Identification and Estimation of Nonlinear Dynamic Panel Data Models with Unobserved Covariates*

Ji-Liang Shiu† and Yingyao Hu‡

April 3, 2010

Abstract

This paper considers nonparametric identification of nonlinear dynamic models for panel data with unobserved voariates. Including such unobserved covariates may control for both the individual-specific unobserved heterogeneity and the endogeneity of the explanatory variables. Without specifying the distribution of the initial condition with the unobserved variables, we show that the models are nonparametrically identified from three periods of data. The main identifying assumption requires the evolution of the observed covariates depends on the unoberved covariates but not on the lagged dependent variable. We also propose a sieve maximum likelihood estimator (MLE) and focus on two classes of nonlinear dynamic panel data models, i.e., dynamic discrete choice models and dynamic censored models. We present the asymptotic property of the sieve MLE and investigate the finite sample properties of these sieve-based estimator through a Monte Carlo study. An intertemporal female labor force participation model is estimated as an empirical illustration using a sample from the Panel Study of Income Dynamics (PSID).

Keywords: dynamic nonlinear panel data model, dynamic discrete choice model, dynamic censored model, nonparametric identification, initial condition, random effects, unobserved heterogeneity, unobserved covariate, endogeneity, intertemporal labor force participation,

---

*The authors thank Chin-Wei Yang for proofreading the draft. All errors remain our own.

†Department of Economics, National Chung-Cheng University, 168 University Rd. Min-Hsiung Chia-Yi, Taiwan. Email: jishiu@ccu.edu.tw.

‡Department of Economics, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218. Email: yhu@jhu.edu.

1. Introduction

This paper considers nonlinear dynamic models for panel data with unobserved covariates. These models take into account the dynamic processes by allowing the lagged value of the dependent variable as one of the explanatory variables as well as containing observed and unobserved permanent (heterogeneous) or transitory (serially-correlated) individual differences. Let Y_{it} be the dependent variable at period t and X_{it} be a vector of observed covariates for individual i. We consider nonlinear dynamic panel data models of the form:

$$Y_{it} = g\left(Y_{it-1}, X_{it}, U_{it}, \xi_{it}\right), \tag{1}$$

where g is an unknown nonstochastic function, U_{it} is an unobserved covariate correlated with other observed explanatory variables $\left(Y_{it-1}, X_{it}\right)$, and ξ_{it} stands for a random shock independent of all other explanatory variables $(Y_{it-1}, X_{it}, U_{it})$. The unobserved covariate U_{it} may contain two components as follows:

$$U_{it} = V_i + \eta_{it},$$

where V_i is the unobserved heterogeneity or the random effects correlated with the observed covariates X_{it} and η_{it} is an unobserved serially-correlated component.[1] The transitory component η_{it} may be a function of all the time-varying RHS variables in the history, i.e., $\eta_{it} = \varphi\left(\{Y_{i\tau-1}, X_{i\tau}, \xi_{i\tau}\}_{\tau=0,1,\ldots,t-1}\right)$ for some function φ. Both observed explanatory variables Y_{it-1} and X_{it} become endogeneous if the unobserved covariate U_{it} is ignored. In this paper, we provide reasonable assumptions under which the distribution of Y_{it} conditional on $(Y_{it-1}, X_{it}, U_{it})$, i.e., $f_{Y_{it}|Y_{it-1},X_{it},U_{it}}$, is nonparametrically identified. The nonparametric identification of $f_{Y_{it}|Y_{it-1},X_{it},U_{it}}$ may lead to that of the general form of our model in equation (1) under certain specifications of the distribution of the random shock ξ_{it}.

In the econometric literature, there are two approaches to tackling the unobserved heterogeneity V_i: random effects and fixed effects. In the fixed effect approach, much attention has been devoted to linear models with an additive unobserved effect. The problem can be solved

[1] The random effect V_i is usually assumed to be independent of other covariates for convenience so that it mainly causes an efficiency problem instead of an endogeneity problem especially in a linear panel data model. In this paper, we consider a more realistic setting where the individual-specific heterogeneity may be correlated with the covariates from the same individual.

by first applying an appropriate transformation to eliminate the unobserved effect and then implementing instrument variables (IV) in a generalized method of moments (GMM) framework. Anderson and Hsiao (1982), Arellano and Bond (1991), Arellano and Bover (1995) and Ahn and Schmidt (1995) employ an IV estimator on a transformation equation through first-differencing. Eliminating the unobserved effects is notably more difficult in nonlinear models and some progress has been made in this area. Chamberlain (1984) considers a conditional likelihood approach for logit models with strictly exogenous assumption. Honoré and Kyriazidou (2000) generalize the conditional probability approach to estimate the unknown parameters without formulating the distribution of the unobserved individual effects or the probability distribution of the initial observations for certain types of discrete choice logit models. However, their results have to rely on a very strong assumption to match the explanatory variables in different time-periods. Their estimator is consistent and asymptotically normal but the rate of convergence is not the inverse of the square root of the sample size. Honoré (1993), Hu (2002) and Honoré and Hu (2004) obtain moment conditions for estimating dynamic censored regression panel data models.

On the other hand, it is often appealing to take a random effect specification by making assumptions on the distribution of the individual effects. The main difficulty of this approach is the so-called initial conditions problem.[2] With a relatively short panel, the initial conditions have a very strong impact on the entire path of the observations but they may not be observed in the sample. One remedy to this problem is to specify the distribution of the initial conditions given the unobserved heterogeneity. The drawbacks of this approach are that the corresponding likelihood functions typically involve high order integration and misspecification of the distributions generally results in inconsistent parameter estimates. The

[2] The random effect approach for dynamic models requires the specification on the initial conditions of the process. Specifically, consider a special case of our model (1), dynamic discrete choice models without observed covariates X_{it}, in the following form:

$$Y_{it} = 1\left(\gamma Y_{it-1} + V_i + \xi_{it} \geq 0\right).$$

Then the conditional distribution $f_{Y_{it}|Y_{it-1},V_i}$ can be specified and the corresponding likelihood function has the structure

$$\mathcal{L} = \int f_{Y_{i0}|V_i} \prod_{t=1}^{T-1} f_{Y_{it}|Y_{it-1},V_i} f_{V_i} dv_i,$$

where $f_{Y_{i0}|V_i}$ denotes the marginal probability of Y_{i0} given V_i. If the process is not observed from the start then the initial state for individual i, y_{i0} cannot be assumed fixed. However, it is not clear that how to derive the initial condition $f_{Y_{i0}|V_i}$ from $f_{Y_{it}|Y_{it-1},V_i}$ so it could be internally inconsistent across different time periods if the evolution of these two process can not be connected. Heckman (1981b) suggested that using a flexible functional form to approximate the initial conditions.

associated computational burden of high order integration has been reduced significantly by recent advances in simulation techniques.[3] Hyslop (1999) analyzes the intertemporal labor force participation behavior of married women using maximum simulated likelihood (MSL) estimator to simulate the likelihood function of dynamic probit models with a nontrivial error structure. Wooldridge (2005) suggests a general method for handling the initial conditions problem by using a joint density conditional on the strictly exogenous variables and the initial condition. Honoré and Tamer (2006) relax the distributional assumption of the initial condition and calculate bounds on parameters of the interest in panel dynamic discrete choice models.

In this paper we adopt the random effect approach for nonlinear dynamic panel data models without specifying the distribution of the initial condition. We treat the unobserved covariate in nonlinear dynamic panel data models as the latent true values in nonlinear measurement error models and the observed covariates as the measurement of the latent true values.[4] We then utilize the identification results in Hu and Schennach (2008a), where the measurement error is not assumed to be independent of the latent true values. Their results rely on a unique eigenvalue-eigenfunction decomposition of an integral operator associated with joint densities of observable variables and unobservable variables.

The strength of our approach is that we provide nonparametric identification of nonlinear dynamic panel data model using three periods of data without specifying initial conditions. The model may be described in,$f_{Y_{it}|Y_{it-1},X_{it},U_{it}}$, the conditional distribution of the dependent variable of interest for an individual i, Y_{it}, conditional on a lagged value of that variable Y_{it-1}, explanatory variables X_{it}, and an unobserved covariate U_{it}. We show that $f_{Y_{it}|Y_{it-1},X_{it},U_{it}}$ can be nonparametrically identified from a sample of $\left\{X_{it+1}, Y_{it}, X_{it}, Y_{it-1}, X_{it-1}\right\}$ without parametric assumptions on the distribution of the individuals' dependent variable conditional on the unobserved covariate in the initial period. The main identifying assumption requires that the dynamic process of the covariates X_{it+1} depends on the unoberved covariate U_{it} but is independent of the lagged dependent variables Y_{it}, Y_{it-1}, and X_{it-1} conditional on X_{it} and U_{it}.

[3]See Gourieroux and Monfort (1993), Hajivassiliou (1993), Hajivassiliou and Ruud (1994) and Keane (1993) for the reviews of the literature.

[4]An ideal candidate for the "measurement" of the latent covariate would be the dependent variable because it is inherently correlated with the latent covariate. However, such a measurement is not informative enough when the dependent variable is discrete and the latent covariate is continuous.

The identification of $f_{Y_{it}|Y_{it-1},X_{it},U_{it}}$ leads to the identification of the general form of our model in equation (1).[5] We present below two motivating examples in the existing literature. The specifications in these two types of models can be used to distinguish between dynamic responses to lagged dependent variables, observed covariates, and unobserved covariates. While the state dependence Y_{it-1} reflects that experiencing the event in one period should affect the probability of the event in the next period, the unobserved heterogeneity V_i represents individual's inherent ability to resist the transitory shocks η_{it}.

Example 1 (Dynamic Discrete-choice Model with an Unobserved Covariate): A binary case of dynamic discrete choice models is as follows:

$$Y_{it} = 1\left(X_{it}'\beta + \gamma Y_{it-1} + V_i + \varepsilon_{it} \geq 0\right) \quad \text{with} \quad \forall i = 1, ..., n; t = 1, ..., T-1,$$

where $1(\cdot)$ is the 0-1 indicator function and the error ε_{it} follows an AR(1) process $\varepsilon_{it} = \rho\varepsilon_{it-1} + \xi_{it}$ for some constant ρ. The conditional distribution of the interest is then:

$$f_{Y_{it}|X_{it},Y_{it-1},U_{it}} = \left(1 - F_{\xi_{it}}\left[-\left(X_{it}'\beta + \gamma Y_{it-1} + U_{it}\right)\right]\right)^{Y_{it}} F_{\xi_{it}}\left[-\left(X_{it}'\beta + \gamma Y_{it-1} + U_{it}\right)\right]^{1-Y_{it}},$$

where $F_{\xi_{it}}$ is the CDF of the random shock ξ_{it}, $U_{it} = V_i + \eta_{it}$, and $\eta_{it} = \rho\varepsilon_{it-1}$. Empirical applications of the dynamic discrete-choice model above have been studied in a variety of contexts, such as health status (Contoyannis, Jones, and Rice (2004), Halliday (2002)), brand loyalty (Chintagunta, Kyriazidou, and Perktold (2001)) , welfare participation (Chay, Hoynes, and Hyslop (2001)), and labor force participation (Heckman and Willis (1977), Hyslop (1999)). Among these studies, the intertemporal labor participation behavior of married women is a natural illustration of the dynamic discrete choice model. In such a model, the dependent variable Y_{it} denotes the t-th period participation decision and the covariate X_{it} is the wage or other observable characteristics in that period. The heterogeneity V_i is the unobserved individual skill level or motivation, while the idiosyncratic disturbance ε_{it} denotes the luck and the measurement error. Heckman (1978, 1981a,b) has termed the presence of Y_{it-1} "true" state dependence and V_i "spurious" state dependence.

[5] Evdokimov (2009) considers a nonparametric panel data model with nonadditive unobserved heterogeneity: $Y_{it} = m(X_{it}, V_i) + \varepsilon_{it}$. The model has a different focus since our model explicitly includes lags of the endogenous dependent variable Y_{it-1} and a nonadditive ε_{it}.

Example 2 (Dynamic Censored Model with an Unobserved Covariate): In many applications, we may have

$$Y_{it} = \max\left\{X_{it}'\beta + \gamma Y_{it-1} + V_i + \varepsilon_{it}, 0\right\} \quad \text{with} \quad \forall i = 1, ..., n; t = 1, ..., T-1,$$

with $\varepsilon_{it} = \rho\varepsilon_{it-1} + \xi_{it}$. It follows that

(2)

$$f_{Y_{it}|X_{it},Y_{it-1},U_{it}} = F_{\xi_{it}}\left[-\left(X_{it}'\beta + \gamma Y_{it-1} + U_{it}\right)\right]^{\mathbf{1}(Y_{it}=0)} f_{\xi_{it}}\left[Y_{it} - X_{it}'\beta - \gamma Y_{it-1} - U_{it}\right]^{\mathbf{1}(Y_{it}>0)}.$$

where $F_{\xi_{it}}$ and $f_{\xi_{it}}$ are the CDF and the PDF of the random shock ξ_{it} respectively. The dependent variable Y_{it} may stand for the amount of insurance coverage chosen by an individual or a firm's expenditures on R&D. In each case, an economic agent solves an optimization problem and $Y_{it} = 0$ may be an optimal corner solution. For this reason, this type of censored regression models is also called a corner solution model or a censored model with lagged censored dependent variables.[6] Honoré (1993) and Honoré and Hu (2004) use a method of moments framework to estimate the model without making distributional assumptions about V_i.

Based on our nonparametric identification results, we propose a semiparametric sieve maximum likelihood estimator (MLE) for the model. We show the consistency of our estimator and the asymptotic normality of its parametric components. The finite sample properties of the proposed sieve MLE are investigated through Monte Carlo simulations of dynamic discrete choice models and dynamic censored models. Our empirical application focuses on how the labor participation decisions of married women respond to their previous participation states, fertility decisions, and nonlabor incomes. We develop and test a variety of dynamic econometric models using a seven year longitudinal sample from the Panel Study of Income Dynamics (PSID) in order to compare the results with those in Hyslop (1999). In the empirical application, we examine three different model specifications, i.e., a static probit model, a maximum simulation likelihood (MSL) model, and a semi-parametric dynamic probit model. Our results find a smaller significant negative effects on nonlabor income variables and a

[6] This setting rules out certain types of data censoring. For example, if the censoring is due to top-coding, then it makes sense to consider a lagged value of the latent variable, i.e., $Y_{it}^* = X_{it}'\beta + \gamma Y_{it-1}^* + v_i + \varepsilon_{it}$ and $Y_{it} = \max[Y_{it}^*, c_t]$. This top-coded dynamic censored model has been considered in Hu (2000, 2002).

estimated negative effect of children age 0-2 in the the current period and previous period increases by 30% and decline by 50% respectively.

The paper is organized as follows. We present the nonparametric identification of nonlinear dynamic panel data models in Section 2. Section 3 discusses our proposed sieve MLE. Section 4 provides the Monte Carlo study. Section 5 presents an empirical application describing the intertemporal labor participation of married women. Section 6 concludes. Appendices include proofs of consistency and asymptotic normality of the proposed sieve MLE and discussions on how to impose restrictions on sieve coefficients in the sieve MLE.

2. Nonparametric Identification

In this section, we present the assumptions under which the distribution of the dependent variable Y_{it} conditional on Y_{it-1}, covariates X_{it}, and the unobserved covariate U_{it}, i.e., $f_{Y_{it}|Y_{it-1},X_{it},U_{it}}$, is nonparametrically identified. We start with a panel data containing three periods, $\{X_{it+1}, Y_{it}, X_{it}, Y_{it-1}, X_{it-1}\}$ for $i = 1, 2, ..., n$. The law of total probability leads to

$$f_{X_{it+1},Y_{it},X_{it},Y_{it-1},X_{it-1}} = \int f_{X_{it+1}|Y_{it},X_{it},Y_{it-1},X_{it-1},U_{it}} f_{Y_{it}|X_{it},Y_{it-1},X_{it-1},U_{it}} f_{X_{it},Y_{it-1},X_{it-1},U_{it}} dU_{it},$$

where we omit the arguments in the density function to make the expressions concise.

We assume

Assumption 2.1. *(Exogeneous shocks) the random shock ξ_{it} is independent of $\xi_{i\tau}$ for any $\tau \neq t$ and $\{Y_{i\tau-1}, X_{i\tau}, U_{i\tau}\}$ for any $\tau \leq t$.*

As shown in the two examples above, this assumption has been used in many existing studies in the literature. However, it is still stronger than necessary. For the nonparametric identification of $f_{Y_{it}|Y_{it-1},X_{it},U_{it}}$, we only need $f_{Y_{it}|X_{it},Y_{it-1},X_{it-1},U_{it}} = f_{Y_{it}|X_{it},Y_{it-1},U_{it}}$, which is implied by Assumption 2.1. Given equation 1, the condition $f_{Y_{it}|X_{it},Y_{it-1},X_{it-1},U_{it}} = f_{Y_{it}|X_{it},Y_{it-1},U_{it}}$ holds if the random shock ξ_{it} is independent of the covariate X_{it-1}. Assumption 2.1 then implies

$$f_{X_{it+1},Y_{it},X_{it},Y_{it-1},X_{it-1}} = \int f_{X_{it+1}|Y_{it},X_{it},Y_{it-1},X_{it-1},U_{it}} f_{Y_{it}|X_{it},Y_{it-1},U_{it}} f_{X_{it},Y_{it-1},X_{it-1},U_{it}} dU_{it}.$$

Furthermore, we simplify the evolution of the observed covariates X_{it} as follows:

Assumption 2.2. *(Covariate evolution)* $f_{X_{it+1}|Y_{it},X_{it},Y_{it-1},X_{it-1},U_{it}} = f_{X_{it+1}|X_{it},U_{it}}$.

Assumption 2.2 may be decomposed into two steps. The first step is a Markov-type assumption $f_{X_{it+1}|Y_{it},X_{it},Y_{it-1},X_{it-1},U_{it}} = f_{X_{it+1}|Y_{it},X_{it},U_{it}}$, which implies that the evolution of the observed covariate X_{it+1} only depends on all the explanatory variables in the previous period (Y_{it}, X_{it}, U_{it}). Such Markov-type assumptions make the model tractable without losing economic intuitions, and therefore, are widely used in dynamic models. The second step is a simplification, i.e., $f_{X_{it+1}|Y_{it},X_{it},U_{it}} = f_{X_{it+1}|X_{it},U_{it}}$. In nonlinear models, the dependent variable Y_{it} may either be discrete or truncated, while at least part of covariates X_{it} is continuous. In those cases, such a simplification may not lose too much generality. Without this simplication, the nonparametric identification of the model is still feasible using another identification strategy (see Hu and Shum (2009)), which we do not pursue here.

Assumption 2.2 then implies that

$$f_{X_{it+1},Y_{it},X_{it},Y_{it-1},X_{it-1}} = \int f_{X_{it+1}|X_{it},U_{it}} f_{Y_{it}|X_{it},Y_{it-1},U_{it}} f_{X_{it},Y_{it-1},X_{it-1},U_{it}} dU_{it}. \tag{3}$$

Based on this equation, we may apply the identification results in Hu and Schennach (2008) to show the all the unknown densities on the RHS are identified from the observed density on the LHS. Let $\mathcal{L}^p(\mathcal{X}), 1 \leq p < \infty$ stand for the space of function $h(\cdot)$ with $\int_{\mathcal{X}} |h(x)|^p dx < \infty$. For any $1 \leq p \leq \infty$ and any given $(y_{it}, x_{it}, y_{it-1})$, we define operators as follows:

$$\begin{aligned} L_{X_{it+1},y_{it},x_{it},y_{it-1},X_{it-1}} &: \mathcal{L}^p(\mathcal{X}_{t-1}) \to \mathcal{L}^p(\mathcal{X}_{t+1}) \\ (L_{X_{it+1},y_{it},x_{it},y_{it-1},X_{it-1}} h)(u) &= \int f_{X_{it+1},Y_{it},X_{it},Y_{it-1},X_{it-1}}(u, y_{it}, x_{it}, y_{it-1}, x) h(x) dx, \end{aligned}$$

and

$$\begin{aligned} D_{y_{it}|x_{it},y_{it-1},U_{it}} &: \mathcal{L}^p(\mathcal{U}) \to \mathcal{L}^p(\mathcal{U}) \\ (D_{y_{it}|x_{it},y_{it-1},U_{it}} h)(u) &= f_{Y_{it}|X_{it},Y_{it-1},U_{it}}(y_{it}|x_{it}, y_{it-1}, u) h(u). \end{aligned}$$

Similarly, we define

$$(L_{X_{it+1},x_{it},y_{it-1},X_{it-1}}h)(u) = \int f_{X_{it+1},X_{it},Y_{it-1},X_{it-1}}(u,x_{it},y_{it-1},x)h(x)dx,$$
$$(L_{X_{it+1}|x_{it},U_{it}}h)(x) = \int f_{X_{it+1}|X_{it},U_{it}}(x|x_{it},u)h(u)du,$$
$$(L_{x_{it},y_{it-1},X_{it-1},U_{it}}h)(u) = \int f_{X_{it},Y_{it-1},X_{it-1},U_{it}}(x_{it},y_{it-1},x,u)h(x)dx.$$

Eq. (3) is equivalent to the following operator relationship:

$$L_{X_{it+1},y_{it},x_{it},y_{it-1},X_{it-1}} = L_{X_{it+1}|x_{it},U_{it}} D_{y_{it}|x_{it},y_{it-1},U_{it}} L_{x_{it},y_{it-1},X_{it-1},U_{it}}.$$

Integrating out Y_{it} in Eq. (3) leads to $f_{X_{it+1},X_{it},Y_{it-1},X_{it-1}} = \int f_{X_{it+1}|X_{it},U_{it}} f_{X_{it},Y_{it-1},X_{it-1},U_{it}} dU_{it}$, which is equivalent to

$$L_{X_{it+1},x_{it},y_{it-1},X_{it-1}} = L_{X_{it+1}|x_{it},U_{it}} L_{x_{it},y_{it-1},X_{it-1},U_{it}}.$$

with $(L_{X_{it-1},x_{it},y_{it-1},X_{it+1}}h)(u) = \int f_{X_{it-1},X_{it},Y_{it-1},X_{it+1}}(u,x_{it},y_{it-1},x)h(x)dx$. We may then apply the results in Hu and Schennach (2008) to identify $f_{X_{it+1}|X_{it},U_{it}}$, $f_{Y_{it}|X_{it},Y_{it-1},U_{it}}$, and $f_{X_{it},Y_{it-1},X_{it-1},U_{it}}$ from $f_{X_{it+1},Y_{it},X_{it},Y_{it-1},X_{it-1}}$.

We assume

Assumption 2.3. *(Invertibility) For any $(x_{it}, y_{it-1}) \in \mathcal{X}_{it} \times \mathcal{Y}_{it-1}$, $L_{X_{it-1},x_{it},y_{it-1},X_{it+1}}$ and $L_{X_{it+1}|x_{it},U_{it}}$ are invertible.*

Intuitively, this assumption guarantees that the observables contains enough information on the unobserved covariate U_{it}. The invertibility of $L_{X_{it-1},x_{it},y_{it-1},X_{it+1}}$ is imposed on observables so that it is testable in principle. The invertibility of $L_{X_{it+1}|x_{it},U_{it}}$ requires the covariates in period $t+1$, i.e., X_{it+1} contain enough information on the unobserved covariate U_{it} conditional on X_{it}. For example, we may have $X_{it+1} = X_{it} + U_{it} + h(X_{it})\epsilon_{it}$, where ϵ_{it} is independent of X_{it} and U_{it} and has a nonvanishing characteristic function on the real line. We use X_{it+1} instead of Y_{it+1} for the information on U_{it} because the dependent variable Y_{it+1} is discrete and U_{it} is continuous in many interesting applications. In that case, the operator mapping from functions of U_{it} to those of Y_{it+1} can't be invertible. On the other hand, when Y_{it+1} is continuous, it would be more reasonable to impose invertibility on the operator mapping from functions of U_{it} to those of Y_{it+1}, while U_{it} or V_i is allowed to be independent of the observed covariates X_{it}.

Note that when the unobserved component U_{it} is continuous-valued, the invertibility of $L_{X_{it+1}|x_{it},U_{it}}$ implies that the set of the explanatory variables X_{it} contains a continuous element Z_{it}. The existence of the continuous component, Z_{it} is essential. It is impossible to nonparametrically identify a distribution of a continuous unobservable variable only by observed discrete variables. The restriction imposed on the continuous Z_{it+1} guarantees that the explanatory variables X_{it+1} contains enough information to identify unobserved component U_{it}.

This assumption enables us to have

$$L_{X_{it+1},y_{it},x_{it},y_{it-1},X_{it-1}} L^{-1}_{X_{it+1},x_{it},y_{it-1},X_{it-1}} = L_{X_{it+1}|x_{it},U_{it}} D_{y_{it}|x_{it},y_{it-1},U_{it}} L^{-1}_{X_{it+1}|x_{it},U_{it}},$$

which implies a spectral decomposition of the observed operators on the LHS. The eigenvalues are the kernel function of the diagonal operator $D_{y_{it}|x_{it},y_{it-1},U_{it}}$ and the eigenfunctions are the kernel function $f_{X_{it+1}|X_{it},U_{it}}$ of the operator $L_{X_{it+1}|x_{it},U_{it}}$. In order to make the eigenvalues distinctive, we assume

Assumption 2.4. *(Distinctive eigenvalues) there exist a known function $\omega(\cdot)$ such that $E[\omega(Y_{it})\,|x_{it}, y_{it-1}, u_{it}]$ is monotonic in u_{it} for any given (x_{it}, y_{it-1}).*

The function $\omega(\cdot)$ may be specified by users, such as $\omega(y) = y$, $\omega(y) = I(y > 0)$, or $\omega(y) = y^2$. For example, we may have $\omega(y) = I(y = 0)$ in the two examples above. In both cases, $E[I(Y_{it} = 0)\,|x_{it}, y_{it-1}, u_{it}] = F_{\xi_{it}}[-(x'_{it}\beta + \gamma y_{it-1} + u_{it})]$, which is monotonic in u_{it}. Since the identification from the spectral decomposition is only identified up to u_{it} and its monotone transformation, we make a normalization assumption to pins down the unobserved covariate u_{it}.

Assumption 2.5. *(Normalization) For any given $x_{it} \in \mathcal{X}_{it}$, there exists a known functional G such that $G\left[f_{X_{it+1}|X_{it},U_{it}}(\cdot|x_{it}, u_{it})\right] = u_{it}$.*

The functional G may be the mean, the mode, median, or a quantile. For example, we may have $X_{it+1} = X_{it} + U_{it} + h(X_{it})\epsilon_{it}$ with an unknown function $h(\cdot)$ and a zero median independent error ϵ_{it}. Then U_{it} is the median of the density function $f_{(X_{it+1}-X_{it})|X_{it},U_{it}}(\cdot|x_{it}, u_{it})$.

Notice that Theorem 1 in Hu and Schennach (2008) implies that all three densities $f_{X_{it+1}|X_{it},U_{it}}$, $f_{Y_{it}|X_{it},Y_{it-1},U_{it}}$, and $f_{X_{it},Y_{it-1},X_{it-1},U_{it}}$ are identified under the assumptions introduced above. The model of interest is described in the density $f_{Y_{it}|X_{it},Y_{it-1},U_{it}}$. The initial condition at period $t-1$ is contained in the joint distribution $f_{X_{it},Y_{it-1},X_{it-1},U_{it}}$. The evolution of the covariates X_{it} is described in $f_{X_{it+1}|X_{it},U_{it}}$. We summarize our identification results as follows:

Theorem 2.1. *Under Assumptions 2.1, 2.2, 2.3, 2.4, 2.5, the joint distribution* $f_{X_{it+1},Y_{it},X_{it},Y_{it-1},X_{it-1}}$ *uniquely determines the model of interest* $f_{Y_{it}|X_{it},Y_{it-1},U_{it}}$*, together with the evolution density of observed covariates* $f_{X_{it+1}|X_{it},U_{it}}$ *and the initial joint distribution* $f_{X_{it},Y_{it-1},X_{it-1},U_{it}}$.

3. Estimation

The dynamic panel data model (1) specifies the relationship between a dependent variable of interest for an individual i, Y_{it}, and the explanatory variables including a lagged dependent variable, a set of possibly time-varying explanatory variables X_{it}, an unobserved covariate U_{it}. If we are willing to make a normality assumption on ξ_{it}, then the model in example 1 becomes a probit model and the model in example 2 becomes a tobit model. The general specification here covers a number of other dynamic nonlinear panel data model in one framework.

Given that the random shocks $\{\xi_{it}\}_{t=0}^{T}$ is exogenous, the conditional distribution $f_{Y_{it}|X_{it},Y_{it-1},U_{it}}$ is a combination of the function g and the distribution of ξ_{it}. In most applications, the function g and the distribution of ξ_{it} have a parametric form. That means the model may be parameterized in the following form,

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

where θ includes the unknown parameters in both the function g and the distribution of ξ_{it}. Under the rank condition in the regular identification of parametric models, the nonparametric identification of $f_{Y_{it}|X_{it},Y_{it-1},U_{it}}$ implies that of the parameter θ, and therefore, the identification of the function g and the distribution of ξ_{it}. In general, we may allow $\theta = (b, \lambda)^T$, where b is a finite-dimensional parameter vector of interest and λ is a potentially infinite-dimensional nuisance parameter or nonparametric component.[7] What is not specified in the

[7] A participation of α_0 into finite-dimensional parameters and infinite-dimensional parameters does not affect

model is the evolution of the covariate X_{it}, together with the unobserved component U_{it}, i.e., $f_{X_{it+1}|X_{it},U_{it}}$, and the initial joint distribution of all the variables $f_{X_{it},Y_{it-1},X_{it-1},U_{it}}$. We consider the nonparametric elements $(f_{X_{it+1}|X_{it},U_{it}},\ \lambda,\ f_{X_{it},Y_{it-1},X_{it-1},U_{it}})^T$ as infinite-dimensional nuisance parameters in our semiparametric estimator.

Our semiparametric sieve maximum likelihood estimator (sieve MLE) does not require the initial condition assumption for the widely used panel data models, such as dynamic discrete-response models and dynamic censored models. In section 2, we have shown equation (3) uniquely determines $(f_{X_{it+1}|X_{it},U_{it}},\ f_{Y_{it}|X_{it},Y_{it-1},U_{it}},\ f_{X_{it},Y_{it-1},X_{it-1},U_{it}})^T$. While the dynamic panel data model component $f_{Y_{it}|X_{it},Y_{it-1},U_{it}}$ will be parameterized, the other components are treated as nonparametric nuisance functions. Eq. (3) implies

$$\begin{aligned}\alpha_0 &\equiv (f_{X_{it+1}|X_{it},U_{it}}, \theta, f_{X_{it},Y_{it-1},X_{it-1},U_{it}})^T \\ &= \arg \max_{(f_1,\theta,f_2)^T\in\mathcal{A}} E \ln \int f_1(x_{it+1}|x_{it},u_{it}) f_{Y_{it}|X_{it},Y_{it-1},U_{it}}(y_{it}|x_{it},y_{it-1},u_{it};\theta) \\ &\qquad\qquad \times f_2(x_{it},y_{it-1},x_{it-1},u_{it})du_{it},\end{aligned}$$

which suggests a corresponding semiparametric sieve MLE using an i.i.d. sample $\{x_{it+1}, y_{it}, x_{it}, y_{it-1}, x_{it-1}\}_{i=1}^n$

$$\begin{aligned}\widehat{\alpha}_n &\equiv \left(\hat{f}_1, \hat{\theta}, \hat{f}_2\right)^T \\ &= \arg \max_{(f_1,\theta,f_2)^T\in\mathcal{A}^n} \frac{1}{n}\sum_{i=1}^{n} \ln \int f_1(x_{it+1}|x_{it},u_{it}) f_{Y_{it}|X_{it},Y_{it-1},U_{it}}(y_{it}|x_{it},y_{it-1},u_{it};\theta) \\ &\qquad\qquad \times f_2(x_{it},y_{it-1},x_{it-1},u_{it})du_{it}.\end{aligned}$$

The function space $\mathcal{A}$ contains the corresponding true densities and $\mathcal{A}^n$ is a sequence of approximating sieve spaces.

Our estimator is a direct application of the general semi-parametric sieve MLE in Shen (1997), Chen and Shen (1998), and Ai and Chen (2003). In the appendix, we provide sufficient conditoins for the consistency of our semiparametric estimator $\widehat{\alpha}_n$ and those for the $\sqrt{n}$ asymptotic normality of the parametric component $\widehat{b}$. The asymptotic theory of the proposed sieve MLE and the detailed development of sieve approximations of the nonparametric components are also provided in Appendix A.

---

our sieve MLE. More examples of a partition can be found in Shen (1997).

4. Monte Carlo Evidence

In this section we present a Monte Carlo study that investigates the finite sample properties of the proposed sieve MLE estimators in the two different settings, dynamic discrete-choice models and dynamic censored models. We start with the specification of the models as follows.

Semi-parametric Dynamic Probit Models

First, we adopt a parametric assumption for ε_{it}. Suppose that ε_{it} has a stationary AR(1) with an independent Gaussian white noise process, $\varepsilon_{it} = \rho\varepsilon_{it-1} + \xi_{it}$, $\xi_{it} \sim N(0, 1/2)$. We have

$$f_{Y_{it}|X_{it},Y_{it-1},U_{it}} = \Phi_{\xi_{it}}\left(X_{it}'\beta + \gamma Y_{it-1} + U_{it}\right)^{Y_{it}}\left[1 - \Phi_{\xi_{it}}\left(X_{it}'\beta + \gamma Y_{it-1} + U_{it}\right)\right]^{1-Y_{it}},$$

with $U_{it} = V_i + \rho\varepsilon_{it-1}$.

The density $f_{Y_{it}|X_{it},Y_{it-1},U_{it}}$ is fully parameterized and θ only contain the parametric component $b = (\gamma, \beta)$. We approximate $f_{X_{it+1}|X_{it},U_{it}}$, and $f_{X_{it},Y_{it-1},X_{it-1},U_{it}}$ by truncated series in the estimation.

Semi-parametric Dynamic Tobit Models:

We also assume that ε_{it} has a stationary AR(1) with an independent Gaussian white noise process, $\varepsilon_{it} = \rho\varepsilon_{it-1} + \xi_{it}$. This gives

(4)

$$\begin{aligned}
f_{Y_{it}|X_{it},Y_{it-1},U_{it}} &= \left[1 - \Phi_{\varepsilon_{it}}\left(X_{it}'\beta + \gamma Y_{it-1} + U_{it}\right)\right]^{\mathbf{1}(Y_{it}=0)}\phi_{\varepsilon_{it}}(y_{it} - X_{it}'\beta - \gamma Y_{it-1} - U_{it})^{\mathbf{1}(Y_{it}>0)} \\
&= \left[1 - \Phi\left(\frac{X_{it}'\beta + \gamma Y_{it-1} + U_{it}}{\sigma_\xi}\right)\right]^{\mathbf{1}(Y_{it}=0)} \times \\
&\qquad \left[\frac{1}{\sigma_\xi}\phi\left(\frac{y_{it} - X_{it}'\beta - \gamma Y_{it-1} - U_{it}}{\sigma_\xi}\right)\right]^{\mathbf{1}(Y_{it}>0)},
\end{aligned}$$

and the parameter is $\theta = b = (\gamma, \beta, \sigma_\xi^2)$.

The data generating process (DGP) for dynamic discrete choice models and dynamic censored models in the Monte Carlo experiments are generated according to the following processes respectively:

$$
\begin{aligned}
(5) \qquad & Y_{it} = 1\left(\beta_0 + \beta_1 X_{it} + \gamma Y_{it-1} + U_{it} + \xi_{it} \geq 0\right) \quad \text{with} \\
& U_{it} = V_i + \rho\varepsilon_{it-1} \quad \forall \quad i = 1, ..., N; t = 1, ..., T-1.
\end{aligned}
$$

and

$$
\begin{aligned}
(6) \qquad & Y_{it} = \max\left\{\beta_0 + \beta_1 X_{it} + \gamma Y_{it-1} + U_{it} + \xi_{it}, 0\right\} \quad \text{with} \\
& U_{it} = V_i + \rho\varepsilon_{it-1} \quad \forall \quad i = 1, ..., N; t = 1, ..., T-1.
\end{aligned}
$$

where $V_i \sim N(1, 1/2)$. For simplicity in the implementation, the distribution of U_{it} is truncated on $[0, 2]$, and our generating processes of covariate evolution have the following form $X_{it+1} = X_{it} + h(X_{it})\epsilon_{it} + U_{it}$ or

$$
f_{X_{it+1}|X_{it},U_{it}}(x_{t+1}|x_t, u) = \frac{1}{h(x_t)} f_\epsilon\left(\frac{x_{t+1} - x_t - u}{h(x_t)}\right),
$$

where f_ϵ is a density function that can be specified under different identification conditions of Assumption 2.5.[8] We consider the mode condition in this paper, and $f_\epsilon(x) = exp(x - e^x)$ in all simulated data. In addition, we set $h(x) = 0.3\exp(-x)$ to allow heterogeneity and assume the initial observation (y_0, x_0) and the initial component ξ_0 $(=\epsilon_{i0})$ equal to zero.

We consider five different values of $(\gamma, \sigma_\xi^2, \rho)$ in the experiments: $(\gamma, \sigma_\xi^2, \rho)$ = (0,0.5,0), (0,0.5,0.5), (1,0.5,0), (1,0.5,0.5), (1,0.5,-0.5) and the parameters in the intercept and the exogenous variable are held fixed: $\beta_0 = 0$ and $\beta_1 = -1$. In summary, the data generating processes are as follows:

$$
\begin{aligned}
\text{DGP I:} \quad & (\beta_0, \beta_1, \gamma, \sigma_\xi^2, \rho) = \quad (0, -1, 0, 0.5, 0) \\
\text{DGP II:} \quad & (\beta_0, \beta_1, \gamma, \sigma_\xi^2, \rho) = \quad (0, -1, 0, 0.5, 0.5) \\
\text{DGP III:} \quad & (\beta_0, \beta_1, \gamma, \sigma_\xi^2, \rho) = \quad (0, -1, 1, 0.5, 0) \\
\text{DGP IV:} \quad & (\beta_0, \beta_1, \gamma, \sigma_\xi^2, \rho) = \quad (0, -1, 1, 0.5, 0.5) \\
\text{DGP V:} \quad & (\beta_0, \beta_1, \gamma, \sigma_\xi^2, \rho) = \quad (0, -1, 1, 0.5, -0.5).
\end{aligned}
$$

[8]This generating process is also adopted in Hu and Schennach (2008a) and it can be adjusted to a variety of identification conditions, the mean, the mode, median, or a quantile.

The first two DGPs are not state dependence ($\gamma = 0$) while the rest are state dependence with $\gamma = 1$. Three different sample sizes N are considered: 250, 500, 1000. To secure a more stationary sample, the sampling data are drawn over $T = 7$ periods but only last three periods are utilized. A detailed derivation of sieve MLE method is shown in Appendix B.

Tables 1, 2, and 3 present simulation results under the semi-parametric probit model. The simulation results of DGP I (only allows for unobserved heterogeneity) show that generally downward bias in the structural model coefficients (β_1, γ) for sample sizes N=250 and 500 but upward bias for N=1000. For DGP II, the results have downward bias in the structural model coefficients in $(\beta_0, \beta_1, \gamma)$. In addition, with nontrivial transitory component ($\rho \neq 0$) in DGP II, the standard errors of β_1 are larger except for N=1000. As for DGPs with nontrivial state dependence, there is less bias for $(\beta_0, \beta_1, \gamma)$ for DGP III for sample size N=1000 but for DGP IV & V our results show less bias in sample size N=500. The coefficient estimators of γ in DGP IV & V have very small bias for sample sizes N=500 and 1000, which means that our estimation for state dependence is very precise among processes with serial correlation ($\rho \neq 0$). In general, there are smaller standard errors in sample size N=1000. For DGPs in this sample size, the parameters are much less precisely estimated, and the means and medians of (β_1, γ) are quite different, reflecting some skewness in their respective distributions. Our estimator provides fairly consistent results in this case.

Tables 4, 5, and 6 report the results of estimates for the semi-parametric tobit model. In the tobit model, there is positive bias in β_1 and γ for all DGPs with trivial state dependence except for γ in N=1000. In tobit case, we have additional parameters to estimate, σ_ξ^2. There is downward bias of the parameter in DGP I & II and their standard errors become smaller as numbers of simulation increase. As for other DGPs in positive state dependence, estimation results of γ show that there are small bias and precision increase with numbers of simulation. Also, for those DGPs, the means and medians of β_1 are quite different, reflecting some skewness in distributions. Our sieve MLE estimators tend to perform well. In general, bias and standard errors in DGPs with state dependence are smaller than those without state dependence.

There are two nuisance parameters, $f_{X_{t+1}|X_t,U_t}$ and $f_{X_t,Y_{t-1},X_{t-1},U_t}$ in our Monte Carlo simulation and we use Fourier series to approximate the evolution density and the square root of the initial joint distribution. Since a higher dimensional sieve space is constructed by tensor product of univariate sieve series, approximation series can be formed from several univariate Fourier series. In the semi-parametric probit model, while in the approximation of the evolution densities we use three univariate Fourier series with the number of term, $i_n = 5$, $j_n = 2$, and $k_n = 2$, in the approximation of the initial joint distribution we have $i_n = 5$, $j_n = 2$, $k_n = 2$, and $l_n = 2$. As for the semi-parametric tobit model, we have similar choices of approximation series. The detailed sieve expression of those nuisance parameters can be found in Appendix B.

In summary, the Monte Carlo study shows that our semiparametric sieve MLE performs well with a finite sample.

5. Empirical Example

In this section, we apply our estimator to a dynamic discrete choice model, which discribes the labor-force participation decisions of married women given their past participation state and other covariates. The advantage of our estimator is that our model may include (i) arbitrary and unspecified correlated random effects between unobserved time invariant factors such as skill level or motivation and time-varying X_{it}'s, and (ii) no initial conditions assumption.[9] We will compare our estimates with those in Hyslop (1999), which studied a similar empirical model with less general assumptions.

5.1. Data Descriptive

In order to provide comparison of the models developed in this paper and by Hyslop (1999), we also used the data related to waves 12-19 of the the Michigan Panel Survey of Income Dynamics from the calendar years 1979-85 to study married women's employment decisions. The seven-year sample consists of women aged 18-60 in 1980, continuously married, and the

[9] In Hyslop (1999), a correlated random-effects (CRE) specification for v_i is:

$$v_i = \sum_{s=0}^{T} (\delta_{1s} \cdot (\#\text{Kids0-2})_{is} + \delta_{2s} \cdot (\#\text{Kids3-5})_{is} + \delta_{3s} \cdot (\#\text{Kids6-17})_{is}) + \sum_{s=0}^{T-1} \delta_{4s} \cdot y_{mis} + \eta_i.$$

An alternative CRE specification can be:

$$v_i = \delta_1 \cdot (\overline{\#\text{Kids0-2}})_i + \delta_2 \cdot (\overline{\#\text{Kids3-5}})_i + \delta_3 \cdot (\overline{\#\text{Kids6-17}})_i + \delta_4 \cdot \bar{y}_{mi} + \eta_i,$$

where $\bar{x}_i = \sum_{t=0}^{T} x_{it}$.

husband is a labor force participant in each of the sample years from. A woman is defined to be a labor market participant if she works for money any time in the sample year.[10] The sample contains 1752 married women and also includes both the random Census subsample of families and the nonrandom Survey of Economic Opportunities (SEO) subsample of families.[11]

The number of possible binary participation sequences over a 7-period panel is $2^7 = 128$ and the sequences can expressed as sequences of zeros and ones of the length 7.[12] If we partition the full sample based on all the observed annual participation outcomes of women during the seven-year period, the number of subsamples is up to 128. To provide a useful analysis of the differences of women's work propensity due to the number of years worked and the associated participation sequences, we choose a small group of dividing criteria. The mutually exclusive sub-sample partition is as follows: we have in column (2) women who work in each year corresponding to a sequence '1111111'; in column (3), women who never work during the sample period corresponding to a sequence '0000000'; in column (4), women who experience a single transition from employment to nonemployment-that is, six participation sequences '1000000',...,'1111110'; in column (5), women who experience a single transition from nonemployment to employment corresponding to another six participation sequences '1000000',...,'1111110'; and in column (6), women who experience more than a single transition in their participation status corresponding to the rest of participation sequences.

Table 7 reports the descriptive statistics from the resulting subsamples. The selection of variables of interest in the table is close to the sample characteristics in Hyslop (1999) and the variables show similar trends and features. Column (1) presents the characteristics of the variables for the whole sample. Comparison of the observed annual participation outcomes with individual's independent participation decision form a binomial distribution with fixed probability of 0.7 (the average participation rate) indicates there is strong persistence in the married women's annual participation decisions. If there does not exist any persistence, then about 8 percent of the sample would be expected to work each year, and only 0.02 percent would not work at all, which are quite different from the sample relative frequencies, 47 percent

[10] A standard definition of a participant ia that an individual reports both positive annual hours worked and annual earnings. Hyslop (1995) provided a description of the extent of aggregation bias which results from ignoring intra-year labor force transition.

[11] Hyslop (1995) obtains a sample consisted of 1812 observations. The PSID contains an over-sample of low-income families called the Survey of Economic Opportunity (SEO).

[12] An '1' in the t-th position of the sequence denotes participation in year t, while a '0' denotes nonparticipation.

and 9 percent respectively. In addition, the rest of the columns demonstrate the difference in the observable variables across the subsamples. For 825 women in the sample whom we observe employment in each of the seven years (column (2)), they are more likely to be better educated than average, or be Black, have fewer dependent children (especially children's age under 6 years), and their husbands' labor incomes are lower than average. Women who are never employed (column (3)) are older, less educated, and their husbands' labor income are higher than average. The women in this group have slightly fewer young children, reflecting the older age of the group. In column (4), women who make a single transition from employment to nonemployment have fewer dependent children but are more likely to have infant children (aged 0-2 years), and their husbands have above average earnings. Women who experience a single transition from nonemployment to employment (column (5)) are less likely to be black and have significantly more children (aged 0-17 years). The last column (6) indicates women who experience multiple employment transitions are younger, have more dependent children of all ages, and their husbands have below-average labor income.

The description of the sample characteristics according to various subsamples suggests that there are several patterns between observable characteristics of individuals and their participation behavior. First, there is a negative income effect from husband's labor income on women's willingness for labor market participation (column (2) vs column (3)). Secondly, In general, the presence of children, especially young children, tends to reduce the participation of women, except for women who never work in the sample (column (3)). The numbers of very young and older children between the single-transition subsamples in columns (4) and (5) are 0.34 and 0.24 (aged 0-2 years) and 0.67 and 1.21 (aged 6-17 years) respectively. The differences suggest that women leave employment to have children and re-enter employment as their children reach school age. The life-cycle interpretation is plausible by slight age difference between women in these groups (35.66 and 35.81). However, the age differences between these groups in Hyslop (1999) suggesting that the composition of these samples is determined by more than simply fertility considerations. Finally, the column (6) indicates that the presence of children in all age group (aged 0-17 years), together with low husband's labor income, increases the number of employment transitions of women.

5.2. Specifications and Estimation Results

According to a theoretic model in Hyslop (1999), the labor-force participation decisions of married women depend on whether or not their market wage offer exceeds their reservation wage, which in turn may depend on their past participation state, namely, suppose Y_t is the t-th period participation decision, W_t is the wage, and W_{0t}^* is a reservation wage then period t participation decision can be formulated by

$$Y_t = 1(W_t > W_{0t}^* - \gamma Y_{t-1}) \tag{7}$$

where $1(\cdot)$ denotes an indicator function that is equal to 1 if the expression is true and 0 otherwise. An empirical reduced form specification for Eq. (7) is the following

$$Y_{it} = 1(\gamma Y_{t-1} + X_{it}'\beta + U_{it} + \xi_{it} > 0) \qquad \forall i = 1, ..., N; t = 1, ..., T-1$$

where X_{it} is a vector of observed demographic and family structure variables U_{it} captures the effects of unobserved factors, and β and γ are parameters. There are two latent sources for the unobserved term U_{it}:

$$U_{it} = V_i + \rho\varepsilon_{it-1}$$

where V_i is an individual-specific component, which captures unobserved time invariant factors possibly correlated with the time-varying X_{it}'s such as skill level or motivation; and ε_{it} is a serially correlated error term, which captures factors such as transitory wage movements.

The estimation results for the various models of labor force participation are presented in Table 8 which includes estimates from static probit models with random effect (column 1), maximum simulation likelihood (MSL) models with random effect[13] (column 2), semi-parametric dynamic probit models (column 3). While the first two models in columns are estimated using full seven years of data, the last one is estimated over three-period data. In addition, the last model is the dynamic models without an initial conditions specification. The static probit model is estimated by MSL with 200 replications. It allows for individual-specific random effects but ignores possible dynamic effects of the past employment and potential

[13] A detailed discussion of MSL models can be found in Hyslop (1999). There are more different specifications in the paper. Here we only compare the models allowing the three sources of persistence.

correlation between the unobserved heterogeneity and the regressors. The results are that permanent nonlabor income has a significantly negative effect, transitory income reduces the contemporaneous participation, and preschool children have substantially negative effect. In addition, the variance of unobserved heterogeneity is 0.786. We now turn to dynamic specifications. The specifications in MSL model contain random effects, a stationary AR(1) error component, and first-order state dependence (SD(1)). The results show a large and significant first-order state dependence effect (1.117). The addition of SD(1) and AR(1) error component reduced the effects of nonlabor income variables largely (-0.007 & -0.004) and the contemporaneous fertility variables like #Kid3-5_t and #Kid6-17_t by approximately 50 percent. But the estimated effects of younger kids in the past and current periods #Kid0-2_{t-1} and #Kid0-2_t have stronger negative effects on women's participation decisions (-0.117 & -0.380). Including state dependence and serial correlation error component reduce the error variance (0.313) due to unobserved heterogeneity. The estimated AR(1) coefficient ρ is -0.146.[14]

Next, consider the specifications without an initial conditions assumption. The results also show that large first-order state dependence effects on the semi-parametric model (1.089). There exists a strong dependence between married women's current labor force participation and past labor force participation and relaxing the initial conditions assumption increase the negative effects of nonlabor income variables and their significance in the dynamic models. Permanent income and transitory income both reduce the probability of participation but the effect of permanent nonlabor income has substantially greater magnitude.

The fertility variables in the model are generally similar to those in column (1) and (2) but with less magnitude. That is: each of them has a significantly negative effect on married women's current labor force participation status, and younger children have stronger effect than older. In our semi-parametric Probit model, the unobserved heterogeneity and the AR(1) component have been mixed into the unobserved covariate U_{it}. They are not identified so there are no any estimation results.

Although the model allows more flexible approach, its ability to predict the observed participation outcomes does not increase much. We compare frequencies of the participation outcomes predicted by the models in Table 9 to assess their fitting ability. Table 9 presents

[14] A correlated random-effects (CRE) is adopted in Hyslop (1999) to test the exogeneity of fertility with respect to participation decisions. His results show that there is no evidence against the exogeneity of fertility decision in dynamic model specifications.

the frequencies of sample distribution and these predicted outcomes by the various estimated models over seven years period. Column (2) presents the predicted frequency from static probit model with random effect. The fraction of the predicted outcomes greatly over-predicts the frequencies of zero, one and six years worked and greatly under-predicts the frequencies of seven years worked. The results from the model MSL greatly under-predicts the frequencies of zero and seven year worked. As a result, the model substantially under-predicts the frequencies of the outcomes with no change in participation status over periods.

The final column in Table 9 contains the predicted frequencies from the semi-parametric probit model. The model predicts the frequencies in each participation outcome adequately for never work, and always work. It over-predicts the frequency of one and six years worked and under-predict the frequency of two, three, four, and five years worked. This is expected if there are larger lagged effect of participation decisions. Without initial conditions assumption, the model predicts the distribution of the number of years worked reasonably well. However, the predictive power from the model (column 3) without initial conditions assumption relative to the dynamic model with initial conditions assumption (column 2) is relatively small. One possible explanation of this is that the source of the serial persistence in participation outcomes over time is not well identified by those regressors. We might need other important regressors like child-care cost or welfare benefit from working.

In comparison to the results in the dynamic probit models allowing for CRE, AR(1), and SD(1) in Hyslop (1999), adding unspecified CRE and avoiding initial conditions have significant effect on the model. Our results find a smaller significant negative effects on nonlabor income variables (-0.221 and -0.106 v.s. -0.285 and -0.140, respectively) and a estimated negative effect of children age 0-2 in the the current period and previous period increases by 30 % (from -0.252 to -0.316) and decline by 50% (from -0.115 to -0.055) respectively. The effects of relaxing assumptions in Hyslop (1999) are similar to the comparison here.

6. Conclusion

This paper presents the nonparametric identification of nonlinear dynamic panel data models with unobserved covariates. We show the models are identified using only three periods of data without initial conditions assumption, and we propose a sieve MLE estimator, which is applied to two examples, a dynamic discrete-choice model and a dynamic censored model. Both of them allow for three sources of persistence, "true" state dependence, unobserved individual heterogeneity ("spurious" state dependence), and possible serially correlated transitory error. Monte Carlo experiments have shown that how to deal with specific implementation issues and the sieve MLE estimators perform well for these models. Our sieve MLE is shown to be root n consistent and asymptotically normal. Finally, we apply our estimator to an intertemporal female labor force participation model using a sample from the Panel Study of Income Dynamics (PSID).

Appendix

A. Asymptotic Properties of the Sieve Maximum Likelihood Estimator

This appendix presents the consistency of our estimator and the asymptotic normality of the parametric component of our estimator. Furthermore, we provide further details on the implementation of the semiparametric sieve estimator, i.e., how to impose restrictions on the sieve coefficients.

Our asymptotic analysis relies on regularity restrictions on function containing the parameters of interest α. Frist, we introduce a typical space of smooth functions, Hölder space. Given a $d \times 1$ vector of nonnegative integers, $a = (a_1, ..., a_d)'$ and denote $[a] = a_1 + ... + a_d$ and let D^a denote the differential operator defined by $D^a = \frac{\partial^{[a]}}{\partial \xi_1^{a_1} ... \partial \xi_d^{a_d}}$. Let m denote the largest integer satisfying $\gamma > \underline{\gamma}$ and set $\gamma = \underline{\gamma} + p$. The Hölder space $\Lambda^\gamma(\nu)$ of order $\gamma > 0$ is a collection of functions which are m times continuously differentiable on ν and the $\underline{\gamma}-$th derivative are Hölder continuous with the exponent p. The Hölder space becomes a Banach space with the Hölder norm, i.e., $\forall g \in \Lambda^\gamma(\nu)$

$$\|g\|_{\Lambda^\gamma} = \sup_{\xi \in \nu} |g(\xi)| + \max_{a_1+...+a_d=\underline{\gamma}} \sup_{\xi \neq \xi' \in \nu} \frac{|D^a g(\xi) - D^a g(\xi')|}{\|\xi - \xi'\|_E^p}. \tag{8}$$

The weighted Hölder norm is defined as $\|g\|_{\Lambda^{\gamma,\omega}} \equiv \|\widetilde{g}\|_{\Lambda^\gamma}$ for $\widetilde{g}(\xi) \equiv g(\xi)\omega(\xi)$ and the corresponding weighted Hölder space is $\Lambda^{\gamma,\omega}(\nu)$. Define a weighted Hölder ball as $\Lambda_c^{\gamma,\omega}(\nu) \equiv \{g \in$

$\Lambda^{\gamma,\omega}(\nu) : \|g\|_{\Lambda^{\gamma,\omega}} \leq c < \infty\}$. Let $\varepsilon \in \mathbb{R}$, and $W \in \mathcal{W}$ with $\mathcal{W}$ a compact convex subset in $\mathbb{R}^{d_w}$. Define the following spaces:

$$\mathcal{F}_1 = \{f_1(\cdot|\cdot,\cdot) \in \Lambda_c^{\gamma_1,\omega}(\mathcal{X} \times \mathcal{X} \times \mathcal{U}) : f_1(\cdot) \geq 0 \text{ and } \int_{\mathcal{X}} f_1(\cdot|x,u)dx = 1, \forall (x,u) \in \mathcal{X} \times \mathcal{U}\},$$
$$\mathcal{M} = \{\lambda(\cdot) \in \Lambda_c^{\gamma_m,\omega}(\mathbb{R}) : 0 \leq \lambda(\cdot) \leq 1, \lambda(-\infty) = 0 \text{ and } \lambda(\infty) = 1\},$$
$$\mathcal{F}_2 = \{f(\cdot)_2^{1/2} \in \Lambda_c^{\gamma_3,\omega}(\mathcal{X} \times \mathcal{Y} \times \mathcal{X} \times \mathcal{U}) : f(\cdot)_2 \geq 0 \text{ and } \int_{\mathcal{X}\times\mathcal{Y}\times\mathcal{X}\times\mathcal{U}} f_2(\cdot,...,\cdot)dxdydxdu = 1\},$$

where $\gamma_i > 1$ $\forall i = 1,2$, and $\gamma_m > 1$. Recall that the parameter of the dynamic panel data model is $\theta = (b, \lambda)$. Suppose that $\mathcal{B}$ is a compact set such that its interior containing the true parametric component of the dynamic panel data model b_0. If the dynamic panel data model component $f_{Y_t|X_t,Y_{t-1},U_t}$ is fully parameterized, then we do not need the infinite-dimensional function space $\mathcal{M}$. In the case, the parameter θ of the dynamic panel data model only contain a finite-dimensional parameter vector b. Without loss of generality, we assume that θ contains an unknown function λ. We assume that the parameters of interest $f_{X_{t+1}|X_t,U_t}$, θ, and $f_{X_t,Y_{t-1},X_{t-1},U_t}$ belong to the spaces, $\mathcal{F}_1$, $\Theta \equiv \mathcal{B} \times \mathcal{M}$, and $\mathcal{F}_2$ respectively. The following smoothness and boundedness restrictions to limit the size of the parameter spaces.

Assumption A.1. *With $\gamma_i > 1$ $\forall i = 1,2$, and $\gamma_m > 1$, we have (i) $f_1(\cdot|\cdot,\cdot) \in \mathcal{F}_1$, (ii) $\lambda(\cdot) \in \mathcal{M}$, (iii) $f(\cdot)_2^{1/2} \in \mathcal{F}_2$.*

Set $\mathcal{A} = \mathcal{F}_1 \times \Theta \times \mathcal{F}_2$ and $\alpha \equiv (f_1, \theta, f_2)'$. Then the true parameter α_0 maximizes:

$$\text{(9)} \qquad \sup_{\alpha \in \mathcal{A}} E\left[\ln \int f_1(x_{t+1}|x_t,u_t) f_{Y_t|X_t,Y_{t-1},U_t}(y_t|x_t,y_{t-1},u_t;\theta) f_2(x_t,y_{t-1},x_{t-1},u_t)du_t\right].$$

An estimator could then be obtained by maximizing the sample analog of Eq. (9). Define

$$\text{(10)} \qquad \widehat{Q}_n(z_t;\alpha) = \frac{1}{n}\sum_{i=1}^{n} \ln f_{Z_t}(z_{it};\alpha) \text{ with}$$
$$\ln f_{Z_t}(z_{it};\alpha) \equiv \ln \int f_1(x_{it+1}|x_{it},u_t) f_{Y_{it}|X_{it},Y_{it-1},U_t}(y_{it}|x_{it},y_{it-1},u_t;\theta) \times f_2(x_{it},y_{it-1},x_{it-1},u_t)du_t,$$

where z_t is a realization of a random variable $Z_t \equiv (X_{t+1}, Y_t, X_t, Y_{t-1}, X_{t-1})$. However, when the function spaces $\mathcal{A}$ is large, the estimation method could yield an inconsistent estimator or a consistent estimator which converges very slowly. Denote $\Theta^n \equiv \mathcal{B} \times \mathcal{M}^n$. The sieve spaces $\mathcal{A}^n \equiv \mathcal{F}_1^n \times \Theta^n \times \mathcal{F}_2^n$ will be introduced to replace the function spaces $\mathcal{A}$ to overcome the problem, namely, maximizing $\widehat{Q}_n(z_t;\alpha)$ over $\mathcal{A}^n$, a sequence of approximation spaces to $\mathcal{A}$. In the sieve approximation, we consider a finite-dimensional sieve $\mathcal{A}^n$ as follows. Let $p^k(\cdot) = (p_1(\cdot), ..., p_k(\cdot))'$ be a vector of some known univariate basis function and $p^k(\cdot, ..., \cdot) = (p_1(\cdot, ..., \cdot), ..., p_k(\cdot, ..., \cdot))'$ be multivariate basis function generated by tensor product construction. The sieve spaces are

$$\begin{aligned}
\mathcal{F}_1^n &= \{f_1(x_{t+1}|x_t, v) = p^{k_{n1}}(x_{t+1}, x_t, u_t)'\beta_1 \in \mathcal{F}_1\}, \\
\mathcal{M}^n &= \{\lambda(\varepsilon) = p^{k_{n\lambda}}(\varepsilon)'\beta_\lambda \in \mathcal{M}\}, \\
\mathcal{F}_2^n &= \{f_2(x_t, y_{t-1}, x_{t-1}, u_t)^{1/2} = p^{k_{n2}}(x_t, y_{t-1}, x_{t-1}, u_t)'\beta_2 \in \mathcal{F}_2\}.
\end{aligned}$$

A consistent sieve MLE $\widehat{\alpha}_n$ is given by

$$\widehat{\alpha}_n = \arg\max_{\alpha \in \mathcal{A}^n} \widehat{Q}_n(z_t;\alpha). \tag{11}$$

The rest of this appendix show the consistency of $\widehat{\alpha}_n$ and its convergence rate under different metrics and the $\sqrt{n}$ asymptotic normality of the parametric component b.

A.1. Consistency and Convergence Rates

In this section, we first introduce a strong norm $\|\cdot\|_s$ in Newey and Powell (2003) which would be used to show the consistency of the sieve estimator and then the Fisher norm, $\|\cdot\|$, in which the sieve estimator is consistent with a rate faster than $n^{-1/4}$.

For $\alpha \equiv (f_1, \theta, f_2)'$,

$$\|\alpha\|_s = \|b\|_E + \|\lambda\|_{s,\omega} + \sum_{s=1}^{2} \|f_s\|_{s,\omega} \tag{12}$$

where $\|b\|_E$ is the Euclidean norm and $\|f_s\|_{s,\omega} \equiv \sup_\xi |f_s(\xi)\omega(\xi)|$ with $\omega(\xi) = (1 + \|\xi\|_E^2)^{-\varsigma/2}$, $\varsigma > 0$. Since the supports of the unobserved variables v and ε could be unbounded, the weighting function w is introduced to deal with unbounded support and has been used in Chen, Hansen, and Scheinkman (1997), Chen, Hong, and Tamer (2005) and Hu and Schennach

(2008a). We make the following assumptions:

Assumption A.2. *(i) The data* $\{(Z_{it})_{i=1}^n\}$ *are i.i.d.; (ii) The density of* Z_t*,* f_{Z_t}*, satisfies* $\int \omega(\xi)^{-2} f_{Z_t}(\xi) d\xi < \infty$*.*

Assumption A.3. *(i)* $b_0 \in \mathcal{B}$*, a compact subset of* $\mathbb{R}^b$*; (ii) Assumption A.1 holds under the norm* $\|\alpha\|_s$*.*

Assumption A.4. *(i) For any* $\alpha \in \mathcal{A}$*, there exists* $\Pi_n \alpha \in \mathcal{A}_n$ *such that* $\|\Pi_n \alpha - \alpha\|_s = o(1)$*; (ii)* $k_{ni} \to +\infty$ *and* $k_{ni}/n \to 0$ *for* $i = 1, \lambda, 2$*.*

Definition A.1. $\ln f_{Z_t}(z_t; \alpha)$ *is Hölder continuous with respect to* $\alpha \in \mathcal{A}_n$ *if there exists a measurable function* $c_h(Z_t)$ *with* $E\{c_h(Z_t)^2\} < \infty$ *such that, for all* $\alpha_1, \alpha_2 \in \mathcal{A}$*, and* Z_t*, we have*

$$|\ln f_{Z_t}(z_t; \alpha_1) - \ln f_{Z_t}(z_t; \alpha_2)| \leq c_h(Z_t) \|\alpha_1 - \alpha_2\|_s. \tag{13}$$

Assumption A.5. *(i)* $E\{|\ln f_{Z_t}(z_t; \alpha)|^2\}$ *is bounded; (ii) There exits a measurable function* $\widetilde{h}(Z_t)$ *with* $E\{\widetilde{h}(Z_t)^2\} < \infty$ *such that, for any* $\bar{\alpha}_{12} = (\bar{f}_1, \bar{\theta}, \bar{f}_2, \bar{f}_3)$ *and* $\bar{\omega}(z_t, \varepsilon) = \left[1, \omega^{-1}(x_{t+1}, x_t, u_t), \omega^{-1}(\varepsilon), \omega^{-1}(x_t, y_{t-1}, x_{t-1}, u_t))\right]'$*, we have* $|h_1(z_t, \bar{\alpha}_{12}, \bar{\omega})| < \widetilde{h}(Z_t)$*. (The definition of* $h_1(z_t, \bar{\alpha}_{12}, \bar{\omega})$ *can be found in Eq. (14)).*

Applying Theorem 4.1 in Newey and Powell (2003) or Theorem 3.1 of Chen (2007), we obtain the following lemma.

Lemma A.1. *Let* $\widehat{\alpha}_n$ *be the sieve MLE for* α_0 *identified in section 2 and Assumptions A.1-A.5 holds, then we have* $\|\widehat{\alpha}_n - \alpha_0\|_s = o_p(1)$*.*

The proof of Lemma A.1. After checking the conditions in Theorem 4.1 in Newey and Powell (2003), the only thing we have to show is that $\ln f_{Z_t}(z_t; \alpha)$ is Hölder continuous in α. The difference of $\ln f_{Z_t}(z_t; \cdot)$ at α_1 and α_2 is given by

$$\begin{aligned}
&\ln f_{Z_t}(z_t; \alpha_1) - \ln f_{Z_t}(z_t; \alpha_2) \\
&= \frac{d}{d\alpha} \ln f_{Z_t}(z_t; \bar{\alpha}_{12})[\alpha_1 - \alpha_2] \\
&= \frac{d}{dt} \ln f_{Z_t}(z_t; \bar{\alpha}_{12} + t(\alpha_1 - \alpha_2))\Big|_{t=0},
\end{aligned}$$

where $\bar{\alpha}_{12} = (\bar{f}_1, \bar{\theta}, \bar{f}_2)$, a mean value between α_1 and α_2, and $\bar{\alpha}_{12} + t(\alpha_1 - \alpha_2) = (\bar{f}_1 + t(f_{11} - f_{12}), \bar{\theta} + t(\theta_1 - \theta_2), \bar{f}_2 + t(f_{21} - f_{22}))$. Consider

$$\begin{aligned}
&\frac{d}{dt} \ln f_{Z_t}(z_t; \bar{\alpha}_{12} + t(\alpha_1 - \alpha_2))\Big|_{t=0} \\
&= \frac{1}{f_{Z_t}(z_t; \bar{\alpha}_{12})} \bigg(\int (f_{11} - f_{12}) f_{Y_{it}|X_{it}, Y_{it-1}, U_t}(y_{it}|x_{it}, y_{it-1}, u_t; \bar{\theta}) \bar{f}_2 du_t \\
&\quad + \int \bar{f}_1 \frac{d}{d\theta} f_{Y_{it}|X_{it}, Y_{it-1}, U_t}(y_{it}|x_{it}, y_{it-1}, u_t; \bar{\theta})(\theta_1 - \theta_2) \bar{f}_2 du_t \\
&\quad + \int \bar{f}_1 f_{Y_{it}|X_{it}, Y_{it-1}, U_t}(y_{it}|x_{it}, y_{it-1}, u_t; \bar{\theta})(f_{21} - f_{22}) du_t.
\end{aligned}$$

Then we can obtain the bounds for Hölder continuous as follows:

$$\begin{aligned}
&\left|\frac{d}{dt} \ln f_{Z_t}(z_t; \bar{\alpha}_{12} + t(\alpha_1 - \alpha_2))\right|_{t=0} \\
&\le \frac{1}{|f_{Z_t}(z_t; \bar{\alpha}_{12})|} \bigg(\int \left|\omega^{-1}(x_{t+1}, x_t, u_t) f_{Y_t|X_t, Y_{t-1}, U_t}(y_t|x_t, y_{t-1}, u_t; \bar{\theta}) \bar{f}_2\right| du_t \|f_{11} - f_{12}\|_s \\
&\quad + \int \left|\bar{f}_1 \frac{d}{d\theta} f_{Y_t|X_t, Y_{t-1}, U_t}(y_t|x_t, y_{t-1}, u_t; \bar{\theta}) \omega^{-1}(\varepsilon) \bar{f}_2\right| du_t \|\theta_1 - \theta_2\|_s \\
&\quad + \int \left|\bar{f}_1 f_{Y_t|X_t, Y_{t-1}, U_t}(y_t|x_t, y_{t-1}, u_t; \bar{\theta}) \omega^{-1}(x_t, y_{t-1}, x_{t-1}, u_t)\right| du_t \|f_{21} - f_{22}\|_s \\
&\le \frac{1}{|f_{Z_t}(z_t; \bar{\alpha}_{12})|} \bigg(\int \left|\omega^{-1}(x_{t+1}, x_t, u_t) f_{Y_t|X_t, Y_{t-1}, U_t}(y_t|x_t, y_{t-1}, u_t; \bar{\theta}) \bar{f}_2\right| du_t \\
&\quad + \int \left|\bar{f}_1 \frac{d}{d\theta} f_{Y_t|X_t, Y_{t-1}, U_t}(y_{it}|x_t, y_{t-1}, u_t; \bar{\theta}) \omega^{-1}(\varepsilon) \bar{f}_2\right| du_t \\
&\quad + \int \left|\bar{f}_1 f_{Y_t|X_t, Y_{t-1}, U_t}(y_t|x_t, y_{t-1}, u_t; \bar{\theta}) \omega^{-1}(x_t, y_{t-1}, x_{t-1}, u_t)\right| du_t \bigg) \|\alpha_1 - \alpha_2\|_s \\
&\equiv h_1(z_t, \bar{\alpha}, \bar{\omega}) \|\alpha_1 - \alpha_2\|_s,
\end{aligned} \tag{14}$$

where $\bar{\omega}(z_t, \varepsilon) \equiv \left[1, \omega^{-1}(x_{t+1}, x_t, v), \omega^{-1}(\varepsilon), \omega^{-1}(x_t, y_{t-1}, x_{t-1}, u_t)\right]'$. The property of Hölder continuity is ensured by Assumption A.5. A similar proof can also be found in that of Lemma 2 in Hu and Schennach (2008a). *Q.E.D.*

Lemma A.1 provides a consistency result under the metric $\|\cdot\|_s$ but the convergence rate under the metric is not faster enough to establish our semi-parametric asymptotic normality and $\sqrt{n}$ consistency result. In order to achieve this we consider the Fisher norm $\|\cdot\|$ in which $\widehat{\alpha}_n$ converges at a rate faster than $n^{-1/4}$. In addition, Lemma A.1 allows us to restrict the sieve estimator $\widehat{\alpha}_n$ to local $\|\cdot\|_s$-neighborhood around the true parameter α_0. For simplicity, we can assume the function space $\mathcal{A}$ is convex. For any $v \in \bar{V}$, we define the pathwise derivative as:

$$\frac{d \ln f_{Z_t}(z_t;\alpha)}{d\alpha}[v] \equiv \left.\frac{d \ln f_{Z_t}(z_t;\alpha+\tau v)}{d\tau}\right|_{\tau=0} \quad \text{a.s. } Z_t.$$

In particular, we can have it at the direction $[\alpha_1 - \alpha_2]$ evaluated at α_0 by:

$$\frac{d \ln f_{Z_t}(z_t;\alpha_0)}{d\alpha}[\alpha_1-\alpha_2] \equiv \frac{d \ln f_{Z_t}(z_t;\alpha_0)}{d\alpha}[\alpha_1-\alpha_0] - \frac{d \ln f_{Z_t}(z_t;\alpha_0)}{d\alpha}[\alpha_2-\alpha_0] \text{ a.s. } Z_t.$$

Expanding the pathwise derivative of $\ln f_{Z_t}(z_t;\alpha_0)$ gives:

$$\begin{aligned}
&\frac{d \ln f_{Z_t}(z_t;\alpha_0)}{d\alpha}[\alpha-\alpha_0] \\
&= \frac{1}{f_{Z_t}(z_t;\alpha_0)}\Bigg(\int (f_1 - f_{X_{t+1}|X_t,U_t}) f_{Y_t|X_t,Y_{t-1},U_t} f_{X_t,Y_{t-1},X_{t-1},U_t} du_t \\
&\quad + \int f_{X_{t+1}|X_t,U_t} \frac{d}{d\theta} f_{Y_{it}|X_{it},Y_{it-1},U_t}(\theta-\theta_0) f_{X_t,Y_{t-1},X_{t-1},U_t} du_t \\
&\quad + \int f_{X_{t+1}|X_t,U_t} f_{Y_{it}|X_{it},Y_{it-1},U_t}(f_2 - f_{X_t,Y_{t-1},X_{t-1},U_t}) du_t\Bigg).
\end{aligned}$$

Following the notation, for any $\alpha_1, \alpha_2 \in \mathcal{A}$ we define the Fisher norm:

$$\|\alpha_1-\alpha_2\|^2 \equiv \mathrm{E}\left\{\left(\frac{d \ln f_{Z_t}(z_t;\alpha_0)}{d\alpha}[\alpha_1-\alpha_2]\right)^2\right\}.$$

We make the following assumptions to obtain a rate faster than $n^{-1/4}$.

Assumption A.6. *Let k_n be the total number of sieve coefficients in the sieve estimator $\widehat{\alpha}_n$, i.e., $k_n = k_{n1} + d_b + k_{n\lambda} + k_{n2}$. $(k_n n^{-1/2} \ln n) \times \sup_{\xi \in (\mathcal{X}\times\mathcal{X}\times\mathcal{U}\cup\mathbb{R}\cup\mathcal{X}\times\mathcal{Y}\times\mathcal{X}\times\mathcal{U})} \|p^{k_n}(\xi)\|_E^2 = o(1)$.*

Assumption A.7. (i) *there exist a measurable function $c(Z_t)$ with $E\{c(Z_t)^4\} < \infty$ such that $|\ln f_{Z_t}(z_t;\alpha)| \le c(Z_t)$ for all Z_t and $\alpha \in \mathcal{A}_n$; (ii) $\ln f_{Z_t}(z_t;\alpha) \in \Lambda_c^{\tau,\omega}(\mathcal{X}\times\mathcal{Y}\times\mathcal{X}\times\mathcal{Y}\times\mathcal{X})$ with $\tau > d_z/2$, for all $\alpha \in \mathcal{A}_n$, where d_z is the dimension of Z_t.*

Assumption A.8. *$\mathcal{A}$ is convex in α_0, and $f_{Y_t|X_t,Y_{t-1},U_t}(Y_t|X_t,Y_{t-1},u_t;\theta)$ is pathwise differentiable at θ_0.*

Assumption A.9. $\ln N(\delta,\mathcal{A}_n) = O(k_n \ln(k_n/\delta))$ *where $N(\delta,\mathcal{A}_n)$ is the minimum number of balls with radius δ under the $\|\cdot\|_s$ norm covering $\mathcal{A}_n$.*

Assumption A.10. *There exists* c_1, $c_2 > 0$,

$$c_1 E\left(\ln \frac{f_{Z_t}(z_t;\alpha_0)}{f_{Z_t}(z_t;\alpha)}\right) \leq \|\alpha-\alpha_0\|^2 \leq c_2 E\left(\ln \frac{f_{Z_t}(z_t;\alpha_0)}{f_{Z_t}(z_t;\alpha)}\right)$$

holds for all $\alpha \in \mathcal{A}_n$ *with* $\|\alpha-\alpha_0\|_s = o(1)$.

Assumption A.11. *For any* $\alpha \in \mathcal{A}$, *there exists* $\Pi_n\alpha \in \mathcal{A}_n$ *such that* $\|\Pi_n\alpha-\alpha\| = o(k_n^{-\mu 1})$ *and* $k_n^{-\mu 1} = o(n^{-1/4})$.

The following lemma is a direct application of Theorem 3.1 of Ai and Chen (2003) and a similar proof can also be found in that of Theorem 2 in Hu and Schennach (2008b); we omit its proof.

Theorem A.1. *Suppose that* α_0 *is identified and Assumptions A.6-A.11 hold, then* $\|\widehat{\alpha}_n - \alpha_0\| = o_p(n^{-1/4})$.

A.2. Asymptotic Normality

In this section, we follow the semiparametric MLE framework of Hu and Schennach (2008b) to show the asymptotic normality of the parametric component b which represents the parameter of interest in dynamic panel data models. Let V be the space spanned by $\mathcal{A}-\alpha_0$ and $\bar{V}$ be completion of V under the Fisher norm $\|\cdot\|$. It follows that $(\bar{V}, \|\cdot\|)$ is a Hilbert space with the inner product

$$\langle v_1, v_2\rangle \equiv E\left\{\left(\frac{d}{d\alpha}\ln f_{Z_t}(z_t;\alpha_0)[v_1]\right)\left(\frac{d}{d\alpha}\ln f_{Z_t}(z_t;\alpha_0)[v_2]\right)\right\},$$

and $\langle v, v\rangle = \|v\|$. For any fixed and nonzero $\kappa \in \mathbb{R}^{d_b}$, $f_\kappa(\alpha-\alpha_0) \equiv \kappa'(b-b_0)$ is linear in $\alpha-\alpha_0$ and $f_\kappa(\alpha-\alpha_0)$ is a linear functional on $(\overline{V}, \|\cdot\|)$. Shen (1997) and Der Vaart (1991) show that $f(\alpha) \equiv \kappa' b$ is a bounded linear functional on $\bar{V}$ under the operator norm. That is:

$$|||f_\kappa||| \equiv \sup_{\{\alpha\in\mathcal{A}:\|\alpha-\alpha_0\|>0\}} \frac{|f_\kappa(\alpha-\alpha_0)|}{\|\alpha-\alpha_0\|} < \infty. \tag{15}$$

By the Riesz representation theorem, there exists $v^* \in \bar{V}$ such that for any $\alpha \in \mathcal{A}$, I have $f_\kappa(\alpha-\alpha_0) = \langle\alpha-\alpha_0, v^*\rangle$. and $\|f_\kappa\| = \|v^*\|$.[15] Denote $\bar{V} = \mathbb{R}^{d_b}\times\overline{W}$ and $\overline{W} \equiv \overline{\mathcal{F}_1^n \times \mathcal{M}^n \times \mathcal{F}_2^n} -$

[15] Stein (1956) pointed out that v^* yields the most difficult one-dimensional sub-problem. Begun, Hall, Huang, and Wellner (1983) mentioned that v^* represents a worst possible direction to nonparametric component for

$(f_{X_{t+1}|X_t,U_t}, \lambda, f_{X_t,Y_{t-1},X_{t-1},U_t})'$.[16] We can expand the first pathwise derivative out as follows:

$$\begin{aligned}\frac{d\ln f_{Z_t}(z_t;\alpha_0)}{d\alpha}[\alpha-\alpha_0] = & \frac{d\ln f_{Z_t}(z_t;\alpha_0)}{df_1}[f_1 - f_{X_{t+1}|X_t,U_t}] + \frac{d\ln f_{Z_t}(z_t;\alpha_0)}{db}[b-b_0] \\ & + \frac{d\ln f_{Z_t}(z_t;\alpha_0)}{d\lambda}[\lambda-\lambda_0] + \frac{d\ln f_{Z_t}(z_t;\alpha_0)}{df_2}[f_2 - f_{X_t,Y_{t-1},X_{t-1},U_t}].\end{aligned}$$

For each component b_j of b, $j = 1, 2, ..., d_b$, we define $w_j^* \in \overline{W}$ to be the solution to the following minimization problem associated with the denominator of the operator norm,

$$\begin{aligned}w_j^* = \arg\min_{w_j=(f_1,\lambda,f_2,f_3)'\in\overline{W}} E\Bigg[\Bigg(&\frac{d\ln f_{Z_t}(z_t;\alpha_0)}{db_j} - \frac{d\ln f_{Z_t}(z_t;\alpha_0)}{df_1}[f_1] \\ & - \frac{d\ln f_{Z_t}(z_t;\alpha_0)}{d\lambda}[\lambda] - \frac{d\ln f_{Z_t}(z_t;\alpha_0)}{df_2}[f_2]\Bigg)\Bigg].\end{aligned}$$

Define $w^* = (w_1^*, ..., w_{d_b}^*)$,

$$\frac{d\ln f_{Z_t}(z_t;\alpha_0)}{df}[w^*] = \left(\frac{d\ln f_{Z_t}(z_t;\alpha_0)}{dh}[w_1], ..., \frac{d\ln f_{Z_t}(z_t;\alpha_0)}{df}[w_{d_b}^*]\right),$$

and

$$D_{w^*}(z_t) \equiv \frac{d\ln f_{Z_t}(z_t;\alpha_0)}{d_b'} - \frac{d\ln f_{Z_t}(z_t;\alpha_0)}{df}[w^*]. \tag{16}$$

With these notation, we have

$$\|f_\kappa\|^2 = \sup_{\{\alpha\in\mathcal{A}:\|\alpha-\alpha_0\|>0\}} \frac{|f_\kappa(\alpha-\alpha_0)|^2}{\|\alpha-\alpha_0\|} = \kappa'\left(E\{D_{w^*}(Z_t)'D_{w^*}(Z_t)\}\right)^{-1}\kappa,$$

$v^* \equiv (v_b^*, v_h^*) \in \overline{V}$ with $v_b^* = \left(E\{D_{w^*}(Z_t)'D_{w^*}(Z_t)\right)^{-1}\kappa$ and $v_h^* = -w^* \times v_b^*$. In addition, $f_\kappa(\alpha - \alpha_0) = \kappa'(b-b_0) = \langle \alpha-\alpha_0, v^*\rangle$ and $\frac{d\ln f_{Z_t}(z_t;\alpha_0)}{d\alpha}[v^*] = D_{w^*}(z_t)v_b^*$. See Chen (2007) for detailed discussion about this linear functional approach. Therefore, the asymptotic distribution of parametric component $\widehat{b}_n$ reduces to when the linear functional f_κ is bounded and what is

estimating parametric component.

[16] $\overline{W}$ is a function space of nonparametric components.

the asymptotic distribution of $\langle \widehat{\alpha}_n - \alpha_0, v^* \rangle$. That is:

$$
\begin{aligned}
\kappa' \left(\widehat{b}_n - b_0 \right) &= \langle \widehat{\alpha}_n - \alpha_0, v^* \rangle \\
&= \frac{1}{n} \sum_{i=1}^{n} \frac{d \ln f_{Z_t}(z_{it}; \alpha_0)}{d\alpha} [v^*] + o_p(n^{-1/2}) \\
&= \frac{1}{n} \sum_{i=1}^{n} \kappa' \left(E\{ D_{w^*}(Z_t)' D_{w^*}(Z_t) \right)^{-1} D_{w^*}(z_{it})' + o_p(n^{-1/2}),
\end{aligned}
$$

and $\sqrt{n}(\widehat{b}_n - b_0) \to N(0, (E\{D_{w^*}(Z_t)' D_{w^*}(Z_t)\})^{-1})$.

We make the following sufficient conditions for the $\sqrt{n}-$normality of $\widehat{b}_n$ which are also conditions in Ai and Chen (2003) and Hu and Schennach (2008b):

Assumption A.12. *(i) $E\{D_{w^*}(Z_t)' D_{w^*}(Z_t)\}$ is positive-definite and bounded; (ii) $b_0 \in int(\mathcal{B})$.*

Assumption A.13. *There is a $v_n^* = (v_b^*, -\Pi_n w^* \times v_b^*) \in \mathcal{A}_n - \alpha_0$ such that $\|v_n^* - v^*\| = o_p(n^{-1/4})$.*

We use the $\sqrt{n}$ consistency results in the previous section to focus on a smaller neighbor of α_0, Define $\mathcal{N}_{on} \equiv \{\alpha \in \mathcal{A}_n : \|\alpha - \alpha_0\|_s = o(1), \|\alpha - \alpha_0\| = o(n^{-1/4})\}$ and $\mathcal{N}_o \equiv \{\alpha \in \mathcal{A} : \|\alpha - \alpha_0\|_s = o(1), \|\alpha - \alpha_0\| = o(n^{-1/4})\}$.

Assumption A.14. *There exits a measurable function $\widehat{h}(Z_t)$ with $E\{\widehat{h}(Z_t)^2\} < \infty$ such that, for any $\bar{\alpha} = (\bar{f}_1, \bar{\theta}, \bar{f}_2)$, we have*

$$
\left| h_2(z_t, \bar{\alpha}, \bar{\omega}) \right| + \left| h_1(z_t, \bar{\alpha}, \bar{\omega}) \right|^2 < \widehat{h}(Z_t). \tag{17}
$$

(The definition of $h_2(z_t, \bar{\alpha}, \bar{\omega})$ can be found in Eq. (20)).

For $\tilde{f} = f_1, \lambda, f_2$, denote

$$\frac{df_{Z_t}(z_t;\alpha_0)}{d\tilde{f}}[p^{k_{ni}}] = \left(\frac{df_{Z_t}(z_t;\alpha_0)}{d\tilde{f}}[p_1^{k_{ni}}], ..., \frac{df_{Z_t}(z_t;\alpha_0)}{d\tilde{f}}[p_{k_{ni}}^{k_{ni}}]\right)' \quad \forall i = 1, \lambda, 2,$$
$$\frac{df_{Z_t}(z_t;\alpha_0)}{db} = \left(\frac{df_{Z_t}(z_t;\alpha_0)}{db_1}, ..., \frac{df_{Z_t}(z_t;\alpha_0)}{db_{d_b}}\right)',$$
$$\frac{df_{Z_t}(z_t;\alpha_0)}{d\alpha}[p^{k_n}] = \left(\left(\frac{df_{Z_t}(z_t;\alpha_0)}{db}\right)', \left(\frac{df_{Z_t}(z_t;\alpha_0)}{df_1}[p^{k_{n1}}]\right)', \left(\frac{df_{Z_t}(z_t;\alpha_0)}{d\lambda}[p^{k_{n\lambda}}]\right)', \left(\frac{df_{Z_t}(z_t;\alpha_0)}{df_2}[p^{k_{n2}}]\right)'\right)', \tag{18}$$

and

$$\Omega_{k_n} = E\left\{\left(\frac{df_{Z_t}(z_t;\alpha_0)}{d\alpha}[p^{k_n}]\right)\left(\frac{df_{Z_t}(z_t;\alpha_0)}{d\alpha}[p^{k_n}]\right)'\right\}.$$

Assumption A.15. *The smallest eigenvalue of the matrix Ω_{k_n} is bounded away from zero, and $\|p_j^{k_{ni}}\|_{s,\omega}$ for $j = 1, 2, ..., k_{ni}$ uniformly in k_{ni}.*

Assumption A.16. *For all $\alpha \in \mathcal{N}_{on}$, there exists a measurable function $h(Z_t)$ with $E|h(Z_t)| < \infty$ such that*

$$\left|\frac{d^4 \ln f_{Z_t}(z_t;\bar{\alpha} + t(\alpha - \alpha_0))}{dt^4}\right|_{t=0} \leq h(Z_t)\|\alpha - \alpha_0\|_s^4. \tag{19}$$

Theorem A.2. *Suppose that α_0 is identified and Assumptions A.6-A.11 and A.12-A.16 hold, then $\sqrt{n}(\widehat{b}_n - b_0) \Rightarrow N(0, V^{-1})$ where $V = E\{D_{w^*}(Z_t)'D_{w^*}(Z_t)\}$.*

The proof of Theorem A.2. The likelihood function $f_{Z_t}(z_t;\alpha)$ has a similar expression as the likelihood function in Hu and Schennach (2008a). The proof there can directly apply to our case. We prove the results by showing an envelop condition on the second derivative of the likelihood function (Assumption A.15). Consider the term

$$
\begin{aligned}
&\left|\sup_{\alpha\in\mathcal{N}_{on}} \frac{d^2 \ln f_{Z_t}(z_t;\bar{\alpha}_1)}{d\alpha d\alpha'}[v_n, \alpha-\alpha_0]\right| \\
\leq{}& \sup_{\alpha\in\mathcal{N}_{on}} \left| \frac{1}{f_{Z_t}(z_t;\bar{\alpha}_1)} \frac{d^2 f_{Z_t}(z_t;\bar{\alpha}_1)}{d\alpha d\alpha'}[v_n, \alpha-\alpha_0] \right. \\
&\qquad \left. - \frac{d \ln f_{Z_t}(z_t;\bar{\alpha}_1)}{d\alpha}[v_n] \frac{d \ln f_{Z_t}(z_t;\bar{\alpha}_1)}{d\alpha}[\alpha-\alpha_0] \right| \\
\leq{}& \sup_{\alpha\in\mathcal{N}_{on}} \left(\left| \frac{1}{f_{Z_t}(z_t;\bar{\alpha}_1)} \frac{d^2 f_{Z_t}(z_t;\bar{\alpha}_1)}{d\alpha d\alpha'}[v_n, \alpha-\alpha_0] \right| \right. \\
&\qquad \left. + \left| \frac{d \ln f_{Z_t}(z_t;\bar{\alpha}_1)}{d\alpha}[v_n] \right| \left| \frac{d \ln f_{Z_t}(z_t;\bar{\alpha}_1)}{d\alpha}[\alpha-\alpha_0] \right| \right).
\end{aligned}
$$

Then

$$
\begin{aligned}
&\frac{d \ln f_{Z_t}(z_t;\bar{\alpha}_1)}{d\alpha}[\alpha-\alpha_0] \\
&= \frac{1}{f_{Z_t}(z_t;\bar{\alpha}_{12})} \left(\int [f_1 - f_{X_{t+1}|X_t,U_t}] f_{Y_t|X_t,Y_{t-1},U_t}(y_t|x_t,y_{t-1},u_t;\bar{\theta}) \bar{f}_2 du_t \right. \\
&\quad + \int \bar{f}_1 \frac{d}{d\theta} f_{Y_t|X_t,Y_{t-1},U_t}(y_t|x_t,y_{t-1},u_t;\bar{\theta})[\theta-\theta_0] \bar{f}_2 du_t \\
&\quad \left. + \int \bar{f}_1 f_{Y_t|X_t,Y_{t-1},U_t}(y_t|x_t,y_{t-1},u_t;\bar{\theta})[f_2 - f_{X_t,Y_{t-1},X_{t-1},U_t}] du_t \right).
\end{aligned}
$$

Therefore, similar to the derivation of Hölder continuity, we obtain

$$
\left| \frac{d \ln f_{Z_t}(z_t;\bar{\alpha}_1)}{d\alpha}[\alpha-\alpha_0] \right| \leq h_1(z_t,\bar{\alpha},\bar{\omega}) \|\alpha_1-\alpha_2\|_s.
$$

and

$$
\left| \frac{d \ln f_{Z_t}(z_t;\bar{\alpha}_1)}{d\alpha}[v_n] \right| \leq h_1(z_t,\bar{\alpha},\bar{\omega}) \|v_n\|_s,
$$

where $h_1(z_t,\bar{\alpha}_1,\bar{\omega})$ is defined in Eq. (14) and Assumption A.1. We expand out the term

$\frac{1}{f_{Z_t}(z_t;\bar{\alpha}_1)}\frac{d^2 f_{Z_t}(z_t;\bar{\alpha}_1)}{d\alpha d\alpha'}[v_n, \alpha-\alpha_0]$:

$$
\begin{aligned}
&\frac{1}{f_{Z_t}(z_t;\bar{\alpha}_1)}\frac{d^2 f_{Z_t}(z_t;\bar{\alpha}_1)}{d\alpha d\alpha'}[v_n, \alpha-\alpha_0] \\
=&\frac{1}{f_{Z_t}(z_t;\bar{\alpha}_{12})}\bigg\{ \int [v_n]_{f_1}[\theta-\theta_0]\frac{d}{d\theta}f_{Y_t|X_t,Y_{t-1},U_t}(y_t|x_t,y_{t-1},v;\bar{\theta})\bar{f}_2 dv \\
&+\int [v_n]_{f_1}[f_2 - f_{X_t,Y_{t-1},X_{t-1},U_t}]f_{Y_t|X_t,Y_{t-1},U_t}(y_t|x_t,y_{t-1},u_t;\bar{\theta})dv \\
&+\int [v_n]_{\theta}[f_1 - f_{X_{t+1}|X_t,U_t}]\frac{d}{d\theta}f_{Y_t|X_t,Y_{t-1},U_t}(y_t|x_t,y_{t-1},u_t;\bar{\theta})\bar{f}_2 du_t \\
&+\int [v_n]_{\theta}[\theta-\theta_0]\frac{d^2}{d\theta^2}f_{Y_t|X_t,Y_{t-1},V}(y_t|x_t,y_{t-1},u_t;\bar{\theta})\bar{f}_1\bar{f}_2 du_t \\
&+\int [v_n]_{\theta}[f_2 - f_{X_t,Y_{t-1},X_{t-1},U_t}]\frac{d}{d\theta}f_{Y_t|X_t,Y_{t-1},U_t}(y_t|x_t,y_{t-1},u_t;\bar{\theta})\bar{f}_1 du_t \\
&+\int [v_n]_{f_2}[f_1 - f_{X_{t+1}|X_t,U_t}]f_{Y_t|X_t,Y_{t-1},U_t}(y_t|x_t,y_{t-1},u_t;\bar{\theta})du_t \\
&+\int [v_n]_{f_2}[\theta-\theta_0]\frac{d}{d\theta}f_{Y_t|X_t,Y_{t-1},U_t}(y_t|x_t,y_{t-1},u_t;\bar{\theta})\bar{f}_1 du_t\bigg\}.
\end{aligned}
$$

Denote $\left[\omega^{-1}(x_{t+1},x_t,u_t),\omega^{-1}(\varepsilon),\omega^{-1}(x_t,y_{t-1},x_{t-1},u_t)\right] \equiv \left[\bar{\omega}_1^{-1},\bar{\omega}_\lambda^{-1},\bar{\omega}_2^{-1}\right]$.

The absolute value of the term can be bounded through

$$
\begin{aligned}
(20)\quad & \left|\frac{1}{f_{Z_t}(z_t;\bar{\alpha}_1)}\frac{d^2 f_{Z_t}(z_t;\bar{\alpha}_1)}{d\alpha d\alpha'}[v_n,\alpha-\alpha_0]\right| \\
\leq & \frac{1}{f_{Z_t}(z_t;\bar{\alpha}_{12})}\left\{\int\left|\bar{\omega}_1^{-1}\bar{\omega}_\lambda^{-1}\frac{d}{d\theta}f_{Y_t|X_t,Y_{t-1},U_t}(y_t|x_t,y_{t-1},u_t;\bar{\theta})\bar{f}_2\right|du_t\|[v_n]_{f_1}\|_s\|[\theta-\theta_0]\|_s\right. \\
& +\int\left|\bar{\omega}_1^{-1}\bar{\omega}_2^{-1}f_{Y_t|X_t,Y_{t-1},U_t}(y_t|x_t,y_{t-1},u_t;\bar{\theta})\right|du_t\|[v_n]_{f_1}\|_s\|[f_2-f_{X_t,Y_{t-1},X_{t-1},U_t}]\|_s \\
& +\int\left|\bar{\omega}_\lambda^{-1}\bar{\omega}_1^{-1}\frac{d}{d\theta}f_{Y_t|X_t,Y_{t-1},U_t}(y_t|x_t,y_{t-1},u_t;\bar{\theta})\bar{f}_2\right|du_t\|[v_n]_\theta\|_s\|[f_1-f_{X_{t+1}|X_t,U_t}]\|_s \\
& +\int\left|\bar{\omega}_\lambda^{-1}\bar{\omega}_\lambda^{-1}\frac{d^2}{d\theta^2}f_{Y_t|X_t,Y_{t-1},U_t}(y_t|x_t,y_{t-1},u_t;\bar{\theta})\bar{f}_1\bar{f}_2\right|du_t\|[v_n]_\theta\|_s\|[\theta-\theta_0]\|_s \\
& +\int\left|\bar{\omega}_\lambda^{-1}\bar{\omega}_2^{-1}\frac{d}{d\theta}f_{Y_t|X_t,Y_{t-1},U_t}(y_t|x_t,y_{t-1},u_t;\bar{\theta})\bar{f}_1\right|du_t\|[v_n]_\theta\|_s\|[f_2-f_{X_t,Y_{t-1},X_{t-1},U_t}]\|_s \\
& +\int\left|\bar{\omega}_2^{-1}\bar{\omega}_1^{-1}f_{Y_t|X_t,Y_{t-1},U_t}(y_t|x_t,y_{t-1},u_t;\bar{\theta})\right|du_t\|[v_n]_{f_2}\|_s\|[f_1-f_{X_{t+1}|X_t,U_t}]\|_s \\
& \left.+\int\left|\bar{\omega}_2^{-1}\bar{\omega}_\lambda^{-1}\frac{d}{d\theta}f_{Y_t|X_t,Y_{t-1},U_t}(y_t|x_t,y_{t-1},u_t;\bar{\theta})\bar{f}_1\right|du_t\|[v_n]_{f_2}\|_s\|[\theta-\theta_0]\|_s\right\} \\
\leq & \frac{1}{f_{Z_t}(z_t;\bar{\alpha}_{12})}\left\{\int\left|\bar{\omega}_1^{-1}\bar{\omega}_\lambda^{-1}\frac{d}{d\theta}f_{Y_t|X_t,Y_{t-1},U_t}(y_t|x_t,y_{t-1},u_t;\bar{\theta})\bar{f}_2\right|du_t\right. \\
& +\int\left|\bar{\omega}_1^{-1}\bar{\omega}_2^{-1}f_{Y_t|X_t,Y_{t-1},U_t}(y_t|x_t,y_{t-1},u_t;\bar{\theta})\right|du_t \\
& +\int\left|\bar{\omega}_\lambda^{-1}\bar{\omega}_1^{-1}\frac{d}{d\theta}f_{Y_t|X_t,Y_{t-1},U_t}(y_t|x_t,y_{t-1},u_t;\bar{\theta})\bar{f}_2\right|du_t \\
& +\int\left|\bar{\omega}_\lambda^{-1}\bar{\omega}_\lambda^{-1}\frac{d^2}{d\theta^2}f_{Y_t|X_t,Y_{t-1},U_t}(y_t|x_t,y_{t-1},u_t;\bar{\theta})\bar{f}_1\bar{f}_2\right|du_t \\
& +\int\left|\bar{\omega}_\lambda^{-1}\bar{\omega}_2^{-1}\frac{d}{d\theta}f_{Y_t|X_t,Y_{t-1},U_t}(y_t|x_t,y_{t-1},u_t;\bar{\theta})\bar{f}_1\right|du_t \\
& +\int\left|\bar{\omega}_2^{-1}\bar{\omega}_1^{-1}f_{Y_t|X_t,Y_{t-1},U_t}(y_t|x_t,y_{t-1},u_t;\bar{\theta})\right|du_t \\
& \left.+\int\left|\bar{\omega}_2^{-1}\bar{\omega}_\lambda^{-1}\frac{d}{d\theta}f_{Y_t|X_t,Y_{t-1},U_t}(y_t|x_t,y_{t-1},u_t;\bar{\theta})\bar{f}_1\right|du_t\right\}\|[v_n]\|_s\|[\alpha-\alpha_0]\|_s \\
& \equiv h_2(z_t,\bar{\alpha},\bar{\omega})\|[v_n]\|_s\|[\alpha-\alpha_0]\|_s.
\end{aligned}
$$

Combining all the bounds results in

$$
\begin{aligned}
& \left|\sup_{\alpha\in\mathcal{N}_{on}}\frac{d^2\ln f_{Z_t}(z_t;\bar{\alpha}_1)}{d\alpha d\alpha'}[v_n,\alpha-\alpha_0]\right| \\
\leq & \sup_{\alpha\in\mathcal{N}_{on}}\left(\left|h_2(z_t,\bar{\alpha},\bar{\omega})\right|+\left|h_1(z_t,\bar{\alpha},\bar{\omega})\right|^2\right)\|[v_n]\|_s\|[\alpha-\alpha_0]\|_s.
\end{aligned}
$$

Then Assumption A.15 guarantees the envelop condition and help us to control the linear approximation of the likelihood function near α_0. *Q.E.D.*

B. Restrictions on the Sieve Coefficients

This appendix describes the sieve MLE method used to estimate nonlinear dynamic panel data models. We provide detailed derivation of the method based on the likelihood function in Eq. (3). According to Eq. (3), there are several essential parts in the likelihood function, $f_{X_{t+1}|X_t,U_t}$, $f_{Y_t|X_t,Y_{t-1},U_t}$, and $f_{X_t,Y_{t-1},X_{t-1},U_t}$. While the specifications of $f_{Y_t|X_t,Y_{t-1},U_t}$ have been provided in Section 3, $f_{X_{t+1}|X_t,U_t}$, and $f_{X_t,Y_{t-1},X_{t-1},U_t}$ will be treated here. We will show sieve approximations and their constraints of those nonparametric components in the two examples. First, we introduce the sieve estimators for the covariate evolution, $f_{X_{t+1}|X_t,U_t}$, since we can use the same sieve approximates for them in the examples. Suppose that $x_t, u_t \in [0, l_1]$ and $(x_{t+1} - u_t) \in [-l_2, l_2]$.[17] The sieve estimators for the covariate evolution are constructed by Fourier series as follows:

$$f_1(x_{t+1}|x_t, u_t) = \sum_{i=0}^{i_n}\sum_{j=0}^{j_n}\sum_{k=0}^{k_n} \gamma_{ijk} p_{1i}(x_{t+1} - u_t) p_{2j}(x_t) p_{3k}(u_t),$$

where

$$p_{1i}(x_{t+1} - u_t) = \cos\frac{i\pi}{l_2}(x_{t+1} - u_t) \text{ or } \sin\frac{i\pi}{l_2}(x_{t+1} - u_t),$$
$$p_{2j}(x_t) = \cos\frac{j\pi}{l_1}x_t, \text{ and, } p_{3k}(u_t) = \cos\frac{k\pi}{l_1}u_t.$$

The conditional density restrictions $\int f_1(x_{t+1}|x_t, u_t)dx_{t+1} = 1 \ \forall x_t, u_t$ implies that a constant term in the sieve expression $f_1(x_{t+1}|x_t, u_t)$ equals $\frac{1}{2l_2}$.

Next, since $f_{X_{t+1}|X_t,U_t}$ is identified through Assumption 2.5, one thing remained to show is how to implement the normalization assumption in estimation. Consider the zero mode case, we have $\frac{\partial}{\partial x_{t+1}} f_1(x_{t+1}|x_t, u_t)|_{x_{t+1}=u_t} = 0$ for all x_t, u_t. By properties of the trigonometric functions, sieve coefficients related to terms like $\sin\frac{i\pi}{l_2}(x_{t+1} - u_t)$ survive. The identification restrictions impose restrictions on those coefficients.

[17] While the range of x_t can be obtained from data set, the domain of u_t depends on the modeling of unobserved heterogeneity. In our simulation, u_t is truncated on $[0, 2]$.

We consider the following simple case:

$$
\begin{aligned}
& f_1(x_{t+1}|x_t, u_t) \\
= & \left(c_{00} + c_{01}\cos\frac{\pi}{l_1}x_t + c_{02}\cos\frac{2\pi}{l_1}x_t\right)\left(a_{00} + a_{01}\cos\frac{\pi}{l_1}u_t + a_{02}\cos\frac{2\pi}{l_1}u_t\right) \\
& + \sum_{i=1}^{5}\left(c_{i0} + c_{i1}\cos\frac{\pi}{l_1}x_t + c_{i2}\cos\frac{2\pi}{l_1}x_t\right)\left(a_{00} + a_{01}\cos\frac{\pi}{l_1}u_t + a_{02}\cos\frac{2\pi}{l_1}u_t\right) \\
& \quad \times \cos\frac{i\pi}{l_2}(x_{t+1} - u_t) \\
& + \sum_{i=1}^{5}\left(d_{i0} + d_{i1}\cos\frac{\pi}{l_1}x_t + d_{i2}\cos\frac{2\pi}{l_1}x_t\right)\left(a_{00} + a_{01}\cos\frac{\pi}{l_1}u_t + a_{02}\cos\frac{2\pi}{l_1}u_t\right) \\
& \quad \times \sin\frac{i\pi}{l_2}(x_{t+1} - u_t).
\end{aligned}
$$

Then the density restriction gives $c_{00}a_{00} = \frac{1}{2l_2}$ and the identification restriction on the coefficients are

$$
\sum_{i=1}^{5} i d_{i0} = \sum_{i=1}^{5} i d_{i1} = \sum_{i=1}^{5} i d_{i2} = 0.
$$

As for nonparametric series estimator of $f_{X_t,Y_{t-1},X_{t-1},U_t}$, we have to separate it into two cases to fit into our examples. First, we handle with dynamic discrete choice models and a sieve estimator of $f_{X_t,Y_{t-1},X_{t-1},U_t}$ is given by the following:

$$
f_{X_t,Y_{t-1},X_{t-1},U_t} = \left(f_{X_t,X_{t-1},U_t|Y_{t-1}=0} f_{Y_{t-1}=0}\right)^{1-Y_{t-1}} \left(f_{X_t,X_{t-1},U_t|Y_{t-1}=1}(1 - f_{Y_{t-1}=0})\right)^{Y_{t-1}},
$$

where

$$
\left(f_{X_t,X_{t-1},U_t|Y_{t-1}=0}\right)^{1/2} = \sum_{i=0}^{i_n}\sum_{j=0}^{j_n}\sum_{k=0}^{k_n} \hat{a}_{ijk} q_i(x_t - x_{t-1} - u_t) q_j(x_{t-1}) q_k(u_t), \tag{21}
$$

and

$$
\left(f_{X_t,X_{t-1},U_t|Y_{t-1}=1}\right)^{1/2} = \sum_{i=0}^{i_n}\sum_{j=0}^{j_n}\sum_{k=0}^{k_n} \tilde{a}_{ijk} q_i(x_t - x_{t-1} - u_t) q_j(x_{t-1}) q_k(u_t). \tag{22}
$$

Our choice of $q_i's$ and $q_j's$ are the orthonormal Fourier series:

$$q_0(u_t) = \frac{1}{\sqrt{l_1}}, q_k(u_t) = \sqrt{\frac{2}{l_1}}\cos(\frac{k\pi}{l_1}u_t), q_0(x_{t-1}) = \frac{1}{\sqrt{l_1}} \text{ and } q_j(x_{t-1}) = \sqrt{\frac{2}{l_1}}\cos(\frac{j\pi}{l_1}x_{t-1})$$
$$q_0(x_t - x_{t-1} - u_t) = \frac{1}{\sqrt{l_2}} \text{ and } q_i(x_t - x_{t-1} - u_t) = \frac{1}{\sqrt{l_2}}\sin(\frac{i\pi}{l_2}(x_t - x_{t-1} - u_t)) \text{ or}$$
$$q_i(x_t - x_{t-1} - u_t) = \frac{1}{\sqrt{l_2}}\cos(\frac{i\pi}{l_2}(x_t - x_{t-1} - u_t)).$$

The density restrictions $\int f_{X_t,X_{t-1},U_t|Y_{t-1}=0}dx_t dx_{t-1}du_t = 1$ and $\int f_{X_t,X_{t-1},U_t|Y_{t-1}=1}dx_t dx_{t-1}du_t = 1$ amount to

$$\sum_{i=0}^{i_n}\sum_{j=0}^{j_n}\sum_{k=0}^{k_n}(\hat{a}_{ijk})^2 = 1, \text{ and } \sum_{i=0}^{i_n}\sum_{j=0}^{j_n}\sum_{k=0}^{k_n}(\tilde{a}_{ijk})^2 = 1.$$

We consider the case where $i_n = 5$, $j_n = 2$, and $k_n = 2$:

$$\begin{aligned}
&\left(f_{X_t,X_{t-1},U_t|Y_{t-1}=0}\right)^{1/2} \\
&= \left(\hat{c}_{00} + \hat{c}_{01}\cos\frac{\pi}{l_1}x_{t-1} + \hat{c}_{02}\cos\frac{2\pi}{l_1}x_{t-1}\right)\left(\hat{a}_{00} + \hat{a}_{01}\cos\frac{\pi}{l_1}u_t + \hat{a}_{02}\cos\frac{2\pi}{l_1}u_t\right) \\
&\quad + \sum_{i=1}^{5}\left(\hat{c}_{00} + \hat{c}_{01}\cos\frac{\pi}{l_1}x_{t-1} + \hat{c}_{02}\cos\frac{2\pi}{l_1}x_{t-1}\right)\left(\hat{a}_{i0} + \hat{a}_{i1}\cos\frac{\pi}{l_1}u_t + \hat{a}_{i2}\cos\frac{2\pi}{l_1}u_t\right) \\
&\qquad \times \cos\frac{i\pi}{l_2}(x_t - x_{t-1} - u_t) \\
&\quad + \sum_{i=1}^{5}\left(\hat{c}_{00} + \hat{c}_{01}\cos\frac{\pi}{l_1}x_{t-1} + \hat{c}_{02}\cos\frac{2\pi}{l_1}x_{t-1}\right)\left(\hat{b}_{i0} + \hat{b}_{i1}\cos\frac{\pi}{l_1}u_t + \hat{b}_{i2}\cos\frac{2\pi}{l_1}u_t\right) \\
&\qquad \times \sin\frac{i\pi}{l_2}(x_t - x_{t-1} - u_t).
\end{aligned}$$

We can similarly find the sieve expression of $f_{X_{it},Y_{it-1},X_{it-1},U_{it}}$ in dynamic censor models. Eq. (21) is still applicable for $Y_{t-1} = 0$ part,

$$\left(f_{X_t,X_{t-1},U_t|Y_{t-1}=0}\right)^{1/2} = \sum_{i=0}^{i_n}\sum_{j=0}^{j_n}\sum_{k=0}^{k_n}\hat{a}_{ijk}q_i(x_t - x_{t-1} - u_t)q_j(x_{t-1})q_k(u_t).$$

Suppose that $y_{t-1} \in (0, l_3]$. Consider

$$\left(f_{X_t,Y_{t-1}>0,X_{t-1},U_t}\right)^{1/2} = \sum_{i=0}^{i_n}\sum_{j=0}^{j_n}\sum_{k=0}^{k_n}\sum_{l=0}^{l_n}\tilde{a}_{ijkl}\tilde{q}_i(x_t - x_{t-1} - u_t)\tilde{q}_j(x_{t-1})\tilde{q}_k(u_t)\tilde{q}_l(y_{t-1}).$$

The density restriction $\int f_{X_t,X_{t-1},U_t|Y_{t-1}=0} + \left(\int f_{X_t,Y_{t-1}>0,X_{t-1},U_t} dy_{t-1}\right) dx_t dx_{t-1} du_t = 1$ is

$$\sum_{i=0}^{i_n}\sum_{j=0}^{j_n}\sum_{k=0}^{k_n}\left(\hat{a}_{ijk}\right)^2 + \sum_{i=0}^{i_n}\sum_{j=0}^{j_n}\sum_{k=0}^{k_n}\sum_{l=0}^{l_n}\left(\tilde{a}_{ijkl}\right)^2 = 1. \tag{23}$$

In the simulation of the censored tobit model in Section 4, our choice of $Y_{t-1} > 0$ part is $i_n = 5$, $j_n = 2$, $k_n = 2$, and $l_n = 2$.

References

Ahn, S., and P. Schmidt (1995): "Efficient Estimation of Models for Dynamic Panel Data," *Journal of Econometrics*, 68(1), 5–28.

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

Anderson, T., and C. Hsiao (1982): "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of econometrics*, 18(1), 47–82.

Arellano, M., and S. Bond (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *The Review of Economic Studies*, pp. 277–297.

Arellano, M., and O. Bover (1995): "Another Look at the Instrumental Variable Estimation of Error Component Models," *Journal of Econometrics*, 68(1), 29–51.

Begun, J., W. Hall, W. Huang, and J. Wellner (1983): "Information and Asymptotic Efficiency in Parametric-nonparametric Models," *Ann. Statist*, 11(2), 432–452.

Chamberlain, G. (1984): "Panel Data," *Handbook of Econometrics*, II, 1247–1318.

Chay, K., H. Hoynes, and D. Hyslop (2001): "A Non-experimental Analysis of True State Dependence in Monthly Welfare Participation Sequences," University of California, Berkeley.

Chen, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," *Handbook of Econometrics*, 6.

Chen, X., L. Hansen, and J. Scheinkman (1997): "Shape-preserving Estimation of Diffusions," *manuscript, University of Chicago, Dept. of Economics.*

Chen, X., H. Hong, and E. Tamer (2005): "Measurement Error Models with Auxiliary Data," *Review of Economic Studies*, 72(2), 343–366.

Chen, X., and X. Shen (1998): "Sieve Extremum Estimates for Weakly Dependent Data," *Econometrica*, pp. 289–314.

Chintagunta, P., E. Kyriazidou, and J. Perktold (2001): "Panel Data Analysis of Household Brand Choices," *Journal of Econometrics*, 103(1-2), 111–153.

Contoyannis, P., A. Jones, and N. Rice (2004): "Simulation-based Inference in Dynamic Panel Probit Models: An Application to Health," *Empirical Economics*, 29(1), 49–77.

Der Vaart, V. (1991): "AW 1991. On Differentiable Functionals," *Ann. Statist*, 19, 178–204.

Evdokimov, K. (2009): "Identification and Estimation of a Nonparametric Panel Data Model with Unobserved Heterogeneity," .

Gourieroux, C., and A. Monfort (1993): "Simulation-based Inference: A Survey with Special Reference to Panel Data Models," *Journal of Econometrics*, 59(1-2), 5–33.

Hajivassiliou, V. (1993): "Simulation Estimation Methods for Limited Dependent Variable Models," *Handbook of Statistics*, 59.

Hajivassiliou, V., and P. Ruud (1994): "Classical Estimation Methods for LDV Models Using Simulation," *Handbook of econometrics*, 4, 2383–2441.

Halliday, T. (2002): "Heterogeneity, State Dependence and Health," Princeton University.

Heckman, J. (1978): "Simple Statistical Models for Discrete Panel Data Developed and Applied to Test the Hypothesis of True State Dependence Against the Hypothesis of Spurious State Dependence," in *Annales de l'INSEE*, pp. 227–269. Institut national de la statistique et des études économiques.

——— (1981a): "Statistical Models for Discrete Panel Data," *Structural analysis of discrete data with econometric applications.*

——— (1981b): "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process," *Structural Analysis of Discrete Data with Econometric Applications*, pp. 114–178.

Heckman, J., and R. Willis (1977): "A Beta-logistic Model for the Analysis of Sequential Labor Force Participation by Married Women," *The Journal of Political Economy*, pp. 27–58.

Honoré, B. (1993): "Orthogonality Conditions for Tobit Models with Fixed Effects and Lagged Dependent Variables," *Journal of Econometrics*, 59(1-2), 35–61.

Honoré, B., and L. Hu (2004): "Estimation of Cross Sectional and Panel Data Censored Regression Models with Endogeneity," *Journal of Econometrics*, 122(2), 293–316.

Honoré, B., and E. Kyriazidou (2000): "Panel Data Discrete Choice Models with Lagged Dependent Variables," *Econometrica*, pp. 839–874.

Honoré, B., and E. Tamer (2006): "Bounds on Parameters in Panel Dynamic Discrete Choice Models," *Econometrica*, pp. 611–629.

Hu, L. (2000): "Estimating a Censored Dynamic Panel Data Model with an Application to Earnings Dynamics," *Northwestern University.*

——— (2002): "Estimation of a Censored Dynamic Panel Data Model," *Econometrica*, pp. 2499–2517.

Hu, Y., and S. Schennach (2008a): "Instrumental Variable Treatment of Nonclassical Measurement Error models," *Econometrica*, 76(1), 195–216.

——— (2008b): "Supplement to Instrumental Variable Treatment of Nonclassical Measurement Error models," *Econometrica*, 76(1), 195–216.

Hyslop, D. (1995): "State Dependence, Serial Correlation and Heterogeneity in Intertemporal Participation Behavior: Monte Carlo Evidence and Empirical Results for Married Women," *Industrial Relations Section Working Paper #347, Princeton University.*

——— (1999): "State Dependence, Serial Correlation and Heterogeneity in Intertemporal Labor Force Participation of Married Women," *Econometrica*, 67(6), 1255–1294.

KEANE, M. (1993): "Simulation Estimation for Panel Data Models with Limited Dependent Variables," *Handbook of Statistics*, 11, 545–571.

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

SHEN, X. (1997): "On Methods of Sieves and Penalization," *Annals of Statistics*, 25, 2555–2591.

STEIN, C. (1956): "Efficient Nonparametric Testing and Estimation," in *Proc. Third Berkeley Symp. Math. Statist. Probab*, vol. 1, pp. 187–195.

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

Table 1: Monte Carlo Simulation of Semiparametric Probit model (n=250)

		Parameters		
N	DGP	β_0	β_1	γ
DGP I:	true value	0	-1	0
	mean estimate	0.001	-1.091	-0.012
	median estimate	0.014	-1.017	-0.005
	standard error	0.155	0.569	0.127
DGP II:	true value	0	-1	0
	mean estimate	-0.013	-1.117	-0.014
	median estimate	-0.005	-1.099	-0.005
	standard error	0.112	0.621	0.140
DGP III:	true value	0	-1	1
	mean estimate	0.005	-1.281	1.023
	median estimate	0.001	-1.275	0.901
	standard error	0.116	0.605	0.716
DGP IV:	true value	0	-1	1
	mean estimate	-0.015	-1.241	0.923
	median estimate	0.007	-1.256	0.880
	standard error	0.150	0.788	1.060
DGP V:	true value	0	-1	1
	mean estimate	0.003	-1.297	1.154
	median estimate	0.002	-1.182	0.942
	standard error	0.311	0.966	1.350

Note: The simulated date has 7 periods but only last 3 periods are used to construct the sieve MLE in the semi-parametric probit model. Standard errors of the parameters are computed by using sample standard deviation of 100 replications.

Table 2: Monte Carlo Simulation of Semiparametric Probit model (n=500)

		Parameters		
N	DGP	β_0	β_1	γ
DGP I:	true value	0	-1	0
	mean estimate	0.013	-1.097	-0.015
	median estimate	0.011	-1.080	-0.019
	standard error	0.121	0.362	0.111
DGP II:	true value	0	-1	0
	mean estimate	-0.012	-0.983	-0.039
	median estimate	-0.011	-1.037	-0.043
	standard error	0.100	0.529	0.107
DGP III:	true value	0	-1	1
	mean estimate	0.019	-1.173	0.927
	median estimate	0.018	-1.173	0.919
	standard error	0.109	0.803	0.672
DGP IV:	true value	0	-1	1
	mean estimate	0.028	-0.972	1.018
	median estimate	0.022	-0.918	1.020
	standard error	0.106	0.661	0.870
DGP V:	true value	0	-1	1
	mean estimate	0.006	-0.971	1.100
	median estimate	0.011	-1.000	1.037
	standard error	0.089	0.451	0.529

Note: The simulated date has 7 periods but only last 3 periods are used to construct the sieve MLE in the semi-parametric probit model. Standard errors of the parameters are computed by using sample standard deviation of 100 replications.

Table 3: Monte Carlo Simulation of Semiparametric Probit model (n=1000)

		Parameters		
N	DGP	β_0	β_1	γ
DGP I:	true value	0	-1	0
	mean estimate	-0.010	-1.003	0.008
	median estimate	-0.004	-0.995	0.005
	standard error	0.095	0.499	0.128
DGP II:	true value	0	-1	0
	mean estimate	-0.007	-1.010	-0.013
	median estimate	-0.008	-1.042	-0.007
	standard error	0.093	0.482	0.095
DGP III:	true value	0	-1	1
	mean estimate	-0.001	-1.054	1.025
	median estimate	-0.011	-1.149	0.953
	standard error	0.011	0.676	0.805
DGP IV:	true value	0	-1	1
	mean estimate	0.005	-0.801	1.104
	median estimate	0.008	-0.771	1.069
	standard error	0.098	0.600	0.468
DGP V:	true value	0	-1	1
	mean estimate	-0.021	-0.815	1.084
	median estimate	-0.006	-0.850	1.036
	standard error	0.111	0.435	0.529

Note: The simulated date has 7 periods but only last 3 periods are used to construct the sieve MLE in the semi-parametric probit model. Standard errors of the parameters are computed by using sample standard deviation of 100 replications.

Table 4: Monte Carlo Simulation of Semiparametric Tobit model (n=250)

		Parameters			
N	DGP	β_0	β_1	γ	σ_ξ^2
DGP I:	true value	0	-1	0	0.5
	mean estimate	-0.012	-0.858	0.006	0.463
	median estimate	0.001	-0.949	-0.013	0.439
	standard error	0.103	0.396	0.118	0.218
DGP II:	true value	0	-1	0	0.5
	mean estimate	0.016	-0.932	0.025	0.495
	median estimate	0.004	-0.934	0.018	0.489
	standard error	0.101	0.353	0.127	0.182
DGP III:	true value	0	-1	1	0.5
	mean estimate	-0.018	-0.835	0.968	0.458
	median estimate	-0.001	-0.911	0.992	0.453
	standard error	0.147	0.339	0.362	0.189
DGP IV:	true value	0	-1	1	0.5
	mean estimate	-0.02260	-0.906	0.912	0.486
	median estimate	-0.025	-0.891	0.922	0.470
	standard error	0.111	0.357	0.471	0.210
DGP V:	true value	0	-1	1	0.5
	mean estimate	0.013	-0.981	0.934	0.478
	median estimate	0.015	-1.023	0.916	0.489
	standard error	0.120	0.276	0.300	0.151

Note: The simulated date has 7 periods but only last 3 periods are used to construct the sieve MLE in the semi-parametric Tobit models. Standard errors of the parameters are computed by using sample standard deviation of 100 replications.

Table 5: Monte Carlo Simulation of Semiparametric Tobit model (n=500)

		Parameters			
N	DGP	β_0	β_1	γ	σ_ξ^2
DGP I:	true value	0	-1	0	0.5
	mean estimate	0.010	-0.890	0.004	0.483
	median estimate	0.007	-0.873	0.006	0.472
	standard error	0.116	0.465	0.097	0.168
DGP II:	true value	0	-1	0	0.55
	mean estimate	-0.012	-0.885	0.011	0.480
	median estimate	-0.018	-0.896	0.004	0.481
	standard error	0.112	0.356	0.101	0.170
DGP III:	true value	0	-1	1	0.5
	mean estimate	0.021	-0.803	1.016	0.483
	median estimate	0.015	-0.845	1.009	0.498
	standard error	0.106	0.376	0.254	0.180
DGP IV:	true value	0	-1	1	0.5
	mean estimate	0.009	-0.861	0.991	0.443
	median estimate	0.018	-0.908	0.972	0.445
	standard error	0.120	0.262	0.261	0.161
DGP V:	true value	0	-1	1	0.5
	mean estimate	-0.012	-0.935	0.930	0.476
	median estimate	-0.015	-0.965	0.937	0.467
	standard error	0.128	0.278	0.289	0.166

Note: The simulated date has 7 periods but only last 3 periods are used to construct the sieve MLE in the semi-parametric Tobit models. Standard errors of the parameters are computed by using sample standard deviation of 100 replications.

Table 6: Monte Carlo Simulation of Semiparametric Tobit model (n=1000)

		Parameters			
N	DGP	β_0	β_1	γ	σ_ξ^2
DGP I:	true value	0	-1	0	0.5
	mean estimate	0.013	-0.850	0.021	0.479
	median estimate	0.011	-0.947	0.009	0.480
	standard error	0.103	0.377	0.121	0.141
DGP II:	true value	0	-1	0	0.5
	mean estimate	0.017	-0.874	-0.001	0.478
	median estimate	0.021	-0.896	-0.017	0.472
	standard error	0.099	0.291	0.127	0.165
DGP III:	true value	0	-1	1	0.5
	mean estimate	0.001	-0.739	1.002	0.508
	median estimate	-0.006	-0.810	1.011	0.502
	standard error	0.115	0.340	0.313	0.167
DGP IV:	true value	0	-1	1	
	mean estimate	0.010	-0.857	1.019	0.501
	median estimate	0.014	-0.897	1.009	0.496
	standard error	0.088	0.324	0.269	0.152
DGP V:	true value	0	-1	1	0.5
	mean estimate	-0.003	-0.968	0.923	0.500
	median estimate	-0.010	-0.982	0.907	0.478
	standard error	0.092	0.271	0.254	0.153

Note: The simulated date has 7 periods but only last 3 periods are used to construct the sieve MLE in the semi-parametric Tobit models. Standard errors of the parameters are computed by using sample standard deviation of 100 replications.

Table 7: Sample Characteristics

Variables	Full Sample (1)	Employed 7 years (2)	Employed 0 years (3)	Single Transition from Work (4)	Single Transition to Work (5)	Multiple Transitions (6)
Age	35.38	35.22	39.90	35.66	35.81	33.65
	(0.22)	(0.31)	(0.80)	(0.84)	(0.65)	(0.44)
Education[18]	12.99	13.34	11.88	12.85	13.04	12.74
	(0.05)	(0.08)	(0.17)	(0.18)	(0.14)	(0.10)
Race (1=Black)	0.21	0.24	0.23	0.17	0.16	0.19
	(0.01)	(0.01)	(0.03)	(0.03)	(0.02)	(0.02)
No. Children[19]	0.25	0.20	0.24	0.34	0.24	0.33
aged 0-2 years	(0.01)	(0.01)	(0.03)	(0.02)	(0.02)	(0.02)
No. Children	0.27	0.22	0.26	0.26	0.33	0.36
aged 3-5 years	(0.01)	(0.01)	(0.03)	(0.02)	(0.02)	(0.02)
No. Children	0.96	0.92	0.96	0.67	1.21	1.01
aged 6-17 years	(0.02)	(0.03)	(0.07)	(0.06)	(0.06)	(0.04)
Husband's Labor[20]	27.30	25.85	32.59	28.49	29.93	26.40
Income ($1000)	(0.38)	(0.49)	(1.73)	(1.26)	(1.48)	(0.64)
Participation	0.71	1	0	0.51	0.54	0.57
	(0.01)	–	–	(0.02)	(0.02)	(0.01)
No. years worked[21]						
zero	9.34	–	100	–	–	–
one	5.90	–	–	20.79	14.54	10.40
two	5.51	–	–	15.73	15.86	9.98
three	6.29	–	–	12.92	14.54	14.97
four	7.18	–	–	11.24	12.33	20.37
five	9.39	–	–	15.73	20.26	24.32
six	9.29	–	–	23.60	22.47	19.96
seven	47.10	100	–	–	–	–
Sample size	1752	825	164	153	196	414

Note: Standard error of means $\sigma/\sqrt{n}$ in parentheses. Sample selection criteria: continuously married couples, aged 18-60 in 1980, with positive husband's annual earnings and hours worked each year.

[18] Years of Education are imputed from the following categorical scheme: 1 ='0-5 grades' (2.5 years); 2 ='6-8' (7 years); 3 ='9-11' (10 years); 4 ='12' (12 years); 5 ='12 plus non-academic training' (13 years); 6 ='some college' (14 years); 7 ='college degree, not advanced' (16 years); 8 ='college advanced degree' (18 years). Education is measured as the highest level reported in the 1980-86 surveys.

[19] Sample averages: child variables based on 8 observations from waves 12-19 of the PSID; participation and male earnings based on 7 observations form 1979 to 1985.

[20] The amounts are computed in constant (1987) dollars deflated by the consumer price index (CPI).

[21] Column percentages.

Table 8: Estimates of Married Women's Participation Outcomes

	Static Probit+RE (1)	MSL, RE AR(1)+SD(1) (2)	Semi-parametric Probit (3)
y_{t-1}	–	1.117	1.089
	–	(0.528)	(0.077)
y_{mp}	-0.312	-0.007	-0.221
	(0.045)	(0.017)	(0.012)
y_{mt}	-0.1060	-0.004	-0.106
	(0.026)	(0.028)	(0.056)
#Kid0-2$_{t-1}$	-0.022	-0.117	-0.055
	(0.010)	(0.013)	(0.048)
#Kid0-2$_t$	-0.330	-0.380	-0.316
	(0.021)	(0.145)	(0.061)
#Kid3-5$_t$	-0.400	-0.206	-0.137
	(0.015)	(0.027)	(0.028)
#Kid6-17$_t$	-0.120	-0.056	-0.062
	(0.011)	(0.037)	(0.011)
Cov. Parameters			
σ_v^2	0.786	0.313	–
	(0.071)	(0.323)	–
ρ	–	-0.146	–
	–	(0.140)	–

Note: Bootstrap standard errors are reported in parentheses, using 100 bootstrap replications. The models in the first two columns are estimated using full seven years of data but the last two columns are estimated over three-period data.

Table 9: Predicted Frequencies of Married Women's Participation Outcomes

	Sample Distribution (1)	Static Probit+RE (2)	MSL, RE AR(1)+SD(1) (3)	Semi-parametric Probit (4)
No. years worked				
zero	9.34	12.32	6.26	12.25
	—	(0.003)	(0.005)	(0.003)
one	5.90	15.15	5.20	10.07
	—	(0.005)	(0.005)	(0.004)
two	5.51	7.09	5.70	2.96
	—	(0.005)	(0.005)	(0.004)
three	6.29	6.14	7.19	2.69
	—	(0.005)	(0.006)	(0.003)
four	7.18	6.57	9.11	2.74
	—	(0.005)	(0.006)	(0.003)
five	9.39	8.42	12.91	3.72
	—	(0.006)	(0.007)	(0.004)
six	9.29	21.93	19.72	26.42
	—	(0.006)	(0.009)	(0.004)
seven	47.10	22.38	33.92	39.15
	—	(0.004)	(0.008)	(0.004)
Total	100	100	100	100

Note: Frequencies are computed as average values of 1000 predicted outcomes of 7 periods. They are reported in percentages and their standard deviations are reported in parentheses. The unobserved covariate U_{it} in the Semi-parametric Probit model is generated using the estimated parameters (σ_v^2, ρ) in column (2).