UNIVERSITY OF CAMBRIDGE

Faculty of Economics

Cambridge Working Papers in Economics

Cambridge Working Papers in Economics: 1906

SEMIPARAMETRIC NONLINEAR PANEL DATA MODELS WITH MEASUREMENT ERROR

Oliver Linton　　　Ji-Liang Shiu

18 January 2019

This paper develops identification and estimation of the parameters of a nonlinear semi-parametric panel data model with mismeasured variables as well as the corresponding average partial effects using only three periods of data. The past observables are used as instruments to control the measurement error problem, and the time averages of perfectly observed variables are used to restrict the unobserved individual-specific effect by a correlated random effects specification. The proposed approach relies on the Fourier transforms of several conditional expectations of observable variables. We estimate the model via the semi-parametric sieve minimum distance estimator. The finite-sample properties of the estimator are investigated through Monte Carlo simulations. We use our method to estimate the effect of the wage rate on labor supply using PSID data.

Semiparametric Nonlinear Panel Data Models with Measurement Error

Oliver Linton* Ji-Liang Shiu†

January 12, 2019

Abstract

This paper develops identification and estimation of the parameters of a nonlinear semi-parametric panel data model with mismeasured variables as well as the corresponding average partial effects using only three periods of data. The past observables are used as instruments to control the measurement error problem, and the time averages of perfectly observed variables are used to restrict the unobserved individual-specific effect by a correlated random effects specification. The proposed approach relies on the Fourier transforms of several conditional expectations of observable variables. We estimate the model via the semi-parametric sieve minimum distance estimator. The finite-sample properties of the estimator are investigated through Monte Carlo simulations. We use our method to estimate the effect of the wage rate on labor supply using PSID data.

Keywords: Correlated random effects, Measurement error, Nonlinear panel data models, Semi-parametric identification

*Department of Economics, University of Cambridge, Email: obl20@cam.ac.uk.
†Institute for Economic and Social Research, Jinan University. Email: jishiu.econ@gmail.com.

1. Introduction

The availability of panel data allows economists to control for unobservable individual-specific characteristics that may be correlated with explanatory variables. Substantial progress has been made on how to handle linear or nonlinear models ignoring the potential presence of measurement error. However, many economic quantities such as work hours, earnings, fringe benefits, employment, and health in surveys are frequently measured with errors, especially when longitudinal information is collected through one-time retrospective surveys.[1] This concern has been heightened by the increased use of longitudinal data sets; mismeasurement of the panel data may lead to false results or obscure the true economic relationships. The estimation problems caused by the mismeasurement of economic data may be greatly exacerbated when one tries to control for the consequences of unobserved individual effects by using standard fixed effects or first-differenced estimators.

We consider the following semi-parametric nonlinear panel data model with unknown finite-dimensional parameter β_0

$$Y_{it} = m\left(W_{it}, X^*_{it}, C_i; \beta_0\right) + U_{it}, \quad i = 1, \ldots, n, \quad t = 1, \ldots, T. \tag{1}$$

In this model, Y_{it} is an observed scalar dependent variable, W_{it} are perfectly observed explanatory variables, X^*_{it} is a latent continuously distributed mismeasured variable, C_i is an unobserved individual-specific effect, and U_{it} is an unobserved random variable. The function m may not be separable with regard to W_{it}, X^*_{it}, and C_i, but it belongs to a known, finite-dimensional, parametric family. We focus on the case where the data consists of a large number of individuals observed through a small (fixed) number of time periods. The variable X_{it} is a proxy or measure of the unobserved true regressor X^*_{it}.

The model described in Eq. (1) has two aspects that are new in the literature of

[1]The problems of the measurement error have raised great concern in a number of practical applications. Studies in Bollinger (1998), Bound, Brown, Duncan, and Rodgers (1994), and Bound, Brown, and Mathiowetz (2001) provide evidences of the measurement errors in economics data sets.

panel data models with measurement errors. First, the unobserved heterogeneity enters the structural regression function nonseparably and without imposing a linear index structure. Second, the potentially nonlinear regression function also contains a mismeasured variable (nonseparably) along with other explanatory variables. This proposed regression model is consistent with a structural function derived from a dynamic utility maximization problem with flexible preferences. For example, models of this form can arise in the study of life cycle labor supply with individual preference. See e.g., Koebel, Laisney, Pohlmeier, and Staat (2008).[2]

Linear panel data models with measurement error problems have been widely studied in the literature including: Griliches and Hausman (1986), Wansbeek and Koning (1991), Biørn (1992), and Wansbeek (2001). Their approaches involved first applying an appropriate transformation to handle the unobserved effect and then using instruments in a generalized method of moments (GMM) framework. On the other hand, if we ignore the measurement error problem in Eq. (1), then the model belong to the class of nonseparable panel data models, which have been studied in: Evdokimov (2011), Chernozhukov, Fernández-Val, Hahn, and Newey (2013), Hoderlein and White (2012), Chen and Swanson (2012), Hoderlein and Mammen (2007), Altonji and Matzkin (2005), and Chernozhukov, Fernandez-Val, Hoderlein, Holzmann, and Newey (2015). In particular, Chernozhukov, Fernández-Val, Hahn, and Newey (2013), Graham and Powell (2012), and Hoderlein and White (2012) use changes over time in x to obtain the ceteris paribus effect of x on y for identification and estimation of nonseparable models. Wilhelm (2015) considers nonlinear panel data models with measurement error where fixed effects are additively separable. He differences out the fixed effects and provides a nonparametric identification result without requiring any extra variable other than outcomes and observed regressors. However, in nonseparable panel data models it is not clear how to remove the unobserved heterogeneity and address measurement error problems simultaneously (first differencing does not work), so there is a fundamental difference between additively separable models and nonseparable models.

[2]Our model could accommodate Eq. (23.13) in Koebel, Laisney, Pohlmeier, and Staat (2008) with $\delta = \delta_i$ which depends on individual i and thus the equation is a special case of our formula provided.

Besides the short panel data setting considered here, there is a lot of closely related work in the large panel literature, but not allowing for measurement error. Alvarez and Arellano (2003) investigate the linear panel regression models with fixed effects for large n, T, and they find that their GMM estimator has an asymptotic bias of an order $1/n$ and does not cause bias for large T. Akashi and Kunitomo (2012) use the approach in Alvarez and Arellano (2003) to study panel dynamic simultaneous equation models. Hahn and Kuersteiner (2002) characterize the bias of the fixed effect estimator by allowing both n and T to approach infinity and the ratio n/T to approach a constant.

We develop an identification technique that builds on previous work of Schennach (2007), concerning the identification and estimation of nonlinear measurement error models with instruments. The identification strategy is to employ Fourier transforms of conditional expectations of observable variables and to provide a closed form solution to the regression function based on these transforms. We generalize the method of Schennach (2007) by allowing for a measurement error term in the regression function with an additional unobserved individual-specific effect in a panel data setting. The proposed method works in a way that panel data contains enough information on observables to identify the mismeasured variable X_{it}^*, and the unobserved individual-specific effect C_i. While the past observables are used as instruments to control the measurement error problem, the time averages of perfectly observed variables are used to restrict the unobserved individual-specific effect by a correlated random effects specification. Altonji and Matzkin (2005) and Wooldridge (2005) have used correlated random effects (CRE) approaches to nonlinear panel data models to control the unobserved individual-specific effect. Thus, the nonseparable regression function of interest also admits a similar representation of the closed form solution in Schennach (2007) under a mild regularity condition.

We propose an estimation method that closely follows the identification result, in particular it builds on knowledge of the three conditional expectations. We propose a sieve minimum distance (hereafter SMD) estimator for the parameters of interest. Then, estimating the parameters of interest by implementing the methods of series or sieve estimation developed in Ai and Chen (2007), which is an extension of SMD estimation in Ai and Chen (2003) and Newey and Powell (2003). The estimation procedure consists of applying the SMD method to a vector of moment conditions with different conditioning variables related to the identification result. It follows that the SMD estimator for the finite-dimensional parameters of the structural function is $\sqrt{n}$-consistent and asymptotically normally distributed.

The rest of the paper is organized as follows. Section 2 describes the identification assumptions and strategy for nonlinear panel data models with measurement errors. Section 3 covers the SMD estimation procedure based on the identification restrictions in Section 2. Section 4 discusses the implementation of the SMD estimator and presents its Monte Carlo simulation. Section 5 presents our empirical application, the elasticity of labor supply. Section 6 concludes. All proofs are collected in the Appendix.

2. Semiparametric Identification

Without loss of generality, we consider both W_{it} and X_{it}^* to be scalar quantities (a multivariate case can be straightforwardly provided). To avoid confusion, upper case letters are used exclusively for random variables and lower case letters are used exclusively for non-random quantities corresponding to its upper case random variables. The data $\{y_{it}, w_{it}, x_{it}\}$ is independently and identically distributed across i for each t and it is an observable random sample for $\{Y_{it}, W_{it}, X_{it}\}$ for $i = 1, 2, \ldots, n$ and $t = 1, \ldots, T \geq 2$.

Assumption 2.1. *(Correlated Random Effects (CRE)) There exists a nonzero coefficient* λ_0 *such that*

$$C_i = \lambda_0 \overline{W}_i + \eta_i,$$

where $\overline{W}_i = \frac{1}{T}\sum_{t=1}^{T} W_{it}$ *is denoted as the time average of the perfectly observed explanatory variables. The remainder term* η_i *is independent of* $\overline{W}_i$.

Assumption 2.1 can be generalized to include more perfectly observed explanatory variables. For example, if there exists another time-invariant variable $\overline{Z}_i$, we can consider the following CRE specification

$$C_i = \lambda_{01}\overline{W}_i + \lambda_{02}\overline{Z}_i + \eta_i.$$

Including more control variables in the specification may make the independent assumption of the projection error η_i more reasonable.

Assumption 2.2. *(Classical measurement error):*
(i)(Past variables as IV) There exists an unknown function h_t at time t satisfying

$$X_{it}^* = h_t(G_{i,<t}) + V_{it},$$

where $G_{i,<t} = (W_{it-1}, X_{it-1}, \ldots, W_{i1}, X_{i1})$, while V_{it} is independent of $G_{i,<t}$ and $E[V_{it}] = 0$.
(ii)(Measurement error)

$$X_{it} = X_{it}^* + e_{it}, \quad E[e_{it}|W_{it}, G_{i,<t}, V_{it}, \overline{W}_i, \eta_i, U_{it}] = 0$$

(iii)(Conditional mean independence)

$$E[U_{it}|W_{it}, G_{i,<t}, V_{it}, \overline{W}_i] = 0;$$

(iv)(Independent Distribution) The remainder error of CRE η_i and the unobservable V_{it} are independent.

The setting for the measurement errors is the same as Schennach (2007). She uses external instruments to identify her nonlinear errors-in-variables model. Assumption 2.2(i) can be regarded as a control function assumption. It uses the past variables as instruments to construct the estimable $h_t(G_{i,<t})$ thereby to extract the independent unobservable variable V_{it} from the unobservable true regressor X_{it}^*. The assumption is commonly used for identification of nonlinear models.[3] We may further assume that X_{it}^* fol-

[3]Combining Assumption 2.2(i) and (ii) yields $X_{it} = h_t(G_{i,<t}) + V_{it} + e_{it}$. As mentioned in Schennach (2007), an indirect test of the validity of the independence of V_{it} in Assumption 2.2(i) and conditional mean independence of e_{it} in Assumption 2.2(ii) can be conducted by testing the dependence of the estimated residual from regressing X_{it} on $h_t(G_{i,<t})$.

lows a first order stationary (Markov-type) motion by setting $X_{it}^* = h(W_{it-1}, X_{it-1}) + V_{it}$. Assumption 2.2(ii) implies that $\mathrm{E}[X_{it}^* e_{it}] = 0$, i.e., there is no correlation between the unobserved true regressor and the measurement error. Assumption 2.2(iii) only imposes the standard conditional moment restriction that $\mathrm{E}[U_{it}|W_{it}, G_{i,<t}, V_{it}, \overline{W}_i] = 0$; the disturbance U_{it} does not have to be independent of W_{it}, $G_{i,<t}$, V_{it}, and $\overline{W}_i$ and the distribution of U_{it} does not have to be the same across time periods. This implies that U_{it} can have an AR(1) stochastic process, for example.

As mentioned in Eq. (A.3), the measurement error equation and correlated random effects can be defined as follows:

$$X_{it}^* = \widetilde{G}_{i,<t} - \widetilde{V}_{it}, \text{ and } C_i = \lambda_0 \overline{W}_i - \widetilde{\eta}_i,$$

where $h_t(G_{i,<t}) \equiv \widetilde{G}_{i,<t} = \mathrm{E}[X_{it}|G_{i,<t}]$, $\widetilde{V}_{it} = -V_{it}$, and $\widetilde{\eta}_i = -\eta_i$. The following assumption guarantees that the Fourier transforms of the related conditional expectations are well defined.

Assumption 2.3. *Define $R_t(g, \overline{w}; w) = E[Y_{it}|W_{it} = w, \widetilde{G}_{i,<t} = g, \overline{W}_i = \overline{w}]$ and $S_t(g, \overline{w}; w) = E[X_{it}Y_{it}|W_{it} = w, \widetilde{G}_{i,<t} = g, \overline{W}_i = \overline{w}]$, and consider these as functions of $g, \overline{w}$ for fixed values of w. They belong to a function space $\mathscr{S}_\gamma$ that contains functions $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ satisfying*

$$\int (1 + \xi^\top \xi)^r |f(\xi)| d\xi \le A < \infty, \textit{ for some } \gamma > 0.$$

Assumption 2.3 ensures that the Fourier transforms of the conditional expectations to be well defined members of a subclass of locally integrable functions.

Define the Fourier transforms of the function m and the conditional expectations $R_t(g, \overline{w}; w)$ and $S_t(g, \overline{w}; w)$ defined in Assumption 2.3:

$$\mathscr{F}_y(w, \xi_1, \xi_2) = \int\int R_t(g, \overline{w}; w) e^{\mathbf{i}\xi_1 g} e^{\mathbf{i}\xi_2 \overline{w}} dg d\overline{w} \tag{2}$$

$$\mathscr{F}_{xy}(w, \xi_1, \xi_2) = \int\int S_t(g, \overline{w}; w) e^{\mathbf{i}\xi_1 g} e^{\mathbf{i}\xi_2 \overline{w}} dg d\overline{w} \tag{3}$$

$$\mathscr{F}_m(w, \xi_1, \xi_2; \beta_0) = \int\int m\left(w, x, c; \beta_0\right) e^{\mathbf{i}\xi_1 x} e^{\mathbf{i}\xi_2 c} dx dc, \tag{4}$$

where $\mathbf{i} = \sqrt{-1}$. Define also $\phi_v(\xi_1) = \int e^{\mathbf{i}\xi_1\tilde{v}} f_{\tilde{V}_{it}}(\tilde{v})d\tilde{v}$ and $\phi_\eta(\xi_2) = \int e^{\mathbf{i}\xi_2\tilde{\eta}} f_{\tilde{\eta}_i}(\tilde{\eta})d\tilde{\eta}$, where $f_{\tilde{V}_{it}}(\tilde{v})$ and $f_{\tilde{\eta}_i}(\tilde{\eta})$ are the density functions of $\tilde{V}_{it}$ and $\tilde{\eta}_i$, respectively.

Lemma 2.1. *Suppose that Assumptions 2.1, 2.2, and 2.3 hold. Then,*

$$\mathscr{F}_y(w,\xi_1,\xi_2) = \frac{1}{\lambda_0}\mathscr{F}_m(w,\xi_1,\frac{\xi_2}{\lambda_0})\phi_v(\xi_1)\phi_\eta(\frac{\xi_2}{\lambda_0}), \tag{5}$$

$$\mathscr{F}_{xy}(w,\xi_1,\xi_2) = \frac{1}{\lambda_0} - \mathbf{i}\frac{\partial \mathscr{F}_m(w,\xi_1,\frac{\xi_2}{\lambda_0})}{\partial \xi_1}\phi_v(\xi_1)\phi_\eta(\frac{\xi_2}{\lambda_0}). \tag{6}$$

Proof. See the appendix.

Assumption 2.4. *Suppose that: (i)* $\int |\tilde{v}| f_{\tilde{V}_{it}}(\tilde{v})d\tilde{v} \le A < \infty$, $\int |\tilde{\eta}| f_{\tilde{\eta}_i}(\tilde{\eta})d\tilde{\eta} \le A < \infty$; *and (ii) the characteristic functions* $\phi_v(\xi_1) \neq 0$, *and* $\phi_\eta(\xi_2) \neq 0$ *are continuously differentiable for all* $\xi_1, \xi_2 \in \mathbb{R}$.

Assumption 2.5. *Set* Θ *as a parameter space containing* β_0. *There exists a finite or infinite constant* $\bar{\zeta} > 0$ *and some* w_{it} *such that for all* $\beta \in \Theta$: *(i)* $\mathscr{F}_m(w_{it},\xi_1,\xi_2;\beta) \neq 0$ *almost everywhere in* $[-\bar{\zeta},\bar{\zeta}]^2$ *and (ii)* $\mathscr{F}_m(w_{it},\xi_1,\xi_2;\beta) = 0$ *for all* $|\zeta_1|,|\zeta_2| > \bar{\zeta}$.

Assumptions 2.4 and 2.5 are standard in the deconvolution literature. Assumption 2.4(ii) requires that the characteristic functions of V and $\tilde{\eta}$ are non-vanishing, which excludes uniform or triangular distributions, for example.

Exploiting the conditional mean function in Eq. (A.5) by replacing $f_{\tilde{\eta}_i}(\tilde{\eta})$ by $f_{\tilde{\eta}_i;\gamma}(\tilde{\eta})$, we have the following result. Denote $\gamma = (\beta,\lambda)$ and γ is a $(d_\beta + 2)\times 1$-dimensional vector. Consider the parametric conditional mean function in Eq. (A.16):

$$\begin{aligned} R_t(g,\overline{w};w,\gamma) &= \mathrm{E}[Y_{it}|W_{it} = w, \tilde{G}_{i,<t} = g, \overline{W}_i = \overline{w};\gamma] \\ &= \int\int m\left(w, g - \tilde{v}_{it}, \lambda_1\overline{w} - \tilde{\eta}_i;\beta\right) f_{\tilde{V}_{it}}(\tilde{v}) f_{\tilde{\eta}_i;\gamma}(\tilde{\eta})d\tilde{v}d\tilde{\eta}. \end{aligned}$$

Define the gradient of $\mathrm{E}[Y_{it}|W_{it} = w, \tilde{G}_{i,<t} = g, \overline{W}_i = \overline{w};\gamma]$ and the information matrix as follows:

$$\nabla_\gamma \mathrm{E}[Y_{it}|W_{it} = w, \tilde{G}_{i,<t} = g, \overline{W}_i = \overline{w};\gamma] = \left(\frac{\partial R_t(g,\overline{w};w,\gamma)}{\partial \beta_1}, \dots, \frac{\partial R_t(g,\overline{w};w,\gamma)}{\partial \lambda_2}\right)^\intercal.$$

$$I(\gamma) = \mathrm{E}\left[\nabla_\gamma R_t(\widetilde{G}_{i,<t}, \overline{W}_i; W_{it}, \gamma) \cdot \nabla_\gamma R_t(\widetilde{G}_{i,<t}, \overline{W}_i; W_{it}, \gamma)^\top\right].$$

Assumption 2.6. *(Nonsingular Parametric Structure) Set* $\Gamma = \Theta \times \Upsilon$ *as a parameter space containing* (β_0, λ_0)*. The elements of the vector* $\nabla_\gamma E[Y_{it}|w_{it}, \widetilde{g}_{i,<t}, \overline{w}_i; \gamma]$ *exist and are continuous in* Γ *for each* $(w_{it}, \widetilde{g}_{i,<t}, \overline{w}_i)$*, and the matrix* $I(\beta_0, \lambda_0)$ *is nonsingular.*

Theorem 2.1. *Suppose that Assumptions 2.1, 2.2, 2.3, 2.4, 2.5, and 2.6 hold. Then, the three unknown parameters of interest, including the finite-dimensional parameters* β_0 *and* λ_0*, the distribution of the remainder error of control function approach* $f_{\widetilde{V}_{it}}(\widetilde{v})$*, and the distribution of the remainder error of CRE* $\eta_i, f_{\widetilde{\eta}_i}(\widetilde{\eta})$*, are identifiable.*

Proof. See the appendix.

There are two main steps for the identification strategy for Theorem 2.1. In the first step, we use the method of Theorem 1 in Schennach (2007) and of Theorem 3(B) in Zinde-Walsh (2014) to identify the distribution of measurement error. As for the second step we use the CRE specification and the properties of Fourier transforms on convolution functions to connect the distribution of the individual effect to a parametric conditional moment function. Then, the identification is achieved by the nonsingular parametric structure of the information matrix formed by the parametric conditional moment function of Assumption 2.6.

A quantity of interest in many applications is the partial effect. The magnitude of the partial effect evidently cannot be estimated at meaningful values of the individual effect. One solution is to average the partial effects across the distribution of the individual effect; this quantity is also identified by Theorem 2.1. With the identification of the distribution of η_i and the independence assumption of η_i in Assumption 2.1, we have $f(c|\overline{w}_i) = f_{\widetilde{\eta}_i}(-c + \lambda_0 \overline{w}_i)$. Then, the distribution of the individual effect can be identified with the identification of $f(c|\overline{w}_i)$ from the equation

$$f_{C_i}(c) = \int f(c|\overline{w}_i) \cdot \underbrace{f(\overline{w}_i)}_{\substack{\text{estimable} \\ \text{from data}}} \, d\overline{w}_i. \tag{7}$$

Suppose that x^*_{it} takes continuous values. The average partial effect (APE) for x^*_{it} at the point (w_0, x^*_0) is

$$\text{APE}(w_0, x^*_0) = \int_{\mathscr{C}} \frac{\partial m\left(w_{it}, x_{it}, c_i; \beta_0\right)}{\partial x_{it}}\Big|_{(w_{it}, x_{it})=(w_0, x^*_0)} f_{C_i}(c) dc. \tag{8}$$

Corollary 2.1. *Suppose that Assumptions 2.1, 2.2, 2.3, 2.4, 2.5, and 2.6 hold. Then, both the distribution of the individual effect and the average partial effect defined in Eq. (8) are identified.*

3. SMD Estimation

We have shown in Theorem 2.1 that the three unknown parameters of interest, including the finite-dimensional parameters β_0 and λ_0, the distribution of the remainder error of control function approach $f_{\widetilde{V}_{it}}(\widetilde{v})$, and the distribution of the remainder error of CRE η_i, $f_{\widetilde{\eta}_i}(\widetilde{\eta})$, are uniquely identified. The identification is based on knowledge of the three observable conditional expectations $\text{E}[X_{it}|G_{i,<t}]$, $\text{E}[Y_{it}|W_{it}, \widetilde{G}_{i,<t}, \overline{W}_i]$ and $\text{E}[X_{it}Y_{it}|W_{it}, \widetilde{G}_{i,<t}, \overline{W}_i]$, where $\widetilde{G}_{i,<t} = h_t(G_{i,<t})$. In general, the conditioning set is high dimensional and nonparametric estimation procedures will perform poorly. We impose a Markov assumption, which reduces the dimensionality considerably.

Assumption 3.1. *(Stationary Markov motion) The mismeasured covariate X^*_{it} follows a first order stationary Markov process, $X^*_{it} = h(W_{it-1}, X_{it-1}) + V_{it}$ for each t.*

Denote $\widetilde{H}_{i,<t} = h(W_{it-1}, X_{it-1})$ and $D_{it} = (W_{it}, W_{it-1}, X_{it-1}, \overline{W}_i)$. Under the assumptions of Theorem 2.1 and Assumption 3.1, we rewrite these conditional expectations as

follows:[4]

$$
\begin{aligned}
0 &\equiv \mathrm{E}[X_{it}|W_{it-1}, X_{it-1}] - h(W_{it-1}, X_{it-1}),\\
0 &\equiv \mathrm{E}[Y_{it}|D_{it}] - \int\int m\left(W_{it}, \widetilde{H}_{i,<t} - \widetilde{v}, \lambda_0\overline{W}_i - \widetilde{\eta}; \beta_0\right) f_{\widetilde{V}_{it}}(\widetilde{v}_{it}) f_{\widetilde{\eta}_i}(\widetilde{\eta}_i) d\widetilde{v} d\widetilde{\eta},\\
0 &\equiv \mathrm{E}[X_{it}Y_{it}|D_{it}] - \int\int (\widetilde{H}_{i,<t} - \widetilde{v}) m\left(W_{it}, \widetilde{H}_{i,<t} - \widetilde{v}, \lambda_0\overline{W}_i - \widetilde{\eta}; \beta_0\right)\\
&\qquad \times f_{\widetilde{V}_{it}}(\widetilde{v}) f_{\widetilde{\eta}_i}(\widetilde{\eta}) d\widetilde{v} d\widetilde{\eta}.
\end{aligned}
$$

Denote $\alpha_0 = (\beta_0, \lambda_0, f_{\widetilde{V}_{it}}(\cdot), f_{\widetilde{\eta}_i}(\cdot), h(\cdot))^\intercal$. Define the following residual functions:

$$
\begin{aligned}
\rho_1(X_{it}, Y_{it}, D_{it}; \alpha_0) &\equiv X_{it} - h(W_{it-1}, X_{it-1}),\\
\rho_2(X_{it}, Y_{it}, D_{it}; \alpha_0) &\equiv Y_{it} - \int\int m\left(W_{it}, \widetilde{H}_{i,<t} - \widetilde{v}, \lambda_0\overline{W}_i - \widetilde{\eta}; \beta_0\right) f_{\widetilde{V}_{it}}(\widetilde{v}) f_{\widetilde{\eta}_i}(\widetilde{\eta}) d\widetilde{v} d\widetilde{\eta},\\
\rho_3(X_{it}, Y_{it}, D_{it}; \alpha_0) &\equiv X_{it}Y_{it} - \int\int (\widetilde{H}_{i,<t} - \widetilde{v}) m\left(W_{it}, \widetilde{H}_{i,<t} - \widetilde{v}, \lambda_0\overline{W}_i - \widetilde{\eta}; \beta_0\right)\\
&\qquad \times f_{\widetilde{V}_{it}}(\widetilde{v}) f_{\widetilde{\eta}_i}(\widetilde{\eta}) d\widetilde{v} d\widetilde{\eta}.
\end{aligned}
$$

Define the 3×1 vector of residual functions $\rho(X_{it}, Y_{it}, D_{it}; \alpha_0)$ that contains $\rho_j(X_{it}, Y_{it}, D_{it}; \alpha_0)$, $j = 1, 2, 3$. The parameter vector $\alpha = (\beta, \lambda, f_V(\cdot), f_\eta(\cdot), h(\cdot))^\intercal$ has three infinite-dimensional nuisance parameters because of the presence of the unknown functions $\lambda, f_V(\cdot)$, $f_\eta(\cdot)$, and $h(\cdot)$. The conditional moments functions for α_0 can be summarized as the following conditional moment restrictions with different conditioning variables

$$
m(D_{it}; \alpha) \equiv \begin{pmatrix} m_1(W_{it-1}, X_{it-1}; \alpha) \\ m_2(D_{it}; \alpha) \\ m_3(D_{it}; \alpha) \end{pmatrix} \equiv \begin{pmatrix} \mathrm{E}[\rho_1(X_{it}, Y_{it}, D_{it}; \alpha)|W_{it-1}, X_{it-1}] \\ \mathrm{E}[\rho_2(X_{it}, Y_{it}, D_{it}; \alpha)|D_{it}] \\ \mathrm{E}[\rho_3(X_{it}, Y_{it}, D_{it}; \alpha)|D_{it}] \end{pmatrix},
$$

with $m(D_{it}; \alpha_0) = 0$. While the conditioning variable used in the first conditional moment restriction is (W_{it-1}, X_{it-1}), the conditioning variable used in the second and third conditional moment restriction is D_{it}. Therefore, the model fits into the general models of conditional moment restrictions with different conditioning variables in Ai and Chen

[4]The detailed derivations can be found in Eqs. (A.5) and (A.6) in the appendix.

(2007), which contain finite dimensional unknown parameters and infinite dimensional unknown functions.

We consider a nonparametric least squares (LS) regression estimator for each component of $m(D_{it};\alpha)$. Let $p^k(\cdot) = (p_1(\cdot),\dots,p_k(\cdot))^\top$ be a vector of some known univariate basis function and $p^k(\cdot,\dots,\cdot) = (p_1(\cdot,\dots,\cdot),\dots,p_k(\cdot,\dots,\cdot))^\top$ be a multivariate basis function generated by the tensor product construction. Let $p^{k_{1n}}(W_{it-1},X_{it-1}) = (p_1(W_{it-1},X_{it-1}),\dots, p_{k_{1n}}(W_{it-1},X_{it-1}))^\top$ and $H_1 = (p^{k_{1n}}(W_{11},X_{11}),\dots,P^{k_{1n}}(W_{n,T-1},X_{n,T-1}))^\top$. The series LS estimator of $m_1(W_{it-1},X_{it-1};\alpha)$ is given by

$$
\begin{aligned}
(9) \qquad & \widehat{m}_1(W_{it-1},X_{it-1};\alpha) \\
&= p^{k_{1n}}(W_{it-1},X_{it-1})^\top (H_1^\top H_1)^{-1} \sum_{i=1}^{n}\sum_{t=3}^{T} p^{k_{1n}}(W_{it-1},X_{it-1})\rho_1(X_{it},Y_{it},D_{it};\alpha).
\end{aligned}
$$

As for the other conditional moment restrictions, for $j = 2,3$, denote the $k_{jn}\times 1$ vector of approximating functions as $p^{k_{jn}}(D_{it}) = (p_1(D_{it}),\dots,p_{k_{jn}}(D_{it}))^\top$, which is constructed from some known basis functions for any square integrable real-valued function of D_{it}. A linear consistent sieve estimator $\widehat{m}_j(D_{it};\alpha)$ can be obtained by regressing $\rho_j(X_{it},Y_{it},D_{it};\alpha)$ on $p^{k_{jn}}(D_{it})$, whence

$$
(10) \qquad \widehat{m}_j(D_{it};\alpha) = p^{k_{jn}}(D_{it})^\top (H_j^\top H_j)^{-1} \sum_{i=1}^{n}\sum_{t=3}^{T} p^{k_{jn}}(D_{it})\rho_j(X_{it},Y_{it},D_{it};\alpha),
$$

where $H_j = (p^{k_{jn}}(D_{12}),\dots,P^{k_{jn}}(D_{nT}))^\top$. It follows that $\widehat{m}(D_{it};\alpha) \equiv (\widehat{m}_1(W_{it-1},X_{it-1};\alpha), \widehat{m}_2(D_{it};\alpha),\widehat{m}_3(D_{it};\alpha))^\top$ is a consistent estimator for $m(D_{it};\alpha)$ and $\mathscr{A}_n$ is a sequence of approximating sieve spaces for the parameter space $\mathscr{A}$ containing α_0. The SMD estimator $\widehat{\alpha}_n$ minimizes the following sample analog of a minimum distance objective function with the parameters restricted to the sieve spaces, $\mathscr{A}_n$:

$$
\widehat{\alpha}_n = \arg\min_{\alpha\in\mathscr{A}_n} \frac{1}{n(T-1)} \sum_{i=1}^{n}\sum_{t=3}^{T} \widehat{m}(D_{it};\alpha)^\top \widehat{m}(D_{it};\alpha).
$$

For simplicity, we use the identity weighting matrix in the sample objective function. There are two approximations in the optimization problem to make the estimator feasible and consistent. One is that $\widehat{m}(D_{it};\alpha)$ approximates $m(D_{it};\alpha)$ and the other is that $\mathscr{A}_n$ approximates $\mathscr{A}$. This GMM type estimator is proposed by Ai and Chen (2007) and is called a modified SMD estimator comparing to the sieve minimum distance estimation that are identified through a conditional moment restriction model with the same conditioning variables in each conditional moment restriction in Ai and Chen (2003) and Newey and Powell (2003). Ai and Chen (2007) show that the modified S-MD estimator is consistent, and the parametric components of the estimator have an asymptotically normal limiting distribution under suitable regularity conditions.

4. Monte Carlo Simulation

This section presents the finite sample properties of the SMD estimator (defined in Section 3) by a Monte Carlo simulation. We focus on the estimation of β_0 and λ_0, which correspond to the regression function $m\left(W_{it},X_{it}^*,C_i;\beta_0\right)$ and the CRE $C_i = \lambda_{01}\overline{W}_i + \lambda_{02}\overline{Z}_i + \eta_i$, respectively. However, the distributions of $f_{\widetilde{V}_{it}}(\widetilde{v})$ and $f_{\widetilde{\eta}_i}(\widetilde{\eta})$ are treated nonparametrically and will be approximated by a sequence of truncated sieves.

The simulation design is according to the following DGP. Denote Trun($\Phi,[a,b]$) as the distribution of a random variable generated by $\Phi^{-1}(u\cdot(\Phi(b)-\Phi(a))+\Phi(a))$, where Φ is the CDF of standard normal distribution, while Φ^{-1} is the inverse of Φ and u is a uniform random variable on $[0,1]$. Both W_{i1}, and X_{i1}^* are generated from Trun($\Phi,[0,1]$). The covariates (W_{it},X_{it}^*) for $t=2,3$ are generated according to

$$W_{it} = \rho W_{it-1} + U_{W,it-1} \text{ with } U_{W,it-1} \sim \text{Trun}(\Phi,[-2,2]),$$
$$X_{it}^* = \rho X_{it-1}^* + U_{X,it-1} \text{ with } U_{X,it-1} \sim \text{Trun}(\Phi,[-2,2]),$$

where $\rho = 0.8$. The specification for the measurement error problem is:

$$X_{it} = X_{it}^* + e_{it}, \text{ where } e_{it} \sim \text{Trun}(\Phi,[-2,2]).$$

Let $\overline{W}_i = \frac{1}{3}\sum_{t=1}^{3} W_{it}$ and $\overline{Z}_i \sim$ Trun($\Phi,[0,1]$). Then, the specification for the individual effect is

$$C_i = \lambda_{01}\overline{W}_i + \lambda_{02}\overline{Z}_i + \eta_i, \text{ where } (\lambda_{01}, \lambda_{02}) = (-0.5, 0.5), \eta_i \sim \text{Trun}(\Phi, [-2, 2]).$$

Set $\beta_0 = (\beta_{00}, \beta_{01}, \beta_{02}) = (0.5, 0.5, -0.5)$. We consider three specifications for the regression function:

$$\begin{aligned}
\text{Simulation I: } m\left(W_{it}, X_{it}^*, C_i; \beta_0\right) &= \beta_{00} + \beta_{01}W_{it} + \beta_{02}X_{it}^{*2} + C_i, \\
\text{Simulation II: } m\left(W_{it}, X_{it}^*, C_i; \beta_0\right) &= \left(\beta_{00} + \beta_{01}W_{it} + \beta_{02}X_{it}^* + C_i\right)^2, \\
\text{Simulation III: } m\left(W_{it}, X_{it}^*, C_i; \beta_0\right) &= \left(\beta_{00} + \beta_{01}(1+C_i)W_{it} + \beta_{02}(1+C_i)X_{it}^* + C_i\right)^2.
\end{aligned}$$

The SMD procedure requires approximating the three nonparametric parts by sieves, including the conditional expectation function h_t, $f_{\widetilde{V}_{it}}(\widetilde{v})$ and $f_{\widetilde{\eta}_i}(\widetilde{\eta})$. We use the polynomial base in the sieve approximation series for h_t,

$$h_t(w, x) = \gamma_0 + \gamma_1 w + \gamma_2 x + \gamma_3 w^2 + \gamma_4 x^2 + \gamma_5 xw.$$

Let f_1 and f_2 be the nonparametric series estimators for $f_{\widetilde{V}_{it}}(\widetilde{v})$ and $f_{\widetilde{\eta}_i}(\widetilde{\eta})$, respectively. We construct $f_1^{1/2}$ and $f_2^{1/2}$ by univariate Hermite functions,

$$f_1^{1/2}(\widetilde{v}) = \sum_{i=0}^{3} \delta_{1i} H_i(\widetilde{v}), \quad f_2^{1/2}(\widetilde{\eta}) = \sum_{i=0}^{3} \delta_{2i} H_i(\widetilde{\eta}),$$

where $H_0(x) = e^{-\frac{x^2}{2}}$, $H_1(x) = xe^{-\frac{x^2}{2}}$, $H_2(x) = (x^2-1)e^{-\frac{x^2}{2}}$, $H_3(x) = (x^3-3x)e^{-\frac{x^2}{2}}$. The sieve coefficients of both f_1 and f_2 need to satisfy density restrictions. Because the Hermite functions form an orthogonal series that satisfies $\int_{-\infty}^{\infty} H_n(x)H_m(x)dx = \sqrt{2\pi}n!\delta_{nm}$, where $\delta_{nm} = 1$ if $n = m$, and $\delta_{nm} = 0$ otherwise, the density restriction on the sieve coefficients is $\sqrt{2\pi}(\delta_{10}^2 + \delta_{11}^2 + 2!\delta_{12}^2 + 3!\delta_{13}^2) = 1$.

As discussed in Section 3, we use a tensor product polynomial sieve to approximate each component of the conditional mean function $m(D_{it}; \alpha)$, which are the sets of instruments. While we choose the set of instruments for the argument of $p^{k_{1n}}(W_{it-1}, X_{it-1})$

as $\{1, W_{it-1}, X_{it-1}, W_{it-1}^2, W_{it-1}X_{it-1}, X_{it-1}^2\}$ for the first conditional moment restriction, the set of instruments for each argument of $p^{k_{2n}}(D_{it})$, and $p^{k_{3n}}(D_{it})$ for the second and third conditional moment restrictions is being chosen from $\{1, W_{it}, W_{it-1}, X_{it-1}, \overline{W}_i, \overline{Z}_i,$ $W_{it}^2, W_{it}W_{it-1}, W_{it}X_{it-1}, W_{it}\overline{W}_i, W_{it}\overline{Z}_i, W_{it-1}^2, W_{it-1}X_{it-1}, W_{it-1}\overline{W}_i, W_{it-1}\overline{Z}_i, X_{it-1}^2, X_{it-1}\overline{W}_i,$ $X_{it-1}\overline{Z}_i, \overline{W}_i^2, \overline{W}_i\overline{Z}_i, \overline{Z}_i^2\}$. The total number of the instruments is $\sum_{j=1}^{3} k_{jn} = 6 + 2 \times 21 = 48$.

The 150 replications of 500, and 1000 observations are drawn from these three data generating processes corresponding to the different regression function $m(\cdot)$. The simulation results of Tables 1-2 show the proposed SMD estimator performs well in these samples. The mean estimates are almost the same as median estimates of different sample sizes and simulation designs. This implies that there does not exist skewness in their respective distributions. For each estimated coefficient, the RMSE declines as the sample size is increased, as would be expected for this simulation. We can further use Eq. (7) with the estimated coefficient of λ and observation of $\overline{w}_i$ to recover the distribution of the individual effect $f_{C_i}(\cdot)$ and then APEs can be calculated by Eq. (8). Tables 3-4 report the mean, standard deviation (SD) and RMSE of the APE estimation results. All estimations are nearly unbiased and the APE estimator has the best performance in DGP II. In terms of RMSE, the RMSE almost declines as the sample size is increased.

5. Empirical Study

In this section, we apply our proposed nonlinear panel data model to investigate the effect of the hourly wage rate of individuals on their labor supply given their demographic variables. The dependent variable is the log of annual hours of work for those with positive working hours. The variable of interest is the hourly wage rate. Measurement error can be a significant problem for the hourly wage rate in survey data. Our model allows for measurement error of the hourly wage rate and provides consistent estimate of the effects of interest. Our model uses the correlated random effect to control for unobserved time invariant factors such as individual unobserved skill level, ability, or motivation factors which may be correlated with the hourly wage rate.[5] The data is from Ziliak (1997), Waves IXX-XXI of the PSID. Table 5 presents summary statistics for the working hours, the hourly wage rate, and socioeconomic variables. The between and within sample standard deviations are 0.233 and 0.172 for ln($hours$) and 0.432 and 0.118 for ln($wage$), respectively. We have a three-periods of the panel data with a cross-sectional size 532 of males.

We consider the following empirical model for labor supply:

$$\ln(hours_{it}) = \beta_0 + \beta_1(1+c_i)\ln(wage_{it}) + \beta_2 kids_{it} + \beta_3 age_{it} + \beta_4 age_{it}^2 + \beta_5 disab_{it} + \beta_6 t + c_i + u_{it}.$$

This specification allows interactions between observables and unobservables through the term $\beta_1(1+c_i)$; working in differences does not eliminate this effect. The growth of individual hours is allowed to proceed heterogeneously unlike many studies in the literature, MaCurdy (1981). The variable c_i represents unmeasured ability or motivation factors that affect hours of working, while u_{it} may contain time-varying unobserved macro shocks. Because the true wage rate of each individual is subject to a misreporting error, the measurement error of the variable ln($wage_{it}$) is likely to occur.[6] The vector of time-varying covariates is $(kids_{it}, age_{it}, age_{it}^2, disab_{it})^\top$ and the time averages of these variables are used in the CRE specification in this estimation of labor supply elasticity. A theoretical model of labor supply implies that there are two effects of a wage increase on labor supply, one is the income effect and the other is the substitution effect. While the income effect induces less work, the substitution effect increases more work. Because both effects work in opposite directions, the overall effect of a wage increase on labor supply is ambiguous.

The identification assumptions in Section 2 must hold to apply the proposed sieve GMM estimator. The following discussion presents these assumptions for this empirical application. Assumption 2.1 is the modelling of the individual effect and is to

[5] Borjas (2013) reviews the literature on the estimation of the labor supply elasticity and also discusses the problems caused by measurement error.

[6] See detailed discussion in Bound, Brown, and Mathiowetz (2001).

replace the unobserved individual effect with its linear projection onto the time average of explanatory variables. This allows a correlation between C_i and W_{it}, X_{it}^*. If C_i represents the willingness to work long hours, then the modelling indicates that it would depend on the average number of kids, age, and the average status of disability. Assumption 2.2 is based on the validity of using the past variables as IVs. This setup attempts to eliminate the endogeneity bias of the measurement error by exploiting the zero correlation of the measurement error at period t with other explanatory variables in the past, and the past explanatory variables might be related to the current hourly wage rate. Assumptions 2.3-2.5 ensures the Fourier transforms or the characteristic functions of the related conditional mean functions and density functions are well defined for algebraic manipulations, and are technical conditions. Assumption 2.6 implies that there is a nonsingular parametric structure around the population parameters.

Three comparable estimators can be constructed based on this regression model without the $(1+c_i)$ multiplying wage, i.e.,

$$\ln(hours_{it}) = \beta_0 + \beta_1 \ln(wage_{it}) + \beta_2 kids_{it} + \beta_3 age_{it} + \beta_4 age_{it}^2 + \beta_5 disab_{it} + \beta_6 t + c_i + u_{it}.$$

The first estimator (Linear Fixed Effect) is the fixed effect method using within transformation to remove the individual effect C_i, and the second estimator (First Differencing IV) is to use the first-difference and then estimate the parameters by using the past variables as IVs. We use $(kids_{it}, age_{it}, age_{it}^2, disab_{it})^\intercal$ from the periods $t-1$ and $t-2$ and $\ln(wage_{it})$ from $t-2$ as instruments for the contemporaneous period. The third linear correlated random effects model is to estimate the parameters using the CRE specification in Assumption 2.1.

Table 6 reports the estimates obtained with our sieve GMM method and with the other three linear estimates. We find that the estimated coefficients for the elasticity are not much different to both models except for the one using the linear fixed effect method which is negative. The values of the coefficients in these estimates are -9.4%, 4.9%, 4.1%, and 3.3%. However, if we consider the estimates of APE then the estimate for the elasticity in our semi-parametric nonlinear panel data model is twice as the estimates in the linear first differencing IV model and the linear correlated random effect model. A 1% increase in wage exhibits an approximately 9.7% increase in working hours. Given the flexible nature of our estimation approach, the difference implies that the estimate in the other linear models might be biased downward when the measurement error problem is not accounted for. As for the sign of the labor supply elasticity, the estimates are all positive except for the fixed effect method and this indicates that the number of hours worked is increasing in the wage, i.e. the substitution effect is stronger than the income effect.

In Figure 1, the distribution of the error η in the CRE specification does not show any kind of symmetry so it is an asymmetric distribution. On the other hand, the distribution is a bimodal distribution because the distribution has two peaks. This indicates that there are two different groups for the error. The error falls into a bimodal distribution with a lot of values getting zero and a lot getting some value greater than zero.

6. Conclusion

This paper presents the semi-parametric identification and estimation of nonlinear panel data models with mismeasured variables and their corresponding average partial effects using only three periods of data. The approach addresses settings without external information such as a validation or replicate data set. This study was motivated by the richer structure of panel data. We have shown how to use past observables as instruments to identify the nonlinear regression model in the presence of measurement error, while applying the correlated random effects specification to control the unobserved individual heterogeneity.

In simulation experiments we showed that the sieve GMM estimators perform well for both linear and nonlinear panel models with measurement errors. In the application we found that the substitution effect is stronger than the income effect and a 1% increase in wage enhances an approximately 10% increase in working hours.

Appendix

A. Identification Results

The proof of Lemma 2.1: Because both W_{it} and X^*_{it} are a scalar, we can write $C_i = \lambda_0 \overline{W}_i + \eta_i$. Combining Assumptions 2.2(i) and (ii) yields

(A.1)
$$X_{it} = h_t(G_{i,<t}) + V_{it} + e_{it}.$$

Taking conditional expectation with respect to $G_{i,<t}$, and applying zero conditional mean of V_{it}, and e_{it} implies:

(A.2)
$$\mathrm{E}[X_{it}|G_{i,<t}] = h_t(G_{i,<t}) \equiv \widetilde{G}_{i,<t}.$$

Rewrite the measurement error equation and correlated random effects as follows:

(A.3)
$$X^*_{it} = \widetilde{G}_{i,<t} - \widetilde{V}_{it}, \text{ and } C_i = \lambda_0 \overline{W}_i - \widetilde{\eta}_i.$$

Use the relations in Eq. (A.3) to write

(A.4)
$$Y_{it} = m\left(W_{it}, \widetilde{G}_{i,<t} - \widetilde{V}_{it}, \lambda_0 \overline{W}_i - \widetilde{\eta}_i; \beta_0\right) + U_{it}$$

Then, using the conditional mean independence of U_{it} in Assumption 2.2(iii) and independence of $\widetilde{V}_{it}$ and $\widetilde{\eta}_i$ in Assumption 2.2(iv), we obtain

(A.5)
$$\begin{aligned}&\mathrm{E}[Y_{it}|w_{it}, \widetilde{g}_{i,<t}, \overline{w}_i] \\ &= \int\int m\left(w_{it}, \widetilde{g}_{i,<t} - \widetilde{v}_{it}, \lambda_0 \overline{w}_i - \widetilde{\eta}_i; \beta_0\right) f_{\widetilde{V}_{it}}(\widetilde{v}_{it}) f_{\widetilde{\eta}_i}(\widetilde{\eta}_i) d\widetilde{v}_{it} d\widetilde{\eta}_i.\end{aligned}$$

Expanding out the term $X_{it}Y_{it}$ and taking conditional expectation with respect to $(w_{it}, \widetilde{g}_{i,<t}, \overline{w}_i)$ results in

$$
\begin{aligned}
&\mathrm{E}[X_{it}Y_{it}|w_{it}, \widetilde{g}_{i,<t}, \overline{w}_i] \\
&= \mathrm{E}[(\widetilde{G}_{i,<t} - \widetilde{V}_{it})m\left(W_{it}, \widetilde{G}_{i,<t} - \widetilde{V}_{it}, \lambda_0 \overline{W}_i - \widetilde{\eta}_i; \beta_0\right)|w_{it}, \widetilde{g}_{i,<t}, \overline{w}_i] \\
&\quad + \mathrm{E}[\Delta X_{it} m\left(W_{it}, \widetilde{G}_{i,<t} - \widetilde{V}_{it}, \lambda_0 \overline{W}_i - \widetilde{\eta}_i; \beta_0\right)|w_{it}, \widetilde{g}_{i,<t}, \overline{w}_i] \\
&\quad + \mathrm{E}[(\widetilde{G}_{i,<t} - \widetilde{V}_{it})U_{it}|w_{it}, \widetilde{g}_{i,<t}, \overline{w}_i] + \mathrm{E}[\Delta X_{it}U_{it}|w_{it}, \widetilde{g}_{i,<t}, \overline{w}_i] \\
&= \mathrm{E}[(\widetilde{G}_{i,<t} - \widetilde{V}_{it})m\left(W_{it}, \widetilde{G}_{i,<t} - \widetilde{V}_{it}, \lambda_0 \overline{W}_i - \widetilde{\eta}_i; \beta_0\right)|w_{it}, \widetilde{g}_{i,<t}, \overline{w}_i], \\
&= \int\int (\widetilde{g}_{i,<t} - \widetilde{v}_{it})m\left(w_{it}, \widetilde{g}_{i,<t} - \widetilde{v}_{it}, \lambda_0 \overline{w}_i - \widetilde{\eta}_i; \beta_0\right) f_{\widetilde{V}_{it}}(\widetilde{v}_{it}) f_{\widetilde{\eta}_i}(\widetilde{\eta}_i) d\widetilde{v}_{it} d\widetilde{\eta}_i. \qquad \text{(A.6)}
\end{aligned}
$$

where we have used the zero conditional mean of ΔX_{it} in Assumption 2.2(ii), the zero conditional mean of U_{it} in Assumption 2.2(iii), and the law of iterated expectation. Given w_{it}, taking the Fourier transform on both sides of Eqs. (A.5) and (A.6) with respect to $\widetilde{G}_{i,<t}$, and $\overline{W}_i$, we have

$$
\begin{aligned}
&\mathscr{F}_y(w_{it}, \xi_1, \xi_2) \\
&= \int\int \mathrm{E}[Y_{it}|w_{it}, \widetilde{g}_{i,<t}, \overline{w}_i] e^{\mathbf{i}\xi_1 \widetilde{g}_{i,<t}} e^{\mathbf{i}\xi_2 \overline{w}_i} d\widetilde{g}_{i,<t} d\overline{w}_i \\
&= \int\int \left(\int\int m\left(w_{it}, \widetilde{g}_{i,<t} - \widetilde{v}_{it}, \lambda_0 \overline{w}_i - \widetilde{\eta}_i; \beta_0\right) f_{\widetilde{V}_{it}}(\widetilde{v}_{it}) f_{\widetilde{\eta}_i}(\widetilde{\eta}_i) d\widetilde{v}_{it} d\widetilde{\eta}_i\right) e^{\mathbf{i}\xi_1 \widetilde{g}_{i,<t}} e^{\mathbf{i}\xi_2 \overline{w}_i} d\widetilde{g}_{i,<t} d\overline{w}_i \\
&= \frac{1}{\lambda_0}\left(\int\int m\left(w_{it}, x^*_{it}, c_i; \beta_0\right) e^{\mathbf{i}\xi_1 x^*_{it}} e^{\mathbf{i}\xi_2 \frac{c_i}{\lambda_0}} dx^*_{it} dc_i\right)\left(\int e^{\mathbf{i}\xi_1 \widetilde{v}_{it}} f_{\widetilde{V}_{it}}(\widetilde{v}_{it}) d\widetilde{v}_{it}\right)\left(\int e^{\mathbf{i}\xi_2 \frac{\widetilde{\eta}_i}{\lambda_0}} f_{\widetilde{\eta}_i}(\widetilde{\eta}_i) d\widetilde{\eta}_i\right) \\
&= \frac{1}{\lambda_0}\mathscr{F}_m(w_{it}, \xi_1, \frac{\xi_2}{\lambda_0})\phi_v(\xi_1)\phi_\eta(\frac{\xi_2}{\lambda_0}),
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathscr{F}_{xy}(w_{it},\xi_1,\xi_2)\\
&=\int\int \mathrm{E}[X_{it}Y_{it}|w_{it},\widetilde{g}_{i,<t},\overline{w}_i]e^{\mathbf{i}\xi_1\widetilde{g}_{i,<t}}e^{\mathbf{i}\xi_2\overline{w}_i}d\widetilde{g}_{i,<t}d\overline{w}_i\\
&=\int\int\left(\int\int(\widetilde{g}_{i,<t}-\widetilde{v}_{it})m\left(w_{it},\widetilde{g}_{i,<t}-\widetilde{v}_{it},\lambda_0\overline{w}_i-\widetilde{\eta}_i;\beta_0\right)f_{\widetilde{V}_{it}}(\widetilde{v}_{it})f_{\widetilde{\eta}_i}(\widetilde{\eta}_i)d\widetilde{v}_{it}d\widetilde{\eta}_i\right)e^{\mathbf{i}\xi_1\widetilde{g}_{i,<t}}e^{\mathbf{i}\xi_2\overline{w}_i}d\widetilde{g}_{i,<t}d\overline{w}_i\\
&=\frac{1}{\lambda_0}\left(\int\int x^*_{it}m\left(w_{it},x^*_{it},c_i;\beta_0\right)e^{\mathbf{i}\xi_1x^*_{it}}e^{\mathbf{i}\xi_2\frac{c_i}{\lambda_0}}dx^*_{it}dc_i\right)\left(\int e^{\mathbf{i}\xi_1\widetilde{v}_{it}}f_{\widetilde{V}_{it}}(\widetilde{v}_{it})d\widetilde{v}_{it}\right)\left(\int e^{\mathbf{i}\xi_2\frac{\widetilde{\eta}_i}{\lambda_0}}f_{\widetilde{\eta}_i}(\widetilde{\eta}_i)d\widetilde{\eta}_i\right)\\
&=\frac{1}{\lambda_0}-\mathbf{i}\frac{\partial\mathscr{F}_m(w_{it},\xi_1,\frac{\xi_2}{\lambda_0})}{\partial\xi_1}\phi_v(\xi_1)\phi_\eta(\frac{\xi_2}{\lambda_0}).
\end{aligned}
$$

This yields Eqs. (5) and (6). *Q.E.D.*

The proof of Theorem 2.1: We will recover $f_{\widetilde{V}_{it}}(\widetilde{v})$ first. Differentiating the definition of $\mathscr{F}_y(w_{it},\xi_1,\xi_2)$ in Eq. (2) with respect to ξ_1 yields

$$
\begin{aligned}
\frac{\partial}{\partial\xi_1}\mathscr{F}_y(w_{it},\xi_1,\xi_2)&=\frac{\partial}{\partial\xi_1}\int\int \mathrm{E}[Y_{it}|w_{it},\widetilde{g}_{i,<t}]e^{\mathbf{i}\xi_1\widetilde{g}_{i,<t}}e^{\mathbf{i}\xi_2\overline{w}_i}d\widetilde{g}_{i,<t}d\overline{w}_i\\
&=\mathbf{i}\int\int \mathrm{E}[\widetilde{G}_{i,<t}Y_{it}|w_{it},\widetilde{g}_{i,<t}]e^{\mathbf{i}\xi_1\widetilde{g}_{i,<t}}e^{\mathbf{i}\xi_2\overline{w}_i}d\widetilde{g}_{i,<t}d\overline{w}_i. \qquad \text{(A.7)}
\end{aligned}
$$

Notice that Eq. (6) can be written as $\frac{\partial\mathscr{F}_m(w_{it},\xi_1,\frac{\xi_2}{\lambda_0})}{\partial\xi_1}\phi_v(\xi_1)\phi_\eta(\frac{\xi_2}{\lambda_0})=\mathbf{i}\mathscr{F}_{xy}(w_{it},\xi_1,\xi_2)$. On the other hand, differentiating Eq. (5) with respect to ξ_1, we obtain

$$
\begin{aligned}
&\frac{\partial}{\partial\xi_1}\mathscr{F}_y(w_{it},\overline{w}_i,\xi_1,\xi_2)\\
&=\frac{1}{\lambda_0}\Big[\frac{\partial\mathscr{F}_m(w_{it},\xi_1,\frac{\xi_2}{\lambda_0})}{\partial\xi_1}\phi_v(\xi_1)+\mathscr{F}_m(w_{it},\xi_1,\frac{\xi_2}{\lambda_0})\frac{\partial\phi_v(\xi_1)}{\partial\xi_1}\Big]\phi_\eta(\frac{\xi_2}{\lambda_0})\\
&=\mathbf{i}\mathscr{F}_{xy}(w_{it},\overline{w}_i,\xi_1,\xi_2)+\frac{1}{\lambda_0}\mathscr{F}_m(w_{it},\xi_1,\frac{\xi_2}{\lambda_0})\frac{\partial\phi_v(\xi_1)}{\partial\xi_1}\phi_\eta(\frac{\xi_2}{\lambda_0})\\
&=\mathbf{i}\int\int \mathrm{E}[X_{it}Y_{it}|w_{it},\widetilde{g}_{i,<t},\overline{w}_i]e^{\mathbf{i}\xi_1\widetilde{g}_{i,<t}}e^{\mathbf{i}\xi_2\overline{x}_i}d\widetilde{g}_{i,<t}d\overline{x}_i\\
&\qquad+\frac{1}{\lambda_0}\mathscr{F}_m(w_{it},\xi_1,\frac{\xi_2}{\lambda_0})\frac{\partial\phi_v(\xi_1)}{\partial\xi_1}\phi_\eta(\frac{\xi_2}{\lambda_0}). \qquad \text{(A.8)}
\end{aligned}
$$

Combining Eqs. (A.7) and (A.8) yields

$$
\begin{aligned}
&\mathbf{i}\mathscr{F}_{(\widetilde{g}-x)y}(w_{it},\xi_1,\xi_2) \\
&\equiv \mathbf{i}\int\int \mathrm{E}[(\widetilde{G}_{i,<t}-X_{it})Y_{it}|w_{it},\widetilde{g}_{i,<t},\overline{w}_i]e^{\mathbf{i}\xi_1\widetilde{g}_{i,<t}}e^{\mathbf{i}\xi_2\overline{w}_i}d\widetilde{g}_{i,<t}d\overline{w}_i \\
&= \frac{1}{\lambda_0}\mathscr{F}_m(w_{it},\xi_1,\frac{\xi_2}{\lambda_0})\frac{\partial\phi_v(\xi_1)}{\partial\xi_1}\phi_\eta(\frac{\xi_2}{\lambda_0})
\end{aligned}
\tag{A.9}
$$

Because $\phi_v(\xi_1)$, $\phi_\eta(\xi_2)$, and $\mathscr{F}_m(w_{it},\xi_1,\xi_2;\beta_0)$ are all nonzero by Assumptions 2.4(ii) and 2.5, we can divide each side of Eq. (A.9) by the corresponding side of Eq. (5) to obtain

$$
-\mathbf{i}\mathscr{F}_{(\widetilde{g}-x)y}(w_{it},\xi_1,\xi_2)+\frac{\frac{\partial\phi_v(\xi_1)}{\partial\xi_1}}{\phi_v(\xi_1)}\mathscr{F}_y(w_{it},\xi_1,\xi_2)=0. \tag{A.10}
$$

By Theorem 1(b) in Zinde-Walsh (2014), there exists a unique function $Q(\xi_1)\equiv\frac{\frac{\partial\phi_v(\xi_1)}{\partial\xi_1}}{\phi_v(\xi_1)}$ such that

$$
-\mathbf{i}\mathscr{F}_{(\widetilde{g}-x)y}(w_{it},\xi_1,\xi_2)+Q(\xi_1)\mathscr{F}_y(w_{it},\xi_1,\xi_2)=0. \tag{A.11}
$$

Integrating the above equation from 0 to ξ_1 with the boundary condition $\phi_v(0)=\int f_{\widetilde{V}_{it}}(\widetilde{v}_{it})d\widetilde{v}_{it}=1$ yields

$$
\phi_v(\xi_1)=\exp\left(\int_0^{\xi_1}Q(\xi)d\xi\right).
$$

This implies that $\phi_v(\xi_1)$ is identified because it is expressed in terms of the Fourier transforms of observable conditional expectations. It follows that the distribution $f_{\widetilde{V}_{it}}(\widetilde{v}_{it})$ is identified. Rescaling ξ_2 by $\lambda_0\xi_2$ in Eq. (5) and rearranging the terms, we have

$$
\lambda_0\mathscr{F}_y(w_{it},\xi_1,\lambda_0\xi_2)=\mathscr{F}_m(w_{it},\xi_1,\xi_2;\beta_0)\phi_v(\xi_1)\phi_\eta(\xi_2), \tag{A.12}
$$

Solving $\phi_\eta(\xi_2)$ from the above equation yields

$$\phi_\eta(\xi_2) = \frac{\lambda_0 \mathscr{F}_y(w_{it}, \xi_1, \lambda_0 \xi_2)}{\mathscr{F}_m(w_{it}, \xi_1, \xi_2; \beta_0)\phi_v(\xi_1)}. \tag{A.13}$$

Because $\mathscr{F}_y(w_{it}, \xi_1, \xi_2)$, $\mathscr{F}_m(w_{it}, \xi_1, \xi_2; \beta)$ are all known from the data and the proposed semi-parametric regression function, and $\phi_v(\xi_1)$ is identified, we can generalize the relation into the following parametric function:

$$\phi_{\eta;\gamma}(\xi_2) = \frac{\lambda \mathscr{F}_y(w_{it}, \xi_1, \lambda \xi_2)}{\mathscr{F}_m(w_{it}, \xi_1, \xi_2; \beta)\phi_v(\xi_1)}, \tag{A.14}$$

where $\phi_{\eta;\gamma_0}(\xi_2) = \phi_\eta(\xi_2)$. Notice that the identification of the true parameter γ_0 leads to the identification of $\phi_\eta(\xi_2)$. Consider the following parametric function by applying the inverse Fourier transform to $\phi_{\eta;\gamma}(\xi_2)$:

$$f_{\widetilde{\eta}_i;\gamma}(\widetilde{\eta}) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\xi_2\widetilde{\eta}} \phi_{\eta;\gamma}(\xi_2) d\xi_2. \tag{A.15}$$

Evaluating the parametric function at γ_0, we have $f_{\widetilde{\eta}_i;\gamma_0}(\widetilde{\eta}) = f_{\widetilde{\eta}_i}(\widetilde{\eta})$ by the Fourier inversion theorem. Exploiting the conditional mean function in Eq. (A.5) by replacing $f_{\widetilde{\eta}_i}(\widetilde{\eta}_i)$ by $f_{\widetilde{\eta}_i;\gamma}(\widetilde{\eta})$, we have

$$\begin{aligned} &\mathrm{E}[Y_{it}|w_{it}, \widetilde{g}_{i,<t}, \overline{w}_i; \gamma] \\ &= \int\int m\left(w_{it}, \widetilde{g}_{i,<t} - \widetilde{v}_{it}, \lambda_1 \overline{w}_i - \widetilde{\eta}_i; \beta\right) f_{\widetilde{V}_{it}}(\widetilde{v}_{it}) f_{\widetilde{\eta}_i;\gamma}(\widetilde{\eta}_i) d\widetilde{v}_{it} d\widetilde{\eta}_i. \end{aligned} \tag{A.16}$$

with $\mathrm{E}[Y_{it}|w_{it}, \widetilde{g}_{i,<t}, \overline{w}_i; \gamma_0] = \mathrm{E}[Y_{it}|w_{it}, \widetilde{g}_{i,<t}, \overline{w}_i]$. Next, we will show that γ_0 is identifiable. If γ_0 is not locally identifiable. Then there exists a sequence of distinct parameters $\gamma_s \equiv (\beta_s, \lambda_s)$ approaching to $\gamma_0 = (\beta_0, \lambda_0)$ such that $\|(\beta_s, \lambda_s) - (\beta_0, \lambda_0)\| \neq 0$ and $\mathrm{E}[Y_{it}|w_{it}, \widetilde{g}_{i,<t}, \overline{w}_i; \gamma_s] = \mathrm{E}[Y_{it}|w_{it}, \widetilde{g}_{i,<t}, \overline{w}_i]$. Applying the mean value theorem to

$E[Y_{it}|w_{it},\widetilde{g}_{i,<t},\overline{w}_i;\gamma_s]$ around γ_0 yields

$$
\text{(A.17)}\quad \begin{aligned} &E[Y_{it}|w_{it},\widetilde{g}_{i,<t},\overline{w}_i;\gamma_s]-E[Y_{it}|w_{it},\widetilde{g}_{i,<t},\overline{w}_i;\gamma_0]\\ &=\sum_{\tau=1}^{d_\beta}\frac{\partial E[Y_{it}|w_{it},\widetilde{g}_{i,<t},\overline{w}_i;\gamma^*]}{\partial\beta_\tau}(\beta_{s\tau}-\beta_{0\tau})+\sum_{k=1}^{2}\frac{\partial E[Y_{it}|w_{it},\widetilde{g}_{i,<t},\overline{w}_i;\gamma^*]}{\partial\lambda_k}(\lambda_{sk}-\lambda_{0k}),\end{aligned}
$$

where $\gamma^*\equiv(\beta^*,\lambda^*)$ is a parameter between γ_s and γ_0. Combining these relationships yields

$$
\begin{aligned}
0&=\sum_{\tau=1}^{d_\beta}\frac{\partial E[Y_{it}|w_{it},\widetilde{g}_{i,<t},\overline{w}_i;\gamma^*]}{\partial\beta_\tau}\frac{(\beta_{s\tau}-\beta_{0\tau})}{\|(\beta_s,\lambda_s)-(\beta_0,\lambda_0)\|}\\
&\quad+\sum_{k=1}^{2}\frac{\partial E[Y_{it}|w_{it},\widetilde{g}_{i,<t},\overline{w}_i;\gamma^*]}{\partial\lambda_k}\frac{(\lambda_{sk}-\lambda_{0k})}{\|(\beta_s,\lambda_s)-(\beta_0,\lambda_0)\|},\\
&=\nabla_\gamma E[Y_{it}|w_{it},\widetilde{g}_{i,<t},\overline{w}_i;\gamma^*]^T\left[\frac{(\beta_s-\beta_0)'}{\|(\beta_s,\lambda_s)-(\beta_0,\lambda_0)\|}\quad\frac{(\lambda_s-\lambda_0)'}{\|(\beta_s,\lambda_s)-(\beta_0,\lambda_0)\|}\right]^T\\
&\equiv\nabla_\gamma E[Y_{it}|w_{it},\widetilde{g}_{i,<t},\overline{w}_i;\gamma^*]^T S_{\gamma_s} \qquad \text{(A.18)}
\end{aligned}
$$

Because $\|S_{\gamma_s}\|_E^2=1$ for all s, $\{S_{\gamma_s}:s=1,...\}$ is a distinct sequence on the unit sphere. This implies that there exist a convergent subsequence $\{S_{\gamma_{s_j}}:j=1,...\}$ whose limit is also on the unit sphere. Denote the limit as S_{γ_0}. Combining the continuity assumption in Assumption 2.6 and Eq. (A.18), we obtain

$$
0=\nabla_\gamma E[Y_{it}|w_{it},\widetilde{g}_{i,<t},\overline{w}_i;\gamma_0]^T S_{\gamma_0}. \qquad \text{(A.19)}
$$

Multiplying each side by $\nabla_\gamma E[Y_{it}|w_{it},\widetilde{g}_{i,<t},\overline{w}_i;\gamma_0]$ yields

$$
0=\left(\nabla_\gamma E[Y_{it}|w_{it},\widetilde{g}_{i,<t},\overline{w}_i;\gamma_0]\cdot\nabla_\gamma E[Y_{it}|w_{it},\widetilde{g}_{i,<t},\overline{w}_i;\gamma_0]^T\right)S_{\gamma_0}. \qquad \text{(A.20)}
$$

Taking an expectation, we obtain

$$
\begin{aligned}
0&=E\left[\nabla_\gamma E[Y_{it}|w_{it},\widetilde{g}_{i,<t},\overline{w}_i;\gamma_0];\gamma_0]\cdot\nabla_\gamma E[Y_{it}|w_{it},\widetilde{g}_{i,<t},\overline{w}_i;\gamma_0]^T\right]S_{\gamma_0}\\
&=I(\beta_0,\lambda_0)S_{\gamma_0}\text{ with }S_{\gamma_0}\neq 0. \qquad \text{(A.21)}
\end{aligned}
$$

Since $I(\beta_0, \lambda_0)$ is nonsingular by Assumption 2.6, we have to conclude that (β_0, λ_0) is identifiable from this contradiction. *Q.E.D.*

References

AI, C., AND X. CHEN (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71(6), 1795–1843.

——— (2007): "Estimation of Possibly Misspecified Semiparametric Conditional Moment Restriction Models with Different Conditioning Variables," *Journal of Econometrics*, 141(1), 5–43.

AKASHI, K., AND N. KUNITOMO (2012): "Some Properties of the LIML Estimator in a Dynamic Panel Structural Equation," 166, 167–183.

ALTONJI, J., AND R. MATZKIN (2005): "Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors," *Econometrica*, 73(4), 1053–1102.

ALVAREZ, J., AND M. ARELLANO (2003): "The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators," *Econometrica*, 71(4), 1121–1159.

BIØRN, E. (1992): "The Bias of Some Estimators for Panel Data Models with Measurement Errors," *Empirical Economics*, 17(1), 51–66.

BOLLINGER, C. (1998): "Measurement Error in the Current Population Survey: A Nonparametric Look," *Journal of Labor Economics*, 16(3), 576–594.

BORJAS, G. J. (2013): *Labor Economics*. McGraw-Hill/Irwin.

BOUND, J., C. BROWN, G. DUNCAN, AND W. RODGERS (1994): "Evidence on the Validity of Cross-sectional and Longitudinal Labor Market Data," *Journal of Labor Economics*, 12(3), 345–368.

BOUND, J., C. BROWN, AND N. MATHIOWETZ (2001): *Measurement Error in Survey Data*. North-Holland.

CHEN, X., AND N. SWANSON (2012): *Dynamic Nonseparable Panel Data Models, in: Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis: Essays in Honor of Halbert L. White Jr*. Springer.

CHERNOZHUKOV, V., I. FERNÁNDEZ-VAL, J. HAHN, AND W. NEWEY (2013): "Average and Quantile Effects in Nonseparable Panel Models," *Econometrica*, 81(2), 535–580.

CHERNOZHUKOV, V., I. FERNANDEZ-VAL, S. HODERLEIN, S. HOLZMANN, AND W. NEWEY (2015): "Nonparametric Identification in Panels using Quantiles," *Journal of Econometrics*, 188(2), 378–392.

EVDOKIMOV, K. (2011): "Identification and Estimation of a Nonparametric Panel Data Model with Unobserved Heterogeneity," *Working Paper*.

GRAHAM, B., AND J. POWELL (2012): "Identification and Estimation of Average Partial Effects in "Irregular" Correlated Random Coefficient Panel Data Models," *Econometrica*, 80(5), 2105–2152.

GRILICHES, Z., AND J. HAUSMAN (1986): "Errors in Variables in Panel Data," *Journal of Econometrics*, 31(1), 93–118.

HAHN, J., AND G. KUERSTEINER (2002): "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When Both n and T Are Large," *Econometrica*, 70(4), 1639–1657.

HODERLEIN, S., AND E. MAMMEN (2007): "Identification of Marginal Effects in Nonseparable Models without Monotonicity," *Econometrica*, 75(5), 1513–1518.

HODERLEIN, S., AND H. WHITE (2012): "Nonparametric Identification in Nonseparable Panel Data Models with Generalized Fixed Effects," *Journal of Econometrics*, 168(2), 300–314.

KOEBEL, B., F. LAISNEY, W. POHLMEIER, AND M. STAAT (2008): "Life Cycle Labor Supply and Panel Data: a Survey," in *The Econometrics of Panel Data*, pp. 761–794. Springer.

MACURDY, T. E. (1981): "An Empirical Model of Labor Supply in a Life-cycle Setting," *Journal of Political Economy*, 89(6), 1059–1085.

NEWEY, W., AND J. POWELL (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71(5), 1565–1578.

SCHENNACH, S. (2007): "Instrumental Variable Estimation of Nonlinear Errors-in-variables Models," *Econometrica*, 75(1), 201–239.

WANSBEEK, T. J. (2001): "GMM Estimation in Panel Data Models with Measurement Error," *Journal of Econometrics*, 104(2), 259–268.

WANSBEEK, T. J., AND R. H. KONING (1991): "Measurement Error and Panel Data," *Statistica Neerlandica*, 45(2), 85–92.

WILHELM, D. (2015): "Identification and Estimation of Nonparametric Panel Data Regressions with Measurement Error," Discussion paper, Cemmap Working Paper.

WOOLDRIDGE, J. (2005): "Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity," *Journal of Applied Econometrics*, 20(1), 39–54.

ZILIAK, J. (1997): "Efficient Estimation With Panel Data When Instruments are Predetermined: An Empirical Comparison of Moment-Condition Estimators," *Journal of Business & Economic Statistics*, 15(4), 419–431.

ZINDE-WALSH, V. (2014): "Measurement Error and Deconvolution in Spaces of Generalized Functions," *Econometric Theory*, 30(6), 1207–1246.

Table 1: Estimations of Nonlinear Panel Data Models with Measurement Error (n=500)

	$\beta_0 = 0.5$	$\beta_1 = 0.5$	$\beta_2 = -0.5$	$\lambda_1 = -0.5$	$\lambda_2 = 0.5$
Simulation I					
Mean	0.527	0.501	-0.436	-0.493	0.496
Median	0.530	0.493	-0.436	-0.497	0.493
RMSE	0.122	0.109	0.118	0.113	0.106
Simulation II					
Mean	0.533	0.505	-0.425	-0.501	0.525
Median	0.527	0.507	-0.426	-0.519	0.522
RMSE	0.119	0.110	0.105	0.100	0.122
Simulation III					
Mean	0.524	0.501	-0.436	-0.491	0.502
Median	0.521	0.501	-0.437	-0.502	0.505
RMSE	0.114	0.107	0.102	0.101	0.110

Note: Standard deviations of the parameters are computed by the standard deviation of the estimates across 150 simulations and called (simulation) standard deviations.

Table 2: Estimations of Nonlinear Panel Data Models with Measurement Error (n=1000)

	$\beta_0 = 0.5$	$\beta_1 = 0.5$	$\beta_2 = -0.5$	$\lambda_1 = -0.5$	$\lambda_2 = 0.5$
Simulation I					
Mean	0.522	0.503	-0.435	-0.491	0.532
Median	0.522	0.500	-0.434	-0.501	0.524
RMSE	0.116	0.112	0.111	0.100	0.131
Simulation II					
Mean	0.541	0.506	-0.424	-0.512	0.517
Median	0.541	0.506	-0.424	-0.517	0.509
RMSE	0.131	0.110	0.106	0.110	0.109
Simulation III					
Mean	0.525	0.510	-0.434	-0.499	0.503
Median	0.527	0.511	-0.436	-0.510	0.514
RMSE	0.117	0.118	0.101	0.106	0.109

Note: Standard deviations of the parameters are computed by the standard deviation of the estimates across 150 simulations and called (simulation) standard deviations.

Table 3: Estimation of the APEs in Simulations (n=500)

	Infeasible	Sieve MD
Simulation I:		
Mean	-0.250	-0.218
Std. dev.	0.000	0.049
RMSE	–	0.059
Simulation II:		
Mean	-0.375	-0.387
Std. dev.	0.038	0.114
RMSE	–	0.114
Simulation III:		
Mean	-1.662	-1.205
Std. dev.	0.083	0.260
RMSE	–	0.526

Note: Standard deviations of the parameters are computed by the standard deviation of the estimates across 150 simulations.

Table 4: Estimation of the APEs in Simulations (n=1000)

	Infeasible	Sieve MD
Simulation I:		
Mean	-0.250	-0.217
Std. dev.	0.000	0.045
RMSE	–	0.056
Simulation II:		
Mean	-0.375	-0.394
Std. dev.	0.025	0.125
RMSE	–	0.126
Simulation III:		
Mean	-1.662	-1.204
Std. dev.	0.060	0.230
RMSE	–	0.511

Note: Standard deviations of the parameters are computed by the standard deviation of the estimates across 150 simulations.

Table 5: Data Summary

Variable		Mean	Std. Dev.	Min	Max
ln($hours$)	overall	7.671	0.289	2.770	8.560
	between		0.233	4.950	8.407
	within		0.172	5.491	10.011
ln($wage$)	overall	2.614	0.448	-0.220	4.600
	between		0.432	0.877	4.367
	within		0.118	1.274	3.344
kids	overall	1.484	1.218	0	6
	between		1.191	0	5.333
	within		0.257	-0.183	3.150
age	overall	42.415	7.973	29	60
	between		7.933	30	59
	within		0.849	40.748	44.081
age^2	overall	1,862.545	708.068	841	3,600
	between		704.740	900.667	3,481.667
	within		72.973	1,668.212	2,051.545
disab	overall	0.083	0.276	0	1
	between		0.230	0	1
	within		0.153	-0.583	0.750

Note: The data is a three-periods of panel data with a cross-sectional size 532.

Table 6: Estimates for the Elasticity of Labor Supply

	Dependent Variable: ln($hours$)			
	Linear Fixed Effect	First Differencing IV	Linear Corr. Random Effects	Semi-parametric Nonlinear Reg.
ln($wage$)	-0.094	0.049	0.041	0.033
	(0.045)	(0.081)	(0.021)	(0.020)
kids	-0.019	0.000	-0.015	-0.011
	(0.021)	(0.021)	(0.021)	(0.013)
age	-0.008	0.019	-0.011	-0.009
	(0.043)	(0.072)	(0.043)	(0.012)
age^2	0.000	0.000	0.000	-0.002
	(0.000)	(0.001)	(0.000)	(0.003)
disab	-0.042	-0.063	-0.048	-0.040
	(0.035)	(0.091)	(0.035)	(0.044)
time trend	0.002	0.000	0.002	0.003
	(0.029)	(0.004)	(0.029)	(0.002)
$\overline{kids}$	–	–	0.018	0.026
	–	–	(0.024)	(0.014)
$\overline{age}$	–	–	0.016	0.014
	–	–	(0.046)	(0.013)
$\overline{age}^2$	–	–	0.000	0.002
	–	–	(0.000)	(0.004)
$\overline{disab}$	–	–	-0.109	-0.084
	–	–	(0.056)	(0.162)
constant	7.918	–	7.523	8.435
	1.314	–	(0.325)	(5.723)
APE	-0.094	0.049	0.041	0.097
	(0.045)	(0.081)	(0.021)	(0.020)

Note: Bootstrap (simulation) standard errors are reported in parentheses, using 150 bootstrap replications for the semi-parametric nonlinear regression model. The linear fixed effects model, the first differencing IV model and the linear correlated random effects model are the proposed model without the $(1+c_i)$ multiplying wage. The linear correlated random effects model is estimated using the CRE specification in Assumption 2.1. While the APE of labor supply for the linear models is β_1, the APE for the non-linear model is $\beta_1 \int_{\mathscr{C}} (1+c) f_{C_i}(c) dc$.

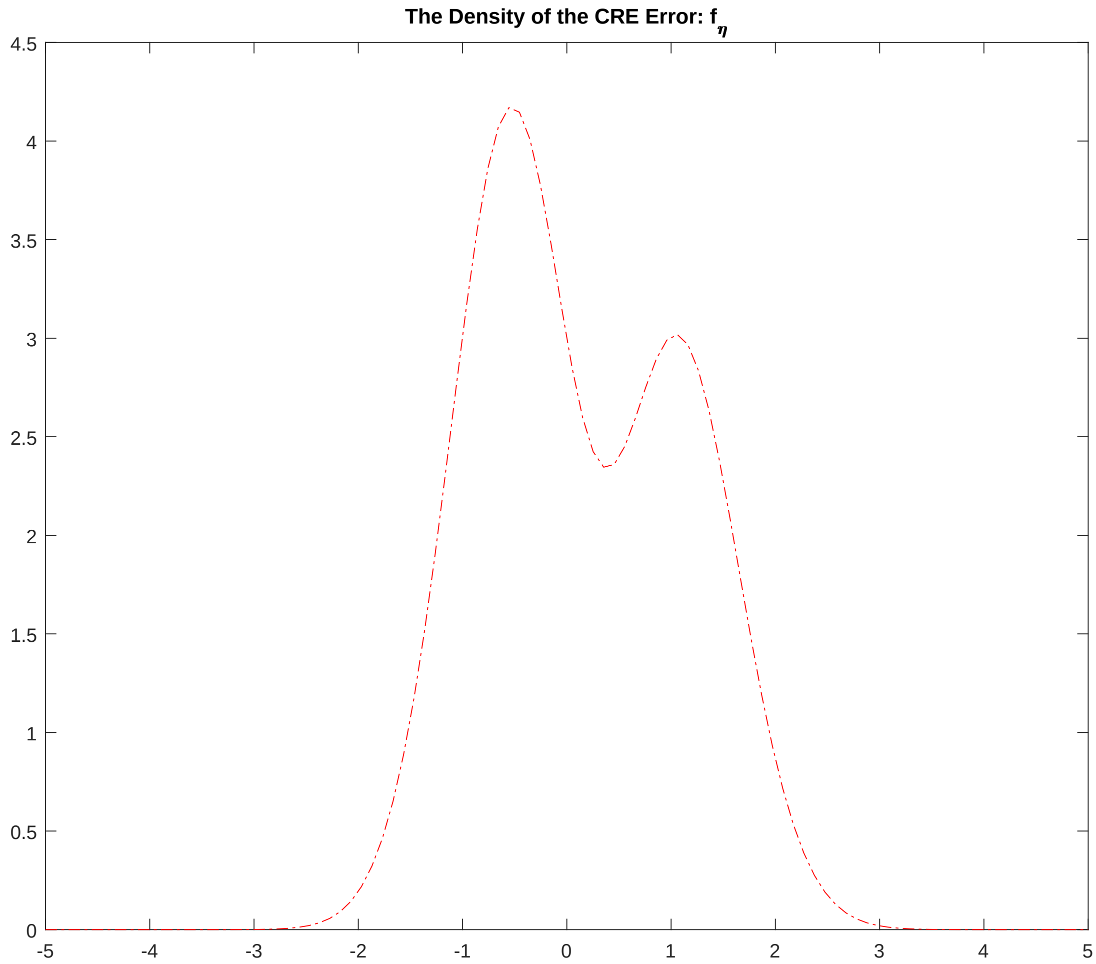

Figure 1: The Estimated Density of the Error in the CRE Specification