Semiparametric Least Squares Estimator of Binary Choice Panel Data Models with Endogeneity

Anastasia Semykina* Yimeng Xie† Cynthia Fan Yang* Qiankun Zhou‡

July 23, 2020

Abstract

In this paper, we consider a semiparametric least squares estimation of binary response panel data models with endogenous regressors. The estimator relies on the correlated random effects model and control function approach to address the endogeneity due to the presence of the unobserved time-invariant effect and nonzero correlation of the idiosyncratic error with one or more explanatory variables. We derive the asymptotic properties of the proposed estimator and use Monte Carlo simulations to show that it performs well in finite samples. As an illustration, the considered method is used for estimating the effect of non-wife income on labor force participation of married women.

Keywords: Panel Binary Model, Mundlak's Projection Method, Semiparametric Estimation, Least Squares Estimaton, Control Function Method, Kernel Regression.

*Department of Economics, Florida State University, Tallahassee, FL 32306.
†Department of Economics, University of Southern California, Los Angeles, CA 90089.
‡Department of Economics, Louisiana State University, Baton Rouge, LA 70803.

1 Introduction

A binary choice panel data model is frequently used in empirical economics research, as it incorporates heterogeneity in individual responses. In practice, a standard conditional exogeneity assumption can be violated because of the measurement error, simultaneity, or omitted variable problem. This paper proposes a nonlinear least squares estimator of such a model, where we furthermore assume the distribution of the error term to be unknown. Our method relies on the semiparametric estimation of the conditional expectation of the choice variable and accounts for both the correlated unobserved heterogeneity and endogeneity due to correlation with an idiosyncratic error.

In this paper, the potential correlation between the unobserved time-invariant effect and exogenous covariates is addressed using a correlated random effects approach, which is free from the incidental parameters problem associated with the fixed effects estimation. Specifically, we model the unobserved effect using the Mundlak (1978) and Chamberlain (1980) specification, which assumes that the unobserved effect is a function of observed independent variables. Although such procedure induces serial correlation at the individual level, we show that in short panels standard asymptotic theory applies. Additionally, following Song (2017), we assume that there is at least one continuous independent variable. Under such a condition, it is shown that unlike Manski (1985), we can identify parameters without necessarily requiring the unboundedness of other independent variables. Because variables in applied economics research are usually bounded, this assumption is more suitable for observational data.

Endogeneity due to correlation with time-varying unobservables is addressed using a control function method. In a parametric setting, this method was initially proposed by Smith and Blundell (1986) and Rivers and Vuong (1988). Blundell and Powell (2004), Rothe (2009), and Song (2017) use the control function technique to correct for endogeneity in semiparametric binary response models for cross section data. Generally, control function methods have been used for estimating a wide variety of linear and nonlinear models (e.g., Wooldridge (2015)).

Our estimator is developed from the least squares estimator of a single index model considered by Ichimura (1993) and can be applied to panel data. The asymptotic theory is developed based on Chen et. al. (2003). This strategy is commonly adopted in the literature related to the present paper, such as Rothe (2009) and Song (2017). However, the latter two papers impose the IID assumption, similar to Chen et. al. (2003) and Blundell and Powell (2004), and focus on cross section data. In this paper, we consider panel data and assume independence across cross-section units, but permit serial correlation at the individual level.

The rest of the paper is organized as follows. The model is presented in Section 2, followed by the discussion of the estimator and identification conditions in Section 3. Asymptotic theory is provided in Section 4. Monte Carlo simulations and empirical application are discussed in Sections 5 and 6, respectively. Section 7 provides concluding remarks.

Here we specify notations used in this paper. $||\cdot||$ denotes the Frobenius norm of vector, $||\cdot||_\infty$ denotes the uniform norm of vector, and vec denotes the vectorization operator. For arbitrary variables a and b, $a \vee b = \max(a, b)$, $\sim$ denotes "is distributed as," $\rightarrow_p$ denotes convergence in probability and $\rightarrow_d$ denotes convergence in distribution. For a sequence of random variables $\{X_N : N = 1, 2, \ldots\}$, $X_N = o_p(a_N)$ means that X_N/a_N converges in probability to zero as N goes to infinity, while $X_N = O_p(a_N)$ means that X_N/a_N is stochastically bounded.

2 The Model

Consider a binary response panel data model of the form

$$\begin{aligned} y^*_{it1} &= \mathbf{y}'_{it2}\boldsymbol{\alpha}_0 + \mathbf{x}'_{it}\boldsymbol{\beta}_0 - c_{i1} - u_{it1}, & (2.1)\\ y_{it1} &= 1\left[y^*_{it1} > 0\right], \quad i = 1, \ldots, N; t = 1, \ldots, T, & (2.2) \end{aligned}$$

where y^*_{it1} is a latent outcome for unit i in period t, y_{it1} is the observed binary outcome, and $1[\cdot]$ is an indicator function, which equals one if the condition in brackets holds and is zero otherwise; $\mathbf{y}_{it2}$ and $\mathbf{x}_{it}$ are $k_e \times 1$ and $k_x \times 1$ vectors of time-varying covariates, respectively, which may be correlated with the unobserved individual effects, c_{i1}. Here, we assume that variables in $\mathbf{y}_{it2}$ are continuous and could be endogenous in the sense that they may be correlated with the idiosyncratic error, u_{it1}, because of simultaneity, measurement error, or an omitted time-varying covariate. The parameters of interest in the above model are $\boldsymbol{\alpha}_0$ and $\boldsymbol{\beta}_0$. We consider the panel structure with large N and fixed T.

Let $\mathbf{z}_{it} = (\mathbf{x}'_{it}, \mathbf{z}'_{it1})'$ be a $k_z \times 1$ vector of instruments, $k_z \geq k_x + k_e$. Assume that endogenous variables, $\mathbf{y}_{it2}$, are determined by the following reduced form equations:

$$\mathbf{y}_{it2} = \mathbf{H}_{it}\boldsymbol{\gamma}_0 + \mathbf{c}_{i2} + \mathbf{u}_{it2}, \tag{2.3}$$

where $\mathbf{H}_{it}$ is a $k_e \times (k_e \cdot k_z)$ block diagonal matrix with vectors $\mathbf{z}'_{it}$ on its principal diagonal, $\boldsymbol{\gamma}_0 = (\boldsymbol{\gamma}'_{01}, \ldots, \boldsymbol{\gamma}'_{0k_e})'$, $\boldsymbol{\gamma}_{0j}$ is $k_z \times 1$ for $j = 1, \ldots, k_e$, and $\mathbf{c}_{i2}$ is a $k_e \times 1$ vector of unobserved effects that may be correlated with $\mathbf{H}_{it}$. In (2.3), instruments are assumed to be strictly exogenous conditional on the unobserved effect, so that for $\mathbf{Z}_i = (\mathbf{z}'_{i1}, \ldots, \mathbf{z}'_{iT})$, we have

$$\mathbf{u}_{it2}|\mathbf{Z}_i, \mathbf{c}_{i2} \sim \mathbf{u}_{it2}|\mathbf{z}_{it}, \mathbf{c}_{i2} \sim \mathbf{u}_{it2}. \tag{2.4}$$

To account for the correlation between covariates and unobserved effects, we model c_{i1} and $\mathbf{c}_{i2}$ as

$$\begin{aligned} c_{i1} &= \eta_1(\mathbf{Z}_i) + a_{i1}, && (2.5)\\ \mathbf{c}_{i2} &= \boldsymbol{\eta}_2(\mathbf{Z}_i) + \mathbf{a}_{i2}, && (2.6) \end{aligned}$$

where $\eta_1(\cdot) = E\left(c_{i1}|\mathbf{Z}_i\right)$, $a_{i1} = c_{i1} - E\left(c_{i1}|\mathbf{Z}_i\right)$, $\boldsymbol{\eta}_2(\cdot) = E\left(\mathbf{c}_{i2}|\mathbf{Z}_i\right)$, and $\mathbf{a}_{i2} = \mathbf{c}_{i2} - E\left(\mathbf{c}_{i2}|\mathbf{Z}_i\right)$. In practice, most studies use Mundlak's (1978) specification,

$$\begin{aligned} \eta_1(\mathbf{Z}_i) &= \bar{\mathbf{z}}_i'\boldsymbol{\xi}_{01}, && (2.7)\\ \boldsymbol{\eta}_2(\mathbf{Z}_i) &= \bar{\mathbf{H}}_i\boldsymbol{\xi}_{02}, && (2.8) \end{aligned}$$

where $\bar{\mathbf{z}}_i = T^{-1}\sum_{t=1}^{T}\mathbf{z}_{it}$, $\bar{\mathbf{H}}_i$ is a $k_e \times (k_e \cdot k_z)$ block diagonal matrix with vectors $\bar{\mathbf{z}}_i'$ on its principal diagonal, $\boldsymbol{\xi}_{02} = \left(\boldsymbol{\xi}_{021}', \ldots, \boldsymbol{\xi}_{02k_e}'\right)'$, $\boldsymbol{\xi}_{02j}$ is $k_z \times 1$, $j = 1, \ldots, k_e$, and Chamberlain's (1981) specification,

$$\begin{aligned} \eta_1(\mathbf{Z}_i) &= \mathbf{Z}_i\boldsymbol{\zeta}_{01}, && (2.9)\\ \boldsymbol{\eta}_2(\mathbf{Z}_i) &= \tilde{\mathbf{H}}_i\boldsymbol{\zeta}_{02}, && (2.10) \end{aligned}$$

where $\tilde{\mathbf{H}}_i$ is a $k_e \times (k_e \cdot k_z \cdot T)$ block diagonal matrix with vectors $\mathbf{Z}_i$ on its principal diagonal, and $\boldsymbol{\zeta}_{02} = \left(\boldsymbol{\zeta}_{021}', \ldots, \boldsymbol{\zeta}_{02k_e}'\right)'$, where ζ_{02j} is $(k_z \cdot T) \times 1$ for $j = 1, \ldots, k_e$. See, for example, Hsiao and Zhou (2018) for a discussion on these two specifications.[1]

Let $v_{it1} = a_{i1} + u_{it1}$ and $\mathbf{v}_{it2} = \mathbf{a}_{i2} + \mathbf{u}_{it2}$. Then, model (2.1) and (2.2) can be rewritten as

$$\begin{aligned} y_{it1} &= 1\left(\mathbf{y}_{it2}'\boldsymbol{\alpha}_0 + \mathbf{x}_{it}'\boldsymbol{\beta}_0 - \eta_1(\mathbf{Z}_i) - v_{it1} \geq 0\right), && (2.11)\\ \mathbf{y}_{it2} &= \mathbf{H}_{it}\boldsymbol{\gamma}_0 + \boldsymbol{\eta}_2(\mathbf{Z}_i) + \mathbf{v}_{it2}, \quad i = 1, \ldots, N; t = 1, \ldots, T. && (2.12) \end{aligned}$$

Remark 2.1 *One could also consider a more general version of the reduced form equations,*

$$\mathbf{y}_{it2} = \mathbf{h}(\mathbf{Z}_{it}) + \mathbf{v}_{it2}. \tag{2.13}$$

This specification only impacts the choice of the method for estimating the reduced form equations and can be easily accommodated within our current framework.

[1] It would be optimal to allow arbitrary correlation between the unobserved effect and explanatory variables. However, Chamberlain (2010) has shown that when the observed covariates have bounded support, identification is only possible when the error term has logistic distribution.

In equations (2.11)-(2.12), following Blundell and Powell (2004) and Semykina and Wooldridge (2018), we assume

$$v_{it1}|\mathbf{Z}_i, \mathbf{y}_{it2} \sim v_{it1}|\mathbf{Z}_i, \mathbf{v}_{it2} \sim v_{it1}|\mathbf{v}_{it2}, \tag{2.14}$$

$$\mathbf{v}_{it2}|\mathbf{Z}_i \sim \mathbf{v}_{it2}, \qquad i = 1, \ldots, N; t = 1, \ldots, T. \tag{2.15}$$

Thus, the endogeneity problem can be addressed using the control function approach. If $(v_{it1}, \mathbf{v}_{it2})$ have a bivariate normal distribution, the system can be estimated by MLE. In this paper, we do not make distributional assumptions, but instead consider a semiparametric approach proposed by Song (2017) and extend his method to panel data models.

3 Identification and Estimation

In this section, we consider the identification and least squares estimation for model (2.11)-(2.12).

3.1 Identification

For illustration purposes, we adapt the Mundlak's (1978) specification in this paper for $\eta_1(\mathbf{Z}_i)$ and $\boldsymbol{\eta}_2(\mathbf{Z}_i)$, i.e., $\eta_1(\mathbf{Z}_i) = \bar{\mathbf{z}}_i'\boldsymbol{\xi}_{01}$ and $\boldsymbol{\eta}_2(\mathbf{Z}_i) = \bar{\mathbf{H}}_i\boldsymbol{\xi}_{02}$, respectively.[2] Let $\mathbf{w}_{it} = (\mathbf{y}_{it2}, \mathbf{x}_{it}', \bar{\mathbf{z}}_i')'$ and $\boldsymbol{\pi}_0 = (\boldsymbol{\alpha}_0', \boldsymbol{\beta}_0', -\boldsymbol{\xi}_{01}')'$ with the dimension $\dim(\mathbf{w}_{it}) = k_w = k_e + k_x + k_z$. Then, model (2.11) can be rewritten as

$$y_{it1} = 1[\mathbf{w}_{it}'\boldsymbol{\pi}_0 - v_{it1} \geq 0], \quad i = 1, \ldots, N; t = 1, \ldots, T. \tag{3.1}$$

Furthermore, define $F_t(\cdot, \mathbf{v}_{it2})$ as a function mapping $\mathbb{R}$ to $[0,1]$. Throughout this paper we use it to define the conditional cumulative distribution function (CDF) of v_{it1} given the value of $\mathbf{v}_{it2}$, such that $F_t(\cdot, \mathbf{v}_{it2})$ is the true conditional CDF of v_{it1} and $\hat{F}_t(\cdot, \mathbf{v}_{it2})$ is the corresponding estimator. Note that the conditional CDF is time-specific, so that the error distribution may change over time (e.g. the error variance may change).

For identification of (3.1), it is required that the population parameter vector $\boldsymbol{\pi}_0 \in \Pi$ is unique and satisfies the following equality:

$$E\left(y_{it1}|\mathbf{w}_{it}, \mathbf{v}_{it2}\right) = E\left(y_{it1}|\mathbf{w}_{it}'\boldsymbol{\pi}_0, \mathbf{v}_{it2}\right). \tag{3.2}$$

To that end, we make the following assumption:

[2] Our approach is applicable for any function that is linear in parameters.

Assumption IC: (1) [*Conditional CDF*] For each t, function $F_t(\cdot, \mathbf{v}_{it2})$ is differentiable and strictly increasing in its first argument on a set $\mathcal{A}$ with positive probability under the distribution of $\mathbf{w}_{it}$;

(2) [*Continuity of* $\mathbf{w}_{it}$] Conditional on the control variable $\mathbf{v}_{it2}$, the vector $\mathbf{w}_{it}$ contains at least one continuously distributed component, $\mathbf{w}_{it}^{(1)}$, with nonzero coefficient;

(3) [*No colinearity*] The span of the remaining components of $\mathbf{w}_{it}$, $\mathbf{w}_{it}^{(-1)}$, contains no proper linear subspace which has probability 1 under the distribution of $\mathbf{w}_{it}$.

The above assumptions are the extensions of those in Rothe (2009). We note that Assumption IC(2) indicates the fact that having continuous endogenous variables in $\mathbf{w}_{it}$ is not sufficient for identification as the continuity might be from $\mathbf{v}_{it2}$. Therefore, it is necessary to have at least one continuous variable in either $\mathbf{x}_{it}$ or $\mathbf{z}_{it}$ to achieve identification.

Theorem 3.1 *If Assumption IC holds, then there exists a unique interior point* $\boldsymbol{\pi}_0 \in \Pi$ *when the relationship* $E\left(y_{it1}|\mathbf{w}_{it}, \mathbf{v}_{it2}\right) = E\left(y_{it1}|\mathbf{w}_{it}'\boldsymbol{\pi}_0, \mathbf{v}_{it2}\right)$ *holds for* $\mathbf{w}_{it} \in \mathcal{A}$ *with positive probability.*

See the Appendix for the proof.

3.2 Semiparametric Least Squares (SLS) Estimation

Once the model is identified, we can apply the least squares approach to estimate the parameters in the model, which is an M-estimator that utilizes the conditional mean of y_{it1} given $(\mathbf{w}_{it}, \mathbf{v}_{it2})$ to obtain moment conditions. Suppose the true value of error $\mathbf{v}_{it2}$ in the reduced form model is known. Then, the SLS estimator can be obtained by minimizing the following function,

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}[y_{it1} - E\left(y_{it1}|\mathbf{w}_{it}, \mathbf{v}_{it2}\right)]^2. \tag{3.3}$$

Based on the fact that the conditional expectation of y_{it1} (3.3) is equal to the conditional CDF of v_{it1}, it can be rewritten as

$$\begin{aligned} E[y_{it1}|\mathbf{w}_{it}, \mathbf{v}_{it2}] &= E(1[\mathbf{w}_{it}'\boldsymbol{\pi}_0 - v_{it1} \geq 0]|\mathbf{v}_{it2}) \\ &= F_t(\mathbf{w}_{it}'\boldsymbol{\pi}_0, \mathbf{v}_{it2}). \end{aligned} \tag{3.4}$$

If $F_t(\cdot, \mathbf{v}_{it2})$ is known, then the objective function for model (3.3) has the form of

$$\tilde{S}_N(\boldsymbol{\pi}) = \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}[y_{it1} - F_t(\mathbf{w}_{it}'\boldsymbol{\pi}, \mathbf{v}_{it2})]^2. \tag{3.5}$$

Unfortunately, the conditional CDF $F_t(\cdot, \mathbf{v}_{it2})$ is unknown in general, and thus $\tilde{S}_N(\boldsymbol{\pi})$ is infeasible in practice.

Following the strategy in Blundell and Powell (2004) and Rothe (2009), we can replace the conditional CDF (3.4) by the Nadaraya-Watson estimator,

$$\hat{F}_t(\mathbf{w}_{it}'\boldsymbol{\pi}, \mathbf{v}_{it2}) = \frac{\hat{p}_t(\mathbf{w}_{it}'\boldsymbol{\pi}, \mathbf{v}_{it2})}{\hat{q}_t(\mathbf{w}_{it}'\boldsymbol{\pi}, \mathbf{v}_{it2})}, \tag{3.6}$$

where

$$\hat{p}_t(\mathbf{w}_{it}'\boldsymbol{\pi}, \mathbf{v}_{it2}) = \frac{1}{N}\sum_{j\neq i}\kappa_1\left(\frac{\mathbf{w}_{it}'\boldsymbol{\pi} - \mathbf{w}_{jt}'\boldsymbol{\pi}}{h_1}\right)\kappa_2\left(\frac{\mathbf{v}_{it2} - \mathbf{v}_{jt2}}{h_2}\right)y_{jt1}, \tag{3.7}$$

$$\hat{q}_t(\mathbf{w}_{it}'\boldsymbol{\pi}, \mathbf{v}_{it2}) = \frac{1}{N}\sum_{j\neq i}\kappa_1\left(\frac{\mathbf{w}_{it}'\boldsymbol{\pi} - \mathbf{w}_{jt}'\boldsymbol{\pi}}{h_1}\right)\kappa_2\left(\frac{\mathbf{v}_{it2} - \mathbf{v}_{jt2}}{h_2}\right), \tag{3.8}$$

and $\kappa_1(\cdot)$ and $\kappa_2(\cdot)$ denote the kernel densities, and we can replace the control variable $\mathbf{v}_{it2}$ by $\hat{\mathbf{v}}_{it2}$ from the initial estimation of the reduced form model (2.12). For instance, we can estimate (2.12) by pooled OLS to obtain estimators $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\xi}}_2$. Then, $\hat{\mathbf{v}}_{it2}$ can be obtained by computing pooled OLS residuals,

$$\hat{\mathbf{v}}_{it2} = \mathbf{y}_{it2} - \mathbf{H}_{it}\hat{\boldsymbol{\gamma}} - \bar{\mathbf{Z}}_i\hat{\boldsymbol{\xi}}_2. \tag{3.9}$$

Consequently, we can replace the objective function (3.5) with

$$\hat{S}_N(\boldsymbol{\pi}) = \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}[y_{it1} - \hat{F}_t(\mathbf{w}_{it}'\boldsymbol{\pi}, \hat{\mathbf{v}}_{it2})]^2, \tag{3.10}$$

and the resulting semiparametric least squares estimator (SLS) of $\boldsymbol{\pi}_0$ is given by

$$\hat{\boldsymbol{\pi}}_{SLS} = \arg\min_{\boldsymbol{\pi}\in R^{k_w}} \hat{S}_N(\boldsymbol{\pi}), \tag{3.11}$$

where $k_w = k_e + k_x + k_z$ denotes the number of parameters in model (3.1).

Although the focus of this paper is on estimating parameter vector $(\boldsymbol{\alpha}_0', \boldsymbol{\beta}_0')$, researchers may also be interested in estimating the average structural function (ASF), which was introduced by Blundell and Powell (2004). The ASF returns the probability that $y_{it1} = 1$ for the given values of explanatory variables. For the panel data model formulated here, the ASF for period t is obtained by averaging over the marginal distribution of error v_{it2},

$$G_t(\mathbf{w}_t'\boldsymbol{\pi}_0) = \int F_t(\mathbf{w}_t'\boldsymbol{\pi}_0, \mathbf{v}_{t2})dF_{v_{t2}}, \tag{3.12}$$

where F_{vt2} is the CDF of $\mathbf{v}_{it2}$ at time t. The derivative of $G_t(\mathbf{w}_t'\boldsymbol{\pi}_0)$ with respect to variable x_k represents the marginal change in the response probability due to an exogenous change in x_k and is analogous to the partial effect of x_k in parametric models. Note that when evaluating partial effects, function $\eta_1(\mathbf{Z}_i)$ should be fixed, along with other observed variables (other than x_k). Moreover, it is possible to average ASF over t:

$$\frac{1}{T}\sum_{t=1}^{T}\int F_t(\mathbf{w}_t'\boldsymbol{\pi}_0, \mathbf{v}_{t2})dF_{v_{t2}}. \tag{3.13}$$

To obtain a consistent estimator of ASF, Blundell and Powell (2004) suggest estimating $F_t(\mathbf{w}_t'\boldsymbol{\pi}_0, \mathbf{v}_{t2})$ using the Nadaraya-Watson estimator, where y_{it1} is nonparametrically regressed on $\mathbf{w}_{it}'\hat{\boldsymbol{\pi}}$ and $\hat{\mathbf{v}}_{it2}$. Rothe (2009) estimates $F_t(\mathbf{w}_t'\boldsymbol{\pi}_0, \mathbf{v}_{t2})$ by the local linear estimator. Then, the ASF for period t can be estimated as

$$\frac{1}{N}\sum_{i=1}^{N}\hat{F}_t(\mathbf{w}_t'\hat{\boldsymbol{\pi}}_{SLS}, \hat{\mathbf{v}}_{it2}) \tag{3.14}$$

for particular values of $\mathbf{w}_t$, which can be sample means, medians, or any other values.

4 Asymptotics

In this section, we provide the asymptotic results for the above SLS estimator.

4.1 Consistency

To show that the SLS estimator in (3.11) is consistent, we assume

Assumption CON: (1) [*Compact space*] Both $\boldsymbol{\pi}_0$ and $\hat{\boldsymbol{\pi}}$ belong to a compact parameter space Π and are interior points;

(2) [*Kernel function*] $\kappa_1(\cdot) : \mathbb{R} \to \mathbb{R}$ and $\kappa_2(\cdot) : \mathbb{R}^{k_e} \to \mathbb{R}$ satisfy: $\int \kappa_j(u)du = 1$; $\int u^s\kappa_j(u)du = 0$, for $s = 1, \cdots, r-1$ for some $r \in \mathbb{N}$; $\int u^r\kappa_j(u)dx < \infty$; $\kappa_j(x)$ is r times continuously differentiable for $j = 1, 2$;

(3) [*Bandwidth*] There exist constants c_1 and δ_1 such that $c_1 > 0$ and $1/(2r+1) \leq \delta_1 < 1/(4k_e)$. Also, there exist $c_{2j} > 0$ and $1/(2r+1) \leq \delta_{2j} < 1/(4k_e)$ for $j = 1, \cdots, k_e$. Then the bandwidth h_1 and h_2 satisfy $h_1 = c_1N^{-\delta_1}$, $h_2 = (h_{21}, h_{22}, \cdots, h_{2k_e})$, where $h_{2j} = c_{2j}N^{-\delta_{2j}}$ for $j = 1, \cdots, k_e$.

(4) [*Lipchitz continuity*] For each t, the conditional CDF $F_t(\cdot, \mathbf{v}_{it2})$ is r times continuously differentiable for some $r \in \mathbb{N}$, and its r-th derivatives are Lipchitz continuous;

(5) [*Boundedness*] The estimator $\hat{\mathbf{v}}_{it2}$ satisfies $\max_{i,t} \|\hat{\mathbf{v}}_{it2} - \mathbf{v}_{it2}\| = o_p(N^{-1/4})$. Also define $\mathbf{D}_{it} = (\mathbf{H}_{it}, \bar{\mathbf{H}}_i)$ and $\hat{\mathbf{v}}_{it2} - \mathbf{v}_{it2} = N^{-1}\sum_{j=1}^{N}\sum_{s=1}^{T} g(\mathbf{D}_{it}, \mathbf{D}_{js})\boldsymbol{\psi}_{js} + \mathbf{r}_{it}$, where $\boldsymbol{\psi}_{js}$ is an influence function with $E(\boldsymbol{\psi}_{js}|\mathbf{D}_{js}) = 0$, $Var(\boldsymbol{\psi}_{js}^2|\mathbf{D}_{js}) < \infty$, $E\left[g(\mathbf{D}_{it}, \mathbf{D}_{jt})^2\right] = o(N)$ and $\max_{i,t}\|\mathbf{r}_{it}\| = o_p(N^{-1/2})$.

Remark 4.1 *Assumptions CON(2)-(3) define a standard bias-reducing kernel of order r, which are used to reduce asymptotic bias in the estimator of $F_t(\cdot, \mathbf{v}_{it2})$. The inequality regarding δ_1 and δ_{2j} indicates $r > 2k_e - 1/2$. Therefore, in practice it is suggested using higher order kernels. CON(4) is standard for kernel smoothing theory. Finally, CON(5) regulates the estimation bias in $\mathbf{v}_{it2}$ in the reduced form equations and requires that it is asymptotically approximated by the influence function. Rothe (2009) and Song (2017) make similar assumptions. Note that $\hat{\mathbf{v}}_{it2} - \mathbf{v}_{it2} = \mathbf{H}_{it}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_{\mathbf{0}}) + (\hat{\boldsymbol{\eta}}_2(\mathbf{Z}_i) - \boldsymbol{\eta}_2(\mathbf{Z}_i))$, so that CON(5) will hold as long as the estimators $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\eta}}_2(\cdot)$ are consistent as $N \to \infty$.*

The consistency of the SLS estimator (3.11) is established in the following theorem.

Theorem 4.2 *If Assumptions IC and CON hold, then as $N \to \infty$, we have*

$$\hat{\boldsymbol{\pi}}_{SLS} \to_p \boldsymbol{\pi}_0. \tag{4.1}$$

See the Appendix for the proof.

4.2 Asymptotic Distribution of the SLS Estimator

Now let's turn to the asymptotic distribution of $\hat{\boldsymbol{\pi}}_{SLS}$. Since the criterion function (3.10) depends on the nonparametric estimators of the conditional CDFs, while those nonparametric estimators themselves depend on the estimators of reduced-form parameters, the usual approach to deriving the asymptotic distribution needs to be modified. Here, we follow the classical analysis of Chen et. al. (2003) to derive the asymptotic distribution of $\hat{\boldsymbol{\pi}}_{SLS}$.

To begin with, define

$$h = \{(\mathbf{v}(\mathbf{y}_{it2}, \mathbf{z}_{it}), F_t(\cdot, \mathbf{v}(\mathbf{y}_{it2}, \mathbf{z}_{it})) : i = 1\cdots, N, t = 1\cdots, T\}$$

where function $\mathbf{v}(\mathbf{y}_{it2}, \mathbf{z}_{it}) = \mathbf{v}_{it2} = \mathbf{y}_{it2} - \mathbf{H}_{it}\boldsymbol{\gamma} - \bar{\mathbf{H}}_i\boldsymbol{\xi}_2$. The parameter space of h is denoted as $\mathcal{H}$, and the norm $\|h\|_{\mathcal{H}}$ is defined as

$$\|h\|_{\mathcal{H}} = \max_{i,t}\left(\|\mathbf{v}_{it2}\|_\infty \vee \|F_t(\cdot, \mathbf{v}_{it2})\|_\infty\right).$$

Then, we generate the moment function that is rooted in our estimator. Namely,

$$M(\boldsymbol{\pi}, h) = E\left[\sum_{t=1}^{T}(y_{it1} - F_t(\mathbf{w}_{it}'\boldsymbol{\pi}, \mathbf{v}_{it2})) \cdot \frac{\partial F_t(\mathbf{w}_{it}'\boldsymbol{\pi}, \mathbf{v}_{it2})}{\partial \boldsymbol{\pi}}\right] = E\left[m_i(\boldsymbol{\pi}, h)\right], \tag{4.2}$$

We also denote the sample analogue of $M(\boldsymbol{\pi}, h)$ as $M_N(\boldsymbol{\pi}, h) = 1/N \sum_{i=1}^{T} m_i(\boldsymbol{\pi}, h)$.

Based on the above definition, we are able to show that the conditions in Chen et. al. (2003) are satisfied for this moment function. Hence, we can establish the asymptotic normality as in Chen et. al. (2003). Before we state the main results, we need to make additional assumptions.

Assumption NR: (1) [*Entropy condition*] $\int_0^\infty \sqrt{\log N(\epsilon, \mathcal{H}, \|\cdot\|_{\mathcal{H}})} d\epsilon \leq \infty$, where $N(\epsilon, \mathcal{H}, \|\cdot\|_{\mathcal{H}})$ is the covering number with respect to $\|\cdot\|_{\mathcal{H}}$ norm of classes of function h;

(2) [*Rank condition*] The matrix $\Omega = E\left[\sum_{t=1}^{T} \frac{\partial F_t(\mathbf{w}_{it}'\boldsymbol{\pi}_0, \mathbf{v}_{it2})}{\partial \boldsymbol{\pi}_0} \frac{\partial F_t(\mathbf{w}_{it}'\boldsymbol{\pi}_0, \mathbf{v}_{it2})}{\partial \boldsymbol{\pi}_0'}\right]$ is of full rank.

We note that Assumption NR(1) regulates the complexity of function space of h, which is a necessary condition for the proof of stochastic equicontinuity. Similar condition has also been imposed by most nonparametric estimators (e.g., Rothe (2009) and Song (2017)). Assumption NR(2) requires the asymptotic variance-covariance matrix to be positive definite, which is a standard assumption in the literature.

The asymptotic normality of $\hat{\boldsymbol{\pi}}_{SLS}$ is based on Chen et. al. (2003), who establish the asymptotic normality for a semiparametric criterion function. Firstly, denote the ordinary derivative of $M(\boldsymbol{\pi}, h)$ with respect to $\boldsymbol{\pi}$ as

$$\Gamma_1(\boldsymbol{\pi}, h) = \frac{\partial M(\boldsymbol{\pi}, h)}{\partial \boldsymbol{\pi}}, \tag{4.3}$$

for all $\boldsymbol{\pi} \in \Pi$ and the pathwise derivative of $M(\boldsymbol{\pi}, h)$ as $\Gamma_2(\boldsymbol{\pi}, h)\left(\bar{h} - h\right)$ for all $\bar{h} \in \mathcal{H}$.

For future purposes, we also define

$$\Psi_i(\boldsymbol{\pi}, h) = \sum_{t=1}^{T} E\left[\sum_{s=1}^{T} \frac{-\partial F_t(\mathbf{w}_{js}'\boldsymbol{\pi}, \mathbf{v}_{js2})}{\partial \boldsymbol{\pi}} \frac{\partial F_t(\mathbf{w}_{js}'\boldsymbol{\pi}, \mathbf{v}_{js2})}{\partial \mathbf{v}_{js2}'} g(\mathbf{D}_{js}, \mathbf{D}_{it}) | \mathbf{D}_{it}\right] \boldsymbol{\psi}_{it}, \tag{4.4}$$

and

$$\boldsymbol{\phi}_i(\boldsymbol{\pi}, h) = m_i(\boldsymbol{\pi}, h) + \Psi_i(\boldsymbol{\pi}, h). \tag{4.5}$$

We also let

$$\mathbf{V} = E\left(\boldsymbol{\phi}_i(\boldsymbol{\pi}_0, h_0)\boldsymbol{\phi}_i(\boldsymbol{\pi}_0, h_0)'\right). \tag{4.6}$$

As a result, provided the existence of both $\Gamma_1(\boldsymbol{\pi}_0, h_0)$ and $\Gamma_2(\boldsymbol{\pi}_0, h_0)[\hat{h} - h_0]$, intuitively we can apply the standard Taylor expansion of the moment function $M_N(\hat{\boldsymbol{\pi}}_{SLS}, \hat{h}) = 0$ around

$(\boldsymbol{\pi}_0, h_0)$ and obtain

$$\sqrt{N}(\hat{\boldsymbol{\pi}}_{SLS} - \boldsymbol{\pi}_0) = -\sqrt{N}\Gamma_1(\boldsymbol{\pi}_0, h_0)^{-1}\left(M_N(\boldsymbol{\pi}_0, h_0) + \Gamma_2(\boldsymbol{\pi}_0, h_0)[\hat{h} - h_0]\right). \qquad (4.7)$$

The asymptotic normality of $\hat{\boldsymbol{\pi}}_{SLS}$ can then be achieved by showing that the right-hand side of (4.7) converges to normal distribution. In the Appendix, it is shown that the sum of $M_N(\boldsymbol{\pi}_0, h_0)$ and $\Gamma_2(\boldsymbol{\pi}_0, h_0)[\hat{h} - h_0]$ is $O_p(N^{-1/2})$, so that the asymptotic normality can be verified by standard central limit theorem. This is summarized in the following theorem:

Theorem 4.3 *Under Assumptions CI, CON and NR, as $N \to \infty$, we have*

$$\sqrt{N}(\hat{\boldsymbol{\pi}}_{SLS} - \boldsymbol{\pi}_0) \to_d \mathcal{N}\left(0, \Omega^{-1}\mathbf{V}\Omega^{-1}\right), \qquad (4.8)$$

where $\Omega = E\left[\sum_{t=1}^{T} \frac{\partial F_t(\mathbf{w}_{it}'\boldsymbol{\pi}_0, \mathbf{v}_{it2})}{\partial \boldsymbol{\pi}_0} \frac{\partial F_t(\mathbf{w}_{it}'\boldsymbol{\pi}_0, \mathbf{v}_{it2})'}{\partial \boldsymbol{\pi}_0'}\right]$, *and* $\mathbf{V}$ *is provided in (4.6).*

See the Appendix for a proof.

To perform inference, one needs to estimate the asymptotic variance-covariance (VC) matrix in (4.8). Unfortunately, the VC matrix in (4.8) depends on a number of unknown and complicated functions. For instance, matrix $\mathbf{V}$ depends on the influence function, which depends on the choice of the estimator of $(\boldsymbol{\gamma}_0', \boldsymbol{\xi}_{02}')'$. In order to obtain a feasible estimator of the asymptotic VC matrix, one can either consider a nonparametric panel bootstrap procedure as in Chen et. al. (2003) or a modified sample moment estimator as in Rothe (2009). For the bootstrap approach, following Chen et. al. (2003), we let $\mathbf{y}_{i1} = (y_{i1,1}, y_{i1,2}, \ldots, y_{i1,T})'$, $\mathbf{Y}_{i2} = (\mathbf{y}_{i2,1}, \mathbf{y}_{i2,2}, \ldots, \mathbf{y}_{i2,T})'$, and $\mathbf{Z}_i = (\mathbf{z}_{i1}, \mathbf{z}_{i2}, \ldots, \mathbf{z}_{iT})'$. Then we can draw $\{(\mathbf{y}_{i1}^*, \mathbf{Y}_{i2}^*, \mathbf{Z}_i^*)\}_{i=1}^N$ randomly with replacement from the original data $\{(\mathbf{y}_{i1}, \mathbf{Y}_{i2}, \mathbf{Z}_i)\}_{i=1}^N$. The SLS estimator using the bootstrap sample is given by

$$\hat{\boldsymbol{\pi}}^* = \arg\min_{\boldsymbol{\pi} \in R^{k_w}} \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[y_{it1}^* - \hat{F}_t(\mathbf{w}_{it}^{*\prime}\boldsymbol{\pi}, \hat{\mathbf{v}}_{it2}^*)\right]^2. \qquad (4.9)$$

Chen et al. (2003) show that $\sqrt{N}(\hat{\boldsymbol{\pi}}^* - \hat{\boldsymbol{\pi}}_{SLS})$ has the same asymptotic distribution as $\sqrt{N}(\hat{\boldsymbol{\pi}}_{SLS} - \boldsymbol{\pi}_0)$.

On the other hand, as argued by Rothe (2009), the disadvantage of the nonparametric bootstrap procedure is that it is extremely computational extensive. Therefore, for practical purposes, here we follow Rothe (2009) to provide an estimator for the VC matrix in (4.8). Following the derivation in the Appendix, we are able to show a uniform convergence of $\hat{F}_t(\cdot, \hat{\mathbf{v}}_{it2})$ to $F_t(\cdot, \mathbf{v}_{it2})$ which also holds for the derivative of F_t. Then Ω can be estimated using a sample analogue,

$$\hat{\Omega} = \frac{1}{N}\sum_{i=1}^{N}\left[\sum_{t=1}^{T}\frac{\partial \hat{F}_t(\mathbf{w}_{it}'\hat{\boldsymbol{\pi}}_{SLS}, \hat{\mathbf{v}}_{it2})}{\partial \hat{\boldsymbol{\pi}}}\frac{\partial \hat{F}_t(\mathbf{w}_{it}'\hat{\boldsymbol{\pi}}_{SLS}, \hat{\mathbf{v}}_{it2})}{\partial \hat{\boldsymbol{\pi}}'}\right]. \tag{4.10}$$

For matrix $\mathbf{V}$ in (4.8), the exact form of the influence function depends on the particular estimator we adopt for equation (2.12). When function $\boldsymbol{\eta}_2(\cdot)$ in (2.12) is unknown, one can consider semi/non-parametric estimation, and the resulting influence function can be constructed based on the semi/non-parametric regression.[3] For illustration purposes, here we consider a simple case of (2.12) (e.g., $\boldsymbol{\eta}_2(\cdot)$ is specified in (2.5)) and use the pooled OLS estimator. The associated terms in the asymptotic expansion of $\hat{\mathbf{v}}_{it2} - \mathbf{v}_{it2}$ are defined as follows:

$$g(\mathbf{D}_{js}, \mathbf{D}_{it}) = -\mathbf{D}_{js}, \tag{4.11}$$

$$\boldsymbol{\psi}_{it} = \left(\frac{1}{N}\sum_{l=1}^{N}\sum_{p=1}^{T}\mathbf{D}_{lp}'\mathbf{D}_{lp}\right)^{-1}\mathbf{D}_{it}'\mathbf{v}_{it2}, \tag{4.12}$$

where $\mathbf{D}_{it} = (\mathbf{H}_{it}, \bar{\mathbf{H}}_i)$. Consequently, $\Psi_i(\boldsymbol{\pi}, h)$ in (4.4) can be rewritten as

$$\Psi_i(\boldsymbol{\pi}, h) = \sum_{t=1}^{T}\sum_{s=1}^{T} E\left[\frac{-\partial F_t(\mathbf{w}_{js}'\boldsymbol{\pi}, \mathbf{v}_{js2})}{\partial \boldsymbol{\pi}}\frac{\partial F_t(\mathbf{w}_{js}'\boldsymbol{\pi}, \mathbf{v}_{js2})}{\partial \mathbf{v}_{js2}'}\mathbf{D}_{js}\left(\frac{1}{N}\sum_{l=1}^{N}\sum_{p=1}^{T}\mathbf{D}_{lp}'\mathbf{D}_{lp}\right)^{-1}\right]\mathbf{D}_{it}'(-\mathbf{v}_{it2}), \tag{4.13}$$

and thus $\mathbf{V}$ can be estimated as follows

$$\hat{\mathbf{V}} = \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\phi}}_i(\hat{\boldsymbol{\pi}}_{SLS}, \hat{h})\hat{\boldsymbol{\phi}}_i'(\hat{\boldsymbol{\pi}}_{SLS}, \hat{h}), \tag{4.14}$$

where

$$\begin{aligned}
&\hat{\boldsymbol{\phi}}_i(\hat{\boldsymbol{\pi}}_{SLS}, \hat{h}) \\
= &\left[\frac{1}{N}\sum_{j=1}^{N}\sum_{s=1}^{T}\frac{-\partial \hat{F}_t(\mathbf{w}_{js}'\hat{\boldsymbol{\pi}}_{SLS}, \hat{\mathbf{v}}_{js2})}{\partial \hat{\boldsymbol{\pi}}}\frac{\partial \hat{F}_t(\mathbf{w}_{js}'\hat{\boldsymbol{\pi}}_{SLS}, \hat{\mathbf{v}}_{js2})}{\partial \hat{\mathbf{v}}_{js2}'}\mathbf{D}_{js}\right]\left(\frac{1}{N}\sum_{l=1}^{N}\sum_{p=1}^{T}\mathbf{D}_{lp}'\mathbf{D}_{lp}\right)^{-1}\sum_{t=1}^{T}\mathbf{D}_{it}'\left(-\hat{\mathbf{v}}_{it2}\right) \\
&+\sum_{t=1}^{T}\left[y_{it1} - \hat{F}_t(\mathbf{w}_{it}'\hat{\boldsymbol{\pi}}_{SLS}, \hat{\mathbf{v}}_{it2})\right]\cdot\frac{\partial \hat{F}_t(\mathbf{w}_{it}'\hat{\boldsymbol{\pi}}_{SLS}, \hat{\mathbf{v}}_{it2})}{\partial \hat{\boldsymbol{\pi}}}.
\end{aligned} \tag{4.15}$$

Combining (4.10)-(4.15) yeilds the estimator for the VC matrix in (4.8).

[3]See Rothe (2009, p. 55) for an example of the influence function based on a kernel regression.

5 Monte Carlo Simulation

To gain insights on the performance of the proposed estimator in finite samples we conduct limited Monte Carlo experiments. We consider the following data generating process (DGP) with a single endogenous variable:

$$y_{it1} = 1[y_{it2}\beta_1 + z_{it1}\beta_2 - c_{i1} > u_{it1}], \tag{5.1}$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$. The endogenous variable, y_{it2}, is generated by

$$y_{it2} = z_{it1}\gamma_1 + z_{it2}\gamma_2 + z_{it3}\gamma_3 + c_{i2} + u_{it2}, \tag{5.2}$$

where $\mathbf{z}_{it} = (z_{it1}, z_{it2}, z_{it3})'$ are exogenous variables. Let $\bar{\mathbf{z}}_i = T^{-1}\sum_{t=1}^{T}\mathbf{z}_{it}$. The unobserved individual effects are generated as

$$\begin{aligned} c_{i1} &= \bar{\mathbf{z}}_i'\boldsymbol{\xi}_1 + a_{i1}, && (5.3)\\ c_{i2} &= \eta_2 + \bar{\mathbf{z}}_i'\boldsymbol{\xi}_2 + a_{i2}. && (5.4) \end{aligned}$$

Note that inserting (5.3) into (5.1) yields

$$y_{it1} = 1[\mathbf{w}_{it}'\boldsymbol{\pi} - v_{it1} \geq 0],$$

where $\mathbf{w}_{it} = (y_{it2}, z_{it1}, \bar{\mathbf{z}}_i')'$, $\boldsymbol{\pi} = (1, \beta, -\boldsymbol{\xi}_1')'$ and $v_{it1} = a_{i1} + u_{it1}$. Substituting (5.4) into (5.2) gives

$$y_{it2} = \eta_2 + \mathbf{z}_{it}'\boldsymbol{\gamma} + \bar{\mathbf{z}}_i'\boldsymbol{\xi}_2 + v_{it2},$$

where $v_{it2} = a_{i2} + u_{it2}$. Following Semykina and Wooldridge (2018), the exogenous variables z_{itj} are generated by

$$z_{itj} = b_{ij} + \varepsilon_{itj}, \text{ for } j = 1, 2, 3,$$

where b_{ij} are independent across i and distributed as $N(0, 1/4)$ with $Corr\left(b_{ij}, b_{ij'}\right) = 1/4$, for $j' = 1, 2, 3$ and $j' \neq j$; ε_{itj} are independent across i and t and distributed as $N(0, 3/4)$. The unobserved effects are generated as $a_{i1} = \rho a_{i2} + \epsilon_{it}$, where $a_{i2} \sim IIDN\left(0, \sigma_a^2\right)$ and $\epsilon_{it} \sim IIDN\left[0, (1-\rho)^2\sigma_a^2\right]$. The idiosyncratic errors are generated as $u_{it1} = \rho u_{it2} + e_{it}$. We consider the following designs for u_{it2} and e_{it}:

- Design I (Gaussian errors): For all i and t,

$$u_{it2} \sim IIDN\left(0, \sigma_u^2\right), \quad e_{it} \sim IIDN\left[0, (1-\rho)^2\sigma_u^2\right]. \tag{5.5}$$

- Design II (Non-Gaussian errors): For all i and t,

$$u_{it2} \sim IID\left\{\sigma_u\left[\chi^2(1)-1\right]/\sqrt{2}\right\}, \quad e_{it} \sim IID\left\{\sigma_e\left[\chi^2(1)-1\right]/\sqrt{2}\right\}, \tag{5.6}$$

where $\chi^2(1)$ denotes a chi-square random variable with one degree of freedom.

The true parameter values are set to $\beta_1 = 1$, $\beta_2 = 1$, $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \gamma_3)' = (2/3, 2/3, 1/3)'$, $\eta_2 = 1$, $\boldsymbol{\xi}_1 = (-1/3, -1/3, -1/3)'$, $\boldsymbol{\xi}_2 = (1/3, 1/3, 1/3)'$, $\rho = 0.5$, $(\sigma_a)^2 = 1/4$, $\sigma_u^2 = 3/4$, and $(\sigma_e)^2 = 5$. We focus on the estimated relative effect, $\hat{\beta}_2/\hat{\beta}_1$.

In computing the SLS estimator, we treat the bandwidths, h_1 and h_2, as additional parameters of the objective function as in Rothe (2009), and perform the following minimization with respect to $\boldsymbol{\pi}$, h_1 and h_2 jointly:

$$\left(\hat{\boldsymbol{\pi}}'_{SLS}, \hat{h}_1, \hat{h}_2\right)' = \arg \min_{(\boldsymbol{\pi}', h_1, h_2)'} \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} [y_{it1} - \hat{F}_t(\mathbf{w}'_{it}\boldsymbol{\pi}, \hat{v}_{it2})]^2, \tag{5.7}$$

where

$$\hat{F}_t(\mathbf{w}'_{it}\boldsymbol{\pi}, \hat{v}_{it2}) = \frac{\sum_{j \neq i} \kappa_1\left(\frac{\mathbf{w}'_{it}\boldsymbol{\pi} - \mathbf{w}'_{jt}\boldsymbol{\pi}}{h_1}\right) \kappa_2\left(\frac{\hat{v}_{it2} - \hat{v}_{jt2}}{h_2}\right) y_{jt1}}{\sum_{j \neq i} \kappa_1\left(\frac{\mathbf{w}'_{it}\boldsymbol{\pi} - \mathbf{w}'_{jt}\boldsymbol{\pi}}{h_1}\right) \kappa_2\left(\frac{\hat{v}_{it2} - \hat{v}_{jt2}}{h_2}\right)}, \quad t = 1, \ldots, T,$$

and $\hat{v}_{it2} = y_{it2} - \hat{\eta}_2 - \mathbf{z}'_{it}\hat{\boldsymbol{\gamma}} - \bar{\mathbf{z}}'_i\hat{\boldsymbol{\xi}}_2$, for $i = 1, \ldots, N$, is the residual from the initial pooled OLS estimation of the reduced form model for y_{it2}. After obtaining $\hat{v}_{it2}$, but prior to solving the minimization problem, all regressors in $\mathbf{w}_{it}$ and $\hat{v}_{it2}$ were orthogonalized by the Cholesky transformation. The estimated parameters are then recovered by the reverse transformation after the minimization.[4] Second-order Gaussian kernels were used in the computation.

For comparison, we computed the Two-step Probit estimator and the Two-step 2SLS estimator for a linear probability model, where the endogeneity was corrected using a control function approach, and the unobserved heterogeneity was modeled using the Mundlak device.

We consider combinations of $N = 250$, 500, $1,000$, and $T = 3$, 5, 10. The number of replication is $R = 1,000$ for each experiment. The estimation results for the relative effects, $\hat{\beta}_2/\hat{\beta}_1$, are presented in Tables 1 and 2. For each estimator of β_2/β_1, we report bias, standard deviation (SD), root mean squared error (RMSE), median absolute deviation (MAD) and interquantile range (IQR). In addition, we compute the bootstrap standard errors averaged across 1,000 Monte Carlo samples. The number of bootstrap replications is $B = 200$ for the Probit

[4]Specifically, let $\mathbf{x}_{it} = (\mathbf{w}'_{it}, \hat{v}_{it2})'$ and $\boldsymbol{\Omega}$ be the sample covariance matrix of $\mathbf{x}_{it}$. By the Cholesky decomposition we have $\boldsymbol{\Omega} = \mathbf{L}\mathbf{L}'$, where $\mathbf{L}$ is a lower triangular matrix. Then we use $\tilde{\mathbf{x}}_{it} = \mathbf{L}'\mathbf{x}_{it}$ in the optimization given by (5.7) and obtain a vector of estimated parameters $\tilde{\boldsymbol{\pi}}$. Finally the SLS estimates, $\hat{\boldsymbol{\pi}}$, are recovered from all but the last parameters in $\mathbf{L}(\tilde{\boldsymbol{\pi}}', 0)'$.

and 2SLS estimators, and set to $B = 100$ for the SLS estimator due to high computational intensity. Analytical standard errors of the SLS estimator, averaged over all replications, are reported in the last column in Tables 1 and 2.

Several patterns emerge. When errors have Gaussian distribution (Table 1), parametric estimators tend to have smaller bias and less dispersion. However, as the sample size grows, the advantages of parametric methods diminish. For all N and T, the Two-step Probit estimator has the smallest RMSE thanks to both less bias and more precision. The latter finding is as expected, because under Design I the Probit model is correctly specified. On the other hand, the proposed SLS has quite comparable performance.

When the error distribution is not Gaussian (Table 2), the SLS estimator tends to have a larger bias when $N = 250$. However, both the bias and the dispersion of the SLS estimator decrease with N. As a result, for each T, it has the smallest bias, RMSE, MAD, and IQR when $N = 1,000$. Note that bootstrap standard errors behave well in all experiments (both tables). Analytical standard errors are reasonably close to standard deviations for $N = 1,000$, but are rather inaccurate for smaller N. Hence, in empirical applications it is better to use bootstrap standard errors when performing the SLS estimation.

6 Empirical Application

As an empirical application we study the labor force participation decisions of married women using the Panel Study of Income Dynamics (PSID) data, years 1982-1984. The sample includes 742 white women, ages 20-57, who were married in all three years. Thus, it is a balanced three-year panel.

We consider the following model:

$$inlf_{it} = 1[\beta_1 nwfinc_{it} + \beta_2 age_{it} + \beta_3 educ_{it} + \beta_4 kids_{it} + \alpha_t - c_{i1} > u_{it1}], \tag{6.1}$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$, where $inlf_{it}$ is an indicator equal to one if woman i reported positive work hours in year t, $nwfinc$ is non-wife income, defined as a difference between the total family income and woman's earnings (measured in thousands of dollars), age and $educ$ are the woman's age and years of schooling, respectively, $kids$ is the number of children less than six years old, α_t is a year-specific intercept, and c_{i1} is the unobserved individual effect. A significant portion of the non-wife income is husband's earnings. As long as spouses' labor market decisions are determined simultaneously, spousal earnings and, therefore, non-wife income is potentially endogenous. The reduced form equation is modeled as

$$nwfinc_{it} = \gamma_1 age_{it} + \gamma_2 educ_{it} + \gamma_3 kids_{it} + \gamma_4 hage_{it} + \gamma_5 heduc_{it} + \eta_t + c_{i2} + u_{it2}, \qquad (6.2)$$

where *hage* and *heduc* are husband's age and years of education, respectively. The instruments (*hage* and *heduc*) are expected to affect non-wife income through their impact on husband's earnings, but should have no partial effect on the woman's labor force participation decision. We assume all explanatory variables except $nwfinc$ are strictly exogenous conditional on the unobserved effect. Table 3 reports the descriptive statistics of the employed variables.

To account for a non-zero correlation between the unobserved effect and covariates, the time means of exogenous variables were included in both (6.1) and (6.2). The time means of $educ_{it}$ and $heduc_{it}$ were not included due to no variation over time. This means that the effect of education cannot be separated from the impact of the unobserved heterogeneity and should be interpreted with caution. However, it does not affect the reliability of the estimated effects of other variables because $educ_{it}$ and $heduc_{it}$ serve as sufficient controls, as they capture both the direct effects and the impact of the unobserved heterogeneity.

We estimate model parameters using the SLS and instrumental variables probit estimator (IV-Probit), which estimates both equations jointly by full MLE. For comparison, we also report results from a usual probit regression and a semiparametric least squares estimation that does not account for endogeneity (SLS-exog.). Because semiparametric methods can only estimate relative effects, we focus on discussing coefficient ratios. Moreover, only bootstrap standard errors are computed for the SLS and SLS-exog., as they behaved better in simulations. Table 4 summarizes estimation results.

One notable finding is that $\hat{\beta}_{nwinc}/\hat{\beta}_{age}$ are very similar for the IV-Probit and SLS. In contrast, the estimators that do not account for endogeneity produce much smaller relative effects of the non-wife income. Similarly, the semiparametric estimates of the coefficient ratios for education and number of children are slightly smaller when not correcting for endogeneity.

Finally, we are interested in how the propensity of working would change as non-wife income changes while fixing all the other explanatory variables at their sample means. To this end, we estimate the average structural function (ASF) derived from the SLS estimation following (3.14), where F_t is estimated by the Nadaraya-Watson estimator with bandwidths selected by the leave-one-out least-squares cross-validation. Figures 1–3 present the estimated ASF for each year of our panel, respectively, and Figure 4 displays the ASF averaged over all three years. In each figure, the ASF is plotted over the 5–95% range of the non-wife income distribution observed in our sample (\$7,721–\$68,200). To examine the effect of correcting for endogeneity on the estimated marginal probability, we also estimate the ASF derived from the SLS-exog. estimation. As can be seen from the graphs, the probability of working decreases monotonically and significantly with non-wife income for all years. When endogeneity is accounted for, the probability of working drops from approximately 0.85 to 0.4, with slower (faster) decline at lower (higher) non-wife income level. In contrast, if endogeneity of non-wife income is neglected, we tend to underestimate (overestimate) the probability of working when non-wife income is low (high). In all but the third year of our panel, ignoring endogeneity results in a much flatter and almost linear estimate of ASF, with estimated probability of working lying in a narrower range of 0.8 to 0.6. These results exemplify the bias in the estimated response probability due to failure to control for endogeneity.

7 Conclusion

In this paper, we considered a semiparametric least squares method for estimating parameters in binary response panel data models with endogenous variables. The method permits a non-zero correlation between the covariates and unobserved time-invariant effect and accounts for endogeneity caused by the correlation with an idiosyncratic error. We show that the estimator is $\sqrt{N}$-consistent and asymptotically normal. The results of Monte Carlo experiments indicate that the estimator performs well in finite samples. When the error distribution is not normal, the SLS estimator has a smaller bias and less variance than a parametric IV-Probit estimator.

An important advantage of the considered estimator is that it can be used for estimating panel data models, where time dependence is present. Moreover, the employed least squares method is less computationally intense than other existing approaches. Future research could consider extending the method to models with cross sectional dependence.

Table 1: Small sample properties of the estimators of β_2/β_1 for Design I (Gaussian errors)

			Bias	SD	RMSE	MAD	IQR	Bootstrap SE	Average SE
$T = 3$	$N = 250$	SLS	0.0580	0.4360	0.4397	0.2330	0.4759	0.4235	0.2270
		Two-step Probit	0.0273	0.2688	0.2701	0.1673	0.3383	0.2499	
		Two-step 2SLS	0.0366	0.3394	0.3412	0.1959	0.3967	0.3130	
	$N = 500$	SLS	0.0436	0.2480	0.2516	0.1569	0.3209	0.2328	0.1796
		Two-step Probit	0.0286	0.1887	0.1908	0.1274	0.2535	0.1745	
		Two-step 2SLS	0.0365	0.2267	0.2295	0.1440	0.2879	0.2164	
	$N = 1,000$	SLS	0.0104	0.1506	0.1509	0.0995	0.2017	0.1438	0.1335
		Two-step Probit	0.0057	0.1199	0.1200	0.0796	0.1619	0.1194	
		Two-step 2SLS	0.0060	0.1494	0.1495	0.0990	0.1986	0.1464	
$T = 5$	$N = 250$	SLS	0.0141	0.2266	0.2269	0.1420	0.2937	0.2307	0.1933
		Two-step Probit	0.0127	0.1699	0.1703	0.1163	0.2324	0.1688	
		Two-step 2SLS	0.0175	0.2098	0.2105	0.1354	0.2813	0.2082	
	$N = 500$	SLS	0.0141	0.1622	0.1627	0.1077	0.2145	0.1480	0.1485
		Two-step Probit	0.0119	0.1242	0.1247	0.0748	0.1500	0.1191	
		Two-step 2SLS	0.0177	0.1478	0.1488	0.0988	0.2009	0.1459	
	$N = 1,000$	SLS	0.0044	0.0993	0.0994	0.0637	0.1277	0.0991	0.1106
		Two-step Probit	0.0038	0.0821	0.0822	0.0514	0.1039	0.0829	
		Two-step 2SLS	0.0026	0.1000	0.1000	0.0655	0.1305	0.1007	
$T = 10$	$N = 250$	SLS	0.0154	0.1448	0.1455	0.0908	0.1824	0.1419	0.1902
		Two-step Probit	0.0116	0.1145	0.1150	0.0738	0.1453	0.1105	
		Two-step 2SLS	0.0153	0.1334	0.1342	0.0881	0.1762	0.1346	
	$N = 500$	SLS	0.0064	0.0986	0.0988	0.0673	0.1368	0.0952	0.1448
		Two-step Probit	0.0045	0.0795	0.0796	0.0526	0.1048	0.0776	
		Two-step 2SLS	0.0050	0.0960	0.0961	0.0650	0.1303	0.0940	
	$N = 1,000$	SLS	-0.0029	0.0670	0.0671	0.0446	0.0875	0.0649	0.1086
		Two-step Probit	-0.0030	0.0533	0.0534	0.0343	0.0683	0.0544	
		Two-step 2SLS	-0.0020	0.0653	0.0653	0.0427	0.0840	0.0658	

Notes: The DGP is given by (5.1), where the errors follow Gaussian distributions under Design I. The true value of β_2/β_1 is 1. The bootstrap standard errors of the Probit and 2SLS estimators are computed using $B = 200$ replications per sample and averaged across $R = 1,000$ Monte Carlo samples. The number of bootstrap replications for the SLS estimator is $B = 100$.

Table 2: Small sample properties of the estimators of β_2/β_1 for Design II (Non-Gaussian errors)

			Bias	SD	RMSE	MAD	IQR	Bootstrap SE	Average SE
$T=3$	$N=250$	SLS	0.1583	0.7957	0.8109	0.2970	0.6329	0.9491	0.3487
		Two-step Probit	0.1060	0.5177	0.5282	0.2853	0.5880	0.4994	
		Two-step 2SLS	0.1086	0.5464	0.5568	0.2796	0.5840	0.5042	
	$N=500$	SLS	0.0334	0.3339	0.3354	0.1977	0.3962	0.3666	0.2279
		Two-step Probit	0.0410	0.3394	0.3417	0.2056	0.4124	0.3203	
		Two-step 2SLS	0.0405	0.3376	0.3398	0.1975	0.3989	0.3200	
	$N=1,000$	SLS	0.0100	0.2104	0.2106	0.1308	0.2642	0.2129	0.1646
		Two-step Probit	0.0193	0.2215	0.2222	0.1475	0.2929	0.2172	
		Two-step 2SLS	0.0166	0.2207	0.2212	0.1435	0.2881	0.2159	
$T=5$	$N=250$	SLS	0.0378	0.3322	0.3342	0.2017	0.4112	0.3945	0.2310
		Two-step Probit	0.0363	0.3191	0.3210	0.2159	0.4280	0.3119	
		Two-step 2SLS	0.0339	0.3192	0.3208	0.2037	0.4144	0.3114	
	$N=500$	SLS	0.0219	0.2173	0.2183	0.1471	0.2981	0.2289	0.1704
		Two-step Probit	0.0257	0.2240	0.2253	0.1436	0.2857	0.2166	
		Two-step 2SLS	0.0263	0.2209	0.2224	0.1401	0.2870	0.2164	
	$N=1,000$	SLS	0.0151	0.1477	0.1484	0.0951	0.1946	0.1470	0.1236
		Two-step Probit	0.0164	0.1538	0.1546	0.0970	0.1949	0.1502	
		Two-step 2SLS	0.0159	0.1523	0.1530	0.0975	0.1964	0.1495	
$T=10$	$N=250$	SLS	0.0179	0.2088	0.2094	0.1275	0.2636	0.2305	0.1744
		Two-step Probit	0.0193	0.2032	0.2040	0.1329	0.2653	0.2004	
		Two-step 2SLS	0.0196	0.2037	0.2046	0.1333	0.2694	0.1998	
	$N=500$	SLS	0.0088	0.1359	0.1361	0.0922	0.1851	0.1448	0.1293
		Two-step Probit	0.0062	0.1405	0.1405	0.0918	0.1851	0.1397	
		Two-step 2SLS	0.0067	0.1388	0.1389	0.0903	0.1810	0.1392	
	$N=1,000$	SLS	0.0074	0.0949	0.0952	0.0648	0.1307	0.0946	0.0973
		Two-step Probit	0.0083	0.0978	0.0981	0.0673	0.1359	0.0986	
		Two-step 2SLS	0.0087	0.0984	0.0987	0.0675	0.1335	0.0981	

Notes: The DGP is given by (5.1), where the errors follow chi-square distributions under Design II. The true value of β_2/β_1 is 1. The bootstrap standard errors of the Probit and 2SLS estimators are computed using $B = 200$ replications per sample and averaged across $R = 1,000$ Monte Carlo samples. The number of bootstrap replications for the SLS estimator is $B = 100$.

Table 3: Descriptive statistics

Variable	Mean	Std. Dev.	Min	Max
Labor force participation	.6999	.4584	0	1
Nonwife income in $1,000	32.9073	25.0175	-14.9979	397.55
Age	37.4061	9.4811	20	57
Educ	12.7008	2.2018	2	17
Number of kids	.5076	.7382	0	3
Husband's age	40.0189	10.3895	20	76
Husband's education	12.9811	2.7409	3	17

Notes: Sample size is $N = 742$ and $T = 3$. Labor force participation is an indicator equal to one if the woman worked (hours>0) in a given year. Age and education are measured in years.

Table 4: Estimation results of labor force participation for married women

	Reduced form for *nwinc*	Probit-exog.	IVProbit	SLS-exog.	SLS
Independent Variable	(1)	(2)	(3)	(4)	(5)
Nonwife income		-0.011***	-0.021***		
		(0.003)	(0.008)		
Age	0.351	0.086	0.088		
	(0.669)	(0.055)	(0.054)		
Education	1.120**	0.147***	0.154***		
	(0.480)	(0.027)	(0.029)		
Number of kids	-0.410	-0.205***	-0.203***		
	(0.926)	(.078)	(.077)		
Husband's age	0.068				
	(0.812)				
Husband's education	2.413***				
	(0.387)				
$\hat{\beta}_{nwinc}/\hat{\beta}_{age}$		-0.128	-0.240	-0.082	-0.239*
		(0.085)	(0.168)		
		[0.087]	[0.171]	[0.133]	[0.133]
$\hat{\beta}_{educ}/\hat{\beta}_{age}$		1.696	1.756	0.918	1.307*
		(1.100)	(1.114)		
		[1.075]	[1.101]	[0.635]	[0.721]
$\hat{\beta}_{kids}/\hat{\beta}_{age}$		-2.370	-2.318	-1.960***	-2.502**
		(1.807)	(1.742)		
		[1.771]	[1.719]	[0.655]	[1.188]
Number of observations			$N = 742$,	$T = 3$	

Notes: The main model and reduced form equation are specified in (6.1) and (6.2), respectively. All estimations include year indicators and time means of exogenous regressors. Time means of education variables were excluded due to no variation over time. Analytical clustered standard errors are in parentheses, and panel boostrapped standard errors based on 1,000 replications are in brackets. *, **, *** indicate significance at the 10, 5, and 1 percent levels, respectively.

Figure 1: Average structural function for the first year of our panel

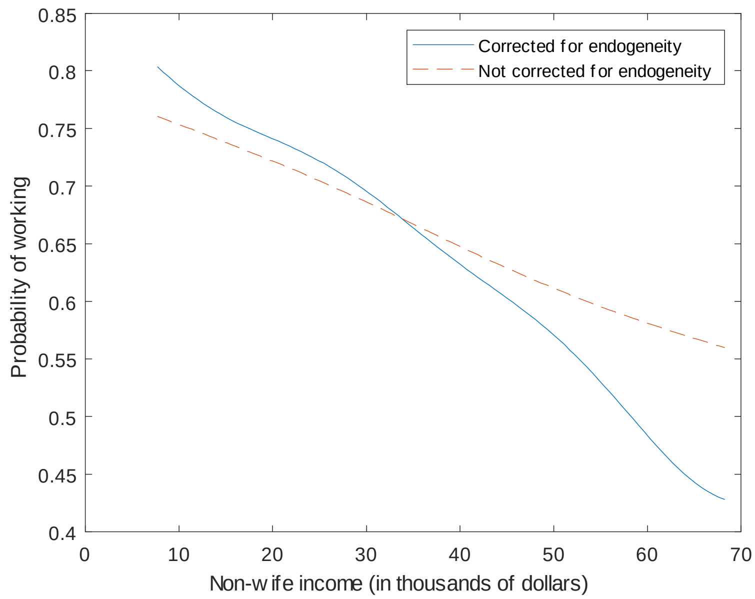


Figure 2: Average structural function for the second year of our panel

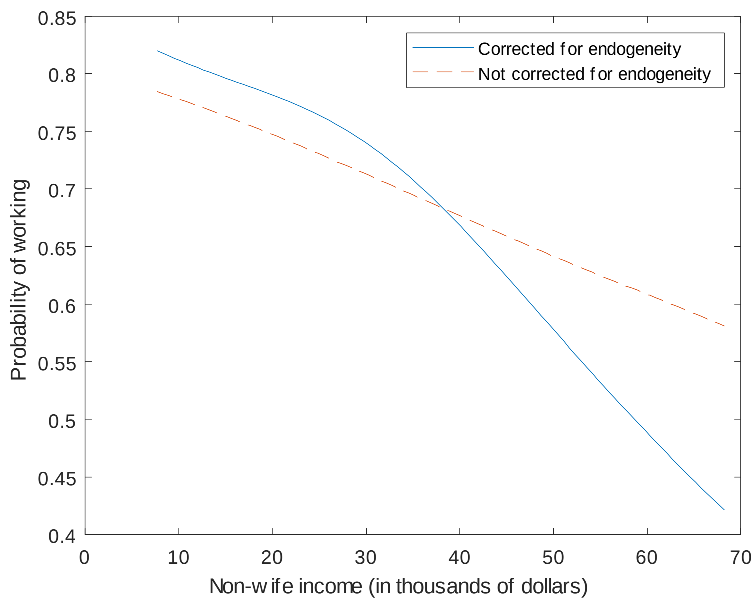


Figure 3: Average structural function for the third year of our panel

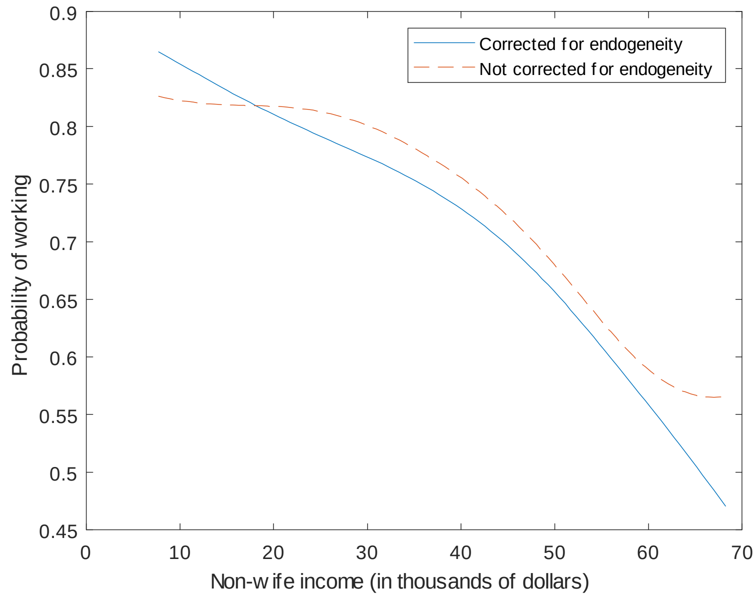
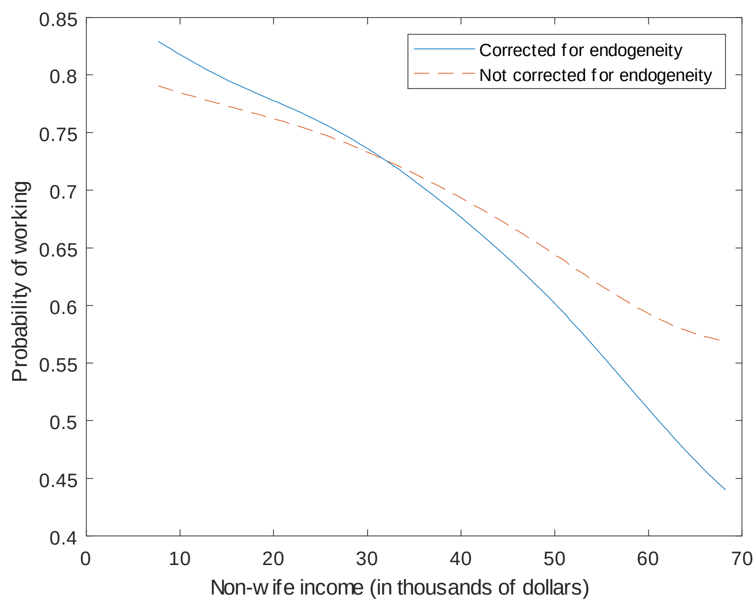


Figure 4: Average structural function averaged over all three yeas of our panel



References

[1] Blundell, R., and J. Powell, 2004, Endogeneity in semiparametric binary response models, *The Review of Economic Studies* 71, 655-679.

[2] Chamberlain, G., 1980, Analysis of Covariance with Qualitative Data, *The Review of Economic Studies* 47, 225-238.

[3] Chamberlain, G., 2010, Binary Response Models for Panel Data: Identification and Information, *Econometrica* 78, 159-168.

[4] Chen, X., O. Linton, and I. Van Keilegom, 2003, Estimation of semiparametric models when the criterion function is not smooth, *Econometrica* 71, 1591-1608.

[5] Hsiao, C., and Q, Zhou, 2018, Incidental parameters, initial conditions and sample size in statistical inference for dynamic panel data models, *Journal of Econometrics* 207, 114-128.

[6] Ichimura, H., 1993, Semiparametric least squares (SLS) and weighted SLS estimation of single-index models, *Journal of Econometrics* 58, 71-120.

[7] Klein, R., and R.H. Spady, 1993, An Efficient Semiparametric Estimator for Binary Response Models, *Econometrica* 61, 387-421.

[8] Manski, C., 1985, Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator, *Journal of Econometrics* 27, 313-333.

[9] Masry, E., 1996, Multivariate regression estimation local polynomial fitting for time series, *Stochastic Processes and their Applications* 65, 81-101.

[10] Mundlak, Y., 1978, On the Pooling of Time Series and Cross Section Data, *Econometrica* 46, 69-85.

[11] Newey, W., and D. McFadden, 1994, Large sample estimation and hypothesis testing, Chapter 36 in *Handbook of Econometrics* Volume 4, 2111-2245, Elsevier, Amsterdam, The Netherlands.

[12] Rivers, D. and Vuong, Q.H., 1988, Limited Information Estimators and Exogeneity Tests for Simultaneous Probit. *Journal of Econometrics* 39, 347-366.

[13] Rothe, C., 2009, Semiparametric estimation of binary response models with endogenous regressors, *Journal of Econometrics* 153, 51-64.

[14] Semykina, A., and J.M. Wooldridge, 2018, Binary response panel data models with sample selection and self selection, *Journal of Applied Econometrics* 33, 179-197.

[15] Smith, R.J. and Blundell, R.W., 1986, An Exogeneity Test for a Simultaneous Equation Tobit Model with an Application to Labor Supply. *Econometrica* 54, 679-685.

[16] Song, W., 2017, Least square estimation of semiparametric binary response model with endogeneity. Working Paper.

[17] Wooldridge J.M., 2015, Control Function Methods in Applied Econometrics. *Journal of Human Resources* 50(2), 420-445.

[18] van der Vaart, A.W., and J.A. Wellner, 1996, *Weak Convergence and Empirical Processes: With Applications to Statistics*, Springer.

A Appendix: Mathematical Derivations

This appendix contains mathematical derivations for the main results in the paper.

A.1 Proof for Theorem 3.1

Proof. The proof is similar to Manski (1985); here we only sketch the proof. By setups of the model, we have

$$E[y_{it1}|\mathbf{w}_{it}, \mathbf{v}_{it2}] = E[y_{it1}|\mathbf{w}_{it}'\boldsymbol{\pi}_0, \mathbf{v}_{it2}]$$

Suppose now there exists parameter $\breve{\boldsymbol{\pi}} \neq \boldsymbol{\pi}_0$, which satisfies

$$E[y_{it1}|\mathbf{w}_{it}, \mathbf{v}_{it2}] = E[y_{it1}|\mathbf{w}_{it}'\breve{\boldsymbol{\pi}}, \mathbf{v}_{it2}].$$

It then implies that we can recover information of $\mathbf{w}_{it}'\boldsymbol{\pi}_0$ from $\mathbf{w}_{it}'\breve{\boldsymbol{\pi}}$ and $\mathbf{v}_{it2}$. Therefore we can find a monotonic function $g(\cdot, \mathbf{v}_{it2})$ which is continuous for all $\mathbf{v}_{it2}$ such that

$$\mathbf{w}_{it}'\breve{\boldsymbol{\pi}} = g(\mathbf{w}_{it}'\boldsymbol{\pi}_0, \mathbf{v}_{it2}). \tag{A.1}$$

By Assumption IC(2) there is one continuous regressor $w_{it}^{(1)}$ for which we can normalize its coefficient as 1. Then we can differentiate (A.1) with respect to $w_{it}^{(1)}$ to obtain

$$1 = \frac{\partial g(\mathbf{w}_{it}'\boldsymbol{\pi}_0, \mathbf{v}_{it2})}{\partial w_{it}^{(1)}}, \tag{A.2}$$

which verifies $g(\cdot, \mathbf{v}_{it2})$ is an identity function. Thus $\mathbf{w}_{it}'\breve{\boldsymbol{\pi}} = \mathbf{w}_{it}'\boldsymbol{\pi}_0$. By Assumption IC(3), this equation holds if and only if $\breve{\boldsymbol{\pi}} = \boldsymbol{\pi}_0$. ■

A.2 Proof for Theorem 4.2

Before showing the proof of Theorem 4.2, we provide some preliminary lemmas.

Lemma A.1 *Under Assumptions CON, we have the following equations hold uniformly for* $\boldsymbol{\pi} \in \Pi$

$$\begin{aligned} \hat{p}_t(\mathbf{w}_{it}'\boldsymbol{\pi}, \hat{\mathbf{v}}_{it2}) &= \hat{p}_t(\mathbf{w}_{it}'\boldsymbol{\pi}, \mathbf{v}_{it2}) + o_p(N^{-\frac{1}{4}}), \\ \hat{q}_t(\mathbf{w}_{it}'\boldsymbol{\pi}, \hat{\mathbf{v}}_{it2}) &= \hat{q}_t(\mathbf{w}_{it}'\boldsymbol{\pi}, \mathbf{v}_{it2}) + o_p(N^{-\frac{1}{4}}). \end{aligned}$$

Proof. Here we just prove for the first equation since the rest can be shown similarly. We note that by applying Taylor expansion on $\hat{p}_t(\mathbf{w}_{it}'\boldsymbol{\pi}, \hat{\mathbf{v}}_{it2})$ around $\mathbf{v}_{it2}$, we obtain

$$\hat{p}_t(\mathbf{w}_{it}'\boldsymbol{\pi}, \hat{\mathbf{v}}_{it2}) = \hat{p}_t(\mathbf{w}_{it}'\boldsymbol{\pi}, \mathbf{v}_{it2}) + \frac{\partial \hat{p}_t(\mathbf{w}_{it}'\boldsymbol{\pi}, \mathbf{v}_{it2})}{\partial \mathbf{v}_{it2}}(\hat{\mathbf{v}}_{it2} - \mathbf{v}_{it2}) + o_p(N^{-1/4}),$$

where the last term is obtained using the assumption $\max_{i,t} \|\hat{\mathbf{v}}_{it2} - \mathbf{v}_{it}\| = o_p\left(N^{-1/4}\right)$. Now for the second term, we can decompose it further as follows, where for convenience we drop the argument for kernel functions κ_1 and κ_2,

$$\begin{aligned}
\frac{\partial \hat{p}_t(\mathbf{w}_{it}'\boldsymbol{\pi}, \mathbf{v}_{it2})}{\partial \mathbf{v}_{it2}}(\hat{\mathbf{v}}_{it2} - \mathbf{v}_{it2}) &= \frac{1}{Nh_2^{k_e}} \sum_{j \neq i} \kappa_1 \frac{\partial \kappa_2}{\partial \mathbf{v}_{it2}} y_{jt1} (\hat{\mathbf{v}}_{it2} - \mathbf{v}_{it2}) \\
&= \frac{1}{Nh_2^{k_e}} \sum_{j \neq i} \kappa_1 \frac{\partial \kappa_2}{\partial \mathbf{v}_{it2}} y_{jt1} \left(\frac{1}{N} \sum_{l=1}^{N} \sum_{s=1}^{T} g(\mathbf{D}_{it}, \mathbf{D}_{ls}) \psi_{ls} + o_p(N^{-1/2}) \right) \\
&= \frac{1}{N^2 h_2^{k_e}} \sum_{l=1}^{N} \sum_{s=1}^{T} \sum_{j \neq i} \kappa_1 \frac{\partial \kappa_2}{\partial \mathbf{v}_{it2}} y_{jt1} g(\mathbf{D}_{it}, \mathbf{D}_{ls}) \psi_{ls} + o_p(N^{-1/2}) \\
&= \frac{1}{Nh_2^{k_e}} \sum_{l=1}^{N} \sum_{s=1}^{T} E\left[E\left(\kappa_1 \frac{\partial \kappa_2}{\partial \mathbf{v}_{it2}} y_{jt1} | \mathbf{D}_{it} \right) g(\mathbf{D}_{it}, \mathbf{D}_{ls}) | \mathbf{D}_{ls} \right] \psi_{ls} + o_p(N^{-1/2}) \\
&= O_p(N^{-1/2} h_2^{-k_e}).
\end{aligned}$$

where the fourth equation is by projection for U-statistics. By Assumption CON(3) for $j = 1, \cdots, k_e$,

$$O_p(N^{-1/2} h_2^{-k_e}) = O_p(N^{k_e \delta_{2j} - 1/2}) < O_p(N^{-1/4}).$$

Therefore $\hat{p}_t(\mathbf{w}_{it}'\boldsymbol{\pi}, \hat{\mathbf{v}}_{it2}) = \hat{p}(\mathbf{w}_{it}'\boldsymbol{\pi}, \mathbf{v}_{it2}) + o_p(N^{-1/4})$. ■

Lemma A.2 *Under Assumptions CON, as N goes to infinity,*

$$\begin{aligned}
\sup_{\boldsymbol{\pi} \in \Pi} \left| \hat{F}_t(\mathbf{w}_{it}'\boldsymbol{\pi}, \hat{\mathbf{v}}_{it2}) - F_t(\mathbf{w}_{it}'\boldsymbol{\pi}, \mathbf{v}_{it2}) \right| &= o_p(N^{-1/4}), \\
\sup_{\boldsymbol{\pi} \in \Pi} \left| \frac{\partial \hat{F}_t(\mathbf{w}_{it}'\boldsymbol{\pi}, \hat{\mathbf{v}}_{it2})}{\partial \boldsymbol{\pi}} - \frac{\partial F_t(\mathbf{w}_{it}'\boldsymbol{\pi}, \mathbf{v}_{it2})}{\partial \boldsymbol{\pi}} \right| &= o_p(N^{-1/4}), \\
\sup_{\boldsymbol{\pi} \in \Pi} \left| \frac{\partial \hat{F}_t(\mathbf{w}_{it}'\boldsymbol{\pi}, \hat{\mathbf{v}}_{it2})}{\partial \hat{\mathbf{v}}_{it2}} - \frac{\partial F_t(\mathbf{w}_{it}'\boldsymbol{\pi}, \mathbf{v}_{it2})}{\partial \mathbf{v}_{it2}} \right| &= o_p(N^{-1/4}).
\end{aligned}$$

Proof. This follows from the standard kernel smoothing theory, here we just show the first equation since the proof is similar for the rest. From Masry (1996) and under Assumption

CON(3), we have

$$\left|\hat{F}_t(\mathbf{w}'_{it}\boldsymbol{\pi}, \mathbf{v}_{it2}) - F_t(\mathbf{w}'_{it}\boldsymbol{\pi}, \mathbf{v}_{it2})\right| \leq O_p(N^{-1/4}).$$

Then by Lemma A.1 and the triangular inequality, we obtain

$$\begin{aligned} & \sup_{\boldsymbol{\pi}\in\Pi}\left|\hat{F}_t(\mathbf{w}'_{it}\boldsymbol{\pi}, \hat{\mathbf{v}}_{it2}) - F_t(\mathbf{w}'_{it}\boldsymbol{\pi}, \mathbf{v}_{it2})\right| \\ \leq & \sup_{\boldsymbol{\pi}\in\Pi}\left|\hat{F}_t(\mathbf{w}'_{it}\boldsymbol{\pi}, \hat{\mathbf{v}}_{it2}) - \hat{F}_t(\mathbf{w}'_{it}\boldsymbol{\pi}, \mathbf{v}_{it2})\right| + \sup_{\boldsymbol{\pi}\in\Pi}\left|\hat{F}_t(\mathbf{w}'_{it}\boldsymbol{\pi}, \mathbf{v}_{it2}) - F_t(\mathbf{w}'_{it}\boldsymbol{\pi}, \mathbf{v}_{it2})\right| \\ = & o_p\left(N^{-1/4}\right). \end{aligned}$$

■

Now we are ready to prove Theorem 4.2 as follows.

Proof of Theorem 4.2. To obtain the consistency of $\hat{\boldsymbol{\pi}}_{SLS}$. we first show the uniform convergence of $\hat{S}_N(\boldsymbol{\pi})$ to $\tilde{S}_N(\boldsymbol{\pi})$. Namely, for any $\boldsymbol{\pi} \in \Pi$ we have

$$\begin{aligned} & \left|\hat{S}_N(\boldsymbol{\pi}) - \tilde{S}_N(\boldsymbol{\pi})\right| \\ = & \left|\frac{1}{N}\sum_{i=1}^N\sum_{t=1}^T 2y_{it1}\left[\hat{F}_t(\mathbf{w}'_{it}\boldsymbol{\pi}, \hat{\mathbf{v}}_{it2}) - F_t(\mathbf{w}'_{it}\boldsymbol{\pi}, \mathbf{v}_{it2})\right] + \left[\hat{F}_t(\mathbf{w}'_{it}\boldsymbol{\pi}, \hat{\mathbf{v}}_{it2})^2 - F_t(\mathbf{w}'_{it}\boldsymbol{\pi}, \mathbf{v}_{it2})^2\right]\right| \\ \leq & \frac{2}{N}\sum_{i=1}^N\sum_{t=1}^T |y_{it1}|\left|\hat{F}_t(\mathbf{w}'_{it}\boldsymbol{\pi}, \hat{\mathbf{v}}_{it2}) - F_t(\mathbf{w}'_{it}\boldsymbol{\pi}, \mathbf{v}_{it2})\right| + \frac{1}{N}\sum_{i=1}^N\sum_{t=1}^T\left|\hat{F}_t(\mathbf{w}'_{it}\boldsymbol{\pi}, \hat{\mathbf{v}}_{it2})^2 - F_t(\mathbf{w}'_{it}\boldsymbol{\pi}, \mathbf{v}_{it2})^2\right|. \end{aligned}$$

By Lemma A.2, we obtain

$$\sup_{\boldsymbol{\pi}\in\Pi}\left|\hat{S}_N(\boldsymbol{\pi}) - \tilde{S}_N(\boldsymbol{\pi})\right| = o_p(1). \tag{A.3}$$

Define $S(\boldsymbol{\pi})$ as the probability limit of $\tilde{S}_N(\boldsymbol{\pi})$. Then, by the uniform law of large numbers,

$$\sup_{\boldsymbol{\pi}\in\Pi}\left|\tilde{S}_N(\boldsymbol{\pi}) - S(\boldsymbol{\pi})\right| = o_p(1). \tag{A.4}$$

and by triangular inequality,

$$\sup_{\boldsymbol{\pi}\in\Pi}\left|\hat{S}_N(\boldsymbol{\pi}) - S(\boldsymbol{\pi})\right| = o_p(1). \tag{A.5}$$

On the other hand, based on the data generating process for y_{it1},

$$
\begin{aligned}
S(\boldsymbol{\pi}) &= \sum_{t=1}^{T}\left\{E\left[y_{it1} - F_t(\mathbf{w}_{it}'\boldsymbol{\pi}, \mathbf{v}_{it2})\right]^2\right\} \\
&= \sum_{t=1}^{T}\left\{E\left[y_{it1}^2 - 2y_{it1}F_t(\mathbf{w}_{it}'\boldsymbol{\pi}, \mathbf{v}_{it2}) + F_t(\mathbf{w}_{it}'\boldsymbol{\pi}, \mathbf{v}_{it2})^2\right]\right\} \\
&= \sum_{t=1}^{T}\left\{E\left[F_t(\mathbf{w}_{it}'\boldsymbol{\pi}_0, \mathbf{v}_{it2})^2 - 2F_t(\mathbf{w}_{it}'\boldsymbol{\pi}_0, \mathbf{v}_{it2})F_t(\mathbf{w}_{it}'\boldsymbol{\pi}, \mathbf{v}_{it2}) + F_t(\mathbf{w}_{it}'\boldsymbol{\pi}, \mathbf{v}_{it2})^2\right]\right\}
\end{aligned}
$$

By taking the first order derivative with respect to $\mathbf{w}_{it}'\boldsymbol{\pi}$, the first order condition becomes

$$F_t(\mathbf{w}_{it}'\boldsymbol{\pi}_0, \mathbf{v}_{it2}) = F_t(\mathbf{w}_{it}'\boldsymbol{\pi}, \mathbf{v}_{it2}),$$

which indicates $E\left(y_{it1}|\mathbf{w}_{it}, \mathbf{v}_{it2}\right) = E\left(y_{it1}|\mathbf{w}_{it}'\boldsymbol{\pi}, \mathbf{v}_{it2}\right)$. By the identification condition of Theorem 3.1, this holds only if $\boldsymbol{\pi} = \boldsymbol{\pi}_0$. Then, following Theorem 2.1 in Newey and McFadden (1994), we obtain $\hat{\boldsymbol{\pi}}_{SLS} = \boldsymbol{\pi}_0 + o_p(1)$ as required. ■

A.3 Proof for Theorem 4.3

The proof of Theorem 4.3 is accomplished by showing that the six conditions in Theorem 2 of Chen et. al. (2003) are satisfied. We prove each condition holds by the following six lemmas.

Lemma A.3 *Under Assumptions CI, CON and NR,*

$$\left\|M_N(\hat{\boldsymbol{\pi}}_{SLS}, \hat{h})\right\| = \inf_{\boldsymbol{\pi}\in\Pi}\left\|M_N(\boldsymbol{\pi}, \hat{h})\right\| + o_p\left(N^{-1/2}\right).$$

Proof. This is obvious since with $\hat{\boldsymbol{\pi}}$ the moment function $M_N(\boldsymbol{\pi}, \hat{h})$ achieves the value of zero. ■

Lemma A.4 *Under Assumptions CI, CON and NR,* $\Gamma_1(\boldsymbol{\pi}, h_0)$ *exists for* $\boldsymbol{\pi} \in \Pi$ *and is continuous at* $\boldsymbol{\pi}_0$*.*

Proof. The first part is obvious by following the definition in (4.3). The second part can be verified by showing

$$
\begin{aligned}
\Gamma_1(\boldsymbol{\pi}_0,h_0) &= E\left(\sum_{t=1}^{T}\left[(y_{it1}-F_t(\mathbf{w}_{it}'\boldsymbol{\pi}_0,\mathbf{v}_{it2}))\cdot\frac{\partial^2 F_t(\mathbf{w}_{it}'\boldsymbol{\pi}_0,\mathbf{v}_{it2})}{\partial\boldsymbol{\pi}_0\partial\boldsymbol{\pi}_0'}-\frac{\partial F_t(\mathbf{w}_{it}'\boldsymbol{\pi}_0,\mathbf{v}_{it2})}{\partial\boldsymbol{\pi}_0}\frac{\partial F_t(\mathbf{w}_{it}'\boldsymbol{\pi}_0,\mathbf{v}_{it2})}{\partial\boldsymbol{\pi}_0'}\right]\right)\\
&= -E\left(\sum_{t=1}^{T}\frac{\partial F_t(\mathbf{w}_{it}'\boldsymbol{\pi}_0,\mathbf{v}_{it2})}{\partial\boldsymbol{\pi}_0}\frac{\partial F_t(\mathbf{w}_{it}'\boldsymbol{\pi}_0,\mathbf{v}_{it2})}{\partial\boldsymbol{\pi}_0'}\right)\\
&= -\Omega,
\end{aligned}
$$

which together with assumption CON(4) indicates the continuity of $\Gamma_1(\boldsymbol{\pi},h_0)$ at $\boldsymbol{\pi}=\boldsymbol{\pi}_0$. ■

Lemma A.5 *Under Assumptions CI, CON and NR, the pathwise derivative* $\Gamma_2(\boldsymbol{\pi},h_0)[\bar{h}-h_0]$ *exists in all directions* $[\bar{h}-h_0]\in\mathcal{H}$, *and for all* $(\boldsymbol{\pi},\bar{h})\in\Pi\times\mathcal{H}$ *with a positive sequence* $\delta_n=o(1)$, *we have:*

(1) $\left\|M(\boldsymbol{\pi},\bar{h})-M(\boldsymbol{\pi},h_0)-\Gamma_2(\boldsymbol{\pi},h_0)[\bar{h}-h_0]\right\|\le c||\bar{h}-h_0||_{\mathcal{H}}^2$ *for a constant* $c\ge 0$;

(2) $\left\|\Gamma_2(\boldsymbol{\pi},h_0)[\bar{h}-h_0]-\Gamma_2(\boldsymbol{\pi}_0,h_0)[\bar{h}-h_0]\right\|\le o(1)\delta_n$.

Proof. For $\bar{h}$ that is close to h_0, the pathwise derivative can be calculated as following

$$
\begin{aligned}
&\Gamma_2(\boldsymbol{\pi},h_0)[\bar{h}-h_0]\\
=\;& E\sum_{t=1}^{T}\Bigg[\frac{-\partial F_t(\mathbf{w}_{it}'\boldsymbol{\pi},\mathbf{v}_{it2})}{\partial\boldsymbol{\pi}}\frac{\partial F_t(\mathbf{w}_{it}'\boldsymbol{\pi},\mathbf{v}_{it2})}{\partial\mathbf{v}_{it2}'}(\bar{\mathbf{v}}_{it2}-\mathbf{v}_{it2})\\
&-\frac{\partial F_t(\mathbf{w}_{it}'\boldsymbol{\pi},\mathbf{v}_{it2})}{\partial\boldsymbol{\pi}}(\bar{F}_t(\mathbf{w}_{it}'\boldsymbol{\pi},\mathbf{v}_{it2})-F_t(\mathbf{w}_{it}'\boldsymbol{\pi},\mathbf{v}_{it2}))\\
&-\left(\left[y_{it1}-F_t(\mathbf{w}_{it}'\boldsymbol{\pi},\mathbf{v}_{it2})\right]\frac{\partial^2 F_t(\mathbf{w}_{it}'\boldsymbol{\pi},\mathbf{v}_{it2})}{\partial\boldsymbol{\pi}\partial\mathbf{v}_{it2}'}(\bar{\mathbf{v}}_{it2}-\mathbf{v}_{it2})\right)\Bigg].
\end{aligned}
$$

And by the law of iterated expectations, we can show

$$
\begin{aligned}
&\Gamma_2(\boldsymbol{\pi}_0,h_0)[\bar{h}-h_0]\\
=\;& E\sum_{t=1}^{T}\Bigg[\frac{-\partial F_t(\mathbf{w}_{it}'\boldsymbol{\pi}_0,\mathbf{v}_{it2})}{\partial\boldsymbol{\pi}_0}\frac{\partial F_t(\mathbf{w}_{it}'\boldsymbol{\pi}_0,\mathbf{v}_{it2})}{\partial\mathbf{v}_{it2}'}(\bar{\mathbf{v}}_{it2}-\mathbf{v}_{it2})\\
&-\frac{\partial F_t(\mathbf{w}_{it}'\boldsymbol{\pi}_0,\mathbf{v}_{it2})}{\partial\boldsymbol{\pi}_0}(\bar{F}_t(\mathbf{w}_{it}'\boldsymbol{\pi}_0,\mathbf{v}_{it2})-F_t(\mathbf{w}_{it}'\boldsymbol{\pi}_0,\mathbf{v}_{it2}))\Bigg].
\end{aligned}
$$

Then by Assumption CON (4), which requires that function F_t is Lipchitz continuous for high orders, the two equalities can be verified. ■

Lemma A.6 *Under Assumptions CI, CON and NR,* $\hat{h}\in\mathcal{H}$ *with probability tending to one, and* $||\hat{h}-h_0||_{\mathcal{H}}=o_p(N^{-1/4})$.

Proof. We note that $\hat{\mathbf{v}}_{it2} - \mathbf{v}_{it2} = o_p(N^{-1/4})$ from Assumption CON(5). As a result, together with Lemma A.2, we obtain $\hat{F}_t(\cdot, \hat{\mathbf{v}}_{it2}) - F_t(\cdot, \mathbf{v}_{it2}) = o_p(N^{-1/4})$ for each t, and $||\hat{h} - h_0||_{\mathcal{H}} = o_p(N^{-1/4})$ as required. ∎

Lemma A.7 *Under Assumptions CI, CON and NR, for all sequences of positive numbers $\{\delta_N\}$ with $\delta_N = o(1)$,*

$$\sup_{||\boldsymbol{\pi}-\boldsymbol{\pi}_0||\leq\delta_N, ||h-h_0||_{\mathcal{H}}\leq\delta_N} \sqrt{N}\, \|M_N(\boldsymbol{\pi}, h) - M(\boldsymbol{\pi}, h) - M_N(\boldsymbol{\pi}_0, h_0)\| = o_p(1).$$

Proof. This result can be verified by showing that Donsker Theorem holds for $M_N(\boldsymbol{\pi}, h) - M(\boldsymbol{\pi}, h)$. Define

$$\begin{aligned} q_{it}(\boldsymbol{\pi}, F_t, \mathbf{v}_{it2}) &= (y_{it1} - F_t(\mathbf{w}_{it}'\boldsymbol{\pi}, \mathbf{v}_{it2})) \cdot \frac{F_t(\mathbf{w}_{it}'\boldsymbol{\pi}, \mathbf{v}_{it2})}{\partial \boldsymbol{\pi}}, \\ m_i(\boldsymbol{\pi}, h) &= \sum_{t=1}^{T} q_{it}(\boldsymbol{\pi}, F_t, \mathbf{v}_{it2}), \end{aligned}$$

so that $M_N(\boldsymbol{\pi}, h) = 1/N \sum_{i=1}^{N} m_i(\boldsymbol{\pi}, h)$. Following Chen et. al. (2003), firstly let's show $L_P(2)$-continuity of $M_N(\boldsymbol{\pi}, h)$ with respect to $\boldsymbol{\pi}$ and h, that is

$$E\left(\sup_{||\boldsymbol{\pi}^*-\boldsymbol{\pi}||\leq\delta, ||h^*-h||_{\mathcal{H}}\leq\delta} \|m_i(\boldsymbol{\pi}^*, h^*) - m_i(\boldsymbol{\pi}, h)\|^2\right) \leq C\delta^2, \tag{A.6}$$

where C is a finite positive constant, $\delta = o(1)$ is a small positive number.

Then equation (A.6) can be written as

$$\begin{aligned} & E\left(\sup_{||\boldsymbol{\pi}^*-\boldsymbol{\pi}||\leq\delta, ||h^*-h||_{\mathcal{H}}\leq\delta} \left\|\sum_{t=1}^{T} q_{it}(\boldsymbol{\pi}^*, F_t^*, \mathbf{v}_{it2}^*) - q_{it}(\boldsymbol{\pi}, F_t, \mathbf{v}_{it2})\right\|^2\right) \\ \leq & \sum_{t=1}^{T} E\left(\sup_{||\boldsymbol{\pi}^*-\boldsymbol{\pi}||\leq\delta, ||h^*-h||_{\mathcal{H}}\leq\delta} \|q_{it}(\boldsymbol{\pi}^*, F_t^*, \mathbf{v}_{it2}^*) - q_{it}(\boldsymbol{\pi}, F_t, \mathbf{v}_{it2})\|^2\right), \end{aligned}$$

Moreover,

$$\begin{aligned} \|q_{it}(\boldsymbol{\pi}^*, F_t^*, \mathbf{v}_{it2}^*) - q_{it}(\boldsymbol{\pi}, F_t, \mathbf{v}_{it2})\|^2 \leq\ & \|q_{it}(\boldsymbol{\pi}^*, F_t^*, \mathbf{v}_{it2}^*) - q_{it}(\boldsymbol{\pi}, F_t^*, \mathbf{v}_{it2}^*)\|^2 + \|q_{it}(\boldsymbol{\pi}, F_t^*, \mathbf{v}_{it2}^*) \\ & - q_{it}(\boldsymbol{\pi}, F_t, \mathbf{v}_{it2}^*)\|^2 + \|q_{it}(\boldsymbol{\pi}, F_t, \mathbf{v}_{it2}^*) - q_{it}(\boldsymbol{\pi}, F_t, \mathbf{v}_{it2})\|^2. \end{aligned}$$

For the first term, by mean value theorem there exists some $\tilde{\boldsymbol{\pi}}$ between $\boldsymbol{\pi}^*$ and $\boldsymbol{\pi}$ such that

$$\begin{aligned}\sup_{||\boldsymbol{\pi}^*-\boldsymbol{\pi}||\leq\delta}\|q_{it}(\boldsymbol{\pi}^*, F_t^*, \mathbf{v}_{it2}^*) - q_{it}(\boldsymbol{\pi}, F_t^*, \mathbf{v}_{it2}^*)\|^2 &= \sup_{||\boldsymbol{\pi}^*-\boldsymbol{\pi}||\leq\delta}\left\|\frac{\partial q_{it}(\tilde{\boldsymbol{\pi}}, F_t^*, \mathbf{v}_{it2}^*)}{\partial\boldsymbol{\pi}}(\boldsymbol{\pi}^*-\boldsymbol{\pi})\right\|^2 \\ &\leq C\delta^2\left\|\frac{\partial q_{it}(\tilde{\boldsymbol{\pi}}, F_t^*, \mathbf{v}_{it2}^*)}{\partial\tilde{\boldsymbol{\pi}}}\right\|^2.\end{aligned}$$

Furthermore, the right-hand side can be decomposed as

$$\left\|\frac{\partial q_{it}(\tilde{\boldsymbol{\pi}}, F_t^*, \mathbf{v}_{it2}^*)}{\partial\tilde{\boldsymbol{\pi}}}\right\| = \left\|-\frac{\partial F_t^*}{\partial\tilde{\boldsymbol{\pi}}}\frac{\partial F_t^*}{\partial\tilde{\boldsymbol{\pi}}'} + (y_{it1} - F_t^*)\frac{\partial^2 F_t^*}{\partial\boldsymbol{\pi}\partial\boldsymbol{\pi}'}\right\|, \tag{A.7}$$

where we drop the arguments in F_t^* on the right-hand side for convenience; the arguments are the same as those in q_{it} on the left-hand side. Then, (A.7) is bounded by the continuity and boundedness of function F_t and its derivatives, as stated in Assumption CON (4). Therefore,

$$E\left(\sup_{||\boldsymbol{\pi}^*-\boldsymbol{\pi}||\leq\delta}\|q_{it}(\boldsymbol{\pi}^*, F_t^*, \mathbf{v}_{it2}^*) - q_{it}(\boldsymbol{\pi}, F_t^*, \mathbf{v}_{it2}^*)\|^2\right) \leq C_1\delta^2,$$

where C_1 is a positive constant. For the second term, we can replace F_t^* and get

$$\begin{aligned}&\sup_{||\boldsymbol{\pi}^*-\boldsymbol{\pi}||\leq\delta, ||h^*-h||_{\mathcal{H}}\leq\delta}\|q_{it}(\boldsymbol{\pi}, F_t^*, \mathbf{v}_{it2}^*) - q_{it}(\boldsymbol{\pi}, F_t, \mathbf{v}_{it2}^*)\|^2 \\ \leq\ &\sup_{||\boldsymbol{\pi}^*-\boldsymbol{\pi}||\leq\delta, ||h^*-h||_{\mathcal{H}}\leq\delta}\left\|(y_{it1} - F_t + \delta/2)\cdot\frac{\partial F_t}{\partial\boldsymbol{\pi}} - (y_{it1} - F_t)\cdot\frac{\partial F_t}{\partial\boldsymbol{\pi}}\right\|^2 \\ =\ &\sup_{||\boldsymbol{\pi}^*-\boldsymbol{\pi}||\leq\delta, ||h^*-h||_{\mathcal{H}}\leq\delta}\frac{\delta^2}{4}\left\|\frac{\partial F_t}{\partial\boldsymbol{\pi}}\right\|^2,\end{aligned}$$

which is bounded almost everywhere. Therefore,

$$E\left(\sup_{||\boldsymbol{\pi}^*-\boldsymbol{\pi}||\leq\delta, ||h^*-h||_{\mathcal{H}}\leq\delta}\|q_{it}(\boldsymbol{\pi}, F_t^*, \mathbf{v}_{it2}^*) - q_{it}(\boldsymbol{\pi}, F_t, \mathbf{v}_{it2}^*)\|^2\right) \leq C_2\delta^2$$

for some $C_2 > 0$. Finally, the third term can be rewritten as

$$\|q_{it}(\boldsymbol{\pi}, F_t, \mathbf{v}_{it2}^*) - q_{it}(\boldsymbol{\pi}, F_t, \mathbf{v}_{it2})\|^2 = \left\|\frac{\partial q_{it}(\boldsymbol{\pi}, F_t, \mathbf{v}_{it2})}{\partial\mathbf{v}_{it2}}(\mathbf{v}_{it2}^* - \mathbf{v}_{it2})\right\|^2,$$

by the mean value theorem. Since

$$\frac{\partial q_{it}}{\partial \mathbf{v}_{it2}} = -\frac{\partial F_t}{\partial \boldsymbol{\pi}}\frac{\partial F_t}{\partial \mathbf{v}_{it2}} + (y_{it1} - F_t)\frac{\partial^2 F_t}{\partial \boldsymbol{\pi}\partial \mathbf{v}_{it2}},$$

is bounded almost everywhere,

$$\sup_{||h^* - h||_{\mathcal{H}} \leq \delta} \|q_{it}(\boldsymbol{\pi}, F_t, \mathbf{v}^*_{it2}) - q_{it}(\boldsymbol{\pi}, F_t, \mathbf{v}_{it2})\|^2 \leq C_3\delta^2,$$

for some $C_3 > 0$. Therefore (A.6) is proved.

Under Assumptions CON(4) and NR(1), it can be verified that Donsker Theorem holds by Corollary 2.7.4 in Van der Vaart and Wellner (1996). Therefore, the stated stochastic equicontinuity then immediately follows. ■

Lemma A.8 *Under Assumptions CI, CON and NR, as $N \to \infty$, T fixed, we have*

$$\sqrt{N}\left(M_N(\boldsymbol{\pi}_0, h_0) + \Gamma_2(\boldsymbol{\pi}_0, h_0)\left[\hat{h} - h_0\right]\right) \to_d \mathcal{N}\left(0, \mathbf{V}\right),$$

where $\mathbf{V} = E\left(\boldsymbol{\phi}_i\boldsymbol{\phi}_i'\right)$ with $\boldsymbol{\phi}_i$ provided in (A.8).

Proof. Since it is already shown that $\hat{F}_t(\cdot, \hat{\mathbf{v}}_{it2})$ converges uniformly to $F_t(\cdot, \mathbf{v}_{it2})$, we can simplify $\Gamma_2(\boldsymbol{\pi}, h)[\hat{h} - h]$ as

$$\Gamma_2(\boldsymbol{\pi}_0, h_0)[\hat{h} - h_0] = E\left[\sum_{t=1}^{T} \frac{-\partial F_t(\mathbf{w}'_{it}\boldsymbol{\pi}_0, \mathbf{v}_{it2})}{\partial \boldsymbol{\pi}_0}\frac{\partial F_t(\mathbf{w}'_{it}\boldsymbol{\pi}_0, \mathbf{v}_{it2})}{\partial \mathbf{v}'_{it2}}(\hat{\mathbf{v}}_{it2} - \mathbf{v}_{it2})\right].$$

Assuming that reduced form parameters are estimated by pooled OLS, we can rewrite it using a specific influence function and obtain

$$\begin{aligned}
&\Gamma_2(\boldsymbol{\pi}_0, h_0)[\hat{h} - h_0] \\
= \;& E\left[\sum_{t=1}^{T} \frac{-\partial F_t(\mathbf{w}'_{it}\boldsymbol{\pi}_0, \mathbf{v}_{it2})}{\partial \boldsymbol{\pi}_0}\frac{\partial F_t(\mathbf{w}'_{it}\boldsymbol{\pi}_0, \mathbf{v}_{it2})}{\partial \mathbf{v}'_{it2}}\frac{1}{N}\sum_{j=1}^{N}\sum_{s=1}^{T} g(\mathbf{D}_{it}, \mathbf{D}_{js})\psi_{js}\right] + o_p(N^{-1/2}) \\
= \;& \frac{1}{N}\sum_{j=1}^{N} E\left[\sum_{s=1}^{T}\sum_{t=1}^{T} \frac{-\partial F_t(\mathbf{w}'_{it}\boldsymbol{\pi}_0, \mathbf{v}_{it2})}{\partial \boldsymbol{\pi}_0}\frac{\partial F_t(\mathbf{w}'_{it}\boldsymbol{\pi}_0, \mathbf{v}_{it2})}{\partial \mathbf{v}'_{it2}} g(\mathbf{D}_{it}, \mathbf{D}_{js})|\mathbf{D}_{js}\right]\psi_{js} + o_p(N^{-1/2}) \\
= \;& \frac{1}{N}\sum_{j=1}^{N} \Psi_j(\boldsymbol{\pi}_0, h_0) + o_p(N^{-1/2}).
\end{aligned}$$

By change of variables, now we can show

$$\begin{aligned}
& \sqrt{N}\{M_N(\boldsymbol{\pi}_0, h_0) + \Gamma_2(\boldsymbol{\pi}_0, h_0)[\hat{h} - h_0]\} \\
= & \sqrt{N}\left[\frac{1}{N}\sum_{i=1}^{N} m_i(\boldsymbol{\pi}_0, h_0) + \frac{1}{N}\sum_{i=1}^{N}\Psi_i(\boldsymbol{\pi}_0, h_0) + o_p(N^{-1/2})\right] \\
= & \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{\phi}_i(\boldsymbol{\pi}_0, h_0) + o_p(1). \qquad (A.8)
\end{aligned}$$

Following Chen et al. (2003) and Rothe (2009), and applying standard central limit theorem, we obtain

$$\sqrt{N}\{M_N(\boldsymbol{\pi}_0, h_0) + \Gamma_2(\boldsymbol{\pi}_0, h_0)[\hat{h} - h_0]\} \to_d N(0, \mathbf{V}), \qquad (A.9)$$

where $\mathbf{V} = E(\boldsymbol{\phi}_i(\boldsymbol{\pi}_0, h_0)\boldsymbol{\phi}_i(\boldsymbol{\pi}_0, h_0)')$. ■

Now we prove Theorem 4.3 as follows.

Proof of Theorem 4.3. Based on Chen et. al. (2003), by applying stochastic equicontinuity in Lemma A.8 and Taylor expansion of moment function $M_N(\hat{\boldsymbol{\pi}}_{SLS}, \hat{h})$ around $(\boldsymbol{\pi}_0, h_0)$,

$$\sqrt{N}(\hat{\boldsymbol{\pi}}_{SLS} - \boldsymbol{\pi}_0) = \Gamma_1^{-1}(\boldsymbol{\pi}_0, h_0) \cdot \sqrt{N}(M_N(\boldsymbol{\pi}_0, h_0) + \Gamma_2(\boldsymbol{\pi}_0, h_0)[\hat{h} - h_0]). \qquad (A.10)$$

Then by Lemmas A.4 and A.8,

$$\sqrt{N}(\hat{\boldsymbol{\pi}}_{SLS} - \boldsymbol{\pi}_0) \to_d N(0, \Omega^{-1}\mathbf{V}\Omega^{-1}), \qquad (A.11)$$

as required. ■