statistically significant at the level .05. Covariates (1) include treatments and school dummies, covariates (2) include covariates (1) and student characteristics. Covariates (3) include covariates(2) and teacher characteristics and, finally, covariates (4) include covariates (3) and peer characteristics.

Table 2.43. Upper bound estimates for cross-section IV estimator

	reduced-from							
	K	G1	G2	G3	K	G1	G2	G3
Initial small	**5.205**	**6.621**	**5.279**	**5.247**	**5.224**	**6.490**	**5.064**	**4.979**
	(.749)	(.758)	(.786)	(.814)	(.749)	(.767)	(.786)	(.817)
Initial aide	.259	1.509	1.175	-.072	.434	1.560	1.156	-.078
	(.698)	(.679)	(.700)	(.726)	(.698)	(.680)	(.699)	(.728)
R-squared	.309	.295	.276	.220	.311	.296	.279	.221
# of sample	5840	6430	5919	6086	5840	6430	5919	6086
Other covariates	School dummy, student and teacher characteristics				School dummy, student, teacher peer(classmates) characteristics			

(Note) Students cluster-robust standard errors are provided in parentheses. Student characteristics include (i) race (ii) gender (iii) free lunch status and teacher characteristics include (i) race (ii) years of experience (iii) highest degree. Peer characteristics include (i) fraction of classmates with free lunch (ii) fraction of minority classmates (iii) fraction of female classmates.

2.8 conclusion

Two formal tests of the missing completely at random (MCAR) assumption are introduced in unbalanced panel data. The powers of tests for strict exogeneity increase with cross-section size n and missing proportion of data. For instance, the power of Hausman-type tests which is based on the difference of pooled LS(or RE) and FE estimators with a data set of $n \geq$5,000 and $T = 4$ is 1. The power of the variable addition test is also 1 with a data set of $n \geq$5,000 and $T = 4$.

We compare the estimates of unweighted, IPW and MI-IPW methods in unbalanced panel data using a Monte Carlo simulation. The results show that the IPW method under MAR works well if the probability of selection is correctly specified. We also study the robustness of the IPW estimator with the probability of selection model misspecification. Simulation shows that finite sample bias of the IPW estimator is smaller than that of an unweighted estimator unless misspecification is severe.

We provide the MI-IPW method under MAR to improve efficiency by combining the MI and IPW methods together in an unbalanced panel data model. The estimated effect of class size reduction (CSR) on student academic achievement for students in grades K-3 is about 4 to 6 percent with both IPW LS and FD and MI-IPW LS and FD estimators while the estimated effect for unweighted LS and FD estimators with complete-case is about 6 to 7 percent. Estimators ignoring missing data overestimate the effect of CSR on student score by about 1 to 2 percentage points.

Chapter 3

Cluster robust inference in unbalanced panel data when T is large

3.1 Introduction

A robust inference method in the linear panel data model with a long T has become more relevant recently as more panel data are available with long time dimensions. An inference method which ignores heterogeneity and dependence can severely distort the inference about parameters of interest. For instance, if errors are correlated across observations either in the time or cross-section dimension, a covariance matrix using OLS standard error is biased. Recently, Wooldridge [2003], Hansen [2007], Bester et al. [2009], Vogelsang [2008] and Ibragimov and Muller [2009] provide robust inference methods with dependent data in a balanced panel data model.[1] However, attrition in a panel data and unbalancedness in a two-level data

[1] Arellano [1987] provides asymptotic properties of a covariance matrix in the linear panel models with random sampling (over cross-section unit) but it is limited only to the case when T fixed and n large. Moreover, cluster covariance estimators are used with data that

occur frequently in a large T data or a two-level data. For instance, unbalancedness of data occurs when each school has different numbers of classes or each region has different numbers of states or starting periods of available data that can differ across cross-section units. Therefore, this paper focuses on robust inferences in unbalanced panel data and the effect of an unbalancedness on the validity for the inference methods in balanced panel data. In particular, unbalancedness causes heterogeneity even if attrition or unbalancedness occurs completely at random and heterogeneity can cause distortion in inference. For example, the methods in Hansen [2007] and Vogelsang [2008] require homogeneity in cross-section units to use t-distribution and F-distribution as reference distributions for hypothesis tests. Therefore, unfortunately, robust inference methods in balanced panel data cannot be directly extended to unbalanced panel data.

Among those robust inference methods in balanced panel data, the Ibragimov and Muller [2009] approach provides a t-statistic based inference method. However, this t-statistic based inference method uses only between variations so there is room for improving efficiency. Bester et al. [2009] shows simulation results that the t-statistic based method of Ibragimov and Muller [2009] performs poorly if the estimator presents substantial finite sample bias using the example of 2SLS simulation with weak IV.

On the other hand, Hansen [2007] provides methods based on constructing t and Wald statistics using a cluster covariance matrix estimator (CCE). Under homogeneity in both covariates and errors, t and Wald statistics based on CCE follow standard t and F distri-

has a group structure with independence across groups. In our panel data setting, cross-section units represent groups and time series represents group members and we assume independence between groups. See Wooldridge [2003], Liang and Zeger [1986] and Bertrand et al. [2004] for more details.

bution with scale factors and provide valid inference. Under homogenous covariates and heterogenous variance, t statistic using CCE with scaling factors converges t-distribution so the t-test provides conservative inference in Bakirov and Szekely [2006]. When heterogeneity exists for both covariates and errors, the t statistic using CCE converges to a limiting distribution which is neither standard nor pivotal in both Hansen [2007] and Ibragimov and Muller [2009]. Hansen [2007]'s method without homogeneity can not perform valid robust inference using standard reference distributions of t and F while Ibragimov and Muller [2009]'s approach provides conservative inference of Bakirov and Szekely [2006] in the presence of heterogenous variance.

Vogelsang [2008] provides a robust inference method based on constructing t and Wald statistics with HACs of a covariance matrix where the HAC smoothing parameter grows proportionally to the sample size (this is called fixed-b inference). Vogelsang [2008] also shows convergence of t and Wald statistics with HACs covariance matrix based on fixed-b smoothing parameter to t and F distribution with scaling factors under homogeneity. Both in Hansen [2007] and Vogelsang [2008], we need to calculate limiting distributions for t and Wald statistics treating the covariance matrix as random variables in the limit. This causes heterogeneous variance for the data with different time periods of attrition across different cross-section units.

Our objective, in this paper, is to propose a method which provides a robust inference with heterogenous variance and autocorrelation in an unbalanced linear panel data model. We compare our method to those in Hansen [2007] and Ibragimov and Muller [2009] which are developed for balanced panel data. Our proposed inference method is an extension of the OLS method in Hansen [2007] to weighted least squares (WLS) in constructing t and Wald

statistics using CCE where attrition periods are used as weights. We show that, using the result of Bakirov and Szekely [2006], t-test with WLS provide conservative inference result. To our knowledge, there has been no study that focuses on the asymptotic properties of a variance matrix in unbalanced panel data in the presence of heterogeneity and autocorrelation. For the sake of simplicity, when doing the simulation we focus on random attrition for panel data and random unbalancedness for two-level data.

In the MC experiment, heterogeneity in variance and covariates are induced through random attrition in unbalanced panel data. Our proposed inference method based on WLS which extends Hansen [2007] is compared to the t-statistic based method of Ibragimov and Muller [2009] and the method of Hansen [2007] with OLS. The WLS method which uses the attrition periods information as weights, in comparison to the OLS method in Hansen [2007], shows significantly smaller size distortion. Simulation results also reveal that our method shows power improvement over the t-statistic based method in Ibragimov and Muller [2009]. We also extend our proposed method with WLS to the case where the heterogeneity by both attrition and heteroskedastic variance is present. We compare the power of our method to that of Ibragimov and Muller [2009] and provide evidence for improvement of power performance through a MC simulation.

Unbalancedness which is caused by attrition introduces heterogeneity and this forces conservative t-tests for the method in Ibragimov and Muller [2009]. Thus, we examine the tradeoff for the t-tests based on the method in Ibragimov and Muller [2009] as we drop data on purpose to make the data balanced. Dropping available data inevitably introduces the loss of power while the balancedness of data eliminates heterogeneity so it leads to no size distortion in inference. Simulation results show that the cost of losing power dominates the

benefit of eliminating size distortion as the loss of power is substantial while size distortion is not significant.

The rest of the paper is organized as follows. In section 2, we introduce models with individual fixed effects in unbalanced panel data and propose a robust inference method which extends the method of Hansen [2007] with WLS in unbalanced panel data. Section 3 contains MC experiment results and, in section 4, we provide a brief summary of results and the conclusion.

3.2 Model

This section presents a method for conducting robust inference with attrition in linear unbalanced panel data model.

$$y_{it} = \mathbf{z}_{it}\beta_0 + c_i + f_t + u_{it} \tag{3.1}$$

with $i = 1, 2, ..., n$ and $t = 1, 2, ..., T$.

y_{it} is continuous and $\mathbf{z}_{it}$ is a vector of covariates, c_i is individual fixed effects, f_t is time fixed effects and u_{it} is weakly dependent error. We assume random sampling for cross-section units but allow weak dependence in time dimension.[2] For the sake of simplicity of notation,

[2] The application of balanced panel data model of (3.1) with large T has been appeared in empirical finance literature, in equity returns and earnings surprises, and oftentimes in these models standard errors were estimated with the adjustment to reflect correlation across observations. Typical standard error adjustments was based on Fama and MacBeth [1974] procedure or the method of clustering across time is most common. Recently, the use of panel data and linear model with outcome of serial dependence becomes quite common in finance and macroeconomics applications but Petersen [2009] surveys that the methods each researcher uses to adjust the bias of standard errors are quite various.

we rewrite (3.1) as in (3.2).

$$y_{it} = \mathbf{x}_{it}\beta + c_i + u_{it} \tag{3.2}$$

where $\mathbf{x}_{it} = (\mathbf{z}_{it}, f_t)$ and $\beta = (\beta_0', 1)'$.

Now we define selection indicator (s_{it}) to express the model with unbalanced panel data by random attrition. $s_{it} = 1$ if both dependent variable (y_{it}) and all of covariates ($\mathbf{x}_{it}$) are available and $s_{it} = 0$ otherwise.

Consider the following fixed-effects (FE) transformation of the original model of (3.1) to (3.3) with balanced panel data.

$$\tilde{y}_{it} = \tilde{\mathbf{x}}_{it}\beta + \tilde{u}_{it} \tag{3.3}$$

where

$$\tilde{y}_{it} = y_{it} - \frac{1}{T_i}\sum_{t=1}^{T} s_{it} \cdot y_{it},\ T_i = \sum_{t=1}^{T} s_{it},\ \tilde{\mathbf{x}}_{it} = \mathbf{x}_{it} - \frac{1}{T_i}\sum_{t=1}^{T} s_{it} \cdot \mathbf{x}_{it},\ \tilde{u}_{it} = u_{it} - \frac{1}{T_i}\sum_{t=1}^{T} s_{it}u_{it} \tag{3.4}$$

Correspondingly, fixed-effects (FE) transformed model for unbalanced panel data can be expressed with selection indicator as follows.

$$s_{it}\tilde{y}_{it} = s_{it}\tilde{\mathbf{x}}_{it}\beta + s_{it}\tilde{u}_{it} \tag{3.5}$$

The focus is on inference about β. OLS estimator for β in (3.5) is given by (3.6).(We denote this FE estimator as OLS hereafter.)

$$\hat{\beta}_{ols} = (\sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\tilde{\mathbf{x}}_{it}'\tilde{\mathbf{x}}_{it})^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\tilde{\mathbf{x}}_{it}'\tilde{y}_{it} \tag{3.6}$$

where we use $s_{it}^2 = s_{it}$.

$$\hat{\beta}_{ols} - \beta = (\sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\tilde{\mathbf{x}}_{it}'\tilde{\mathbf{x}}_{it})^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\tilde{\mathbf{x}}_{it}'\tilde{u}_{it} = (\sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\tilde{\mathbf{x}}_{it}'\tilde{\mathbf{x}}_{it})^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\tilde{\mathbf{x}}_{it}'u_{it} \tag{3.7}$$

3.2.1 Robust inference with t-test in unbalanced panel data

In this section, we present fixed-n, large-T asymptotic properties of the CCE for unbalanced panel data. Under fixed-n and large T, the CCE is not consistent but converges in distribution to a limiting random variable. In the CCE with attrition, T_i plays the role of the HAC smoothing parameter as T_i determines the bandwidth for each cross-section unit i. In particular, smoothing parameter, $\lambda_i = \frac{T_i}{T}$ which changes for each cross-section unit is equivalent to b in HAR covariance estimator in fixed-b asymptotic theory (Kiefer and Vogelsang [2002] and Kiefer and Vogelsang [2005]).

For balanced panel data, Hansen [2007] shows that we can perform t and F tests with scaled t and Wald statistics under the assumption that covariates and error terms are independent of cross-section unit (hereafter we call this assumption as cross-section homogeneity).

Theorem 4 (Hansen 2007) *Under cross-section homogeneity and balanced panel, standard t-statistic for $\widehat{\beta}_{ols}$ with cluster robust standard error converges to a t-distribution with $n-1$ degree of freedom, scaled by $\sqrt{\frac{n}{n-1}}$, under the null hypothesis.*

However, the methods which ignore attrition cannot be directly applied with unbalanced panel data since attrition introduces heterogeneity in both covariates and variance so this leads to distortion in inference for unbalanced panel data.

We extend the result of theorem 4 to unbalanced panel data using weights and the result of Bakirov and Szekely [2006]. We use weight for each cross-section unit to recover original homogeneity in covariates of balanced panel data. The idea of recovering homogeneity is that cross-section units with fewer time-series get higher weights for each observation to contribute evenly across cross-section.

With the following assumptions, under cross-section homogeneity of original balanced panel data, t-test with weighted least squares (WLS) of (3.8) and CCE of $\widehat{V}_{w-clus}$ in (3.9) provides a conservative inference result of Bakirov and Szekely [2006] for unbalanced panel data.

$$\widehat{\beta}_{wls} = (\sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\frac{\widetilde{\mathbf{x}_{it}}'}{\sqrt{\lambda_i}}\frac{\widetilde{\mathbf{x}_{it}}}{\sqrt{\lambda_i}})^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\frac{\widetilde{\mathbf{x}_{it}}'}{\sqrt{\lambda_i}}\frac{\widetilde{y_{it}}}{\sqrt{\lambda_i}} \tag{3.8}$$

$$\widehat{V}_{w-clus} = (\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{1}{\lambda_i}s_{it}\widetilde{\mathbf{x}_{it}}'\widetilde{\mathbf{x}_{it}})^{-1}\widehat{\Omega}_{wls}(\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{1}{\lambda_i}s_{it}\widetilde{\mathbf{x}_{it}}'\widetilde{\mathbf{x}_{it}})^{-1}, \tag{3.9}$$

where

$$\widehat{\Omega}_{wls} = \sum_{i=1}^{n}[(\sum_{t=1}^{T}\frac{1}{\lambda_i}s_{it}\widetilde{\mathbf{x}_{it}}'\widehat{u}_{it})(\sum_{t=1}^{T}\frac{1}{\lambda_i}s_{it}\widetilde{\mathbf{x}_{it}}\widehat{u}_{it})]$$

and $\widehat{u}_{it} = \widetilde{y_{it}} - \widetilde{\mathbf{x}_{it}}\widehat{\beta}_{WLS}$.

Assumption 5 *1. Independent sampling across cross-section, n is fixed and T_i goes infinity for each i*

2. Weak dependence for $\boldsymbol{x}_{it}, u_{it}$

3. (WLLN) $\frac{1}{T}\sum_{t=1}^{T} s_{it}\widetilde{\boldsymbol{x}_{it}}'\widetilde{\boldsymbol{x}_{it}} \xrightarrow{p} \lambda_i Q_i$ and Q_i^{-1} exists, where $\frac{T_i}{T} = \lambda_i$ $\lambda_i \in (0,1]$

4. (FCLT) $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} s_{it}\widetilde{\boldsymbol{x}_{it}}u_{it} \Rightarrow \Lambda_i W(\lambda_i)$, where $W(\lambda_i)$ is brownian motion with mean zero and variance λ_i

5. (Homogeneity across cross-section) $Q_i = Q$ *and* $\Lambda_i = \Lambda$

6. (Missing Completely At Random) $E(u_{it}|\boldsymbol{x}_{it}, \boldsymbol{s}_i) = 0$ *where* $\boldsymbol{s}_i = (s_{i1}, s_{i2}, ..., s_{iT})$

Consider testing linear hypothesis about β of the form

$$H_0 : R\beta = r,\ H_1 : R\beta \neq r \tag{3.10}$$

where R is a $q \times k$ matrix of known constants with full rank and r is a $q \times 1$ vector of known constants.

In the case $q = 1$ we can define the t-statistics

$$\widehat{t} = \frac{\sqrt{T}(R\widehat{\beta}_{ols} - r)}{\sqrt{R \cdot T \cdot \widehat{V}_{clus} R'}}\ ,\ t^*_{wls} = \frac{\sqrt{T}(R\widehat{\beta}_{wls} - r)}{\sqrt{R \cdot T \cdot \widehat{V}_{w-clus} R'}} \tag{3.11}$$

Using following two theorems 6 and 7, we can obtain corollary 8 which allows conservative inference for t-test with weight.

Theorem 6 (Bakirov and Szekely [2006]) *Let* $X_1, X_2, ..., X_n$ *be an i.i.d. sample from normal scale mixture, and let* $y_1, y_2, ..., y_n$ *be independent normal* $(0, \sigma_i)$ *random variables,* $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, *and* $S_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2$. *Then*

$$P\{\sqrt{n}\frac{\overline{X} - \mu}{S_X} > x\} = \sup_{\sigma_1, \sigma_2, ... \sigma_n} \int_{R^n} P\{\sqrt{n}\frac{\overline{y}}{S_y} > x\} \prod_{i=1}^{n} dF(\sigma_i)$$

Proof. See Bakirov and Szekely [2006]. ■

Theorem 7 (Bakirov and Szekely [2006]) *Let* $C_{n-1}(\alpha)$ *be the critical value of the usual two-sided t-test based on (3) of level* α *(i.e.* $P(|T_{n-1}| > C_{n-1}(\alpha)) = \alpha$*)*

$$T_{n-1} = \frac{\sqrt{n}\overline{X}}{S_X} \text{ where } X_1, X_2, ..., X_n \text{ be an i.i.d.sample from normal} \tag{3.12}$$

Then, using theorem 6,

(i) If $\alpha \leq 2 \cdot (1 - \Phi(\sqrt{3})) = 0.0832....$, *then* $\forall n \geq 2$,

$$\sup_{\sigma_1,\sigma_2,...\sigma_n} P(|\widehat{T}_{n-1}| > C_{n-1}(\alpha)) \leq P(|T_{n-1}| > C_{n-1}(\alpha)) = \alpha \tag{3.13}$$

(i.e. $\alpha \leq 0.0832$ *and* $\forall n \geq 2$, $\widehat{T}_{n-1} \Rightarrow T_{n-1}$ *where* T_{n-1} *is t-distribution with* $n-1$ *degree of freedom)*

(ii) Usual one-sided t-test of nominal level 0.05 or lower remains valid as long as $n \leq 14$. *For* $n > 15$, *usual two-sided tests of nominal level 0.1(or one-sided tests of nominal level 0.05) are not automatically conservative for large* n.

(iii) For $\alpha \leq 0.0832$ *and* $\forall n \geq 2$, *as* $n \to \infty$,

$$\sup_{\sigma_1,\sigma_2,...\sigma_n} P(|\widehat{T}_{n-1}| > C_{n-1}(\alpha)) \leq P(Z > C_{n-1}(\alpha)) = \alpha$$

where $Z \sim N(0, I)$

Proof. See Bakirov and Szekely [2006]. ■

Corollary 8 *Under assumptions 5 and attrition, the inference based on* t^*_{wls} *in (3.11) scaled by* $\sqrt{\frac{n}{n-1}}$ *converges to* $\widehat{T}_{n-1}$*-distribution in theorem 7, under the null hypothesis.*

Proof. Here we provide the sketch of the proof. See appendix for details. First, consider the numerator of t^*_{wls}. $\widehat{\beta}_{wls} = (\sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\frac{\widetilde{\mathbf{x}_{it}}'}{\sqrt{\lambda_i}}\frac{\widetilde{\mathbf{x}_{it}}}{\sqrt{\lambda_i}})^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\frac{\widetilde{\mathbf{x}_{it}}'}{\sqrt{\lambda_i}}\frac{\widetilde{y_{it}}}{\sqrt{\lambda_i}}$. Under assumptions 5 and null hypothesis, we can rewrite numerator as

$$\sqrt{T}(R\widehat{\beta}_{wls} - R\beta) = R\sqrt{T}(\widehat{\beta}_{wls} - \beta)$$

$$R\sqrt{T}(\widehat{\beta}_{wls} - \beta) = R\sqrt{T}(\sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\frac{\widetilde{\mathbf{x}_{it}}'}{\sqrt{\lambda_i}}\frac{\widetilde{\mathbf{x}_{it}}}{\sqrt{\lambda_i}})^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\frac{\widetilde{\mathbf{x}_{it}}'}{\sqrt{\lambda_i}}\frac{u_{it}}{\sqrt{\lambda_i}}$$

$$R\sqrt{T}(\widehat{\beta}_{wls}-\beta)=R\sqrt{T}(\sum_{i=1}^{n}\sum_{t=1}^{T}s_{it}\frac{\widetilde{\mathbf{x}_{it}}'}{\sqrt{T_i}}\frac{\widetilde{\mathbf{x}_{it}}}{\sqrt{T_i}})^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}s_{it}\frac{\widetilde{\mathbf{x}_{it}}'}{\sqrt{T_i}}\frac{u_{it}}{\sqrt{T_i}}$$

$$=> R(\sum_{i=1}^{n}\frac{1}{\lambda_i}\lambda_i Q_i)^{-1}\sum_{i=1}^{n}\frac{1}{\lambda_i}\Lambda_i W(\lambda_i)=R(\sum_{i=1}^{n}Q_i)^{-1}\sum_{i=1}^{n}\frac{1}{\lambda_i}\Lambda_i W(\lambda_i)$$

where $W(\lambda_i)$ is wiener process with mean zero and variance λ_i.

Now consider denominator $T\cdot\widehat{V}_{w-clus}$

$$T\cdot\widehat{V}_{w-clus}=(\sum_{i=1}^{n}\frac{1}{T_i}\sum_{t=1}^{T}s_{it}\widetilde{x_{it}}'\widetilde{x_{it}})^{-1}\widehat{\Omega}_{wls}(\sum_{i=1}^{n}\frac{1}{T_i}\sum_{t=1}^{T}s_{it}\widetilde{x_{it}}'\widetilde{x_{it}})^{-1}$$

$$\widehat{\Omega}_{wls}=\sum_{i=1}^{n}[(\frac{1}{\lambda_i}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}s_{it}\widetilde{x_{it}}'\widehat{u}_{it})(\frac{1}{\lambda_i}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}s_{it}\widetilde{x_{it}}'\widehat{u}_{it})']$$

$$\frac{1}{\lambda_i}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}s_{it}\widetilde{x_{it}}'\widehat{u}_{it}=\frac{1}{\lambda_i}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}s_{it}\widetilde{x_{it}}'u_{it}-\frac{1}{\lambda_i}\frac{1}{T}\sum_{t=1}^{T}s_{it}\widetilde{x_{it}}'\widetilde{x_{it}}\sqrt{T}(\widehat{\beta}_{WLS}-\beta)+o_p(1)$$

$$\Longrightarrow\frac{1}{\lambda_i}\Lambda_i W(\lambda_i)-Q_i(\sum_{j=1}^{n}Q_j)^{-1}(\sum_{j=1}^{n}\frac{1}{\lambda_j}\Lambda_j W(\lambda_j))$$

Let's denote $W(\lambda_j)$ as W_j for simplicity of notation.

Under cross-section homogeneity, $T\cdot\widehat{V}_{w-clus}\Rightarrow(\sum_{i=1}^{n}Q_i)^{-1}\widehat{\Omega}_{wls}(\sum_{i=1}^{n}Q_i)^{-1}=(nQ)^{-1}\widehat{\Omega}_{wls}(nQ)^{-1}$ and $\widetilde{\Omega}_{wls}\Rightarrow\Lambda[\sum_{i=1}^{n}\frac{W_i}{\lambda_i}\frac{W_i'}{\lambda_i}-\frac{1}{n}(\sum_{j=1}^{n}\frac{W_j}{\lambda_j})(\sum_{j=1}^{n}\frac{W_j'}{\lambda_j})]\Lambda'$.

Let's define y_i as $y_i\equiv RQ^{-1}\Lambda\frac{W(\lambda_i)}{\lambda_i}$. Since $W_i\equiv W(\lambda_i)$ is brownian motion with mean 0 and the variance λ_i, y_i is mean zero and variance of $\frac{RQ^{-1}\Lambda var(W(\lambda_i))\Lambda'Q^{-1}R'}{\lambda_i^2}$. As the variance of $W(\lambda_i)$ is λ_i*I_q, for the case of $q=1$, $y_1,y_2,y_3,.....,y_n$ is independent and each y_i is normal $(0,\frac{RQ^{-1}\Lambda\Lambda'Q^{-1}R'}{\lambda_i})$. Then, in the case of $q=1$, we have

$$t^*_{wls}\Rightarrow\frac{\sum_{i=1}^{n}y_i}{\sqrt{\sum_{i=1}^{n}y_i^2-\frac{1}{n}\sum_{i=1}^{n}y_i\sum_{j=1}^{n}y_j}}=\frac{n\overline{y}}{\sqrt{\sum_{i=1}^{n}y_i^2-n\overline{y}^2}}=\frac{n\overline{y}}{\sqrt{\sum_{i=1}^{n}(y_i-\overline{y})^2}}$$

$$=\sqrt{\frac{n}{n-1}}\cdot\frac{\sqrt{n}\overline{y}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\overline{y})^2}}\equiv\sqrt{\frac{n}{n-1}}\widehat{T}_{n-1}\text{ where }\widehat{T}_{n-1}\equiv\sqrt{n}\frac{\overline{y}}{S_y}$$

Thus, the proper weight leaves λ_i as the only source of heterogeneity so that we can apply Bakirov and Szekely [2006] result and obtain conservative inference. ■

Remark 9 *Heterogeniety is induced by attrition. Attrition leads to heterogeneity in* $s_{it}\widetilde{\boldsymbol{x}_{it}}'\widetilde{\boldsymbol{x}_{it}}$ *and* $s_{it}\widetilde{\boldsymbol{x}_{it}}'\widetilde{u_{it}}$ *for each* i*. However, using the ratio of the length of time-series to full length of time series as weight,* $\lambda_i = \frac{T_i}{T}$*, we apply WLS and the weight eliminates heterogeneity in covariates.*

Under cross-section homogeneity assumption, as attrition is the unique source of heterogeneity, the use of weights recovers the homogeneity in covariates. Moreover, since empirically most relevant case is two-sided tests with $\alpha \leq 0.05$, we can conduct inference with $t*_{wls}$-statistics by multiplying $\sqrt{\frac{n-1}{n}}$, which is equivalent to replace $\widehat{V}_{w-clus}$ with $\frac{n}{n-1} \cdot \widehat{V}_{w-clus}$. For a significant level .05 or lower, the t-test with $t*_{wls}$-statistic scaled by $\sqrt{\frac{n-1}{n}}$ remains conservative.

3.2.2 WLS with heterogenous covariance matrix in unbalanced panel data

In addition to heterogeneity by attrition, we introduce heteroskedastic variance for each cross-section unit in unbalanced panel data. For balanced panel data, t-statistic based test in Bester et al. [2009] and Ibragimov and Muller [2009] provide conservative inference results of Bakirov and Szekely [2006] in the presence of heteroskedastic variance. For unbalanced panel data, if homogeneity only holds for covariates but not for variance so that $Q_i = Q$ and $\Lambda_i \neq \Lambda$ hold because of heteroskedastic variance, we can still obtain the conservative

inference result in Bakirov and Szekely [2006] with t-test as we use $t*_{wls}$-statistic with scale factor.

Under heteroskedastic variance and homogenous covariates for unbalanced panel data, we have two sources of heterogeneity but we can still apply corollary 8 with $\widehat{\beta}_{WLS}$ and $\widehat{V}_w - clus$ and perform t-test. Only change from homoskedastic variance is that we define y_i in the proof of corollary 8 as $RQ^{-1}\Lambda_i \frac{W(\lambda_i)}{\lambda_i}$ so that its variance is $\frac{RQ^{-1}\Lambda_i\Lambda_i'Q^{-1}R'}{\lambda_i}$. The key for the determination of weights is making each cross-section unit to contribute evenly as in the case of no heteroskedastic variance. Units with fewer time series observation and with smaller variance receive more weights than the units with more time series or large variance. This weighting eliminates heterogeniety in covariates but two sources of heteroskedasticity in error still remains after weighting.

In sum, we show that robust inference method in balanced panel data can be extended to unbalanced panel data with proper choice of weights for cross-section units. Corollary 8 shows that t-test with WLS in unbalanced panel data provides the bound result of conservative inference under heterogenous variance and attrition.

3.3 Monte Carlo experiment

In this section, we examine finite sample properties of proposed WLS method in unbalanced linear panel data model with dependent data. We provide two sets of simulation. The only source of heterogeneity is attrition in the first simulation while there are two sources of heterogeneity, attrition and heteroskedastic variance in the second simulation. Simulation results for the size distortion and the power of test are provided. In each set of simulation,

we compare the size distortion and the power of our method to those of Hansen [2007] and Ibragimov and Muller [2009] methods which were developed for balanced panel data. Attrition time periods are generated in simulation not to be correlated with either outcome or covariates. Thus, each cross-section unit has different attrition time period and this leads to different time-series length which introduces heterogenous variance for each cross-section unit.

3.3.1 The t-test using t^*_{wls}-statistic with scaled factor under homogeniety and attrition

DGP I: Attrition is the only source of heterogeneity

We consider following data generating process for balanced panel data.

$$Y_{it} = x_{it} \cdot \beta + c_i + f_t + u_{it}, \text{ for } i = 1, 2, .., n \text{ and } t = 1, 2, ..., T$$

where x_{it} is continuous variable.

Variables generation

1. $x_{it} \sim iidN(0, 1)$, $\beta = 1$

2. Homogeneous error, $\mathbf{u}_i$, with t=T

$$\begin{bmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{iT} \end{bmatrix} \sim MVN \left(\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \left(\begin{bmatrix} 1 & \rho & \dots & \rho^{T-1} \\ \rho & 1 & & \rho^{T-2} \\ \vdots & & 1 & \vdots \\ \rho^{T-1} & & \dots & 1 \end{bmatrix} \right) \right)$$

where MVN is multivariate normal and ρ=0.75 in the simulation.

3. $c_i = a_i + \bar{x}_i$, $a_i \sim iidN(0,1)$ and $\bar{x}_i = \sum_{t=1}^{T} s_{it} \cdot x_{it}$

4. $f_t = 0$

The determination of attrition for unbalanced panel data is presented in table 3.1. Two key features in the determination of attrition are: (i) the proportion of available data is about 60% and (ii) the length of available time-series for each cross-section units vary so the variation of time length for each cross-section unit induces heterogeneity. We consider various length of time-series, T=50, 100, 300, and 1,000 for balanced panel data. T=50 has relevance for yearly data of 40 years such as data from 1970 to 2009, while T=100 has relevance for quarterly data of 25 years such as from 1985 to 2009.

Table 3.1. selection process

T=50		missing fraction
n=4	50(2), 10(2)	0.4
n=5	50(2), 30(1), 10(2)	0.4
n=6	50(2), 40(1), 20(1), 10(2)	0.4
n=7	50(2), 40(1), 30(1), 20(1), 10(2)	0.4
n=8	50(2), 40(2), 20(2), 10(2)	0.4
n=9	50(2), 40(1), 35(1), 30(1), 25(1), 20(1), 10(2)	0.4
n=10	50(2), 40(1), 35(1), 30(2), 25(1), 20(1), 10(2)	0.4
n=11	50(2), 40(1), 35(1), 30(3), 25(1), 20(1), 10(2)	0.4
n=16	50(2), 40(2), 35(2), 30(4), 25(2), 20(2), 10(2)	0.4
T=100		missing fraction
n=4	100(2), 20(2)	0.4
n=5	100(2), 60(1), 20(2)	0.4
n=6	100(2), 70(1), 50(1), 20(2)	0.4
n=7	100(2), 70(1), 60(1), 50(1), 20(2)	0.4
n=8	100(2), 70(1), 60(2), 50(1), 20(2)	0.4
n=16	100(2), 80(2), 70(2), 60(4), 50(2), 40(2), 20(2)	0.4
T=300		missing fraction
n=4	300(2), 100(2)	0.33
n=5	300(2), 100(3)	0.4
n=6	300(2), 140(2), 100(2)	0.4
n=7	300(2), 180(1), 140(2), 100(2)	0.4
n=8	300(2), 200(2), 160(1), 140(1), 100(2)	0.4
n=16	300(4), 240(1), 210(2), 150(2), 125(2), 90(4)	0.4

(Note) The length of time is reported and number of cross-section is reported in parentheses.

Simulation result: Homogeneity and attrition

In all simulation, we consider inference about parameter β in panel data model of (3.3) using OLS in (3.6) and WLS in (3.8). Table 3.2 shows the size of t-test with nominal level 0.05. The first column shows the dimension of panel data. The second column reports the rejection probability of t-test for balanced panel data using t-statistic in Hansen [2007]. As the homegeneity assumption is satisfied, t-test for balanced panel data shows no size distortion. Coverage rate for rejection probability contain nominal level 0.05 for all n. Third column reports the rejection probability for t-test of complete case from unbalanced panel data. For all simulated cross-section dimension $n \leq 8$, rejection probability shows substantial size distortion by over-rejection and the size distortion is greater for unbalanced panel data with smaller n. Therefore, simulation results verify that t-test using t-statistic in Hansen [2007] show size distortion with unbalanced panel data. The fourth column shows the rejection probability from t-test of Ibragimov and Muller [2009] method. t-statistics in Ibragimov and Muller [2009] is obtained from (3.14).

$$t_\beta = \sqrt{n}\frac{\bar{\hat{\beta}}}{s_{\hat{\beta}}} \tag{3.14}$$

where $\bar{\hat{\beta}} = \frac{1}{n}\sum_{j=1}^{n}\hat{\beta}_j$ and $s^2_{\hat{\beta}} = \frac{1}{n-1}\sum_{j=1}^{n}(\beta_j - \bar{\hat{\beta}})^2$

The rejection probability from t-test of Ibragimov and Muller [2009] method provides the evidence of conservative inference when there is heterogeneity. It also shows that the size distortion tapers off as n increases. Eventually, Ibragimov and Muller [2009] method with heterogeneity and n=16 shows no size distortion at all.

The fifth column in table 3.2 provides the rejection probability of t-test with t^*_{wls}-statistics

where the weights for each cross-section is calculated by $\sqrt{\frac{T}{T_i}}$. As WLS put more weights to cross-section unit with shorter length so that eliminate heterogeneity across cross-section. Simulation results for t-test with t^*_{wls}-statistics in table 3.2 verify the conservative inference of corollary 8 and the size distortion with WLS is substantially smaller than that of OLS especially for small $n \leq 8$. In sum, WLS method reduces type I error. The bound from WLS is very close to .05 and much closer to .05 than the bound from Ibragimov and Muller [2009] method.

Table 3.3 reports the size adjusted power of t-test for T=50. Power is greater for t-test of Hansen [2007] with WLS than for that of Ibragimov and Muller [2009]. An inference of Hansen [2007] with WLS dominates that of Ibragimov and Muller [2009] in all n and the power advantage of Hansen [2007] with WLS comes from using both within and between variations while Ibragimov and Muller [2009] only use between variation. All these imply that a robust inference of Hansen [2007] with WLS is preferred method especially with small $n \leq 8$ when attrition is the sole source of heterogeneity. We obtain the same qualitative implication in table 3.4 with T=100. As n and T increases, the power of tests also increases for all four methods. The order of higher power is balanced data, complete case with OLS, complete case with WLS and complete case with Ibragimov and Muller [2009] method.

Table 3.2. The size of test, $1(p < 0.05)$

T=50	balanced OLS	complete-case OLS	complete-case t-stat inference of IM	complete-case WLS
n=4	.051(.047,.055)	.090(.084,.096)	.032(.028,.035)	.051(.046,.055)
n=5	.053(.048,.057)	.074(.069,.079)	.034(.030,.037)	.051(.046,.055)
n=6	.053(.048,.057)	.068(.063,.073)	.039(.035,.043)	.050(.044,.054)
n=7	.053(.048,.057)	.065(.060,.070)	.039(.035,.043)	.049(.045,.054)
n=8	.049(.045,.053)	.058(.053,.062)	.045(.041,.049)	.049(.044,.053)
n=16	.050(.046,.054)	.049(.045,.054)	.050(.044,.054)	.046(.041,.051)
T=100	balanced OLS	complete-case OLS	complete-case t-stat inference of IM	complete-case WLS
n=4	.050(.046,.054)	.083(.078,.089)	.035(.031,.038)	.044(.040,.048)
n=5	.052(.047,.056)	.090(.085,.096)	.024(.021,.027)	.045(.041,.049)
n=6	.048(.044,.053)	.079(.073,.084)	.029(.026,.032)	.045(.041,.049)
n=7	.049(.045,.054)	.068(.063,.073)	.028(.025,.031)	.043(.040,.047)
n=8	.051(.046,.056)	.061(.056,.066)	.042(.038,.045)	.045(.041,.049)
n=16	.051(.045,.057)	.050(.044,.056)	.051(.045,.058)	.050(.044,.056)
T=300	balanced OLS	complete-case OLS	complete-case t-stat inference of IM	complete-case WLS
n=4	.051(.046,.055)	.075(.069,.080)	.034(.030,.037)	.042(.038,.046)
n=5	.049(.044,.053)	.067(.062,.071)	.040(.036,.044)	.045(.041,.049)
n=6	.052(.047,.056)	.067(.062,.072)	.042(.038,.046)	.047(.043,.051)
n=7	.047(.043,.051)	.062(.058,.067)	.041(.037,.045)	.045(.041,.049)
n=8	.049(.044,.053)	.061(.056,.066)	.042(.038,.045)	.045(.041,.049)
n=16	.048(.044,.052)	.055(.050,.059)	.047(.043,.051)	.049(.045,.054)

(Note) The rejection probability with nominal level .05 is reported. 95 % coverage rate for rejection probability is reported in parenthesis. The number of replications in simulation is 10,000. Balanced OLS and complete case OLS is based on the methods in Hansen [2007].

Table 3.3. The size adjusted power of test

T=50	balanced	complete-case	complete-case	complete-case
n=4	OLS	OLS	t-stat inference of IM	WLS
bias=0	.051(.047,.055)	.090(.084,.096)	.032(.028,.035)	.051(.046,.055)
bias=2%	.056(.051,.060)	.089(.083,.094)	.033(.029,.037)	.050(.045,.054)
bias=5%	.077(.071,.082)	.103(.097,.109)	.039(.035,.042)	.060(.056,.065)
bias=10%	.147(.140,.153)	.152(.145,.159)	.061(.057,.066)	.096(.091,.102)
bias=20%	.395(.385,.404)	.347(.337,.356)	.157(.149,.164)	.244(.236,.253)
T=50	balanced	complete-case	complete-case	complete-case
n=5	OLS	OLS	t-stat inference of IM	WLS
bias=0	.053(.048,.057)	.074(.069,.079)	.034(.030,.037)	.051(.046,.055)
bias=2%	.057(.052,.061)	.082(.077,.087)	.037(.033,.041)	.054(.050,.059)
bias=5%	.084(.078,.089)	.098(.092,.104)	.048(.044,.053)	.067(.062,.072)
bias=10%	.188(.181,.196)	.171(.164,.178)	.087(.082,.093)	.119(.113,.126)
bias=20%	.551(.541,.561)	.429(.420,.439)	.238(.229,.246)	.345(.226,.354)
T=50	balanced	complete-case	complete-case	complete-case
n=6	OLS	OLS	t-stat inference of IM	WLS
bias=0	.053(.048,.057)	.068(.063,.073)	.039(.035,.042)	.050(.045,.054)
bias=2%	.055(.050,.060)	.074(.069,.079)	.045(.041,.049)	.056(.051,.060)
bias=5%	.091(.086,.097)	.096(.090,.101)	.059(.054,.063)	.072(.067,.077)
bias=10%	.227(.219,.235)	.188(.180,.195)	.107(.100,.113)	.149(.142,.156)
bias=20%	.668(.658,.677)	.498(.488,.507)	.307(.298,.316)	.430(.420,.440)
T=50	balanced	complete-case	complete-case	complete-case
n=7	OLS	OLS	t-stat inference of IM	WLS
bias=0	.053(.048,.057)	.065(.060,.070)	.039(.035,.042)	.049(.045,.054)
bias=2%	.058(.053,.062)	.072(.067,.077)	.045(.040,.049)	.056(.052,.061)
bias=5%	.107(.101,.113)	.100(.094,.106)	.064(.059,.069)	.080(.075,.086)
bias=10%	.278(.269,.287)	.206(.198,.214)	.131(.125,.138)	.172(.165,.179)
bias=20%	.763(.754,.771)	.576(.567,.586)	.372(.363,.382)	.513(.503,.523)
T=50	balanced	complete-case	complete-case	complete-case
n=8	OLS	OLS	t-stat inference of IM	WLS
bias=0	.049(.045,.053)	.058(.053,.062)	.045(.041,.049)	.049(.044,.053)
bias=2%	.059(.054,.063)	.066(.061,.071)	.046(.042,.050)	.056(.052,.061)
bias=5%	.118(.111,.124)	.105(.099,.111)	.075(.069,.080)	.089(.084,.095)
bias=10%	.320(.311,.330)	.230(.221,.238)	.158(.151,.165)	.201(.193,.209)
bias=20%	.835(.827,.842)	.641(.631,.650)	.465(.455,.474)	.601(.591,.610)

(Note) The rejection probability is reported. 95 % coverage rate for rejection probability is reported in parenthesis. Generated errors are from AR(1) process with correlation coefficient 0.75. The number of replications is 10,000. Balanced OLS and complete case OLS is based on the methods in Hansen [2007].

Table 3.4. The size adjusted power of test, $1(p < 0.05)$, $T = 100$

T=100	balanced	complete-case	complete-case	complete-case
n=4	OLS	OLS	t-stat inference of IM	WLS
bias=0	.050(.046,.054)	.083(.078,.089)	.035(.031,.038)	.044(.040,.048)
bias=2%	.056(.051,.060)	.083(.078,.089)	.036(.032,.039)	.049(.044,.053)
bias=5%	.097(.091,.104)	.117(.109,.124)	.044(.039,.049)	.067(.061,.073)
bias=10%	.228(.220,.237)	.223(.215,.231)	.097(.091,.103)	.140(.133,.146)
bias=20%	.646(.635,.657)	.549(.537,.560)	.289(.278,.299)	.407(.396,.418)
T=100	balanced	complete-case	complete-case	complete-case
n=5	OLS	OLS	t-stat inference of IM	WLS
bias=0	.052(.047,.056)	.090(.085,.096)	.024(.021,.027)	.045(.041,.049)
bias=2%	.063(.056,.069)	.097(.089,.104)	.027(.023,.031)	.046(.040,.051)
bias=5%	.117(.109,.125)	.141(.132,.149)	.040(.035,.045)	.080(.073,.087)
bias=10%	.320(.309,.330)	.281(.271,.290)	.096(.090,.103)	.176(.168,.185)
bias=20%	.821(.813,.828)	.678(.669,.687)	.285(.276,.293)	.543(.533,.553)
T=100	balanced	complete-case	complete-case	complete-case
n=6	OLS	OLS	t-stat inference of IM	WLS
bias=0	.048(.044,.053)	.079(.073,.084)	.029(.026,.032)	.045(.041,.049)
bias=2%	.058(.054,.063)	.089(.083,.094)	.030(.027,.033)	.056(.052,.060)
bias=5%	.140(.131,.148)	.142(.133,.150)	.056(.050,.061)	.094(.087,.102)
bias=10%	.406(.396,.416)	.315(.306,.324)	.131(.124,.138)	.233(.225,.241)
bias=20%	.916(.909,.923)	.766(.755,.777)	.399(.387,.411)	.679(.667,.691)
T=100	balanced	complete-case	complete-case	complete-case
n=7	OLS	OLS	t-stat inference of IM	WLS
bias=0	.049(.045,.054)	.068(.063,.073)	.028(.025,.031)	.043(.039,.047)
bias=2%	.073(.067,.080)	.086(.078,.093)	.038(.033,.043)	.060(.052,.061)
bias=5%	.172(.161,.182)	.142(.132,.151)	.067(.059,.069)	.102(.093,.111)
bias=10%	.487(.474,.500)	.366(.354,.378)	.171(.161,.180)	.293(.282,.304)
bias=20%	.964(.960,.968)	.850(.842,.858)	.495(.484,.505)	.789(.780,.798)
T=100	balanced	complete-case	complete-case	complete-case
n=8	OLS	OLS	t-stat inference of IM	WLS
bias=0	.051(.046,.056)	.061(.056,.066)	.042(.038,.045)	.045(.041,.049)
bias=2%	.070(.064,.075)	.078(.072,.083)	.049(.044,.053)	.060(.052,.071)
bias=5%	.188(.178,.198)	.163(.153,.173)	.095(.086,.104)	.128(.118,.139)
bias=10%	.554(.544,.564)	.408(.398,.418)	.248(.238,.258)	.348(.338,.358)
bias=20%	.986(.982,.989)	.893(.883,.902)	.670(.659,.680)	.855(.844,.865)

(Note) The rejection probability is reported. 95 % coverage rate for rejection probability is reported in parenthesis. Generated errors are from AR(1) process with correlation coefficient 0.75. The number of replications is 10,000. Balanced OLS and complete case OLS is based on the methods in Hansen [2007].

Table 3.5. Configuration of heteroskedastic variance

	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
n=4	1.5	.5	2	.8				
n=5	1.5	.5	2	.8	.9			
n=6	1.5	.5	2	.8	.9	1.1		
n=7	1.5	.5	2	.8	.9	1.1	1.2	
n=8	1.5	.5	2	.8	.9	1.1	1.2	.7

3.3.2 The t-test using t^*_{wls}-statistic with scaled factor under heterogenous variance and attrition

DGP II: Attrition and heterokedastic variance

We use the same data generating process as in DGP I where attrition is the only source of heterogeneity and attrition configurations as in table 3.1. In DGP II, the only difference in DGP is coming from error process $\mathbf{u}_i$. As in DGP I, each u_{it} is initially drawn from standard normal with AR(1) process of serial correlation coefficient, ρ=0.75. Then, we introduce heteroskedastic variance by replace u_{it} with $a_i * u_{it}$ where, for the case of n=4, we assign a_1=1.5, a_2=.5, a_3=2 and a_4=.8. Heteroskedastic variance configurations in DGPs are provided in table 3.5.

Simulation result: Heteroskedastic variance and attrition

Table 3.6 shows the size of t-test at the nominal level .05 for data with n=4,5,6,7,8,16 and T=100. The t-test inference of Hansen [2007] results in balanced panel data are reported in the second column and it shows conservative inference for all n. As n increases, the rejection frequency approaches nominal level .05 so that size distortion disappears. Third column shows the results of t-test with complete case in unbalanced panel data. There are

two sources of heterogeneity for the t-test with complete case: attrition and heteroskedastic variance. Both sources of heterogeneity are not controlled for t-test with complete case. Results in third column shows that two sources of heterogeneity may work against each other as attrition mitigates standard error while heteroskedastic variance increases standard error used in t-statistic. Thus, there are no consistent pattern of over-rejection or under-rejection as we change n. The fourth column reports the results of t-tests using the method in Ibragimov and Muller [2009] with complete case in unbalanced panel data. The t-tests provide conservative inference for all n while rejection frequency approaches to true nominal level .05 as n increases up to n=16. The mean value of MC rejection frequency for Ibragimov and Muller [2009] is lower than that for Hansen [2007] in balanced panel data. Fifth column reports the results of t-test with complete case in unbalanced panel using WLS which is proposed in this paper. For this method, weights used in WLS eliminate heterogeneity in covariates due to attrition while heterogeneity in errors by two sources still remains. The results in fifth column shows conservative inference for all n. The rejection frequency approaches to true nominal level .05 as n increases up to n=16. The mean value of MC rejection frequency is lower than that in balanced panel data for all n.

Table 3.7 shows the power of t-tests for the method in Hansen [2007] with balanced panel, the method in Hansen [2007] with unbalanced panel, the method in Ibragimov and Muller [2009] with unbalanced panel, and our suggested WLS method with unbalanced panel. The power increases with n for all methods. Our WLS method has higher power than the method in Ibragimov and Muller [2009] for all n and all magnitude of bias. The third column shows the power for the methods in Hansen [2007] while the fifth column reports the power for our WLS method in unbalanced panel. The WLS method has lower power than the method in

Table 3.6. The size of test, $1(p < 0.05)$

T=100	balanced OLS	complete-case OLS	complete-case t-stat inference of IM	complete-case WLS
n=4	.033(.030,.037)	.031(.028,.034)	.017(.014,.020)	.018(.015,.020)
n=5	.036(.031,.040)	.049(.044,.054)	.027(.023,.031)	.028(.024,.031)
n=6	.038(.033,.042)	.051(.045,.056)	.027(.023,.031)	.033(.028,.037)
n=7	.042(.038,.046)	.051(.047,.056)	.029(.025,.032)	.037(.033,.040)
n=8	.042(.038,.046)	.055(.050,.060)	.036(.032,.040)	.042(.037,.046)
n=16	.047(.041,.053)	.058(.052,.064)	.042(.036,.048)	.046(.040,.052)

(Note) The rejection probability with nominal level .05 is reported. 95 % coverage rate for rejection probability is reported in parenthesis. The number of replications in simulation is 2,000. Balanced OLS and complete case OLS is based on the methods in Hansen [2007].

Hansen [2007] but the difference is quite small overall. However, compared to homogeneous case, the introduction of heteroskedastic variance leads to loss of power for all methods we considered in simulation.

3.3.3 Trade-off calculation: Dropping observation to make balanced panel data

As unbalancedness of panel data causes heterogeneity and it forces conservative t-tests for the methods in Ibragimov and Muller [2009] and Bester et al. [2009] and leads to t-tests invalid when these methods use t-distribution as reference distribution. Thus, we examine the tradeoff of dropping some observations on purpose to make data balanced.[3] Dropping observations inevitably introduces the loss of power while balanced panel data eliminates

[3] We observe this practice of balancing data occasionally especially in international macroeconomic data. Many data are available from 1970 and the number of countries analyzed sometimes less than 50. It is quite common that researchers have shorter time-series for developing countries and they use only balanced data from 1980 or later period.

Table 3.7. The size adjusted power of test, $1(p < 0.05)$, $T = 100$

T=100	balanced	complete-case	complete-case	complete-case
n=4	OLS	OLS	t-stat inference of IM	WLS
bias=0	.033(.030,.037)	.031(.028,.034)	.017(.014,.020)	.018(.015,.020)
bias=2%	.037(.028,.046)	.033(.030,.037)	.017(.014,.020)	.019(.016,.021)
bias=5%	.069(.061,.077)	.059(.051,.064)	.028(.024,.033)	.034(.027,.041)
bias=10%	.190(.182,.198)	.153(.145,.161)	.064(.059,.069)	.093(.087,.100)
bias=20%	.497(.489,.505)	.422(.414,.430)	.208(.203,.213)	.300(.293,.307)
T=100	balanced	complete-case	complete-case	complete-case
n=5	OLS	OLS	t-stat inference of IM	WLS
bias=0	.036(.031,.044)	.049(.044,.054)	.027(.023,.031)	.028(.024,.031)
bias=2%	.045(.037,.053)	.097(.089,.104)	.027(.023,.031)	.046(.040,.053)
bias=5%	.081(.073,.089)	.141(.133,.149)	.040(.035,.045)	.080(.073,.087)
bias=10%	.260(.252,.268)	.281(.273,.289)	.096(.091,.101)	.176(.169,.184)
bias=20%	.681(.673,.689)	.678(.670,.686)	.285(.276,.290)	.543(.535,.550)
T=100	balanced	complete-case	complete-case	complete-case
n=6	OLS	OLS	t-stat inference of IM	WLS
bias=0	.038(.033,.042)	.051(.045,.056)	.027(.023,.031)	.033(.028,.037)
bias=2%	.051(.043,.059)	.058(.053,.063)	.035(.030,.040)	.043(.037,.050)
bias=5%	.114(.106,.122)	.107(.100,.114)	.051(.047,.056)	.077(.070,.084)
bias=10%	.316(.308,.324)	.254(.246,.262)	.111(.106,.116)	.181(.174,.188)
bias=20%	.786(.778,.794)	.622(.614,.630)	.349(.344,.354)	.565(.558,.572)
T=100	balanced	complete-case	complete-case	complete-case
n=7	OLS	OLS	t-stat inference of IM	WLS
bias=0	.042(.038,.046)	.051(.047,.056)	.029(.025,.032)	.037(.033,.041)
bias=2%	.059(.051,.067)	.063(.056,.071)	.036(.031,.040)	.045(.038,.052)
bias=5%	.137(.129,.144)	.117(.110,.125)	.060(.055,.065)	.085(.078,.092)
bias=10%	.391(.383,.399)	.311(.303,.319)	.134(.129,.140)	.241(.234,.248)
bias=20%	.870(.863,.878)	.717(.709,.725)	.409(.404,.415)	.655(.648,.662)
T=100	balanced	complete-case	complete-case	complete-case
n=8	OLS	OLS	t-stat inference of IM	WLS
bias=0	.042(.038,.046)	.055(.050,.060)	.036(.032,.040)	.042(.037,.046)
bias=2%	.070(.064,.075)	.078(.072,.083)	.049(.044,.053)	.060(.052,.071)
bias=5%	.188(.178,.198)	.163(.153,.173)	.095(.086,.104)	.128(.118,.139)
bias=10%	.554(.544,.564)	.408(.398,.418)	.248(.238,.258)	.348(.338,.358)
bias=20%	.986(.982,.989)	.893(.883,.902)	.670(.659,.680)	.855(.844,.865)

(Note) The rejection probability is reported. 95 % coverage rate for rejection probability is reported in parenthesis. The number of replications is 2,000. Balanced OLS and complete case OLS is based on the methods in Hansen [2007].

heterogeneity and the size distortion in inference. Thus, we can quantify the tradeoff between the benefit of eliminating the rejection probability bias and the cost of losing power using DGP I.

Table 3.8 reports the size of t-tests for a method in Ibragimov and Muller [2009]. The first column shows cross-section dimension of data. The second column reports the rejection probability of t-test for balanced panel data without attrition. The fourth and fifth columns report the rejection probability of t-test for balanced panels by dropping observations (reduced balanced panel) from complete case for time dimension and cross-section dimension respectively. These three columns show that there is no size distortion of inferences for balanced panels in simulation. The third column reports the rejection probability of t-test for complete case of unbalanced panel data. Different length of data due to missing process configuration of table 3.1 induces heterogeneity and heteroskedastic variance leads to conservative inference of t-test as theory prediction of Ibragimov and Muller [2009] approach indicates. As n increases the degree of conservativeness of inference is mitigated and, for n=16, inference exhibits no size distortion regardless of size of time dimensions.

Table 3.9 and appendix I report the size adjusted power of t-tests in Ibragimov and Muller [2009] for T=50, 100, 300 and 1,000 respectively. Regardless of T, the loss of power is quite significant for reduced balanced panel. Especially for large $n \geq 8$ and bias=20%, the power of t-test for reduced balanced panel is less than 50% of the power of unbalanced panel of complete case. This is noticeable decline of power by dropping observations. As reported in tables 3.3 and 3.4, the difference of the power of t-tests of between complete case with OLS and complete case with WLS is negligibly small. The power analysis for reduced balanced panel data shows that the cost of losing power overwhelms the benefit of no size distortion.

Table 3.8. The size of test, $1(p < 0.05)$ where AR(1) error with correlation coefficient 0.75

T=50	balanced	complete-case	balanced keep $\forall i$ drop t	balanced keep $\forall t$ drop i
n=4	.053(.045,.059)	.031(.025,.036)	.050(.042,.056)	.050(.043,.056)
n=5	.054(.046,.060)	.034(.029,.040)	.045(.039,.051)	.055(.048,.063)
n=6	.047(.040,.054)	.037(.031,.043)	.049(.042,.056)	.050(.043,.056)
n=7	.046(.039,.053)	.042(.036,.048)	.048(.041,.054)	.049(.042,.055)
n=8	.051(.044,.058)	.039(.033,.045)	.050(.043,.056)	.052(.045,.059)
n=9	.052(.045,.059)	.040(.034,.046)	.047(.040,.053)	.052(.045,.059)
n=10	.047(.040,.054)	.042(.036,.048)	.049(.042,.056)	.053(.046,.060)
n=11	.049(.042,.056)	.048(.041,.055)	.051(.044,.058)	.052(.045,.059)
n=16	.050(.043,.057)	.048(.041,.055)	.054(.047,.062)	.048(.041,.054)
T=100	balanced	complete-case	balanced-drop t	balanced -drop i
n=4	.048(.041,.054)	.028(.023,.033)	.046(.040,.053)	.053(.046,.060)
n=5	.050(.043,.056)	.023(.019,.028)	.051(.044,.058)	.047(.040,.053)
n=6	.046(.040,.053)	.027(.022,.032)	.046(.039,.053)	.048(.041,.054)
n=7	.054(.047,.061)	.030(.025,.036)	.047(.040,.054)	.048(.041,.054)
n=8	.048(.042,.055)	.038(.032,.044)	.050(.043,.057)	.053(.046,.060)
n=16	.048(.041,.055)	.049(.040,.058)	.050(.042,.057)	.050(.043,.057)
T=300	balanced	complete-case	balanced-drop t	balanced -drop i
n=4	.046(.039,.053)	.029(.022,.035)	.053(.042,.063)	.048(.039,.057)
n=5	.051(.044,.058)	.032(.026,.038)	.051(.045,.058)	.046(.040,.053)
n=6	.053(.046,.060)	.039(.032,.045)	.050(.043,.057)	.052(.046,.059)
n=7	.045(.038,.052)	.038(.032,.044)	.051(.044,.059)	.049(.041,.056)
n=8	.053(.042,.055)	.042(.035,.049)	.049(.040,.057)	.050(.042,.058)
n=16	.046(.041,.055)	.050(.043,.057)	.050(.043,.057)	.048(.041,.055)
T=1,000	balanced	complete-case	balanced-drop t	balanced -drop i
n=4	.056(.049,.063)	.033(.026,.040)	.042(.034,.049)	.057(.050,.064)
n=5	.051(.044,.058)	.039(.033,.046)	.045(.038,.052)	.049(.042,.055)
n=6	.050(.043,.057)	.040(.034,.047)	.050(.043,.057)	.050(.043,.057)
n=7	.047(.040,.053)	.038(.032,.044)	.047(.040,.053)	.053(.046,.060)
n=8	.046(.038,.055)	.037(.031,.044)	.052(.045,.058)	.055(.048,.063)
n=9	.051(.041,.058)	.043(.036,.051)	.049(.042,.056)	.045(.038,.052)
n=16	.050(.043,.057)	.046(.039,.053)	.050(.043,.057)	.048(.041,.055)

(Note) The rejection probability is reported. 95 % coverage rate for rejection probability is reported in parenthesis. The number of replications is 4,000.

Table 3.9. The size adjusted power of test, $1(p < 0.05)$, $T = 50$

T=50, n=4	balanced	complete-case	balanced keep $\forall i$ drop t	balanced keep $\forall t$ drop i
bias=0	.053(.045,.059)	.031(.025,.036)	.050(.042,.056)	.050(.043,.056)
bias=2%	.058(.047,.068)	.043(.034,.051)	.047(.037,.057)	.057(.046,.067)
bias=5%	.074(.063,.085)	.049(.040,.058)	.049(.040,.058)	.066(.056,.076)
bias=10%	.137(.121,.152)	.072(.060,.083)	.058(.047,.068)	.082(.070,.094)
bias=20%	.391(.369,.412)	.174(.157,.190)	.101(.088,.114)	.121(.107,.135)
T=50, n=8	balanced	complete-case	balanced-drop t	balanced -drop i
bias=0	.051(.044,.058)	.039(.033,.045)	.050(.043,.056)	.052(.045,.059)
bias=2%	.058(.048,.068)	.052(.042,.062)	.054(.044,.064)	.059(.049,.069)
bias=5%	.114(.100,.128)	.071(.059,.082)	.060(.049,.070)	.064(.053,.074)
bias=10%	.307(.287,.327)	.152(.136,.167)	.081(.069,.092)	.083(.071,.095)
bias=20%	.812(.794,.829)	.467(.445,.488)	.204(.186,.222)	.132(.117,.146)
T=50, n=16	balanced	complete-case	balanced-drop t	balanced -drop i
bias=0	.050(.043,.057)	.048(.041,.055)	.054(.047,.062)	.048(.041,.054)
bias=2%	.074(.063,.085)	.053(.043,.062)	.053(.043,.063)	.054(.049,.064)
bias=5%	.191(.173,.208)	.102(.089,.115)	.083(.071,.095)	.073(.062,.084)
bias=10%	.613(.591,.634)	.306(.286,.326)	.176(.159,.193)	.132(.117,.147)
bias=20%	.993(.989,.997)	.810(.792,.827)	.503(.481,.525)	.383(.361,.404)

(Note) The rejection probability is reported. 95 % coverage rate for rejection probability is reported in parenthesis. Generated errors are from AR(1) process with correlation coefficient 0.75. The number of replications is 2,000.

3.4 Conclusion

In this paper, we extend the robust inference method with heterogenous variance and autocorrelation in Hansen [2007] to unbalanced panel data. In unbalanced panel data, heterogeneity is induced by attrition and heterogeneity in both covariate and variance makes the t-test using OLS with scaled t-statistic in Hansen [2007] invalid. We derive robust inference result using weighted least squares (WLS) for unbalanced panel data where t^*_{wls}-statistic with scaling factor provides conservative inference results with standard t-test in the presence of both attrition and heteroskedastic variance.

We examine our proposed WLS inference method and compare to the methods in Hansen [2007] and Ibragimov and Muller [2009] via MC simulation. Simulation results verify the bound result of t-test for the method with WLS when attrition is the only source of heterogeneity and compared to the method in Hansen [2007], the size distortion is significantly reduced with WLS especially for small n. Moreover, compared to t-statistic based method of Ibragimov and Muller [2009], WLS method performs better in power and the bound of type I error is closer to .05. Simulation also shows that the proposed WLS method in an unbalanced panel provides conservative inference but shows greater power than the method in Ibragimov and Muller [2009] in the presence of both heterogenous variance and attrition. Finally, we examine the tradeoff of dropping some observations on purpose to make the data balanced. Balanced panel data eliminate heterogeneity and the size distortion in inference but the loss of power due to dropping observations is quite substantial. Overall, the cost of losing power dominates the benefit of eliminating the rejection probability bias.

APPENDICES

Appendix A

Gaussian copula density

Consider following joint logistic distribution for $(u_1, u_2, ..., u_T)$.

$$F(u_1, u_2, ..., u_T) = C(F_1(u_1), ..., F_t(u_t), ..., F_T(u_T); \rho)$$

where $F(\cdot)$ is joint logistic distribution and $F_t(\cdot)$ is univariate logistic distribution.

Then, we can obtain joint logistic density as following.

$$f(u_1, u_2, ..., u_T) = \frac{\partial^T C(\varepsilon_1, \varepsilon_2, ..., \varepsilon_T; \rho)}{\partial \varepsilon_1 \partial \varepsilon_2 \cdots \partial \varepsilon_T} \frac{\partial \varepsilon_1}{\partial u_1} \frac{\partial \varepsilon_2}{\partial u_2} \cdots \frac{\partial \varepsilon_T}{\partial u_T} \tag{A.1}$$

where $F_t(u_t) = \varepsilon_t$, $\frac{\partial \varepsilon_t}{\partial u_t} = f_t(u_t)$ for all t and let $\frac{\partial^T C(\varepsilon_1, \varepsilon_2, ..., \varepsilon_T); \rho)}{\partial \varepsilon_1 \partial \varepsilon_2 \cdots \partial \varepsilon_T} = c(\varepsilon_1, \varepsilon_2, ..., \varepsilon_T; \rho)$.

Since $C(\cdot)$ is gaussian copula, we can rewrite $c(\cdot)$ as follow using the probability integral transformation(Hoel et al. [1972]) which transforms normal CDF to logistic CDF.

$$c(\varepsilon_1, \varepsilon_2, ..., \varepsilon_T; \rho) = \frac{\phi(\Phi_1^{-1}(\varepsilon_1), \Phi_2^{-1}(\varepsilon_2), ..., \Phi_T^{-1}(\varepsilon_T); \rho)}{\phi_1(\Phi_1^{-1}(\varepsilon_1)) \cdot \phi_2(\Phi_2^{-1}(\varepsilon_2)) \cdots \phi_T(\Phi_T^{-1}(\varepsilon_T))} \tag{A.2}$$

where $\phi(\cdot)$ is joint normal density and $\phi_t(\cdot)$ is a univariate normal density.

Using the equations (A.1) and (A.2), we obtain following logistic density. In applications, we can use equation (A.3) to estimate model parameters by maximum likelihood estimation.

$$f(u_1, u_2, ..., u_T) = \phi(\Phi_1^{-1}(F_1(u_1)), \Phi_2^{-1}(F_2(u_2)), ..., \Phi_T^{-1}(F_T(u_T)); \rho) \prod_{t=1}^{T} \frac{f_t(u_t)}{\phi_t(\Phi_t^{-1}(F_t(u_t)))} \tag{A.3}$$

where joint normal density is given by the following.

$$\phi(g_1, g_2, ..., g_T) = (\frac{1}{2\pi})^{\frac{T}{2}} |\Sigma|^{-\frac{1}{2}} exp(-\frac{1}{2}\mathbf{g}'(|\Sigma|^{-1} - I_T)\mathbf{g})$$

where $\mathbf{g} = (g_1, g_2, ..., g_T)'$ and Σ is covariance matrix of $\mathbf{g}$, cov($\mathbf{g}$).

Appendix B

Proof for the consistency of conditional logit estimator in chapter 1

Likelihood of joint T events $y_{i1}, y_{i2}...., y_{iT}$ can be obtained from the following.

$$P(y_{i1} = y_1, ..., y_{iT} = y_T|\mathbf{x}_i, c_i, n_i = n) = \frac{P(y_{i1} = y_1, ..., y_{iT} = y_T|\mathbf{x}_i, c_i)}{P(n_i = n|\mathbf{x}_i, c_i)}$$

$$\underbrace{=}_{CI-A4} \frac{P(y_{i1} = y_1|\mathbf{x}_i, c_i)P(y_{i2} = y_2|\mathbf{x}_i, c_i) \cdot \cdot \cdot \cdot \cdot P(y_{iT} = y_T|\mathbf{x}_i, c_i)}{P(n_i = n|\mathbf{x}_i, c_i)}$$

$$\underbrace{=}_{SE-A2} \frac{P(y_{i1} = y_1|x_{i1}, c_i)P(y_{i2} = y_2|x_{i2}, c_i) \cdot \cdot \cdot \cdot \cdot P(y_{iT} = y_T|x_{i3}, c_i)}{P(n_i = n|\mathbf{x}_i, c_i)}$$

$$\underbrace{=}_{A3} \frac{\exp(\sum_{t=1}^{T} y_{it}x_{it}\beta)}{\sum_{a \in R_i} [\exp(\sum_{t=1}^{T} a_t x_{it}\beta)]} \tag{B.1}$$

where $a = (a_1, a_2, ..., a_T)$, $a_t \in \{0, 1\}$ $a_t = 1$ if $y_{it} = 1$, $\sum_{t=1}^{T} a_t = n_i$ and R_i is the subset of R^T defined as $\{a \in R^T : a_t \in \{0, 1\}$ and $\sum_{t=1}^{T} a_t = n_i\}$.

$$\begin{aligned} \widehat{\beta}_{c-\log it} &= \arg\max_{b} \sum_{i=1}^{n} l_i(b) \qquad \text{(B.2)} \\ l_i(b) &= \log\{\frac{\exp(\sum_{t=1}^{T} y_{it} x_{it} b)}{\sum_{a \in R_i} [\exp(\sum_{t=1}^{T} a_t x_{it} b)]}\} \end{aligned}$$

Appendix C

Monte Carlo simulation results:

Coefficient β for continuous covariate

Table C.1. Unconditional logit estimates for β with no serial correlation

ρ=0, n=100	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	2.46	1.20	0.86	0.30	(0.28,0.32)
$t = 3$	1.75	0.46	0.44	0.34	(0.32,0.36)
$t = 4$	1.48	0.31	0.29	0.33	(0.31,0.35)
$t = 5$	1.36	0.23	0.22	0.32	(0.30,0.34)
$t = 6$	1.28	0.18	0.18	0.30	(0.28,0.32)
$t = 7$	1.23	0.16	0.16	0.29	(0.27,0.31)
$t = 8$	1.20	0.14	0.14	0.27	(0.25,0.29)
ρ=0, n=200	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	2.19	0.59	0.54	0.60	(0.58,0.62)
$t = 3$	1.67	0.30	0.29	0.63	(0.61,0.65)
$t = 4$	1.45	0.20	0.20	0.61	(0.59,0.63)
$t = 5$	1.33	0.16	0.15	0.57	(0.54,0.59)
$t = 6$	1.26	0.13	0.13	0.52	(0.50,0.55)
$t = 7$	1.22	0.11	0.11	0.50	(0.47,0.52)
$t = 8$	1.18	0.10	0.10	0.47	(0.45,0.49)
ρ=0, n=400	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	2.07	0.37	0.36	0.89	(0.88,0.90)
$t = 3$	1.63	0.21	0.20	0.92	(0.90,0.93)
$t = 4$	1.42	0.14	0.14	0.90	(0.89,0.91)
$t = 5$	1.32	0.11	0.11	0.87	(0.85,0.88)
$t = 6$	1.25	0.09	0.09	0.82	(0.81,0.83)
$t = 7$	1.20	0.08	0.08	0.77	(0.75,0.79)
$t = 8$	1.17	0.07	0.07	0.74	(0.71,0.77)
ρ=0, n=600	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	2.03	0.30	0.29	0.99	(0.98,0.99)
$t = 3$	1.62	0.17	0.16	0.98	(0.98,0.99)
$t = 4$	1.42	0.11	0.11	0.98	(0.98,0.99)
$t = 5$	1.31	0.09	0.09	0.96	(0.95,0.98)
$t = 6$	1.24	0.07	0.07	0.95	(0.94,0.96)
$t = 7$	1.20	0.06	0.06	0.92	(0.91,0.94)
$t = 8$	1.17	0.05	0.05	0.90	(0.88,0.92)

*Note: r is the rejection rate for the null of H_0:β=1 with nominal value 0.05. ρ is AR(1) serial correlation coefficient for the errors.

Table C.2. Conditional logit estimates for β with no serial correlation

ρ=0, n=100	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	1.23	0.60	0.45	0.03	(0.03,0.04)
$t = 3$	1.08	0.26	0.25	0.04	(0.03,0.05)
$t = 4$	1.04	0.20	0.19	0.06	(0.05,0.07)
$t = 5$	1.04	0.16	0.16	0.05	(0.04,0.06)
$t = 6$	1.03	0.14	0.14	0.04	(0.03,0.05)
$t = 7$	1.03	0.13	0.12	0.05	(0.04,0.06)
$t = 8$	1.02	0.11	0.11	0.05	(0.04,0.06)
ρ=0, n=200	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	1.09	0.29	0.27	0.04	(0.03,0.05)
$t = 3$	1.03	0.17	0.17	0.05	(0.04,0.06)
$t = 4$	1.02	0.13	0.13	0.05	(0.04,0.06)
$t = 5$	1.02	0.11	0.11	0.05	(0.04,0.06)
$t = 6$	1.01	0.10	0.10	0.05	(0.04,0.06)
$t = 7$	1.01	0.09	0.09	0.05	(0.04,0.06)
$t = 8$	1.01	0.08	0.08	0.05	(0.04,0.06)
ρ=0, n=400	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	1.03	0.19	0.18	0.04	(0.03,0.05)
$t = 3$	1.01	0.12	0.12	0.05	(0.04,0.06)
$t = 4$	1.01	0.09	0.09	0.05	(0.04,0.06)
$t = 5$	1.00	0.08	0.08	0.05	(0.04,0.06)
$t = 6$	1.00	0.07	0.07	0.05	(0.04,0.06)
$t = 7$	1.00	0.06	0.06	0.05	(0.04,0.06)
$t = 8$	1.00	0.05	0.05	0.05	(0.04,0.06)
ρ=0, n=600	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	1.01	0.15	0.15	0.05	(0.04,0.06)
$t = 3$	1.01	0.10	0.09	0.05	(0.04,0.06)
$t = 4$	1.00	0.07	0.07	0.05	(0.04,0.06)
$t = 5$	1.00	0.06	0.06	0.06	(0.05,0.07)
$t = 6$	1.00	0.05	0.05	0.05	(0.04,0.06)
$t = 7$	1.00	0.05	0.05	0.05	(0.04,0.06)
$t = 8$	1.00	0.05	0.05	0.05	(0.04,0.06)

*Note: r is the rejection rate for the null of H_0:β=1 with nominal value 0.05. ρ is AR(1) serial correlation coefficient for the errors.

Table C.3. Unconditional logit estimates for β with serial correlation

ρ=0.2, n=100	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	2.88	1.89	1.04	0.36	(0.34,0.38)
$t = 3$	1.94	0.53	0.48	0.45	(0.43,0.48)
$t = 4$	1.60	0.34	0.31	0.46	(0.44,0.48)
$t = 5$	1.45	0.24	0.24	0.46	(0.44,0.48)
$t = 6$	1.35	0.20	0.19	0.41	(0.39,0.43)
$t = 7$	1.28	0.17	0.16	0.39	(0.37,0.41)
$t = 8$	1.25	0.15	0.14	0.38	(0.36,0.40)
ρ=0.2, n=200	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	2.50	0.72	0.62	0.72	(0.70,0.74)
$t = 3$	1.84	0.33	0.32	0.80	(0.78,0.82)
$t = 4$	1.56	0.21	0.21	0.78	(0.76,0.80)
$t = 5$	1.41	0.16	0.16	0.75	(0.73,0.76)
$t = 6$	1.32	0.13	0.13	0.70	(0.68,0.72)
$t = 7$	1.27	0.12	0.11	0.67	(0.65,0.69)
$t = 8$	1.22	0.10	0.10	0.62	(0.60,0.64)
ρ=0.2, n=400	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	2.37	0.42	0.42	0.97	(0.96,0.98)
$t = 3$	1.80	0.23	0.22	0.98	(0.97,0.99)
$t = 4$	1.54	0.15	0.15	0.98	(0.97,0.99)
$t = 5$	1.40	0.12	0.11	0.97	(0.96,0.98)
$t = 6$	1.32	0.09	0.09	0.95	(0.93,0.96)
$t = 7$	1.26	0.08	0.08	0.92	(0.91,0.93)
$t = 8$	1.22	0.07	0.07	0.90	(0.88,0.92)
ρ=0.2, n=600	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	2.31	0.35	0.33	0.99	(0.99,1)
$t = 3$	1.79	0.19	0.18	1	(1,1)
$t = 4$	1.53	0.12	0.12	1	(1,1)
$t = 5$	1.40	0.09	0.09	1	(1,1)
$t = 6$	1.31	0.07	0.08	1	(1,1)
$t = 7$	1.25	0.07	0.07	0.99	(0.98,0.99)
$t = 8$	1.22	0.06	0.06	0.98	(0.97,0.99)

*Note: r is the rejection rate for the null of H_0:β=1 with nominal value 0.05. ρ is AR(1) serial correlation coefficient for the errors.

Table C.4. Conditional logit estimates for β with serial correlation

ρ=0.2, n=100	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	1.44	0.94	0.56	0.02	(0.01,0.02)
$t = 3$	1.18	0.30	0.27	0.05	(0.04,0.06)
$t = 4$	1.12	0.22	0.20	0.07	(0.06,0.08)
$t = 5$	1.10	0.17	0.17	0.07	(0.06,0.08)
$t = 6$	1.08	0.15	0.14	0.07	(0.06,0.08)
$t = 7$	1.07	0.13	0.13	0.07	(0.06,0.08)
$t = 8$	1.06	0.12	0.12	0.07	(0.06,0.08)
ρ=0.2, n=200	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	1.25	0.36	0.32	0.05	(0.04,0.06)
$t = 3$	1.14	0.19	0.18	0.07	(0.06,0.08)
$t = 4$	1.09	0.14	0.14	0.08	(0.07,0.09)
$t = 5$	1.08	0.11	0.12	0.08	(0.07,0.09)
$t = 6$	1.06	0.10	0.10	0.08	(0.07,0.09)
$t = 7$	1.06	0.09	0.09	0.09	(0.08,0.10)
$t = 8$	1.05	0.09	0.08	0.09	(0.08,0.10)
ρ=0.2, n=400	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	1.18	0.21	0.21	0.08	(0.06,0.09)
$t = 3$	1.11	0.13	0.13	0.11	(0.10,0.13)
$t = 4$	1.08	0.10	0.10	0.12	(0.10,0.13)
$t = 5$	1.07	0.08	0.08	0.13	(0.12,0.15)
$t = 6$	1.06	0.07	0.07	0.12	(0.10,0.14)
$t = 7$	1.05	0.06	0.06	0.10	(0.08,0.11)
$t = 8$	1.04	0.06	0.06	0.10	(0.08,0.12)
ρ=0.2, n=600	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	1.16	0.17	0.17	0.10	(0.08,0.12)
$t = 3$	1.11	0.11	0.10	0.15	(0.13,0.17)
$t = 4$	1.08	0.08	0.08	0.16	(0.13,0.18)
$t = 5$	1.06	0.06	0.07	0.14	(0.12,0.17)
$t = 6$	1.05	0.06	0.06	0.14	(0.12,0.16)
$t = 7$	1.04	0.05	0.05	0.13	(0.11,0.15)
$t = 8$	1.03	0.05	0.05	0.13	(0.11,0.15)

*Note: r is the rejection rate for the null of H_0:β=1 with nominal value 0.05. ρ is AR(1) serial correlation coefficient for the errors.

Table C.5. Unconditional logit estimates for β with serial correlation

ρ=0.4, n=100	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	3.59	2.65	1.32	0.44	(0.42,0.46)
$t = 3$	2.26	0.65	0.57	0.60	(0.58,0.63)
$t = 4$	1.81	0.40	0.36	0.63	(0.61,0.65)
$t = 5$	1.60	0.27	0.26	0.65	(0.63,0.67)
$t = 6$	1.47	0.21	0.21	0.62	(0.60,0.65)
$t = 7$	1.38	0.18	0.18	0.57	(0.54,0.59)
$t = 8$	1.33	0.16	0.15	0.57	(0.55,0.59)
ρ=0.4, n=200	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	2.99	0.92	0.77	0.84	(0.83,0.86)
$t = 3$	2.12	0.39	0.38	0.93	(0.91,0.94)
$t = 4$	1.75	0.25	0.24	0.93	(0.91,0.94)
$t = 5$	1.56	0.18	0.18	0.92	(0.91,0.94)
$t = 6$	1.44	0.15	0.14	0.89	(0.88,0.91)
$t = 7$	1.36	0.12	0.12	0.87	(0.86,0.89)
$t = 8$	1.31	0.11	0.11	0.83	(0.81,0.84)
ρ=0.4, n=400	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	2.79	0.54	0.50	0.99	(0.99,1)
$t = 3$	2.07	0.27	0.26	1	(1,1)
$t = 4$	1.73	0.17	0.17	1	(1,1)
$t = 5$	1.54	0.13	0.13	1	(1,1)
$t = 6$	1.42	0.10	0.10	1	(1,1)
$t = 7$	1.35	0.09	0.09	0.99	(0.99,1)
$t = 8$	1.30	0.08	0.08	0.98	(0.98,0.99)
ρ=0.4, n=600	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	2.72	0.43	0.40	1	(1,1)
$t = 3$	2.06	0.21	0.21	1	(1,1)
$t = 4$	1.72	0.14	0.14	1	(1,1)
$t = 5$	1.53	0.10	0.10	1	(1,1)
$t = 6$	1.42	0.08	0.08	1	(1,1)
$t = 7$	1.35	0.07	0.07	1	(1,1)
$t = 8$	1.30	0.06	0.06	1	(1,1)

*Note: r is the rejection rate for the null of H_0:β=1 with nominal value 0.05. ρ is AR(1) serial correlation coefficient for the errors.

Table C.6. Conditional logit estimates for β with serial correlation

ρ=0.4, n=100	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	1.80	1.32	0.77	0.004	(0.01,0.02)
$t = 3$	1.35	0.35	0.32	0.11	(0.04,0.06)
$t = 4$	1.25	0.25	0.22	0.14	(0.06,0.08)
$t = 5$	1.20	0.19	0.18	0.15	(0.06,0.08)
$t = 6$	1.17	0.16	0.15	0.16	(0.14,0.18)
$t = 7$	1.13	0.14	0.14	0.15	(0.13,0.16)
$t = 8$	1.13	0.13	0.13	0.15	(0.13,0.16)
ρ=0.4, n=200	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	1.50	0.46	0.39	0.12	(0.10,0.13)
$t = 3$	1.28	0.21	0.21	0.21	(0.19,0.23)
$t = 4$	1.21	0.16	0.15	0.25	(0.23,0.27)
$t = 5$	1.18	0.13	0.12	0.26	(0.23,0.27)
$t = 6$	1.15	0.11	0.11	0.25	(0.22,0.26)
$t = 7$	1.13	0.09	0.10	0.24	(0.21,0.26)
$t = 8$	1.11	0.09	0.09	0.23	(0.21,0.25)
ρ=0.4, n=400	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	1.39	0.27	0.25	0.27	(0.25,0.29)
$t = 3$	1.26	0.15	0.14	0.40	(0.38,0.42)
$t = 4$	1.20	0.11	0.11	0.45	(0.42,0.47)
$t = 5$	1.16	0.09	0.09	0.46	(0.43,0.48)
$t = 6$	1.14	0.08	0.08	0.41	(0.38,0.44)
$t = 7$	1.12	0.07	0.07	0.42	(0.39,0.45)
$t = 8$	1.10	0.06	0.06	0.40	(0.37,0.43)
ρ=0.4, n=600	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	1.36	0.21	0.20	0.40	(0.37,0.43)
$t = 3$	1.26	0.12	0.12	0.59	(0.56,0.62)
$t = 4$	1.19	0.09	0.09	0.64	(0.61,0.67)
$t = 5$	1.16	0.07	0.07	0.61	(0.58,0.64)
$t = 6$	1.13	0.06	0.06	0.59	(0.56,0.62)
$t = 7$	1.11	0.06	0.05	0.56	(0.52,0.59)
$t = 8$	1.10	0.05	0.05	0.53	(0.50,0.56)

*Note: r is the rejection rate for the null of H_0:β=1 with nominal value 0.05. ρ is AR(1) serial correlation coefficient for the errors.

Table C.7. Unconditional logit estimates for β with serial correlation

ρ=0.6, n=100	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	5.32	10.91	1.89	0.57	(0.55,0.59)
$t = 3$	2.87	0.94	0.76	0.79	(0.77,0.81)
$t = 4$	2.21	0.49	0.45	0.86	(0.84,0.87)
$t = 5$	1.90	0.33	0.32	0.88	(0.87,0.90)
$t = 6$	1.71	0.25	0.25	0.87	(0.85,0.88)
$t = 7$	1.58	0.21	0.20	0.85	(0.84,0.87)
$t = 8$	1.49	0.18	0.17	0.83	(0.82,0.85)
ρ=0.6, n=200	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	3.94	1.66	1.09	0.91	(0.90,0.93)
$t = 3$	2.66	0.52	0.49	0.99	(0.98,0.99)
$t = 4$	2.13	0.32	0.30	0.99	(0.99,1)
$t = 5$	1.84	0.22	0.22	0.99	(0.99,1)
$t = 6$	1.67	0.17	0.17	0.99	(0.99,1)
$t = 7$	1.55	0.15	0.14	0.99	(0.99,1)
$t = 8$	1.47	0.13	0.12	0.98	(0.98,0.99)
ρ=0.6, n=400	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	3.55	0.74	0.67	1	(1,1)
$t = 3$	2.56	0.35	0.33	1	(1,1)
$t = 4$	2.09	0.21	0.21	1	(1,1)
$t = 5$	1.82	0.15	0.15	1	(1,1)
$t = 6$	1.64	0.12	0.12	1	(1,1)
$t = 7$	1.53	0.10	0.10	1	(1,1)
$t = 8$	1.46	0.09	0.09	1	(1,1)
ρ=0.6, n=600	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	3.43	0.58	0.52	0.99	(0.99,1)
$t = 3$	2.55	0.27	0.27	1	(1,1)
$t = 4$	2.08	0.16	0.17	1	(1,1)
$t = 5$	1.80	0.12	0.12	1	(1,1)
$t = 6$	1.64	0.10	0.10	1	(1,1)
$t = 7$	1.53	0.08	0.08	1	(1,1)
$t = 8$	1.45	0.07	0.07	1	(1,1)

*Note: r is the rejection rate for the null of H_0:β=1 with nominal value 0.05. ρ is AR(1) serial correlation coefficient for the errors.

Table C.8. Conditional logit estimates for β with serial correlation

ρ=0.6, n=100	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	2.86	11.06	3048013	0.001	(0,0.03)
$t = 3$	1.68	0.50	0.41	0.27	(0.25,0.29)
$t = 4$	1.49	0.29	0.27	0.39	(0.37,0.41)
$t = 5$	1.40	0.22	0.21	0.45	(0.42,0.47)
$t = 6$	1.34	0.18	0.18	0.47	(0.44,0.49)
$t = 7$	1.29	0.16	0.15	0.45	(0.43,0.47)
$t = 8$	1.25	0.14	0.14	0.44	(0.42,0.46)
ρ=0.6, n=200	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	1.97	0.83	0.58	0.23	(0.21,0.25)
$t = 3$	1.57	0.28	0.26	0.59	(0.57,0.61)
$t = 4$	1.44	0.19	0.18	0.72	(0.70,0.74)
$t = 5$	1.36	0.15	0.14	0.73	(0.71,0.75)
$t = 6$	1.31	0.13	0.12	0.75	(0.73,0.77)
$t = 7$	1.27	0.11	0.11	0.71	(0.69,0.73)
$t = 8$	1.24	0.10	0.09	0.71	(0.69,0.73)
ρ=0.6, n=400	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	1.77	0.37	0.34	0.68	(0.67,0.70)
$t = 3$	1.52	0.19	0.18	0.89	(0.88,0.90)
$t = 4$	1.42	0.13	0.13	0.95	(0.94,0.96)
$t = 5$	1.35	0.10	0.10	0.96	(0.95,0.97)
$t = 6$	1.30	0.09	0.08	0.95	(0.94,0.96)
$t = 7$	1.26	0.07	0.07	0.95	(0.94,0.96)
$t = 8$	1.23	0.07	0.07	0.96	(0.94,0.97)
ρ=0.6, n=600	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	1.71	0.29	0.27	0.85	(0.83,0.87)
$t = 3$	1.51	0.15	0.14	0.98	(0.98,0.99)
$t = 4$	1.41	0.10	0.10	1	(0.99,1)
$t = 5$	1.34	0.08	0.08	1	(0.99,1)
$t = 6$	1.29	0.07	0.07	0.99	(0.99,1)
$t = 7$	1.25	0.06	0.06	0.99	(0.99,1)
$t = 8$	1.22	0.06	0.05	0.99	(0.98,0.99)

*Note: r is the rejection rate for the null of H_0:β=1 with nominal value 0.05. ρ is AR(1) serial correlation coefficient for the errors.

Table C.9. Unconditional logit estimates for β with serial correlation

ρ=0.8, n=100	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	8.70	17.11	3.05	0.67	(0.65,0.68)
$t = 3$	6.53	25.88	1.67	0.91	(0.89,0.92)
$t = 4$	3.34	0.90	0.75	0.98	(0.98,0.99)
$t = 5$	2.72	0.52	0.50	1	(0.99,1)
$t = 6$	2.36	0.39	0.36	1	(0.99,1)
$t = 7$	2.13	0.31	0.29	1	(0.99,1)
$t = 8$	1.96	0.25	0.24	1	(0.99,1)
ρ=0.8, n=200	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	7.12	12.38	2.33	0.93	(0.92,0.94)
$t = 3$	4.05	0.92	0.83	1	(1,1)
$t = 4$	3.13	0.53	0.48	1	(1,1)
$t = 5$	2.61	0.34	0.33	1	(1,1)
$t = 6$	2.26	0.25	0.24	1	(1,1)
$t = 7$	2.06	0.20	0.20	1	(1,1)
$t = 8$	1.91	0.17	0.17	1	(1,1)
ρ=0.8, n=400	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	5.50	2.28	1.25	1	(0.99,1)
$t = 3$	3.81	0.58	0.54	1	(1,1)
$t = 4$	3.03	0.34	0.33	1	(1,1)
$t = 5$	2.55	0.23	0.23	1	(1,1)
$t = 6$	2.24	0.18	0.17	1	(1,1)
$t = 7$	2.04	0.14	0.14	1	(1,1)
$t = 8$	1.89	0.12	0.12	1	(1,1)
ρ=0.8, n=600	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	5.09	1.07	0.92	1	(1,1)
$t = 3$	3.74	0.46	0.43	1	(1,1)
$t = 4$	3.00	0.27	0.26	1	(1,1)
$t = 5$	2.52	0.18	0.18	1	(1,1)
$t = 6$	2.23	0.14	0.14	1	(1,1)
$t = 7$	2.03	0.11	0.11	1	(1,1)
$t = 8$	1.89	0.09	0.09	1	(1,1)

*Note: r is the rejection rate for the null of H_0:β=1 with nominal value 0.05. ρ is AR(1) serial correlation coefficient for the errors.

Table C.10. Conditional logit estimates for β with serial correlation

ρ=0.8, n=100	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	5.52	29.96	6296147	0	(0,0)
$t = 3$	4.03	19.14	380015	0.61	(0.59,0.63)
$t = 4$	2.12	0.50	0.42	0.89	(0.87,0.90)
$t = 5$	1.91	0.32	0.30	0.95	(0.94,0.97)
$t = 6$	1.78	0.26	0.24	0.96	(0.95,0.97)
$t = 7$	1.69	0.22	0.20	0.96	(0.95,0.96)
$t = 8$	1.61	0.19	0.18	0.97	(0.96,0.97)
ρ=0.8, n=200	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	3.69	11.54	1053342	0.27	(0.25,0.29)
$t = 3$	2.29	0.48	0.44	0.98	(0.98,0.99)
$t = 4$	2.01	0.29	0.27	1	(1,1)
$t = 5$	1.84	0.21	0.20	1	(1,1)
$t = 6$	1.72	0.17	0.16	1	(1,1)
$t = 7$	1.64	0.14	0.14	1	(1,1)
$t = 8$	1.57	0.13	0.12	1	(1,1)
ρ=0.8, n=400	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	2.75	1.14	0.68	0.98	(0.98,0.99)
$t = 3$	2.17	0.30	0.28	1	(1,1)
$t = 4$	1.95	0.19	0.18	1	(1,1)
$t = 5$	1.81	0.14	0.14	1	(1,1)
$t = 6$	1.70	0.12	0.11	1	(1,1)
$t = 7$	1.63	0.10	0.10	1	(1,1)
$t = 8$	1.56	0.09	0.08	1	(1,1)
ρ=0.8, n=600	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	2.55	0.54	0.47	1	(1,1)
$t = 3$	2.13	0.24	0.22	1	(1,1)
$t = 4$	1.94	0.15	0.15	1	(1,1)
$t = 5$	1.79	0.11	0.11	1	(1,1)
$t = 6$	1.69	0.09	0.09	1	(1,1)
$t = 7$	1.62	0.08	0.08	1	(1,1)
$t = 8$	1.56	0.07	0.07	1	(1,1)

*Note: r is the rejection rate for the null of H_0:β=1 with nominal value 0.05. ρ is AR(1) serial correlation coefficient for the errors.

Appendix D

Monte Carlo simulation results: APEs for continuous covariate

Table D.1. LPM, unconditional logit, CRE logit APE estimates for β

n=100	ρ=0	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t=2$	LPM	0.157	0.040	0.040	0.058	(0.047,0.068)
$t=2$	uncon logit	0.348	0.081	0.048	0.869	(0.854,0.884)
$t=2$	CRE logit	0.158	0.040	0.040	0.059	(0.048,0.068)
$t=3$	LPM	0.161	0.028	0.028	0.058	(0.048,0.068)
$t=2$	uncon logit	0.276	0.044	0.034	0.833	(0.816,0.849)
$t=3$	CRE logit	0.162	0.028	0.028	0.057	(0.056,0.068)
$t=4$	LPM	0.163	0.023	0.023	0.049	(0.040,0.058)
$t=4$	uncon logit	0.235	0.032	0.027	0.692	(0.672,0.712)
$t=4$	CRE logit	0.163	0.023	0.023	0.053	(0.043,0.063)
$t=5$	LPM	0.164	0.020	0.020	0.056	(0.046,0.066)
$t=5$	uncon logit	0.215	0.025	0.022	0.586	(0.048,0.068)
$t=5$	CRE logit	0.164	0.020	0.019	0.060	(0.050,0.071)
$t=6$	LPM	0.165	0.018	0.018	0.050	(0.040,0.059)
$t=6$	uncon logit	0.202	0.021	0.019	0.453	(0.432,0.475)
$t=6$	CRE logit	0.165	0.018	0.018	0.069	(0.040,0.058)
$t=7$	LPM	0.166	0.017	0.016	0.053	(0.041,0.060)
$t=7$	uncon logit	0.194	0.019	0.017	0.392	(0.370,0.413)
$t=7$	CRE logit	0.166	0.017	0.016	0.059	(0.049,0.069)
$t=8$	LPM	0.166	0.016	0.015	0.051	(0.041,0.060)
$t=8$	uncon logit	0.194	0.017	0.015	0.332	(0.311,0.352)
$t=8$	CRE logit	0.166	0.015	0.015	0.056	(0.046,0.066)

*Note: r is the rejection rate for the null of H_0:β=1 . ρ is AR(1) serial correlation coefficient for the errors. True APE is 0.162($t=2$), 0.162($t=3$), 0.165($t=4$), 0.165($t=5$), 0.167($t=6$), 0.166($t=7$), and 0.166($t=8$).

Table D.2. LPM, unconditional logit, CRE logit APE estimates for β

n=200	ρ=0	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	LPM	0.157	0.028	0.028	0.059	(0.049,0.069)
$t = 2$	uncon logit	0.349	0.050	0.032	0.979	(0.973,0.985)
$t = 2$	CRE logit	0.157	0.028	0.028	0.060	(0.049,0.070)
$t = 3$	LPM	0.161	0.020	0.020	0.053	(0.043,0.063)
$t = 3$	uncon logit	0.273	0.031	0.024	0.973	(0.966,0.980)
$t = 3$	CRE logit	0.161	0.020	0.020	0.055	(0.045,0.065)
$t = 4$	LPM	0.163	0.016	0.016	0.053	(0.043,0.062)
$t = 4$	uncon logit	0.236	0.022	0.019	0.927	(0.915,0.938)
$t = 4$	CRE logit	0.163	0.016	0.016	0.058	(0.048,0.068)
$t = 5$	LPM	0.164	0.014	0.014	0.057	(0.047,0.067)
$t = 5$	uncon logit	0.214	0.028	0.015	0.851	(0.834,0.866)
$t = 5$	CRE logit	0.164	0.014	0.014	0.056	(0.045,0.066)
$t = 6$	LPM	0.165	0.013	0.013	0.054	(0.044,0.063)
$t = 6$	uncon logit	0.201	0.015	0.013	0.753	(0.734,0.772)
$t = 6$	CRE logit	0.165	0.013	0.013	0.055	(0.045,0.065)
$t = 7$	LPM	0.166	0.011	0.011	0.047	(0.037,0.056)
$t = 7$	uncon logit	0.194	0.013	0.012	0.658	(0.637,0.678)
$t = 7$	CRE logit	0.166	0.011	0.011	0.049	(0.039,0.058)
$t = 8$	LPM	0.166	0.011	0.011	0.060	(0.049,0.070)
$t = 8$	uncon logit	0.188	0.011	0.011	0.622	(0.601,0.643)
$t = 8$	CRE logit	0.166	0.011	0.011	0.059	(0.049,0.069)

*Note: r is the rejection rate for the null of H_0:β=1 . ρ is AR(1) serial correlation coefficient for the errors. True APE is 0.163($t = 2$), 0.163($t = 3$), 0.164($t = 4$), 0.165($t = 5$), 0.165($t = 6$), 0.166($t = 7$), and 0.164($t = 8$).

Table D.3. LPM, unconditional logit, CRE logit APE estimates for β

n=400	ρ=0	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t=2$	LPM	0.156	0.020	0.020	0.063	(0.051,0.073)
$t=2$	uncon logit	0.347	0.033	0.023	1	(1,1)
$t=2$	CRE logit	0.156	0.020	0.020	0.063	(0.052,0.073)
$t=3$	LPM	0.160	0.014	0.013	0.060	(0.049,0.070)
$t=3$	uncon logit	0.273	0.022	0.017	1	(1,1)
$t=3$	CRE logit	0.160	0.014	0.014	0.055	(0.045,0.065)
$t=4$	LPM	0.163	0.011	0.011	0.056	(0.046,0.066)
$t=4$	uncon logit	0.235	0.015	0.013	0.997	(0.994,0.999)
$t=4$	CRE logit	0.163	0.011	0.011	0.056	(0.046,0.066)
$t=5$	LPM	0.164	0.010	0.010	0.056	(0.046,0.067)
$t=5$	uncon logit	0.214	0.012	0.011	0.986	(0.980,0.991)
$t=5$	CRE logit	0.164	0.010	0.010	0.056	(0.046,0.067)
$t=6$	LPM	0.165	0.009	0.009	0.044	(0.035,0.060)
$t=6$	uncon logit	0.202	0.010	0.009	0.962	(0.953,0.970)
$t=6$	CRE logit	0.165	0.009	0.009	0.048	(0.039,0.062)
$t=7$	LPM	0.165	0.008	0.008	0.054	(0.044,0.064)
$t=7$	uncon logit	0.193	0.009	0.008	0.847	(0.831,0.865)
$t=7$	CRE logit	0.165	0.008	0.008	0.056	(0.046,0.066)
$t=8$	LPM	0.166	0.007	0.007	0.050	(0.040,0.060)
$t=8$	uncon logit	0.188	0.008	0.008	0.800	(0.780,0.820)
$t=8$	CRE logit	0.166	0.007	0.007	0.053	(0.043,0.063)

*Note: r is the rejection rate for the null of H_0:β=1 . ρ is AR(1) serial correlation coefficient for the errors. True APE is 0.161($t=2$), 0.162($t=3$), 0.166($t=4$), 0.166($t=5$), 0.165($t=6$), 0.167($t=7$), and 0.167($t=8$).

Table D.4. LPM, unconditional logit, CRE logit APE estimates for β

n=600	ρ=0	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t=2$	LPM	0.156	0.016	0.016	0.101	(0.082,0.119)
$t=2$	uncon logit	0.346	0.028	0.018	1	(1,1)
$t=2$	CRE logit	0.156	0.016	0.016	0.097	(0.078,0.115)
$t=3$	LPM	0.160	0.012	0.011	0.062	(0.047,0.076)
$t=3$	uncon logit	0.273	0.018	0.014	1	(1,1)
$t=3$	CRE logit	0.160	0.012	0.011	0.062	(0.047,0.076)
$t=4$	LPM	0.163	0.009	0.009	0.056	(0.041,0.072)
$t=4$	uncon logit	0.235	0.013	0.011	1	(1,1)
$t=4$	CRE logit	0.163	0.009	0.009	0.060	(0.045,0.076)
$t=5$	LPM	0.164	0.008	0.008	0.059	(0.044,0.073)
$t=5$	uncon logit	0.214	0.010	0.009	1	(1,1)
$t=5$	CRE logit	0.164	0.008	0.008	0.059	(0.045,0.073)
$t=6$	LPM	0.165	0.007	0.007	0.047	(0.034,0.060)
$t=6$	uncon logit	0.202	0.008	0.007	0.998	(0.995,1)
$t=6$	CRE logit	0.165	0.007	0.007	0.047	(0.034,0.060)
$t=7$	LPM	0.165	0.007	0.007	0.055	(0.041,0.069)
$t=7$	uncon logit	0.194	0.007	0.007	0.997	(0.967,0.986)
$t=7$	CRE logit	0.165	0.007	0.007	0.058	(0.044,0.073)
$t=8$	LPM	0.166	0.006	0.006	0.049	(0.036,0.062)
$t=8$	uncon logit	0.188	0.006	0.006	0.931	(0.915,0.947)
$t=8$	CRE logit	0.166	0.006	0.006	0.055	(0.040,0.069)

*Note: r is the rejection rate for the null of H_0:β=1 . ρ is AR(1) serial correlation coefficient for the errors. True APE is 0.165($t=2$), 0.163($t=3$), 0.167($t=4$), 0.165($t=5$), 0.164($t=6$), 0.167($t=7$), and 0.167($t=8$).

Table D.5. LPM, unconditional logit, CRE logit APE estimates for β

n=100	ρ=0.2	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t=2$	LPM	0.156	0.039	0.038	0.068	(0.057,0.078)
$t=2$	uncon logit	0.362	0.083	0.047	0.906	(0.893,0.918)
$t=2$	CRE logit	0.157	0.039	0.038	0.070	(0.060,0.081)
$t=3$	LPM	0.160	0.027	0.027	0.056	(0.045,0.066)
$t=2$	uncon logit	0.289	0.047	0.034	0.889	(0.875,0.903)
$t=3$	CRE logit	0.161	0.027	0.027	0.057	(0.046,0.067)
$t=4$	LPM	0.162	0.023	0.023	0.057	(0.046,0.067)
$t=4$	uncon logit	0.246	0.033	0.027	0.802	(0.785,0.820)
$t=4$	CRE logit	0.163	0.023	0.023	0.057	(0.047,0.067)
$t=5$	LPM	0.164	0.020	0.020	0.053	(0.043,0.064)
$t=5$	uncon logit	0.223	0.025	0.022	0.713	(0.693,0.733)
$t=5$	CRE logit	0.164	0.020	0.019	0.053	(0.043,0.063)
$t=6$	LPM	0.165	0.018	0.018	0.055	(0.045,0.065)
$t=6$	uncon logit	0.208	0.021	0.019	0.614	(0.477,0.521)
$t=6$	CRE logit	0.165	0.018	0.018	0.063	(0.052,0.074)
$t=7$	LPM	0.165	0.016	0.016	0.050	(0.040,0.060)
$t=7$	uncon logit	0.198	0.018	0.017	0.499	(0.477,0.521)
$t=7$	CRE logit	0.165	0.016	0.016	0.056	(0.046,0.067)
$t=8$	LPM	0.166	0.015	0.015	0.046	(0.036,0.055)
$t=8$	uncon logit	0.193	0.016	0.015	0.426	(0.404,0.448)
$t=8$	CRE logit	0.166	0.015	0.015	0.050	(0.040,0.059)

*Note: r is the rejection rate for the null of H_0:β=1 . ρ is AR(1) serial correlation coefficient for the errors. True APE is 0.163($t=2$), 0.162($t=3$), 0.164($t=4$), 0.165($t=5$), 0.165($t=6$), 0.166($t=7$), and 0.166($t=8$).

Table D.6. LPM, unconditional logit, CRE logit APE estimates for β

n=200	ρ=0.2	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	LPM	0.157	0.027	0.028	0.059	(0.049,0.069)
$t = 2$	uncon logit	0.363	0.050	0.032	0.979	(0.973,0.985)
$t = 2$	CRE logit	0.157	0.028	0.028	0.060	(0.049,0.070)
$t = 3$	LPM	0.161	0.019	0.019	0.049	(0.039,0.058)
$t = 3$	uncon logit	0.289	0.031	0.024	0.995	(0.992,0.998)
$t = 3$	CRE logit	0.161	0.019	0.019	0.054	(0.044,0.064)
$t = 4$	LPM	0.163	0.016	0.016	0.051	(0.041,0.061)
$t = 4$	uncon logit	0.246	0.022	0.019	0.978	(0.972,0.984)
$t = 4$	CRE logit	0.163	0.016	0.016	0.049	(0.039,0.058)
$t = 5$	LPM	0.164	0.014	0.014	0.049	(0.040,0.058)
$t = 5$	uncon logit	0.222	0.017	0.015	0.946	(0.936,0.955)
$t = 5$	CRE logit	0.164	0.014	0.014	0.050	(0.040,0.059)
$t = 6$	LPM	0.165	0.013	0.013	0.050	(0.040,0.059)
$t = 6$	uncon logit	0.208	0.014	0.013	0.858	(0.842,0.872)
$t = 6$	CRE logit	0.165	0.013	0.012	0.055	(0.045,0.065)
$t = 7$	LPM	0.166	0.011	0.011	0.051	(0.041,0.060)
$t = 7$	uncon logit	0.199	0.013	0.012	0.776	(0.758,0.794)
$t = 7$	CRE logit	0.166	0.011	0.011	0.056	(0.046,0.066)
$t = 8$	LPM	0.166	0.011	0.011	0.054	(0.044,0.064)
$t = 8$	uncon logit	0.192	0.012	0.011	0.664	(0.643,0.685)
$t = 8$	CRE logit	0.166	0.011	0.011	0.056	(0.046,0.066)

*Note: r is the rejection rate for the null of H_0:β=1 . ρ is AR(1) serial correlation coefficient for the errors. True APE is 0.163($t = 2$), 0.162($t = 3$), 0.164($t = 4$), 0.165($t = 5$), 0.166($t = 6$), 0.166($t = 7$), and 0.164($t = 8$).

Table D.7. LPM, unconditional logit, CRE logit APE estimates for β

n=400	ρ=0.2	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	LPM	0.156	0.019	0.019	0.056	(0.046,0.066)
$t = 2$	uncon logit	0.364	0.032	0.020	1	(1,1)
$t = 2$	CRE logit	0.156	0.019	0.019	0.056	(0.046,0.066)
$t = 3$	LPM	0.160	0.014	0.014	0.053	(0.043,0.063)
$t = 3$	uncon logit	0.288	0.022	0.017	1	(1,1)
$t = 3$	CRE logit	0.160	0.014	0.013	0.052	(0.042,0.062)
$t = 4$	LPM	0.163	0.011	0.011	0.055	(0.045,0.065)
$t = 4$	uncon logit	0.246	0.015	0.013	1	(1,1)
$t = 4$	CRE logit	0.163	0.011	0.011	0.058	(0.048,0.068)
$t = 5$	LPM	0.164	0.010	0.010	0.052	(0.041,0.063)
$t = 5$	uncon logit	0.222	0.012	0.011	0.998	(0.996,1)
$t = 5$	CRE logit	0.164	0.010	0.010	0.056	(0.046,0.067)
$t = 6$	LPM	0.165	0.010	0.009	0.052	(0.038,0.066)
$t = 6$	uncon logit	0.208	0.010	0.009	0.990	(0.984,0.996)
$t = 6$	CRE logit	0.165	0.009	0.009	0.054	(0.039,0.068)
$t = 7$	LPM	0.165	0.008	0.008	0.050	(0.040,0.060)
$t = 7$	uncon logit	0.198	0.009	0.008	0.972	(0.964,0.979)
$t = 7$	CRE logit	0.165	0.008	0.008	0.052	(0.042,0.062)
$t = 8$	LPM	0.166	0.008	0.008	0.058	(0.044,0.072)
$t = 8$	uncon logit	0.192	0.008	0.008	0.923	(0.907,0.939)
$t = 8$	CRE logit	0.166	0.008	0.007	0.056	(0.042,0.070)

*Note: r is the rejection rate for the null of H_0:β=1 . ρ is AR(1) serial correlation coefficient for the errors. True APE is 0.162($t = 2$), 0.162($t = 3$), 0.166($t = 4$), 0.165($t = 5$), 0.165($t = 6$), 0.165($t = 7$), and 0.166($t = 8$).

Table D.8. LPM, unconditional logit, CRE logit APE estimates for β

n=600	ρ=0.2	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t=2$	LPM	0.155	0.015	0.016	0.062	(0.047,0.077)
$t=2$	uncon logit	0.363	0.028	0.017	1	(1,1)
$t=2$	CRE logit	0.155	0.015	0.015	0.060	(0.045,0.075)
$t=3$	LPM	0.160	0.011	0.011	0.052	(0.039,0.066)
$t=3$	uncon logit	0.288	0.018	0.014	1	(1,1)
$t=3$	CRE logit	0.160	0.011	0.011	0.052	(0.039,0.066)
$t=4$	LPM	0.163	0.009	0.009	0.060	(0.045,0.075)
$t=4$	uncon logit	0.246	0.013	0.011	1	(1,1)
$t=4$	CRE logit	0.163	0.009	0.009	0.060	(0.045,0.075)
$t=5$	LPM	0.164	0.008	0.008	0.046	(0.033,0.059)
$t=5$	uncon logit	0.222	0.010	0.009	1	(1,1)
$t=5$	CRE logit	0.164	0.008	0.008	0.050	(0.036,0.064)
$t=6$	LPM	0.165	0.007	0.007	0.046	(0.033,0.059)
$t=6$	uncon logit	0.208	0.008	0.008	1	(1,1)
$t=6$	CRE logit	0.165	0.007	0.007	0.049	(0.036,0.062)
$t=7$	LPM	0.165	0.007	0.007	0.057	(0.042,0.071)
$t=7$	uncon logit	0.198	0.008	0.007	0.990	(0.984,1)
$t=7$	CRE logit	0.165	0.007	0.007	0.057	(0.042,0.071)
$t=8$	LPM	0.166	0.006	0.006	0.045	(0.032,0.058)
$t=8$	uncon logit	0.192	0.006	0.006	0.975	(0.965,0.985)
$t=8$	CRE logit	0.166	0.006	0.006	0.045	(0.032,0.058)

*Note: r is the rejection rate for the null of H_0:β=1 . ρ is AR(1) serial correlation coefficient for the errors. True APE is 0.165($t=2$), 0.163($t=3$), 0.165($t=4$), 0.164($t=5$), 0.164($t=6$), 0.167($t=7$), and 0.167($t=8$).

Table D.9. LPM, unconditional logit, CRE logit APE estimates for β

n=100	ρ=0.4	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t=2$	LPM	0.155	0.037	0.037	0.074	(0.063,0.086)
$t=2$	uncon logit	0.379	0.094	0.047	0.936	(0.925,0.946)
$t=2$	CRE logit	0.155	0.037	0.036	0.078	(0.066,0.089)
$t=3$	LPM	0.160	0.027	0.026	0.057	(0.047,0.068)
$t=2$	uncon logit	0.310	0.047	0.033	0.955	(0.946,0.964)
$t=3$	CRE logit	0.161	0.027	0.026	0.058	(0.048,0.068)
$t=4$	LPM	0.163	0.022	0.022	0.058	(0.048,0.068)
$t=4$	uncon logit	0.262	0.034	0.027	0.910	(0.897,0.923)
$t=4$	CRE logit	0.163	0.022	0.022	0.059	(0.049,0.070)
$t=5$	LPM	0.164	0.019	0.019	0.055	(0.045,0.065)
$t=5$	uncon logit	0.235	0.025	0.022	0.859	(0.844,0.874)
$t=5$	CRE logit	0.164	0.019	0.019	0.059	(0.049,0.069)
$t=6$	LPM	0.165	0.018	0.018	0.053	(0.043,0.063)
$t=6$	uncon logit	0.218	0.021	0.019	0.762	(0.743,0.781)
$t=6$	CRE logit	0.165	0.018	0.018	0.058	(0.049,0.069)
$t=7$	LPM	0.165	0.016	0.016	0.045	(0.036,0.054)
$t=7$	uncon logit	0.206	0.018	0.017	0.669	(0.648,0.690)
$t=7$	CRE logit	0.165	0.016	0.016	0.050	(0.041,0.060)
$t=8$	LPM	0.166	0.015	0.015	0.055	(0.045,0.065)
$t=8$	uncon logit	0.198	0.017	0.016	0.561	(0.539,0.583)
$t=8$	CRE logit	0.166	0.015	0.015	0.057	(0.047,0.067)

*Note: r is the rejection rate for the null of H_0:β=1 . ρ is AR(1) serial correlation coefficient for the errors. True APE is 0.161($t=2$), 0.162($t=3$), 0.164($t=4$), 0.165($t=5$), 0.166($t=6$), 0.165($t=7$), and 0.167($t=8$).

Table D.10. LPM, unconditional logit, CRE logit APE estimates for β

n=200	ρ=0.4	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t=2$	LPM	0.157	0.026	0.026	0.057	(0.046,0.067)
$t=2$	uncon logit	0.378	0.048	0.027	0.998	(0.996,1)
$t=2$	CRE logit	0.157	0.026	0.025	0.059	(0.048,0.069)
$t=3$	LPM	0.160	0.018	0.019	0.048	(0.039,0.057)
$t=3$	uncon logit	0.308	0.031	0.023	0.995	(0.996,1)
$t=3$	CRE logit	0.160	0.018	0.018	0.053	(0.043,0.063)
$t=4$	LPM	0.162	0.016	0.016	0.060	(0.049,0.070)
$t=4$	uncon logit	0.262	0.023	0.019	0.998	(0.982,0.995)
$t=4$	CRE logit	0.162	0.016	0.016	0.057	(0.047,0.067)
$t=5$	LPM	0.164	0.014	0.014	0.055	(0.040,0.070)
$t=5$	uncon logit	0.235	0.018	0.016	0.989	(0.982,0.995)
$t=5$	CRE logit	0.164	0.014	0.013	0.051	(0.047,0.076)
$t=6$	LPM	0.165	0.012	0.012	0.058	(0.047,0.068)
$t=6$	uncon logit	0.218	0.015	0.013	0.953	(0.943,0.962)
$t=6$	CRE logit	0.165	0.012	0.012	0.056	(0.046,0.076)
$t=7$	LPM	0.165	0.011	0.011	0.053	(0.043,0.063)
$t=7$	uncon logit	0.206	0.013	0.012	0.906	(0.893,0.919)
$t=7$	CRE logit	0.165	0.011	0.011	0.051	(0.041,0.060)
$t=8$	LPM	0.166	0.011	0.011	0.055	(0.045,0.065)
$t=8$	uncon logit	0.198	0.012	0.011	0.834	(0.818,0.851)
$t=8$	CRE logit	0.166	0.011	0.011	0.059	(0.049,0.070)

*Note: r is the rejection rate for the null of H_0:β=1 . ρ is AR(1) serial correlation coefficient for the errors. True APE is 0.163($t=2$), 0.161($t=3$), 0.165($t=4$), 0.163($t=5$), 0.166($t=6$), 0.165($t=7$), and 0.164($t=8$).

Table D.11. LPM, unconditional logit, CRE logit APE estimates for β

n=400	ρ=0.4	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t=2$	LPM	0.156	0.019	0.018	0.068	(0.056,0.079)
$t=2$	uncon logit	0.379	0.032	0.018	1	(1,1)
$t=2$	CRE logit	0.156	0.019	0.018	0.066	(0.056,0.076)
$t=3$	LPM	0.160	0.013	0.013	0.050	(0.040,0.060)
$t=3$	uncon logit	0.307	0.022	0.016	1	(1,1)
$t=3$	CRE logit	0.160	0.013	0.013	0.054	(0.044,0.064)
$t=4$	LPM	0.163	0.011	0.011	0.046	(0.037,0.055)
$t=4$	uncon logit	0.262	0.016	0.013	1	(1,1)
$t=4$	CRE logit	0.163	0.011	0.011	0.048	(0.038,0.057)
$t=5$	LPM	0.164	0.010	0.010	0.054	(0.044,0.064)
$t=5$	uncon logit	0.234	0.012	0.011	1	(1,1)
$t=5$	CRE logit	0.164	0.010	0.010	0.056	(0.046,0.067)
$t=6$	LPM	0.165	0.009	0.009	0.048	(0.035,0.060)
$t=6$	uncon logit	0.217	0.010	0.010	1	(1,1)
$t=6$	CRE logit	0.165	0.009	0.009	0.048	(0.035,0.060)
$t=7$	LPM	0.165	0.008	0.008	0.050	(0.040,0.060)
$t=7$	uncon logit	0.206	0.009	0.008	1	(1,1)
$t=7$	CRE logit	0.165	0.008	0.008	0.051	(0.041,0.061)
$t=8$	LPM	0.166	0.008	0.008	0.056	(0.046,0.076)
$t=8$	uncon logit	0.199	0.008	0.008	0.982	(0.972,0.991)
$t=8$	CRE logit	0.166	0.008	0.007	0.055	(0.045,0.075)

*Note: r is the rejection rate for the null of H_0:β=1 . ρ is AR(1) serial correlation coefficient for the errors. True APE is 0.162($t=2$), 0.162($t=3$), 0.166($t=4$), 0.165($t=5$), 0.166($t=6$), 0.166($t=7$), and 0.165($t=8$).

Table D.12. LPM, unconditional logit, CRE logit APE estimates for β

n=600	ρ=0.4	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t=2$	LPM	0.155	0.015	0.015	0.071	(0.055,0.086)
$t=2$	uncon logit	0.378	0.026	0.014	1	(1,1)
$t=2$	CRE logit	0.155	0.015	0.015	0.074	(0.057,0.090)
$t=3$	LPM	0.160	0.011	0.011	0.052	(0.039,0.066)
$t=3$	uncon logit	0.309	0.017	0.013	1	(1,1)
$t=3$	CRE logit	0.160	0.011	0.011	0.054	(0.041,0.068)
$t=4$	LPM	0.163	0.009	0.009	0.045	(0.032,0.057)
$t=4$	uncon logit	0.262	0.012	0.011	1	(1,1)
$t=4$	CRE logit	0.163	0.009	0.009	0.048	(0.035,0.061)
$t=5$	LPM	0.164	0.008	0.008	0.045	(0.032,0.057)
$t=5$	uncon logit	0.234	0.010	0.009	1	(1,1)
$t=5$	CRE logit	0.164	0.008	0.008	0.048	(0.035,0.061)
$t=6$	LPM	0.165	0.007	0.007	0.046	(0.033,0.059)
$t=6$	uncon logit	0.217	0.008	0.008	1	(1,1)
$t=6$	CRE logit	0.165	0.007	0.007	0.043	(0.030,0.056)
$t=7$	LPM	0.165	0.007	0.007	0.054	(0.040,0.068)
$t=7$	uncon logit	0.206	0.007	0.007	1	(1,1)
$t=7$	CRE logit	0.165	0.007	0.007	0.057	(0.042,0.071)
$t=8$	LPM	0.166	0.006	0.006	0.044	(0.031,0.057)
$t=8$	uncon logit	0.199	0.006	0.006	1	(1,1)
$t=8$	CRE logit	0.166	0.006	0.006	0.043	(0.031,0.055)

*Note: r is the rejection rate for the null of H_0:β=1 . ρ is AR(1) serial correlation coefficient for the errors. True APE is 0.165($t=2$), 0.162($t=3$), 0.163($t=4$), 0.165($t=5$), 0.164($t=6$), 0.166($t=7$), and 0.166($t=8$).

Appendix E

Monte Carlo simulation results:

Coefficient δ for binary covariate

Table E.1. Unconditional logit for δ with low serial correlation for $n = 100$

ρ=0	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	2.57	3.04	1.22	0.16	(0.13,0.18)
$t = 3$	1.65	0.63	0.62	0.13	(0.11,0.15)
$t = 4$	1.41	0.43	0.42	0.14	(0.12,0.16)
$t = 5$	1.29	0.33	0.34	0.12	(0.10,0.14)
$t = 6$	1.26	0.29	0.29	0.14	(0.12,0.16)
$t = 7$	1.20	0.25	0.25	0.11	(0.09,0.13)
$t = 8$	1.18	0.23	0.23	0.11	(0.00,0.13)
ρ=0.2	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	3.05	3.80	7.93	0.18	(0.16,0.21)
$t = 3$	1.83	0.69	0.66	0.18	(0.16,0.21)
$t = 4$	1.51	0.45	0.45	0.17	(0.15,0.20)
$t = 5$	1.37	0.35	0.35	0.15	(0.13,0.17)
$t = 6$	1.32	0.30	0.30	0.17	(0.15,0.19)
$t = 7$	1.26	0.26	0.36	0.16	(0.13,0.18)
$t = 8$	1.22	0.24	0.24	0.14	(0.12,0.16)
ρ=0.4	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	4.69	30.20	1.73	0.21	(0.18,0.24)
$t = 3$	2.12	0.77	0.75	0.26	(0.24,0.29)
$t = 4$	1.69	0.50	0.48	0.27	(0.24,0.30)
$t = 5$	1.50	0.37	0.37	0.24	(0.22,0.26)
$t = 6$	1.41	0.32	0.32	0.25	(0.23,0.28)
$t = 7$	1.35	0.28	0.28	0.23	(0.20,0.25)
$t = 8$	1.29	0.25	0.25	0.20	(0.18,0.23)

*Note: r is the rejection rate for the null of H_0:δ=1 with nominal value 0.05. ρ is AR(1) serial correlation coefficient for the errors.

Table E.2. Unconditional logit for δ with high serial correlation for $n = 100$

ρ=0.6	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	7.43	35.86	6.24	0.27	(0.24,0.30)
$t = 3$	2.69	1.52	0.92	0.40	(0.37,0.43)
$t = 4$	2.04	0.57	0.56	0.45	(0.42,0.48)
$t = 5$	1.76	0.43	0.42	0.42	(0.39,0.45)
$t = 6$	1.62	0.35	0.34	0.43	(0.40,0.46)
$t = 7$	1.52	0.30	0.30	0.41	(0.38,0.44)
$t = 8$	1.43	0.26	0.26	0.37	(0.34,0.40)
ρ=0.8	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	146.09	46.19	20.32	0.34	(0.31,0.38)
$t = 3$	6.15	23.83	1.77	0.63	(0.60,0.66)
$t = 4$	3.03	1.37	0.81	0.78	(0.75,0.80)
$t = 5$	2.45	0.58	0.56	0.80	(0.77,0.82)
$t = 6$	2.18	0.47	0.44	0.80	(0.77,0.83)
$t = 7$	2.00	0.38	0.37	0.79	(0.76,0.81)
$t = 8$	1.86	0.33	0.32	0.77	(0.74,0.80)

*Note: r is the rejection rate for the null of H_0:δ=1 with nominal value 0.05. ρ is AR(1) serial correlation coefficient for the errors.

Table E.3. Conditional logit for δ with low serial correlation for $n = 100$

ρ=0	Mean	SD	SE	r $\equiv 1$(P<0.05)	95-CI
$t = 2$	1.29	1.55	19.40	0.03	(0.01,0.04)
$t = 3$	1.06	0.39	0.38	0.05	(0.03,0.06)
$t = 4$	1.04	0.31	0.30	0.06	(0.04,0.07)
$t = 5$	1.02	0.26	0.26	0.04	(0.03,0.05)
$t = 6$	1.04	0.23	0.23	0.05	(0.04,0.07)
$t = 7$	1.02	0.21	0.21	0.05	(0.03,0.06)
$t = 8$	1.02	0.19	0.19	0.04	(0.03,0.06)
ρ=0.2	Mean	SD	SE	r $\equiv 1$(P<0.05)	95-CI
$t = 2$	1.53	1.94	38.55	0.02	(0.01,0.03)
$t = 3$	1.17	0.43	0.18	0.06	(0.04,0.07)
$t = 4$	1.10	0.32	0.14	0.05	(0.04,0.06)
$t = 5$	1.08	0.27	0.27	0.05	(0.03,0.06)
$t = 6$	1.07	0.24	0.24	0.06	(0.04,0.07)
$t = 7$	1.06	0.22	0.22	0.05	(0.04,0.06)
$t = 8$	1.06	0.21	0.20	0.05	(0.03,0.06)
ρ=0.4	Mean	SD	SE	r $\equiv 1$(P<0.05)	95-CI
$t = 2$	4.32	75.84	68.47	0.02	(0.01,0.03)
$t = 3$	1.34	0.46	0.45	0.07	(0.05,0.08)
$t = 4$	1.23	0.35	0.34	0.08	(0.07,0.10)
$t = 5$	1.17	0.28	0.29	0.08	(0.06,0.10)
$t = 6$	1.16	0.25	0.25	0.09	(0.07,0.10)
$t = 7$	1.14	0.23	0.23	0.08	(0.07,0.10)
$t = 8$	1.12	0.21	0.21	0.08	(0.06,0.10)

*Note: r is the rejection rate for the null of H_0:δ=1 with nominal value 0.05.
ρ is AR(1) serial correlation coefficient for the errors.

Table E.4. Conditional logit for δ with high serial correlation for $n = 100$

ρ=0.6	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	5.23	45.89	326.60	0.02	(0.01,0.02)
$t = 3$	1.67	0.82	3.24	0.17	(0.14,0.19)
$t = 4$	1.47	0.39	0.38	0.19	(0.17,0.22)
$t = 5$	1.37	0.32	0.32	0.19	(0.17,0.22)
$t = 6$	1.33	0.28	0.27	0.20	(0.17,0.23)
$t = 7$	1.28	0.25	0.25	0.19	(0.17,0.21)
$t = 8$	1.24	0.23	0.23	0.17	(0.14,0.19)
ρ=0.8	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t = 2$	253.9	80.97	1255596	0.003	(0.00,0.006)
$t = 3$	3.48	11.39	155.10	0.38	(0.35,0.41)
$t = 4$	2.11	0.78	2.58	0.56	(0.53,0.59)
$t = 5$	1.88	0.41	0.40	0.59	(0.56,0.62)
$t = 6$	1.76	0.36	0.34	0.61	(0.58,0.64)
$t = 7$	1.68	0.31	0.30	0.63	(0.60,0.66)
$t = 8$	1.60	0.28	0.27	0.62	(0.59,0.65)

*Note: r is the rejection rate for the null of H_0:δ=1 with nominal value 0.05. ρ is AR(1) serial correlation coefficient for the errors.

Appendix F

Monte Carlo simulation results: APEs for binary covariate

Table F.1. LPM, unconditional logit, CRE logit APE estimates for δ

n=100	ρ=0	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t=2$	LPM	0.184	0.089	0.084	0.063	(0.048,0.086)
$t=2$	uncon logit	0.454	0.373	1834.10	0.564	(0.534,0.595)
$t=2$	CRE logit	0.174	0.084	0.036	0.065	(0.049,0.080)
$t=3$	LPM	0.176	0.059	0.059	0.050	(0.036,0.063)
$t=2$	uncon logit	0.311	0.100	0.078	0.483	(0.453,0.513)
$t=3$	CRE logit	0.168	0.055	0.055	0.052	(0.038,0.065)
$t=4$	LPM	0.174	0.049	0.049	0.061	(0.046,0.076)
$t=4$	uncon logit	0.260	0.069	0.059	0.391	(0.361,0.421)
$t=4$	CRE logit	0.168	0.046	0.046	0.053	(0.038,0.068)
$t=5$	LPM	0.167	0.041	0.041	0.056	(0.042,0.070)
$t=5$	uncon logit	0.231	0.053	0.048	0.331	(0.302,0.360)
$t=5$	CRE logit	0.162	0.039	0.039	0.056	(0.042,0.070)
$t=6$	LPM	0.170	0.038	0.037	0.051	(0.037,0.065)
$t=6$	uncon logit	0.221	0.046	0.041	0.306	(0.277,0.335)
$t=6$	CRE logit	0.166	0.036	0.035	0.052	(0.038,0.066)
$t=7$	LPM	0.163	0.033	0.034	0.043	(0.030,0.056)
$t=7$	uncon logit	0.205	0.039	0.037	0.253	(0.226,0.280)
$t=7$	CRE logit	0.160	0.031	0.031	0.048	(0.035,0.061)
$t=8$	LPM	0.164	0.031	0.031	0.051	(0.037,0.065)
$t=8$	uncon logit	0.198	0.035	0.033	0.246	(0.219,0.273)
$t=8$	CRE logit	0.160	0.030	0.030	0.058	(0.043,0.073)

*Note: r is the rejection rate for the null of H_0:APE for δ=true APE . ρ is AR(1) serial correlation coefficient for the errors. True APE is 0.172($t=2$), 0.166($t=3$), 0.168($t=4$), 0.164($t=5$), 0.166($t=6$), 0.162($t=7$), and 0.160($t=8$).

Table F.2. LPM, unconditional logit, CRE logit APE estimates for δ

n=100	ρ=0.2	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t=2$	LPM	0.184	0.081	0.080	0.068	(0.052,0.084)
$t=2$	uncon logit	0.501	0.397	13231	0.616	(0.586,0.647)
$t=2$	CRE logit	0.174	0.077	0.075	0.068	(0.052,0.084)
$t=3$	LPM	0.176	0.057	0.057	0.050	(0.036,0.064)
$t=2$	uncon logit	0.335	0.104	0.080	0.547	(0.516,0.578)
$t=3$	CRE logit	0.168	0.054	0.053	0.050	(0.036,0.064)
$t=4$	LPM	0.172	0.047	0.047	0.051	(0.038,0.065)
$t=4$	uncon logit	0.273	0.069	0.060	0.439	(0.408,0.470)
$t=4$	CRE logit	0.168	0.044	0.044	0.058	(0.043,0.073)
$t=5$	LPM	0.167	0.040	0.041	0.050	(0.037,0.064)
$t=5$	uncon logit	0.243	0.053	0.049	0.280	(0.350,0.410)
$t=5$	CRE logit	0.163	0.038	0.038	0.056	(0.043,0.070)
$t=6$	LPM	0.170	0.037	0.037	0.056	(0.037,0.070)
$t=6$	uncon logit	0.228	0.046	0.042	0.360	(0.330,0.390)
$t=6$	CRE logit	0.166	0.035	0.035	0.051	(0.037,0.065)
$t=7$	LPM	0.164	0.032	0.033	0.046	(0.033,0.059)
$t=7$	uncon logit	0.213	0.040	0.037	0.298	(0.270,0.326)
$t=7$	CRE logit	0.160	0.031	0.031	0.051	(0.037,0.065)
$t=8$	LPM	0.164	0.031	0.031	0.049	(0.036,0.062)
$t=8$	uncon logit	0.204	0.036	0.033	0.294	(0.266,0.322)
$t=8$	CRE logit	0.160	0.029	0.029	0.051	(0.037,0.065)

*Note: r is the rejection rate for the null of H_0:APE for δ=true APE . ρ is AR(1) serial correlation coefficient for the errors. True APE is 0.172($t=2$), 0.166($t=3$), 0.168($t=4$), 0.164($t=5$), 0.166($t=6$), 0.161($t=7$), and 0.159($t=8$).

Table F.3. LPM, unconditional logit, CRE logit APE estimates for δ

n=100	ρ=0.4	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t=2$	LPM	0.182	0.076	0.075	0.062	(0.047,0.077)
$t=2$	uncon logit	0.566	0.528	17430	0.677	(0.647,0.706)
$t=2$	CRE logit	0.173	0.072	0.070	0.062	(0.047,0.077)
$t=3$	LPM	0.176	0.053	0.054	0.050	(0.039,0.061)
$t=2$	uncon logit	0.372	0.105	0.082	0.652	(0.627,0.677)
$t=3$	CRE logit	0.169	0.050	0.050	0.049	(0.038,0.060)
$t=4$	LPM	0.173	0.045	0.045	0.058	(0.043,0.073)
$t=4$	uncon logit	0.298	0.071	0.061	0.555	(0.524,0.586)
$t=4$	CRE logit	0.167	0.042	0.042	0.058	(0.043,0.073)
$t=5$	LPM	0.167	0.038	0.039	0.044	(0.031,0.057)
$t=5$	uncon logit	0.261	0.054	0.050	0.477	(0.446,0.508)
$t=5$	CRE logit	0.162	0.036	0.037	0.058	(0.034,0.061)
$t=6$	LPM	0.169	0.036	0.036	0.059	(0.044,0.074)
$t=6$	uncon logit	0.242	0.046	0.042	0.436	(0.405,0.467)
$t=6$	CRE logit	0.165	0.034	0.034	0.059	(0.044,0.074)
$t=7$	LPM	0.164	0.032	0.032	0.046	(0.033,0.059)
$t=7$	uncon logit	0.225	0.041	0.038	0.417	(0.386,0.448)
$t=7$	CRE logit	0.161	0.031	0.031	0.050	(0.036,0.064)
$t=8$	LPM	0.163	0.031	0.030	0.051	(0.037,0.065)
$t=8$	uncon logit	0.213	0.036	0.034	0.379	(0.349,0.409)
$t=8$	CRE logit	0.159	0.029	0.029	0.051	(0.037,0.065)

*Note: r is the rejection rate for the null of H_0:APE for δ=true APE . ρ is AR(1) serial correlation coefficient for the errors. True APE is 0.172($t=2$), 0.166($t=3$), 0.168($t=4$), 0.164($t=5$), 0.166($t=6$), 0.162($t=7$), and 0.158($t=8$).

Table F.4. LPM, unconditional logit, CRE logit APE estimates for δ

n=100	ρ=0.6	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t=2$	LPM	0.182	0.069	0.069	0.058	(0.043,0.073)
$t=2$	uncon logit	0.695	0.729	95297	0.737	(0.708,0.765)
$t=2$	CRE logit	0.172	0.065	0.064	0.066	(0.050,0.082)
$t=3$	LPM	0.176	0.050	0.050	0.049	(0.036,0.062)
$t=2$	uncon logit	0.432	0.163	0.781	0.785	(0.759,0.810)
$t=3$	CRE logit	0.169	0.047	0.047	0.052	(0.038,0.066)
$t=4$	LPM	0.173	0.041	0.042	0.050	(0.037,0.064)
$t=4$	uncon logit	0.341	0.072	0.063	0.737	(0.710,0.765)
$t=4$	CRE logit	0.168	0.022	0.022	0.053	(0.039,0.067)
$t=5$	LPM	0.167	0.037	0.037	0.056	(0.042,0.070)
$t=5$	uncon logit	0.294	0.058	0.052	0.673	(0.644,0.702)
$t=5$	CRE logit	0.163	0.035	0.035	0.054	(0.040,0.068)
$t=6$	LPM	0.169	0.034	0.034	0.052	(0.038,0.065)
$t=6$	uncon logit	0.268	0.048	0.044	0.633	(0.603,0.663)
$t=6$	CRE logit	0.164	0.032	0.032	0.056	(0.042,0.070)
$t=7$	LPM	0.164	0.031	0.031	0.054	(0.040,0.068)
$t=7$	uncon logit	0.247	0.042	0.039	0.591	(0.560,0.621)
$t=7$	CRE logit	0.161	0.029	0.029	0.054	(0.040,0.068)
$t=8$	LPM	0.162	0.029	0.029	0.054	(0.040,0.068)
$t=8$	uncon logit	0.232	0.037	0.035	0.568	(0.536,0.599)
$t=8$	CRE logit	0.158	0.028	0.028	0.056	(0.041,0.070)

*Note: r is the rejection rate for the null of H_0:APE for δ=true APE . ρ is AR(1) serial correlation coefficient for the errors. True APE is 0.172($t=2$), 0.166($t=3$), 0.168($t=4$), 0.164($t=5$), 0.166($t=6$), 0.162($t=7$), and 0.158($t=8$).

Table F.5. LPM, unconditional logit, CRE logit APE estimates for δ

n=100	ρ=0.8	Mean	SD	SE	r $\equiv$ 1(P<0.05)	95-CI
$t=2$	LPM	0.174	0.061	0.061	0.060	(0.043,0.077)
$t=2$	uncon logit	0.892	1.112	173194	0.739	(0.708,0.770)
$t=2$	CRE logit	0.165	0.058	0.058	0.055	(0.038,0.073)
$t=3$	LPM	0.176	0.045	0.045	0.045	(0.032,0.058)
$t=2$	uncon logit	0.638	0.600	40290	0.902	(0.883,0.920)
$t=3$	CRE logit	0.168	0.042	0.042	0.050	(0.036,0.064)
$t=4$	LPM	0.175	0.037	0.039	0.052	(0.038,0.066)
$t=4$	uncon logit	0.434	0.127	0.329	0.946	(0.933,0.961)
$t=4$	CRE logit	0.168	0.032	0.032	0.046	(0.033,0.059)
$t=5$	LPM	0.168	0.034	0.034	0.044	(0.032,0.057)
$t=5$	uncon logit	0.366	0.061	0.054	0.934	(0.919,0.950)
$t=5$	CRE logit	0.163	0.032	0.032	0.044	(0.032,0.057)
$t=6$	LPM	0.169	0.032	0.032	0.051	(0.037,0.067)
$t=6$	uncon logit	0.329	0.052	0.046	0.911	(0.893,0.929)
$t=6$	CRE logit	0.165	0.031	0.030	0.054	(0.039,0.069)
$t=7$	LPM	0.164	0.029	0.029	0.047	(0.034,0.060)
$t=7$	uncon logit	0.301	0.044	0.041	0.906	(0.888,0.924)
$t=7$	CRE logit	0.161	0.028	0.028	0.052	(0.038,0.065)
$t=8$	LPM	0.162	0.028	0.028	0.050	(0.037,0.064)
$t=8$	uncon logit	0.279	0.040	0.037	0.867	(0.845,0.888)
$t=8$	CRE logit	0.159	0.026	0.026	0.048	(0.035,0.061)

*Note: r is the rejection rate for the null of H_0:APE for δ=true APE . ρ is AR(1) serial correlation coefficient for the errors. True APE is 0.171($t=2$), 0.166($t=3$), 0.168($t=4$), 0.164($t=5$), 0.166($t=6$), 0.162($t=7$), and 0.159($t=8$).

Appendix G

Proof of Hausman-type test in chapter 2

We start with rewriting stacked estimator to obtain asymptotic variance.

$$\widehat{\theta}_a = \begin{bmatrix} \sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\mathbf{x}'_{it}\mathbf{x}_{it} & 0 \\ 0 & \sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\ddot{\mathbf{x}}'_{it}\ddot{\mathbf{x}}_{it} \end{bmatrix}^{-1}_{2k\times 2k} \begin{bmatrix} \sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\mathbf{x}'_{it}y_{it} \\ \sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\ddot{\mathbf{x}}'_{it}\ddot{y}_{it} \end{bmatrix}_{2k\times 1}$$

$$\sqrt{n}(\widehat{\theta}_a - \theta_a) = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}\mathbf{x}'_{it}\mathbf{x}_{it} & 0 \\ 0 & \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\ddot{\mathbf{x}}'_{it}\ddot{\mathbf{x}}_{it} \end{bmatrix}^{-1} \begin{bmatrix} (\frac{1}{\sqrt{n}}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}\mathbf{x}'_{it}u_{it}) \\ (\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\ddot{\mathbf{x}}'_{it}\ddot{u}_{it}) \end{bmatrix}$$

$$\sqrt{n}(\widehat{\theta}_a - \theta_a) = \begin{bmatrix} \sqrt{n}(\widehat{\theta}_{pls} - \theta) \\ \sqrt{n}(\widehat{\theta}_{FE} - \theta) \end{bmatrix} = \begin{bmatrix} (\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\mathbf{x}'_{it}\mathbf{x}_{it})^{-1}(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\mathbf{x}'_{it}u_{it}) \\ (\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\ddot{\mathbf{x}}'_{it}\ddot{\mathbf{x}}_{it})^{-1}(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\ddot{\mathbf{x}}'_{it}\ddot{u}_{it}) \end{bmatrix}$$

$\widehat{\theta}_a$ is a consistent estimator for θ_a by rank conditions and exogeneity conditions for pooled LS and FE estimators.

$$(p\lim \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\mathbf{x}_{it}'\mathbf{x}_{it})^{-1} = E(\sum_{t=1}^{T} s_{it}\mathbf{x}_{it}'\mathbf{x}_{it})^{-1} \equiv A_{11}^{-1}$$

$$(p\lim \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\ddot{\mathbf{x}}_{it}'\ddot{\mathbf{x}}_{it})^{-1} = E(\sum_{t=1}^{T} s_{it}\ddot{\mathbf{x}}_{it}'\ddot{\mathbf{x}}_{it})^{-1} \equiv A_{22}^{-1}$$

$$\begin{aligned}(p\lim \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\mathbf{x}_{it}'u_{it}) &= E(\sum_{t=1}^{T} s_{it}\mathbf{x}_{it}'u_{it}) = 0\\ (p\lim \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\ddot{\mathbf{x}}_{it}'\ddot{u}_{it}) &= E(\sum_{t=1}^{T} s_{it}\ddot{\mathbf{x}}_{it}'\ddot{u}_{it}) = 0\end{aligned}$$

Consider an estimator for the variance of $\sqrt{n}(\widehat{\theta}_a - \theta)$. we use following notations and obtain the asymptotic variance estimator as in (G.1).

$$\widehat{G}_a^{-1} = \begin{bmatrix} \widehat{A}_{11} & 0 \\ 0 & \widehat{A}_{22} \end{bmatrix}^{-1} \quad \widehat{D}_a = \begin{bmatrix} \widehat{D}_{11} & \widehat{D}_{12} \\ \widehat{D}_{21} & \widehat{D}_{22} \end{bmatrix}$$

$$\widehat{A}_{11} = \sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\mathbf{x}_{it}'\mathbf{x}_{it},\ \widehat{A}_{22} = \sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\ddot{\mathbf{x}}_{it}'\ddot{\mathbf{x}}_{it}$$

$$\widehat{D}_{11} = \frac{1}{n}\sum_{i=1}^{n}[(\sum_{t=1}^{T} s_{it}\mathbf{x}_{it}'\widehat{u}_{it})(\sum_{t=1}^{T} s_{it}\mathbf{x}_{it}'\widehat{u}_{it})'],\ \widehat{D}_{12} = \frac{1}{n}\sum_{i=1}^{n}[(\sum_{t=1}^{T} s_{it}\mathbf{x}_{it}'\widehat{u}_{it})((\sum_{t=1}^{T} s_{it}\ddot{\mathbf{x}}_{it}'\widetilde{u}_{it})']$$

$$\widehat{D}_{21} = \frac{1}{n}\sum_{i=1}^{n}[(\sum_{t=1}^{T} s_{it}\ddot{\mathbf{x}}_{it}'\widetilde{u}_{it})(\sum_{t=1}^{T} s_{it}\mathbf{x}_{it}'\widehat{u}_{it})'], \text{and } \widehat{D}_{22} = \frac{1}{n}\sum_{i=1}^{n}[(\sum_{t=1}^{T} s_{it}\ddot{\mathbf{x}}_{it}'\widetilde{u}_{it})(\sum_{t=1}^{T} s_{it}\ddot{\mathbf{x}}_{it}'\widetilde{u}_{it})']$$

where $\widetilde{u}_{it}$: residual from FE estimation and $\widehat{u}_{it}$ residual from pooled LS estimation.

$$\widehat{var}(\sqrt{n}(\widehat{\theta}_a - \theta_a)) = \begin{bmatrix} \widehat{A}_{11} & 0 \\ 0 & \widehat{A}_{22} \end{bmatrix}^{-1} \begin{bmatrix} \widehat{D}_{11} & \widehat{D}_{12} \\ \widehat{D}_{21} & \widehat{D}_{22} \end{bmatrix} \begin{bmatrix} \widehat{A}_{11} & 0 \\ 0 & \widehat{A}_{22} \end{bmatrix}^{-1} \tag{G.1}$$

$$\widehat{D}_{11} = \frac{1}{N}\sum_{i=1}^{n}[(\sum_{t=1}^{T} s_{it}x'_{it}\widehat{u}_{it})(\sum_{t=1}^{T} s_{it}x'_{it}\widehat{u}_{it})'] \xrightarrow{p} E[((\sum_{t=1}^{T} s_{it}x'_{it}u_{it})(\sum_{t=1}^{T} s_{it}x'_{it}u_{it})'] = D_{11}$$

$$\widehat{D}_{12} = \frac{1}{N}\sum_{i=1}^{n}[(\sum_{t=1}^{T} s_{it}x'_{it}\widehat{u}_{it})((\sum_{t=1}^{T} s_{it}\ddot{x}'_{it}\widetilde{u}_{it})'] \xrightarrow{p} E[(\sum_{t=1}^{T} s_{it}x'_{it}u_{it})((\sum_{t=1}^{T} s_{it}\ddot{x}'_{it}u_{it})'] = D_{12}$$

$$\widehat{D}_{21} = \frac{1}{N}\sum_{i=1}^{n}[(\sum_{t=1}^{T} s_{it}\ddot{x}'_{it}\widetilde{u}_{it})(\sum_{t=1}^{T} s_{it}x'_{it}\widehat{u}_{it})'] \xrightarrow{p} E[(\sum_{t=1}^{T} s_{it}\ddot{x}'_{it}u_{it})(\sum_{t=1}^{T} s_{it}x'_{it}u_{it})'] = D_{21}$$

$$\widehat{D}_{22} = \frac{1}{N}\sum_{i=1}^{n}[(\sum_{t=1}^{T} s_{it}\ddot{x}'_{it}\widetilde{u}_{it})(\sum_{t=1}^{T} s_{it}\ddot{x}'_{it}\widetilde{u}_{it})'] \xrightarrow{p} E[(\sum_{t=1}^{T} s_{it}\ddot{x}'_{it}u_{it})(\sum_{t=1}^{T} s_{it}\ddot{x}'_{it}u_{it})'] = D_{22}$$

We obtain (23) from (21) and (22).

$$\widehat{A}_{11}^{-1} \xrightarrow{p} A_{11}^{-1},\ \widehat{A}_{22}^{-1} \xrightarrow{p} A_{22}^{-1} \text{ where } \xrightarrow{p} \text{ denote probability limit} \tag{G.2}$$

$$\begin{bmatrix} \widehat{D}_{11} & \widehat{D}_{12} \\ \widehat{D}_{21} & \widehat{D}_{22} \end{bmatrix} \xrightarrow{p} \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix} \tag{G.3}$$

$$\sqrt{N}(\widehat{\theta}_a - \theta) \Rightarrow N\left(0, \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix}^{-1} \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix} \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix}^{-1}\right) \tag{G.4}$$

where $\Rightarrow$ implies weak convergence (i.e. convergence in distribution)

$$\begin{aligned} & \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix}^{-1} \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix} \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix}^{-1} \\ = & \begin{bmatrix} A_{11}^{-1} & 0 \\ 0 & A_{22}^{-1} \end{bmatrix} \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix} \begin{bmatrix} A_{11}^{-1} & 0 \\ 0 & A_{22}^{-1} \end{bmatrix} \\ = & \begin{bmatrix} A_{11}^{-1}D_{11}A_{11}^{-1} & A_{a1}^{-1}D_{12}A_{22}^{-1} \\ A_{22}^{-1}D_{21}A_{11}^{-1} & A_{22}^{-1}D_{22}A_{22}^{-1} \end{bmatrix} \end{aligned}$$

Thus, we obtain (24) from (23).

$$R\sqrt{N}(\widehat{\theta}_a - \theta) \underbrace{=}_{R\theta=0} R\sqrt{N}\widehat{\theta}_a \Rightarrow N\left(0, \begin{bmatrix} R & \begin{matrix} A_{11}^{-1}D_{11}A_{11}^{-1} & A_{a1}^{-1}D_{12}A_{22}^{-1} \\ A_{22}^{-1}D_{21}A_{11}^{-1} & A_{22}^{-1}D_{22}A_{22}^{-1} \end{matrix} & R' \end{bmatrix}\right) \tag{G.5}$$

Denote following notations ,

$$\widehat{G}_a^{-1} = \begin{bmatrix} \widehat{A}_{11} & 0 \\ 0 & \widehat{A}_{22} \end{bmatrix}^{-1} \quad \widehat{D}_a = \begin{bmatrix} \widehat{D}_{11} & \widehat{D}_{12} \\ \widehat{D}_{21} & \widehat{D}_{22} \end{bmatrix}$$

Then, we have (25) from (20) and (23).

$$\widehat{var}(\sqrt{N}(\widehat{\theta}_a - \theta)) = \widehat{G}_a^{-1}\widehat{D}_a\widehat{G}_a^{-1} = G_a^{-1}D_aG_a^{-1} + o_p(1), \text{ (i.e. } \widehat{G}_a^{-1}\widehat{D}_a\widehat{G}_a^{-1} \xrightarrow{p} G_a^{-1}D_aG_a^{-1}) \tag{G.6}$$

where

$$\begin{aligned} G_a^{-1}D_aG_a^{-1} &= \begin{bmatrix} A_{11}^{-1} & 0 \\ 0 & A_{22}^{-1} \end{bmatrix} \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix} \begin{bmatrix} A_{11}^{-1} & 0 \\ 0 & A_{22}^{-1} \end{bmatrix} \\ &= \begin{bmatrix} A_{11}^{-1}D_{11}A_{11}^{-1} & A_{a1}^{-1}D_{12}A_{22}^{-1} \\ A_{22}^{-1}D_{21}A_{11}^{-1} & A_{22}^{-1}D_{22}A_{22}^{-1} \end{bmatrix} \end{aligned}$$

Thus, from (25) we can construct a statistic $\widehat{W}_1 = [R\sqrt{N}\widehat{\theta}_a]'[R\widehat{G}_a^{-1}\widehat{D}_a\widehat{G}_a^{-1}R']^{-1}R\sqrt{N}\widehat{\theta}_a$ which converges to chi-squared distribution.

$$R\sqrt{N}\widehat{\theta}_a \overset{a}{\sim} N(0, RG_a^{-1}D_aG_a^{-1}R')$$

$$[R\sqrt{N}\widehat{\theta}_a]'[R\widehat{G}_a^{-1}\widehat{D}_a\widehat{G}_a^{-1}R']^{-1}R\sqrt{N}\widehat{\theta}_a \sim \chi_k^2.$$

Appendix H

Proof of Corollary 3 in Chapter 3: Robust inference in unbalanced panel data

Proof First, we consider WLS estimator.

$$\widehat{\beta}_{wls} = (\sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\frac{1}{\sqrt{\lambda_i}}\widetilde{x_{it}}'\frac{1}{\sqrt{\lambda_i}}\widetilde{x_{it}})^{-1}(\sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\frac{1}{\sqrt{\lambda_i}}\widetilde{x_{it}}'\frac{1}{\sqrt{\lambda_i}}\widetilde{y_{it}})$$
$$= (\sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\frac{1}{\sqrt{T_i}}\widetilde{x_{it}}'\frac{1}{\sqrt{T_i}}\widetilde{x_{it}})^{-1}(\sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\frac{1}{\sqrt{T_i}}\widetilde{x_{it}}'\frac{1}{\sqrt{T_i}}\widetilde{y_{it}}).$$

Then we have $\sqrt{T}(\widehat{\beta}_{wls}-\beta) = \sqrt{T}(\sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\frac{1}{\sqrt{T_i}}\widetilde{x_{it}}'\frac{1}{\sqrt{T_i}}\widetilde{x_{it}})^{-1}(\sum_{i=1}^{n}\sum_{t=1}^{T} s_{it}\frac{1}{\sqrt{T_i}}\widetilde{x_{it}}'\frac{1}{\sqrt{T_i}}u_{it})$

$$= (\sum_{i=1}^{n}\frac{1}{T_i}\sum_{t=1}^{T} s_{it}\widetilde{x_{it}}'\widetilde{x_{it}})^{-1}(\sum_{i=1}^{n}\frac{1}{\sqrt{T_i\lambda_i}}\sum_{t=1}^{T} s_{it}\widetilde{x_{it}}'u_{it}) =$$
$$(\sum_{i=1}^{n}\frac{T}{T_i}\frac{1}{T}\sum_{t=1}^{T} s_{it}\widetilde{x_{it}}'\widetilde{x_{it}})^{-1}(\sum_{i=1}^{n}\frac{\sqrt{T}}{\sqrt{T_i\lambda_i}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} s_{it}\widetilde{x_{it}}'u_{it})$$
$$\Rightarrow (\sum_{i=1}^{n}\frac{1}{\lambda_i}\lambda_i Q_i)^{-1}(\sum_{i=1}^{n}\frac{1}{\lambda_i}\Lambda_i W_i(\lambda_i)) = (\sum_{i=1}^{n} Q_i)^{-1}(\sum_{i=1}^{n}\frac{1}{\lambda_i}\Lambda_i W(\lambda_i)) =$$
$$(\sum_{i=1}^{n} Q_i)^{-1}(\sum_{i=1}^{n}\frac{1}{\lambda_i}\Lambda_i W_i)$$

$$\sqrt{T}(\widehat{\beta}_{wls}-\beta) \Rightarrow (\sum_{i=1}^{n} Q_i)^{-1}(\sum_{i=1}^{n}\frac{1}{\lambda_i}\Lambda_i W_i) \tag{H.1}$$

$$T\cdot\widehat{var}(\widehat{\beta}_{wls}) = T\cdot\widehat{V}_{w-clus} = (\sum_{i=1}^{n}\frac{1}{T_i}\sum_{t=1}^{T} s_{it}\widetilde{x_{it}}'\widetilde{x_{it}})^{-1}\widehat{\Omega}_{wls}(\sum_{i=1}^{n}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} s_{it}\widetilde{x_{it}}'\widetilde{x_{it}})^{-1}$$

where

$$\widehat{Q}_i = \frac{1}{T_i}\sum_{t=1}^{T} s_{it}\widetilde{x_{it}}'\widetilde{x_{it}}\ ,\ \widehat{\Omega}_{wls} = \sum_{i=1}^{n}[(\frac{1}{\lambda_i}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} s_{it}\widetilde{x_{it}}'\widehat{u}_{it})(\frac{1}{\lambda_i}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} s_{it}\widetilde{x_{it}}'\widehat{u}_{it})']$$

$$\frac{1}{\lambda_i}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} s_{it}\widetilde{x_{it}}'\widehat{u}_{it} = \frac{1}{\lambda_i}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} s_{it}\widetilde{x_{it}}' u_{it} - \frac{1}{\lambda_i}\frac{1}{T}\sum_{t=1}^{T} s_{it}\widetilde{x_{it}}'\widetilde{x_{it}}\sqrt{T}(\widehat{\beta}_{wls}-\beta) + o_p(1)$$

$$\Longrightarrow \frac{1}{\lambda_i}\Lambda_i W_i - \frac{1}{\lambda_i}\lambda_i Q_i(\sum_{j=1}^{n} Q_j)^{-1}(\sum_{j=1}^{n}\frac{1}{\lambda_j}\Lambda_j W_j) = \frac{1}{\lambda_i}\Lambda_i W_i - Q_i(\sum_{j=1}^{n} Q_j)^{-1}(\sum_{j=1}^{n}\frac{1}{\lambda_j}\Lambda_j W_j)$$

$$\widehat{\Omega}_{wls} = \sum_{i=1}^{n}[(\frac{1}{\lambda_i}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} s_{it}\widetilde{x_{it}}'\widehat{u}_{it})(\frac{1}{\lambda_i}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} s_{it}\widetilde{x_{it}}'\widehat{u}_{it})']$$

$$\Longrightarrow \sum_{i=1}^{n}[\tfrac{1}{\lambda_i}\Lambda_i B_i - Q_i(\sum_{j=1}^{n} Q_j)^{-1}(\sum_{j=1}^{n}\tfrac{1}{\lambda_j}\Lambda_j W_j)]\{[\tfrac{1}{\lambda_i}\Lambda_i B_i - Q_i(\sum_{j=1}^{n} Q_j)^{-1}(\sum_{j=1}^{n}\tfrac{1}{\lambda_j}\Lambda_j W_j)]'\}$$

$$= \sum_{i=1}^{n}[\tfrac{1}{\lambda_i}\Lambda_i W_i W_i'\Lambda_i'\tfrac{1}{\lambda_i} - Q_i(\sum_{j=1}^{n} Q_j)^{-1}(\sum_{j=1}^{n}\tfrac{1}{\lambda_j}\Lambda_j W_j)\tfrac{1}{\lambda_i}W_i'\Lambda_i' - \tfrac{1}{\lambda_i}\Lambda_i W_i(\sum_{j=1}^{n}\tfrac{1}{\lambda_j}W_j'\Lambda_j')(\sum_{j=1}^{n} Q_j)^{-1}Q_i + Q_i(\sum_{j=1}^{n} Q_j)^{-1}(\sum_{j=1}^{n}\tfrac{1}{\lambda_j}\Lambda_j W_j)(\sum_{j=1}^{n}\tfrac{1}{\lambda_j}W_j'\Lambda_j')(\sum_{j=1}^{n} Q_j)^{-1}Q_i]$$

Using homogeneity, $\Lambda_j = \Lambda$ and $Q_j = Q$ for all $j = 1, 2, .., n$

$$= \sum_{i=1}^{n}[\tfrac{1}{\lambda_i}\Lambda W_i W_i'\Lambda'\tfrac{1}{\lambda_i} - \tfrac{1}{n}(\Lambda\sum_{j=1}^{n}\tfrac{1}{\lambda_j}W_j)\tfrac{1}{\lambda_i}W_i'\Lambda' - \Lambda\tfrac{1}{\lambda_i}W_i(\sum_{j=1}^{n}\tfrac{1}{\lambda_j}W_j')\Lambda'\tfrac{1}{n} + \tfrac{1}{n}\Lambda(\sum_{j=1}^{n}\tfrac{1}{\lambda_j}W_j)(\sum_{j=1}^{n}\tfrac{1}{\lambda_j}W_j')\Lambda'\tfrac{1}{n}]$$

$$= \Lambda[\sum_{i=1}^{n}\frac{W_i}{\lambda_i}\frac{W_i'}{\lambda_i} - \tfrac{1}{n}(\sum_{j=1}^{n}\frac{W_j}{\lambda_j})(\sum_{i=1}^{n}\frac{W_i'}{\lambda_i}) - (\sum_{i=1}^{n}\frac{W_i}{\lambda_i})(\sum_{j=1}^{n}\frac{W_j'}{\lambda_j})\tfrac{1}{n} + \sum_{i=1}^{n}\tfrac{1}{n}(\sum_{j=1}^{n}\frac{W_j}{\lambda_j})(\sum_{j=1}^{n}\frac{W_j'}{\lambda_j})\tfrac{1}{n}]\Lambda'$$

$$= \Lambda[\sum_{i=1}^{n}\frac{W_i}{\lambda_i}\frac{W_i'}{\lambda_i} - \tfrac{1}{n}(\sum_{j=1}^{n}\frac{W_j}{\lambda_j})(\sum_{i=1}^{n}\frac{W_i'}{\lambda_i})]\Lambda'$$

$$\widehat{\Omega}_{wls} \Longrightarrow \Lambda[\sum_{i=1}^{n} \frac{W_i}{\lambda_i}\frac{W_i'}{\lambda_i} - \frac{1}{n}(\sum_{j=1}^{n}\frac{W_j}{\lambda_j})(\sum_{j=1}^{n}\frac{W_j'}{\lambda_j})]\Lambda'$$

$$\begin{aligned} T\cdot\widehat{V}_{w-clus} &= (\sum_{i=1}^{n}\frac{1}{T_i}\sum_{t=1}^{T} s_{it}\widetilde{x_{it}}\widetilde{x_{it}}')^{-1}\widehat{\Omega}_{wls}(\sum_{i=1}^{n}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} s_{it}\widetilde{x_{it}}\widetilde{x_{it}}')^{-1} \\ &= (\sum_{i=1}^{n}\widehat{Q_i})^{-1}\widehat{\Omega}_{wls}(\sum_{i=1}^{n}\widehat{Q_i})^{-1} \\ &\Rightarrow (nQ)^{-1}\Lambda[\sum_{i=1}^{n}\frac{W_i}{\lambda_i}\frac{W_i'}{\lambda_i} - \frac{1}{n}(\sum_{j=1}^{n}\frac{W_j}{\lambda_j})(\sum_{j=1}^{n}\frac{W_j'}{\lambda_j})]\Lambda'(nQ)^{-1} \qquad \text{(H.2)} \end{aligned}$$

Now consider t-statistic based on WLS which we use for t-test.

$$t^*_{wls} = \frac{\sqrt{T}(R\widehat{\beta}_{wls} - r)}{\sqrt{R\cdot T\cdot\widehat{V}_{w-clus}R'}} = \frac{R\sqrt{T}(\widehat{\beta}_{wls}-\beta)}{\sqrt{R\cdot T\cdot\widehat{V}_{w-clus}R'}}$$

For the numerator, using (H.1), we have

$$\begin{aligned} R\sqrt{T}(\widehat{\beta}_{wls}-\beta) &\Rightarrow R(nQ)^{-1}(\sum_{i=1}^{n}\frac{1}{\lambda_i}\Lambda W_i) = \frac{1}{n}RQ^{-1}\Lambda(\sum_{i=1}^{n}\frac{W_i}{\lambda_i}) \\ &= \frac{1}{n}(\sum_{i=1}^{n}\frac{RQ^{-1}\Lambda W_i}{\lambda_i}) \equiv \frac{1}{n}(\sum_{i=1}^{n} y_i) \qquad \text{(H.3)} \end{aligned}$$

For the denominator, using (H.2), we have

$$\begin{aligned} R\cdot T\cdot\widehat{V}_{w-clus}R' &\Rightarrow R(nQ)^{-1}\Lambda[\sum_{i=1}^{n}\frac{W_i}{\lambda_i}\frac{W_i'}{\lambda_i} - \frac{1}{n}(\sum_{j=1}^{n}\frac{W_j}{\lambda_j})(\sum_{j=1}^{n}\frac{W_j'}{\lambda_j})]\Lambda'(nQ)^{-1}R' \\ &= \frac{1}{n^2}[\sum_{i=1}^{n}\frac{RQ^{-1}\Lambda W_i}{\lambda_i}\frac{W_i'\Lambda'Q^{-1}R'}{\lambda_i} - \frac{1}{n}(\sum_{i=1}^{n}\frac{RQ^{-1}\Lambda W_i}{\lambda_i})(\sum_{j=1}^{n}\frac{W_j'\Lambda'Q^{-1}R'}{\lambda_j})] \\ &= \frac{1}{n^2}[\sum_{i=1}^{n} y_i y_i' - \frac{1}{n}\sum_{i=1}^{n} y_i\sum_{j=1}^{n} y_j'] \qquad \text{(H.4)} \end{aligned}$$

Using (H.3) and (H.4) together, we have

$$
\begin{aligned}
t^{*}_{wls} &= \frac{R\sqrt{T}(\widehat{\beta}_{wls}-\beta)}{\sqrt{R\cdot T\cdot \widehat{V}_{w-clus}R'}} \Rightarrow \frac{\frac{1}{n}(\sum_{i=1}^{n} y_i)}{\sqrt{\frac{1}{n^2}[\sum_{i=1}^{n} y_i y_i' - \frac{1}{n}\sum_{i=1}^{n} y_i \sum_{j=1}^{n} y_j']}} = \frac{(\sum_{i=1}^{n} y_i)}{\sqrt{[\sum_{i=1}^{n} y_i y_i' - \frac{1}{n}\sum_{i=1}^{n} y_i \sum_{j=1}^{n} y_j']}} \\
&= \frac{n\overline{y}}{\sqrt{\sum_{i=1}^{n}(y_i-\overline{y})^2}} = \sqrt{\frac{n}{n-1}}\cdot\frac{\sqrt{n}\overline{y}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\overline{y})^2}} \equiv \sqrt{\frac{n}{n-1}}\widehat{T}_{n-1}
\end{aligned}
$$

Appendix I

Monte Carlo simulation results in chapter 3: Size-adjusted power for t-test in Ibragimov and Muller [2009]

Table I.1. The size adjusted power of test, $1(p < 0.05)$, $T = 100$

T=100, n=4	balanced	complete-case	balanced keep $\forall i$ drop t	balanced keep $\forall t$ drop i
bias=0	.048(.041,.054)	.028(.023,.033)	.046(.040,.053)	.053(.046,.060)
bias=2%	.059(.048,.069)	.038(.029,.046)	.048(.039,.057)	.063(.052,.073)
bias=5%	.092(.079,.105)	.052(.042,.061)	.050(.040,.060)	.070(.059,.081)
bias=10%	.219(.200,.237)	.100(.087,.113)	.063(.052,.074)	.093(.080,.105)
bias=20%	.626(.605,.647)	.302(.282,.322)	.136(.120,.151)	.159(.143,.175)
T=100, n=8	balanced	complete-case	balanced-drop t	balanced -drop i
bias=0	.048(.042,.055)	.038(.032,.044)	.050(.040,.059)	.053(.046,.060)
bias=2%	.073(.061,.084)	.054(.044,.063)	.047(.037,.056)	.052(.042,.061)
bias=5%	.167(.150,.183)	.099(.085,.112)	.062(.051,.072)	.060(.049,.070)
bias=10%	.517(.495,.539)	.243(.224,.262)	.108(.094,.122)	.089(.077,.101)
bias=20%	.977(.970,.983)	.671(.650,.692)	.311(.290,.331)	.156(.140,.171)
T=100, n=16	balanced	complete-case	balanced-drop t	balanced -drop i
bias=0	.048(.041,.055)	.050(.043,.057)	.050(.043,.057)	.048(.041,.055)
bias=2%	.110(.096,.123)	.069(.057,.080)	.057(.047,.067)	.059(.048,.069)
bias=5%	.362(.340,.383)	.191(.174,.208)	.124(.109,.138)	.094(.081,.106)
bias=10%	.887(.873,.901)	.574(.552,.595)	.330(.309,.351)	.225(.206,.243)
bias=20%	1(1,1)	.984(.978,.989)	.839(.823,.855)	.628(.607,.649)

(Note) The rejection probability is reported. 95 % coverage rate for rejection probability is reported in parenthesis. Generated errors are from AR(1) process with correlation coefficient 0.75. The number of replications is 2,000.

Table I.2. The size adjusted power of test, $1(p < 0.05)$, $T = 300$

T=300, n=4	balanced	complete-case	balanced keep $\forall i$ drop t	balanced keep $\forall t$ drop i
bias=0	.046(.036,.055)	.029(.022,.037)	.058(.047,.068)	.048(.039,.077)
bias=2%	.067(.056,.078)	.034(.026,.041)	.053(.043,.063)	.055(.046,.065)
bias=5%	.169(.153,.185)	.088(.075,.100)	.072(.061,.083)	.079(.067,.090)
bias=10%	.520(.498,.541)	.237(.218,.255)	.143(.127,.158)	.145(.130,.160)
bias=20%	.968(.960,.976)	.642(.621,.663)	.367(.346,.388)	.262(.243,.281)
T=300, n=8	balanced	complete-case	balanced-drop t	balanced -drop i
bias=0	.053(.046,.060)	.042(.034,.050)	.049(.040,.059)	.050(.043,.058)
bias=2%	.122(.107,.136)	.089(.076,.101)	.070(.059,.081)	.053(.043,.063)
bias=5%	.435(.413,.457)	.270(.250,.289)	.179(.162,.196)	.070(.059,.081)
bias=10%	.941(.931,.951)	.748(.729,.767)	.548(.526,.569)	.126(.111,.141)
bias=20%	1(1,1)	.999(.998,1)	.986(.980,.991)	.249(.230,.268)
T=300, n=16	balanced	complete-case	balanced-drop t	balanced -drop i
bias=0	.046(.036,.055)	.050(.040,.060)	.050(.040,.059)	.048(.038,.057)
bias=2%	.187(.170,.204)	.135(.120,.150)	.084(.072,.096)	.060(.050,.071)
bias=5%	.798(.780,.815)	.502(.480,.523)	.294(.274,.313)	.180(.163,.197)
bias=10%	1(1,1)	.960(.951,.969)	.782(.763,.800)	.530(.508,.551)
bias=20%	1(1,1)	1(1,1)	1(.999,1)	.971(.963,.978)

(Note) The rejection probability is reported. 95 % coverage rate for rejection probability is reported in parenthesis. Generated errors are from AR(1) process with correlation coefficient 0.75. The number of replications is 2,000.

Table I.3. The size adjusted power of test, $1(p < 0.05)$, $T = 1,000$

T=1,000, n=4	balanced	complete-case	balanced keep $\forall i$ drop t	balanced keep $\forall t$ drop i
bias=0	.056(.045,.066)	.033(.025,.041)	.042(.033,.050)	.061(.050,.071)
bias=2%	.136(.121,.151)	.067(.056,.078)	.047(.038,.057)	.076(.064,.087)
bias=5%	.484(.462,.506)	.220(.201,.238)	.050(.040,.060)	.130(.115,.144)
bias=10%	.942(.931,.952)	.603(.581,.624)	.072(.061,.083)	.250(.231,.269)
bias=20%	1(1,1)	.959(.950,.967)	.147(.131,.163)	.480(.458,.502)
T=1,000, n=8	balanced	complete-case	balanced-drop t	balanced -drop i
bias=0	.046(.038,.054)	.036(.027,.044)	.052(.042,.061)	.055(.045,.065)
bias=2%	.273(.253,.292)	.144(.129,.159)	.092(.079,.104)	.066(.055,.077)
bias=5%	.906(.893,.918)	.577(.555,.598)	.281(.261,.300)	.130(.115,.144)
bias=10%	1(1,1)	.963(.955,.971)	.740(.721,.759)	.240(.221,.258)
bias=20%	1(1,1)	1(1,1)	.999(.997,1)	.442(.420,.464)
T=1,000, n=16	balanced	complete-case	balanced-drop t	balanced -drop i
bias=0	.050(.045,.059)	.046(.037,.055)	.050(.040,.060)	.050(.040,.060)
bias=2%	.523(.046,.060)	.285(.265,.305)	.187(.169,.204)	.133(.118,.147)
bias=5%	1(1,1)	.939(.929,.949)	.739(.721,.758)	.469(.447,.491)
bias=10%	1(1,1)	1(1,1)	.999(.997,1)	.942(.932,.952)
bias=20%	1(1,1)	1(1,1)	1(1,1)	1(1,1)

(Note) The rejection probability is reported. 95 % coverage rate for rejection probability is reported in parenthesis. Generated errors are from AR(1) process with correlation coefficient 0.75. The number of replications is 2,000.

BIBLIOGRAPHY

Bibliography

J. Abrevaya. The equivalence of two estimators of the fixed-effects logit model. *Economics Letters*, 55:41–43, 1997.

E. Andersen. *Conditional Inference and Models for Measuring*. Mentalhygienjnisk Forlag, Copenhague, 1973.

J. Angrist and J. Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton Univ. Press, Princeton, New Jersey, 2009.

M. Arellano. Computing robust standard errors for within-groups estimators. *Oxford Bulletin of Economics and Statistics*, 49:431–34, 1987.

M. Arellano. Discrete choices with panel data. *Investigaciones Economicas*, 27:423–458, 2003.

M. Arellano and J. Hahn. *Understanding Bias in Nonlinear Panel Models: Some Recent Developments. In Advances in Economics and Econometrics, Blundell R, Newey W, Persson T (eds)*. Cambridge University Press, Cambridge, 2007.

N. Bakirov and G. Szekely. Student's t-test for gaussian scale mixtures. *Journal of Mathematical Science*, 139:6497–6505, 2006.

H. Bang and J. Robins. Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61:692–697, 2005.

G. Becker. *Human Capital: A Theoretical and Empirical Ananlysis, with Special Reference to Education*. The university of Chicago Press, Chicago and London, 1993.

M. Bertrand, E. Duflo, and S. Mullainathan. How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, 119:249–275, 2004.

A. Bester, T. Conley, and C. Hansen. Inference with dependent data using cluster covariance esteimators. mimeo, 2009.

P. Bickel, C. Klaassen, Y. Ritov, and J. Wellner. *Efficient and Adaptive Estimation in Semiparametric Models*. Springer Verlag, New York, 1998.

S. Van Buuren, H. Boshuizen, and D. Knook. Multiple imputation of missing blood pressure covariates in survival analysis. *Statistics in Medicine*, 18:681–694, 1999.

J. Carpenter, M. Kenward, and S. Vansteelandt. comparison of multiple imputation and doubly robust estimation for analyses with missing data. *Journal of the Royal Statistical Society, Series A(Statistics in Society)*, 169:571–584, 2006.

G. Chamberlain. Analysis of covariance with qualitative data. *Review of Economic Studies*, 47:225–238, 1980.

G. Chamberlain. "panel data", in eds z. griliches and m. intriligator. *Handbook of Econometrics*, 2:1248–1313, 1984.

G. Chamberlain. Binary response models for panel data: Identification and information. *Econometrica*, 78:159–168, 2010.

K. Chay and D. Hyslop. Identification and estimation of dynamic binary panel data models: Empirical evidence using alternative approaches. 2001.

D. Clayton, D. Spiegelhalter, and A. Pickles. Analysis of longitudinal binary data from multi-phase sampling (with discussion). *Journal of the Royal Statistical Society, Series B(statistical methodology)*, 60:71–87, 1999.

M. Collado and M. Browning. Habits and heterogeneity in demands: a panel data analysis. *Journal of Applied Econometrics*, 22:625–640, 2007.

L. Devroye. *Non-Uniform Random Variate Generation*. Springer-Verlag, New York, 1986.

E. Fama and J. MacBeth. Long-term growth in a short-term market. *Journal of Finance*, 29(3):857–85, 1974.

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

B. Graham, C. Campos de Xavier Pinto, and D. Egel. Inverse probability tilting for moment condition models with missing data. NBER Working Paper No. 13981, 2008.

W. Green. The behavior of the fixed effects estimator in nonlinear models. *The Econometrics Journal*, 7:98–119, 2004.

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

V. Hajivassiliou and D. McFadden. The method of simulated scores with application to models of external debt crises. *Econometrica*, 66:863–896, 1998.

P. Hall and B. Presnell. Biased bootstrap methods for reducign the effects of contamination. *Journal of the Royal Statistical Society, Series B(statistical methodology)*, 61:661–680, 1999.

C. Hansen. Asymptotic properties of a robust variance matrix estimator for panel data when t is large. *Journal of Econometrics*, 141:597–620, 2007.

E. Hanushek. Instrumental variables estimates of the effect of subsidized traning on the quantiles of traninee earnings. *Econometrica*, 70:91–117, 1986.

E. Hanushek. The effects of birth inputs on birthweight: Evidence from quantile estimation on panel data. *Journal of Business and Economic Statistics*, 26:379–397, 1997.

J. Hausman. Specification tests in econometrics. *Econometrica*, 46:1251–1271, 1978.

J. Heckman. *The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Descrete Time-Discrete Data Stochastic Process and Some Monte Carlo Evidence*. MIT Press, Cambridge, Massachusetts, 1981.

K. Hirano, G. Imbens, and G. Ridder. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, 71:1161–1189, 2003.

P. Hoel, S. Port, and C. Stone. *Introduction to Probability Theory*. Brooks Cole, Boston, 1972.

D. Horvitz and D. Thompson. A generalization of sampling without replacement from a finite universe. *Journal of American Statistical Association*, 47:663–685, 1952.

R. Ibragimov and U. Muller. t-statistic based correlation and heterogeneity robust inference. *Journal of Business and Economic Statistics*, 28:453–468, 2009.

M. Johnson. *Multivariate Statistical Simulation*. John Wiley and Sons, New York, 1987.

N. Kiefer and T. Vogelsang. Heteroskedasticity-autocorrelation robust standard errors using the bartlett kernel without truncation. *Econometrica*, 70(5):2093–2095, 2002.

N. Kiefer and T. Vogelsang. A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory*, 21(06):1130–1164, 2005.

A. Krueger. Experimental estimates of education production functions. *Quarterly Journal of Economics*, 114:497–532, 1999.

A. Krueger and D. Whitmore. The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from project star. *Economic Journal*, 111:1–28, 2001.

T. Lancaster. The incidental parameters problem since 1948. *Journal of Econometrics*, 95: 391–413, 2000.

L. Lee. On comparisions of normal and logistic models in the bivariate dichotomous analysis. *Journal of Econometrics*, 4:151–155, 1979.

K. Liang and S. Zeger. Longitudinal data analysis using generalized linear model. *Biometrika*, 73:13–22, 1986.

R. Little and D. Rubin. *Statistical Analysis with Missing Data*. J. Wiley and Sons, New York, 2002.

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

M. Petersen. Estimating standard errors in finance panel data sets: Comparing approaches. *The Review of Financial Studies*, 22:435–480, 2009.

T. Ragnuthan, J. Lepkowski, J. van Hoewyk, and P. Solenberger. A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology*, 27:85–95, 2001.

J. Robins and A. Rotnitzky. Semiparametric efficiency in multivariate regression models with missing data. *Journal of American Statistical Association*, 90:122–129, 1995.

J. Robins, A. Rotnitzky, and L. Zhao. Estimation of regression coefficients when some regressors are not always observed. *Journal of American Statistical Association*, 89:846–866, 1994.

P. Royston. Multiple imputation of missing values. *Stata Journal*, 4:227–241, 2004.

P. Royston. Multiple imputation of missing values: update. *Stata Journal*, 5:188–201, 2005.

D. Rubin. Inference and missing data. *Biometrika*, 63:581–592, 1976.

D. Rubin. *Semiparametric Theory and Missing Data*. J. Wiley and Sons, New York, 1987.

J. Schafer and J. Graham. Missing data: Our view of the state of the art. *Psychological Methods*, 7:147–177, 2002.

D. Scharfstein, A. Rotnitzky, and J. Robins. Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of American Statistical Association*, 94: 1096–1120 with Rejoinder, 1135–1146, 1999.

B. Schweizer and A. Sklar. *Probabilistic Metric Spaces*. North Holland, New York, 1983.

A. Sklar. *Fonctions de repartition a n dimensions et leurs marges*. Pul Inst Statist Univ, Paris, 1959.

T. Trivedi and D. Zimmer. *Copula Modeling: an Introduction for Practitioners*. Now Publishers, 2007.

A. Tsiatis. *Semiparametric Theory and Missing Data*. Springer, New York, 2006.

T. Vogelsang. Heteroskedasticity, autocorrelation, and spatial correlation robust inference in linear panel models with fixed-effects. 2008.

J. Wooldridge. Cluster-sample methods in applied econometrics. *American Economic Review*, 93(2):133–138, 2003.

J. Wooldridge. Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics*, 20:39–54, 2005.

J. Wooldridge. Inverse probability weighted estimation for general missing data problems. *Journal of Econometrics*, 141:1281–1301, 2007.

J. Wooldridge. Correlated random effects models with unbalanced panels. mimeo, 2009.

J. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, Massachustte, 2010.

E. Word, J. Johnston, H. Bain, B. Fulton, C. Achilles, M. Lintz, J. Folger, and C. Breda. The state of tennessee's student/teacher achievement ratio (star) project: Technical report 1985-1990. Tennessee State Department of Education, 1990.