Journal of Econometrics 175 (2013) 116–131

Contents lists available at SciVerse ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Identification and estimation of nonlinear dynamic panel data models with unobserved covariates[☆]

Ji-Liang Shiu[a,*], Yingyao Hu[b]

[a] Hanqing Advanced Institute of Economics and Finance, Renmin University of China, Beijing 100872, PR China

[b] Department of Economics, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218, United States

ARTICLE INFO

Article history:
Received 8 July 2010
Received in revised form
29 January 2013
Accepted 5 March 2013
Available online 19 March 2013

Keywords:
Nonlinear dynamic panel data model
Dynamic discrete choice model
Dynamic censored model
Nonparametric identification
Initial condition
Correlated random effects
Unobserved heterogeneity
Unobserved covariate
Endogeneity

ABSTRACT

This paper considers nonparametric identification of nonlinear dynamic models for panel data with unobserved covariates. 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 two periods of the dependent variable Y_{it} and three periods of the covariate X_{it}. The main identifying assumptions include high-level injectivity restrictions and require that the evolution of the observed covariates depends on the unobserved 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 properties of the sieve MLE and investigate the finite sample properties of these sieve-based estimators 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).

1. Introduction

There are very few papers that provide full nonparametric identification of panel data models in the existing literature. This paper provides sufficient conditions for nonparametric identification of 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:

$$Y_{it} = g\left(X_{it}, Y_{it-1}, U_{it}, \xi_{it}\right), \quad \forall i = 1, \ldots, N;\ t = 1, \ldots, T-1, \tag{1}$$

where g is an unknown nonstochastic function, U_{it} is an unobserved covariate correlated with other observed explanatory variables (X_{it}, Y_{it-1}), and ξ_{it} stands for a random shock independent of all other explanatory variables $(X_{it}, Y_{it-1}, U_{it})$. The focuses of the above model are on the cases in which the time dimension, T, is fixed and the cross section dimension, N, grows without bound. 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.

If the unobserved heterogeneity V_i is treated as a parameter for each i, then both V_i and other unknown parameters need to be estimated for the model (1). When T tends to infinity, the MLE provides a consistent estimator for V_i and other unknown parameters. However, T is fixed and usually small for the panel data model considered here, and therefore, there are not enough observations to estimate these parameters. The model suffers from an incidental parameters problem (Neyman and Scott, 1948). In this paper, the unobserved heterogeneity, V_i, is treated as an unobservable random variable which may be correlated with observed covariates from the same individual. This correlated random

☆ Ji-Liang Shiu acknowledges support from the National Science Council of Taiwan via Grant 98-2410-H-194-118. The authors would like to thank Ryan Bush and Chin-Wei Yang for proofreading the draft. Helpful comments by Cheng Hsiao and three anonymous referees are acknowledged. All errors remain our own.

* Corresponding author. Tel.: +886 5 2720411; fax: +886 5 2720816.
E-mail addresses: jishiu@ruc.edu.cn (J.-L. Shiu), yhu@jhu.edu (Y. Hu).

0304-4076/$ – see front matter Crown Copyright © 2013 Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jeconom.2013.03.001

effect[1] approach (treating V_i as a random variable correlated with the covariates) allows us to integrate out unobserved variables to construct sieve MLE. This reduces potential computational burden from the incidental parameters problem for sieve MLE estimators in the estimation.[2] The transitory component η_{it} may be a function of all the time-varying RHS variables in the history, i.e., $\eta_{it} = \varphi\left(\{X_{i\tau}, Y_{i\tau-1}, \xi_{i\tau}\}_{\tau=0,1,\ldots,t-1}\right)$ for some function φ.[3] Both observed explanatory variables X_{it} and Y_{it-1} become endogenous if the unobserved covariate U_{it} is ignored. In this paper, we provide assumptions, including high-level injectivity restrictions, under which the distribution of Y_{it} conditional on $(X_{it}, Y_{it-1}, U_{it})$, i.e., $f_{Y_{it}|X_{it},Y_{it-1},U_{it}}$, is nonparametrically identified. The nonparametric identification of $f_{Y_{it}|X_{it},Y_{it-1},U_{it}}$ may lead to that of the general form of our model (1) under certain specifications of the distribution of the random shock ξ_{it}.

In this paper we adopt the correlated 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. Hu and Shum (2010) uses an identification technique described in Carroll et al. (2010). The two identification strategies are different although both use the spectral decomposition of linear operators. The discussion of the difference in the two techniques can be found in Carroll et al. (2010). The conditional independence assumptions in Hu and Shum (2010) are more general than those here but their results require five periods of data in the comparable setting. Our assumptions are more suitable for panel data models. Although some of our assumptions are stronger, our estimator requires only two periods of the dependent variable Y_{it} and three periods of the covariate X_{it}. This advantage is important because semi-nonparametric estimators usually require the sample size to be large.

The strength of our approach is that we provide nonparametric identification of nonlinear dynamic panel data model using two periods of the dependent variable Y_{it} and three periods of the covariate X_{it} without specifying initial conditions. The model may be described by, $f_{Y_{it}|X_{it},Y_{it-1},U_{it}}$, 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 $f_{Y_{it}|X_{it},Y_{it-1},U_{it}}$ 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 unobserved 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}.

The identification of $f_{Y_{it}|X_{it},Y_{it-1},U_{it}}$ leads to the identification of the general form of our model in Eq. (1). 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:

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

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

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

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 et al., 2004; Halliday, 2002), brand loyalty (Chintagunta et al., 2001), welfare participation (Chay et al., 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 covariates X_{it} are the nonlabor income or other observable characteristics in that period. The heterogeneity V_i is the unobserved individual skill level or motivation, while the idiosyncratic disturbance ξ_{it} denotes unexpected change of child-care cost or fringe benefit for married women from working. Heckman (1978, 1981a,b) has termed the presence of Y_{it-1} "true" state dependence and V_i "spurious" state dependence.

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

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

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

$$f_{Y_{it}|X_{it},Y_{it-1},U_{it}} = F_{\xi_{it}}\left[-\left(X_{it}'\beta + \gamma Y_{it-1} + U_{it}\right)\right]^{\mathbf{1}(Y_{it}=0)} \times f_{\xi_{it}}\left[Y_{it} - X_{it}'\beta - \gamma Y_{it-1} - U_{it}\right]^{\mathbf{1}(Y_{it}>0)} \quad (2)$$

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

1 In several studies, random effect means V_i is a random variable independent of the explanatory variables. The discussion here is based on definitions on p. 286 of Wooldridge (2010).

2 The estimation of an individual parameter V_i along with other model parameters leads to an incidental parameters problem. Our sieve MLE has a feature of random effect, treating V_i as a random variable and integrating out a composite unobserved variable to construct a likelihood function. Thus, the proposed sieve MLE has a computational advantage over a fixed effect approach because the individual parameter V_i does not appear in the likelihood function.

3 By the definition of η_{it}, U_{it} might not only contain the error terms in panels but also some unobserved covariates from the past. Hence, U_{it} denotes an unobserved covariate in this paper.

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.

dependent variables.[5] 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 semi-parametric sieve 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 estimation specifications, i.e., a static probit model, a maximum simulated likelihood (MSL) estimator, and the sieve MLE estimator. Our results find a large significant state dependence of labor force participation, smaller significant negative effects on nonlabor income variables, and also negative effects of children aged 0–2 in the current period and past period.

The paper is organized as follows. Section 2 provides a brief review of studies in the context of dynamic panel data models. We present the nonparametric identification of nonlinear dynamic panel data models in Section 3. Section 4 discusses our proposed sieve MLE. Section 5 provides the Monte Carlo study. Section 6 presents an empirical application describing the intertemporal labor participation of married women. Section 7 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. Related studies

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 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. Rasch (1960) and Chamberlain (1980, 1984) consider a conditional likelihood approach for logit models. 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. Their results rely on matching the explanatory variables in different time-periods. Honoré (1993), Hu (2002) and Honoré and Hu (2004) obtain moment conditions for estimating dynamic censored regression panel data models. Altonji and Matzkin (2005) develop two estimators for panel data models with nonseparable unobservable errors and endogenous explanatory variables.

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 condition problem.[6] 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 that misspecification of the distributions generally results in inconsistent parameter estimates. The associated computational burden of high order integration has been reduced significantly by recent advances in simulation techniques.[7] 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 condition 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 interest in panel dynamic discrete choice models. Evdokimov (2009) considers a nonparametric panel data model with nonadditive unobserved heterogeneity: $Y_{it} = m(X_{it}, V_i) + \varepsilon_{it}$ where individual-specific effects are allowed to be correlated with the covariates in an arbitrary manner. That model has a different focus from ours since our model includes lags of the endogenous dependent variable Y_{it-1} and a nonadditive ε_{it}.

While the proposed model (1) focuses on nonlinear dynamic panel data models, there are several studies on panel data models that are close in spirit to our work. Chernozhukov et al. (2009) derive bounds for marginal effects in nonlinear panel models and show that they can tighten rapidly as the number of time series observations grows. They also provide two novel inference methods that produce uniformly valid confidence regions in large samples. Hoderlein and White (2009) consider identification of marginal effects in general nonseparable models with unrestricted correlated unobserved effects and without lagged dependent variables, even if there are only two time periods. Arellano and Bonhomme (2009) provide a characterization of the class of weights for nonlinear panel data models that produce first-order unbiased estimators. Although the focus of the models in this paper is on the fixed time dimension, the results can be generalized to large T cases. The recent large-T literature for dynamic panel models can be found in Hahn and Kuersteiner (2004), Carro (2007) and Fernández-Val (2009).

5 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).

6 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 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 cannot be connected. Heckman (1981b) suggested the use of a flexible functional form to approximate the initial conditions.

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

In this paper, we provide nonparametric identification of nonlinear dynamic panel data models with unobserved covariates, show that the models are identified using only two periods of the dependent variable Y_{it} and three periods of the covariate X_{it} without initial conditions assumptions, and propose a sieve MLE estimator. The advantages of our results include that the point identification results are nonparametric and global, the model is quite general compared with the existing ones and makes use of the recently developed techniques, and the proposed sieve estimator is known to be convenient in computation. Meanwhile, our results have their disadvantages. The general nonparametric identification requires high-level technical assumptions. In particular, the injectivity assumption is not testable and its implication is still an active research area. The proposed sieve estimator also has its known shortcomings, such as the difficulty in choosing nuisance parameters.

3. Nonparametric identification

3.1. Main assumptions

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., $f_{Y_{it}|X_{it},Y_{it-1},U_{it}}$, is nonparametrically identified. As discussed above, some of our assumptions are high-level because we are providing nonparametric identification of the model. We have the following assumption.

Assumption 3.1 (*Exogenous Shocks*).

$f_{Y_{it}|X_{it},Y_{it-1},X_{it-1},U_{it}} = f_{Y_{it}|X_{it},Y_{it-1},U_{it}}$.

A sufficient condition for Assumption 3.1 is that the random shock ξ_{it} is independent of $\xi_{i\tau}$ for any $\tau \neq t$ and $\{X_{i\tau}, U_{i\tau}\}$ for any $\tau \leq t$. Given Eq. (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}. This assumption can be called an exogenous shocks condition. As shown in the two examples above, this sufficient assumption has been used in many existing studies.

Both ξ_{it} and U_{it} are scalar unobservables in the latent variable formulation of the dependent variable Y_{it} and account for the particular error structure in the formulation. While ξ_{it} is an exogenous random shock in period t, $U_{it} = V_i + \eta_{it}$ is the sum of the time-invariant heterogeneity and a function of all time-varying variables in the past.

The exogeneity of ξ_{it} can be relaxed to allow some dependence between ξ_{it} and (X_{it}, Y_{it-1}). For example, for some positive function h, write $\xi_{it} = h(X_{it}, Y_{it-1})^{1/2} e_{it}$ for an exogenous random shock e_{it} with unit variance. Hence, ξ_{it} contains heteroskedasticity and $\mathrm{Var}(\xi_{it}|X_{it}, Y_{it-1}) = h(X_{it}, Y_{it-1})$. In this case, the conditional distribution of the interest in Example 1 changes into

$$f_{Y_{it}|X_{it},Y_{it-1},U_{it}} = \left(1 - F_{\xi_{it}}\left[\frac{-\left(X_{it}'\beta + \gamma Y_{it-1} + U_{it}\right)}{h(X_{it}, Y_{it-1})^{1/2}}\right]\right)^{Y_{it}} \times F_{\xi_{it}}\left[\frac{-\left(X_{it}'\beta + \gamma Y_{it-1} + U_{it}\right)}{h(X_{it}, Y_{it-1})^{1/2}}\right]^{1-Y_{it}}.$$

Making ξ_{it} heteroskedastic generalizes the functional form of the dynamic panel data models considered in this paper. However, for simplicity we assume ξ_{it} is exogenous with a constant variance.

The existence of the exogenous random shock ξ_{it} in the error term of the latent variable formulation means that (X_{it}, Y_{it-1}) fully capture the dynamics conditional on U_{it} since further lags of Y_{it-1} or lags of X_{it} are not important once $(X_{it}, Y_{it-1}, U_{it})$ have been controlled for. To some extent, Assumption 3.1 has assumed dynamic completeness since

$$f_{Y_{it}|X_{it},Y_{it-1},U_{it}} = f_{Y_{it}|X_{it},Y_{it-1},U_{it},X_{it-1},Y_{it-2},U_{it-1},\ldots,X_{i1},Y_{i0},U_{i1}}, \quad t = 1, \ldots, T-1,$$

and once U_{it} is controlled for no past values of X_{it} or Y_{it-1} appear in the conditional density in the RHS of the above equation.

We simplify the evolution of the observed covariates X_{it} as follows.

Assumption 3.2 (*Covariate Evolution*)**.** The evolution of the observed covariates satisfies the equation $f_{X_{it+1}|Y_{it},X_{it},Y_{it-1},X_{it-1},U_{it}} = f_{X_{it+1}|X_{it},U_{it}}$.

Note that the assumption can be written as $X_{it+1} \perp (Y_{it}, Y_{it-1}, X_{it-1}) | (X_{it}, U_{it})$ and the lagged effects of Y_{it} such as $Y_{it-1}, Y_{it-2}, \ldots$ enter the evolution of X_{it+1} through the unobserved covariate U_{it}. A sufficient condition for Assumption 3.2 is that X_{it+1} is strictly exogenous and follows a first order Markov, conditional on U_{it}. Another sufficient condition for Assumption 3.2 is constituted of three steps, (i) (Markov evolution of X_{it+1}) $f_{X_{it+1}|Y_{it},X_{it},Y_{it-1},X_{it-1},U_{it}} = f_{X_{it+1}|Y_{it},X_{it},U_{it}}$, (ii) (No impact of ξ_{it} on X_{it+1}) $f_{X_{it+1}|Y_{it},Y_{it-1},X_{it},U_{it}} = f_{X_{it+1}|Y_{it-1},X_{it},U_{it}}$, and (iii) (Limited feedback) $f_{X_{it+1}|Y_{it-1},X_{it},U_{it}} = f_{X_{it+1}|X_{it},U_{it}}$.

The first step (i) 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}). The implication of the Markov assumption is that while the time-varying component of U_{it}, η_{it}, captures all the serially-correlated variation in the process of X_{it+1}, the corresponding time-invariant component V_i controls the time-invariant part of X_{it+1}. If X_{it+1} contains a time-invariant component other than V_i then the Markov assumption may fail. For example, suppose that we have[8]

$$X_{it+1} = \rho X_{it} + W_i + V_i + \upsilon_{it},$$

where υ_{it} are i.i.d., and a latent W_i is not perfectly correlated with V_i. In this case, given U_{it}, X_{it-1} will contain some information about W_i, even given X_{it}. Thus, X_{it-1} can be informative on X_{it+1} given (Y_{it}, X_{it}, U_{it}) and the Markov condition does not hold. However, the composite error U_{it} is a scalar unobservable in the latent variable formulation of the dependent variable Y_{it} and should also take account of the variation of X_{it}. If the time-varying component of U_{it} contains υ_{it} and its time-invariant component has $W_i + V_i$, the Markov assumption may hold. Our assumption rules out the situation that the evolution of X_{it} depends on other time-invariant element not in the latent variable formulation of Y_{it}.

The second step (ii) is that conditional on Y_{it-1}, X_{it} and U_{it}, X_{it+1} is independent of the exogenous shock ξ_{it}. Since U_{it} is a function of all past shocks $\{\xi_{i\tau}\}_{\tau<t}$, this step only excludes the immediate effect of the current shock ξ_{it} on the future covariate X_{it+1}.[9] This implies that $f_{X_{it+1}|X_{it},Y_{it-1},U_{it},\xi_{it}} = f_{X_{it+1}|X_{it},Y_{it-1},U_{it}}$. The third step (iii) is a limited feedback assumption, i.e., $f_{X_{it+1}|X_{it},Y_{it-1},U_{it},} = f_{X_{it+1}|X_{it},U_{it}}$ which rules out direct feedback from the lagged dependent variable Y_{it-1} on the future value of the observed covariate X_{it+1}. The effect of Y_{it-1} on X_{it+1} is indirectly through X_{it}, and U_{it}.

---

8 We thank an anonymous referee for suggesting this example.

9 The assumption imposes some restriction to regressors in panel data setting. For example, suppose that $U_{it} = V_i$. The assumption that X_{it+1} is independent of ξ_{it} given X_{it} and V_i implies that $E[X_{it+1}\xi_{it}] \neq 0$. If the future covariate X_{it+1} is predetermined, in the sense that $E[X_{it+1}\xi_{is}] \neq 0$ for $s < t+1$ and zero otherwise, then the assumption fails when the X_{it+1} is predetermined. However, the assumption permits a weaker version of a predetermined variable such as $E[X_{it+1}\xi_{is}] \neq 0$ for $s < t$ and zero otherwise.

Overall, Assumption 3.2 implies that conditional on X_{it} and U_{it}, X_{it+1} is independent of the exogenous shock ξ_{it}. In other words, conditional on the past information, the future covariate X_{it+1} rules out the immediate effect of the current shock ξ_{it} of the dependent variable Y_{it}.

Let $\mathcal{L}^p(\mathcal{X})$, $1 \leq p < \infty$ stand for the space of function $h(\cdot)$ with $\int_{\mathcal{X}} |h(x)|^p dx < \infty$. Suppose $\mathcal{X}_t$, and $\mathcal{U}_t$ be the supports of the random variables X_{it} and U_{it}, respectively. For any $1 \leq p \leq \infty$ and we define operators as follows: for any given (x_{it}, y_{it-1}),

$$L_{X_{it+1},x_{it},y_{it-1},X_{it-1}} : \mathcal{L}^p(\mathcal{X}_{t-1}) \to \mathcal{L}^p(\mathcal{X}_{t+1})$$

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

and for any given x_{it},

$$L_{X_{it+1}|x_{it},U_{it}} : \mathcal{L}^p(\mathcal{U}_t) \to \mathcal{L}^p(\mathcal{X}_{t+1})$$

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

Assumption 3.3 (*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.

This is a high-level assumption, which is hard to avoid for nonparametric identification. Intuitively, this assumption guarantees that the observables contain enough information on the unobserved covariate U_{it} and the covariates in period $t+1$, X_{it+1}, depend on X_{it}. However, the invertibility of $L_{X_{it+1},x_{it},y_{it-1},X_{it-1}}$, which is equivalent to a completeness condition on an observed distribution $f_{X_{it+1},X_{it},Y_{it-1},X_{it-1}}$, is not testable in a nonparametric setting with continuous variables as shown in Canay et al. (2011).

If an operator is constructed by a density of independent variables, the operator certainly fails to be invertible. Since $f_{X_{it+1},X_{it},Y_{it-1},X_{it-1}}$ is the density of correlated variables, it provides at least some justification for the completeness property.[10] Thus, the invertibility may require functional form restrictions on $f_{X_{it+1},X_{it},Y_{it-1},X_{it-1}}$. For example, if $\mathcal{X}_{t+1}$ contains an open set then $f_{X_{it+1},X_{it},Y_{it-1},X_{it-1}} = \phi(X_{it-1} - \alpha_1 X_{it+1} - \alpha_2 X_{it} - \alpha_3 Y_{it-1})$ satisfies Assumption 3.3 where ϕ is the standard normal pdf and $\alpha_i \neq 0$.[11] Besides a linear process, another example may be that $f_{X_{it+1},X_{it},Y_{it-1},X_{it-1}}$ belongs to an exponential family. Given a fixed (x_{it}, y_{it-1}). Suppose that

$$f_{X_{it+1},x_{it},y_{it-1},X_{it-1}} = s(x_{it}, y_{it-1}, X_{it-1})t(X_{it+1}, x_{it}, y_{it-1}) \times \exp[\mu(X_{it+1}, x_{it}, y_{it-1})\tau(x_{it}, y_{it-1}, X_{it-1})]$$

where $s(x_{it}, y_{it-1}, X_{it-1}) > 0$, $\tau(x_{it}, y_{it-1}, X_{it-1})$ is one-to-one in X_{it-1}, and support of $\mu(X_{it+1}, x_{it}, y_{it-1}) \in \mathcal{X}_{t+1}$ is an open set. Theorem 2.2 in Newey and Powell (2003) shows the family of the joint density functions $\{f_{X_{it+1},x_{it},y_{it-1},X_{it-1}} : X_{it+1} \in \mathcal{X}_{t+1}\}$ is complete over $\mathcal{L}^p(\mathcal{X}_{t-1})$ for each (x_{it}, y_{it-1}).[12] This also implies the invertibility of $L_{X_{it+1},x_{it},y_{it-1},X_{it-1}}$ in Assumption 3.2.

On the other hand, the invertibility of $L_{X_{it+1}|x_{it},U_{it}}$ requires the covariates in period $t+1$, X_{it+1}, contains enough information on the unobserved covariate U_{it} conditional on X_{it}. Hahn (2001) considers a dynamic logit model with individual effects where the regressors include the lag dependent variable, time dummies and possibly other strictly exogenous variables. He shows that the semi-parametric information bound for any estimator of the state dependence coefficient is zero. Our results do not cover the dynamic logit model in Hahn (2001) because the invertibility of $L_{X_{it+1}|x_{it},U_{it}}$ in Assumption 3.3 requires some dependence between U_{it} and X_{it+1}. If X_{it+1} only contains time dummies and possibly other strictly exogenous variables, the condition will fail to hold. This is intuitive: the existence of a degree of dependence between U_{it} and X_{it+1} allows us to control the unobservable U_{it} from the observable X_{it+1}. It reflects the methodology of our identification method that provides an alternative way to deal with an unobservable term inside a nonlinear econometric model, tackling down an unobserved effect with an observable correlated covariate instead of eliminating the unobserved effect by transformations. 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} cannot be invertible. Additionally, 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}.[13] Necessary conditions for Assumption 3.3 include that $f_{X_{it+1},Y_{it-1},X_{it}|X_{it-1}} \neq f_{X_{it+1},Y_{it-1},X_{it}}$ and $f_{X_{it+1}|X_{it},U_{it}} \neq f_{X_{it+1}|X_{it}}$. These necessary conditions rule out the case where X_{it+1} and X_{it-1} are independent or X_{it+1} and U_{it} are independent. In other words, Assumption 3.3 permits the existence of serial correlation among X_{it} and correlation between X_{it+1} and U_{it}.

The invertibility of the integral operator $L_{X_{it+1}|x_{it},U_{it}}$ is equivalent to saying that the family $\{f_{X_{it+1}|X_{it},U_{it}}(x_{it+1}|x_{it}, u_{it}) : x_{it+1} \in \mathcal{X}_{t+1}\}$ is complete over $\mathcal{L}^p(\mathcal{U}_t)$. Hu and Shiu (2011) showed that if the conditional density $f(x|z)$ can form a linearly independent sequence and coincides with a known complete density at a limit point in the support of z, then $f(x|z)$ itself is complete. They also provide examples of complete families other than trivial linear/exponential family cases. For example, suppose ϕ is the standard normal pdf, consider

$$f(x|z) = \lambda(z)h(x|z) + [1 - \lambda(z)]\phi(x - z), \tag{3}$$

which is a mixture of two continuous conditional densities, h and ϕ, and the weight λ in the mixture depends on z.[14] Sufficient conditions for the completeness of $f(x|z)$ are (i) $\lim_{z_k \to z_0} \lambda(z) = 0$; and (ii) $\lim_{x \to -\infty} \frac{h(x|z_k)}{\phi(x-z_k)} < \infty$. Following this result, construct

$$f_{X_{it+1}|X_{it},U_{it}}(x_{it+1}|x_{it}, u_{it}) = \lambda(x_{it+1})h(x_{it+1}, x_{it}, u_{it}) + [1 - \lambda(x_{it+1})]\phi(x_{it+1} - \psi(x_{it}) - u_{it}),$$

with $\lim_{x_{it+1,k} \to x_{it+1,0}} \lambda(x_{it+1}) = 0$ and (ii) $\lim_{u_{it} \to -\infty} \frac{h(x_{it+1,k}, x_{it}, u_{it})}{\phi(x_{it+1,k} - \psi(x_{it}) - u_{it})} < \infty$. The completeness of $\{f_{X_{it+1}|X_{it},U_{it}}(x_{it+1}|x_{it}, u_{it}) : x_{it+1} \in \mathcal{X}_{t+1}\}$ implies that the operator $L_{X_{it+1}|x_{it},U_{it}}$ is invertible. In this case, there is only the tail condition on the function $h(x_{it+1,k}, x_{it}, u_{it})$ and we can regard h as

10 That the variables X_{it+1}, X_{it}, Y_{it-1}, and X_{it-1} are highly correlated can be justified by the fact that most variables in economics are correlated across time which reveal a pattern of serial correlation or autocorrelation.

11 The result is from Theorem 2.3 in Newey and Powell (2003). Suppose that the distribution of x conditional on z is $N(a + bz, \sigma^2)$ for $\sigma^2 > 0$ and the support of z contains an open set, then the integral operator corresponding to $\frac{1}{\sigma}\phi(\frac{x-a-bz}{\sigma})$ is invertible from $\mathcal{L}^p(\mathcal{X})$ to $\mathcal{L}^p(\mathcal{Z})$ where ϕ is the standard normal PDF. There are more detailed discussions and general conditions for an invertible integral operator or complete conditional distributions in $\mathcal{L}^p(\mathcal{X})$ in Hu and Shiu (2011).

12 The whole statement of the theorem is the following: *Let $f(x|z) = s(x)t(z)\exp[\mu(z)\tau(x)]$, where $s(x) > 0$, $\tau(x)$ is one-to-one in x, and support of $\mu(z)$, $\mathcal{Z}$, is an open set, then $E[h(\cdot)|z] = 0$ for any $z \in \mathcal{Z}$ implies $h(x) = 0$ almost everywhere in $\mathcal{X}$; equivalently, the family of conditional density functions $\{f(x|z) : z \in \mathcal{Z}\}$ is complete in $L^p(\mathcal{X})$.*

13 Assumption 3.3 requires $L_{X_{it+1}|x_{it},U_{it}}$ is invertible and it demands the unobservable U_{it} to be correlated with the observed X_{it+1}. This case is complementary to the existing models where U_{it} is independent of X_{it+1}. Honoré and Kyriazidou (2000) and Honoré and Tamer (2006) identify the parameters under certain assumptions on the strictly exogenous covariates.

14 The choice of ϕ is for simplicity. Please see Hu and Shiu (2011) for general results.

nonparametric deviation or oscillation from the normal ϕ. Therefore, the invertibility of the integral operator $L_{X_{it+1}|x_{it},U_{it}}$ is appropriate in a nonparametric setting. The condition contains a restriction on the unobservable and it cannot be verified. A way to justify the condition is invoking the central limit theorem to conclude that $f_{X_{it+1}|X_{it},U_{it}}(x_{it+1}|x_{it},u_{it})$ has an approximate normal distribution and the invertibility permits nontrivial variation around a normal distribution.

In addition, the invertibility of the operator $L_{X_{it+1},x_{it},y_{it-1},X_{it-1}} = L_{X_{it+1}|x_{it},U_{it}}L_{x_{it},y_{it-1},X_{it-1},U_{it}}$ does imply restrictions on the initial condition through the operator $L_{x_{it},y_{it-1},X_{it-1},U_{it}}$. For example, in a case where X_{it} and U_{it} are discrete and the linear operators are matrices, the invertibility of these operators is equivalent to the invertibility of corresponding matrices. However, the operators or matrices may still have a flexible form so that there is no need to specify the initial condition.

Note that when the unobserved component U_{it} is continuous, the invertibility of $L_{X_{it+1}|x_{it},U_{it}}$ implies that the explanatory variables X_{it} contain 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} contain enough information to identify unobserved component U_{it}. A sufficient condition for identification with continuous U_{it} can be obtained from the well-known completeness property of exponential families.[15] Thus, if $\mathcal{U}_{it}$ is an open set, then $\mathcal{X}_{it+1}$ must be an open set.[16] In the case of the intertemporal labor force participation behavior of married women, the covariates X_{it} contain wage and U_{it} includes the unobserved individual skill level or motivation.

Assumption 3.4 (*Distinctive eigenvalues*)**.** There exists 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, $E[I(Y_{it} = 0)|x_{it}, y_{it-1}, u_{it}] = F_{\xi_{it}}\left[-\left(x'_{it}\beta + \gamma y_{it-1} + u_{it}\right)\right]$, which is monotonic in u_{it}. Assumption 3.4 implies that for all $\hat{U}_{it}, \tilde{U}_{it} \in \mathcal{U}$, the set $\{y : f_{Y_{it}|X_{it},Y_{it-1},\hat{U}_{it}} \neq f_{Y_{it}|X_{it},Y_{it-1},\tilde{U}_{it}}\}$ for any given (x_{it}, y_{it-1}) has a positive probability whenever $\hat{U}_{it} \neq \tilde{U}_{it}$.

Assumption 3.5 (*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 $f_{(X_{it+1}-X_{it})|X_{it},U_{it}}(\cdot|x_{it}, u_{it})$. The purpose of Assumption 3.5 is to normalize $f_{X_{it+1}|X_{it},U_{it}}$ to be unique in the spectral decomposition and it requires the functional G to map the eigenfunction to a real number. The condition can also be written as $G\left[f_{X_{it+1}|X_{it},U_{it}}(\cdot|x_{it}, u_{it})\right] = l(u_{it})$ for some one–one function $l(\cdot)$ and thus it is not very restrictive.

This assumption imposes a restriction on the covariate evolution. A choice of G depends on how the covariate X_{it} changes over time given the unobserved covariate U_{it}. Hence, observations on the conditional temporal correlation of X_{it} may shed a light on the pick of G. In the case of the intertemporal labor force participation behavior of married women, X_{it} may include annual family income, which often varies with the unobserved time-invariant family characteristics and past economy shock. In this case, setting G as the mode functional seems appropriate.

3.2. Main identification results

We start our identification with a panel data containing two periods of the dependent variable Y_{it} and three periods of the covariate X_{it}, $\{X_{it+1}, Y_{it}, X_{it}, Y_{it-1}, X_{it-1}\}$ for $i = 1, 2, \ldots, n$. The law of total probability leads to

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

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

Assumption 3.1 implies

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

Then, Assumption 3.2 suggests that

$$f_{X_{it+1},Y_{it},X_{it},Y_{it-1},X_{it-1}} = \int f_{X_{it+1}|X_{it},U_{it}} f_{Y_{it}|X_{it},Y_{it-1},U_{it}} \times f_{X_{it},Y_{it-1},X_{it-1},U_{it}} dU_{it}. \quad (4)$$

Based on this equation, we may apply the identification results in Hu and Schennach (2008a,b) to show that all the unknown densities on the RHS are identified from the observed density on the LHS. For any given $(y_{it}, x_{it}, y_{it-1})$, we define operators as follows.

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

and

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

Similarly, define

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

Eq. (4) 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. (4) 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 spectral decomposition results in Hu and Schennach (2008a,b) 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 $f_{X_{it+1},Y_{it},X_{it},Y_{it-1},X_{it-1}}$. Assumptions 3.1–3.3 enable us to have

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

15 See Newey and Powell (2003) for details.

16 Assumption 3.3 impose the invertibility of the linear operator $L_{X_{it+1}|x_{it},U_{it}}$ which maps from the domain space $\mathcal{L}^p(\mathcal{U}_t)$ to the range space $\mathcal{L}^p(\mathcal{X}_{t+1})$. The invertibility implies a cardinality relation, the cardinality of $\mathcal{U}_t$ is smaller than the cardinality of $\mathcal{X}_{t+1}$. If U_{it} takes continuous values, then X_{it+1} must continuous values.

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 $L_{X_{it+1}|x_{it},U_{it}}$. Assumption 3.4 make the eigenvalues distinctive. Since the identification from the spectral decomposition is only identified up to u_{it} and its monotone transformation, we make a normalization assumption, Assumption 3.5, to pin down the unobserved covariate u_{it}.

Notice that Theorem 1 in Hu and Schennach (2008a,b) 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 by the density $f_{Y_{it}|X_{it},Y_{it-1},U_{it}}$. While 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 by $f_{X_{it+1}|X_{it},U_{it}}$.

We summarize our identification results as follows:

Theorem 3.1. *Under Assumptions 3.1–3.5, the observable 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 $f_{X_{it},Y_{it-1},X_{it-1},U_{it}}$.*[17]

Since the unobserved covariate U_{it} appearing in $f_{Y_{it}|X_{it},Y_{it-1},U_{it}}$ does not have natural units of measurement or it is unclear which values are appropriate for U_{it}, the partial effects averaged across the distribution of U_{it} are more appealing. The average partial effects are based on the effect on a mean response after averaging the unobserved heterogeneity across the population. Theorem 3.1 allows us to obtain the marginal distribution of U_{it},

$$f_{U_{it}}=\int_{\mathcal{X}_{it}}\int_{\mathcal{Y}_{it-1}}\int_{\mathcal{X}_{it-1}} f_{X_{it},Y_{it-1},X_{it-1},U_{it}}\,dX_{it}\,dY_{it}\,dX_{it-1}.$$

Suppose that we are interested in the conditional mean of $\omega(y_t)$, which is a scalar function of y_t. Given (X_{it}, Y_{it-1}) the average structural function (ASF) is defined by

$$\mathrm{ASF}(X_{it},Y_{it-1})=\int_{\mathcal{U}_{it}}\left[\int_{\mathcal{Y}_{it}}\omega(y_t)\,f_{Y_{it}|X_{it},Y_{it-1},U_{it}}\,dY_{it}\right]f_{U_{it}}\,dU_{it}, \qquad (5)$$

whose identification can be shown by Theorem 3.1. Then the average partial effect (APE) can be defined by taking derivatives or differences of the above expression (5) with respect to elements of (X_{it}, Y_{it-1}). These discussions lead to the following result.

Corollary 3.1. *Under Assumptions 3.1–3.5, average structural function (ASF) defined in Eq. (5) and the average partial effect (APE) can be identified and estimated by a panel data containing two periods of the dependent variable Y_{it} and three periods of the covariate X_{it}, $\{X_{it+1}, Y_{it}, X_{it}, Y_{it-1}, X_{it-1}\}$ for $i = 1, 2, \dots, n$.*

3.3. Discussion of assumptions

We discussed the identification assumptions separately in Section 3.1 and now we illustrate these assumptions jointly for the models in Examples 1 and 2. These models can be used to describe the following economic behaviors. While Y_{it} denotes the t-th period labor force participation decision and the amount of insurance coverage chosen by an individual for Examples 1 and 2 respectively, the covariate X_{it} is the nonlabor income in both models. Assumption 3.1 allows us to separate the exogenous random shock of the dependent variable in period t, ξ_{it}, from all time-varying error term in the past. It follows that ξ_{it} and U_{it} can be used to decompose the particular error structure in the latent variable formulation of the dependent variable Y_{it}. While ξ_{it} is an exogenous random shock in period t, $U_{it} = V_i + \eta_{it}$ is the sum of the time-invariant heterogeneity and a function of all time-varying variables in the past. This implies that both time-invariant and the past time-varying information are in U_{it}, and the observed (X_{it}, Y_{it-1}) has completely captured the contemporaneous information of Y_{it} other than ξ_{it}. Hence, the present time-varying shocks of labor force participation decision or the amount of insurance coverage are independent of the lagged dependent variables, the nonlabor income, and U_{it}.

The definition of U_{it} indicates that conditional on U_{it}, the variation of all past shocks before period t $\{\xi_{i\tau}\}_{\tau<t}$ become trivial.[18] Thus, Assumption 3.2 only rules out the immediate effect of the current shock ξ_{it} on the future covariate X_{it+1}. In the economic contexts, the assumption reflects the current exogenous shocks of labor force participation decision or the amount of insurance coverage do not affect the nonlabor income in the next period.

The linear independence interpretation for the invertibility of an operator in Hu and Shiu (2011) suggests that the invertibility of $L_{X_{it+1},x_{it},y_{it-1},X_{it-1}}$ in Assumption 3.3 can be stated as (1) the family of the joint distributions $\{f_{X_{it+1},X_{it},Y_{it-1},X_{it-1}}(u,x_{it},y_{it-1},x) : u \in \widetilde{\mathcal{X}}_{t+1}\}$ where $\widetilde{\mathcal{X}}_{t+1} \subset \mathcal{X}_{t+1}$ has nontrivial variation over the index u in $\widetilde{\mathcal{X}}_{t+1}$ in the function space $\mathcal{L}^p(\mathcal{X}_{t-1})$, and (2) the variation is big enough that every function in $\mathcal{L}^p(\mathcal{X}_{t-1})$ can be approximated by the distributions in the family. The assumption requires some dependence of observed covariates over time. If X_{it} is constant across time, then it violates the condition. In this case, the serially correlated nature of $\mathcal{X}_t$ over time can provide some support of statement (1) but statement (2) is the key assumption to make the invertibility hold.

Next, we discuss the invertibility in Example 1 or the empirical application using the linear independence interpretation. Recall that the dynamic discrete-choice model with an unobserved covariate U_{it}: $Y_{it} = 1\left(X_{it}'\beta+\gamma Y_{it-1}+U_{it}+\xi_{it}\geq 0\right)$ where Y_{it} denotes the t-th period participation decision, and X_{it} is the wage or income variable in that period. First, the invertibility of $L_{X_{it+1},x_{it},y_{it-1},X_{it-1}}$ implies that the conditional distribution of wage or income variables $f_{X_{it+1},X_{it},Y_{it-1},X_{it-1}}(u,x_{it},y_{it-1},x)$ over some subset of $\mathcal{X}_{t+1}$ can approximate distributions of wage or income in period $t-1$ well and hence, any income or wage distribution in period $t-1$ has been accounted for by this functional form using the variation in period $t+1$. The independence of income or wage variables over time clearly cause the invertibility to fail. Second, if the unobserved covariate U_{it} contains time-invariant heterogeneity such as motivation or inherent health, the invertibility of $L_{X_{it+1}|x_{it},U_{it}}$ suggests that given the current income variable x_{it} the variation of the functional form $f_{X_{it+1}|X_{it},U_{it}}$ over the future income variables can fully capture the changes or movement of unobserved motivation or inherent health.

These two models have a point mass at $y = 0$, so we can choose $\omega(y) = I(y = 0)$. Assumption 3.4 is automatically satisfied for these limited dependent variable models. Finally, the covariate evolution represents the changing of the nonlabor income over time in these models. As mentioned in Assumption 3.5, the functional G can be the mean, mode, median, or a quantile. Thus, one of the conditions for Assumption 3.5 is that the mode of the distribution of the nonlabor income in the next period conditional on the current nonlabor income and unobserved covariate u_{it} is equal to the unobserved covariate. Since the unobserved covariate U_{it} contains time-invariant heterogeneity such as motivation or

17 The identification techniques are illustrated in Appendix A using a finite dimensional discrete example where the linear operators are matrices.

18 Recall $U_{it} = V_i + \eta_{it}$ and $\eta_{it} = \varphi\left(\{X_{i\tau}, Y_{i\tau-1}, \xi_{i\tau}\}_{\tau=0,1,\dots,t-1}\right)$, η_{it} is a function of all time-varying variables in the past.

inherent health, it means that the value of nonlabor income that occurs most frequently around the location of true level of unobserved motivation or inherent health.

Set $\varepsilon_{it} = \rho\varepsilon_{it-1} + \xi_{it}$ and $\xi_{it} \sim N(0, \sigma_\xi^2)$. Consider the following data generating process (DGP):

$$Y_{it} = g(\beta_0 + \beta_1 X_{it} + \gamma Y_{it-1} + U_{it} + \xi_{it} \geq 0)$$
$$\text{with } U_{it} = V_i + \rho\varepsilon_{it-1}\ \forall i = 1, \ldots, N;\ t = 1, \ldots, T-1, \tag{6}$$

where $g(\cdot)$ can be the 0–1 indicator function or $g(\cdot) = \max(0, \cdot)$ and $V_i \sim N(\mu_v, \sigma_v^2)$. The generating process of covariate evolution has the following form $X_{it+1} = X_{it} + h(X_{it})\epsilon_{it} + U_{it}$ or

$$f_{X_{it+1}|X_{it},U_{it}}(x_{t+1}|x_t, u) = \frac{1}{h(x_t)} f_\epsilon\left(\frac{x_{t+1} - x_t - u}{h(x_t)}\right), \tag{7}$$

where f_ϵ is a density function that can be specified under different identification conditions of Assumption 3.5.[19] For example, take $f_\epsilon(x) = \exp(x - e^x)$ and the mode as the choice of G for Assumption 3.5. We will use these settings in the Monte Carlo simulation.

It is straightforward to verify the assumptions with the specific data generating processes except for Assumption 3.3. The invertibility of $L_{X_{it+1}|x_{it},U_{it}}$ is equivalent to the completeness of the family $\{f_{X_{it+1}|X_{it},U_{it}}(x_{t+1}|x_t, u) : x_{t+1} \in \mathcal{X}_{it+1}\}$. When $f_\epsilon(x) = \exp(x - e^x)$, the covariate evolution belongs to one of exponential families and it is complete by Theorem 2.2 in Newey and Powell (2003). Therefore, $L_{X_{it+1}|x_{it},U_{it}}$ is invertible. Applying the invertibility of $L_{X_{it+1}|x_{it},U_{it}}$ to the integral relation $L_{X_{it+1},x_{it},y_{it-1},X_{it-1}} = L_{X_{it+1}|x_{it},U_{it}} L_{x_{it},y_{it-1},X_{it-1},U_{it}}$ implies that the invertibility of $L_{X_{it+1},x_{it},y_{it-1},X_{it-1}}$ is equivalent to the invertibility of $L_{x_{it},y_{it-1},X_{it-1},U_{it}}$. Utilize Theorem 2.2 in Newey and Powell (2003) again to the family $\{f_{U_{it-1}|X_{it},X_{it-1}} = \frac{1}{h(x_{t-1})} f_\epsilon\left(\frac{-u+x_t-x_{t-1}}{h(x_{t-1})}\right) : u \in \mathcal{U}_{it-1}\}$ for each given x_t and then use it to obtain the completeness of the family $\{f_{X_{it},X_{it-1},U_{it-1}}(x_t, x_{t-1}, u) : u \in \mathcal{U}_{it-1}\}$.[20] Next, pass the completeness of $\{f_{X_{it},X_{it-1},U_{it-1}}(x_t, x_{t-1}, u) : u \in \mathcal{U}_{it-1}\}$ to $\{f_{X_{it},X_{it-1},U_{it}}(x_t, x_{t-1}, u) : u \in \mathcal{U}_{it}\}$ using an integral equation

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

Since $U_{it} = U_{it-1} +$ a normal error, $f_{U_{it}|U_{it-1}}$ is a complete distribution by the normality. We can express the integral equation as an operator relationship and show the operator using $f_{X_{it},X_{it-1},U_{it}}$ as a kernel is invertible. This implies $\{f_{X_{it},X_{it-1},U_{it}}(x_t, x_{t-1}, u) : u \in \mathcal{U}_{it}\}$ is complete and then the family $\{f_{X_{it},Y_{it-1},X_{it-1},U_{it}}(x_t, y_{t-1}, x_{t-1}, u) : u \in \mathcal{U}_{it}\}$ is also complete over $\mathcal{L}^p(\mathcal{X}_{it-1})$. We have reached $L_{X_{it+1},x_{it},y_{it-1},X_{it-1}}$ is invertible.

4. Estimation

The dynamic panel data model (1) specifies the relationship between the dependent variable of interest for an individual i, Y_{it}, and the explanatory variables including a lagged dependent variable Y_{it-1}, a set of possibly time-varying explanatory variables X_{it}, and 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 $\{\xi_{it}\}_{t=0}^T$ are 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.[21] What is not specified in the 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 semi-parametric estimator.

Our semi-parametric 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 3, we have shown Eq. (4) 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. (4) implies

$$\begin{aligned}\alpha_0 &\equiv (f_{X_{it+1}|X_{it},U_{it}}, \theta_0, f_{X_{it},Y_{it-1},X_{it-1},U_{it}})^T \\ &= \arg\max_{(f_1,\theta,f_2)^T \in \mathcal{A}} E \ln \int f_1(x_{it+1}|x_{it}, u_{it}) f_{Y_{it}|X_{it},Y_{it-1},U_{it}} \\ &\quad \times (y_{it}|x_{it}, y_{it-1}, u_{it}; \theta) f_2(x_{it}, y_{it-1}, x_{it-1}, u_{it}) du_{it},\end{aligned}$$

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

$$\begin{aligned}\widehat{\alpha}_n &\equiv \left(\hat{f}_1, \hat{\theta}, \hat{f}_2\right)^T \\ &= \arg\max_{(f_1,\theta,f_2)^T \in \mathcal{A}_n} \frac{1}{n}\sum_{i=1}^n \ln \int f_1(x_{it+1}|x_{it}, u_{it}) f_{Y_{it}|X_{it},Y_{it-1},U_{it}} \\ &\quad \times (y_{it}|x_{it}, y_{it-1}, u_{it}; \theta) f_2(x_{it}, y_{it-1}, x_{it-1}, u_{it}) du_{it}.\end{aligned} \tag{8}$$

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 A, we provide sufficient conditions for the consistency of our semi-parametric 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 Online Appendix.

[19] 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.

[20] Suppose that $h \in \mathcal{L}^p(\mathcal{X}_{it-1})$ and $\int h(x_{t-1}) f_{X_{it},X_{it-1},U_{it-1}}(x_t, x_{t-1}, u) dx_{t-1} = 0$ for any x_t. The equation can be rewritten as $\int h(x_{t-1}) f_{X_{it},X_{it-1}} f_{U_{it-1}|X_{it},X_{it-1}} dx_{t-1} = 0$ for any u_{it-1}. The completeness of $\{f_{U_{it-1}|X_{it},X_{it-1}} : u \in \mathcal{U}_{it-1}\}$ implies that $h(x_{t-1}) f_{X_{it},X_{it-1}} = 0$ and then $h = 0$. We obtain the completeness of the family $\{f_{X_{it},X_{it-1},U_{it-1}}(x_t, x_{t-1}, u) : u \in \mathcal{U}_{it-1}\}$ over $\mathcal{L}^p(\mathcal{X}_{it-1})$.

[21] A partition of θ into finite-dimensional parameters and infinite-dimensional parameters does not affect our sieve MLE. More examples of a partition can be found in Shen (1997).

With the consistency of the semi-parametric estimator $\widehat{\alpha}_n$, a consistent estimator of the average structural function (ASF) can be obtained by

$$\mathrm{ASF}(X_t, Y_{t-1}) = \int_{\mathcal{U}_t} \left[\int_{\mathcal{Y}_t} \omega(y_t) f_{Y_t|X_t,Y_{t-1},U_t} \times (y_t|x_t, y_{t-1}, u_t; \hat{\theta}) dY_t \right] \hat{f}_2(u_t) du_t, \tag{9}$$

where $\hat{f}_2(U_t) = \int_{\mathcal{X}_t} \int_{\mathcal{Y}_{t-1}} \int_{\mathcal{X}_{t-1}} \hat{f}_2(X_t, Y_{t-1}, X_{t-1}, U_t) dX_t dY_{t-1} dX_{t-1}$. Thus, the average partial effects of the state dependence at interesting values of the explanatory variables can be computed by changes or derivatives of Eq. (9) with respect to Y_{t-1}.

Note that the proposed sieve MLE only needs 3 periods. This means that when a DGP is generated through the dynamic process (1), three-periods data are enough to recovery the parameter of the interest θ. When there are more periods of data, the approach is still tractable. For example, if $T = 4$ and we assume the dynamic panel data specification (1), then estimation results from periods 1, 2, and 3 should be the same as ones from 2, 3, and 4. If the estimated results are significantly different, we would suspect model misspecification. Under the assumptions of stationary and ergodicity, an alternative way to deal with data more than 3 periods is to transform the data into 3 periods of data by rearranging them as 3 periods of data and stacking them into a larger cross-sectional data. For example, suppose that there are 5 periods of data $\{D_t, D_{t+1}, D_{t+2}, D_{t+3}, D_{t+4}\}$. It can be transformed into three observations of three periods of data, i.e., $\{D_t, D_{t+1}, D_{t+2}\}$, $\{D_{t+1}, D_{t+2}, D_{t+3}\}$, and $\{D_{t+2}, D_{t+3}, D_{t+4}\}$.

For a model with a larger number of observed covariates, we can consider a single-index response model with $X_{it}'\beta$. That is X_{it} is a d-dimensional vector of explanatory variables, and $X_{it}'\beta$ is the index, the scalar product of X_{it} with β, a vector of parameters whose values are unknown. Since our assumptions do not exclude time dependence in covariates, time dummies are allowed to be in X_{it}. Many widely used parametric models have this form. In our empirical application, we adopt this approach to deal with a case of many observed covariates. The part (ii) of Assumption B.4. Requires that $k_{ni}/n \to 0$ for $i = 1, \lambda, 2$. Thus, the rate of convergence depends on the degree of the sieve approximations since higher degree of sieve spaces provide better approximations. When X_{it} is a d-dimensional vector and the index form is not used, the degree of approximation has to be increased proportionally in order to get better approximation of these nuisance component, $f_1(x_{t+1}|x_t, u_t; \delta_1)$ and $f_2(x_t, y_{t-1}, x_{t-1}, u_t; \delta_2)$. It follows that the larger the dimension of X_{it}, the slower the rate of convergence. Thus, the curse of dimensionality may be an issue if researchers are interested in the nuisance component $f_1(x_{t+1}|x_t, u_t; \delta_1)$ and $f_2(x_t, y_{t-1}, x_{t-1}, u_t; \delta_2)$, but the convergence speed of the parametric part is still root-n.

4.1. Implementation

As we discussed above, we propose a semi-parametric sieve MLE using an i.i.d. sample $\{x_{it+1}, y_{it}, x_{it}, y_{it-1}, x_{it-1}\}$ for $i = 1, 2, \ldots, n$. The unknown densities are associated with the observed distribution as follows:

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

The parametric part is the model of interest $f_{Y_{it}|X_{it},Y_{it-1},U_{it}}(y_{it}|x_{it}, y_{it-1}, u_{it}; \theta)$. The two nonparametric nuisance functions include $f_{X_{it+1}|X_{it},U_{it}}$ and $f_{X_{it},Y_{it-1},X_{it-1},U_{it}}$. The sieve MLE transforms a semi-parametric MLE to a parametric MLE by replacing the nonparametric nuisance functions with their Fourier approximations. For example, the sieve estimator for the covariate evolution may be constructed by the Fourier series as follows:

$$f_1(x_{t+1}|x_t, u_t; \delta_1) = \sum_{i=0}^{i_n} \sum_{j=0}^{j_n} \sum_{k=0}^{k_n} \delta_{1,ijk} \varphi_{1i}(x_{t+1} - u_t) \times \varphi_{2j}(x_t) \varphi_{3k}(u_t),$$

where i_n, j_n, k_n are smoothing parameters and $\varphi_{1i}, \varphi_{2j}, \varphi_{3k}$ are known basis functions. Similarly, we may have a sieve approximation of the initial joint density, $f_2(x_{it}, y_{it-1}, x_{it-1}, u_{it}; \delta_2)$, where δ_2 is a vector of all the sieve coefficients. The fact that the parametric functions $f_1(x_{t+1}|x_t, u_t; \delta_1)$ and $f_2(x_{it}, y_{it-1}, x_{it-1}, u_{it}; \delta_2)$ are approximations of probability density functions implies certain restrictions on the sieve coefficients (δ_1, δ_2), which is discussed in Online Appendix. In the sieve MLE, we may estimate $(\theta, \delta_1, \delta_2)$ as a parametric MLE with a density function as follows:

$$f(x_{it+1}, y_{it}, x_{it}, y_{it-1}, x_{it-1}; \theta, \delta_1, \delta_2) = \int f_1(x_{it+1}|x_{it}, u_t; \delta_1) f_{Y_{it}|X_{it},Y_{it-1},U_{it}} \times (y_{it}|x_{it}, y_{it-1}, u_{it}; \theta) f_2(x_{it}, y_{it-1}, x_{it-1}, u_{it}; \delta_2) du_{it}.$$

In Online Appendix, we show the consistency and asymptotic normality as sample size goes to infinity.

5. 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, $\varepsilon_{it} = \rho\varepsilon_{it-1} + \xi_{it}$, $\xi_{it} \sim N(0, 1/2)$. Denote $\Phi_{\xi_{it}}$ and $\phi_{\xi_{it}}$ as the CDF and PDF of the independent error ξ_{it}, respectively. We have

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

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 $b = (\gamma, \beta)^T$. 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. The estimator of average structural function (ASF) in the dynamic probit model is

$$\mathrm{ASF}(X_t, Y_{t-1}) = \int_{\mathcal{U}_t} \Phi_{\xi_{it}}\left(X_t'\beta + \gamma Y_{t-1} + U_t\right) f_2(U_t) dU_t, \tag{10}$$

which represents the conditional mean of $\omega(y_t) = y_t$.

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

$$\begin{aligned} f_{Y_{it}|X_{it},Y_{it-1},U_{it}} &= \left[1 - \Phi_{\xi_{it}}\left(X_{it}'\beta + \gamma Y_{it-1} + U_{it}\right)\right]^{\mathbf{1}(Y_{it}=0)} \times \phi_{\xi_{it}}(y_{it} - X_{it}'\beta - \gamma Y_{it-1} - U_{it})^{\mathbf{1}(Y_{it}>0)} \\ &= \left[1 - \Phi\left(\frac{X_{it}'\beta + \gamma Y_{it-1} + U_{it}}{\sigma_\xi}\right)\right]^{\mathbf{1}(Y_{it}=0)} \times \left[\frac{1}{\sigma_\xi}\phi\left(\frac{y_{it} - X_{it}'\beta - \gamma Y_{it-1} - U_{it}}{\sigma_\xi}\right)\right]^{\mathbf{1}(Y_{it}>0)}, \end{aligned} \tag{11}$$

and the parameter is $\theta = b = (\gamma, \beta, \sigma_\xi^2)^T$. Since $\xi_{it} \sim N(0, \sigma_\xi)$, $E_{Y_t}[y_t|X_t, Y_{t-1}, U_t] = \Phi\left(\frac{X_t'\beta+\gamma Y_{t-1}+U_t}{\sigma_\xi}\right)(X_t'\beta + \gamma Y_{t-1} + U_t) + \sigma_\xi \phi\left(\frac{X_t'\beta+\gamma Y_{t-1}+U_t}{\sigma_\xi}\right)$. The estimator of ASF in the dynamic tobit model is

$$\begin{aligned} &\mathrm{ASF}(X_t, Y_{t-1}) \\ &= \int_{\mathcal{U}_t} \left[\Phi\left(\frac{X_t'\beta + \gamma Y_{t-1} + U_t}{\sigma_\xi}\right)(X_t'\beta + \gamma Y_{t-1} + U_t) \right. \\ &\quad \left. + \sigma_\xi \phi\left(\frac{X_t'\beta + \gamma Y_{t-1} + U_t}{\sigma_\xi}\right)\right] f_2(U_t)dU_t. \qquad (12) \end{aligned}$$

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

$$\begin{aligned} &Y_{it} = 1(\beta_0 + \beta_1 X_{it} + \gamma Y_{it-1} + U_{it} + \xi_{it} \ge 0) \\ &\quad \text{with } U_{it} = V_i + \rho\varepsilon_{it-1}\ \forall i = 1, \ldots, N;\ t = 1, \ldots, T-1 \qquad (13) \end{aligned}$$

and

$$\begin{aligned} &Y_{it} = \max\{\beta_0 + \beta_1 X_{it} + \gamma Y_{it-1} + U_{it} + \xi_{it}, 0\} \\ &\quad \text{with } U_{it} = V_i + \rho\varepsilon_{it-1}\ \forall i = 1, \ldots, N;\ t = 1, \ldots, T-1 \qquad (14) \end{aligned}$$

where $V_i \sim N(1, 1/2)$. To construct the sieve MLE, it is necessary to integrate out the unobserved covariate U_{it}. Here U_{it} has an unbounded domain $(-\infty, \infty)$ and we adopted Gauss–Hermite quadrature for approximating the value of the integral. We consider the mode condition for Assumption 3.5 and use $f_\epsilon(x) = \exp(x - e^x)$ in Eq. (7) for 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. As discussed in Section 3.3, these data generating processes satisfy the identification Assumptions 3.1–3.5.

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 of 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:

$$\mathrm{DGP\ I}: (\beta_0, \beta_1, \gamma, \sigma_\xi^2, \rho) = (0, -1, 0, 0.5, 0)$$

$$\mathrm{DGP\ II}: (\beta_0, \beta_1, \gamma, \sigma_\xi^2, \rho) = (0, -1, 0, 0.5, 0.5)$$

$$\mathrm{DGP\ III}: (\beta_0, \beta_1, \gamma, \sigma_\xi^2, \rho) = (0, -1, 1, 0.5, 0)$$

$$\mathrm{DGP\ IV}: (\beta_0, \beta_1, \gamma, \sigma_\xi^2, \rho) = (0, -1, 1, 0.5, 0.5)$$

$$\mathrm{DGP\ V}: (\beta_0, \beta_1, \gamma, \sigma_\xi^2, \rho) = (0, -1, 1, 0.5, -0.5).$$

The first two DGPs are not state dependent ($\gamma = 0$) while the rest are state dependent with $\gamma = 1$. A sample size $N = 500$ is considered.[22] To secure a more stationary sample, the sampling data are drawn over $T = 7$ periods but only last three periods are utilized. Hundred simulation replications are conducted at each estimation.

Table 1 presents simulation results under the semi-parametric probit model. The simulation results of DGP I (only allows for unobserved heterogeneity) show small standard deviations exist in the structural model coefficients (β_0, γ) comparing to the benchmark estimator. For DGP II, the results have downward bias in the structural model coefficient β_1. In addition, with nontrivial transitory component ($\rho \neq 0$) in DGP II, the standard deviations of $(\beta_0, \beta_1, \gamma)$ are not much different from DGP I. As for DGPs with nontrivial state dependence, bias for $(\beta_0, \beta_1, \gamma)$ for these DGPs is around 0.01 or less and their standard deviations are around 0.1. The coefficient estimators of γ in these DGPs have very small bias for all sample sizes, which means that our estimation for state dependence is very precise among processes with serial correlation ($\rho \neq 0$). In general, the means and medians of (β_1, γ) are very close to each other, reflecting little skewness in their respective distributions. Table 2 shows the simulation of the average partial effects in dynamic probit models in these DGPs. When there is no state dependence (DGP I & II), the estimates for average partial effects do not vary much with the lagged value Y_{t-1}. However, when DGPs contain state dependence, the difference in the average responses are up to 0.12. Results using the benchmark estimator have much larger standard deviations than ones using the proposed estimator.

Table 3 reports the results of estimates for the semi-parametric tobit model. In the tobit model, there is negative bias in β_1 for all DGPs. In tobit case, we have additional parameters to estimate, σ_ξ^2. There is upward bias of the parameter in all DGPs and their standard deviations are a little bit higher in DGPs with nontrivial state dependence. For these DGPs with positive state dependence, estimation results of γ show that there is small bias and precision is within 0.05. Also, the means and medians of all model parameters are not much different, reflecting low degree of skewness in distributions. Table 4 shows the results of the average

Table 1
Monte Carlo simulation of the semi-parametric probit model ($n = 500$).

	DGP	Parameters		
		β_0	β_1	γ
DGP I:	True value	0	−1	0
	Mean benchmark	−0.033	−1.011	0.059
	Median benchmark	0.017	−1.016	0.008
	Standard deviation	0.387	0.065	0.452
	Mean estimate	0.008	−0.994	−0.013
	Median estimate	0.010	−1.006	−0.002
	Standard deviation	0.086	0.103	0.108
DGP II:	True value	0	−1	0
	Mean benchmark	0.015	−1.013	0.024
	Median benchmark	0.006	−1.011	0.021
	Standard deviation	0.125	0.065	0.101
	Mean estimate	−0.003	−1.010	0.007
	Median estimate	−0.012	−1.004	0.011
	Standard deviation	0.087	0.095	0.110
DGP III:	True value	0	−1	1
	Mean benchmark	0.002	−1.004	0.998
	Median benchmark	−0.001	−1.005	0.997
	Standard deviation	0.134	0.071	0.093
	Mean estimate	0.008	−0.991	0.997
	Median estimate	0.016	−0.994	1.000
	Standard deviation	0.093	0.105	0.106
DGP IV:	True value	0	−1	1
	Mean benchmark	−0.052	−0.999	1.056
	Median benchmark	−0.014	−1.000	1.015
	Standard deviation	0.412	0.055	0.411
	Mean estimate	−0.005	−1.005	1.008
	Median estimate	0.003	−1.024	1.010
	Standard deviation	0.092	0.104	0.121
DGP V:	True value	0	−1	1
	Mean benchmark	0.012	−1.010	1.000
	Median benchmark	0.001	−1.011	1.001
	Standard deviation	0.112	0.066	0.096
	Mean estimate	−0.001	−0.996	0.996
	Median estimate	0.012	−1.002	0.982
	Standard deviation	0.112	0.095	0.093

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. The benchmark estimator is an unfeasible MLE using the unobserved covariate U_{it}. Standard deviations of the parameters are computed by the standard deviation of the estimates across 100 simulations and called (simulation) standard deviations.

22 Simulation results for other two different sample sizes, $N = 250, 1000$ are online.

Table 2
Simulation of average structural functions in the probit model ($n = 500$).

	State dependence	Average structural functions	
DGP I:	$Y_{t-1} = 0$	Mean benchmark	0.281
		Standard deviation	(0.214)
		Mean estimate	0.574
		Standard deviation	(0.029)
	$Y_{t-1} = 1$	Mean benchmark	0.281
		Standard deviation	(0.214)
		Mean estimate	0.572
		Standard deviation	(0.035)
DGP II:	$Y_{t-1} = 0$	Mean benchmark	0.307
		Standard deviation	(0.216)
		Mean estimate	0.582
		Standard deviation	(0.029)
	$Y_{t-1} = 1$	Mean benchmark	0.307
		Standard deviation	(0.216)
		Mean estimate	0.580
		Standard deviation	(0.034)
DGP III:	$Y_{t-1} = 0$	Mean benchmark	0.301
		Standard deviation	(0.219)
		Mean estimate	0.572
		Standard deviation	(0.021)
	$Y_{t-1} = 1$	Mean benchmark	0.640
		Standard deviation	(0.204)
		Mean estimate	0.696
		Standard deviation	(0.028)
DGP IV:	$Y_{t-1} = 0$	Mean benchmark	0.265
		Standard deviation	(0.220)
		Mean estimate	0.584
		Standard deviation	(0.036)
	$Y_{t-1} = 1$	Mean benchmark	0.587
		Standard deviation	(0.233)
		Mean estimate	0.707
		Standard deviation	(0.045)
DGP V:	$Y_{t-1} = 0$	Mean benchmark	0.282
		Standard deviation	(0.203)
		Mean estimate	0.586
		Standard deviation	(0.036)
	$Y_{t-1} = 1$	Mean benchmark	0.614
		Standard deviation	(0.218)
		Mean estimate	0.717
		Standard deviation	(0.048)

Note: The average structural functions are reported at the mean value of the explanatory variable and two different outcomes of Y_{t-1}, 0 and 1. Standard deviations of these average structural functions are computed by the standard deviation of the estimates across 100 simulations and called (simulation) standard deviations. The true values of ASF are computed using the unobserved covariate U_{it}. Average partial effects of Y_{t-1} can be obtained by taking differences of average structural functions at $Y_{t-1} = 0$, and $Y_{t-1} = 1$.

Table 3
Monte Carlo simulation of semi-parametric tobit model ($n = 500$).

	DGP	Parameters			
		β_0	β_1	γ	σ_ξ^2
DGP I:	True value	0	−1	0	0.5
	Mean benchmark	0.001	−1.002	−0.023	0.502
	Median benchmark	−0.003	−1.003	0.002	0.502
	Standard deviation	0.103	0.064	0.289	0.084
	Mean estimate	0.007	−1.006	0.002	0.525
	Median estimate	0.006	−0.992	0.009	0.523
	Standard deviation	0.092	0.111	0.103	0.031
DGP II:	True value	0	−1	0	0.5
	Mean benchmark	−0.013	−0.994	−0.009	0.494
	Median benchmark	−0.003	−0.991	−0.009	0.496
	Standard deviation	0.088	0.049	0.128	0.065
	Mean estimate	0.001	−1.009	0.017	0.526
	Median estimate	−0.014	−1.009	0.019	0.524
	Standard deviation	0.112	0.096	0.098	0.030
DGP III:	True value	0	−1	1	0.5
	Mean benchmark	0.001	−1.004	1.002	0.499
	Median benchmark	0.004	−1.000	1.000	0.500
	Standard deviation	0.096	0.057	0.052	0.052
	Mean estimate	0.015	−1.011	0.989	0.528
	Median estimate	0.014	−1.003	0.994	0.526
	Standard deviation	0.100	0.112	0.114	0.035
DGP IV:	True value	0	−1	1	0.5
	Mean benchmark	−0.001	−1.006	1.004	0.501
	Median benchmark	−0.002	−1.013	1.001	0.506
	Standard deviation	0.084	0.056	0.047	0.051
	Mean estimate	0.007	−1.015	0.988	0.501
	Median estimate	0.017	−1.023	0.986	0.523
	Standard deviation	0.093	0.103	0.101	0.036
DGP V:	True value	0	−1	1	0.5
	Mean benchmark	0.001	−1.007	1.005	0.502
	Median benchmark	0.003	−1.003	1.007	0.505
	Standard deviation	0.072	0.045	0.057	0.055
	Mean estimate	−0.002	−1.030	0.997	0.528
	Median estimate	0.008	−1.026	0.996	0.527
	Standard deviation	0.108	0.099	0.120	0.035

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. The benchmark estimator is an unfeasible MLE using the unobserved covariate U_{it}. Standard deviations of the parameters are computed by the standard deviation of the estimates across 100 simulations and called (simulation) standard deviations.

partial effects in dynamic tobit models. There are larger standard deviations of average structural functions and state dependence in DGPs with positive state dependence. Similar to the results in Table 2, results using the benchmark estimator have much larger standard deviations than ones in the proposed estimator.

In some estimation results of parameters, the simulation standard deviation is smaller for the proposed semi-parametric estimator than for the benchmark parametric MLE. An explanation for this observation is that we have adopted Gauss–Hermite quadrature for approximating the value of the integral in the sieve MLE and the distribution of the weights of Gauss–Hermite quadrature is close to a normal distribution. On the other hand, in our simulation design, the unobserved covariate U_{it} is normally distributed. This may reduce the simulated standard deviation because in this case the weight function used in numerical integration has the same functional form as a normal PDF.

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 $l_n = 2$.[23] While a formal selection rule for these smoothing parameters would be desirable, it is difficult to provide a general guideline. From our experience, the estimation of the finite-dimensional parameters θ is not very sensitive to these smoothing parameters. If one cares about estimation of nonparametric density functions, one should pick the smoothing parameters to minimize the approximate mean squared errors of the estimator. In the Monte Carlo study, this is relatively easy to do because the true values are known. But in empirical applications where the true values of the parameters are unknown, it is still a difficult task. A rule of thumb is to pick the smoothing parameters such that the estimates are not sensitive to small variations in the smoothing

23 The numbers of term, i_n, j_n, and k_n represent the length of three univariate Fourier series. See Online Appendix for details.

Table 4
Simulation of average effects in tobit model ($n = 500$).

	Average structural functions		State dependence	
DGP I:	Mean benchmark	0.171	Mean benchmark	0.314
	Standard deviation	(0.243)	Standard deviation	(0.253)
	Mean estimate	0.357	Mean estimate	0.399
	Standard deviation	(0.080)	Standard deviation	(0.069)
DGP II:	Mean benchmark	0.131	Mean benchmark	0.263
	Standard deviation	(0.164)	Standard deviation	(0.210)
	Mean estimate	0.360	Mean estimate	0.407
	Standard deviation	(0.096)	Standard deviation	(0.086)
DGP III:	Mean benchmark	0.474	Mean benchmark	1.189
	Standard deviation	(0.353)	Standard deviation	(0.519)
	Mean estimate	0.620	Mean estimate	1.016
	Standard deviation	(0.146)	Standard deviation	(0.187)
DGP IV:	Mean benchmark	0.437	Mean benchmark	1.116
	Standard deviation	(0.353)	Standard deviation	(0.537)
	Mean estimate	0.652	Mean estimate	1.045
	Standard deviation	(0.106)	Standard deviation	(0.143)
DGP V:	Mean benchmark	0.468	Mean benchmark	1.159
	Standard deviation	(0.361)	Standard deviation	(0.535)
	Mean estimate	0.655	Mean estimate	1.073
	Standard deviation	(0.149)	Standard deviation	(0.180)

Note: The average structural functions are reported at the mean value of the explanatory variable including the lagged dependent variable. Standard deviations of these estimation results are computed by the standard deviation of the estimates across 100 simulations and called (simulation) standard deviations. Average partial effects of Y_{t-1} or state dependence can be obtained by taking the derivative of ASF at means. The true values of ASF and state dependence are computed using the unobserved covariate U_{it}.

parameter.[24] While the Fourier approximations to the evolution density $f_{X_{t+1}|X_t,U_t}$ have the density restriction and the identification restriction, there exists only a density restriction for the approximations to the square root of the initial joint distribution $f^{1/2}_{X_t,Y_{t-1},X_{t-1},U_t}$ using Fourier basis.[25] The semi-parametric sieve MLE using this construction does not encounter any negative integral inside the logarithm in Eq. (8) in our Monte Carlo study. As for the semi-parametric tobit model, we have similar choices of approximation series. The detailed sieve expression of these nuisance parameters can be found in Online Appendix.

The standard deviations can be computed from bootstraps from draws of the original sample. The use of nonparametric bootstrap provides an asymptotically valid standard deviations for the sieve MLE estimate for the finite dimensional parameters θ. The discussion of the consistency of the ordinary nonparametric bootstrap for θ can be found in Chen et al. (2003). Set $Z_{ti} = (X_{it+1}, Y_{it}, X_{it}, Y_{it-1}, X_{it-1})$ and then define a moment function as $m(Z_t, \theta, f_1, f_2) = \ln f\,(x_{t+1}, y_t, x_t, y_{t-1}, x_{t-1}; \theta, \delta_1, \delta_2)$, where

$$f\,(x_{t+1}, y_t, x_t, y_{t-1}, x_{t-1}; \theta, \delta_1, \delta_2) = \int f_1(x_{t+1}|x_t, u_t; \delta_1) f_{Y_t|X_t,Y_{t-1},U_t}(y_t|x_t, y_{t-1}, u_t; \theta) \times f_2(x_t, y_{t-1}, x_{t-1}, u_t; \delta_2) du_t.$$

The notation connects the proposed sieve MLE to the setting in Chen et al. (2003). Sufficient conditions for the bootstrap validity in Chen et al. (2003) include the identification of a parameter, the approximation of a sequence of sieve spaces to infinite dimensional parameters, and the regularity conditions of the moment function. These conditions are close to conditions of the consistency and asymptotic normality in the Appendix B of Online Appendix.[26] In a sieve related estimation method, Ai and Chen (2003) also adopted bootstrap standard deviations as standard deviations of their sieve minimum distance estimator in the simulation study.

In summary, the Monte Carlo study shows that our semi-parametric sieve MLE performs well with a finite sample since mean and median estimates are close to the true values with reasonable standard deviations.

6. Empirical example

In this section, we apply our estimator to a dynamic discrete choice model, which describes 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) we require no initial conditions assumption.[27] Hyslop (1999) also studied a similar empirical model with less general assumptions but specified parametric forms of the unobserved heterogeneity V_i and AR(1) time dependence ρ of the transitory error component ε_{it}. Since these two terms are not separately identified from our main result Theorem 3.1, the empirical study here will focus on the parameters of exogenous explanatory variables and lagged dependent variable not the distributions of the error terms. On the other hand, these estimations might not be comparable across specifications, because of the estimator-specific normalizations in binary choice models. Since the average partial effect is identified in Corollary 3.1, the empirical study also focuses on comparable average partial effects.

24 There is no justified general rule on the choice of number of series terms. For each smoothing parameter, a minimum choice of number of terms is 2 because a sieve series with each smoothing parameter less than 2 is too restrictive and may not approximate well. Thus, each smoothing parameter should be at least 2. Start with an approximation series whose smoothing parameter is 2 in each univariate series and construct a corresponding likelihood to conduct Monte Carlo experiment. If the result of the simulation based on the approximation series is not satisfactory, then try to add more terms. In this case, we added more terms in p_{1i} and q_i while fixing other univariate series because it is easier to add terms in one particular univariate series without changing the whole structure of the approximation series. If this does not work well then do the adding and fixing step to other univariate series. The process can continue to an approximation series whose smoothing parameter is at least 3 in each univariate series. Therefore, the search procedure is complete and help us determine the number of series terms. In addition, a discussion in Hu and Schennach (2008a) suggests that a suitable choice of the smoothing parameters lies between short series and long series where the smoothing bias and the statistical noise dominate respectively.

25 An approximation series to a positive density function may take negative values. A natural log value of a negative value is infinity and this may make the construction of log likelihood function infeasible. Using an approximation series to the square root of the initial joint distribution yields a positive approximation to the positive density function.

26 For example, we have Assumption B.5 for that $\ln f_{Z_t}(z_t; \alpha)$ is Hölder continuous and Chen et al. (2003) provided Hölder continuity as one of primitive sufficient assumptions for their bootstrap result. Therefore, we may not need to impose extra assumptions on the validity of bootstrapping standard errors.

27 In Hyslop (1999), a correlated random-effects (CRE) specification for v_i is

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

where y_{mtis} is i's transitory nonlabor income in year s. An alternative CRE specification can be

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

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

6.1. Specifications and estimation results

According to a theoretical 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. 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

$$Y_t = 1(W_t > W_{0t}^* - \gamma Y_{t-1}) \tag{15}$$

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. (15) is as follows:

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

where X_{it} is a vector of observed demographic and family structure variable, 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; ε_{it} is a serially correlated error term, which captures factors such as transitory wage movements.

In order to provide comparison of the estimators developed in this paper and by Hyslop (1999), we also use the data related to waves 12–19 of 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 husband is a labor force participant in each of the sample years. A woman is defined to be a labor market participant if she works for money any time in the sample year.[28] We obtain a sample having 1752 married women.[29]

As the identification of the models hinges on assumptions in Section 3, a careful discussion of them in this labor force application is necessary, while we realize that testing these assumptions is not feasible as discussed before. Assumption 3.1 is a model specification and it implies that regardless of whatever is in X_{it}, Y_{it-1}, and U_{it}, enough information has been included so that further lags of participation decision and the explanatory variables including nonlabor income, fertility status, etc., do not matter for explaining the current participation decision Y_{it} directly. Assumption 3.3 imposes functional form restrictions on the covariate evolution and the initial joint distribution. Assumption 3.4 in the empirical application may be $E\left[I\left(Y_{it}=0\right)|x_{it}, y_{it-1}, u_{it}\right] = F_{\xi_{it}}\left[-\left(x_{it}'\beta + \gamma y_{it-1} + u_{it}\right)\right]$, which is decreasing in u_{it}. Since u_{it} can represent or contain unobserved heterogeneity such as individual ability or motivation, the assumption suggests that the conditional expectation of the absence from labor force decreases with ability or motivation. Our choice of G in Assumption 3.5 is the mode since the covariate X_{it} contains income variables. In Current Population Survey (CPS), it was found that the mode of misreported income conditional on true income is equal to the true income (see Bound and Krueger (1991) and Chen et al. (2008). Using the mode condition may relieve concerns on measurement errors. Obviously, this is not the only choice of the functional G. As discussed before, we may use mean or median as well.

We then focus on Assumption 3.2. The discussion of the assumption in Section 3 suggests that it imposes the key restriction that conditional on X_{it} and U_{it}, X_{it+1} is independent of the exogenous shock ξ_{it} and the lagged effects of Y_{it} enter the evolution of X_{it+1} through U_{it}. The regressors of interest in this empirical application are the nonlabor income variables and the fertility variables. There are several scenarios for the exogenous participation shock ξ_{it}. First, if ξ_{it} denotes the measurement error, then the conditional independence between ξ_{it} and the future nonlabor income and fertility variables is plausible. Second, if ξ_{it} represents luck in labor markets such as unexpected change of child-care cost or fringe benefit for married women from working, the assumption rules out the immediate effect of the current shock ξ_{it} on the future nonlabor income and fertility variables. This implies that married women do not adjust their nonlabor income and fertility variables to the latest participation shock ξ_{it} but consider all other past period information. If there was a negative shock on participation, married women's nonlabor income and fertility decisions would wait one period to respond to it. Therefore, Assumption 3.2 may be plausible in our model of the intertemporal labor force participation behavior of married women. Nevertheless, Assumption 3.2 does rule out the possible correlation between the fertility decisions in X_{it+1} and a negative shock on labor force participation ξ_{it} even conditioning on the fertility decisions in the previous period in X_{it}. While the lagged effects of Y_{it} enter the evolution of X_{it+1} indirectly here, our identification strategy still applies with $f_{X_{it+1}|Y_{it},X_{it},Y_{it-1},X_{it-1},U_{it}} = f_{X_{it+1}|X_{it},Y_{it-1},U_{it}}$ in Assumption 3.2 if Y_{it-1} has direct influence on X_{it+1}. This alternative specification implies that the labor force participation in period $t-1$ affect married women's future nonlabor income and fertility decisions.

We then apply the sieve MLE method introduced in Sections 4 and 5 and maintain a single-index form and a mode condition. The estimation results for the various models of labor force participation are presented in Table 5 which includes estimates from static probit models with random effect (column 1), a maximum simulated likelihood (MSL) estimator[30] (column 2), and the sieve MLE estimator (column 3) for dynamic models. All specifications include unrestricted time effects, a quadratic in age, race, years of education, permanent and transitory nonlabor income y_{mp} and y_{mt}, current realizations of the number of children aged 0–2, 3–5, and 6–17, and lagged realizations of the number of children aged 0–2.[31] While the first two estimators are estimated using full seven years of data, the last one is estimated over three periods of data. In addition, the last estimator is for the dynamic model without an initial condition 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 correlation between the unobserved heterogeneity and the regressors. The estimation results of coefficients and APEs indicate 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

[28] A standard definition of a participant is 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.

[29] Hyslop (1995) obtains a sample consisted of 1812 observations. The descriptive statistics of our sample is very close to Hyslop (1995).

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

[31] The labor earnings of the husband are used as a proxy for nonlabor income. Permanent nonlabor income y_{mp} is estimated by the sample average, and transitory income y_{mt} is measured as deviations from the sample average.

Table 5
Estimates of married women's participation outcomes.

	Static probit + RE (1)		MSL, RE AR(1) + SD(1) (2)		Semi-parametric probit (3)	
	Coefficient	APE	Coefficient	APE	Coefficient	APE
y_{t-1}	–	–	1.117	0.325	1.089	0.225
	–	–	(0.528)	(0.015)	(0.077)	(0.014)
y_{mp}	−0.312	−0.070	−0.007	−0.002	−0.221	−0.048
	(0.045)	(0.005)	(0.017)	(0.001)	(0.012)	(0.003)
y_{mt}	−0.106	−0.024	−0.004	−0.001	−0.106	−0.023
	(0.026)	(0.002)	(0.028)	(0.001)	(0.056)	(0.001)
#Kid0-2_{t-1}	−0.022	−0.005	−0.117	−0.036	−0.055	−0.012
	(0.010)	(0.001)	(0.013)	(0.002)	(0.048)	(0.001)
#Kid0-2_t	−0.330	−0.070	−0.380	−0.112	−0.316	−0.065
	(0.021)	(0.005)	(0.145)	(0.006)	(0.061)	(0.004)
#Kid3-5_t	−0.400	−0.086	−0.206	−0.062	−0.137	−0.029
	(0.015)	(0.007)	(0.027)	(0.003)	(0.028)	(0.002)
#Kid6-17_t	−0.120	−0.028	−0.056	−0.018	−0.062	−0.014
	(0.011)	(0.002)	(0.037)	(0.001)	(0.011)	(0.001)
Cov. parameters						
σ_v^2	0.786	–	0.313	–	–	–
	(0.071)	–	(0.323)	–	–	–
ρ	–	–	−0.146	–	–	–
	–	–	(0.140)	–	–	–

Note: Bootstrap (simulation) 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. APEs are reported by taking derivatives or differences of ASF at the sample mean of (x_t, y_{t-1}).

variance of unobserved heterogeneity is 0.786. We now turn to dynamic specifications. The specifications in the MSL estimator contain random effects, a stationary AR(1) error component, and first-order state dependence (SD(1)). The estimated coefficients and APEs share a similar pattern. The APE estimates show a large and significant first-order state dependence effect reduces the labor force participation probability by about 0.325. The addition of SD(1) and AR(1) error component greatly reduced the effects of nonlabor income variables (−0.002 and −0.001) and the contemporaneous fertility variables like #Kid3-5_t and #Kid6-17_t. 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 the probability of women's participation decisions (−0.036 and −0.112). 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.[32]

The results also show that first-order state dependence has a significant positive effect on the probability of participation (0.225). There exists a strong dependence between married women's current labor force participation and past labor force participation, and relaxing the initial condition 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 estimation are generally similar to those in column (1) and (2) but with smaller 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 estimator, the unobserved heterogeneity and the AR(1) component have been mixed into the unobserved covariate U_{it}. They are not identified so there are not any estimation results.

In comparison to the results across specifications allowing for CRE, AR(1), and SD(1), using unspecified CRE and avoiding initial conditions have significant effect on the estimation. The APE estimates find a larger significant negative effects on nonlabor income variables (−0.048 and −0.002 vs. −0.023 and −0.001, respectively) and negative effects of children aged 0–2 in the current period and past period which decreases by 42% (from −0.112 to −0.065) and decline by 66% (from −0.036 to −0.012) respectively.

7. Conclusion

This paper presents the nonparametric identification of nonlinear dynamic panel data models with unobserved covariates. We show that the models are identified using only two periods of the dependent variable Y_{it} and three periods of the covariate X_{it} without initial conditions assumptions. 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 errors. Monte Carlo experiments have shown 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. Identification in the discrete case

We will show how to utilize the identification techniques in Section 2 for the discrete case. The discrete case refers to that the variables X_{it} and U_{it} is discrete:

$$X_{it} \in \mathcal{X}_t \equiv \{1, 2, \ldots, J_1\} \quad \text{and} \quad U_{it} \in \mathcal{U} \equiv \{1, 2, \ldots, J_2\}.$$

In this finite dimensional discrete example, linear integral operators are matrices, which might be useful to give some intuition about how the identification is achieved. For simplicity, assume

[32] 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.

that $J_1 = J_2 = J$. Based on Eq. (4) which is the consequence of Assumptions 3.1 and 3.2, the key equation of the discrete case is

$$f_{X_{it+1},Y_{it},X_{it},Y_{it-1},X_{it-1}} = \sum_{U_{it}=1}^{J} f_{X_{it+1}|X_{it},U_{it}} f_{Y_{it}|X_{it},Y_{it-1},U_{it}} \times f_{X_{it},Y_{it-1},X_{it-1},U_{it}}. \quad (16)$$

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

$$L_{X_{it+1},y_{it},x_{it},y_{it-1},X_{it-1}} = \left[f_{X_{it+1},Y_{it},X_{it},Y_{it-1},X_{it-1}} \times (u, y_{it}, x_{it}, y_{it-1}, x)\right]_{u,x}$$

$$L_{X_{it+1},x_{it},y_{it-1},X_{it-1}} = \left[f_{X_{it+1},X_{it},Y_{it-1},X_{it-1}}(u, x_{it}, y_{it-1}, x)\right]_{u,x}$$

$$L_{X_{it+1}|x_{it},U_{it}} = \left[f_{X_{it+1}|X_{it},U_{it}}(x|x_{it}, u)\right]_{x,u}$$

$$L_{x_{it},y_{it-1},X_{it-1},U_{it}} = \left[f_{X_{it},Y_{it-1},X_{it-1},U_{it}}(x_{it}, y_{it-1}, x, u)\right]_{u,x}$$

and a J-by-J diagonal matrix

$$D_{y_{it}|x_{it},y_{it-1},U_{it}} = \begin{bmatrix} f_{Y_{it}|X_{it},Y_{it-1},U_{it}}(y_{it}|x_{it}, y_{it-1}, 1) & 0 & 0 \\ 0 & \cdots & 0 \\ 0 & 0 & f_{Y_{it}|X_{it},Y_{it-1},U_{it}}(y_{it}|x_{it}, y_{it-1}, J) \end{bmatrix}.$$

Using these matrices, Eq. (16) can be expressed into a matrix notation as

$$L_{X_{it+1},y_{it},x_{it},y_{it-1},X_{it-1}} = L_{X_{it+1}|x_{it},U_{it}} D_{y_{it}|x_{it},y_{it-1},U_{it}} L_{x_{it},y_{it-1},X_{it-1},U_{it}}. \quad (17)$$

Integrating out Y_{it} in Eq. (16) leads to

$$f_{X_{it+1},X_{it},Y_{it-1},X_{it-1}} = \sum_{u_{it}=1}^{J} f_{X_{it+1}|X_{it},U_{it}} f_{X_{it},Y_{it-1},X_{it-1},U_{it}} \quad (18)$$

which is equivalent to

$$L_{X_{it+1},x_{it},y_{it-1},X_{it-1}} = L_{X_{it+1}|x_{it},U_{it}} L_{x_{it},y_{it-1},X_{it-1},U_{it}}. \quad (19)$$

Assumption 3.3 guarantees that the above matrix $L_{X_{it+1},x_{it},y_{it-1},X_{it-1}}$ is invertible. It follows that

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

The observed matrix on the LHS has a matrix factorization, the product of a diagonal matrix with a matrix of eigenvectors. Uniqueness of the factorization requires the distinct eigenvalues and normalization of the unobserved covariate U_{it}. Assumptions 3.4 and 3.5 are imposed to make these conditions hold. Since the eigenvalues and eigenvectors in the matrix factorization are $f_{Y_{it}|X_{it},Y_{it-1},U_{it}}$ and $f_{X_{it+1}|X_{it},U_{it}}$ respectively, the identification of the model is reached. By Eq. (17), the initial joint distribution $f_{X_{it},Y_{it-1},X_{it-1},U_{it}}$ is also identified.

Appendix B. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.jeconom.2013.03.001.

References

Ahn, S., Schmidt, P., 1995. Efficient estimation of models for dynamic panel data. Journal of Econometrics 68 (1), 5–28.

Ai, C., Chen, X., 2003. Efficient estimation of models with conditional moment restrictions containing unknown functions. Econometrica 71 (6), 1795–1843.

Altonji, J., Matzkin, R., 2005. Cross section and panel data estimators for nonseparable models with endogenous regressors. Econometrica 73 (4), 1053–1102.

Anderson, T., Hsiao, C., 1982. Formulation and estimation of dynamic models using panel data. Journal of Econometrics 18 (1), 47–82.

Arellano, M., Bond, S., 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. The Review of Economic Studies 58 (2), 277–297.

Arellano, M., Bonhomme, S., 2009. Robust priors in nonlinear panel data models. Econometrica 77 (2), 489–536.

Arellano, M., Bover, O., 1995. Another look at the instrumental variable estimation of error component models. Journal of Econometrics 68 (1), 29–51.

Bound, J., Krueger, A., 1991. The extent of measurement error in longitudinal earnings data: Do two wrongs make a right?. Journal of Labor Economics 9 (1), 1–24.

Canay, I., Santos, A., Shaikh, A., 2011. On the testability of identification in some nonparametric models with endogeneity. Working Paper.

Carro, J., 2007. Estimating dynamic panel data discrete choice models with fixed effects. Journal of Econometrics 140 (2), 503–528.

Carroll, R., Chen, X., Hu, Y., 2010. Identification and estimation of nonlinear models using two samples with nonclassical measurement errors. Journal of Nonparametric Satistics 22 (4), 379–399.

Chamberlain, G., 1980. Analysis of covariance with qualitative data. The Review of Economic Studies 47 (1), 225–238.

Chamberlain, G., 1984. Panel data. In: Griliches, Z., Intriligator, M.D. (Eds.), Handbook of Econometrics, Vol. 2. North-Holland, Amsterdam, pp. 1247–1318.

Chay, K., Hoynes, H., Hyslop, D., 2001. A non-experimental analysis of true state dependence in monthly welfare participation sequences. University of California, Berkeley.

Chen, X., Hong, H., Tarozzi, A., 2008. Semiparametric efficiency in GMM models with auxiliary data. The Annals of Statistics 36 (2), 808–843.

Chen, X., Linton, O., Van Keilegom, I., 2003. Estimation of semiparametric models when the criterion function is not smooth. Econometrica 71 (5), 1591–1608.

Chen, X., Shen, X., 1998. Sieve extremum estimates for weakly dependent data. Econometrica 66 (2), 289–314.

Chernozhukov, V., Fernández-Val, I., Hahn, J., Newey, W., 2009. Identification and estimation of marginal effects in nonlinear panel models. CeMMAP Working Paper.

Chintagunta, P., Kyriazidou, E., Perktold, J., 2001. Panel data analysis of household brand choices. Journal of Econometrics 103 (1–2), 111–153.

Contoyannis, P., Jones, A., Rice, N., 2004. Simulation-based inference in dynamic panel probit models: an application to health. Empirical Economics 29 (1), 49–77.

Evdokimov, K., 2009. Identification and estimation of a nonparametric panel data model with unobserved heterogeneity. Working Paper.

Fernández-Val, I., 2009. Fixed effects estimation of structural parameters and marginal effects in panel probit models. Journal of Econometrics 150 (1), 71–85.

Gourieroux, C., Monfort, A., 1993. Simulation-based inference: a survey with special reference to panel data models. Journal of Econometrics 59 (1–2), 5–33.

Hahn, J., 2001. The information bound of a dynamic panel logit model with fixed effects. Econometric Theory 17 (5), 913–932.

Hahn, J., Kuersteiner, G., 2004. Bias reduction for dynamic nonlinear panel models with fixed effects. Unpublished Manuscript.

Hajivassiliou, V., 1993. Simulation estimation methods for limited dependent variable models. In: Handbook of Statistics, Vol. 59.

Hajivassiliou, V., Ruud, P., 1994. Classical estimation methods for LDV models using simulation. In: Engle, R., McFadden, D. (Eds.), Handbook of Econometrics, Vol. 4. North-Holland, Amsterdam, pp. 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. pp. 227–269.

Heckman, J., 1981a. Statistical models for discrete panel data. In: Manski, C., McFadden, D. (Eds.), Structural Analysis of Discrete Panel Data with Econometric Applications. MIT Press, Cambridge, pp. 114–178.

Heckman, J., 1981b. The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. In: Manski, C., McFadden, D. (Eds.), Structural Analysis of Discrete Panel Data with Econometric Applications. MIT Press, Cambridge, pp. 179–195.

Heckman, J., Willis, R., 1977. A beta-logistic model for the analysis of sequential labor force participation by married women. The Journal of Political Economy 85 (1), 27–58.

Hoderlein, S., White, H., 2009. Nonparametric identification in nonseparable panel data models with generalized fixed effects. CeMMAP Working Paper.

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., Hu, L., 2004. Estimation of cross sectional and panel data censored regression models with endogeneity. Journal of Econometrics 122 (2), 293–316.

Honoré, B., Kyriazidou, E., 2000. Panel data discrete choice models with lagged dependent variables. Econometrica 68 (4), 839–874.

Honoré, B., Tamer, E., 2006. Bounds on parameters in panel dynamic discrete choice models. Econometrica 74 (3), 611–629.

Hu, L., 2000. Estimating a censored dynamic panel data model with an application to earnings dynamics. Unpublished Manuscript, Northwestern University.

Hu, L., 2002. Estimation of a censored dynamic panel data model. Econometrica 70 (6), 2499–2517.

Hu, Y., Schennach, S., 2008a. Instrumental variable treatment of nonclassical measurement error models. Econometrica 76 (1), 195–216.

Hu, Y., Schennach, S., 2008b. Supplement to instrumental variable treatment of nonclassical measurement error models. Econometrica 76 (1), 195–216.
Hu, Y., Shiu, J., 2011. Nonparametric identification using instrumental variables: sufficient conditions for completeness. Department of Economics Working Paper, 581, Jonhs Hopkins University.
Hu, Y., Shum, M., 2010. Nonparametric identification of dynamic models with unobserved state variables. Department of Economics Working Paper, 543, Jonhs Hopkins University.
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.
Hyslop, D., 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. In: Handbook of Statistics, Vol. 11. pp. 545–571.
Newey, W., Powell, J., 2003. Instrumental variable estimation of nonparametric models. Econometrica 71 (5), 1565–1578.
Neyman, J., Scott, E., 1948. Consistent estimates based on partially consistent observations. Econometrica 16, 1–32.
Rasch, G., 1960. Probabilistic Models for Some Intelligence and Attainment Tests. Danmarks Paedagogiske Institut., Copenhagen.
Shen, X., 1997. On Methods of Sieves and Penalization. The Annals of Statistics 25, 2555–2591.
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.
Wooldridge, J., 2010. Econometric Analysis of Cross Section and Panel Data. The MIT press.