Semiparametric nonlinear panel data models with measurement error

Oliver Linton
Ji-Liang Shiu

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP09/18
Semiparametric Nonlinear Panel Data Models with Measurement Error

Oliver Linton*  Ji-Liang Shiu†

January 10, 2018

Abstract

This paper develops the identification and estimation of nonlinear semi-parametric panel data models with mismeasured variables and their 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 then estimate the model via the semi-parametric sieve Generalized Method of Moments 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.

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 in the model. Substantial progress has been made to handle linear or nonlinear models ignoring the potential presence of measurement error. However, many economic quantities such as work hours, earnings, fringe benefits and employment in surveys are frequently measured with errors, if longitudinal information is collected through one-time retrospective surveys. This concern has been heightened by the increased use of longitudinal data sets and mismeasurement of the panel data may lead to false results or obscures true economic relationships. The estimation problems caused by the mismeasurement of economic data may be greatly exacerbated when economists exploit panel data to control for the effects of unobserved individual effect using standard fixed effects or first-differenced estimators.

Consider the following semi-parametric nonlinear panel data model with unknown finite-dimensional parameter $\beta_0$

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

In this model, $Y_{it}$ is an observed scalar dependent variable, $W_{it}$ is a perfectly observed explanatory variable, $X^*_{it}$ is a latent continuous mismeasured variable, $C_i$ is an unobserved individual-specific effect, and $U_{it}$ is an unobserved random variable. The function $m$ may be inseparable in $W_{it}$, $X^*_{it}$, and $C_i$, and 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 which are distinct in the literature.

---

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.
of panel data models with measurement errors. First, the unobserved heterogeneity enters the structural regression function nonseparably without imposing a linear index structure. Second, the potential nonlinear regression function also contains a mismeasured variable nonseparably along with other explanatory variables. This suggests that the proposed regression model can be a structural function derived from a dynamic utility maximization problem with flexible preferences.

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 involve 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 models belong to 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 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 so there is a fundamental difference between additively separable models and nonseparable models.

Besides short panel data considered here, there are a lot of closely related work in the existing 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 $T$ approach to infinity and the ratio $n/T$ approach to a constant.

The identification technique developed in this paper 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 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. 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.

The estimation method closely follows the construction of the identification analysis because the identification result is established from knowledge of the three conditional expectations. Based on this identification result, 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 (2003) and Newey and Powell (2003). The estimation procedure consists of applying the SMD method to a vector of the moment conditions 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
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 sieve minimum distance (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 a scalar and a multivariate case can be straightforwardly extended. 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 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. \textit{(Correlated Random Effects (CRE))} There exists a nonzero coefficient $\lambda_0$ such that

\begin{equation}
C_i = \lambda_0 \bar{W}_i + \eta_i,
\end{equation}

where $\bar{W}_i = \frac{1}{T} \sum_{t=1}^{T} W_{it}$ is denoted as the time average of the perfectly observed explanatory variables. In particular, the remainder term $\eta_i$ is independent of $\bar{W}_i$.

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

\begin{equation}
C_i = \lambda_{01} \bar{W}_i + \lambda_{02} \bar{Z}_i + \eta_i.
\end{equation}
Including more control variables in the specification may make the independent assumption of the projection error $\eta_i$ more reasonable.

**Assumption 2.2.** (Classical measurement error) Assume

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

(ii) (Measurement error)

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

(iii) (Conditional mean independence)

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

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

The setting for the measurement errors is the same as Schennach (2007), which uses instruments to identify nonlinear errors-in-variables models. Assumption 2.2(i) can be regarded as a control function assumption which uses the past variable as IV to construct the estimable $h_t(G_{i<t})$ to extract the independent unobservable variable $V_{it}$ from the unobservable true regressor $X^*_{it}$ affecting the response. The assumption is commonly used for identification of nonlinear models. We can further assume $X^*_{it}$ follows a first order stationary Markov motion by setting $X^*_{it} = h(W_{it-1}, X_{it-1}) + V_{it}$. Assumption 2.2(ii) implies that $E[X^*_{it} \Delta X_{it}] = 0$ or there is no correlation between the estimated residual from regressing $X_{it}$ on $h_t(G_{i<t})$.

\[2\] Combining Assumption 2.2(i) and (ii) yields $X_{it} = h_t(G_{i<t}) + V_{it} + \Delta X_{it}$. 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 $\Delta X_{it}$ in Assumption 2.2(ii) can be conducted by testing the dependence of the estimated residual from regressing $X_{it}$ on $h_t(G_{i<t})$. 
unobserved true regressor and the measurement error. Assumption 2.2(iii) only imposes the standard orthogonality restriction that \( E[U_{it}|W_{it}, G_{it}<t, V_{it}, \bar{W}_i] = 0 \) and suggests that the disturbance \( U_{it} \) does not have to be independent of \( W_{it}, G_{it}<t, V_{it}, \) and \( \bar{W}_i \) 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.

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

\[
X_{it}^* = \tilde{G}_{it} - \bar{V}_{it}, \quad \text{and} \quad C_i = \lambda_0 \bar{W}_i - \bar{\eta}_i,
\]

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

**Assumption 2.3.** Consider \( E[Y_{it}|w_{it}, \tilde{G}_{it}<t, \bar{W}_i], E[X_{it}Y_{it}|w_{it}, \tilde{G}_{it}<t, \bar{W}_i] \) for a fixed \( w_{it}. \) These conditional expectations are functions in \( \mathbb{R}^2 \) and belong to a function space \( \mathcal{S} \) which contains functions \( f(\xi) \) satisfying

\[
\int (1 + \xi^\top \xi)^\gamma |f(\xi)| d\xi < \infty, \text{ for some } \gamma > 0.
\]

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 characteristic functions of the conditional expectations \( E[Y_{it}|w_{it}, \tilde{G}_{it}<t, \bar{W}_i], \) \( E[X_{it}Y_{it}|w_{it}, \tilde{G}_{it}<t, \bar{W}_i], \) and \( m \left( w_{it}, x_{it}^*, c_i; \beta_0 \right) \) for a fixed \( w_{it} \) as follows:

\[
\begin{align*}
\varphi_y(w_{it}, \xi_1, \xi_2) &= \int \int E[Y_{it}|w_{it}, \tilde{G}_{it}<t, \bar{W}_i] e^{i \xi_1 \tilde{G}_{it}<t} e^{i \xi_2 \bar{W}_i} d\tilde{G}_{it} < t d\bar{W}_i \\
\varphi_{xy}(w_{it}, \xi_1, \xi_2) &= \int \int E[X_{it}Y_{it}|w_{it}, \tilde{G}_{it}<t, \bar{W}_i] e^{i \xi_1 \tilde{G}_{it}<t} e^{i \xi_2 \bar{W}_i} d\tilde{G}_{it} < t d\bar{W}_i \\
\varphi_m(w_{it}, \xi_1, \xi_2; \beta_0) &= \int \int m \left( w_{it}, x_{it}^*, c_i; \beta_0 \right) e^{i \xi_1 \tilde{G}_{it}<t} e^{i \xi_2 \bar{W}_i} d\tilde{G}_{it} < t d\bar{W}_i,
\end{align*}
\]

where \( i = \sqrt{-1}. \) Define also \( \phi_{\nu}(\xi_1) = \int e^{i \xi_1 \tilde{G}_{it}} f_{\tilde{Y}_{it}}(\tilde{G}_{it}) d\tilde{G}_{it} \) and \( \phi_{\bar{\eta}}(\xi_2) = \int e^{i \xi_2 \bar{W}_i} f_{\bar{\eta}_{it}}(\bar{W}_i) d\bar{\eta}_{it}, \)
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,

$$
\mathcal{F}_y(w_{it}, \xi_1, \xi_2) = \frac{1}{\lambda_0} \mathcal{F}_m(w_{it}, \xi_1, \frac{\xi_2}{\lambda_0}) \phi_v(\xi_1) \phi_\eta(\frac{\xi_2}{\lambda_0}),
$$

$$
\mathcal{F}_{xy}(w_{it}, \xi_1, \xi_2) = \frac{1}{\lambda_0} - i \frac{\partial \mathcal{F}_m(w_{it}, \xi_1, \frac{\xi_2}{\lambda_0})}{\partial \xi_1} \phi_v(\xi_1) \phi_\eta(\frac{\xi_2}{\lambda_0}).
$$

**Proof.** See the appendix.

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

**Assumption 2.5.** Set $\Theta$ as a parameter space containing $\beta_0$. There exists a finite or infinite constant $\bar{\xi} > 0$ and some $w_{it}$ such that for all $\beta \in \Theta$ (i) $\mathcal{F}_m(w_{it}, \xi_1, \xi_2; \beta) \neq 0$ almost everywhere in $[-\bar{\xi}, \bar{\xi}]^2$ and (ii) $\mathcal{F}_m(w_{it}, \xi_1, \xi_2; \beta) = 0$ for all $|\xi_1|, |\xi_2| > \bar{\xi}$.

Assumptions [2.4] and [2.5] are standard in the deconvolution literature. Assumption [2.4(ii)] requires that the characteristic functions of $V$ and $\bar{\eta}$ to be non-vanishing which excludes uniform or triangular distributions. Exploiting the conditional mean function in Eq. [A.5] by replacing $f_{\bar{\eta}_i}(\bar{\eta}_i)$ by $f_{\bar{\eta}_i;\gamma}(\bar{\eta})$, we have the following.

Denote $\gamma = (\beta, \lambda)$ and $\gamma$ is a $(d_\beta + 2) \times 1$-dimensional vector. Consider the parametric conditional mean function in Eq. [A.16]:

$$
E[Y_{it}|w_{it}, \bar{g}_{i< t}, \bar{\omega}_i; \gamma] = \int \int m(w_{it}, \bar{g}_{i< t} - \bar{v}_{it}, \lambda_1 \bar{w}_i - \bar{\eta}_i; \beta) f_{\tilde{V}_{it}}(\tilde{v}_{it}) f_{\bar{\eta}_i;\gamma}(\bar{\eta}_i)d\tilde{v}_{it}d\bar{\eta}_i.
$$

Define the gradient of $E[Y_{it}|w_{it}, \bar{g}_{i< t}, \bar{\omega}_i; \gamma]$ as follows,

$$
\nabla_\gamma E[Y_{it}|w_{it}, \bar{g}_{i< t}, \bar{\omega}_i; \gamma] = \left( \frac{\partial E[Y_{it}|w_{it}, \bar{g}_{i< t}, \bar{\omega}_i; \gamma]}{\partial \beta_1}, \ldots, \frac{\partial E[Y_{it}|w_{it}, \bar{g}_{i< t}, \bar{\omega}_i; \gamma]}{\partial \lambda_2} \right)^\top.
$$
Define the information matrix as follows:

\[ I(\gamma) = E \left[ \nabla_{\gamma} E[Y_{it}|w_{it}, \bar{g}_{i<t}, \bar{w}_i; \gamma] \cdot \nabla_{\gamma} E[Y_{it}|w_{it}, \bar{g}_{i<t}, \bar{w}_i; \gamma] \right]. \]

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

**Theorem 2.1.** Under Assumptions 2.1, 2.2, 2.3, 2.4, 2.5, and 2.6, the three unknown parameters of interest, including the finite-dimensional parameters \( \beta_0 \) and \( \lambda_0 \), the distribution of the remainder error of control function approach \( f_{\tilde{V}_{it}}(\bar{v}) \), and the distribution of the remainder error of CRE \( \eta_i, f_{\tilde{\eta}_i}(\bar{\eta}) \) 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 CRE specification and the properties of Fourier transforms on convolution functions to connect the distribution of 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.

Other quantity of interest is on estimating partial effects. 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 which is also identified by Theorem 2.1. With the identification of the distribution of \( \eta_i \) and the independence assumption of \( \eta_i \) in Assumption 2.1, we have

\[ f(c|\bar{w}_i) = f_{\tilde{\eta}_i}(-c + \lambda_0 \bar{w}_i). \]

Then, the distribution of the individual effect can be identified.
with the identification of \( f(c|\bar{w}_i) \) from the following equation:

\[
(8) \quad f_{C_i}(c) = \int f(c|\bar{w}_i) \cdot \underbrace{f(\bar{w}_i)}_{\text{estimable from data}} \, d\bar{w}_i.
\]

Suppose \( x^*_i \) takes continuous values. Given \((w_0, x^*_0)\), the average partial effect (APE) for \( x^*_i \) at the point is defined as

\[
(9) \quad \text{APE}(w_0, x^*_0) = \int \frac{\partial m(w_{it}, x^*_it, c_i; \beta_0)}{\partial x^*_it} \bigg|_{(w_{it}, x^*_it) = (w_0, x^*_0)} f_{C_i}(c) \, dc.
\]

**Corollary 2.1.** Under Assumptions 2.7, 2.2, 2.3, 2.4, 2.5, and 2.6 the distribution of the individual effect and the average partial effect defined in Eq. (9) is identified.

### 3. SMD Estimation

In Section 2, we have shown in Theorem 2.1, the three unknown parameters of interest, including the finite-dimensional parameters \( \beta_0 \) and \( \lambda_0 \), the distribution of the remainder error of control function approach \( f_{\tilde{V}_{it}}(\tilde{v}) \), and the distribution of the remainder error of CRE \( \eta_i, f_{\tilde{\eta}_i}(\tilde{\eta}) \) are uniquely identified. The identification is based on knowledge of the three observable conditional expectations \( \mathbb{E}[X_{it}|G_{i<t}], \mathbb{E}[Y_{it}|W_{it}, \tilde{G}_{i<t}, \bar{W}_i] \) and \( \mathbb{E}[X_{it}Y_{it}|W_{it}, \tilde{G}_{i<t}, \bar{W}_i] \), where \( \tilde{G}_{i<t} = h_t(G_{i<t}) \). 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 \( \tilde{H}_{i<t} = h(W_{it-1}, X_{it-1}), \) and \( D_{it} = (W_{it}, W_{it-1}, X_{it-1}, \bar{W}_i) \). Under the assumptions of Theorem 2.1 and Assumption 3.1, we rewrite these conditional expectations as
follows\textsuperscript{3}

\[ 0 \equiv E[X_{it}|W_{it-1},X_{it-1}] - h(W_{it-1},X_{it-1}), \]
\[ 0 \equiv E[Y_{it}|D_{it}] - \int \int m(W_{it},\bar{H}_{i<it} - \bar{v}_{it},\lambda_0 \bar{W}_i - \bar{\eta}_i;\beta_0) f_{\bar{v}_{it}}(\bar{v}_{it}) f_{\bar{\eta}_i}(\bar{\eta}_i) d\bar{v}_{it} d\bar{\eta}_i, \]
\[ 0 \equiv E[X_{it}Y_{it}|D_{it}] - \int \int (\bar{H}_{i<it} - \bar{v}_{it}) m(W_{it},\bar{H}_{i<it} - \bar{v}_{it},\lambda_0 \bar{W}_i - \bar{\eta}_i;\beta_0) \]
\[ \times f_{\bar{v}_{it}}(\bar{v}_{it}) f_{\bar{\eta}_i}(\bar{\eta}_i) d\bar{v}_{it} d\bar{\eta}_i. \]

Denote \( \alpha_0 = (\beta_0, \lambda_0, f_{V_{it}}(\cdot), f_{\eta_i}(\cdot), h(\cdot)) \). Define the following residual functions:

\[ \rho_1(X_{it},Y_{it},D_{it};\alpha_0) \equiv X_{it} - h(W_{it-1},X_{it-1}), \]
\[ \rho_2(X_{it},Y_{it},D_{it};\alpha_0) \equiv Y_{it} - \int \int m(W_{it},\bar{H}_{i<it} - \bar{v}_{it},\lambda_0 \bar{W}_i - \bar{\eta}_i;\beta_0) f_{\bar{v}_{it}}(\bar{v}_{it}) f_{\bar{\eta}_i}(\bar{\eta}_i) d\bar{v}_{it} d\bar{\eta}_i, \]
\[ \rho_3(X_{it},Y_{it},D_{it};\alpha_0) \equiv X_{it}Y_{it} - \int \int (\bar{H}_{i<it} - \bar{v}_{it}) m(W_{it},\bar{H}_{i<it} - \bar{v}_{it},\lambda_0 \bar{W}_i - \bar{\eta}_i;\beta_0) \]
\[ \times f_{\bar{v}_{it}}(\bar{v}_{it}) f_{\bar{\eta}_i}(\bar{\eta}_i) d\bar{v}_{it} d\bar{\eta}_i. \]

Define the 3 \times 1 vector of residual functions by

\[ \rho(X_{it},Y_{it},D_{it};\alpha_0) = \begin{pmatrix} 
\rho_1(X_{it},Y_{it},D_{it};\alpha_0) \\
\rho_2(X_{it},Y_{it},D_{it};\alpha_0) \\
\rho_3(X_{it},Y_{it},D_{it};\alpha_0)
\end{pmatrix}. \]

The parameter vector \( \alpha = (\beta, \lambda, f_{V}(\cdot), f_{\eta}(\cdot), h(\cdot)) \) 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 \( \alpha_0 \) can be summarized as the following conditional moment restrictions

\[ m(D_{it};\alpha) \equiv E[\rho(X_{it},Y_{it},D_{it};\alpha)|D_{it}], \]

with \( m(D_{it};\alpha) = 0 \). Therefore, the model fits into the general models of conditional

\textsuperscript{3}The detailed derivations can be found in Eqs. (A.5) and (A.6) in the appendix.
moment restrictions in Ai and Chen (2003), which contain finite dimensional unknown parameters and infinite dimensional unknown functions.

Suppose \( \hat{m}(D_{it}; \alpha) \) is a consistent estimator for \( m(D_{it}; \alpha) \) and \( A_n \) is a sequence of approximating sieve spaces for the parameter space \( \mathcal{A} \) containing \( \alpha_0 \). The SMD estimator \( \hat{\alpha}_n \) minimizes the following sample analog of a minimum distance objective function with the parameters restricted to the sieve spaces, \( A_n \):

\[
\hat{\alpha}_n = \arg \min_{\alpha \in A_n} \frac{1}{n(T-1)} \sum_{i=1}^{n} \sum_{t=2}^{T} \hat{m}(D_{it}; \alpha) \top \left[ \hat{\Sigma}(D_{it}) \right]^{-1} \hat{m}(D_{it}; \alpha),
\]

where \( \hat{\Sigma}(D_{it}) \) is some positive \( 3 \times 3 \) weighting matrix. There are two approximations in the optimization problem to make the estimator feasible and consistent. One is \( \hat{m}(D_{it}; \alpha) \) approximates \( m(D_{it}; \alpha) \) and the other is \( A_n \) approximates \( \mathcal{A} \).

Let \( p^k(\cdot) = (p_1(\cdot), \ldots, p_k(\cdot))^\top \) be a vector of some known univariate basis function and \( p^k(\cdot, \ldots, \cdot) = (p_1(\cdot, \ldots, \cdot), \ldots, p_k(\cdot, \ldots, \cdot))^\top \) be multivariate basis function generated by tensor product construction. Denote the \( k_n \times 1 \) vector of approximating functions as \( p^{k_n}(D_{it}) = (p_1(D_{it}), \ldots, p_{k_n}(D_{it}))^\top \) which is constructed from some known basis functions for any square integrable real-valued function of \( D_{it} \). A linear consistent sieve estimator \( \hat{m}(D_{it}; \alpha) \) can be obtained by regressing \( \rho(X_{it}, Y_{it}, D_{it}; \alpha) \) on \( p^{k_n}(D_{it}) \),

\[
\hat{m}(D_{it}; \alpha) = p^{k_n}(D_{it}) \top (H \top H)^{-1} \sum_{i=1}^{n} \sum_{t=2}^{T} p^{k_n}(D_{it}) \rho(X_{it}, Y_{it}, D_{it}; \alpha),
\]

where \( H = (p^{k_n}(D_{12}), \ldots, p^{k_n}(D_{nT}))^\top \). This GMM type estimator is proposed by Ai and Chen (2003) and is called a sieve minimum distance (hereafter SMD) estimator. Ai and Chen (2003) show that the SMD 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 derived in Section 3 by a Monte Carlo simulation. We focus on the estimation of $\beta_0$ and $\lambda_0$ which correspond to the regression function $m(W_{it}, X_{it}^{*}, C_i; \beta_0)$ and the CRE $C_i = \lambda_{01} \bar{W}_i + \lambda_{02} \bar{Z}_i + \eta_i$, respectively. However, the distributions of $f_{\tilde{\eta}_i}(\tilde{v})$ and $f_{\tilde{\eta}_i}(\tilde{\eta})$ are treated nonparametrically and will be approximated by a sequence of truncated sieve series.

The simulation design is according to the following DGP: Denote Trun($\Phi,[a,b]$) as a distribution of a random variable generated by $\Phi^{-1}(u \cdot (\Phi(b) - \Phi(a)) + \Phi(a))$ where $\Phi$ is the CDF of standard normal distribution, $\Phi^{-1}$ is the inverse of $\Phi$ 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

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

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

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

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

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

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

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

Simulation I: \( m(W_{it}, X_{it}^*, C_i; \beta_0) = \beta_{00} + \beta_{01}W_{it} + \beta_{02}X_{it}^{*2} + C_i, \)

Simulation II: \( m(W_{it}, X_{it}^*, C_i; \beta_0) = (\beta_{00} + \beta_{01}W_{it} + \beta_{02}X_{it}^{*2} + C_i)^2, \)

Simulation III: \( m(W_{it}, X_{it}^*, C_i; \beta_0) = (\beta_{00} + \beta_{01}(1 + C_i)W_{it} + \beta_{02}(1 + C_i)X_{it}^{*} + C_i)^2. \)

The SMD procedure requires approximating the three nonparametric parts by sieves, including the conditional expectation function \( h_t, f_{\tilde{W}_t}(\tilde{v}) \) and \( f_{\tilde{\eta}}(\tilde{\eta}) \). Let \( f_1 \) and \( f_2 \) be the nonparametric series estimators for \( f_{\tilde{W}_t}(\tilde{v}) \) and \( f_{\tilde{\eta}}(\tilde{\eta}) \), respectively. We construct \( f_1^{1/2} \) and \( f_2^{1/2} \) by univariate Hermite functions,

\[
\begin{align*}
    f_1^{1/2}(\tilde{v}) & = \sum_{i=0}^{3} \delta_{1i} H_i(\tilde{v}), \\
    f_2^{1/2}(\tilde{\eta}) & = \sum_{i=0}^{3} \delta_{2i} H_i(\tilde{\eta}),
\end{align*}
\]

where \( H_0(x) = e^{-x^2/2}, H_1(x) = xe^{-x^2/2}, H_2(x) = (x^2 - 1)e^{-x^2/2}, H_3(x) = (x^3 - 3x)e^{-x^2/2} \). 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 \( \sqrt{2\pi}(\delta_{10}^2 + \delta_{11}^2 + 2!\delta_{12}^2 + 3!\delta_{13}^2) = 1 \).

We use a tensor product polynomial sieve to approximate the conditional mean function, which is the set of instruments. In other words, each argument of \( p^{k_n}(D_{it}) \) is in the following set: \( \{1, W_{it}, W_{it-1}, X_{it-1}, \bar{W}_i, \bar{Z}_i, W_{it}^2, W_{it}W_{it-1}, W_{it}X_{it-1}, W_{it}\bar{W}_i, W_{it}\bar{Z}_i, W_{it-1}^2, W_{it-1}X_{it-1}, W_{it-1}\bar{W}_i, W_{it-1}\bar{Z}_i, X_{it-1}^2, X_{it-1}\bar{W}_i, X_{it-1}\bar{Z}_i, \bar{W}_i^2, \bar{Z}_i^2 \} \) and the total number of the instruments is 21. As an illustration, we use the identity weighting, \( \tilde{\Sigma}(\tilde{D}_{it}) = I \) for the SMD estimator.

The 200 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 and 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. (8) with the estimated coefficient of $\lambda$ and observation of $\bar{w}_i$ to recover the distribution of the individual effect $f_{C_i}(\cdot)$ and then APEs can be calculated by Eq. (9). Tables 3-3 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 Application

In this section, we apply our proposed nonlinear panel data model to investigate the effect of the hourly wage rate of individuals on labor supply given their demographic variables. The dependent variables are the log values of annual hours of work for those with positive working hours. The variable of interest is the hourly wage rate and measurement error may be greater for the hourly wage rate in the survey. Quality of the variable is a critical issue for studies of labor supply. The proposed empirical nonlinear panel data model can examine the measurement error of the hourly wage rate and provides consistent estimate of the effect. The panel data model fits to this labor supply topic naturally. In the panel data setting, our model uses the correlated random effect to control unobserved time invariant factors such as individual unobserved skill level, ability, or motivation factors which may be correlated to the hourly wage rate. The data format we used is from Ziliak (1997). Table 5 presents summary statistics for the working hours, the hourly wage rate, and socioeconomic variables. The between observations with zero working hours. The logarithmic transformation is well defined and still effectively capture the movement of working hours. Borjas (2009) reviews the literature on the estimation of the labor supply elasticity and also discusses the problems caused by measurement error.
and within sample standard deviations are 0.233 and 0.172 for \( \ln(\text{hours}) \) and 0.432 and 0.118 for \( \ln(\text{wage}) \), respectively. We have a three-periods of the panel data with a cross-sectional size 532 of males.

Consider an empirical model for labor supply elasticity:

\[
\ln(\text{hours}_{it}) = \beta_1(1 + c_i)\ln(\text{wage}_{it}) + \beta_2 \text{kids}_{it} + \beta_3 \text{age}_{it} + \beta_4 \text{age}_{it}^2 + \beta_5 \text{disab}_{it} + c_i + u_{it}.
\]

This specification allows the interactions between observables and unobservables. That is the random coefficient term of \( \beta_1(1 + c_i) \). In this empirical example, we can treat \( c_i \) as unmeasured ability or motivation factors that affect hours of working and \( u_{it} \) as a time-varying macro shock for labor market. Because the true wage rate of each individual is subject to a misreporting error, the measurement error of the variable \( \ln(\text{wage}_{it}) \) is likely to occur. The vector of time-varying covariates is \((\text{kids}_{it}, \text{age}_{it}, \text{age}_{it}^2, \text{disab}_{it})\) and the time averages of these variables are used in the CRE specification in this estimation of labor supply elasticity. A theoretic model of labor supply implies that there are two effects of a wage increase on labor supply, one is income effect and the other is 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.

Table 6 reports the estimates obtained with our sieve GMM method and with the linear correlated random effect estimates. We find that the estimated coefficients for the elasticity are not much different to both models. The values of the coefficients in these estimates are 4.1\%, and 3.9\%. 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 estimate in the linear correlated random effect model. A 1\% increase in wage exhibits an approximately 9\% increase in working hours. Given the flexible nature of our estimation approach, the difference implies that the estimate in the linear correlated random effect model might be biased downward when the measurement

\[\text{See detailed discussion in Bound, Brown, and Mathiowetz (2001).}\]
error problem is not accounted for. As for the sign of the labor supply elasticity, both estimates are positive 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.

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 the models without external information such as a validation or replicate data set. This study was motivated by a richer structure of panel data. It is shown that using the past observables as instruments to permit identification of nonlinear regression models in the presence of measurement error and also applying the correlated random effects specification to control the unobserved individual heterogeneity.

The identification equation is a system of three functional equations that relate conditional expectations of observed variables to the regression function of interest and distributions of unobservables. The identification strategy contains two steps. While in the first step we use the method of Schennach (2007) to identify the distribution of measurement error, in the second step we use CRE specification and the properties of Fourier transforms on convolution functions to connect the distribution of 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. Using these conditional expectations of observed variables for identification conditions, this study provides a semi-parametric sieve-based GMM estimator and shows that this estimator is consistent and asymptotically normal. Simulation experiments show that the sieve GMM estimators perform well for both linear and nonlinear panel models with measurement errors. We illustrate the performance of this estimator by estimating the elasticity of labor supply and find that the substitution effect is stronger than the income effect and a 1% increase in
wage enhances an approximately 9% 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 $C_i = \lambda_0 \bar{W}_i + \eta_i$. Combining Assumptions 2.2(i) and (ii) yields

(A.1) $\quad X_{it} = h_t(G_{i < t}) + V_{it} + \Delta X_{it}.$

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

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

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

(A.3) $\quad X_{it}^* = \tilde{G}_{i < t} - \tilde{V}_{it},$ and $C_i = \lambda_0 \bar{W}_i - \tilde{\eta}_i.$

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

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

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

(A.5) $\quad E[Y_{it}|w_{it}, \tilde{g}_{i < t}, \bar{w}_i]$

$\quad = \int \int m \left( w_{it}, \tilde{g}_{i < t} - \tilde{v}_{it}, \lambda_0 \bar{w}_i - \tilde{\eta}_i; \beta_0 \right) f_{\tilde{v}_{it}}(\tilde{v}_{it}) f_{\tilde{\eta}_i}(\tilde{\eta}_i) d\tilde{v}_{it} d\tilde{\eta}_i.$
Expanding out the term $X_{it}Y_{it}$ and taking conditional expectation with respect to $(w_{it}, \tilde{g}_{i<t}, \tilde{W}_i)$ results in

$$E[X_{it}Y_{it}|w_{it}, \tilde{g}_{i<t}, \tilde{W}_i]$$

$$= E[(\tilde{G}_{i<t} - \tilde{V}_{it})m\left(W_{it}, \tilde{G}_{i<t} - \tilde{V}_{it}, \lambda_0 \tilde{W}_i - \tilde{\eta}_i; \beta_0 \right) | w_{it}, \tilde{g}_{i<t}, \tilde{W}_i]$$

$$+ E[\Delta X_{it}m\left(W_{it}, \tilde{G}_{i<t} - \tilde{V}_{it}, \lambda_0 \tilde{W}_i - \tilde{\eta}_i; \beta_0 \right) | w_{it}, \tilde{g}_{i<t}, \tilde{W}_i]$$

$$+ E[(\tilde{G}_{i<t} - \tilde{V}_{it})U_{it}|w_{it}, \tilde{g}_{i<t}, \tilde{W}_i] + E[\Delta X_{it}U_{it}|w_{it}, \tilde{g}_{i<t}, \tilde{W}_i]$$

$$= E[(\tilde{G}_{i<t} - \tilde{V}_{it})m\left(W_{it}, \tilde{G}_{i<t} - \tilde{V}_{it}, \lambda_0 \tilde{W}_i - \tilde{\eta}_i; \beta_0 \right) | w_{it}, \tilde{g}_{i<t}, \tilde{W}_i]$$

(A.6)

$$= \int \int (\tilde{g}_{i<t} - \tilde{v}_{it})m\left(w_{it}, \tilde{g}_{i<t} - \tilde{v}_{it}, \lambda_0 \tilde{w}_i - \tilde{\eta}_i; \beta_0 \right) f_{\tilde{V}_{it}}(\tilde{v}_{it}) f_{\tilde{\eta}_i}(\tilde{\eta}_i) d\tilde{v}_{it} d\tilde{\eta}_i.$$

where we have used the zero conditional mean of $\Delta X_{it}$ 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 $\tilde{G}_{i<t}$, and $\tilde{W}_i$, we have

$$F_{\gamma}(w_{it}, \xi_1, \xi_2)$$

$$= \int \int E[Y_{it}|w_{it}, \tilde{g}_{i<t}, \tilde{W}_i] e^{ik_1 \tilde{g}_{i<t} e^{ik_2 \tilde{W}_i} d\tilde{g}_{i<t} d\tilde{W}_i}$$

$$= \int \int \left( \int \int m\left(w_{it}, \tilde{g}_{i<t} - \tilde{v}_{it}, \lambda_0 \tilde{w}_i - \tilde{\eta}_i; \beta_0 \right) f_{\tilde{V}_{it}}(\tilde{v}_{it}) f_{\tilde{\eta}_i}(\tilde{\eta}_i) d\tilde{v}_{it} d\tilde{\eta}_i \right) e^{ik_1 \tilde{g}_{i<t} e^{ik_2 \tilde{W}_i} d\tilde{g}_{i<t} d\tilde{W}_i}$$

$$= \frac{1}{\lambda_0} \left( \int \int m\left(w_{it}, x_{it}^*, c_i; \beta_0 \right) e^{ik_1 \tilde{x}_{it}^*} e^{ik_2 \xi_0} dx_{it}^* d\xi_2 \right) \left( \int e^{ik_1 \tilde{v}_{it}} f_{\tilde{V}_{it}}(\tilde{v}_{it}) d\tilde{v}_{it} \right) \left( \int e^{ik_2 \xi_0} f_{\tilde{\eta}_i}(\tilde{\eta}_i) d\tilde{\eta}_i \right)$$

$$= \frac{1}{\lambda_0} F_m(w_{it}, \xi_1, \xi_2; \phi_0(\xi_1) \phi_\eta(\xi_2; \lambda_0), \xi_2)$$

19
\[ F_{xy}(w_{it}, \xi_1, \xi_2) \]
\[ = \int \int E[X_{it}Y_{it}|w_{it}, \bar{g}_{i<1}, \bar{w}_i]e^{i\xi_1 \bar{g}_{i<1}}e^{i\xi_2 \bar{w}_i}d\bar{g}_{i<1}d\bar{w}_i \]
\[ = \int \int (\int (\bar{g}_{i<1} - \bar{v}_{it})m(w_{it}, \bar{g}_{i<1} - \bar{v}_{it}, \lambda_0 \bar{w}_i - \bar{\eta}_i; \beta_0) f_{\bar{v}_{it}}(\bar{v}_{it})f_{\bar{\eta}_i}(\bar{\eta}_i)d\bar{v}_{it}d\bar{\eta}_i) \int e^{i\xi_1 \bar{g}_{i<1}}e^{i\xi_2 \bar{w}_i}d\bar{g}_{i<1}d\bar{w}_i \]
\[ = \frac{1}{\lambda_0} \int \int \int x_{it}e^{i\xi_1 x_{it}e^{i\xi_2 x_{it}}}dx_{it}dc_i \int \int e^{i\xi_1 \bar{v}_{it}}f_{\bar{v}_{it}}(\bar{v}_{it})d\bar{v}_{it} \int e^{i\xi_2 \bar{\eta}_i}f_{\bar{\eta}_i}(\bar{\eta}_i)d\bar{\eta}_i \]
\[ = \frac{1}{\lambda_0} - i \frac{\partial F_m(w_{it}, \xi_1, \xi_2)}{\partial \xi_1} \phi_v(\xi_1)\phi_\eta(\frac{\xi_2}{\lambda_0}). \]

This yields Eqs. (6) and (7). Q.E.D.

**The proof of Theorem 2.1** We will recover \( f_{\bar{v}_{it}}(\bar{v}) \) first. Differentiating the definition of \( F_y(w_{it}, \xi_1, \xi_2) \) in Eq. (3) with respect to \( \xi_1 \) yields

\[
\frac{\partial}{\partial \xi_1} F_y(w_{it}, \xi_1, \xi_2) = \frac{\partial}{\partial \xi_1} \int \int E[Y_{it}|w_{it}, \bar{g}_{i<1}]e^{i\xi_1 \bar{g}_{i<1}}e^{i\xi_2 \bar{w}_i}d\bar{g}_{i<1}d\bar{w}_i \\
= i \int \int E[\bar{G}_{i<1}Y_{it}|w_{it}, \bar{g}_{i<1}]e^{i\xi_1 \bar{g}_{i<1}}e^{i\xi_2 \bar{w}_i}d\bar{g}_{i<1}d\bar{w}_i.
\]

Notice that Eq. (7) can be written as \( \frac{\partial F_m(w_{it}, \xi_1, \xi_2)}{\partial \xi_1} \phi_v(\xi_1)\phi_\eta(\frac{\xi_2}{\lambda_0}) = iF_{xy}(w_{it}, \xi_1, \xi_2) \). On the other hand, differentiating Eq. (6) with respect to \( \xi_1 \), we obtain

\[
\frac{\partial}{\partial \xi_1} F_y(w_{it}, \bar{w}_i, \xi_1, \xi_2) \\
= \frac{1}{\lambda_0} \left[ \frac{\partial F_m(w_{it}, \xi_1, \xi_2)}{\partial \xi_1} \phi_v(\xi_1) + F_m(w_{it}, \xi_1, \xi_2, \phi_\eta(\frac{\xi_2}{\lambda_0}) \right] \phi_\eta(\frac{\xi_2}{\lambda_0}) \\
= iF_{xy}(w_{it}, \bar{w}_i, \xi_1, \xi_2) + \frac{1}{\lambda_0} F_m(w_{it}, \xi_1, \xi_2, \phi_\eta(\frac{\xi_2}{\lambda_0}) \right] \phi_\eta(\frac{\xi_2}{\lambda_0}) \\
= i \int \int E[X_{it}Y_{it}|w_{it}, \bar{g}_{i<1}, \bar{w}_i]e^{i\xi_1 \bar{g}_{i<1}}e^{i\xi_2 \bar{w}_i}d\bar{g}_{i<1}d\bar{w}_i \\
+ \frac{1}{\lambda_0} F_m(w_{it}, \xi_1, \xi_2, \phi_\eta(\frac{\xi_2}{\lambda_0}) \right] \phi_\eta(\frac{\xi_2}{\lambda_0}).
\]
Combining Eqs. (A.7) and (A.8) yields

\[ i\mathcal{F}(\tilde{g} - x)y(w_{it}, \xi_1, \xi_2) \]

\[ = i \int \int \mathbb{E}[\tilde{G}_{i<t} - X_{it}]Y_{it}|w_{it}, \tilde{g}_{i<t}, \bar{w}_i]e^{i\xi_1\tilde{g}_{i<t}}e^{i\xi_2\bar{w}_i}d\tilde{g}_{i<t}d\bar{w}_i \]

(A.9)

Because \( \phi_v(\xi_1) \), \( \phi_\eta(\xi_2) \), and \( \mathcal{F}_m(w_{it}, \xi_1, \xi_2; \beta_0) \) 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. (6) to obtain

\[ -i\mathcal{F}(\tilde{g} - x)y(w_{it}, \xi_1, \xi_2) + \frac{\partial \phi_v(\xi_1)}{\phi_v(\xi_1)} \mathcal{F}_y(w_{it}, \xi_1, \xi_2) = 0. \]

(A.10)

By Theorem 1(b) in Zinde-Walsh (2014), there exists a unique function \( Q(\xi_1) \) such that

\[ -i\mathcal{F}(\tilde{g} - x)y(w_{it}, \xi_1, \xi_2) + Q(\xi_1)\mathcal{F}_y(w_{it}, \xi_1, \xi_2) = 0. \]

(A.11)

Integrating the above equation from 0 to \( \xi_1 \) with the boundary condition \( \phi_v(0) = \int f_{\tilde{V}_{it}}(\tilde{v}_{it})d\tilde{v}_{it} = 1 \) 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 \( f_{\tilde{V}_{it}}(\tilde{v}_{it}) \) is identified. Rescaling \( \xi_2 \) by \( \lambda_0 \xi_2 \) in Eq. (6) and rearranging the terms, we have

\[ \lambda_0 \mathcal{F}_y(w_{it}, \xi_1, \lambda_0 \xi_2) = \mathcal{F}_m(w_{it}, \xi_1, \xi_2; \beta_0)\phi_v(\xi_1)\phi_\eta(\xi_2), \]

(A.12)
Solving $\phi_\eta(\xi_2)$ from the above equation yields

(A.13) \[ \phi_\eta(\xi_2) = \frac{\lambda \mathcal{F}_y(w_{it}, \xi_1, \lambda_0 \xi_2)}{\mathcal{F}_m(w_{it}, \xi_1, \xi_2; \beta_0) \phi_v(\xi_1)}. \]

Because $\mathcal{F}_y(w_{it}, \xi_1, \xi_2)$, $\mathcal{F}_m(w_{it}, \xi_1, \xi_2; \beta)$ 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:

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

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

(A.15) \[ f_{\tilde{\eta}; \gamma}(\tilde{\eta}) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\xi_2 \tilde{\eta}} \phi_{\eta; \gamma}(\xi_2) d\xi_2. \]

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

(A.16) \[ \mathbb{E}[Y_{it}|w_{it}, \bar{g}_{i<t}, \bar{w}_i; \gamma] = \int \int m(w_{it}, \bar{g}_{i<t} - \bar{v}_{it}, \lambda_1 \bar{w}_i - \bar{\eta}_i; \beta) f_{\tilde{v}_{it}}(\bar{v}_{it}) f_{\tilde{\eta}; \gamma}(\tilde{\eta}_i) d\bar{v}_{it} d\tilde{\eta}_i. \]

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

\begin{equation}
(A.17) \quad E[Y_{it}|w_{it}, \tilde{g}_{i<t}, \overline{w}_i; \gamma_s] - E[Y_{it}|w_{it}, \tilde{g}_{i<t}, \overline{w}_i; \gamma_0] = \sum_{t=1}^{d_h} \frac{\partial E[Y_{it}|w_{it}, \tilde{g}_{i<t}, \overline{w}_i; \gamma_s]}{\partial \beta_t} (\beta_{st} - \beta_0) + \sum_{k=1}^{2} \frac{\partial E[Y_{it}|w_{it}, \tilde{g}_{i<t}, \overline{w}_i; \gamma_s]}{\partial \lambda_k} (\lambda_{sk} - \lambda_{0k}),
\end{equation}

where \(\gamma^* = (\beta^*, \lambda^*)\) is a parameter between \(\gamma_s\) and \(\gamma_0\). Combining these relationships yields

\begin{equation}
(A.18) \quad 0 = \sum_{t=1}^{d_h} \frac{\partial E[Y_{it}|w_{it}, \tilde{g}_{i<t}, \overline{w}_i; \gamma_s]}{\partial \beta_t} (\beta_{st} - \beta_0) \frac{(\beta_{st} - \beta_0)}{\| (\beta_s, \lambda_s) - (\beta_0, \lambda_0) \|}
+ \sum_{k=1}^{2} \frac{\partial E[Y_{it}|w_{it}, \tilde{g}_{i<t}, \overline{w}_i; \gamma_s]}{\partial \lambda_k} (\lambda_{sk} - \lambda_{0k}) \frac{(\lambda_{sk} - \lambda_{0k})}{\| (\beta_s, \lambda_s) - (\beta_0, \lambda_0) \|},
\end{equation}

\begin{align*}
&= \nabla_\gamma E[Y_{it}|w_{it}, \tilde{g}_{i<t}, \overline{w}_i; \gamma_s]^T \left[ \frac{(\beta_s - \beta_0)}{\| (\beta_s, \lambda_s) - (\beta_0, \lambda_0) \|} \frac{(\lambda_s - \lambda_0)}{\| (\beta_s, \lambda_s) - (\beta_0, \lambda_0) \|} \right]^T S_{\gamma_s}.
\end{align*}

Because \(\| S_{\gamma_s} \|^2_E = 1 \) for all \(s\), \(S_{\gamma_s} : s = 1, \ldots\) is a distinct sequence on the unit sphere. This implies that there exist a convergent subsequence \(S_{\gamma_{s_j}} : j = 1, \ldots\) whose limit is also on the unit sphere. Denote the limit as \(S_{\gamma_0}\). Combining the continuity assumption in Assumption 2.6 and Eq. (A.18), we obtain

\begin{equation}
(A.19) \quad 0 = \nabla_\gamma E[Y_{it}|w_{it}, \tilde{g}_{i<t}, \overline{w}_i; \gamma_0]^T S_{\gamma_0}.
\end{equation}

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

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

Taking an expectation, we obtain

\begin{align*}
0 &= E \left[ \nabla_\gamma E[Y_{it}|w_{it}, \tilde{g}_{i<t}, \overline{w}_i; \gamma_0] \cdot \nabla_\gamma E[Y_{it}|w_{it}, \tilde{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.
\end{align*}
Since $I(\beta_0, \lambda_0)$ is nonsingular by Assumption 2.6, we have to conclude that $(\beta_0, \lambda_0)$ is identifiable from this contradiction. \hfill Q.E.D.

References


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

<table>
<thead>
<tr>
<th>Simulation I</th>
<th>( \beta_0 )</th>
<th>( \beta_1 )</th>
<th>( \beta_2 )</th>
<th>( \lambda_1 )</th>
<th>( \lambda_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.557</td>
<td>0.511</td>
<td>-0.409</td>
<td>-0.498</td>
<td>0.514</td>
</tr>
<tr>
<td>Median</td>
<td>0.559</td>
<td>0.509</td>
<td>-0.420</td>
<td>-0.492</td>
<td>0.516</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.154</td>
<td>0.121</td>
<td>0.162</td>
<td>0.111</td>
<td>0.126</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Simulation II</th>
<th>( \beta_0 )</th>
<th>( \beta_1 )</th>
<th>( \beta_2 )</th>
<th>( \lambda_1 )</th>
<th>( \lambda_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.501</td>
<td>0.508</td>
<td>-0.499</td>
<td>-0.498</td>
<td>0.502</td>
</tr>
<tr>
<td>Median</td>
<td>0.504</td>
<td>0.508</td>
<td>-0.503</td>
<td>-0.500</td>
<td>0.508</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.104</td>
<td>0.108</td>
<td>0.100</td>
<td>0.093</td>
<td>0.108</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Simulation III</th>
<th>( \beta_0 )</th>
<th>( \beta_1 )</th>
<th>( \beta_2 )</th>
<th>( \lambda_1 )</th>
<th>( \lambda_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.528</td>
<td>0.552</td>
<td>-0.507</td>
<td>-0.506</td>
<td>0.524</td>
</tr>
<tr>
<td>Median</td>
<td>0.530</td>
<td>0.552</td>
<td>-0.504</td>
<td>-0.498</td>
<td>0.526</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.118</td>
<td>0.133</td>
<td>0.103</td>
<td>0.100</td>
<td>0.120</td>
</tr>
</tbody>
</table>

Note: Standard deviations of the parameters are computed by the standard deviation of the estimates across 200 simulations and called (simulation) standard deviations.

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

<table>
<thead>
<tr>
<th>Simulation I</th>
<th>( \beta_0 )</th>
<th>( \beta_1 )</th>
<th>( \beta_2 )</th>
<th>( \lambda_1 )</th>
<th>( \lambda_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.536</td>
<td>0.506</td>
<td>-0.430</td>
<td>-0.485</td>
<td>0.516</td>
</tr>
<tr>
<td>Median</td>
<td>0.524</td>
<td>0.504</td>
<td>-0.423</td>
<td>-0.488</td>
<td>0.513</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.125</td>
<td>0.111</td>
<td>0.121</td>
<td>0.101</td>
<td>0.120</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Simulation II</th>
<th>( \beta_0 )</th>
<th>( \beta_1 )</th>
<th>( \beta_2 )</th>
<th>( \lambda_1 )</th>
<th>( \lambda_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.502</td>
<td>0.506</td>
<td>-0.500</td>
<td>-0.499</td>
<td>0.502</td>
</tr>
<tr>
<td>Median</td>
<td>0.509</td>
<td>0.506</td>
<td>-0.498</td>
<td>-0.502</td>
<td>0.506</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.104</td>
<td>0.109</td>
<td>0.100</td>
<td>0.093</td>
<td>0.107</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Simulation III</th>
<th>( \beta_0 )</th>
<th>( \beta_1 )</th>
<th>( \beta_2 )</th>
<th>( \lambda_1 )</th>
<th>( \lambda_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.530</td>
<td>0.544</td>
<td>-0.502</td>
<td>-0.507</td>
<td>0.525</td>
</tr>
<tr>
<td>Median</td>
<td>0.531</td>
<td>0.539</td>
<td>-0.499</td>
<td>-0.509</td>
<td>0.521</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.117</td>
<td>0.129</td>
<td>0.097</td>
<td>0.099</td>
<td>0.120</td>
</tr>
</tbody>
</table>

Note: Standard deviations of the parameters are computed by the standard deviation of the estimates across 200 simulations and called (simulation) standard deviations.
Table 3: Estimation of the APEs in Simulations (n=500)

<table>
<thead>
<tr>
<th></th>
<th>Infeasible</th>
<th>Sieve GMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation I:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>-0.250</td>
<td>-0.203</td>
</tr>
<tr>
<td>Std. dev.</td>
<td>0.000</td>
<td>0.072</td>
</tr>
<tr>
<td>RMSE</td>
<td>–</td>
<td>0.086</td>
</tr>
<tr>
<td>Simulation II:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>-0.375</td>
<td>-0.387</td>
</tr>
<tr>
<td>Std. dev.</td>
<td>0.038</td>
<td>0.117</td>
</tr>
<tr>
<td>RMSE</td>
<td>–</td>
<td>0.117</td>
</tr>
<tr>
<td>Simulation III:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>-1.661</td>
<td>-1.273</td>
</tr>
<tr>
<td>Std. dev.</td>
<td>0.083</td>
<td>0.251</td>
</tr>
<tr>
<td>RMSE</td>
<td>–</td>
<td>0.461</td>
</tr>
</tbody>
</table>

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)

<table>
<thead>
<tr>
<th></th>
<th>Infeasible</th>
<th>Sieve GMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation I:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>-0.250</td>
<td>-0.216</td>
</tr>
<tr>
<td>Std. dev.</td>
<td>0.000</td>
<td>0.049</td>
</tr>
<tr>
<td>RMSE</td>
<td>–</td>
<td>0.059</td>
</tr>
<tr>
<td>Simulation II:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>-0.375</td>
<td>-0.388</td>
</tr>
<tr>
<td>Std. dev.</td>
<td>0.025</td>
<td>0.118</td>
</tr>
<tr>
<td>RMSE</td>
<td>–</td>
<td>0.118</td>
</tr>
<tr>
<td>Simulation III:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>-1.662</td>
<td>-1.266</td>
</tr>
<tr>
<td>Std. dev.</td>
<td>0.060</td>
<td>0.225</td>
</tr>
<tr>
<td>RMSE</td>
<td>–</td>
<td>0.454</td>
</tr>
</tbody>
</table>

Note: Standard deviations of the parameters are computed by the standard deviation of the estimates across 150 simulations.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>ln(hours) overall</td>
<td>7.671</td>
<td>0.289</td>
<td>2.770</td>
<td>8.560</td>
</tr>
<tr>
<td>between</td>
<td>0.233</td>
<td>4.950</td>
<td>8.407</td>
<td></td>
</tr>
<tr>
<td>within</td>
<td>0.172</td>
<td>5.491</td>
<td>10.011</td>
<td></td>
</tr>
<tr>
<td>ln(wage) overall</td>
<td>2.614</td>
<td>0.448</td>
<td>-0.220</td>
<td>4.600</td>
</tr>
<tr>
<td>between</td>
<td>0.432</td>
<td>0.877</td>
<td>4.367</td>
<td></td>
</tr>
<tr>
<td>within</td>
<td>0.118</td>
<td>1.274</td>
<td>3.344</td>
<td></td>
</tr>
<tr>
<td>kids overall</td>
<td>1.484</td>
<td>1.218</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>between</td>
<td>1.191</td>
<td>0</td>
<td>5.333</td>
<td></td>
</tr>
<tr>
<td>within</td>
<td>0.257</td>
<td>-0.183</td>
<td>3.150</td>
<td></td>
</tr>
<tr>
<td>age overall</td>
<td>42.415</td>
<td>7.973</td>
<td>29</td>
<td>60</td>
</tr>
<tr>
<td>between</td>
<td>7.933</td>
<td>30</td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>within</td>
<td>0.849</td>
<td>40.748</td>
<td>44.081</td>
<td></td>
</tr>
<tr>
<td>age$^2$ overall</td>
<td>1,862.545</td>
<td>708.068</td>
<td>841</td>
<td>3,600</td>
</tr>
<tr>
<td>between</td>
<td>704.740</td>
<td>900.667</td>
<td>3,481.667</td>
<td></td>
</tr>
<tr>
<td>within</td>
<td>72.973</td>
<td>1,668.212</td>
<td>2,051.545</td>
<td></td>
</tr>
<tr>
<td>disb overall</td>
<td>0.083</td>
<td>0.276</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>between</td>
<td>0.230</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>within</td>
<td>0.153</td>
<td>-0.583</td>
<td>0.750</td>
<td></td>
</tr>
</tbody>
</table>

Note: The data is a three-periods of panel data with a cross-sectional size 532.
<table>
<thead>
<tr>
<th></th>
<th>Linear Correlated Random Effects</th>
<th>Semi-parametric Nonlinear Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>ln(wage)</td>
<td>0.041 (0.021)</td>
<td>0.039 (0.017)</td>
</tr>
<tr>
<td>kids</td>
<td>-0.015 (0.021)</td>
<td>-0.019 (0.007)</td>
</tr>
<tr>
<td>age</td>
<td>-0.009 (0.034)</td>
<td>-0.007 (0.004)</td>
</tr>
<tr>
<td>age²</td>
<td>0.000 (0.000)</td>
<td>-0.001 (0.001)</td>
</tr>
<tr>
<td>disab</td>
<td>-0.048 (0.035)</td>
<td>-0.024 (0.027)</td>
</tr>
<tr>
<td>kids</td>
<td>0.018 (0.024)</td>
<td>0.020 (0.045)</td>
</tr>
<tr>
<td>age</td>
<td>0.015 (0.037)</td>
<td>0.015 (0.052)</td>
</tr>
<tr>
<td>age²</td>
<td>0.000 (0.000)</td>
<td>0.001 (0.002)</td>
</tr>
<tr>
<td>disab</td>
<td>-0.109 (0.056)</td>
<td>-0.072 (0.089)</td>
</tr>
<tr>
<td>constant</td>
<td>7.526 (0.319)</td>
<td>2.957 (1.957)</td>
</tr>
<tr>
<td>APE</td>
<td>–</td>
<td>0.090 (0.057)</td>
</tr>
</tbody>
</table>

Note: Bootstrap (simulation) standard errors are reported in parentheses, using 200 bootstrap replications.