Identification and Estimation of Semi-parametric Censored Dynamic Panel Data Models*

Yingyao Hu and Ji-Liang Shiu

November 2012

Abstract

This study presents a semiparametric identification and estimation method for censored dynamic panel data models and their average partial effects using only two-period data. The proposed method transforms the semi-parametric specification of censored dynamic panel data models into a valid semi-parametric family of PDFs of observables without modeling the distribution of the initial condition. Then the censored dynamic panel data models can be identified by a standard maximum likelihood estimation (MLE). The identifying assumptions are related to the completeness of the families of known semiparametric PDFs corresponding to censored dynamic panel data models and observed conditional density functions between the dependent and explanatory variables. This study shows that the families of PDFs corresponding to dynamic tobit models and dynamic lognormal hurdle models satisfy the identification assumptions with two types of data generating process (DGP). This study proposes a sieve maximum likelihood estimator (sieve MLE) and investigates the finite sample properties of these sieve-based estimators through Monte Carlo analysis. This study presents the dynamic behavior of annual individual health expenditures estimated as an empirical illustration using the dynamic tobit model and data from the Medical Expenditure Panel Survey (MEPS).

Keywords: dynamic lognormal hurdle model, dynamic tobit model, initial condition, nonlinear dynamic panel data model, unobserved covariate, unobserved heterogeneity, endogeneity

*The authors are solely responsible for any remaining errors. The authors can be reached at jishiu@ccu.edu.tw and yhu@jhu.edu.
1. Introduction

The identification and estimation of dynamic panel data models is one of the main challenges in econometrics. These models are appealing in applied research because they consider the lagged value of the dependent variable as one of the explanatory variables and contain observed and unobserved permanent (heterogeneous) or transitory (serially-correlated) individual differences. In dynamic linear panel data models, researchers have developed and compared several instrumental variables (IV) estimators and generalized method of moments (GMM) estimators in the literature (Anderson and Hsiao (1982), Arellano and Bond (1991), Arellano and Bover (1995), Ahn and Schmidt (1995), Kiviet (1995), Blundell and Bond (1998), Hahn (1999), and Hsiao, Hashem Pesaran, and Kamil Talmiscioğlu (2002)).

When the time dimension, T, is fixed in nonlinear panel data models, the presence of the unobserved effect prevents the construction of a log-likelihood function that can be used to estimate structural parameters consistently. This is the so-called incidental parameters problem discussed by Neyman and Scott (1948). However, the dynamic nature of the models leads to the initial conditions problem because integrating the individual unobserved effect out of the distribution raises the issue of how to specify the distribution of the initial condition given unobserved heterogeneity. Wooldridge (2005) proposed finding the distribution conditional on the initial value and the observed history of strictly exogenous explanatory variables to solve the initial conditions problem. Shiu and Hu (2010) adopted the correlated random effect approach for nonlinear dynamic panel data models without specifying the distribution of the initial condition. They used the identification results of the nonclassical measurement error models of Hu and Schennach (2008) to achieve nonparametric identification of nonlinear dynamic panel data model with three periods of data. Honoré (1993), Hu (2002) and Honoré and Hu (2004) used moment restrictions to identify and estimate the parameters of censored dynamic panel data models. Their results were achieved without making distributions of unobserved heterogeneity and the disturbance, but they failed to identify the average partial effects.

Other quantities of interest in nonlinear panel data applications include the partial effects on the mean response, averaged across the population distribution of the unobserved heterogeneity. Chernozhukov, Fernández-Val, Hahn, and Newey (2009) derived bounds for average
effects in nonseparable panel data models and showed that they can tighten considerably for
semiparametric discrete choice models. Graham and Powell (2008) studied the average partial
effect over the distribution of unobserved heterogeneity, which represents the causal effect of a
small change in an endogenous regressor on a continuously-valued outcome of interest. Hoder-
lein and White (2009) considered identification of distributional effects and average effects in
general nonseparable models, allowing for arbitrary dependence between the persistent un-
observables and the regressors of interest even if there are only two time periods. However,
their approach explicitly rules out lagged dependent variables. Dynamic models focus on the
effects of the lagged dependent variables on the current dependent variable, whereas we want
to account or control for the influence of all other variables. The effect of a lagged dependent
variables reflects the persistence of the dependent variables over time and the amount of this
state dependence can be measured by the average partial effect.

This study focuses the identification and estimation of semi-parametric censored dynamic
panel data models and their average partial effects with two periods of data. Compared to
the identification results in Shiu and Hu (2010), the proposed approach requires a somewhat
stronger but places less demand on the time dimension of data. Under a semi-parametric
specification, the models are connected to the semi-parametric distribution of \( Y_{it} \) conditional
on \( (X_{it}, Y_{it-1}, U_{it}) \), i.e., \( f_{Y_{it}|X_{it}, Y_{it-1}, U_{it}; \theta} \) which is called the semi-parametric censored density
function. Then, the identification of \( \theta \) in \( f_{Y_{it}|X_{it}, Y_{it-1}, U_{it}; \theta} \) may lead to that of the proposed
semi-parametric censored dynamic panel data models. As mentioned earlier, the maximum
likelihood estimator (MLE) for the semi-parametric censored density function \( f_{Y_{it}|X_{it}, Y_{it-1}, U_{it}; \theta} \)
is inconsistent if the time-series dimension is finite and the unobserved covariate \( U_{it} \) is corre-
lated with explanatory variables. This study presents an identification strategy similar to the
approach in Hu and Shiu (2011a). The identification method transforms the semi-parametric
censored density function \( f_{Y_{it}|X_{it}, Y_{it-1}, U_{it}; \theta} \) into a valid semi-parametric family of PDFs of
observable variables. This identification technique involves three steps of nontrivial transfor-
mation associate with the completeness of known PDFs. The first step is to apply the inverse
of an integral operator using \( f_{Y_{it}|X_{it}, Y_{it-1}, U_{it}; \theta} \) as a kernel. The second step is to integrate out
the unobserved covariate. The last step is to normalize the integrated semi-parametric density
function created in the second step. The true value of structural parameters can be uniquely
determined by maximizing the likelihood function of the transformed semi-parametric fami-
ily of the PDFs of observables. This process also identifies the average partial effect of the censored dynamic panel data models. The nontrivial transformation steps rely on the completeness of the families of known PDFs \( f_{Y_{it}|X_{it}, Y_{it-1}, U_{it}; \theta} \) corresponding to censored dynamic panel data models and observed conditional density functions between the dependent and explanatory variables \( f_{Y_{it}|X_{it}, Y_{it-1}, X_{it-1}} \). The completeness assumptions in this study are not restrictive and applicable to certain types of censored dynamic panel data models with common DGPs. Examples include dynamic tobit models and dynamic lognormal hurdle models for the observed conditional density functions \( f_{Y_{it}|X_{it}, Y_{it-1}, X_{it-1}} \) following normal and distributions or the distributions of exponential families.

These identification results suggest a semi-parametric sieve maximum likelihood estimator (sieve MLE) for the proposed model. The consistency of the sieve MLE estimator and the asymptotic normality of its parametric components can be directly obtained from the standard treatment in the sieve MLE literature. This study shows how to implement sieve MLE estimators for dynamic tobit models and dynamic lognormal hurdle models. Combining the estimated parametric components with the nuisance parameter for the initial joint distribution makes it possible to derive a consistent estimator for the average partial effect. An apparent advantage of the proposed sieve MLE procedure is that we can estimate these nonlinear dynamic panel data models using two periods of data without specifying initial conditions. This is beneficial because semi-nonparametric estimators usually require a large sample. Another benefit is that the proposed method allows for time dummies, flexible functional forms of state dependence \( Y_{it-1} \) such as quadratics or interaction terms, and parametric heteroskedasticity.

The rest of the article is organized as follows. Section 2 presents the identification of censored dynamic panel data models through several nontrivial transformations. Section 3 shows the identification assumptions hold for dynamic tobit models and dynamic lognormal hurdle models for two types of DGPs. Section 4 presents the proposed sieve MLE and inference. Section 5 presents the results of Monte Carlo experiments for dynamic tobit models. Section 6 shows the application of the sieve MLE to a dynamic tobit model describing the dynamic behaviors of annual individual health expenditures using data from the Medical Expenditure Panel Survey (MEPS). Finally, Section 7 provides concluding remarks. Appendices include proofs of each transformation step and a discrete case.

\footnote{More discussions of the completeness condition can be found in \cite{D'Haultfoeuille2011}, \cite{Andrews2011}, and \cite{HuShiu2011b}.}
2. Identification of Censored Dynamic Panel Data Models

Suppose \( g_1(\cdot, \cdot; \theta_1), g_2(\cdot, \cdot; \theta_2) \) are known parametric functions and \( g_1 \) is strictly increasing in its second argument. Consider the following censored dynamic panel data model:

\[
Y_{it} = g_1 \left( g_2 \left( X'_{it}, Y_{it-1}; \theta_2 \right), V_i + \varepsilon_{it}; \theta_1 \right), \quad \forall i = 1, \ldots, N; t = 1, \ldots, T - 1, \tag{1}
\]

where \( Y_{it} \) is the dependent variable, \( X_{it} \) is a vector of observed explanatory variables, \( \varepsilon_{it} \) is a transitory error term, \( V_i \) is an unobservable individual-specific effect, and \( \theta \equiv (\theta_1, \theta_2) \) is the parameter to be estimated. The functions \( g_1 \) and \( g_2 \) may be specified by users, such as \( g_1(\chi, \nu; \theta_1) = \max(0, \chi + \nu) \) and \( g_2(X'_{it}, Y_{it-1}; \theta_2) = X'_{it} \beta + \gamma Y_{it-1}, \) etc. The specifications of \( g_2 \) can contain time trends, allowing nonlinear relationships such as quadratics or interactions terms. One of the difficulties of distinguishing between structures of Model (1) from observed samples is that the explanatory variables, \( (X'_{it}, Y_{it-1}, V_i) \), and the transitory error term, \( \varepsilon_{it} \), are not independently distributed. Assume that a function of variables in the past can purge the statistical dependence that may exist between the explanatory variables and the transitory error term \( \varepsilon_{it} \) in Model (1).

Assumption 2.1. (Exogenous Shocks) Set \( \eta_{it} \) as an unobserved serially-correlated component in the past such that \( \eta_{it} = \varphi \left( \{X_{i\tau}, Y_{i\tau-1}, \varepsilon_{i\tau}\}_{\tau=0,1,\ldots,t-1} \right) \) for some function \( \varphi \). Assume that a transitory random shock \( \xi_{it} \) is independent of \( \{X_{i\tau}, Y_{i\tau-1}, V_i, \varepsilon_{i\tau-1}\} \) for any \( \tau \leq t \). Then, the transitory error term \( \varepsilon_{it} \) has the following decomposition

\[
\varepsilon_{it} = \eta_{it} + \xi_{it}. \tag{2}
\]

Plugging Eq. (2) into Model (1) leads to

\[
Y_{it} = g_1 \left( g_2 \left( X'_{it}, Y_{it-1}; \theta_2 \right), V_i + \eta_{it} + \xi_{it}; \theta_1 \right) \\
\equiv g_1 \left( g_2 \left( X'_{it}, Y_{it-1}; \theta_2 \right), U_{it} + \xi_{it}; \theta_1 \right), \tag{3}
\]

where \( U_{it} = V_i + \eta_{it} \) is an unobserved covariate. To describe every structure of Model (3) by a parameter, assume that the distribution of \( \xi_{it} \) has a semi-parametric representation. This
effectively reduces the identification problem to identify a set of parameters. This framework leads to the following definitions.

**Definition 2.1.** Let $\Theta_\alpha$ be a parameter space and let $F(\xi; \alpha)$ be a proper distribution function. If $dF(\xi; \alpha_0)$ is the true distribution, then $dF(\xi; \alpha)$ is correctly specified at $\alpha_0$. The parameter point $\alpha_0$ is globally identifiable if there exists no other $\alpha \in \Theta_\alpha$ such that with probability 1, $dF(\xi; \alpha) = dF(\xi; \alpha_0)$, where the measure is taken with respect to $\alpha_0$.

**Definition 2.2.** The parameter point $\alpha_0$ is locally identifiable if there exists an open neighborhood of $\alpha_0$ containing no other $\alpha$ such that with probability 1, $dF(\xi; \alpha) = dF(\xi; \alpha_0)$, where the measure is taken with respect to $\alpha_0$.

If $\alpha_0$ is globally identifiable then it is locally identifiable.

**Assumption 2.2.** (Semi-parametric Distribution) The semi-parametric distribution of the transitory random shock $dF(\xi_{it}; \alpha)$ is known and is correctly specified at an unknown $\alpha_0$. The parameter point $\alpha_0$ is locally identifiable.

Because $g_1$ is strictly increasing in its second argument, the exogeneity of $\xi_{it}$ makes it possible to obtain

$$F_{\xi_{it}; \alpha}(\xi) = F_{Y_{it}|X_{it}, Y_{it-1}, U_{it}}(g_1(g_2(X'_{it}, Y_{it-1}; \theta_2), U_{it} + \xi; \theta_1) | X_{it}, Y_{it-1}, U_{it}), \quad (4)$$

or

$$g_1(g_2(X'_{it}, Y_{it-1}; \theta_2), U_{it} + \xi; \theta_1) = F_{Y_{it}|X_{it}, Y_{it-1}, U_{it}}^{-1}(F_{\xi_{it}; \alpha}(\xi) | X_{it}, Y_{it-1}, U_{it}; \theta), \quad (5)$$

if the inverse of $F_{Y_{it}|X_{it}, Y_{it-1}, U_{it}}$ exists and $\theta \equiv (\theta_1, \theta_2, \alpha)'$. Thus, according to Assumptions 2.1 and 2.2, there is a unique conditional distribution associated with each structure in the censored dynamic panel data Model (3) and the identification of the censored dynamic panel data models (1) is implied by that of the distribution of $Y_{it}$ conditional on $(X_{it}, Y_{it-1}, U_{it})$ (i.e., $f_{Y_{it}|X_{it}, Y_{it-1}, U_{it}}$). Given this semi-parametric representation, the identification problem is to find conditions such that a true underlying parameter $\theta_0$ can be distinguished on the basis of

---

2 The parameters considered are potentially infinite-dimensional ones. We assume a parameter is constituted of two components, a finite-dimensional parameter vector, and a potentially infinite-dimensional nuisance parameter.

3 The definitions can be found in [Bowden, 1973](#).
sample observations. The conditional PDF $f_{Y|X,Y_{t-1},U;\theta}$ corresponding to $F_{Y|X,Y_{t-1},U;\theta}$ is called the semi-parametric censored density function in this paper. We introduce two examples to highlight this important connection. Suppose $F$ and $f$ denote the CDF and the PDF of an independent random shock, respectively.

**Example 1** (Dynamic Censored Model with the Lagged Dependent Variable): Assume $g_1(\chi, \nu; \theta_1) = \max(0, \chi + \nu)$.

$$Y_{it} = \max\{0, g_2(X'_{it}, Y_{it-1}; \theta_2) + U_{it} + \xi_{it}\} \quad \text{with} \quad \forall i = 1, \ldots, N; t = 1, \ldots, T - 1. \quad (6)$$

The semi-parametric censored density function is

$$f_{Y|X,Y_{t-1},U;\theta} = F_{\xi;\alpha} \left(-g_2(X'_{it}, Y_{it-1}; \theta_2) - U_{it}\right)^{Y_{it}=0} \times f_{\xi;\alpha} \left(Y_{it} - g_2(X'_{it}, Y_{it-1}; \theta_2) - U_{it}\right)^{Y_{it}>0}. \quad (7)$$

**Example 2** (Dynamic Log Hurdle Model with the Lagged Dependent Variable): Define a binary indicator variable $d_{it} = 1 \left(g_3(X'_{it}, Y_{it-1}; \theta_1) + \varsigma_{it} \geq 0\right)$ where $1(\cdot)$ is the 0-1 indicator function. Suppose that $Y_{it} > 0$ is observed for $d_{it} = 1$ and $Y_{it} = 0$ for $d_{it} = 0$. When $Y_{it} > 0,$

$$Y_{it} = g_2(X'_{it}, Y_{it-1}; \theta_2) + U_{it} + \xi_{it} \quad \text{with} \quad \forall i = 1, \ldots, N; t = 1, \ldots, T - 1. \quad (8)$$

The conditional distribution of interest is

$$f_{Y|X,Y_{t-1},U;\theta} = F_{\varsigma;\alpha} \left(-g_3(X'_{it}, Y_{it-1}; \theta_1) - U_{it}\right)^{Y_{it}=0} \times f_{\varsigma;\alpha} \left(\log(Y_{it}) - g_2(X'_{it}, Y_{it-1}; \theta_2) - U_{it}\right)^{Y_{it}>0} \times f_{\xi;\alpha} \left(Y_{it} - g_2(X'_{it}, Y_{it-1}; \theta_2) - U_{it}\right)^{Y_{it}>0}. \quad (9)$$

Many nonlinear dynamic panel data models, such as dynamic discrete choice models, can be converted into their corresponding semi-parametric density functions. However, a discrete dependent variable may incur a very strong restriction that forbids a continuously distributed unobserved covariate. These examples satisfy the identification assumptions when the unobserved covariate is assumed to be continuously distributed.

A number of economic optimization studies have presented empirical applications of these
censored dynamic panel data models, where the dependent variables $Y_{it}$ represents the amount of insurance coverage chosen by an individual, annual women’s labor supply, a firm’s expenditures on R&D, or annual individual health expenditures. In both examples, the models contain the lagged censored dependent variables in the right-hand side (RHS), ruling out top-coded censored models that contain a lagged value of a latent variable. Because piles of the dependent variable at zero can be regarded as optimal solutions of utility maximizing behavior, these models are often called corner solution models with lagged censored dependent variables.

### 2.1. General Identification

Consider the semi-parametric censored density function:

$$f_{Y_{it}|X_{it},X_{i,t-1},U_{it};\theta}(y_{it}|x_{it},y_{i,t-1},u_{it}),$$

where $Y_{it}$ is the dependent variable for an individual $i$, and the explanatory variables include a lagged dependent variable, a set of possibly time-varying explanatory variables $X_{it}$, and the unobserved covariate $U_{it}$. As discussed earlier, the identification of this semi-parametric censored density function leads to the identification of certain types of censored dynamic panel data models. Assume that $\theta_0 \in \Theta$ is local identifiable. In other words, $\theta_0$ is a unique value of $\theta$ in an open neighborhood of $\theta_0$, which specifies the exact structure of the model. Consider a panel data containing two periods, $\{Y_{it}, X_{it}, Y_{i,t-1}, X_{i,t-1}\}$ for $i = 1, 2, ..., N$. Assume that for each $i$, $(Y_{it}, X_{it}, Y_{i,t-1}, X_{i,t-1})$ is an independent random draw from a bounded distribution $f_{Y_{it}, X_{it}, Y_{i,t-1}, X_{i,t-1}}$. The law of total probability leads to the following,

$$f_{Y_{it}, X_{it}, Y_{i,t-1}, X_{i,t-1}} = \int f_{Y_{it}|X_{it}, Y_{i,t-1}, U_{it};\theta_0} f_{X_{it}, Y_{i,t-1}, X_{i,t-1};\theta_0} du_{it}$$

(11)

where $f_{X_{it}, Y_{i,t-1}, X_{i,t-1};\theta_0} = f_{X_{it}, Y_{i,t-1}, X_{i,t-1};U_{it}}$ is the joint density function of variables $(x_{it}, y_{i,t-1}, x_{i,t-1}, u_{it})$. Let $\mathcal{Y}_{it}$, $\mathcal{X}_{it}$, and $\mathcal{U}_{it}$ be the support of random variables $Y_{it}$, $X_{it}$, and $U_{it}$, respectively. Set $L^2(\mathcal{Y}) = \{h(\cdot) : \int_{\mathcal{Y}} |h(y)|^2 dy < \infty\}$ and $L^2(\mathcal{U}, \omega) = \{h(\cdot) : \int_{\mathcal{U}} |h(u)|^2 \omega(u) du < \infty, \text{and } \int_{\mathcal{U}} \omega(u) du < \infty\}$. Note the weighted $L^2$-space, $L^2(\mathcal{U}, \omega)$, contains a constant function (i.e., $c(u) = c \ \forall u \in \mathcal{U}$). The key idea of this identification method is to extend Eq. (11) to the semi-parametric censored density function (10) over a proper subset of $\mathcal{Y}_{it}$. The first
step is to construct $f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}; \theta}$ and avoid an unwanted restriction. Set $Y_{it}^+$ as the argument of $y_{it}$ in $\tilde{Y}_{it} \subseteq Y_{it}$.

\[
\int_{\text{Observed from Data}} \frac{f_{Y_{it}^+, X_{it}, Y_{it-1}, X_{it-1}}}{f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}; \theta}} \ f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}; \theta} \, du_{it}. \tag{12}
\]

Because the observable density function $f_{Y_{it}^+, X_{it}, Y_{it-1}, X_{it-1}}$ in the LHS and the semi-parametric censored density function $f_{Y_{it}^+ | X_{it}, Y_{it-1}, U_{it}; \theta}$ are known, it is possible to construct a semi-parametric joint density function $f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}; \theta}$ using Eq. (12). Given $(x_{it}, y_{it-1})$ and a parameter $\theta$, define an integral operator as follows:

\[
L_{f_{Y_{it}^+ | X_{it}, Y_{it-1}, U_{it}; \theta}} : \mathcal{L}^2(U_{it}, \omega) \to \mathcal{L}^2(\tilde{Y}_{it}) \quad \tag{13}
\]

\[
(L_{f_{Y_{it}^+ | X_{it}, Y_{it-1}, U_{it}; \theta}} \omega \ h)(y_{it}^+) = \int L_{f_{Y_{it}^+ | X_{it}, Y_{it-1}, U_{it}; \theta}}(y_{it}^+) \ f_{Y_{it}^+ | X_{it}, Y_{it-1}, U_{it}; \theta}(y_{it}^+ | x_{it}, y_{it-1}, u_{it}) \ \frac{\omega(u_{it})}{\omega(u_{it})} \ h(u_{it}) \, du_{it}.
\]

If the integral operator $L_{f_{Y_{it}^+ | X_{it}, Y_{it-1}, U_{it}; \theta}}$ is invertible for each $\theta$, then Eq. (12) suggests that the semi-parametric joint density function can be obtained by

\[
f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}; \theta} \equiv L^{-1}_{f_{Y_{it}^+ | X_{it}, Y_{it-1}, U_{it}; \theta}}(f_{Y_{it}^+ | X_{it}, Y_{it-1}, X_{it-1}}).
\tag{14}
\]

Plugging the true parameter $\theta_0$ into this equation results in

\[
f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}; \theta_0} \equiv L^{-1}_{f_{Y_{it}^+ | X_{it}, Y_{it-1}, U_{it}; \theta_0}}(f_{Y_{it}^+ | X_{it}, Y_{it-1}, X_{it-1}})
\]

by Eq. (11). The semi-parametric joint density function still achieves the true joint density function at the population parameter $\theta_0$ or it is correctly specified at $\alpha_0$. The concept of completeness provides a sufficient condition for the invertibility of the integral operator using the semi-parametric censored density function as a kernel. The following definition presents this completeness.

**Definition 2.3.** A density function $f(y | u)$ satisfies a completeness condition if for $h(u) \in \mathcal{L}^2(U_{it}, \omega)$

\[\text{Extend Eq. (11) to } \theta \text{ by } f_{Y_{it}, X_{it}, Y_{it-1}, X_{it-1}} = \int f_{Y_{it} | X_{it}, Y_{it-1}, U_{it}; \theta} f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}; \theta} \, du_{it} \text{ over the whole } Y_{it}. \]

\[\text{Integrating out } y_{it} \text{ over } Y_{it} \text{ results in } f_{X_{it}, Y_{it-1}, X_{it-1}} = \int f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}; \theta} \, du_{it}. \]

This suggests that $f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}; \theta}$ loses the variation over $\theta$ after integrating out the observed covariate $U_{it}$.
\[ \mathcal{L}^2(\mathcal{U}, \omega) \text{ such that} \]
\[ \int h(u)f(y|u)\omega(u)du = 0 \quad \text{for all } y \]  \hspace{1cm} (15)

then \( h(u) = 0 \) almost everywhere. In other words, there is no nonzero function in \( \mathcal{L}^2(\mathcal{U}, \omega) \) with zero integration for each function in the family of the density functions \( \{ f(y|u) : y \in \mathcal{Y} \} \).

By switching the roles of \( y \) and \( u \) and dropping \( \omega \), it is possible to define \( \{ f(y|u) : u \in \mathcal{U} \} \) as complete in \( \mathcal{L}^2(\mathcal{Y}) \), and this definition can be generalized to function forms such as \( f(y,u) \).

**Assumption 2.3.** (Dependence between \( Y_{it} \) and \( U_{it} \)) For each \( \theta \in \Theta \) and fixed \( (x_{it}, y_{it-1}) \), the family of the semi-parametric censored density functions \( \{ \frac{1}{\omega(u_{it})}f_{Y_{it}^+|X_{it},Y_{it-1},U_{it};\theta} : y_{it}^+ \in \tilde{\mathcal{Y}}_{it} \} \) is complete over \( \mathcal{L}^2(\mathcal{U}_{it}, \omega) \).

Assumption 2.3 implies that a cardinality restriction in that the cardinality of \( \mathcal{U}_{it} \) is less than the cardinality of \( \tilde{\mathcal{Y}}_{it} \). Thus, if \( \mathcal{U}_{it} \) is a finite discrete set, then the proposed method may apply to a dynamic discrete choice model in which the dependent variable \( Y_{it} \) takes more discrete values. However, because of inaccessibility of units of measurement of the unobserved covariate \( U_{it} \), to some extent it is restrictive to assume \( \mathcal{U}_{it} \) is discrete. Therefore, allowing the unobserved covariate \( U_{it} \) to take continuous values is more appealing and this study focuses on censored dynamic panel data models.

Suppose that \( (L_{f_{Y_{it}^+|X_{it},Y_{it-1},U_{it};\theta}h_1})(y_{it}^+) = (L_{f_{Y_{it}^+|X_{it},Y_{it-1},U_{it};\theta}h_2})(y_{it}^+) \) for all \( y_{it}^+ \in \tilde{\mathcal{Y}}_{it} \). Assumption 2.3 guarantees that \( h_1 = h_2 \) (i.e., \( L_{f_{Y_{it}^+|X_{it},Y_{it-1},U_{it};\theta} \) is one-to-one). Hence, the operators \( L_{f_{Y_{it}^+|X_{it},Y_{it-1},U_{it};\theta}} \) are invertible for each \( \theta \) and \( (x_{it}, y_{it-1}) \). This assumption requires dependence between \( Y_{it} \) and \( U_{it} \) because the independence between \( Y_{it} \) and \( U_{it} \) violates Assumption 2.3. Although Assumption 2.3 ensures the existence of the semi-parametric joint density function, \( f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta} \), it may not be identifiable. The variation of the parameter \( \theta \) in \( f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta} \) might be lost in the sense that for any open neighborhood of \( \theta_0 \), there exists some \( \theta_1 \neq \theta_0 \) such that \( f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_1} = f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0} \). In other words, \( f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta} \) is not locally identifiable. In this case, applying the inverse transformation is useless because the parameter of interest \( \theta \) is not distinguished in the new transformed joint density functions, preventing the identification of the parameter. The assumption prevents this loss.

\footnote{The term \( \mathcal{U}_{it} \) is a composite error term whose support is involved with the supports of \( V_i \) and \( \eta_{it} \). Assuming discreteness suggests that both \( V_i \) and \( \eta_{it} \) are discrete, which may be a strong assumption.}
Assumption 2.4. Given each \((x_{it}, y_{it-1})\), suppose \(f_{X_{it},Y_{it-1},X_{it-1}} > 0\). Assume the following conditions:

(i) (Dependence between \(Y_{it}\) and \(X_{it-1}\)) The family of the observable conditional density functions over \(X_{it-1}\), \(\{f_{Y_{it}^+|X_{it},Y_{it-1},X_{it-1}} : x_{it-1} \in X_{it-1}\}\), is complete over \(L^2(\tilde{Y}_{it})\).

(ii) (Dependence between \(Y_{it}\) and \(U_{it}\)) The family of the semi-parametric censored density function over \(U_{it}\), \(\{f_{Y_{it}^+|X_{it},Y_{it-1},U_{it};\theta_0} : u_{it} \in U_{it}\}\), is complete over \(L^2(\tilde{Y}_{it})\).

This assumption warrants several comments. First, the conditional density functions in the statement are both observable. Second, part (i) suggests that \(X_{it}\) cannot be constant over time. If \(X_{it}\) is constant across time, then \(X_{it} = X_{it-1}\) and \(f_{Y_{it}^+|X_{it},Y_{it-1},X_{it-1}} = f_{Y_{it}^+|X_{it},Y_{it-1}}\) which clearly violates the completeness in part (i). Finally, similar to Assumption 2.3 part (ii) requires that, the cardinality of \(\tilde{Y}_{it}\) is less than the cardinality of \(U_{it}\). Combining the cardinality restrictions in Assumption 2.3 and Assumption 2.4 (ii) shows that the cardinality of \(\tilde{Y}_{it}\) is equal to the cardinality of \(U_{it}\). This restriction is compatible with both the dependent variable \(Y_{it}\) and the unobserved covariate \(U_{it}\) taking continuous values.

Lemma 2.1. (Applying the inverse) Under Assumptions 2.3 and 2.4, the semi-parametric joint density \(f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta}\) is correctly specified at \(\theta_0\) and the parameter \(\theta_0\) is locally identifiable (i.e., there is an open neighborhood of \(\theta_0\) containing no other \(\theta\) such that \(f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta} = f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0}\)).

Proof: See the appendix.

Because \(f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta}\) contains the unobserved component \(U_{it}\), we need to integrate it out to acquire an observable density function. Set

\[
\tilde{f}_{X_{it},Y_{it-1},X_{it-1};\theta}(x_{it}, y_{it-1}, x_{it-1}) \equiv \int f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta}(x_{it}, y_{it-1}, x_{it-1}, u_{it}) du_{it}. \tag{16}
\]

To identify \(\theta\) from the integrated semi-parametric density function \(\tilde{f}_{X_{it},Y_{it-1},X_{it-1};\theta}\), it is necessary to examine whether \(\tilde{f}_{X_{it},Y_{it-1},X_{it-1};\theta}\) can be correctly specified at \(\theta_0\) and the parameter \(\theta_0\) is locally identifiable after applying the integration. This integration step might impose too many restrictions on the parameters and degenerate the variation of the function over its parameter space. Thus, it is necessary to rule out these degenerated cases. The following condition maintains the nontrivial semi-parametric representation of \(\tilde{f}_{X_{it},Y_{it-1},X_{it-1};\theta}\).
Assumption 2.5. (Variation of parameters around $\theta_0$) The family of the derivative of the semi-parametric censored density functions with respect to $\theta$, $\{\frac{\partial}{\partial \theta} f_{Y_{it}^+ | X_{it}, Y_{it-1}, U_{it}; \theta_0} : u_{it} \in U_{it}\}$, is complete over $L^2(\tilde{Y}_{it})$.

We summarize the results of the nontrivial transformation of the semi-parametric censored density function after the integration.

Lemma 2.2. (Integrating out) Under Assumptions 2.3-2.5, the integrated semi-parametric joint density $\tilde{f}_{X_{it}, Y_{it-1}, X_{it-1}; \theta}$ is correctly specified at $\theta_0$ and the parameter $\theta_0$ is locally identifiable.

Proof: See the appendix.

At this point in the process, the unobserved component of the semi-parametric censored density function (10) has been transformed out and the parameter $\theta$ of the function becomes the parameter of the observable semi-parametric function $\tilde{f}_{X_{it}, Y_{it-1}, X_{it-1}; \theta}$. However, if $\tilde{f}_{X_{it}, Y_{it-1}, X_{it-1}; \theta}$ does not integrate to unity (with respect to the measure $dx_{it}dy_{it-1}dx_{it-1}$), it is not a candidate of the semi-parametric family of PDFs for $f_{X_{it}, Y_{it-1}, X_{it-1}; \theta}$ and the standard MLE cannot be applied to $\tilde{f}_{X_{it}, Y_{it-1}, X_{it-1}; \theta}$. To obtain a valid semi-parametric family of PDFs, perform the following normalization step

$$f_{X_{it}, Y_{it-1}, X_{it-1}; \theta} \equiv \frac{\tilde{f}_{X_{it}, Y_{it-1}, X_{it-1}; \theta}}{\int \int \int \tilde{f}_{X_{it}, Y_{it-1}, X_{it-1}; \theta} dx_{it}dy_{it-1}dx_{it-1}}. \quad (17)$$

Similar to the previous discussion, it is necessary to show that the PDF of observables $f_{X_{it}, Y_{it-1}, X_{it-1}; \theta}$ is correctly specified at $\theta_0$ and the parameter $\theta_0$ is locally identifiable after this normalization. The following assumption and lemma demonstrate the existence of a nontrivial $\theta_0$ after normalization.

Assumption 2.6. (Dependence between $Y_{it}$ and $X_{it-1}$) Assume that the family of the observable conditional density functions $\{\frac{\partial}{\partial x_{it-1}} f_{Y_{it}^+ | X_{it}, Y_{it-1}, X_{it-1}} : x_{it-1} \in X_{it-1}\}$ is complete over $L^2(\tilde{Y}_{it})$ for each $x_{it}, y_{it-1}$.

If $X_{it-1}$ is discrete, the derivative can be replaced with the difference. Notice that both Assumption 2.4(i) and Assumption 2.6 are related to the observable conditional distribution.

---

6The appendix presents the discussion of a discrete case.
\( f_{Y_{it}^+|X_{it},Y_{it-1},X_{it-1}} \) and Assumption \( 2.6 \) implies Assumption \( 2.4(iii) \). Hence, the two assumptions are compatible and it is only necessary to verify the completeness in Assumption \( 2.6 \).

The assumption rules out the cases that \( f_{Y_{it}^+|X_{it},Y_{it-1},X_{it-1}} = f_1(Y_{it}^+, X_{it}, Y_{it-1})f_2(X_{it-1}) \) or \( f_{Y_{it}^+|X_{it},Y_{it-1},X_{it-1}} = f_1(Y_{it}^+)f_2(X_{it}, Y_{it-1}, X_{it-1}) \).

**Lemma 2.3.** (Normalization) Under Assumptions \( 2.3, 2.4(ii) \) and \( 2.5, 2.6 \) the PDF of observables after normalization, \( f_{X_{it},Y_{it-1},X_{it-1};\theta} \), is correctly specified at \( \theta_0 \) and the parameter \( \theta_0 \) is locally identifiable.

**Proof:** See the appendix.

The nontrivial transformation includes applying the inverse of an integral operator using \( f_{Y_{it}|X_{it},Y_{it-1},U_{it};\theta} \) as a kernel, integrating out the unobserved covariate, and normalization. After these three steps of transformation associated with the completeness of PDFs, the semi-parametric PDFs of observables \( \{f_{X_{it},Y_{it-1},X_{it-1};\theta} : \theta \in \Theta \} \) is correctly specified at \( \theta_0 \) and the parameter \( \theta_0 \) is locally identifiable under Assumptions \( 2.3, 2.6 \). To distinguish the parameters of interest \( \theta_0 \) from the parameter space \( \Theta \) on the basis of sample information, use the Kullback-Leibler information criterion

\[
K(\theta) = E \left[ \log \left( \frac{f_{X_{it},Y_{it-1},X_{it-1};\theta}(x_{it}, y_{it-1}, x_{it-1})}{f_{X_{it},Y_{it-1},X_{it-1};\theta}(x_{it}, y_{it-1}, x_{it-1})} \right) \right]
\]

where expectation is taken with respect to \( f_{X_{it},Y_{it-1},X_{it-1}} \). Applying the standard framework of the identifiability criterion of maximum likelihood estimation (MLE) produces the following results.

**Theorem 2.1.** Suppose that \( K(\theta) = 0 \) has a unique solution at \( \theta = \theta_0 \) in \( \Theta \). Under Assumptions \( 2.3, 2.4(ii) \) and \( 2.5, 2.6 \) the semi-parametric censored density function \( f_{Y_{it}|X_{it},Y_{it-1},U_{it};\theta} \) and the joint density function \( f_{X_{it},Y_{it-1},X_{it-1},U_{it}} \) can then be identified given the distribution of the two-period observable variables \( (y_{it}, x_{it}, y_{it-1}, x_{it-1}) \) for \( i = 1, 2, ..., N \).

In addition to the Kullback-Leibler information of classical statistics, the identification result is based on the completeness of the families of known PDFs \( f_{Y_{it}|X_{it},Y_{it-1},U_{it};\theta} \) corresponding

---

\(^7\)Suppose that \( h \in L^2(\bar{Y}_{it}) \) and \( \int h(y_{it}^+)f_{Y_{it}^+|X_{it},Y_{it-1},X_{it-1}} dy_{it}^+ = 0 \) for any \( X_{it-1} \). The integration does not involve \( X_{it-1} \) and we can take derivative with respect to \( X_{it-1} \). This leads to \( \int h(y_{it}^+)\pi_{x_{it-1}} f_{Y_{it}^+|X_{it},Y_{it-1},X_{it-1}} dy_{it}^+ = 0 \) for any \( X_{it-1} \). If \( \pi_x f_{Y_{it}^+|X_{it},Y_{it-1},X_{it-1}} : x_{it-1} \in X_{it-1} \) satisfies Assumption \( 2.6 \) then \( h = 0 \), which implies that Assumption \( 2.4(iii) \) holds.
to censored dynamic panel data models and observed conditional density functions between the dependent and explanatory variables \( f_{Y_{it}|X_{it},Y_{it-1},X_{it-1}} \). For some semi-parametric specifications, the conditions of completeness are easy to verify or draw inferences from samples. The following sections provide detailed discussions.

Theorem 2.1 provides the identification of the parameter \( \theta \). However, because \( U_{it} \) does not have meaningful units of measurement, it is not apparent what values of \( U_{it} \) we should use.

In nonlinear models, estimating the average partial effects of explanatory variables is more attractive than estimating parameters. Thus, this study introduces the average structure function (ASF) by averaging a scalar function of \( y_{it}, \omega(y_{it}) \), across the distribution of \( U_{it} \) in the population. Let \((X_{it}, Y_{it-1})\) be a given value of the explanatory variables, whose average structure function is

\[
\mu(X_{it}, Y_{it-1}) = E_{U_{it}} [E_{Y_{it}} [\omega(y_{it}) | X_{it}, Y_{it-1}, U_{it}]] \\
= \int_{U_{it}} \left( \int_{Y_{it}} \omega(y_{it}) f_{Y_{it}|X_{it},Y_{it-1},U_{it}} dy_{it} \right) f_{U_{it}} du_{it}.
\]  

(19)

The marginal distribution of the unobserved covariate \( U_{it} \) is also identified by the integration of the joint density function:

\[
f_{U_{it}} = \int_{X_{it}} \int_{Y_{it-1}} \int_{X_{it-1}} f_{X_{it},Y_{it-1},Y_{it-1},U_{it}} dx_{it} dy_{it-1} dx_{it-1}.
\]  

(20)

Combining the identification results of \( f_{Y_{it}|X_{it},Y_{it-1},U_{it}} \) and \( f_{U_{it}} \) provides the identification of the average structure function \( \mu(X_{it}, Y_{it-1}) \). This indicates that the average partial effect is also identified because the average partial effect can be defined by taking derivatives or differences of ASF in Eq. (19) with respect to elements of \((X_{it}, Y_{it-1})\). This yields the identification of the average partial effect.

Corollary 2.1. Under Assumptions 2.3, 2.4(ii) and 2.5-2.6, the average partial effect defined as derivatives or differences of Eq. (19) is identified by a two-period panel data, \(\{Y_{it}, X_{it}, Y_{it-1}, X_{it-1}\} \) for \(i = 1, 2, ..., N\).

The identification condition in Theorem 2.1 that the log-likelihood has a unique maximum

---

*Section 3 focuses on normal distributed \( \xi_{it} \) in Examples 1 & 2 satisfying the completeness assumptions with two type of DGP's, whereas subsection 4.3 presents a test concerning the completeness of \( f_{Y_{it}|X_{it},Y_{it-1},X_{it-1}} \) in Assumptions 2.6*
at the true parameter $\theta_0$. In a parametric likelihood case the local identifiability of the unknown parameter vector is equivalent to the non-singularity of the information matrix under weak regularity conditions. If the true parameter $\theta_0$ is a critical point of $K(\theta_0)$, then a sufficient condition of the uniqueness of $\theta_0$ is $K''(\theta_0)$ is negative semidefinite. The second derivative of $K(\theta_0)$ in the scalar case is

$$K''(\theta_0) = -E \left[ \left( \frac{\partial}{\partial \theta} f_{X_{it}, Y_{it-1}, X_{it-1}; \theta} \bigg| \theta = \theta_0 \right)^2 \right]$$

(21)

where the expression of $\frac{\partial}{\partial \theta} f_{X_{it}, Y_{it-1}, X_{it-1}; \theta} \bigg| \theta = \theta_0$ appears in Eq. (67). The vector case of $K''(\theta_0)$ is $K''(\theta_0) = [K_{lm}]$, where

$$K_{lm} = -E \left[ \left( \frac{\partial}{\partial \theta_l} f_{X_{it}, Y_{it-1}, X_{it-1}; \theta} \bigg| \theta = \theta_0 \right) \frac{\partial}{\partial \theta_m} f_{X_{it}, Y_{it-1}, X_{it-1}; \theta} \bigg| \theta = \theta_0 \right]$$

(22)

where $\frac{\partial}{\partial \theta_l} f_{X_{it}, Y_{it-1}, X_{it-1}; \theta} \bigg| \theta = \theta_0$ is equal to the term in Eq. (67) after replacing with the partial derivative $\frac{\partial}{\partial \theta_l}$. These results are sufficient conditions for the identification.

**Corollary 2.2.** Suppose that in an open neighbor of $\theta_0$ in $\Theta$, the second derivative of the Kullback-Leibler function $K(\theta)$ in Eq. (21) or (22) is negative definite. Under Assumptions 2.3, 2.4(ii) and 2.5-2.6, the semi-parametric censored density function $f_{Y_{it}|X_{it}, Y_{it-1}, U_{it}; \theta}$ and the joint density function $f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}}$ can be identified given the distribution of the two-period observable variables $(y_{it}, x_{it}, y_{it-1}, x_{it-1})$ for $i = 1, 2, ..., N$.

**Proof:** See the appendix.

### 3. Examples

Consider the two examples presented at the beginning of Section 2. This section shows when the completeness conditions in Section 2 hold in these cases. Assumptions 2.3, 2.4(ii), and 2.5 are related to the completeness of the variant forms of the semi-parametric censored density function $f_{Y_{it}|X_{it}, Y_{it-1}, U_{it}; \theta}$. Equations 7 and 9 show that the completeness of the semi-parametric censored density functions over positive $Y_{it}$ in the two motivating examples are connected to the PDF of the random shock $\xi_{it}$. Therefore, this section focuses on what kind of

---

9The parameter $\theta$ only contains a finite-dimensional component.
semi-parametric distribution assumptions in $\xi_{it}$ make these examples satisfy the completeness assumptions. For simplicity, assume the domains of $\xi_{it}$ and $U_{it}$ are $\mathbb{R}$.

Most of the interesting leading cases for Models (7) and (9) occur when the random shock $\xi_{it}$ is assumed to have an independent Gaussian white noise process. For simplicity, assume $g_2(X'_{it}, Y_{it-1}; \theta_2) = X'_{it}\beta + \gamma Y_{it-1}$. In this case, the semi-parametric censored density function $f_{Y_{it}|X_{it}, Y_{it-1}, U_{it}}$ is fully parameterized and correctly specified at $\theta_0$. The specifications of the models under the normality assumption are as follows:

**Semi-parametric Dynamic Tobit Models:**
Assuming that $\xi_{it} \sim N(0, \sigma_\xi)$, Eq. (7) leads to

$$f_{Y_{it}|X_{it}, Y_{it-1}, U_{it}; \theta} = \left[1 - \Phi\left(\frac{X'_{it}\beta + \gamma Y_{it-1} + U_{it}}{\sigma_\xi}\right)\right]^{1(Y_{it}=0)} \times \left[1 + \frac{1}{\sigma_\xi} \phi\left(\frac{Y_{it} - X'_{it}\beta - \gamma Y_{it-1} - U_{it}}{\sigma_\xi}\right)\right]^{1(Y_{it}>0)}, \tag{23}$$

where $\theta = (\beta, \gamma, \sigma_\xi^2)^T$.

**Semi-parametric Dynamic Lognormal Hurdle Models:**
Let $g_3(X'_{it}, Y_{it-1}; \theta_1) = X'_{it}\beta_d + \gamma_d Y_{it-1}$. Suppose that $\varsigma_{it} \sim N(0, 1)$ and $\xi_{it} \sim N(0, \sigma_\xi)$. Equation (9) then becomes

$$f_{Y_{it}|X_{it}, Y_{it-1}, U_{it}; \theta} = \left(1 - \Phi\left(X'_{it}\beta_d + \gamma_d Y_{it-1} + U_{it}\right)\right)^{1(Y_{it}=0)} \left\{\Phi\left(X'_{it}\beta_d + \gamma_d Y_{it-1} + U_{it}\right) \times \frac{1}{\sigma_\xi} \phi\left(\frac{\log(Y_{it}) - X'_{it}\beta_d - \gamma Y_{it-1} - U_{it}}{\sigma_\xi}\right)\right\}^{1(Y_{it}>0)}, \tag{24}$$

where $\theta = (\beta_d, \gamma_d, \beta, \gamma, \sigma_\xi)^T$.

The normality assumption makes it possible to verify the completeness of the semi-parametric censored density function $f_{Y_{it}^+|X_{it}, Y_{it-1}, U_{it}; \theta}$ in Assumptions 2.3, 2.4(ii), and 2.5 directly. It is then necessary to show that the semi-parametric censored density functions (23) and (24) satisfy these completeness assumptions. To do this, introduce the completeness of normal distributions and exponential families in $L^2$ from Hu and Shiu (2011b) which are
Lemma 3.1. Suppose that the distribution of $u$ conditional on $y$ is $N(a+by, \sigma^2)$ for $b, \sigma^2 > 0$ and the support of $y$ contains an open set. In this case, $E[h(\cdot)|y] = 0$ for any $x \in Y$ implies $h(u) = 0$ almost everywhere in $U$; equivalently, $\{f(u|y) : y \in Y\}$ is complete in $L^2(U)$.

Lemma 3.2. Let $f(u|y) = s(u)t(y)\exp[\mu(y)\tau(u)]$, where $s(u) > 0$, $\tau(u)$ is one-to-one in $u$, and the support of $\mu(y), Y$, contains an open set. In this case, $E[h(\cdot)|y] = 0$ for any $y \in Y$ implies $h(u) = 0$ almost everywhere in $U$; equivalently, the family of conditional density functions $\{f(u|y) : y \in Y\}$ is complete in $L^2(U)$.

Lemma 3.1 implies that for an open set $O_y \subset Y$, $\{\phi\left(\frac{u-(a+by)}{\sigma}\right) : y \in O_y\}$ is complete in $L^2(U)$. This completeness can be extended to a weighted space $L^2(U, \omega)$ for an appropriately chosen $\omega$. Set $\omega(u) = e^{-\frac{u^2}{2\sigma^2}}$. Suppose that $h \in L^2(U, \omega)$ such that for $y \in O_y$,

$$
\int h(u)\phi\left(\frac{u-(a+by)}{\sigma}\right)\omega(u)du = 0.
$$

Multiplying the equation by $e^{\frac{1}{4\sigma^2}(a+by)^2}$ results in

$$
0 = \int h(u)\phi\left(\frac{u-(a+by)}{\sigma}\right)\omega(u)e^{\frac{1}{4\sigma^2}(a+by)^2}du.
$$

It follows that

$$
0 = \int h(u)\omega(u)^{\frac{1}{2}}\phi\left(\frac{u-\frac{1}{2}(a+by)}{\sigma/\sqrt{2}}\right)du
$$

for $y \in O_y$. Note $h(u)\omega(u)^{\frac{1}{2}} \in L^2(U)$ because $h \in L^2(U, \omega)$. Lemma 3.1 also implies that $\{\phi\left(\frac{u-\frac{1}{2}(a+by)}{\sigma/\sqrt{2}}\right) : y \in O_y\}$ is complete $L^2(U)$. Applying this result to the equation suggests that $h(u) = 0$ almost everywhere. Therefore, $\{\phi\left(\frac{u-(a+by)}{\sigma}\right) : y \in O_y\}$ is complete in $L^2(U, \omega)$.

Based on the information about the completeness of normal distributions, it is possible to investigate the completeness condition of Models (23) and (24).

Semi-parametric Dynamic Tobit Models:

\[\text{11}\text{See Theorems 2.2 and 2.3 in Newey and Powell (2003) for details. More general discussions of completeness can be found in D’Haultfoeuille (2011), Andrews (2011), and Hu and Shiu (2011b).}\]
Set $\tilde{Y}_{it} = \mathbb{R}^+$. Given $\theta \in \Theta$, and $(x_{it}, y_{it-1})$,

$$f_{Y_{it}^+ | X_{it}, Y_{it-1}, U_{it}; \theta} = \frac{1}{\sigma \xi} \phi \left( \frac{Y_{it}^+ - X_{it}' \beta - \gamma Y_{it-1} - U_{it}}{\sigma \xi} \right). \quad (25)$$

**Semi-parametric Dynamic Lognormal Hurdle Models:**

Given $\theta \in \Theta$, and $(x_{it}, y_{it-1})$,

$$f_{Y_{it}^+ | X_{it}, Y_{it-1}, U_{it}; \theta} = \Phi \left( X_{it}' \beta d + \gamma_d Y_{it-1} + U_{it} \right) \phi \left( \frac{\log(Y_{it}^+) - X_{it}' \beta - \gamma Y_{it-1} - U_{it}}{\sigma \xi} \right) \frac{1}{\sigma \xi Y_{it}^+} \quad (26)$$

The completeness conditions in Section 2 are all associated with the dependent variables $Y_{it}^+$ and the unobserved covariate $U_{it}$. Therefore, it is necessary to investigate which functional forms connect these two variables. In these models, the dependent variables $Y_{it}^+$ and the unobserved covariate $U_{it}$ are both inside the standard normal PDF $\phi$. It follows that semi-parametric dynamic tobit models satisfies Assumptions 2.3 and 2.4(ii) by switching the role between $Y_{it}^+$ and $U_{it}$ to the result of Lemma 3.1. Because the standard normal CDF $\Phi$ is positive, the semi-parametric dynamic lognormal hurdle models also fulfill Assumptions 2.3 and 2.4(ii) using Lemma 3.1 and a change of variable.$^{12}$

Assumption 2.5 requires that the partial derivatives of the semi-parametric censor density function with respect to all components of the parameter $\theta$ be complete. According to the functional forms in Eqs. (25) and (26) and use of a change of variable, two types of the partial derivatives of $f_{Y_{it}^+ | X_{it}, Y_{it-1}, U_{it}; \theta_0}$ should be considered. The first one is the partial derivative with respect to the components of $\beta$ and $\gamma$, and the second one is $\sigma \xi$. The completeness of the first type can be reduced to the completeness of the family of $\{(y - c - u) \phi \left( \frac{y-c-u}{\sigma \xi} \right) : u \in \mathcal{U} \}$ in $L^2(\tilde{Y})$ for some constant $c$. Similarly, the completeness of the second type depends on the family of $\{ \left( \sigma \xi^2 - (y - c - u)^2 \right) \phi \left( \frac{y-c-u}{\sigma \xi} \right) : u \in \mathcal{U} \}$ in $L^2(\tilde{Y})$ for some constant $c$. The following lemma provides the completeness of the families of variant of the normal PDF $\phi$.

**Lemma 3.3.** Suppose the domain $\mathcal{U}$ contains an open set. For a constant $c$, the families of functions $\{(y - c - u) \phi \left( \frac{y-c-u}{\sigma \xi} \right) : u \in \mathcal{U} \}$ and $\{ \left( \sigma \xi^2 - (y - c - u)^2 \right) \phi \left( \frac{y-c-u}{\sigma \xi} \right) : u \in \mathcal{U} \}$ are

$^{12}$Use $\dot{Y}_{it} = \log(Y_{it}^+)$ and then $d\dot{Y}_{it} = \frac{1}{Y_{it}^+} dY_{it}^+$. 

18
both the dependent variables.

Adding the heterogeneous structure does not affect the functional form, which dominates such that \( \xi_{it} \) complete in \( L^2(\hat{Y}) \).

**Proof:** See the appendix.

This discussion also applies to models with heteroskedasticity, which allow more general functional form in corner solution models. If \( \xi_{it} \) has a heteroskedastic normal distribution such that \( \xi_{it} \sim N(0, h(X'_{it}, Y_{it-1}; \sigma_\xi)) \) then the semi-parametric censored density functions in Eqs. (23) and (24) respectively become

\[
f_{Y_{it}|X_{it}, Y_{it-1}, U_{it}; \theta} = \left[ 1 - \Phi \left( \frac{X'_{it}\beta + \gamma Y_{it-1} + U_{it}}{h(X'_{it}, Y_{it-1}; \sigma_\xi)} \right) \right] 1(Y_{it}=0) \times \\
\left[ \frac{1}{h(X'_{it}, Y_{it-1}; \sigma_\xi)} \phi \left( \frac{Y_{it} - X'_{it}\beta - \gamma Y_{it-1} - U_{it}}{h(X'_{it}, Y_{it-1}; \sigma_\xi)} \right) \right] 1(Y_{it}>0),
\]

(27)

and

\[
f_{Y_{it}|X_{it}, Y_{it-1}, U_{it}; \theta} = (1 - \Phi \left( X'_{it}\beta_d + \gamma_d Y_{it-1} + U_{it} \right)) 1(Y_{it}=0) \left\{ \Phi \left( X'_{it}\beta_d + \gamma_d Y_{it-1} + U_{it} \right) \right. \\
\times \phi \left( \frac{\log(Y_{it}) - X'_{it}\beta - \gamma Y_{it-1} - U_{it}}{h(X'_{it}, Y_{it-1}; \sigma_\xi)} \right) \left. \right\} 1(Y_{it}>0).
\]

(28)

Adding the heterogeneous structure does not affect the functional form, which dominates both the dependent variables \( Y_{it}^+ \) and the unobserved covariate \( U_{it} \). The derivations in homoskedastic cases can be extended to heteroskedastic cases without difficulty.

The assumptions not related to the semi-parametric censored density functions include Assumptions 2.4(i) and 2.6. These assumptions require functional form restrictions on the conditional density function \( f_{Y_{it}^+|X_{it}, Y_{it-1}, X_{it-1}} \) of observables. With the well-known completeness from the normal distributions and the exponential families in Lemma 3.1 and Lemma 3.2, it is possible to construct two types of \( f_{Y_{it}^+|X_{it}, Y_{it-1}, X_{it-1}} \) satisfying Assumptions 2.4(i) and 2.6. Assumption 2.6 implies Assumption 2.4(i). Given a fixed \( (X_{it}, Y_{it-1}) \), suppose \( X_{it-1} \) contains an open set. If \( f_{Y_{it}^+|X_{it}, Y_{it-1}, X_{it-1}} = \phi(Y_{it}^+ - \psi_1(X_{it}, Y_{it-1}; \theta_1) - \beta \psi_2(X_{it-1})) \), then

\[
\frac{\partial}{\partial X_{it-1}} f_{Y_{it}^+|X_{it}, Y_{it-1}, X_{it-1}} = \beta \psi'_2(X_{it-1}) (Y_{it}^+ - \psi_1(X_{it}, Y_{it-1}; \theta_1) - \beta \psi_2(X_{it-1})) \phi(Y_{it}^+ - \psi_1(X_{it}, Y_{it-1}; \theta_1) - \beta \psi_2(X_{it-1}));
\]

(29)
A sufficient condition to satisfy Assumption 2.6 for this specification of $f_{Y_{it} | X_{it}, Y_{it-1}, X_{it-1}}$ is $\beta \psi_2'(X_{it-1}) \neq 0$ and the range of $\psi_2$ contains an open set, according to Lemma 3.3. On the other hand, consider

$$f_{Y_{it} | X_{it}, Y_{it-1}, X_{it-1}} = s(Y_{it}^+, X_{it}, Y_{it-1}) \tau(Y_{it}^+, X_{it}, Y_{it-1}) \tau(Y_{it}^+, X_{it}, Y_{it-1}) \tau(Y_{it}^+, X_{it}, Y_{it-1})$$

where $s(Y_{it}^+, X_{it}, Y_{it-1}) > 0$, $\tau(Y_{it}^+, X_{it}, Y_{it-1}) \neq 0$ is one-to-one in $Y_{it}^+$, and the support of $\mu$ contains an open set such that $\frac{\partial}{\partial X_{it-1}} \mu(X_{it}, Y_{it-1}, X_{it-1}) \neq 0$. This conditional density function also fulfills Assumption 2.4(i) and Assumption 2.6.

The examples in this section rely on the normality of the random shock $\xi_{it}$ and it is possible to relax the normality assumption. However, in limited dependent variable models, the key issue is comparing estimated average partial effects across different models rather than parameter estimates. These models are likely to do an appropriate job of providing average partial effects under more general settings.

4. Semiparametric Estimation and Inference

The semi-parametric censored density function (10) identified in Theorem 2.1 can be determined using Eq. (12). Optimizing certain empirical criteria in general parameter spaces produces a sieve maximum likelihood estimator (sieve MLE). The integral Eq. (12) suggests a corresponding sieve MLE:

$$\hat{(\theta, f_1)}^T = \arg \max_{(\theta, f_1) \in A_n} \frac{1}{N} \sum_{i=1}^{N} \ln \int f_{Y_{it} | X_{it}, Y_{it-1}, U_{it}; \theta}(y_{it} | x_{it}, y_{it-1}, u_{it}) f_1(x_{it}, y_{it-1}, x_{it-1}, u_{it}; \theta) du_{it},$$

using a two-period i.i.d. sample $\{ y_{it}, x_{it}, y_{it-1}, x_{it-1} \}_{i=1}^{N}$. The space $A_n$ is a sequence of approximating sieve spaces containing sieve approximations of the parameter because maximization over the whole parameter space $A$ is undesirable. In addition, $\theta$ is a finite-dimensional parameter of interest and $f_1$ is a potentially infinite-dimensional nuisance parameter or non-

\footnote{A general review of semi-parametric sieve MLE appears in Shen (1997), Chen and Shen (1998), and Ai and Chen (2003).}
parametric component that varies with $\theta$. The following subsection provides a detailed implementation of sieve approximations of the nonparametric component $f_1$.

### 4.1. Restrictions on Sieve Coefficients

As for a nonparametric series estimator of $f_{X_{it-1},X_{it-1},U_{it};\theta}$, constructing a sieve approximating series that varies with the model parameter $\theta$ is an essential issue for the proposed sieve MLE. The sieve expression of $f_{X_{it-1},X_{it-1},U_{it};\theta}$ in dynamic censored models with a lagged dependent variable consists of two different parts, $Y_{it-1} = 0$ and $Y_{it-1} > 0$, and these parts can be built according to their numerical structures. Set $f_{Y_{it-1}=0} = \text{Prob}(Y_{it-1} = 0)$.

A way to split these two parts is

$$f_{X_{it-1},X_{it-1},U_{it};\theta,\delta_1} = \begin{cases} f_{X_{it-1},U_{it}|Y_{it-1}=0} & \text{if } y = 0, \\ f_{X_{it-1},U_{it}} & \text{if } y > 0. \end{cases}$$

The corresponding density restrictions are

$$\int f_{X_{it-1},U_{it}|Y_{it-1}=0} dx_{it} dx_{it-1} du_{it} = 1, \quad \text{and} \quad \int f_{X_{it-1},U_{it}} dy_{it-1} dx_{it-1} du_{it} = 1.$$

Set $z_{1,\sigma} = \frac{x_{it-1}\beta - x_{it}\beta - u_{it}}{\sigma}$, and $z_{2,\sigma} = \frac{x_i\beta - u_{it}}{\sigma}$. For the $Y_{it-1} = 0$ part, consider

$$(f_{X_{it-1},U_{it}|Y_{it-1}=0})^{1/2} = \sum_{i=0}^{i_n} \sum_{j=0}^{j_n} \sum_{k=0}^{k_n} \hat{a}_{ijk} q_i(z_{1,\sigma}) q_j(z_{2,\sigma}) q_k(u_{it}).$$

where $q_i's$, $q_j's$, and $q_k's$ represent the orthonormal Fourier series:

$$q_0(z_1) = 1 \sqrt{1}, \quad q_i(z_1) = 1 \sqrt{1} \sin\left(\frac{i\pi}{l_1} z_1\right) \text{ or } q_i(z_1) = 1 \sqrt{1} \cos\left(\frac{i\pi}{l_1} z_1\right),$$

$$q_0(z_2) = 1 \sqrt{1}, \quad q_j(z_2) = 1 \sqrt{1} \sin\left(\frac{j\pi}{l_2} z_2\right) \text{ or } q_j(z_2) = 1 \sqrt{1} \cos\left(\frac{j\pi}{l_2} z_2\right),$$

$$q_0(u_{it}) = 1 \sqrt{1}, \quad q_k(u_{it}) = 1 \sqrt{1} \cos\left(\frac{k\pi}{l_3} u_{it}\right).$$
On the other hand, suppose that $y_{it-1} \in (0, l_4]$. Write

$$
(f_{X_{it}, Y_{it-1}>0, X_{it-1}, U_{it}})^{1/2} = \sum_{i=0}^{i_n} \sum_{j=0}^{j_n} \sum_{k=0}^{k_n} \sum_{l=0}^{l_n} \tilde{a}_{ijkl} \tilde{q}_l(z'_{1,\sigma^2}) \tilde{q}_j(z'_{2,\sigma^2}) \tilde{q}_k(u_{it} \sigma^2) \tilde{q}_l(y_{it-1} / l_4),
$$

where $z'_{1,\sigma^2} \equiv x'_{it} \beta - \gamma y_{it-1} - u_{it} \sigma^2$, $z'_{2,\sigma^2} \equiv x'_{it} \beta - \gamma y_{it-1} - u_{it} \sigma^2$, and $q_0(z_4) = \frac{1}{\sqrt{l_4}}$, $q_l(z_4) = \sqrt{\frac{2}{l_4}} \cos(l \pi z_4)$.

The density restrictions for these sieve coefficients are

$$
\sum_{i=0}^{i_n} \sum_{j=0}^{j_n} \sum_{k=0}^{k_n} (\tilde{a}_{ijkl})^2 = 1 \text{ and } f_{Y_{it-1}=0} + \sum_{i=0}^{i_n} \sum_{j=0}^{j_n} \sum_{k=0}^{k_n} \sum_{l=0}^{l_n} (\tilde{a}_{ijkl})^2 = 1. \tag{31}
$$

### 4.2. Estimating Average Partial Effects

Denote $f_{X_{it}, Y_{it-1}=0, X_{it-1}, U_{it}; \hat{\theta}, \hat{\delta}_1}$ as the sieve MLE of the initial joint distribution $f_{X_{it}, Y_{it-1}=0, X_{it-1}, U_{it}}$ in the dynamic tobit model, where $\hat{\theta}$ is the estimated finite dimensional parameter of the proposed sieve MLE. This parameter can be used to obtain the sieve approximations of the structural functions in the dynamic tobit models are based on

$$
\hat{f}_{U_{it}} = \int_{X_{it}} \int_{Y_{it-1}} \int_{X_{it-1}} f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}; \hat{\theta}, \hat{\delta}_1} dx_{it} dy_{it-1} dx_{it-1} \tag{32}
$$

Therefore, under the assumptions made in Theorem 2.1 it is possible to consistently estimate average partial effects at interesting values of the explanatory variables. The average structural functions in the dynamic tobit models are based on

$$
\hat{\mu}(X_{it}, Y_{it-1}) = \int_{U_{it}} \left( \int_{Y_{it}} \max (0, y_{it}) f_{Y_{it}|X_{it}, Y_{it-1}, U_{it}; \hat{\theta}, \hat{\delta}_1} dY_{it} \right) \hat{f}_{1, U_{it}} dU_{it} \\
= \int_{U_{it}} \Phi \left( X_{it} \hat{\beta} + \hat{\gamma} Y_{it-1} + U_{it} \right) \left( X_{it} \hat{\beta} + \hat{\gamma} Y_{it-1} + U_{it} \right) + \sigma^2 \phi \frac{X_{it} \hat{\beta} + \hat{\gamma} Y_{it-1} + U_{it}}{\sigma^2} \right] \hat{f}_{1, U_{it}} dU_{it}. \tag{33}
$$

The magnitude of state dependence or average partial effect from $Y_0 = 0$ to $Y_1$ at interesting values of the explanatory variable $X_{it}$ can be measured by the difference

$$
\hat{\mu}(X_{it}, Y_1) - \hat{\mu}(X_{it}, Y_0 = 0). \tag{34}
$$
On the other hand, the average partial effect (APE) of a continuous explanatory variable can be defined using derivatives of the average structural functions in Eq. (33).

### 4.3. Inference

This study addresses two inference problems. One is to provide standard errors of the estimated finite dimensional parameter \( \hat{\theta} \) and the other one is to test the completeness of Assumptions 2.4(i) and 2.6. The parameter \( \hat{\theta} \) is estimated by optimization methods for the integrated nonlinear objective function in Eq. (30). Obtaining a formula of standard errors for this sieve MLE estimator is somewhat complicated, but it can be done using bootstrapping. Chen, Linton, and Van Keilegom (2003) studies sufficient conditions for the consistency and asymptotic normality of a class of semiparametric optimization estimators in which the criterion function does not obey the standard smoothness condition. Their results prove the validity of the bootstrap method for estimating correct confidence regions for the finite dimensional parameter \( \theta \) asymptotically.

On the other hand, because the key conditional distribution \( f_{Y_{it}^+ | X_{it}, Y_{it-1}, X_{it-1}} \) in Assumptions 2.4(i) and 2.6 is observable, the conditions seem to be directly testable using both two periods of data. However, in most cases, the exact form of the conditional distribution is not obvious and it is difficult to test nonparametrically with continuous variables. To provide more support for the validity of the condition, this study presents a test based on a parametric setting and uses the data to estimate unknown parameters associated with the completeness under parametric specifications. Although there are many ways to model completeness, this study adopts a fairly flexible approach. The previous section provides an example of completeness in which \( f_{Y_{it}^+ | X_{it}, Y_{it-1}, X_{it-1}} \) is normally distributed. Given functions \( \psi_1 \) and \( \psi_2 \) such that \( \psi_2' \neq 0 \), suppose \( \psi_2(X_{it-1}) \) contains an open set. Consider

\[
f_{Y_{it}^+ | X_{it}, Y_{it-1}, X_{it-1}} = \phi(Y_{it}^+ - \psi_1(X_{it}, Y_{it-1}; \theta_1) - \beta \psi_2(X_{it-1})).
\]

As shown in Eq. (29), under the parametric representation, the completeness in both Assumptions 2.4(i) and 2.6 depends on whether the parameter \( \beta \psi_2'(X_{it-1}) \) is equal to zero. The

\[14\text{See Canay, Santos, and Shaikh (2011) for details.}\]
null hypothesis that $f_{Y_{it}|X_{it},Y_{it-1},X_{it-1}}$ fails Assumptions $2.4(i)$ and $2.6$ is
\[ H_0 : \beta = 0. \]

When $X_{it}, X_{it-1}$ are vectors of covariates, it is possible to conduct these tests on continuously distributed covariates. If we model $\psi_1(X_{it}, Y_{it-1}; \theta_1)$ as a linear function of the regressors, a standard $t$ statistics to test $H_0$ can be conducted using OLS regression.

5. Monte Carlo Simulation

This study presents the finite sample property of the proposed sieve MLE estimators based on a Monte Carlo simulation study. The simulation design in this study is similar to the dynamic panel data models discussed by [Shiu and Hu (2010)](#). The proposed model adopts the following procedure:

**Dynamic tobit Models with AR(1) Transitory Error:**

\[ Y_{it} = \max \{ \beta_0 + \beta_1 X_{it} + \gamma Y_{it-1} + V_i + \varepsilon_{it}, 0 \} \quad i = 1, ..., N; t = 1, ..., T - 1. \]  
(35)

where $V_i \sim N(1, 1/2)$ and $\varepsilon_{it} = \rho \varepsilon_{it-1} + \xi_{it}$ with $\xi_{it} \sim N(0, \sigma^2_\xi)$. The unobserved covariate $U_{it} = V_i + \rho \varepsilon_{it-1}$. As discussed earlier, the models can be transformed into $f_{Y_{it}|X_{it}, Y_{it-1}, U_{it}; \theta}$ by Eq. (23). Set $h(x) = 0.2 \exp(-x)$. The generating processes of the covariate evolution have the following form $X_{it+1} = X_{it} + h(X_{it}) \eta_{it} + U_{it}$ with $\eta_{it} \sim N(0, 1)$ or

\[ f_{X_{it+1}|X_{it}, U_{it}}(x_{it+1}|x_{it}, u_{it}) = \frac{1}{h(x_{it})} \phi_\eta \left( \frac{x_{it+1} - x_{it} - u_{it}}{h(x_{it})} \right), \]

where $\phi_\eta$ is the standard normal distribution. Four different data generating processes (DGP) are as follows:

**DGP I:** $(\beta_0, \beta_1, \gamma, \sigma^2_\xi, \rho) = (0.2, -1, 0, 0.5, 0)$

**DGP II:** $(\beta_0, \beta_1, \gamma, \sigma^2_\xi, \rho) = (0.2, -1, 0, 0.5, 0.5)$

**DGP III:** $(\beta_0, \beta_1, \gamma, \sigma^2_\xi, \rho) = (0.2, -1, 1, 0.5, 0.5)$

**DGP IV:** $(\beta_0, \beta_1, \gamma, \sigma^2_\xi, \rho) = (0.2, -1, 1, 0.5, -0.5)$. 

24
In the all designs, set $\beta_0 = 0.2$, $\beta_1 = -1$ and $\sigma^2_\xi = 0.5$. These designs focus on different values for state dependence, $\gamma$, and AR(1) coefficient of the serially correlated error term, $\rho$. The simulation designs in DGPs I & II do not have state dependence ($\gamma = 0$), but the simulation designs in DGPs III & IV show strong persistent effects from the past dependent variable ($\gamma = 1$). This study assumes that the panel data is set in three different observations, 250, 500, 1000, and presents experiments for small $T$ for which the sampling data are drawn over $T = 3$ periods.

The two main differences between this experiment and the study by Shiu and Hu (2010) is that 1) at least three periods of data are needed for the simulation in Shiu and Hu (2010), and 2) the generating processes of the covariate evolution $f_{X_{it+1}|X_{it},U_{it}}$ is required to satisfy a mode condition, which is one of the nonparametric identification assumptions in Shiu and Hu (2010). Thus, the current estimation results cannot be compared to the results of Shiu and Hu (2010) if we only use a two-period simulated sample. Another practical advantage of the proposed sieve MLE estimator is that it does not require sieve implementation of the covariate evolution. Hence, the implementation is easier but the normality assumption of $\xi_{it}$ is required and essential to the estimation. The proposed method uses the Fourier series in Subsection 4.2 with the number of term, $i_n = 5$, $j_n = 2$, $k_n = 2$, and $l_n = 2$, to approximate the initial joint distribution $f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta}$. Two-period or three-period simulated samples are used to conduct estimations for 100 replications.

Tables 1-3 present simulation results for the model parameters of the dynamic tobit model. These tables present the means and the medians in estimating $\beta_0$, $\beta_1$, $\sigma^2_\xi$, and $\gamma$ together with their standard errors. In calculating the standard error of coefficient estimators, use the variance of coefficients estimated from the 100 replications as a measure of true variance. Observe the following patterns in the simulation results for the coefficient estimators. First, there generally exists downward bias in estimating the autoregressive parameter $\gamma$ in all sample sizes. Second, the bias in estimating all parameters ($\beta_0, \beta_1, \sigma^2_\xi$) is small, suggesting that the proposed sieve MLE estimator achieves consistent estimation results. Finally, all standard errors do not vary much. This suggests that for these samples, the statistical performances of the proposed sieve MLE estimator are very close. This study presents a comparison of the proposed sieve MLE estimators with the benchmark estimator and the three-period estimator of Shiu and Hu (2010). The benchmark estimator treats the unobserved $U_{it}$ as a covariate.
and applies a MLE method. As expected, this benchmark estimator performs better than the other two estimators, and the biases and standard errors decrease in larger samples. In DPGs with nontrivial state dependence, the three-period estimators provide better estimators for \( \gamma \). This suggests that in the same sample sizes, an additional period of data may help reveal the dynamic structure of the data.

Table 4 presents the magnitude of the state dependence \( SD(\bar{X}_t) \). The results in DGP I & II imply that the proposed sieve MLE performs well because the parameter \( \gamma = 0 \) in these cases and the estimation results are close to zero. The DGPs with nontrivial positive state dependence exhibit significant variation in the estimation results of \( SD(\bar{X}_t) \), and the average response of the state dependence for these DGPs is at least approximately 0.5. In addition, the means and medians of \( SD(\bar{X}_t) \) are similar, reflecting some symmetry in their respective distributions.

6. Empirical Application

This study reports the application of the proposed sieve MLE estimator to a censored dynamic tobit model describing the annual health expenditures of individuals given their past health expenditures and other covariates. In this case, the dependent variables represent the log values of annual individual medical expenditures plus one. To accommodate the piles of the corner outcomes, this censored dynamic tobit model is a natural fit for this health expenditure topic. Identification results show that the proposed model has some advantages: (i) arbitrary correlation between unobserved time invariant factors, such as individual inherent health and other explanatory variables, and (ii) allowing the absence of initial observations of individual health expenditures. In addition, the proposed sieve MLE estimator only requires two periods of data and provides average partial effects.

The empirical analysis in this study is based on the Medical Expenditure Panel Survey (MEPS) Panel 4. The MEPS data provides nationally representative information on health care use, expenditures, sources of payment, and insurance coverage for the U.S. population from 1,999 to 2,000. The MEPS, which contains detailed data on annual total health care expenditure, demographic characteristics, health conditions, health status, use of medical care services, and income, is appropriate for our empirical application. Table 5 presents
summary statistics of health insurance variables, socioeconomic variables, and health status regressors for the first-year and the second-year of the data. We have a two-periods of the data with 7,669 cross-sectional observations. There are sizable fractions of the sample with zero medical expenditure, 18.646% (1,430/7,669) and 20.576% (1,578/7,669) in Periods 1 and 2, respectively.

The estimated equation of dynamic health expenditures is

\[
Y_{it} = \max \left\{ 0, X_{it}'\beta + \gamma Y_{it-1} + V_i + \eta_{it} + \xi_{it} \right\} \quad \forall i = 1, \ldots, N; t = 1, 2, \tag{36}
\]

\[
= \max \left\{ 0, X_{it}'\beta + \gamma Y_{it-1} + U_{it} + \xi_{it} \right\}
\]

The dependent variable \( Y_{it} = \ln exp_{it} \) is the natural logarithm of health expenditure plus one. The covariate \( X_{it} = \left( \ln inc_{it}, \ln fam_{it}, Age_{it}, Male_{it}, Black_{it}, Education_{it}, Phsical_{it}, Ndental_{it}, Good_{it}, Fair_{it}, Poor_{it}, ..., Time \; dummies \right) \). The unobserved heterogeneity \( V_i \) represents time-invariant individual heterogeneity factors, such as inherent ability or personal regimen to resist negative health shock. Assume that Assumptions 2.1 and 2.2 split the transitory error term \( \varepsilon_{it} \) into \( \eta_{it} \) and \( \xi_{it} \), and that \( \xi_{it} \) is normally distributed. This normality assumption guarantees that Assumptions 2.3, 2.4(ii), and 2.5 are fulfilled, as Section 3 shows. Assumptions 2.4(i) and 2.6 demand the completeness conditions related to the family of conditional distribution of positive health expenditure \( y_{it}^+ \) over \( x_{it-1} \), \( \{ f_{Y_{it}^+\mid X_{it}, Y_{it-1}, X_{it-1}} : x_{it-1} \in X_{it-1} \} \). Choose \( \psi_1 \) as a linear function and \( \psi_2 \) as a linear function of squares to conduct the testing proposed in Subsection 4.3. The estimated coefficient of the squared \( \ln inc_{it-1} \) is -0.0014 with a p-value of 0.052. In an intuitive sense, the testing result suggests that the covariate at the previous period \( x_{it-1} \) containing income squares has enough variation such that the conditional distribution of positive health expenditure can cover all variation of positive health expenditures. These assumptions, along with the mild regularity condition stated in Theorem 2.1, provide the identification of Model (36) and the sieve MLE developed in Section 4 is applicable.

Table 6 shows the results of the estimation of panel data Model (36) using three specifications, including a static linear fixed effect model (Column 1), a static tobit model with random effect (Column 2), and the semiparametric dynamic tobit model (Column 3). The
three sets of estimates present similar results in terms of directions of effects and estimated coefficients. As expected, there are differences in the magnitudes of the estimated APEs in the RE tobit and semiparametric dynamic tobit specifications. The APEs of semiparametric dynamic tobit specifications have greater effects after controlling for the dynamic effect of health expenditures. The coefficient estimates of state dependence effect of health expenditures is up to 1.052. As a result, the effect of previous health expenditures on the future health expenditures is estimated to be, in APE, 1.448. The estimated coefficient shows that the previous health expenditures have persistent effects or there is large first order state dependence of health expenditures. One of the variables of interest is $Lninc_{it}$, the natural logarithm of the family income plus one. The coefficient of $Lninc_{it}$ in regression on $Lninc_{it}$ represents the income elasticity of demand for health care. The result of the semiparametric dynamic tobit specifications indicates that individuals consume more health care when their incomes go up after controlling for the past health expenditures.

7. Conclusion

This study presents identification results for the semi-parametric censored dynamic panel data models and their corresponding average partial effects. The main assumptions of the proposed method include the existence of an independent random shock, a semi-parametric specification of the random shock, and the completeness of families of known PDFs corresponding to censored dynamic panel data models and observed conditional density functions between the dependent and explanatory variables. The completeness of the families of PDFs is equivalent to the invertibility of operators using these PDFs as kernel functions. Invertibility permits the nontrivial transformation of semi-parametric censored dynamic panel data models into a valid semi-parametric family of PDFs of observables. Then, identification can be achieved under the MLE framework. The dynamic tobit models and dynamic lognormal hurdle models with two common types of DGPs satisfy these completeness conditions. This identification leads to the proposed sieve MLE, which is consistent and asymptotically normal. The advantage of the proposed approach is that it does not rely on the availability of initial period data, provides average partial effects, and requires only two-period data. In addition, this semi-parametric method allows for time dummies, nonlinear functions of state dependence $Y_{it-1}$.
such as quadratics or interaction terms, and parametric heteroskedasticity. These features make the sieve MLE desirable in semi-parametric censored dynamic panel data models for microeconometric applications.

Appendix

A. Proof of Lemma 2.1

Proof: First, we have shown \( f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0} = f_{X_{it},Y_{it-1},X_{it-1},U_{it}} \). Next, given \((x_{it}, y_{it-1})\), define integral operators

\[
L_{f_{y_{it}^+,X_{it},Y_{it-1},X_{it-1}}}: \mathcal{L}^2(\tilde{Y}_{it}) \to \mathcal{L}^2(\mathbb{X}_{it-1}) \quad \text{with} \quad (L_{f_{y_{it}^+,X_{it},Y_{it-1},X_{it-1}}}(h))(x_{it-1}) = \int f_{y_{it}^+,X_{it},Y_{it-1},X_{it-1}}(y_{it}^+,x_{it},y_{it-1},x_{it-1})h(y_{it}^+)dy_{it}^+,
\]

\[
\tilde{L}_{f_{y_{it}^+|X_{it},Y_{it-1},U_{it};\theta_0}}: \mathcal{L}^2(\tilde{Y}_{it}) \to \mathcal{L}^2(U_{it}, \omega) \quad \text{with} \quad (\tilde{L}_{f_{y_{it}^+|X_{it},Y_{it-1},U_{it};\theta_0}}(h))(u_{it}) = \int f_{y_{it}^+|X_{it},Y_{it-1},U_{it};\theta_0}(y_{it}^+|x_{it},y_{it-1},u_{it})h(y_{it}^+)dy_{it}^+,
\]

\[
L_{f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0}}: \mathcal{L}^2(U_{it}, \omega) \to \mathcal{L}^2(\mathbb{X}_{it-1}) \quad \text{with} \quad (L_{f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0}}(h))(x_{it-1}) = \int \frac{f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0}}{\omega(u_{it})} \rho(u_{it})\omega(u_{it})du_{it}.
\]

For each \( h \in \mathcal{L}^2(\tilde{Y}_{it}) \),

\[
\left( L_{f_{y_{it}^+,X_{it},Y_{it-1},X_{it-1}}} \right)(h)(x_{it-1}) = \int \tilde{Y}_{it} f_{y_{it}^+,X_{it},Y_{it-1},X_{it-1}} h(y_{it}^+)dy_{it}^+
\]

\[
= \int \tilde{Y}_{it} \left( \int f_{y_{it}^+|X_{it},Y_{it-1},U_{it};\theta_0} \frac{f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0}}{\omega(u_{it})} \omega(u_{it})du_{it} \right) h(y_{it}^+)dy_{it}^+
\]

\[
= \int f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0} \frac{f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0}}{\omega(u_{it})} \left( \int \tilde{Y}_{it} f_{y_{it}^+|X_{it},Y_{it-1},U_{it};\theta_0} h(y_{it}^+)dy_{it}^+ \right) \omega(u_{it})du_{it}
\]

\[
= \left( L_{f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0}} \tilde{L}_{f_{y_{it}^+|X_{it},Y_{it-1},U_{it};\theta_0}} \right)(h)(x_{it-1}),
\]

29
based on Eq. (12). Because this derivation holds for arbitrary \( h \), this amounts to the operator relationship

\[
\frac{L_f^{x_{it}}}{y_{it}, Y_{it-1}, X_{it-1}} = L_f^{x_{it}}, Y_{it-1}, X_{it-1}, U_{it}, \theta_0 \left( \frac{\tilde{f}_Y^{x_{it}}}{Y_{it}, Y_{it-1}, U_{it}, \theta_0} \right) 
\]

Combining the condition \( f_{X_{it}, Y_{it-1}, X_{it-1}} > 0 \) and Assumption 2.4(i) results in \( \{ f_{Y_{it}}, X_{it}, Y_{it-1}, X_{it-1} : x_{it-1} \in X_{it-1} \} \), is complete over \( L^2(\tilde{Y}_{it}) \) and then \( Lf_{Y_{it}}, X_{it}, Y_{it-1}, X_{it-1} \) is invertible. In addition, because Assumption 2.4(ii) ensures the operator \( \tilde{f}_Y^{x_{it}} \) invertible, the operator relationship implies that the invertibility of the operator \( Lf_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}, \theta_0} \), i.e.,

\[
\{ \frac{f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}, \theta_0}}{\omega(u_{it})} : x_{it-1} \in X_{it-1} \} \text{ is complete over } L^2(U_{it}, \omega).
\]

Suppose that the parameter \( \theta_0 \) is not locally identifiable. Then, there exists \( \theta_k \neq \theta_0 \) and \( \theta_k \rightarrow \theta_0 \) such that \( f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}, \theta_0} = f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}, \theta_k} \). Using the definition of \( f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}, \theta_0} \) and \( f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}, \theta_k} \),

\[
f_{Y_{it}}^{x_{it}, Y_{it-1}, X_{it-1}} = \int f_{Y_{it}}^{x_{it}, Y_{it-1}, U_{it}, \theta_0} f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}, \theta_0} du_{it}, \tag{40}
\]

\[
f_{Y_{it}}^{x_{it}, Y_{it-1}, X_{it-1}} = \int f_{Y_{it}}^{x_{it}, Y_{it-1}, U_{it}, \theta_k} f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}, \theta_k} du_{it}. \tag{41}
\]

By subtracting Eq. (41) from Eq. (40), it follows that

\[
0 = \int f_{Y_{it}}^{x_{it}, Y_{it-1}, U_{it}, \theta_k} f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}, \theta_k} - f_{Y_{it}}^{x_{it}, Y_{it-1}, U_{it}, \theta_0} f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}, \theta_0} du_{it},
\]

\[
= \int f_{Y_{it}}^{x_{it}, Y_{it-1}, U_{it}, \theta_k} f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}, \theta_k} - f_{Y_{it}}^{x_{it}, Y_{it-1}, U_{it}, \theta_0} f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}, \theta_0} du_{it}
\]

\[
+ \int f_{Y_{it}}^{x_{it}, Y_{it-1}, U_{it}, \theta_k} f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}, \theta_0} - f_{Y_{it}}^{x_{it}, Y_{it-1}, U_{it}, \theta_0} f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}, \theta_0} du_{it},
\]

\[
= \int f_{Y_{it}}^{x_{it}, Y_{it-1}, U_{it}, \theta_k} (f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}, \theta_0} - f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}, \theta_0}) du_{it}
\]

\[
+ \int (f_{Y_{it}}^{x_{it}, Y_{it-1}, U_{it}, \theta_k} - f_{Y_{it}}^{x_{it}, Y_{it-1}, U_{it}, \theta_0}) f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}, \theta_0} \frac{f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}, \theta_0}}{\omega(u_{it})} du_{it}. \tag{42}
\]

Plugging the relation \( f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}, \theta_0} = f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}, \theta_k} \) into the above equation yields

\[
0 = \int (f_{Y_{it}}^{x_{it}, Y_{it-1}, U_{it}, \theta_k} - f_{Y_{it}}^{x_{it}, Y_{it-1}, U_{it}, \theta_0}) \frac{f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}, \theta_0}}{\omega(u_{it})} \omega(u_{it}) du_{it},
\]
for all $x_{it-1}$ in $X_{it-1}$. Because Assumptions 2.4(i) & (ii) implies that $\{ \frac{f_{X_{it},Y_{it-1|X_{it-1}},U_{it};\theta_0}}{\omega(u_{it})} : x_{it-1} \in X_{it-1} \}$ is complete over $L^2(U_{it}, \omega)$, we obtain $f_{Y_{it+1}^+|X_{it},Y_{it-1},U_{it};\theta_0} = f_{Y_{it+1}^+|X_{it},Y_{it-1},U_{it};\theta_k}$ for $\theta_k \neq \theta_0$ and $\theta_k \mapsto \theta_0$. This contradicts to the local identifiability of $\theta_0$ in $f_{Y_{it+1}^+|X_{it},Y_{it-1},U_{it};\theta}$ proving the lemma. 

Q.E.D.

B. Proof of Lemma 2.2

Proof: Because $f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta}$ is correctly specified at $\theta_0$ by Lemma 2.1 $\tilde{f}_{X_{it},Y_{it-1},X_{it-1};\theta}$ is also correctly specified at $\theta_0$ after integrating out. On the other hand, denote two integral kernels as $K_{A;\theta_0}(x_{it}, y_{it-1}, x_{it-1}, u_{it}) \equiv \frac{1}{\omega(u_{it})} \frac{\partial}{\partial \theta_0} f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0}$ and $K_{B;\theta_0}(y_{it}^+, x_{it}, y_{it-1}, u_{it}) \equiv \frac{\partial}{\partial \theta_0} f_{Y_{it+1}^+|X_{it},Y_{it-1},U_{it};\theta_0}$. Divide Eq. (42) by $\theta - \theta_0 \neq 0$ and rewrite it as follows:

$$0 = \int f_{Y_{it+1}^+|X_{it},Y_{it-1},U_{it};\theta} \frac{1}{\omega(u_{it})} \frac{f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta} - f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0}}{\theta - \theta_0} \omega(u_{it})du_{it}$$

$$+ \int f_{Y_{it+1}^+|X_{it},Y_{it-1},U_{it};\theta} - f_{Y_{it+1}^+|X_{it},Y_{it-1},U_{it};\theta_0} \frac{f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0}}{\omega(u_{it})} \omega(u_{it})du_{it}.$$ 

If $\theta \mapsto \theta_0$ then the above equation implies

$$0 = \int f_{Y_{it+1}^+|X_{it},Y_{it-1},U_{it};\theta_0} K_{A;\theta_0}(x_{it}, y_{it-1}, x_{it-1}, u_{it}) \omega(u_{it})du_{it}$$

$$+ \int K_{B;\theta_0}(y_{it}^+, x_{it}, y_{it-1}, u_{it}) \frac{f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0}}{\omega(u_{it})} \omega(u_{it})du_{it}. \quad (43)$$

This equation can be used to establish an operator relationship. For each given $(x_{it}, y_{it-1})$, define integral operators as follows

$$L_{K_{A;\theta_0}} : L^2(U_{it}, \omega) \rightarrow L^2(X_{it-1}) \quad \text{with} \quad (L_{K_{A;\theta_0}} h)(x_{it-1}) = \int \frac{1}{\omega(u_{it})} \frac{\partial}{\partial \theta_0} f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0}(x_{it}, y_{it-1}, x_{it-1}, u_{it}) h(u_{it}) \omega(u_{it})du_{it}, \quad (44)$$

$$L_{K_{B;\theta_0}} : L^2(\tilde{Y}_{it}) \rightarrow L^2(U_{it}, \omega) \quad \text{with} \quad (L_{K_{B;\theta_0}} h)(u_{it}) = \int \frac{\partial}{\partial \theta_0} f_{Y_{it+1}^+|X_{it},Y_{it-1},U_{it};\theta_0}(y_{it}^+|x_{it}, y_{it-1}, u_{it}) h(y_{it}^+)dy_{it}. \quad (45)$$

31
Set \( h \in \mathcal{L}^2(\mathcal{X}_{it-1}) \). Given each \((x_{it}, y_{it-1})\),

\[
\left( L_{K_{A \theta_0}} \bar{L} f_{Y_{it}^+ | X_{it}, Y_{it-1}, U_{it}; \theta_0} \right)(h)(x_{it-1})
= \int_{U_{it}} K_{A \theta_0}(x_{it}, y_{it-1}, x_{it-1}, u_{it}) \left( \int_{Y_{it}} \bar{f}_{Y_{it}^+ | X_{it}, Y_{it-1}, U_{it}; \theta_0} h(y_{it}^+) dy_{it}^+ \right) \omega(u_{it}) du_{it}
= \int_{Y_{it}} \left( \int_{U_{it}} f_{Y_{it}^+ | X_{it}, Y_{it-1}, U_{it}; \theta_0} K_{A \theta_0}(x_{it}, y_{it-1}, x_{it-1}, u_{it}) \omega(u_{it}) du_{it} \right) h(y_{it}^+) dy_{it}^+
= -\int_{Y_{it}} \left( \int_{U_{it}} K_{B \theta_0}(y_{it}^+, x_{it}, y_{it-1}, u_{it}) \frac{f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}; \theta_0}}{\omega(u_{it})} \omega(u_{it}) du_{it} \right) h(y_{it}^+) dy_{it}^+
= -\int_{U_{it}} \int_{Y_{it}} \left( \int_{Y_{it}} K_{B \theta_0}(y_{it}^+, x_{it}, y_{it-1}, u_{it}) h(y_{it}^+) dy_{it}^+ \right) \omega(u_{it}) du_{it}
= -\left( L f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}; \theta_0} L_{K_{B \theta_0}} \right)(h)(x_{it-1})
\]

where we have used (i) an interchange of the order of integration (justified by Fubini’s theorem), (ii) Eq. \(43\), (iii) the definitions of these operators in Eqs. \(38\), \(39\), \(44\), and \(45\). This derivation yields the following operator relationship

\[
L_{K_{A \theta_0}} \bar{L} f_{Y_{it}^+ | X_{it}, Y_{it-1}, U_{it}; \theta_0} + L f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}; \theta_0} L_{K_{B \theta_0}} = 0. \quad (46)
\]

Whereas Assumptions \(2.4\) (i) & (ii) imply that \( \bar{L} f_{Y_{it}^+ | X_{it}, Y_{it-1}, U_{it}; \theta_0} \) and \( L f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}; \theta_0} \) are invertible,\(^{15}\) Assumption \(2.5\) guarantees that \( L_{K_{B \theta_0}} \) is invertible. Because the operators other than \( L_{K_{A \theta_0}} \) in Eq. \(46\) are all invertible, the integral operator \( L_{K_{A \theta_0}} \) is also invertible. This implies that the family of its corresponding kernel functions \( \{ \frac{1}{\omega(u_{it})} \frac{\partial}{\partial \theta} f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}; \theta_0} : x_{it-1} \in \mathcal{X}_{it-1} \} \) is complete over \( \mathcal{L}^2(\mathcal{U}_{it}, \omega) \).

Suppose \( \theta_0 \) is not locally identifiable in \( f_{X_{it}, Y_{it-1}, X_{it-1}; \theta} \). This implies that there exists \( \theta_k \neq \theta_0 \) and \( \theta_k \mapsto \theta_0 \) such that \( \int_{X_{it}, Y_{it-1}, X_{it-1}; \theta_k} (x_{it}, y_{it-1}, x_{it-1}) = \int_{X_{it}, Y_{it-1}, X_{it-1}; \theta_0} (x_{it}, y_{it-1}, x_{it-1}). \)

This implies that \( \int f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}; \theta_k} du_{it} = \int f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}; \theta_0} du_{it} \) for each \( \theta_k \). It follows that for each \( \theta_k \)

\[
\int \frac{1}{\omega(u_{it})} \left( f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}; \theta_k} - f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}; \theta_0} \right) \frac{\omega(u_{it}) du_{it}}{\theta_k - \theta_0} = 0 \quad \text{for all } x_{it-1}.
\]

\(^{15}\)Assumptions \(2.4\) (i) & (ii) imply that the invertibility of \( L f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}; \theta_0} \) is in the proof of Lemma 2.1.
If \( \theta_k \mapsto \theta_0 \), the equation becomes

\[
\int \left( \frac{1}{\omega(u_{it})} \frac{\partial}{\partial \theta} f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}; \theta_0} \right) \omega(u_{it}) du_{it} = 0 \quad \text{for all } x_{it-1}.
\] (47)

Because \( \mathcal{L}^2(\mathcal{U}_{it}, \omega) \) contains the constant function, Eq. (47) is in contradiction with the completeness of \( \left\{ \frac{1}{\omega(u_{it})} \frac{\partial}{\partial x_{it-1}} f_{U_{it}|X_{it}, Y_{it-1}, X_{it-1}} : x_{it-1} \in \mathcal{X}_{it-1} \right\} \) over \( \mathcal{L}^2(\mathcal{U}_{it}, \omega) \). Therefore, under Assumptions 2.3-2.5, \( \theta_0 \) is locally identifiable. Q.E.D.

C. Proof of Lemma 2.3

Before proving Lemma 2.3, consider the following result as the cornerstone of the proof of Lemma 2.3.

**Lemma C.1.** Under Assumptions 2.3-2.6, the family of functions \( \left\{ \frac{1}{\omega(u_{it})} \frac{\partial}{\partial x_{it-1}} f_{U_{it}|X_{it}, Y_{it-1}, X_{it-1}} : x_{it-1} \in \mathcal{X}_{it-1} \right\} \) is complete over \( \mathcal{L}^2(\mathcal{U}_{it}, \omega) \).

**Proof:** In a similar manner to Eq. (12), write the conditional version of Eq. (12) for \( \theta = \theta_0 \),

\[
f_{Y_{it}^+|X_{it}, Y_{it-1}, X_{it-1}} = \int f_{Y_{it}^+|X_{it}, Y_{it-1}, U_{it}} f_{U_{it}|X_{it}, Y_{it-1}, X_{it-1}} du_{it}.
\] (48)

Taking derivative with respect to \( X_{it-1} \) results in

\[
\frac{\partial}{\partial x_{it-1}} f_{Y_{it}^+|X_{it}, Y_{it-1}, X_{it-1}} = \int f_{Y_{it}^+|X_{it}, Y_{it-1}, U_{it}} \frac{\partial}{\partial x_{it-1}} f_{U_{it}|X_{it}, Y_{it-1}, X_{it-1}} du_{it}.
\] (49)

Set \( \kappa_1 = \frac{\partial}{\partial x_{it-1}} f_{Y_{it}^+|X_{it}, Y_{it-1}, X_{it-1}} \) and \( \phi = \frac{\partial}{\partial x_{it-1}} f_{U_{it}|X_{it}, Y_{it-1}, X_{it-1}} \). For each \( (x_{it}, y_{it-1}) \), define operators

\[
L_{\kappa_1} : \mathcal{L}^2(\tilde{Y}_{it}) \to \mathcal{L}^2(\mathcal{X}_{it-1}) \quad \text{with} \quad (L_{\kappa_1} h)(x_{it-1}) = \int \kappa_1(y_{it}^+, x_{it}, y_{it-1}, x_{it-1}) h(y_{it}^+) dy_{it}^+,
\]

\[
L_{\phi} : \mathcal{L}^p(\mathcal{U}_{it}, \omega) \to \mathcal{L}^2(\mathcal{X}_{it-1}) \quad \text{with} \quad (L_{\phi} h)(x_{it-1}) = \int \frac{1}{\omega(u_{it})} \phi(u_{it}, x_{it}, y_{it-1}, x_{it-1}) h(u_{it}) \omega(u_{it}) du_{it}.
\]
For $h \in L^2(\tilde{Y}_{it})$.

$$(L_{\kappa_1})(h)(x_{it-1})$$

$$= \int_{\tilde{Y}_{it}} \kappa_1(y_{it}^+, x_{it}, y_{it-1}, x_{it-1})h(y_{it}^+)dy_{it}^+)$$

$$= \int_{\tilde{Y}_{it}} \left( \int_{U_{it}} F_{it}^+ |X_{it},Y_{it-1},U_{it},\phi(u_{it},x_{it},y_{it-1},x_{it-1})du_{it} \right) h(y_{it}^+)y_{it}^+ dy_{it}^+$$

$$= \int_{U_{it}} \frac{1}{\omega(u_{it})} \phi(u_{it},x_{it},y_{it-1},x_{it-1}) \left( \int_{\tilde{Y}_{it}} F_{it}^+ |X_{it},Y_{it-1},U_{it}h(y_{it})dy_{it} \right) \omega(u_{it})du_{it}$$

$$= \left( L_{\phi} L_{F_{it}^+ |X_{it},Y_{it-1},U_{it},\theta_0} \right)(h)(x_{it-1}),$$

where Eq. (38) defines the operator $L_{F_{it}^+ |X_{it},Y_{it-1},U_{it},\theta_0}$. With the definitions of the operators, this equation can be rewritten as an operator relationship

$$L_{\kappa_1} = L_{\phi} L_{F_{it}^+ |X_{it},Y_{it-1},U_{it},\theta_0}.$$

Assumptions 2.4(ii) and 2.6 guarantee the invertibility of the operators $L_{F_{it}^+ |X_{it},Y_{it-1},U_{it},\theta_0}$ and $L_{\kappa_1}$, respectively. Applying this invertibility to Eq. (50) results in the invertibility of $L_{\phi}$. Thus, the family $\{\frac{1}{\omega(u_{it})} \phi(u_{it},x_{it},y_{it-1},x_{it-1}) : x_{it-1} \in X_{it-1}\}$ is complete over $L^2(U_{it}, \omega)$ for each $x_{it}, y_{it-1}$.

**Proof of Lemma 2.3**: First, $f_{X_{it},Y_{it-1},X_{it-1} | \theta}$ is correctly specified at $\theta_0$ because by Lemma 2.2 $\tilde{f}_{X_{it},Y_{it-1},X_{it-1} | \theta}$ is correctly specified at $\theta_0$. Suppose that $\theta_0$ is not locally identifiable in the observable joint density function $f_{X_{it},Y_{it-1},X_{it-1} | \theta}$. There exists $\theta_k \neq \theta_0$ and $\theta_k \rightarrow \theta_0$ such that $f_{X_{it},Y_{it-1},X_{it-1} | \theta_k} = f_{X_{it},Y_{it-1},X_{it-1} | \theta_0}$. This implies that

$$\int \int \tilde{f}_{X_{it},Y_{it-1},X_{it-1} | \theta_k} dx_{it}dy_{it-1}dx_{it-1} = \frac{\int \int \tilde{f}_{X_{it},Y_{it-1},X_{it-1} | \theta_0} dx_{it}dy_{it-1}dx_{it-1}}{1} = f_{X_{it},Y_{it-1},X_{it-1}}. \quad (51)$$

This equation can be expressed as

$$\int f_{X_{it},Y_{it-1},X_{it-1} | \theta_k} d\theta_k \int \int \tilde{f}_{X_{it},X_{it-1} | \theta_k} dx_{it}dy_{it-1}dx_{it-1} = \int \int \tilde{f}_{X_{it},Y_{it-1},X_{it-1} | \theta_0} dx_{it}dy_{it-1}dx_{it-1}. \quad (52)$$

The multiple integral in the RHS of Eq. (52) only depends on the parameter $\theta_k$, and is
independent of $x_{it-1}$. This suggests that given $x_{1t-1} \neq x_{2t-1},$

$$\int \frac{f_{X_{it},Y_{it-1},X_{1t-1},U_{it}}|\theta_k}}{f_{X_{it},Y_{it-1},X_{1t-1}}} du_{it} = \int \frac{f_{X_{it},Y_{it-1},X_{2t-1},U_{it}}|\theta_k}}{f_{X_{it},Y_{it-1},X_{2t-1}}} du_{it}.$$

If $\theta_k \mapsto \theta_0$, this yields

$$0 = \int \left( f_{U_{it}|X_{it},Y_{it-1},X_{1t-1}} - f_{U_{it}|X_{it},Y_{it-1},X_{2t-1}} \right) du_{it}$$

Divide the equation by $X_{1t-1} - X_{2t-1}$ and let $X_{1t-1} - X_{2t-1} \to 0$. This equation then changes into

$$0 = \int \frac{\partial}{\partial x_{it-1}} f_{U_{it}|X_{it},Y_{it-1},X_{1t-1}} du_{it} = \int \frac{1}{\omega(u_{it})} \frac{\partial}{\partial x_{it-1}} f_{U_{it}|X_{it},Y_{it-1},X_{1t-1}} \omega(u_{it}) du_{it},$$

which contradicts the completeness in Lemma C.1. Therefore, the parameter $\theta_0$ is locally identifiable in the observable joint density function $f_{X_{it},Y_{it-1},X_{1t-1};\theta}$. Q.E.D.

**D. Identification in the Discrete Case**

This section presents a simple case in which the observed variables $Y_{it}, X_{it}, Y_{it-1}, X_{it-1}$ and the unobserved covariate $U_{it}$ are all discrete. This section shows how to use the identification techniques in Theorem 2.1 for this discrete case. For simplicity, assume that the variables $Y^+_{it}, X_{it-1}$ and $U_{it}$ have the same size $J$ (i.e., $Y^+_{it}, X_{it-1}, U_{it} \in \{1, 2, ..., J\}$). For this setting, the integral operators used previously can be represented by $J$-by-$J$ matrices. The idea of using the identification strategy in the discrete case for ease of exposition is because a complete integral operators is associated with an invertible matrix.\(^{16}\)

\(^{16}\)If $y, u \in \{1, 2\}$ and $\int \frac{1}{h(u)} f(y|u) du = 0$, then the condition is equivalent to

$$\begin{bmatrix} f_{y|u}(1|1) & f_{y|u}(1|2) \\ f_{y|u}(2|1) & f_{y|u}(2|2) \end{bmatrix} \begin{bmatrix} h(1) \\ h(2) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$  

The function $h$ can be uniquely determined as $h = 0$ iff the first matrix representing $f_{y|u}$ is invertible.
Equation (12) in the discrete case is

\[ f_{Y_{it+1}^+,X_{it},Y_{it-1},X_{it-1}} = \sum_{U_{it}=1}^J f_{Y_{it+1}^+ | X_{it},Y_{it-1},U_{it}; \theta} f_{X_{it},Y_{it-1},X_{it-1},U_{it}; \theta}. \]  

(53)

Given \((x_{it}, y_{it-1})\), define \(J\)-by-\(J\) matrices

\[
\begin{align*}
M_{f_{Y_{it+1}^+,x_{it},y_{it-1},x_{it-1}}} & = \left[ f_{Y_{it+1}^+ | X_{it},Y_{it-1},X_{it-1}; \theta} (Y_{it+1}^+, x_{it}, y_{it-1}, X_{it-1}) \right]_{y_{it+1}^+,x_{it-1}} \\
L_{f_{Y_{it+1}^+,x_{it},y_{it-1},U_{it}; \theta}} & = \left[ f_{Y_{it+1}^+ | X_{it},Y_{it-1},U_{it}; \theta} (Y_{it+1}^+, x_{it}, y_{it-1}, U_{it}) \right]_{y_{it+1}^+,x_{it}} \\
M_{f_{x_{it},y_{it-1},x_{it-1},U_{it}; \theta}} & = \left[ f_{X_{it},Y_{it-1},X_{it-1},U_{it}; \theta} (x_{it}, y_{it-1}, X_{it-1}, U_{it}) \right]_{x_{it},x_{it-1}}.
\end{align*}
\]

Rewrite the equality (53) in terms of these matrices as follows:

\[
\frac{M_{f_{Y_{it+1}^+,x_{it},y_{it-1},x_{it-1}}} \text{ Observed from Data}}{L_{f_{Y_{it+1}^+,x_{it},y_{it-1},U_{it}; \theta}} \text{ Model Specification}} = \left( L_{f_{Y_{it+1}^+,x_{it},y_{it-1},U_{it}; \theta}} \right)^{-1} M_{f_{x_{it},y_{it-1},x_{it-1},U_{it}; \theta}}. \tag{54}
\]

Assumption 2.3 implies that the square matrix \(L_{f_{Y_{it+1}^+,x_{it},y_{it-1},U_{it}; \theta}}\) is invertible, leading to

\[
M_{f_{x_{it},y_{it-1},x_{it-1},U_{it}; \theta}} = \left( L_{f_{Y_{it+1}^+,x_{it},y_{it-1},U_{it}; \theta}} \right)^{-1} M_{f_{Y_{it+1}^+,x_{it},y_{it-1},x_{it-1}}}. \tag{55}
\]

As discussed earlier, it is necessary to ensure that \(M_{f_{x_{it},y_{it-1},x_{it-1},U_{it}; \theta}}\) is identifiable at \(\theta_0\). According to the proof of Lemma 2.1, there are two steps for identifiability. First, given \((x_{it}, y_{it-1})\), define

\[
\begin{align*}
L_{f_{Y_{it+1}^+,x_{it},y_{it-1},x_{it-1}}} & = \left[ f_{Y_{it+1}^+ | X_{it},Y_{it-1},X_{it-1}; \theta} (Y_{it+1}^+, x_{it}, y_{it-1}, X_{it-1}) \right]_{x_{it-1},y_{it+1}} \\
\bar{L}_{f_{Y_{it+1}^+,x_{it},y_{it-1},U_{it}; \theta}} & = \left[ f_{Y_{it+1}^+ | X_{it},Y_{it-1},U_{it}; \theta} (Y_{it+1}^+, x_{it}, y_{it-1}, U_{it}) \right]_{y_{it},x_{it}} \\
L_{f_{x_{it},y_{it-1},x_{it-1},U_{it}; \theta}} & = \left[ f_{X_{it},Y_{it-1},X_{it-1},U_{it}; \theta} (x_{it}, y_{it-1}, X_{it-1}, U_{it}) \right]_{x_{it},x_{it-1}}.
\end{align*}
\]

Equality (53) can then be expressed by these matrices as follows:

\[
\frac{L_{f_{Y_{it+1}^+,x_{it},y_{it-1},x_{it-1}}} \text{ Assumption 2.4(i)}}{\bar{L}_{f_{Y_{it+1}^+,x_{it},y_{it-1},U_{it}; \theta}} \text{ Assumption 2.4(ii)}} = L_{f_{x_{it},y_{it-1},x_{it-1},U_{it}; \theta_0}} \bar{L}_{f_{Y_{it+1}^+,x_{it},y_{it-1},U_{it}; \theta_0}}. \tag{56}
\]

36
Notice that in this simple case, \( L f_{Y_t^+|x_t,y_{it-1},1} \cdot x_{it-1} = M f_{Y_t^+|x_t,y_{it-1},1} \cdot x_{it-1} \cdot x_{it-1} \cdot x_{it-1} \cdot U_{it\theta} \), and \( L f_{Y_t^+|x_t,y_{it-1},1} \cdot x_{it-1} \cdot x_{it-1} \cdot x_{it-1} \cdot U_{it\theta} = M f_{Y_t^+|x_t,y_{it-1},1} \cdot x_{it-1} \cdot x_{it-1} \cdot x_{it-1} \cdot U_{it\theta} \) which we may not have for a general continuous case. The matrix notations used here are based on integral operators in the proofs of lemmas. Assumption 2.4 makes \( L f_{Y_t^+|x_t,y_{it-1},1} \cdot x_{it-1} \cdot x_{it-1} \cdot x_{it-1} \cdot U_{it\theta} \) invertible. Hence, its transpose \( M f_{Y_t^+|x_t,y_{it-1},1} \cdot x_{it-1} \cdot x_{it-1} \cdot x_{it-1} \cdot U_{it\theta} \) is also invertible. Then, suppose that there exists \( \theta_k \neq \theta_0 \) and \( \theta_k \rightarrow \theta_0 \) such that

\[
M f_{X_{it-1},U_{it\theta_k}} = M f_{X_{it-1},U_{it\theta_0}}. \tag{57}
\]

Following the derivation in Eq. (42), we have a matrix expression

\[
0 = L f_{Y_t^+|x_t,y_{it-1},1} \cdot x_{it-1} \cdot x_{it-1} \cdot x_{it-1} \cdot U_{it\theta_k} \left( M f_{X_{it-1},U_{it\theta_k}} - M f_{X_{it-1},U_{it\theta_0}} \right) + \left( L f_{Y_t^+|x_t,y_{it-1},1} \cdot x_{it-1} \cdot x_{it-1} \cdot U_{it\theta_k} - L f_{Y_t^+|x_t,y_{it-1},1} \cdot x_{it-1} \cdot x_{it-1} \cdot U_{it\theta_0} \right) M f_{X_{it-1},U_{it\theta_0}}.
\]

The invertibility of \( M f_{X_{it-1},U_{it\theta_0}} \) and Eq. (57) implies that \( L f_{Y_t^+|x_t,y_{it-1},1} \cdot x_{it-1} \cdot x_{it-1} \cdot U_{it\theta} \) is not identifiable at \( \theta_0 \), which is a contradiction.

Set \( J \times 1 \)-vector \( J_1 = (1, 1, ..., 1)^T \). Integrating out the unobserved covariate in the discrete case leads to \( J_1^T M f_{X_{it-1},U_{it\theta}} \). Suppose that there exists \( \theta_k \neq \theta_0 \) and \( \theta_k \rightarrow \theta_0 \) such that \( J_1^T M f_{X_{it-1},U_{it\theta_k}} = J_1^T M f_{X_{it-1},U_{it\theta_0}} \). It then follows that

\[
0 = \frac{\partial}{\partial \theta} \left( M f_{X_{it-1},U_{it\theta}} - M f_{X_{it-1},U_{it\theta_0}} \right) \frac{\partial \theta}{\partial \theta} \frac{\partial x_{it-1},y_{it-1},X_{it-1},U_{it,1}}{\partial \theta} \tag{58}
\]

If \( \theta \rightarrow \theta_0 \), the above equation implies \( 0 = J_1^T M f_{X_{it-1},U_{it\theta}} \), where

\[
M f_{X_{it-1},U_{it\theta}} = \left[ \frac{\partial}{\partial \theta} f_{X_{it-1},U_{it\theta}} \right] \frac{\partial x_{it-1},y_{it-1},X_{it-1},U_{it,1}}{\partial \theta}. \tag{59}
\]

Rewrite Eq. (43) in the discrete case as

\[
0 = \sum_{U_{it} = 1}^J f_{Y_{it}^+|x_{it},y_{it-1},U_{it\theta_0}} \frac{\partial}{\partial \theta} f_{Y_{it}^+|x_{it},y_{it-1},U_{it\theta_0}} + \sum_{U_{it} = 1}^J \frac{\partial}{\partial \theta} f_{Y_{it}^+|x_{it},y_{it-1},U_{it\theta_0}} f_{Y_{it}^+|x_{it},y_{it-1},U_{it\theta_0}} \tag{59}
\]

37
This leads to the following matrix expression

\[
M \frac{\partial}{\partial \theta} f_{x_{it},y_{it-1},X_{it-1},U_{it};\theta_0} + L f_{x_{it},y_{it-1},X_{it-1},U_{it};\theta_0} M \frac{\partial}{\partial \theta} f_{y_{it}^+|x_{it},y_{it-1},U_{it};\theta_0} = 0, \quad (60)
\]

where \( M \frac{\partial}{\partial \theta} f_{y_{it}^+|x_{it},y_{it-1},U_{it};\theta_0} = \left[ \frac{\partial}{\partial \theta} f_{y_{it}^+|x_{it},y_{it-1},U_{it};\theta_0} \right] y_{it}. \) Applying assumptions to Eq. \((60)\) shows that \( M \frac{\partial}{\partial \theta} f_{y_{it}^+|x_{it},y_{it-1},X_{it-1},U_{it};\theta_0} \) is invertible, which contradicts \( 0 = J^T M \frac{\partial}{\partial \theta} f_{x_{it},y_{it-1},X_{it-1},U_{it};\theta_0}. \)

Finally, the normalization in the discrete case is equivalent to

\[
V_{f_{x_{it},y_{it-1},X_{it-1};\theta}} = \frac{J^T M_{f_{x_{it},y_{it-1},X_{it-1};\theta}}}{\sum_{x_{it}=1}^{J_{x_{it}}} \sum_{y_{it-1}=1}^{J_{y_{it-1}}} \left( J^T M_{f_{x_{it},y_{it-1},X_{it-1};\theta}} \right)} = V_{f_{x_{it},y_{it-1},X_{it-1};\theta}} \quad (61)
\]

where \( J_{x_{it}} \) and \( J_{y_{it-1}} \) represent the sizes of the discrete variables \( x_{it} \) and \( y_{it-1} \), respectively. Suppose the normalization step does not lead to local identifiability at \( \theta_0 \). This implies that there exists \( \theta_k \neq \theta_0 \) and \( \theta_k \mapsto \theta_0 \) such that

\[
\frac{J^T M_{f_{x_{it},y_{it-1},X_{it-1};\theta_k}}}{\sum_{x_{it}=1}^{J_{x_{it}}} \sum_{y_{it-1}=1}^{J_{y_{it-1}}} \left( J^T M_{f_{x_{it},y_{it-1},X_{it-1};\theta_k}} \right)} = \frac{J^T M_{f_{x_{it},y_{it-1},X_{it-1};\theta_0}}}{\sum_{x_{it}=1}^{J_{x_{it}}} \sum_{y_{it-1}=1}^{J_{y_{it-1}}} \left( J^T M_{f_{x_{it},y_{it-1},X_{it-1};\theta_0}} \right)} = V_{f_{x_{it},y_{it-1},X_{it-1};\theta_0}} \quad (62)
\]

Rearrange the term

\[
\left( J^T M_{f_{x_{it},y_{it-1},X_{it-1};\theta_k}} \right) / V_{f_{x_{it},y_{it-1},X_{it-1};\theta_0}} = \left( \sum_{x_{it}=1}^{J_{x_{it}}} \sum_{y_{it-1}=1}^{J_{y_{it-1}}} \left( J^T M_{f_{x_{it},y_{it-1},X_{it-1};\theta_k}} \right) \right) J_1,
\]

where the notation \( / \) divides two \( 1 \times J \)-vectors element-wise. The right-hand side of this equation is constant in \( x_{it-1} \). Hence, if \( x_{1t-1} \neq x_{2t-1} \), we have

\[
\left( J^T M_{f_{x_{it},y_{it-1},X_{it-1};\theta_k}} \right) / V_{f_{x_{it},y_{it-1},X_{it-1};\theta_0}} = \left( J^T M_{f_{x_{it},y_{it-1},X_{2t-1};\theta_k}} \right) / V_{f_{x_{it},y_{it-1},X_{2t-1};\theta_0}}.
\]

Using Eq. \((55)\), rewrite this equation as

\[
0 = J^T \left( L f_{y_{it}^+|x_{it},y_{it-1},U_{it};\theta_k} \right)^{-1} \left( M_{f_{y_{it}^+|x_{it},y_{it-1},X_{it-1};\theta_k}} - M_{f_{y_{it}^+|x_{it},y_{it-1},X_{2t-1};\theta_k}} \right) / V_{f_{x_{it},y_{it-1},X_{2t-1};\theta_0}}.
\]

Denote \( M_{f_{y_{it}^+|x_{it},y_{it-1},X_{it-1};\theta}} \) as a matrix of the difference of \( f_{y_{it}^+|x_{it},y_{it-1},X_{it-1};\theta} \) with respect to
If \( \theta_k \mapsto \theta_0 \), then

\[
0 = J_1^{T} \left( L_{f_{Y_t |x_{it},y_{it-1},u_{it};\theta_0}} \right)^{-1} \left( M_{f_{Y_t |x_{it},y_{it-1},X_{it-1}} - M_{f_{Y_t |x_{it},y_{it-1},X_{it-1}} \triangle X_{it-1}}} \right)
\]

where \( M_{f_{Y_t |x_{it},y_{it-1},X_{it-1}}} \equiv \left[ f_{Y_t |x_{it},y_{it-1}}(Y_{it} | x_{it}, y_{it-1}, X_{it-1}) \right]_{y_{it-1},x_{it-1}} \). This contradicts the invertibility under Assumptions 2.3 and 2.6, showing that the density function is still locally identifiable at \( \theta_0 \).

**E. Proof of a Sufficient Condition of the Uniqueness of \( K(\theta) \)**

**Proof:** Start with a scalar \( \theta \), and combine Eqs. (16) and (17),

\[
f_{X_{it},Y_{it-1},X_{it-1};\theta} \equiv \frac{\int \int \int f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta} d\theta f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta} d\theta f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta} d\theta f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta} d\theta}{\int \int \int f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta} d\theta f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta} d\theta f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta} d\theta f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta} d\theta}.
\]

and recall Eq. (43),

\[
0 = \int f_{Y_t |x_{it},y_{it-1},U_{it};\theta_0} \frac{\partial}{\partial \theta} f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta} \big|_{\theta=\theta_0} d\theta d\theta f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta} d\theta f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta} d\theta f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta} d\theta.
\]

It follows that \( \frac{\partial}{\partial \theta} f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta} \big|_{\theta=\theta_0} \) is implicitly defined in Eq. (64) by

\[
\frac{\partial}{\partial \theta} f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0} = L_{f_{Y_t |x_{it},y_{it-1},U_{it};\theta_0}}^{-1} \left( \int \frac{\partial}{\partial \theta} f_{Y_t |x_{it},y_{it-1},U_{it};\theta_0} f_{X_{it},Y_{it-1},X_{it-1},U_{it}} d\theta_0 \right).
\]
These relationships can help expand the Kullback-Leibler function $K(\theta)$. Differentiating w.r.t. $\theta$ at $\theta = \theta_0$ leads to

$$K'(\theta_0) = \int \int \int \frac{\partial}{\partial \theta} \log (f_{X_{it},Y_{it-1},X_{it-1};\theta}) \bigg|_{\theta=\theta_0} f_{X_{it},Y_{it-1},X_{it-1}} \, dx_{it} \, dy_{it-1} \, dx_{it-1}$$

$$= \int \int \int \frac{\partial}{\partial \theta} f_{X_{it},Y_{it-1},X_{it-1};\theta} \bigg|_{\theta=\theta_0} f_{X_{it},Y_{it-1},X_{it-1}} \, dx_{it} \, dy_{it-1} \, dx_{it-1}$$

$$= \int \int \int \frac{\partial}{\partial \theta} f_{X_{it},Y_{it-1},X_{it-1};\theta} \, dx_{it} \, dy_{it-1} \, dx_{it-1}$$

$$= \frac{\partial}{\partial \theta} \left( \int \int \int f_{X_{it},Y_{it-1},X_{it-1};\theta} \, dx_{it} \, dy_{it-1} \, dx_{it-1} \right) \bigg|_{\theta=\theta_0}$$

$$= 0.$$

Thus, the true parameter $\theta_0$ is a critical point of $K(\theta_0)$ and a sufficient condition of the uniqueness of $\theta_0$ is $K''(\theta_0)$ is negative semidefinite. Denote $E$ as the expectation w.r.t. $(x_{it},y_{it-1},x_{it-1})$. The second derivative of $K'(\theta_0)$ in the scalar case is

$$K''(\theta_0)$$

$$= \int \int \int \frac{\partial^2}{\partial \theta^2} f_{X_{it},Y_{it-1},X_{it-1};\theta} \bigg|_{\theta=\theta_0} \, dx_{it} \, dy_{it-1} \, dx_{it-1}$$

$$- \int \int \int \frac{\partial}{\partial \theta} \left( \frac{\partial}{\partial \theta} f_{X_{it},Y_{it-1},X_{it-1};\theta} \right)^2 \bigg|_{\theta=\theta_0} f_{X_{it},Y_{it-1},X_{it-1}} \, dx_{it} \, dy_{it-1} \, dx_{it-1}$$

$$= \frac{\partial^2}{\partial \theta^2} \left( \int \int \int f_{X_{it},Y_{it-1},X_{it-1};\theta} \, dx_{it} \, dy_{it-1} \, dx_{it-1} \right) \bigg|_{\theta=\theta_0} - E \left[ \left( \frac{\partial}{\partial \theta} f_{X_{it},Y_{it-1},X_{it-1};\theta} \bigg|_{\theta=\theta_0} \right)^2 \right]$$

$$= -E \left[ \left( \frac{\partial}{\partial \theta} f_{X_{it},Y_{it-1},X_{it-1};\theta} \bigg|_{\theta=\theta_0} \right)^2 \right] \quad (66)$$
where

\[
\frac{\partial}{\partial \theta} f_{X_t, Y_{t-1}, X_{t-1}; \theta} \bigg|_{\theta = \theta_0} = \frac{\partial}{\partial \theta} \int \int \int f_{X_t, Y_{t-1}, X_{t-1}; \theta} dy dt dX_{t-1} \bigg|_{\theta = \theta_0} = \int \frac{\partial}{\partial \theta} f_{X_t, Y_{t-1}, X_{t-1}; \theta} dy dt dX_{t-1} 
\]

defined in Eq. \ref{65},

\[-\left(\int \int \int \frac{\partial}{\partial \theta} f_{X_t, Y_{t-1}, X_{t-1}; \theta} dy dt dX_{t-1} d\theta \right) \int \int \int f_{X_t, Y_{t-1}, X_{t-1}} \bigg|_{\theta = \theta_0} \]

defined in Eq. \ref{65}.

A similar derivation can be applied to the vector case and the form \(K''(\theta_0)\), as Eq. \ref{22} shows.

Q.E.D.

F. Proof of Lemma 3.3

Proof: First, suppose \(\tilde{Y}\) is a domain such that \(\tilde{Y} \subset \mathcal{R}\). Let the family \(\{f(y|u) : u \in \mathcal{U}\}\) be complete in \(L^2(\mathcal{R})\). For each \(h \in L^2(\tilde{Y})\) such that \(\int_{\tilde{Y}} h(y)f(y|u)dy = 0\) for all \(u\). Extend \(h\) to a function in \(L^2(\mathcal{R})\) by \(\tilde{h}(x) = \begin{cases} h & \text{if } x \in \tilde{Y}, \\ 0 & \text{otherwise.} \end{cases}\) It follows that \(\int_{\mathcal{R}} \tilde{h}(y)f(y|u)dy = 0\) for all \(u\). By the completeness of \(f(y|u)\) over \(L^2(\mathcal{R})\), \(\tilde{h} = 0\). Thus, \(h = 0\) and \(f(y|u)\) is complete over \(L^2(\mathcal{R})\). Thus, the completeness of a function over a smaller domain is implied by the completeness of the function over a larger domain, and sufficient conditions for the completeness of these two families can be reduced to the completeness in \(L^2(\mathcal{R})\).

The family of functions \(\{(y-c-u)\phi\left(\frac{y-c-u}{\sigma\xi}\right) : u \in \mathcal{U}\}\) is complete in \(L^2(\mathcal{R})\). Let \(h(y) \in L^2(\mathcal{R})\) and \(\int h(y)(y-c-u)\phi\left(\frac{y-c-u}{\sigma\xi}\right)dy = 0\) for all \(u \in \mathcal{U}\). Because \(\frac{\partial}{\partial y} \phi\left(\frac{y-c-u}{\sigma\xi}\right) = -\frac{y-c-u}{\sigma\xi} \phi\left(\frac{y-c-u}{\sigma\xi}\right)\), it follows that \(\int h(y)\frac{\partial}{\partial y} \phi\left(\frac{y-c-u}{\sigma\xi}\right)dy = 0\) for all \(u \in \mathcal{U}\). Using the integration by part for each \(u\) leads to

\[
\int h(y)\frac{\partial}{\partial y} \phi\left(\frac{y-c-u}{\sigma\xi}\right)dy = h(y)\phi\left(\frac{y-c-u}{\sigma\xi}\right) \bigg|_{\infty}^{\infty} - \int \frac{\partial}{\partial y} h(y)\phi\left(\frac{y-c-u}{\sigma\xi}\right)dy
\]

\[
= -\int \frac{\partial}{\partial y} h(y)\phi\left(\frac{y-c-u}{\sigma\xi}\right)dy.
\]

41
Applying the completeness of \( \{ \phi \left( \frac{y-c-u}{\sigma_{\xi}} \right) : u \in \mathcal{U} \} \) to this equation yields \( \frac{\partial}{\partial y} h(y) = 0 \), which implies that \( h(y) \) is a constant function. The condition \( h(y) \in L^2(\mathcal{R}) \) makes \( h(y) = 0 \), proving the first completeness. As for the second completeness, suppose \( h(y) \in L^2(\mathcal{R}) \) such that \( \int h(y) \left( \sigma_{\xi}^2 - (y-c-u)^2 \right) \phi \left( \frac{y-c-u}{\sigma_{\xi}} \right) dy = 0 \) for all \( u \in \mathcal{U} \). Using \( \frac{\partial^2}{\partial y^2} \phi \left( \frac{y-c-u}{\sigma_{\xi}} \right) = -\frac{1}{\sigma_{\xi}^3} \left( \sigma_{\xi}^2 - (y-c-u)^2 \right) \phi \left( \frac{y-c-u}{\sigma_{\xi}} \right) \) and the integration by part, rewrite the equation as

\[
0 = \int h(y) \frac{\partial^2}{\partial y^2} \phi \left( \frac{y-c-u}{\sigma_{\xi}} \right) dy
= h(y) \frac{\partial}{\partial y} \phi \left( \frac{y-c-u}{\sigma_{\xi}} \right) \bigg|_{-\infty}^{\infty} - \int \frac{\partial}{\partial y} h(y) \frac{\partial}{\partial y} \phi \left( \frac{y-c-u}{\sigma_{\xi}} \right) dy
= -\frac{\partial}{\partial y} h(y) \phi \left( \frac{y-c-u}{\sigma_{\xi}} \right) \bigg|_{-\infty}^{\infty} + \int \frac{\partial^2}{\partial y^2} h(y) \phi \left( \frac{y-c-u}{\sigma_{\xi}} \right) dy
= \int \frac{\partial^2}{\partial y^2} h(y) \phi \left( \frac{y-c-u}{\sigma_{\xi}} \right) dy
\]

The completeness of \( \{ \phi \left( \frac{y-c-u}{\sigma_{\xi}} \right) : u \in \mathcal{U} \} \) implies that \( h \) satisfies the second order differential equation, \( \frac{\partial^2}{\partial y^2} h(y) = 0 \). The characteristic equation of the differential equation is \( r^2 = 0 \). This suggests that the general solution of the differential equation is \( h(y) = c_1 + c_2 y \), where \( c_1 \) and \( c_2 \) are constants. The condition \( h(y) \in L^2(\mathcal{R}) \) indicates \( h(y) = 0 \), reaching the second completeness.

\[ Q.E.D. \]

**References**


**Arellano, M., and S. Bond (1991):** “Some Tests of Specification for Panel Data: Monte...


<table>
<thead>
<tr>
<th>DGP</th>
<th>Parameters</th>
<th>( \beta_0 )</th>
<th>( \beta_1 )</th>
<th>( \gamma )</th>
<th>( \sigma^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>DGP I:</td>
<td>True value</td>
<td>0.2</td>
<td>-1</td>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>Infeasible Mean</td>
<td>0.221</td>
<td>-1.004</td>
<td>-0.006</td>
<td>0.478</td>
</tr>
<tr>
<td></td>
<td>Infeasible Median</td>
<td>0.234</td>
<td>-1.002</td>
<td>-0.005</td>
<td>0.470</td>
</tr>
<tr>
<td></td>
<td>Standard error</td>
<td>0.139</td>
<td>0.116</td>
<td>0.108</td>
<td>0.086</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>0.202</td>
<td>-1.009</td>
<td>-0.003</td>
<td>0.499</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.211</td>
<td>-1.006</td>
<td>-0.013</td>
<td>0.500</td>
</tr>
<tr>
<td></td>
<td>Standard error</td>
<td>0.093</td>
<td>0.099</td>
<td>0.095</td>
<td>0.033</td>
</tr>
<tr>
<td></td>
<td>Three-period Mean</td>
<td>0.201</td>
<td>-0.993</td>
<td>0.020</td>
<td>0.531</td>
</tr>
<tr>
<td></td>
<td>Three-period Median</td>
<td>0.202</td>
<td>-1.009</td>
<td>0.008</td>
<td>0.497</td>
</tr>
<tr>
<td></td>
<td>Standard error</td>
<td>0.097</td>
<td>0.109</td>
<td>0.104</td>
<td>0.105</td>
</tr>
<tr>
<td>DGP II:</td>
<td>True value</td>
<td>0.2</td>
<td>-1</td>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>Infeasible Mean</td>
<td>0.204</td>
<td>-1.011</td>
<td>0.007</td>
<td>0.504</td>
</tr>
<tr>
<td></td>
<td>Infeasible Median</td>
<td>0.190</td>
<td>-0.996</td>
<td>0.012</td>
<td>0.498</td>
</tr>
<tr>
<td></td>
<td>Standard error</td>
<td>0.129</td>
<td>0.098</td>
<td>0.086</td>
<td>0.085</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>0.185</td>
<td>-0.998</td>
<td>0.028</td>
<td>0.501</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.186</td>
<td>-0.985</td>
<td>0.033</td>
<td>0.503</td>
</tr>
<tr>
<td></td>
<td>Standard error</td>
<td>0.084</td>
<td>0.101</td>
<td>0.094</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td>Three-period Mean</td>
<td>0.210</td>
<td>-1.018</td>
<td>0.010</td>
<td>0.552</td>
</tr>
<tr>
<td></td>
<td>Three-period Median</td>
<td>0.210</td>
<td>-1.013</td>
<td>0.004</td>
<td>0.543</td>
</tr>
<tr>
<td></td>
<td>Standard error</td>
<td>0.096</td>
<td>0.094</td>
<td>0.095</td>
<td>0.092</td>
</tr>
<tr>
<td>DGP III:</td>
<td>True value</td>
<td>0.2</td>
<td>-1</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>Infeasible Mean</td>
<td>0.205</td>
<td>-1.012</td>
<td>1.008</td>
<td>0.501</td>
</tr>
<tr>
<td></td>
<td>Infeasible Median</td>
<td>0.202</td>
<td>-1.009</td>
<td>1.004</td>
<td>0.497</td>
</tr>
<tr>
<td></td>
<td>Standard error</td>
<td>0.113</td>
<td>0.087</td>
<td>0.080</td>
<td>0.057</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>0.201</td>
<td>-1.059</td>
<td>0.876</td>
<td>0.518</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.196</td>
<td>-1.052</td>
<td>0.878</td>
<td>0.526</td>
</tr>
<tr>
<td></td>
<td>Standard error</td>
<td>0.100</td>
<td>0.096</td>
<td>0.106</td>
<td>0.027</td>
</tr>
<tr>
<td></td>
<td>Three-period Mean</td>
<td>0.210</td>
<td>-1.035</td>
<td>0.977</td>
<td>0.548</td>
</tr>
<tr>
<td></td>
<td>Three-period Median</td>
<td>0.205</td>
<td>-1.036</td>
<td>0.982</td>
<td>0.544</td>
</tr>
<tr>
<td></td>
<td>Standard error</td>
<td>0.095</td>
<td>0.107</td>
<td>0.122</td>
<td>0.038</td>
</tr>
<tr>
<td>DGP IV:</td>
<td>True value</td>
<td>0.2</td>
<td>-1</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>Infeasible Mean</td>
<td>0.203</td>
<td>-1.011</td>
<td>1.008</td>
<td>0.505</td>
</tr>
<tr>
<td></td>
<td>Infeasible Median</td>
<td>0.201</td>
<td>-1.005</td>
<td>0.997</td>
<td>0.504</td>
</tr>
<tr>
<td></td>
<td>Standard error</td>
<td>0.111</td>
<td>0.090</td>
<td>0.079</td>
<td>0.058</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>0.197</td>
<td>-1.062</td>
<td>0.881</td>
<td>0.548</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.197</td>
<td>-1.057</td>
<td>0.889</td>
<td>0.551</td>
</tr>
<tr>
<td></td>
<td>Standard error</td>
<td>0.096</td>
<td>0.119</td>
<td>0.116</td>
<td>0.035</td>
</tr>
<tr>
<td></td>
<td>Three-period Mean</td>
<td>0.191</td>
<td>-1.057</td>
<td>0.964</td>
<td>0.565</td>
</tr>
<tr>
<td></td>
<td>Three-period Median</td>
<td>0.185</td>
<td>-1.071</td>
<td>0.964</td>
<td>0.557</td>
</tr>
<tr>
<td></td>
<td>Standard error</td>
<td>0.103</td>
<td>0.126</td>
<td>0.146</td>
<td>0.075</td>
</tr>
</tbody>
</table>

Note: Standard errors of the parameters are based on the standard error of the estimates across 100 simulations. The three-period results are estimated using the sieve MLE in Shiu and Hu (2010).
Table 2: Simulation of Dynamic tobit model (N=500)

<table>
<thead>
<tr>
<th>DGP</th>
<th>Parameters</th>
<th>( \beta_0 )</th>
<th>( \beta_1 )</th>
<th>( \gamma )</th>
<th>( \sigma^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DGP I:</strong></td>
<td>True value</td>
<td>0.2</td>
<td>-1</td>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>Infeasible Mean</td>
<td>0.220</td>
<td>-1.017</td>
<td>0.007</td>
<td>0.505</td>
</tr>
<tr>
<td></td>
<td>Infeasible Median</td>
<td>0.217</td>
<td>-1.018</td>
<td>0.018</td>
<td>0.504</td>
</tr>
<tr>
<td></td>
<td>Standard error</td>
<td>0.123</td>
<td>0.075</td>
<td>0.063</td>
<td>0.074</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>0.192</td>
<td>-1.015</td>
<td>-0.009</td>
<td>0.498</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.192</td>
<td>-1.018</td>
<td>-0.001</td>
<td>0.503</td>
</tr>
<tr>
<td></td>
<td>Standard error</td>
<td>0.098</td>
<td>0.110</td>
<td>0.109</td>
<td>0.036</td>
</tr>
<tr>
<td></td>
<td>Three-period Mean</td>
<td>0.207</td>
<td>-0.986</td>
<td>0.004</td>
<td>0.556</td>
</tr>
<tr>
<td></td>
<td>Three-period Median</td>
<td>0.202</td>
<td>-0.988</td>
<td>-0.003</td>
<td>0.551</td>
</tr>
<tr>
<td></td>
<td>Standard error</td>
<td>0.089</td>
<td>0.121</td>
<td>0.095</td>
<td>0.083</td>
</tr>
<tr>
<td><strong>DGP II:</strong></td>
<td>True value</td>
<td>0.2</td>
<td>-1</td>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>Infeasible Mean</td>
<td>0.196</td>
<td>-1.001</td>
<td>0.004</td>
<td>0.500</td>
</tr>
<tr>
<td></td>
<td>Infeasible Median</td>
<td>0.193</td>
<td>-1.005</td>
<td>0.004</td>
<td>0.497</td>
</tr>
<tr>
<td></td>
<td>Standard error</td>
<td>0.091</td>
<td>0.077</td>
<td>0.069</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>0.179</td>
<td>-0.991</td>
<td>-0.013</td>
<td>0.499</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.181</td>
<td>-0.995</td>
<td>-0.009</td>
<td>0.500</td>
</tr>
<tr>
<td></td>
<td>Standard error</td>
<td>0.099</td>
<td>0.102</td>
<td>0.099</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td>Three-period Mean</td>
<td>0.197</td>
<td>-1.017</td>
<td>0.004</td>
<td>0.553</td>
</tr>
<tr>
<td></td>
<td>Three-period Median</td>
<td>0.208</td>
<td>-1.026</td>
<td>0.014</td>
<td>0.554</td>
</tr>
<tr>
<td></td>
<td>Standard error</td>
<td>0.097</td>
<td>0.105</td>
<td>0.100</td>
<td>0.033</td>
</tr>
<tr>
<td><strong>DGP III:</strong></td>
<td>True value</td>
<td>0.2</td>
<td>-1</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>Infeasible Mean</td>
<td>0.193</td>
<td>-1.001</td>
<td>1.006</td>
<td>0.501</td>
</tr>
<tr>
<td></td>
<td>Infeasible Median</td>
<td>0.196</td>
<td>-0.994</td>
<td>1.007</td>
<td>0.495</td>
</tr>
<tr>
<td></td>
<td>Standard error</td>
<td>0.083</td>
<td>0.066</td>
<td>0.062</td>
<td>0.041</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>0.196</td>
<td>-1.058</td>
<td>0.856</td>
<td>0.519</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.192</td>
<td>-1.058</td>
<td>0.860</td>
<td>0.519</td>
</tr>
<tr>
<td></td>
<td>Standard error</td>
<td>0.105</td>
<td>0.090</td>
<td>0.093</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td>Three-period Mean</td>
<td>0.192</td>
<td>-1.047</td>
<td>0.976</td>
<td>0.564</td>
</tr>
<tr>
<td></td>
<td>Three-period Median</td>
<td>0.193</td>
<td>-1.043</td>
<td>0.975</td>
<td>0.565</td>
</tr>
<tr>
<td></td>
<td>Standard error</td>
<td>0.096</td>
<td>0.129</td>
<td>0.124</td>
<td>0.047</td>
</tr>
<tr>
<td><strong>DGP IV:</strong></td>
<td>True value</td>
<td>0.2</td>
<td>-1</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>Infeasible Mean</td>
<td>0.201</td>
<td>-1.001</td>
<td>1.000</td>
<td>0.498</td>
</tr>
<tr>
<td></td>
<td>Infeasible Median</td>
<td>0.210</td>
<td>-1.001</td>
<td>1.003</td>
<td>0.498</td>
</tr>
<tr>
<td></td>
<td>Standard error</td>
<td>0.090</td>
<td>0.061</td>
<td>0.054</td>
<td>0.039</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>0.186</td>
<td>-1.070</td>
<td>0.888</td>
<td>0.551</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.192</td>
<td>-1.072</td>
<td>0.899</td>
<td>0.554</td>
</tr>
<tr>
<td></td>
<td>Standard error</td>
<td>0.090</td>
<td>0.123</td>
<td>0.103</td>
<td>0.036</td>
</tr>
<tr>
<td></td>
<td>Three-period Mean</td>
<td>0.198</td>
<td>-1.068</td>
<td>0.937</td>
<td>0.573</td>
</tr>
<tr>
<td></td>
<td>Three-period Median</td>
<td>0.198</td>
<td>-1.054</td>
<td>0.947</td>
<td>0.567</td>
</tr>
<tr>
<td></td>
<td>Standard error</td>
<td>0.095</td>
<td>0.126</td>
<td>0.144</td>
<td>0.071</td>
</tr>
</tbody>
</table>

Note: Standard errors of the parameters are based on the standard error of the estimates across 100 simulations. The three-period results are estimated using the sieve MLE in Shiu and Hu (2010).
<table>
<thead>
<tr>
<th>DGP</th>
<th>Parameters</th>
<th>(\beta_0)</th>
<th>(\beta_1)</th>
<th>(\gamma)</th>
<th>(\sigma^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DGP I: True value</td>
<td>0.2</td>
<td>-1</td>
<td>0</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>Infeasible Mean</td>
<td>0.208</td>
<td>-1.006</td>
<td>0.005</td>
<td>0.490</td>
<td></td>
</tr>
<tr>
<td>Infeasible Median</td>
<td>0.206</td>
<td>-1.005</td>
<td>0.001</td>
<td>0.489</td>
<td></td>
</tr>
<tr>
<td>Standard error</td>
<td>0.063</td>
<td>0.049</td>
<td>0.040</td>
<td>0.043</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.189</td>
<td>-0.967</td>
<td>-0.003</td>
<td>0.502</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>0.174</td>
<td>-0.973</td>
<td>-0.008</td>
<td>0.512</td>
<td></td>
</tr>
<tr>
<td>Standard error</td>
<td>0.098</td>
<td>0.095</td>
<td>0.092</td>
<td>0.032</td>
<td></td>
</tr>
<tr>
<td>Three-period Mean</td>
<td>0.208</td>
<td>-0.993</td>
<td>-0.005</td>
<td>0.550</td>
<td></td>
</tr>
<tr>
<td>Three-period Median</td>
<td>0.216</td>
<td>-0.992</td>
<td>-0.004</td>
<td>0.554</td>
<td></td>
</tr>
<tr>
<td>Standard error</td>
<td>0.093</td>
<td>0.112</td>
<td>0.107</td>
<td>0.044</td>
<td></td>
</tr>
</tbody>
</table>

| DGP II: True value | 0.2 | -1 | 0 | 0.5 |
| Infeasible Mean | 0.210 | -1.010 | 0.007 | 0.505 |
| Infeasible Median | 0.202 | -1.002 | 0.005 | 0.496 |
| Standard error | 0.067 | 0.061 | 0.056 | 0.041 |
| Mean | 0.199 | -0.995 | -0.006 | 0.503 |
| Median | 0.200 | -1.002 | -0.008 | 0.509 |
| Standard error | 0.107 | 0.102 | 0.105 | 0.509 |
| Three-period Mean | 0.208 | -0.996 | -0.005 | 0.556 |
| Three-period Median | 0.216 | -0.999 | -0.004 | 0.557 |
| Standard error | 0.094 | 0.108 | 0.107 | 0.031 |

| DGP III: True value | 0.2 | -1 | 1 | 0.5 |
| Infeasible Mean | 0.211 | -1.006 | 1.005 | 0.499 |
| Infeasible Median | 0.205 | -1.004 | 1.005 | 0.498 |
| Standard error | 0.058 | 0.052 | 0.050 | 0.027 |
| Mean | 0.217 | -1.051 | 0.839 | 0.526 |
| Median | 0.230 | -1.053 | 0.836 | 0.528 |
| Standard error | 0.087 | 0.101 | 0.096 | 0.026 |
| Three-period Mean | 0.205 | -1.014 | 0.952 | 0.568 |
| Three-period Median | 0.206 | -1.014 | 0.975 | 0.566 |
| Standard error | 0.092 | 0.114 | 0.136 | 0.047 |

| DGP IV: True value | 0.2 | -1 | 1 | 0.5 |
| Infeasible Mean | 0.208 | -1.004 | 1.005 | 0.499 |
| Infeasible Median | 0.208 | -1.008 | 1.005 | 0.497 |
| Standard error | 0.058 | 0.049 | 0.046 | 0.026 |
| Mean | 0.204 | -1.058 | 0.879 | 0.558 |
| Median | 0.213 | -1.063 | 0.880 | 0.559 |
| Standard error | 0.090 | 0.116 | 0.101 | 0.032 |
| Three-period Mean | 0.210 | -1.067 | 0.917 | 0.590 |
| Three-period Median | 0.213 | -1.052 | 0.930 | 0.577 |
| Standard error | 0.099 | 0.133 | 0.144 | 0.068 |

Note: Standard errors of the parameters are based on the standard error of the estimates across 100 simulations. The three-period results are estimated using the sieve MLE in Shiu and Hu (2010).
Table 4: Simulation of Average Partial Effects

<table>
<thead>
<tr>
<th>DGP</th>
<th>State Dependence $SD(\bar{X}<em>{it}, \bar{Y}</em>{it-1})$</th>
<th>N=250</th>
<th>N=500</th>
<th>N=1000</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DGP I:</strong></td>
<td>True value</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>0.002</td>
<td>-0.002</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.002</td>
<td>-0.003</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>Standard error</td>
<td>0.008</td>
<td>0.008</td>
<td>0.009</td>
</tr>
<tr>
<td><strong>DGP II:</strong></td>
<td>True value</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>-0.001</td>
<td>-0.001</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.002</td>
<td>0.001</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>Standard error</td>
<td>0.008</td>
<td>0.010</td>
<td>0.010</td>
</tr>
<tr>
<td><strong>DGP III:</strong></td>
<td>True value</td>
<td>0.579</td>
<td>0.579</td>
<td>0.579</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>0.660</td>
<td>0.652</td>
<td>0.644</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.665</td>
<td>0.655</td>
<td>0.652</td>
</tr>
<tr>
<td></td>
<td>Standard error</td>
<td>0.115</td>
<td>0.093</td>
<td>0.082</td>
</tr>
<tr>
<td><strong>DGP IV:</strong></td>
<td>True value</td>
<td>0.479</td>
<td>0.479</td>
<td>0.479</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>0.618</td>
<td>0.620</td>
<td>0.646</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.632</td>
<td>0.635</td>
<td>0.645</td>
</tr>
<tr>
<td></td>
<td>Standard error</td>
<td>0.177</td>
<td>0.139</td>
<td>0.145</td>
</tr>
</tbody>
</table>

Note: Standard errors of the parameters are based on the standard error of the estimates across 100 simulations. $SD(\bar{X}_{it}, \bar{Y}_{it-1}) \equiv \mu_1(\bar{X}_{it}, \bar{Y}_{it-1}) - \mu_1(\bar{X}_{it}, 0)$, where $(\bar{X}_{it}, \bar{Y}_{it-1})$ is the mean of $(X_{it}, Y_{it})$ and it is the difference of the average structural functions of two different outcomes of $Y_{it-1}$, 0 and $\bar{Y}_{it-1}$. This represents the magnitude of the state dependence.
Table 5: Sample Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
<th>Period_{it}</th>
<th>Period_{t+1}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lnexp</td>
<td>log(medical expenditures+1)</td>
<td>5.292</td>
<td>5.307</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.903)</td>
<td>(3.038)</td>
</tr>
<tr>
<td>Lninc</td>
<td>ln(family income+1)</td>
<td>9.056</td>
<td>9.217</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.821)</td>
<td>(2.695)</td>
</tr>
<tr>
<td>Lnfm</td>
<td>ln(family size)</td>
<td>1.036</td>
<td>1.034</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.538)</td>
<td>(0.542)</td>
</tr>
<tr>
<td>Age</td>
<td>Age</td>
<td>39.427</td>
<td>40.429</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(12.498)</td>
<td>(12.500)</td>
</tr>
<tr>
<td>Male</td>
<td>=1 if person is male; 0 otherwise</td>
<td>0.469</td>
<td>0.469</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.499)</td>
<td>(0.499)</td>
</tr>
<tr>
<td>Black</td>
<td>=1 if race of household head is black; 0 otherwi</td>
<td>0.148</td>
<td>0.148</td>
</tr>
<tr>
<td></td>
<td>s; 0 otherwise</td>
<td>(0.355)</td>
<td>(0.355)</td>
</tr>
<tr>
<td>Education</td>
<td>Education of the household head</td>
<td>12.599</td>
<td>12.599</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.087)</td>
<td>(3.087)</td>
</tr>
<tr>
<td>Pysical</td>
<td>=1 if the person has a physical limitation; 0 o</td>
<td>0.057</td>
<td>0.059</td>
</tr>
<tr>
<td></td>
<td>therwise; 0 otherwise</td>
<td>(0.231)</td>
<td>(0.235)</td>
</tr>
<tr>
<td>Ndental</td>
<td>Number of dental care visits</td>
<td>0.938</td>
<td>0.857</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.746)</td>
<td>(1.617)</td>
</tr>
<tr>
<td>Good</td>
<td>=1 if self-rated health is good; 0 otherwise</td>
<td>0.266</td>
<td>0.276</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.442)</td>
<td>(0.447)</td>
</tr>
<tr>
<td>Fair</td>
<td>=1 if self-rated health is fair; 0 otherwise</td>
<td>0.086</td>
<td>0.081</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.280)</td>
<td>(0.274)</td>
</tr>
<tr>
<td>Poor</td>
<td>=1 if self-rated health is poor; 0 otherwise</td>
<td>0.026</td>
<td>0.027</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.158)</td>
<td>(0.162)</td>
</tr>
<tr>
<td>Deduction</td>
<td>=1 if the person has nonzero itemized deductions; 0 otherwi</td>
<td>0.057</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>s; 0 otherwise</td>
<td>(0.232)</td>
<td>(0.227)</td>
</tr>
<tr>
<td>Medicare</td>
<td>=1 if the person is covered by Medicare; 0 othe</td>
<td>0.025</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td>rwise; 0 otherwise</td>
<td>(0.156)</td>
<td>(0.182)</td>
</tr>
<tr>
<td>Medicaid</td>
<td>=1 if the person is covered by Medicaid; 0 othe</td>
<td>0.070</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td>rwise; 0 otherwise</td>
<td>(0.255)</td>
<td>(0.253)</td>
</tr>
<tr>
<td>Sample size</td>
<td></td>
<td>7,669</td>
<td>7,669</td>
</tr>
</tbody>
</table>

Note: The variables in Period_{it} and Period_{t+1} refer to the first-year and the second-year values of each participation respectively. There are 1,430 and 1,578 individuals with zero medical expenditures in Period_{it} and Period_{t+1}, respectively. Standard deviations are in parentheses.
## Table 6: Panel Censored Estimates for Health Expenditure

<table>
<thead>
<tr>
<th></th>
<th>Linear RE</th>
<th>Semi-parametric Tobit</th>
<th>Fixed Effects Tobit</th>
<th>Dynamic Tobit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>Coefficient</td>
<td>APE</td>
<td>Coefficient</td>
</tr>
<tr>
<td>Lnexp&lt;sub&gt;it&lt;/sub&gt; - 1</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>1.052***</td>
</tr>
<tr>
<td>Lninc</td>
<td>0.031***</td>
<td>0.042***</td>
<td>0.039***</td>
<td>0.041***</td>
</tr>
<tr>
<td>Lnfnam</td>
<td>-0.252***</td>
<td>-0.299***</td>
<td>-0.276***</td>
<td>-0.301***</td>
</tr>
<tr>
<td>Age</td>
<td>0.046***</td>
<td>0.048***</td>
<td>0.044***</td>
<td>0.050***</td>
</tr>
<tr>
<td>Male</td>
<td>-1.130***</td>
<td>-1.399***</td>
<td>-1.294***</td>
<td>-1.399***</td>
</tr>
<tr>
<td>Black</td>
<td>-0.581***</td>
<td>-0.717***</td>
<td>-0.653***</td>
<td>-0.717***</td>
</tr>
<tr>
<td>Education</td>
<td>0.145***</td>
<td>0.184***</td>
<td>0.170***</td>
<td>0.181***</td>
</tr>
<tr>
<td>Pysical</td>
<td>0.806***</td>
<td>0.854***</td>
<td>0.788***</td>
<td>0.855***</td>
</tr>
<tr>
<td>Ndental</td>
<td>0.442***</td>
<td>0.496***</td>
<td>0.458***</td>
<td>0.502***</td>
</tr>
<tr>
<td>Good</td>
<td>0.342***</td>
<td>0.391***</td>
<td>0.362***</td>
<td>0.392***</td>
</tr>
<tr>
<td>Fair</td>
<td>1.037***</td>
<td>1.180***</td>
<td>1.111***</td>
<td>1.178***</td>
</tr>
<tr>
<td>Poor</td>
<td>1.777***</td>
<td>1.956***</td>
<td>1.865***</td>
<td>1.957***</td>
</tr>
<tr>
<td>Deduction</td>
<td>0.384***</td>
<td>0.432***</td>
<td>0.402***</td>
<td>0.429***</td>
</tr>
<tr>
<td>Medicare</td>
<td>0.906***</td>
<td>0.995***</td>
<td>0.936***</td>
<td>0.995***</td>
</tr>
<tr>
<td>Medicaid</td>
<td>1.138***</td>
<td>1.346***</td>
<td>1.270***</td>
<td>1.343***</td>
</tr>
</tbody>
</table>

Note: Bootstrap (simulation) standard errors are reported in parentheses, using 100 bootstrap replications. APEs are reported by taking derivatives or differences of ASF at the sample mean of \((x_{it}, y_{it-1})\).