The Econometrics Journal

Econometrics Journal (2018), volume **21**, pp. 55–85.
doi: 10.1111/ectj.12086

Identification and estimation of semi-parametric censored dynamic panel data models of short time periods

YINGYAO HU[†] AND JI-LIANG SHIU[‡]

[†]*Department of Economics, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218, USA.*
E-mail: yhu@jhu.edu

[‡]*Institute for Economic and Social Research, Jinan University, Guangzhou, China.*
E-mail: jishiu.econ@gmail.com

First version received: January 2017; final version accepted: February 2017

Summary In this paper, we present a semi-parametric identification and estimation method for censored dynamic panel data models of short time periods and their average partial effects with only two periods of data. The proposed method transforms the semi-parametric specification of censored dynamic panel data models into a parametric family of distribution functions of observables without specifying the distribution of the initial condition. Then the censored dynamic panel data models are globally identified under a standard maximum likelihood estimation framework. The identifying assumptions are related to the completeness of the families of known parametric distribution functions corresponding to censored dynamic panel data models. Dynamic tobit models and two-part dynamic regression models satisfy the key assumptions. We propose a sieve maximum likelihood estimator and we investigate the finite sample properties of these sieve-based estimators using Monte Carlo analysis. Our empirical application using the Medical Expenditure Panel Survey shows that individuals consume more health care when their incomes increase, after controlling for past health expenditures.

Keywords: *Dynamic tobit model, Initial condition, Nonlinear dynamic panel data model, Two-part dynamic regression model, Unobserved covariate, Unobserved heterogeneity.*

1. INTRODUCTION

The identification and estimation of dynamic panel data models is one of the main challenges in econometrics. These models are appealing in applied research because they consider the lagged value of the dependent variable as one of the explanatory variables, and they contain observed and unobserved permanent (heterogeneous) or transitory (serially correlated) individual effects. In this paper, we focus on the identification and estimation of semi-parametric censored dynamic panel data models of short time periods and their average partial effects with two periods of data. The observed distribution $f_{Y_{it},X_{it},Y_{it-1},X_{it-1}}$ of $\{Y_{it}, X_{it}, Y_{it-1}, X_{it-1}\}$ is associated with a parametric distribution of the limited dependent variable Y_{it} conditional on $(X_{it}, Y_{it-1}, U_{it})$, i.e. $f_{Y_{it}|X_{it},Y_{it-1},U_{it};\theta_0}$, where X_{it} is an observed explanatory variable and U_{it} is a time-varying unobservable with the unknown parameter of interest θ_0, and a non-parametric distribution

© 2017 Royal Economic Society. Published by John Wiley & Sons Ltd, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA, 02148, USA.

$f_{X_{it},Y_{it-1},X_{it-1},U_{it}}$ as follows:

$$\begin{aligned} f_{Y_{it},X_{it},Y_{it-1},X_{it-1}} &= \int f_{Y_{it}|X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0} f_{X_{it},Y_{it-1},X_{it-1},U_{it}} du_{it} \\ &= \int f_{Y_{it}|X_{it},Y_{it-1},U_{it};\theta_0} f_{X_{it},Y_{it-1},X_{it-1},U_{it}} du_{it}. \end{aligned}$$

Here, we have used $f_{Y_{it}|X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0} = f_{Y_{it}|X_{it},Y_{it-1},U_{it};\theta_0}$. The condition comes from the parametric assumption on the conditional distribution of the dependent variable, which assumes that $(X_{it}, Y_{it-1}, U_{it})$ contains enough information on the dependent variable.

Relying on the parametric specification of the censored dynamic panel data model $f_{Y_{it}|X_{it},Y_{it-1},U_{it};\theta_0}$, we provide sufficient conditions under which both the finite-dimensional parameter θ_0 and the non-parametric density $f_{X_{it},Y_{it-1},X_{it-1},U_{it}}$ are identified from the observed distribution $f_{Y_{it},X_{it},Y_{it-1},X_{it-1}}$.

In dynamic linear panel data models, researchers have developed and compared several instrumental variable (IV) estimators and generalized method of moment (GMM) estimators in the literature; see, e.g. Anderson and Hsiao (1982), Arellano and Bond (1991), Arellano and Bover (1995), Ahn and Schmidt (1995), Kiviet (1995), Blundell and Bond (1998), Hahn (1999) and Hsiao et al. (2002). When the time dimension T is fixed in nonlinear panel data models, the presence of the unobserved effect prevents the construction of a log-likelihood function that can be used to estimate structural parameters consistently. This is the so-called incidental parameters problem discussed by Neyman and Scott (1948). However, the dynamic nature of the models leads to the initial conditions problem because integrating the individual unobserved effect out of the distribution raises the issue of how to specify the distribution of the initial condition given unobserved heterogeneity. Wooldridge (2005) proposed finding the distribution conditional on the initial value and the observed history of strictly exogenous explanatory variables to solve the initial conditions problem. Shiu and Hu (2013) adopted the correlated random effect approach for nonlinear dynamic panel data models without specifying the distribution of the initial condition. They used the identification results of the non-classical measurement error models of Hu and Schennach (2008) to achieve non-parametric identification of nonlinear dynamic panel data model with three periods of data. Honoré (1993), Hu (2002) and Honoré and Hu (2004) used moment restrictions to identify and estimate the parameters of censored dynamic panel data models. Their results were achieved without making distributions of unobserved heterogeneity and the disturbance, but they failed to identify the average partial effects.

Other quantities of interest in nonlinear panel data applications include the partial effects on the mean response, averaged across the population distribution of the unobserved heterogeneity. Chernozhukov et al. (2013) derived bounds for average effects in non-separable panel data models and showed that they can tighten considerably for semi-parametric discrete choice models. Graham and Powell (2012) studied the average partial effect over the distribution of unobserved heterogeneity, which represents the causal effect of a small change in an endogenous regressor on a continuously valued outcome of interest. Hoderlein and White (2012) considered identification of distributional effects and average effects in general non-separable models, allowing for arbitrary dependence between the persistent unobservables and the regressors of interest, even if there are only two time periods. However, their approach explicitly rules out lagged dependent variables. Dynamic models focus on the effects of the lagged dependent variables on the current dependent variable, whereas we want to account or control for the influence of all other variables. The effect of lagged dependent variables reflects the persistence

© 2017 Royal Economic Society.

1368423x, 2018, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ectj.12086 by Johns Hopkins University, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

of the dependent variables over time and the amount of this state dependence can be measured by the average partial effect.

This study focuses on the identification and estimation of semi-parametric censored dynamic panel data models of short time periods and their average partial effects with two periods of data. Under a semi-parametric specification, the model includes the parametric distribution of censored dependent variable Y_{it} conditional on (X_{it}, Y_{it-1}, U_{it}), i.e. $f_{Y_{it}|X_{it},Y_{it-1},U_{it};\theta_0}$ and the non-parametric distribution $f_{X_{it},Y_{it-1},X_{it-1},U_{it}}$. We show that the identification of this semi-parametric model can be reduced to the identification of parameter θ_0 with a parametric density function $f_{X_{it},Y_{it-1},X_{it-1};\theta_0}$. The identification of θ_0 in $f_{Y_{it}|X_{it},Y_{it-1},U_{it};\theta_0}$ can lead to that of the proposed semi-parametric censored dynamic panel data models of short time periods. This identification technique involves three steps of transformations associated with the completeness of known probability density functions (PDFs). The first step is to apply the inverse of an integral operator using $f_{Y_{it}|X_{it},Y_{it-1},U_{it};\theta}$ as a kernel. The second step is to integrate out the unobserved covariate. The last step is to normalize the integrated semi-parametric density function created in the second step. The true value of structural parameters can be uniquely determined by maximizing the likelihood function of the transformed semi-parametric family of the PDFs of observables. This process also identifies the average partial effect of the censored dynamic panel data models. The transformation steps rely on the completeness of the families of known PDFs $f_{Y_{it}|X_{it},Y_{it-1},U_{it};\theta}$ corresponding to censored dynamic panel data models and observed conditional density functions of the dependent variable given the explanatory variables $f_{Y_{it}|X_{it},Y_{it-1},X_{it-1}}$. Although completeness is usually considered to be a high-level technical condition, we show that the completeness assumptions hold in some popular censored dynamic panel data models, such as dynamic tobit models and two-part dynamic regression models. D'Haultfoeuille (2011), Andrews (2011), Hu and Shiu (2017), Hu et al. (2017) and Chen et al. (2014) provided sufficient conditions for L^2 completeness. If a bounded mixing density is assumed, then the condition for identification becomes the bounded completeness conditions, which is much weaker than the L^2 completeness. See Mattner (1993) and Chernozhukov and Hansen (2005) for uses of bounded completeness.

A study close in spirit to our work is that of Schennach (2014), who introduces a general method to convert a model defined by moment conditions involving both observed and unobserved variables into equivalent moment conditions involving only observable variables. Our approach shares the strategy by converting a conditional density function instead of moment conditions.

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

The application of the sieve MLE is to investigate the effects that describe the dynamic behaviour of annual individual health expenditures using the Medical Expenditure Panel Survey (MEPS) Panel 4. The MEPS data record detailed information on health-care use, expenditures,

© 2017 Royal Economic Society.

1368423x, 2018, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ectj.12086 by Johns Hopkins University, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

sources of payment and insurance coverage for the US population from 1999 to 2000. There are sizable fractions of the sample with zero medical expenditure in the data so the dynamic censored model is applicable. The result of the semi-parametric dynamic censored specifications indicates that individuals consume more health care when their incomes increase, after controlling for past health expenditures.

The rest of the paper is organized as follows. In Section 2, we present the identification of censored dynamic panel data models of short time periods through several transformations. In Section 3, we show that the identification assumptions hold for dynamic tobit models and two-part dynamic regression models. In Section 4, we present the proposed sieve MLE. In Section 5, we show the application of the sieve MLE to a dynamic tobit model describing the dynamic behaviours of annual individual health expenditures, using data from the MEPS. Finally, we provide concluding remarks in Section 6. The appendices include proofs of each transformation step and a discrete case.

2. IDENTIFICATION OF CENSORED DYNAMIC PANEL DATA MODELS

Suppose $g_1(\cdot, \cdot; \theta_1)$, and $g_2(\cdot, \cdot; \theta_2)$ are parametric functions known up to the parameter (θ_1, θ_2). Consider the following censored dynamic panel data model:

$$Y_{it} = g_1(g_2(X_{it}, Y_{it-1}; \theta_{20}), V_i + \varepsilon_{it}; \theta_{10}), \qquad \forall i = 1, \ldots, N; t = 1, \ldots, T. \tag{2.1}$$

Here, Y_{it} is the dependent variable, X_{it} is a vector of observed explanatory variables, ε_{it} is a transitory error term, V_i is an unobservable individual-specific effect and $(\theta_{10}, \theta_{20})$ is the true value of parameters to be estimated. The time series T is short, regardless of the number N of cross-sectional units of the panel. The functions g_1 and g_2 can be specified by users, such as $g_1(\chi, \nu; \theta_1) = \max(0, \chi + \nu)$ and $g_2(X_{it}, Y_{it-1}; \theta_2) = X_{it}'\beta + \gamma Y_{it-1}$, etc. The specifications of g_2 can contain time trends, allowing nonlinear relationships such as quadratics or interactions terms. One of the difficulties of identification is that the variables, (X_{it}', Y_{it-1}, V_i), and the transitory error term, ε_{it}, are not independently distributed. We impose following restrictions on the transitory error term ε_{it} in model (2.1).

ASSUMPTION 2.1. (EXOGENOUS SHOCKS) *Let η_{it} be an unobserved serially correlated component in the past such that $\eta_{it} = \varphi(\{X_{i\tau}, Y_{i\tau-1}, \varepsilon_{i\tau}\}_{\tau=0,1,\ldots,t-1})$ for some function φ. Assume that a transitory random shock ξ_{it} is independent of $\{X_{i\tau}, Y_{i\tau-1}, V_i, \varepsilon_{i\tau-1}\}$ for any $\tau \leq t$, and that the transitory error term ε_{it} has the following decomposition:*

$$\varepsilon_{it} = \eta_{it} + \xi_{it}. \tag{2.2}$$

Plugging (2.2) in Assumption 2.1 into model (2.1) leads to

$$\begin{aligned} Y_{it} &= g_1(g_2(X_{it}, Y_{it-1}; \theta_{20}), V_i + \eta_{it} + \xi_{it}; \theta_{10}) \\ &\equiv g_1(g_2(X_{it}, Y_{it-1}; \theta_{20}), U_{it} + \xi_{it}; \theta_{10}), \end{aligned} \tag{2.3}$$

where $U_{it} = V_i + \eta_{it}$ can be considered as an unobserved covariate. To describe every structure of model (2.3) by a parameter, we assume that the distribution of ξ_{it} has a parametric representation. This effectively reduces the identification problem to the identification of a set of parameters. This framework leads to the following definitions.

© 2017 Royal Economic Society.

1368423x, 2018, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ectj.12086 by Johns Hopkins University, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

DEFINITION 2.1. *Let Θ_3 be a parameter space and let $dF(x;\theta_3)$ be a proper distribution function. If $dF(x;\theta_{30})$ is the true distribution, then $dF(x;\theta_3)$ is correctly specified at θ_{30}. The parameter point θ_{30} is globally identifiable if there exists no other $\theta_3 \in \Theta_3$ such that with probability 1, $dF(x;\theta_3) = dF(x;\theta_{30})$, where the measure is taken with respect to θ_{30}.*

DEFINITION 2.2. *The parameter point θ_{30} is locally identifiable if there exists an open neighbourhood of θ_{30} containing no other θ_3 such that with probability 1, $dF(x;\theta_3) = dF(x;\theta_{30})$, where the measure is taken with respect to θ_{30}.*

If θ_{30} is globally identifiable, then it is locally identifiable.

ASSUMPTION 2.2. *The distribution of the transitory random shock $dF(\xi_{it};\theta_3)$ is known and is correctly specified at an unknown θ_{30}. The parameter point θ_{30} is locally identifiable.*

Under some specifications of functions, g_1 and g_2, according to Assumptions 2.1 and 2.2, there is a unique conditional distribution associated with each structure in the censored dynamic panel data model (2.3) and the identification of the censored dynamic panel data models (2.1) is implied by that of the distribution of Y_{it} conditional on $(X_{it}, Y_{it-1}, U_{it})$ (i.e., $f_{Y_{it}|X_{it},Y_{it-1},U_{it};\theta}$). Because the distribution of U_{it} remains unknown, the assumption gives a semi-parametric representation. Given this representation, the identification problem is to find conditions such that a true underlying parameter $\theta_0 := (\theta_{10}, \theta_{20}, \theta_{30})^T$ can be distinguished on the basis of sample observations. The conditional PDF $f_{Y_{it}|X_{it},Y_{it-1},U_{it};\theta}$ corresponding to $F_{Y_{it}|X_{it},Y_{it-1},U_{it};\theta}$ is called the parametric censored density function in this paper. We introduce two examples to highlight this important connection. Suppose $F_{\cdot}$ and $f_{\cdot}$ denote the CDF and the PDF of an independent random shock, respectively.

EXAMPLE 2.1. (DYNAMIC TOBIT MODEL) Assume $g_1(\chi, \nu;\theta_1) = \max(0, \chi + \nu)$

$$Y_{it} = \max\{0, g_2(X_{it}, Y_{it-1};\theta_2) + U_{it} + \xi_{it}\} \qquad \text{with} \quad \forall i = 1, \ldots, N;\ t = 1, \ldots, T.$$

The parametric censored density function is

$$\begin{aligned} f_{Y_{it}|X_{it},Y_{it-1},U_{it};\theta} &= F_{\xi_{it};\theta_3}(-g_2(X_{it}, Y_{it-1};\theta_2) - U_{it})^{\mathbf{1}(Y_{it}=0)} \\ &\quad \times f_{\xi_{it};\theta_3}(Y_{it} - g_2(X_{it}, Y_{it-1};\theta_2) - U_{it})^{\mathbf{1}(Y_{it}>0)}. \end{aligned} \tag{2.4}$$

EXAMPLE 2.2. (TWO-PART DYNAMIC REGRESSION MODEL) Define a binary indicator variable $d_{it} = 1(g_3(X_{it}, Y_{it-1};\theta_1) + \varsigma_{it} \geq 0)$, where $1(\cdot)$ is the 0–1 indicator function and ς_{it} has a known CDF $F_{\varsigma_{it}}$. Suppose that $Y_{it} > 0$ is observed for $d_{it} = 1$ and $Y_{it} = 0$ for $d_{it} = 0$. When $Y_{it} > 0$,

$$\log(Y_{it}) = g_2(X_{it}, Y_{it-1};\theta_2) + U_{it} + \xi_{it} \qquad \text{with} \quad \forall i = 1, \ldots, N;\ t = 1, \ldots, T.$$

The conditional distribution of interest is

$$\begin{aligned} f_{Y_{it}|X_{it},Y_{it-1},U_{it};\theta} &= F_{\varsigma_{it}}(-g_3(X_{it}, Y_{it-1};\theta_1))^{\mathbf{1}(Y_{it}=0)} \Big\{(1 - F_{\varsigma_{it}}(-g_3(X_{it}, Y_{it-1};\theta_1))) \\ &\quad \times f_{\xi_{it};\theta_3}(\log(Y_{it}) - g_2(X_{it}, Y_{it-1};\theta_2) - U_{it})\frac{1}{Y_{it}}\Big\}^{\mathbf{1}(Y_{it}>0)}. \end{aligned} \tag{2.5}$$

© 2017 Royal Economic Society.

1368423x, 2018, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ectj.12086 by Johns Hopkins University, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The model of Example 2.1 is also considered in Shiu and Hu (2013) but the approach of the paper requires one more period of the covariate. The model of Example 2.2 is closely related to the set-up of Kyriazidou (2001) who allows the outcome and selection equations to have different fixed effects. There exist a number of economic applications of these censored dynamic panel data models, where the dependent variables Y_{it} can represent the amount of insurance coverage chosen by an individual, annual women's labour supply, a firm's expenditures on R&D or annual individual health expenditures. In this paper, our models contain the lagged censored dependent variables on the right-hand side. Because piles of the dependent variable at zero can be regarded as optimal solutions of utility maximizing behaviour, these models can also be considered as corner solution models with lagged censored dependent variables.

2.1. General identification

Consider the parametric censored density function:

$$f_{Y_{it}|X_{it},Y_{it-1},U_{it};\theta}(y_{it}|x_{it}, y_{it-1}, u_{it}), \tag{2.6}$$

where Y_{it} is a limited dependent variable for an individual i, and the explanatory variables include a lagged dependent variable, a set of possibly time-varying explanatory variables X_{it} and the unobserved covariate U_{it}. We assume that $\theta_0 \in \Theta$ is local identifiable. In other words, θ_0 is a unique value of θ in an open neighbourhood of θ_0, which specifies the exact structure of the model. Consider panel data containing two periods, $\{Y_{it}, X_{it}, Y_{it-1}, X_{it-1}\}_i$ for $i = 1, 2, \ldots, N$. Assume that for each i, $(Y_{it}, X_{it}, Y_{it-1}, X_{it-1})$ is an independent random draw from a bounded distribution $f_{Y_{it},X_{it},Y_{it-1},X_{it-1}}$. Applying the law of total probability leads to the following,

$$f_{Y_{it},X_{it},Y_{it-1},X_{it-1}} = \int f_{Y_{it}|X_{it},Y_{it-1},U_{it};\theta_0} f_{X_{it},Y_{it-1},X_{it-1},U_{it}} du_{it}, \tag{2.7}$$

where $f_{X_{it},Y_{it-1},X_{it-1},U_{it}}$ is the joint density function of variables $(x_{it}, y_{it-1}, x_{it-1}, u_{it})$. Let $\mathcal{Y}_{it}$, $\mathcal{X}_{it}$ and $\mathcal{U}_{it}$ be the support of random variables Y_{it}, X_{it} and U_{it}, respectively. Set $L^2(\mathcal{Y}) = \{h(\cdot) : \int_{\mathcal{Y}} |h(y)|^2 dy < \infty\}$ and $L^2(\mathcal{U}, \omega) = \{h(\cdot) : \int_{\mathcal{U}} |h(u)|^2 \omega(u) du < \infty,$ and $\int_{\mathcal{U}} \omega(u) du < \infty\}$. Note the weighted L^2-space, $L^2(\mathcal{U}, \omega)$, contains a constant function (i.e. $c(u) = c \;\forall u \in \mathcal{U}$). We consider the identification issue over a proper subset of $\mathcal{Y}_{it}$. The first step is to construct $f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta}$ and to avoid an unwanted restriction. Let $\widetilde{\mathcal{Y}}_{it}$ be a subset of $\mathcal{Y}_{it}$ and denote $\widetilde{Y}_{it}$ as a random variable whose support is in $\widetilde{\mathcal{Y}}_{it}$ and $\widetilde{Y}_{it} = Y_{it}$ over $\widetilde{\mathcal{Y}}_{it}$. Extending (2.7) to a parameter $\theta \in \Theta$ over the domain $\widetilde{\mathcal{Y}}_{it} \times \mathcal{X}_{it} \times \mathcal{Y}_{it-1} \times \mathcal{X}_{it-1}$ yields

$$\underbrace{f_{\widetilde{Y}_{it},X_{it},Y_{it-1},X_{it-1}}}_{\text{observed from data}} = \int \underbrace{f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta}}_{\text{parametric specification}} f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta} du_{it}. \tag{2.8}$$

Because the observable density function $f_{\widetilde{Y}_{it},X_{it},Y_{it-1},X_{it-1}}$ on the left-hand side and the parametric censored density function $f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta}$ are known up to θ, it is possible to construct a parametric joint density function $f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta}$ using (2.8). Given (x_{it}, y_{it-1}) and a parameter θ, define an integral operator as follows:

$$L_{f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it}};\theta} : L^2(\mathcal{U}_{it}, \omega) \to L^2(\widetilde{\mathcal{Y}}_{it}) \tag{2.9}$$

© 2017 Royal Economic Society.

1368423x, 2018, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ectj.12086 by Johns Hopkins University, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

with

$$(L_{f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it}};\theta}h)(\widetilde{y}_{it}) = \int f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta}(\widetilde{y}_{it}|x_{it}, y_{it-1}, u_{it})h(u_{it})du_{it}.$$

If the integral operator $L_{f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it}};\theta}$ is invertible for each θ, then (2.8) suggests that the joint density function $f_{X_{it},Y_{it-1},X_{it-1},U_{it}}$ can be obtained by

$$f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta} \equiv L^{-1}_{f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it}};\theta}(f_{\widetilde{Y}_{it},X_{it},Y_{it-1},X_{it-1}}).$$

Plugging the true parameter θ_0 into this equation results in

$$f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0} \equiv L^{-1}_{f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it}};\theta_0}(f_{\widetilde{Y}_{it},X_{it},Y_{it-1},X_{it-1}})$$

by (2.7). The joint density function $f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta}$ still achieves the true joint density function at the population parameter θ_0. The concept of completeness provides a sufficient condition for the invertibility of the integral operator using the parametric censored density function as a kernel. The following definition presents this completeness.

DEFINITION 2.3. *A density function $f(y|u)$ satisfies a completeness condition for $L^2(\mathcal{U}, \omega)$ if for $h(u) \in L^2(\mathcal{U}, \omega)$ such that $\int (f(y|u)^2/\omega(u))du < \infty$ and*

$$\int h(u)f(y|u)du = 0 \quad \textit{for all } y, \tag{2.10}$$

then $h(u) = 0$ almost everywhere. In other words, there is no non-zero function in $L^2(\mathcal{U}, \omega)$ with zero integration for each function in the family of the density functions $\{f(y|u) : y \in \mathcal{Y}\}$. By switching the roles of y and u and dropping ω, it is possible to define $\{f(y|u) : u \in \mathcal{U}\}$ as complete in $L^2(\mathcal{Y})$, and this definition can be generalized to function forms such as $f(y, u)$.

ASSUMPTION 2.3. (DEPENDENCE BETWEEN Y_{it} AND U_{it}) *For each $\theta \in \Theta$ and fixed (x_{it}, y_{it-1}), the family of the parametric censored density functions $\{f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta} : \widetilde{y}_{it} \in \widetilde{\mathcal{Y}}_{it}\}$ is complete over $L^2(\mathcal{U}_{it}, \omega)$.*

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

Suppose that

$$(L_{f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it}};\theta}h_1)(\widetilde{y}_{it}) = (L_{f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it}};\theta}h_2)(\widetilde{y}_{it}) \quad \text{for all } \widetilde{y}_{it} \in \widetilde{\mathcal{Y}}_{it}.$$

Then, Assumption 2.3 guarantees that $h_1 = h_2$ almost everywhere. Hence, the operator $L_{f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it}};\theta}$ is invertible for each θ and (x_{it}, y_{it-1}). This assumption requires dependence between Y_{it} and U_{it} because the independence between Y_{it} and U_{it} violates Assumption 2.3. Although Assumption 2.3 ensures the existence of the joint density function, $f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta}$, it might not be identifiable. The variation of the parameter θ in $f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta}$ might be lost in the sense that for any open neighbourhood of θ_0, there exists some $\theta_1 \neq \theta_0$ such that $f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_1} = f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0}$. In other words, $f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta}$ is not locally identifiable.

© 2017 Royal Economic Society.

1368423x, 2018, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ectj.12086 by Johns Hopkins University, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

In this case, applying the inverse transformation is useless because the parameter of interest θ is not distinguished in the new transformed joint density functions, preventing the identification of the parameter. The assumption below prevents this loss.

ASSUMPTION 2.4. *Given each* (x_{it}, y_{it-1}), *suppose* $f_{X_{it},Y_{it-1},X_{it-1}} > 0$. *Assume that: (a) for the dependence between* Y_{it} *and* X_{it-1}, *the family of the observable conditional density functions over* $\mathcal{X}_{it-1}$, $\{f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},X_{it-1}} : x_{it-1} \in \mathcal{X}_{it-1}\}$, *is complete over* $L^2(\widetilde{\mathcal{Y}}_{it})$; *(b) for the dependence between* Y_{it} *and* U_{it}, *the family of the parametric censored density function over* $\mathcal{U}_{it}$, $\{f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_0} : u_{it} \in \mathcal{U}_{it}\}$, *is complete over* $L^2(\widetilde{\mathcal{Y}}_{it})$.

This assumption warrants several comments. First, the conditional density functions in Assumption 2.4(a) contain only observables. Second, Assumption 2.4(a) suggests that X_{it} cannot be constant over time. If X_{it} is constant across time, then $X_{it} = X_{it-1}$ and $f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},X_{it-1}} = f_{\widetilde{Y}_{it}|X_{it},Y_{it-1}}$, which clearly violates the completeness in Assumption 2.4(a). Finally, similar to Assumption 2.3, Assumption 2.4(b) requires that the cardinality of $\widetilde{\mathcal{Y}}_{it}$ is less than the cardinality of $\mathcal{U}_{it}$. Combining the cardinality restrictions in Assumptions 2.3 and 2.4(b) shows that $\mathcal{U}_{it}$ and $\widetilde{\mathcal{Y}}_{it}$ must have the same cardinality. This restriction is compatible when both the dependent variable Y_{it} and the unobserved covariate U_{it} take continuous values.

LEMMA 2.1. (PARAMETRIC IDENTIFICATION OF $f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta}$) *Under Assumptions 2.3 and 2.4, the parametric joint density* $f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta}$ *is correctly specified at* θ_0 *and the parameter* θ_0 *is locally identifiable (i.e. there is an open neighbourhood of* θ_0 *containing no other* θ *such that* $f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta} = f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0}$*).*

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

$$\widetilde{f}_{X_{it},Y_{it-1},X_{it-1};\theta}(x_{it}, y_{it-1}, x_{it-1}) \equiv \int f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta}(x_{it}, y_{it-1}, x_{it-1}, u_{it})du_{it}. \tag{2.11}$$

To identify θ from the integrated parametric density function $\widetilde{f}_{X_{it},Y_{it-1},X_{it-1};\theta}$, it is necessary to examine whether $\widetilde{f}_{X_{it},Y_{it-1},X_{it-1};\theta}$ can be correctly specified at θ_0 and the parameter θ_0 is locally identifiable after applying the integration. This integration step might impose too many restrictions on the parameters and degenerate the variation of the function over its parameter space. Thus, it is necessary to rule out these degenerated cases. The following condition maintains the parametric representation of $\widetilde{f}_{X_{it},Y_{it-1},X_{it-1};\theta}$.

ASSUMPTION 2.5. (VARIATION OF PARAMETERS AROUND θ_0) *The family of the derivative of the parametric censored density functions with respect to* θ, $\{(\partial/\partial\theta) f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_0} : u_{it} \in \mathcal{U}_{it}\}$, *is complete over* $L^2(\widetilde{\mathcal{Y}}_{it})$.

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

LEMMA 2.2. (IDENTIFICATION OF θ_0) *Under Assumptions 2.3–2.5, the integrated parametric joint density* $\widetilde{f}_{X_{it},Y_{it-1},X_{it-1};\theta}$ *is correctly specified at* θ_0 *and the parameter* θ_0 *is locally identifiable.*

At this point in the process, the unobserved component of the parametric censored density function in (2.6) has been transformed out and the parameter θ of the function becomes the parameter of the observable parametric function $\widetilde{f}_{X_{it},Y_{it-1},X_{it-1};\theta}$. However, if $\widetilde{f}_{X_{it},Y_{it-1},X_{it-1};\theta}$ does not integrate to unity (with respect to the measure $dx_{it}dy_{it-1}dx_{it-1}$), it is not a candidate

© 2017 Royal Economic Society.

1368423x, 2018, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ectj.12086 by Johns Hopkins University, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

of the parametric family of PDFs for $f_{X_{it},Y_{it-1},X_{it-1}}$ and the standard MLE cannot be applied to $\widetilde{f}_{X_{it},Y_{it-1},X_{it-1};\theta}$. To obtain a valid semi-parametric family of PDFs, perform the following normalization step

$$f_{X_{it},Y_{it-1},X_{it-1};\theta} \equiv \frac{\widetilde{f}_{X_{it},Y_{it-1},X_{it-1};\theta}}{\int\int\int \widetilde{f}_{X_{it},Y_{it-1},X_{it-1};\theta} dx_{it} dy_{it-1} dx_{it-1}}. \tag{2.12}$$

Similar to the previous discussion, it is necessary to show that the PDF of observables $f_{X_{it},Y_{it-1},X_{it-1};\theta}$ is correctly specified at θ_0 and the parameter θ_0 is locally identifiable after this normalization. The following assumption and lemma demonstrate the existence of a non-trivial θ_0 after normalization.

ASSUMPTION 2.6. (DEPENDENCE BETWEEN Y_{it} AND X_{it-1}) *Assume that the family of the observable conditional density functions* $\{(\partial/\partial x_{it-1}) f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},X_{it-1}} : x_{it-1} \in \mathcal{X}_{it-1}\}$ *is complete over* $L^2(\widetilde{\mathcal{Y}}_{it})$ *for each* x_{it}, y_{it-1}.

If $\mathcal{X}_{it-1}$ is discrete, then the derivative can be replaced with the difference. Notice that both Assumptions 2.4(a) and 2.6 are related to the observable conditional distribution $f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},X_{it-1}}$ and Assumption 2.6 implies Assumption 2.4(a). Hence, the two assumptions are compatible and it is only necessary to verify the completeness in Assumption 2.6. The assumption rules out the possible separable cases, such as $f(\widetilde{Y}_{it}|X_{it}, Y_{it-1}, X_{it-1}) = h_1(\widetilde{Y}_{it}, X_{it}, Y_{it-1})h_2(X_{it-1})$ or $f(\widetilde{Y}_{it}|X_{it}, Y_{it-1}, X_{it-1}) = h_1(\widetilde{Y}_{it})h_2(X_{it}, Y_{it-1}, X_{it-1})$.

LEMMA 2.3. (LOCAL IDENTIFICATION) *Under Assumptions 2.3, 2.4(b), 2.5 and 2.6, the PDF of observables after normalization,* $f_{X_{it},Y_{it-1},X_{it-1};\theta}$, *is correctly specified at* θ_0 *and the parameter* θ_0 *is locally identifiable.*

The transformation includes applying the inverse of an integral operator using the function $f_{Y_{it}|X_{it},Y_{it-1},U_{it};\theta}$ as a kernel, integrating out the unobserved covariate, and normalization. After these three steps of transformation associated with the completeness of PDFs, the parametric PDFs of observables $\{f_{X_{it},Y_{it-1},X_{it-1};\theta} : \theta \in \Theta\}$ is correctly specified at θ_0 and the parameter θ_0 is locally identifiable under Assumptions 2.3–2.6. To distinguish the parameters of interest θ_0 from the parameter space Θ on the basis of sample information, use the Kullback–Leibler information criterion

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

where the expectation is taken with respect to $f_{X_{it},Y_{it-1},X_{it-1}}$. The fact that θ_0 is globally identified in Θ is related to the zero set of $K(\theta)$. The following result is a direct application of the results in Bowden (1973), which is the standard framework of the identifiability criterion of maximum likelihood estimation; we omit its proof.

THEOREM 2.1. (GLOBAL IDENTIFICATION) *Suppose that* $K(\theta) = 0$ *has a unique solution at* $\theta = \theta_0$ *in* Θ. *Under Assumptions 2.3, 2.4(b), 2.5 and 2.6, the parametric censored density function* $f_{Y_{it}|X_{it},Y_{it-1},U_{it};\theta}$ *and the joint density function* $f_{X_{it},Y_{it-1},X_{it-1},U_{it}}$ *can then be identified from the joint distribution of two-period observations* $\{Y_{it}, X_{it}, Y_{it-1}, X_{it-1}\}$ *for* $i = 1, 2, \ldots, N$.

In addition to the Kullback–Leibler information of classical statistics, the identification result is based on the completeness of the families of known PDFs $f_{Y_{it}|X_{it},Y_{it-1},U_{it};\theta}$ corresponding to censored dynamic panel data models of short time periods and observed conditional density

© 2017 Royal Economic Society.

1368423x, 2018, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ectj.12086 by Johns Hopkins University, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

functions of the dependent and explanatory variables $f_{Y_{it}|X_{it},Y_{it-1},X_{it-1}}$. In the next section, we show that the completeness condition holds for some popular censored dynamic panel data models.

Theorem 2.1 provides the identification of the parameter θ_0. However, because U_{it} does not have meaningful units of measurement, it is not apparent what values of U_{it} we should use. In nonlinear models, estimating the average partial effects of explanatory variables is more attractive than estimating parameters. Thus, this study introduces the average structure function by averaging a scalar function of y_{it}, $\omega(y_{it})$, across the distribution of U_{it} in the population. Let (X_{it}, Y_{it-1}) be a given value of the explanatory variables, whose average structure function is

$$\begin{aligned}\mu(X_{it}, Y_{it-1}) &\equiv E_{U_{it}}[E_{Y_{it}}[\omega(Y_{it})|X_{it}, Y_{it-1}, U_{it}]]\\ &= \int_{U_{it}} \left(\int_{Y_{it}} \omega(Y_{it}) f_{Y_{it}|X_{it},Y_{it-1},U_{it}} dy_{it} \right) f_{U_{it}} du_{it}. \end{aligned} \tag{2.13}$$

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

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

Combining the identification results of $f_{Y_{it}|X_{it},Y_{it-1},U_{it}}$ and $f_{U_{it}}$ provides the identification of the average structure function $\mu(X_{it}, Y_{it-1})$. This indicates that the average partial effect is also identified because the average partial effect can be defined by taking derivatives or differences of the average structure function in (2.13) with respect to elements of (X_{it}, Y_{it-1}). This yields the identification of the average partial effect.

COROLLARY 2.1. *Under Assumptions 2.3, 2.4(b), 2.5 and 2.6, the average partial effect defined as derivatives or differences of (2.13) is identified from the joint distribution of two-period observations,* $\{Y_{it}, X_{it}, Y_{it-1}, X_{it-1}\}$ *for* $i = 1, 2, \ldots, N$.

In a parametric likelihood case, the local identifiability of unknown parameters is equivalent to the non-singularity of the information matrix under weak regularity conditions. If the true parameter θ_0 is a critical point of $K(\theta_0)$, then a sufficient condition of the uniqueness of θ_0 is $K''(\theta_0)$ is negative semi-definite, where $K''(\theta_0) = [K_{lm}]_{l,m}$ with

$$K_{lm} = -E\left[\left(\frac{(\partial/\partial\theta_l) f_{X_{it},Y_{it-1},X_{it-1};\theta}|_{\theta=\theta_0} (\partial/\partial\theta_m) f_{X_{it},Y_{it-1},X_{it-1};\theta}|_{\theta=\theta_0}}{f^2_{X_{it},Y_{it-1},X_{it-1}}}\right)\right], \tag{2.14}$$

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

COROLLARY 2.2. *Suppose that in an open neighbourhood of* θ_0 *in* Θ*, the second derivative of the Kullback–Leibler function* $K(\theta)$ *with an element in (2.14) is negative definite. Under Assumptions 2.3, 2.4(b), 2.5 and 2.6, the parametric censored density function* $f_{Y_{it}|X_{it},Y_{it-1},U_{it};\theta}$ *and the joint density function* $f_{X_{it},Y_{it-1},X_{it-1},U_{it}}$ *can be identified from the joint distribution of two-period observations* $\{Y_{it}, X_{it}, Y_{it-1}, X_{it-1}\}$ *for* $i = 1, 2, \ldots, N$.

© 2017 Royal Economic Society.

1368423x, 2018, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ectj.12086 by Johns Hopkins University, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3. MOTIVATING EXAMPLES

Consider the two examples presented at the beginning of Section 2. Here, we show that the completeness conditions in Section 2 hold in these cases. Assumptions 2.3, 2.4(b) and 2.5 are related to the completeness of the variant forms of the parametric censored density function $f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta}$. Equations (2.4) and (2.5) show that the completeness of the parametric censored density functions over positive Y_{it} in the two motivating examples are connected to the PDF of the random shock ξ_{it}. Therefore, in this section, we focus on what types of parametric distribution assumptions in ξ_{it} make the completeness assumptions hold. For simplicity, assume the domains of ξ_{it} and U_{it} are the whole real line $\mathbb{R}$.

Most of the interesting leading cases for models in (2.4) and (2.5) occur when the random shock ξ_{it} is assumed to have an independent Gaussian white noise process. For simplicity, we assume $g_2(X_{it}, Y_{it-1}; \theta_2) = X'_{it}\beta + \gamma Y_{it-1}$. In this case, the parametric censored density function $f_{Y_{it}|X_{it},Y_{it-1},U_{it}}$ is fully parameterized and correctly specified at θ_0. The specifications of the models under the normality assumption are as follows.

Semi-parametric dynamic tobit models. Assuming that $\xi_{it} \sim N(0, \sigma_\xi^2)$, (2.4) leads to

$$f_{Y_{it}|X_{it},Y_{it-1},U_{it};\theta} = \Big[1 - \Phi\Big(\frac{X'_{it}\beta + \gamma Y_{it-1} + U_{it}}{\sigma_\xi}\Big)\Big]^{\mathbf{1}(Y_{it}=0)} \times\Big[\frac{1}{\sigma_\xi}\phi\Big(\frac{Y_{it} - X'_{it}\beta - \gamma Y_{it-1} - U_{it}}{\sigma_\xi}\Big)\Big]^{\mathbf{1}(Y_{it}>0)}, \tag{3.1}$$

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

Semi-parametric two-part dynamic regression models. Let $g_3(X_{it}, Y_{it-1}; \theta_1) = X'_{it}\beta_d + \gamma_d Y_{it-1}$. Suppose that $\varsigma_{it} \sim N(0, 1)$ and $\xi_{it} \sim N(0, \sigma_\xi^2)$. Equation (2.5) then becomes

$$f_{Y_{it}|X_{it},Y_{it-1},U_{it};\theta} = (1 - \Phi(X'_{it}\beta_d + \gamma_d Y_{it-1} + U_{it}))^{\mathbf{1}(Y_{it}=0)} \Big\{\Phi(X'_{it}\beta_d + \gamma_d Y_{it-1} + U_{it}) \times\phi\Big(\frac{\log(Y_{it}) - X'_{it}\beta - \gamma Y_{it-1} - U_{it}}{\sigma_\xi}\Big)\frac{1}{\sigma_\xi Y_{it}}\Big\}^{\mathbf{1}(Y_{it}>0)}, \tag{3.2}$$

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

The normality assumption makes it possible to verify the completeness of the parametric censored density function $f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta}$ in Assumptions 2.3, 2.4(b) and 2.5 directly. It is then necessary to show that the parametric censored density functions (3.1) and (3.2) satisfy these completeness assumptions. To do this, we introduce the completeness of normal distributions and exponential families in L^2 from Hu and Shiu (2017), which are variants of the results of Newey and Powell (2003). Denote $\mathcal{X}$, and $\mathcal{Z}$ as the support of random variables x and z, respectively.

LEMMA 3.1. *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 $E[h(x)|z = z_1] = 0$ for any $z_1 \in \mathcal{Z}$ and some $h \in L^2(\mathcal{X})$ implies $h(\cdot) = 0$ almost everywhere in $\mathcal{X}$; equivalently, $\{f(\cdot|z) : z \in \mathcal{Z}\}$ is complete in $L^2(\mathcal{X})$.*

LEMMA 3.2. *Let $f(x|z) = s(x)t(z)\exp(\mu(z)\tau(x))$, where $s(x) > 0$, the mapping from $x \to \tau(x)$ is one-to-one in x, and support of $\mu(z)$ contains an open set, then $E[h(x)|z = z_1] = 0$ for any $z_1 \in \mathcal{Z}$ and some $h \in L^2(\mathcal{X})$ implies $h(\cdot) = 0$ almost everywhere in $\mathcal{X}$; equivalently, the family of conditional density functions $\{f(\cdot|z) : z \in \mathcal{Z}\}$ is complete in $L^2(\mathcal{X})$.*

© 2017 Royal Economic Society.

1368423x, 2018, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ectj.12086 by Johns Hopkins University, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

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

The equation can be rewritten as

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

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

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

It follows that

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

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

The following lemma provides the completeness of the families of variant of the normal PDF ϕ, which is related to the two motivating examples.

LEMMA 3.3. *Suppose the domain $\mathcal{U}$ contains an open set. For a constant c, the families of functions $\{(y-c-u)\phi((y-c-u)/\sigma_\xi) : u \in \mathcal{U}\}$ and $\{(\sigma_\xi^2-(y-c-u)^2)\phi((y-c-u)/\sigma_\xi) : u \in \mathcal{U}\}$ are complete in $L^2(\widetilde{\mathcal{Y}})$.*

Based on the information about the completeness of normal distributions, it is possible to investigate the completeness condition of models in (3.1) and (3.2). Set $\widetilde{\mathcal{Y}}_{it} = \mathbb{R}^+$. Given $\theta \in \Theta$, and (x_{it}, y_{it-1}), the proposed parametric censored density functions over $\widetilde{\mathcal{Y}}_{it}$ in the motivating examples can be written as the following.

Semi-parametric dynamic tobit models.

$$f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta} = \frac{1}{\sigma_\xi}\phi\Big(\frac{\widetilde{Y}_{it}-X'_{it}\beta-\gamma Y_{it-1}-U_{it}}{\sigma_\xi}\Big). \tag{3.3}$$

Semi-parametric two-part dynamic regression models.

$$f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta} = \Phi(X'_{it}\beta_d+\gamma_d Y_{it-1}+U_{it}) \times\phi\Big(\frac{\log(\widetilde{Y}_{it})-X'_{it}\beta-\gamma Y_{it-1}-U_{it}}{\sigma_\xi}\Big)\frac{1}{\sigma_\xi \widetilde{Y}_{it}}. \tag{3.4}$$

[1] We have used the equation $\phi((u-(a+by))/\sigma) = e^{u^2/4\sigma^2}\phi((u-2(a+by))/\sqrt{2}\sigma)e^{(1/2\sigma^2)(a+by)^2}$.

© 2017 Royal Economic Society.

1368423x, 2018, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ectj.12086 by Johns Hopkins University, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

1368423x, 2018, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ectj.12086 by Johns Hopkins University, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The completeness conditions in Assumptions 2.3, 2.4(b) and 2.5 are associated with the dependent variables $\widetilde{Y}_{it}$ and the unobserved covariate U_{it}. Therefore, it is necessary to investigate which functional forms connect these two variables. In these models, the dependent variables $\widetilde{Y}_{it}$ and the unobserved covariate U_{it} are both inside the standard normal PDF ϕ. We can directly apply the result of Lemma 3.1 to (3.3) to show that semi-parametric dynamic tobit models satisfies Assumptions 2.3 and 2.4(b). Because the standard normal CDF Φ is positive, the semi-parametric two-part dynamic regression models also fulfil Assumptions 2.3 and 2.4(b) using Lemma 3.1 and a change of variable.

Assumption 2.5 requires that the partial derivatives of the parametric censor density function with respect to all components of the parameter θ be complete. According to the functional forms in (3.3) and (3.4) and the use of a change of variable, two types of the partial derivatives of $f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_0}$ should be considered: the first is the partial derivative with respect to the components of β and γ, and the second is σ_ξ. The partial derivatives in the dynamic tobit models are

$$\frac{\partial f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta}}{\partial \beta} = \frac{X_{it}}{\sigma_\xi^2}\Big(\frac{\widetilde{Y}_{it} - X_{it}'\beta - \gamma Y_{it-1} - U_{it}}{\sigma_\xi}\Big) \times \phi\Big(\frac{\widetilde{Y}_{it} - X_{it}'\beta - \gamma Y_{it-1} - U_{it}}{\sigma_\xi}\Big),$$

$$\frac{\partial f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta}}{\partial \gamma} = \frac{Y_{it-1}}{\sigma_\xi^2}\Big(\frac{\widetilde{Y}_{it} - X_{it}'\beta - \gamma Y_{it-1} - U_{it}}{\sigma_\xi}\Big) \times \phi\Big(\frac{\widetilde{Y}_{it} - X_{it}'\beta - \gamma Y_{it-1} - U_{it}}{\sigma_\xi}\Big),$$

and

$$\frac{\partial f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta}}{\partial \sigma_\xi} = \frac{-(\sigma_\xi^2 - (\widetilde{Y}_{it} - X_{it}'\beta - \gamma Y_{it-1} - U_{it})^2)}{\sigma_\xi^4} \times \phi\Big(\frac{\widetilde{Y}_{it} - X_{it}'\beta - \gamma Y_{it-1} - U_{it}}{\sigma_\xi}\Big).$$

This implies that the completeness of the first type of the partial derivatives can be reduced to the completeness of the family of $\{(y - c - u)\phi(y - c - u/\sigma_\xi) : u \in \mathcal{U}\}$ in $L^2(\widetilde{\mathcal{Y}})$ for some constant c. Similarly, the completeness of the second type of the partial derivatives depends on the family of $\{(\sigma_\xi^2 - (y - c - u)^2)\phi(y - c - u/\sigma_\xi) : u \in \mathcal{U}\}$ in $L^2(\widetilde{\mathcal{Y}})$ for some constant c. By the completeness results in Lemma 3.3, the semi-parametric dynamic tobit models satisfies Assumption 2.5. The semi-parametric two-part dynamic regression models can be handled in a similar manner and the results are summarized as follows.

PROPOSITION 3.1. *The two motivating models, the semi-parametric dynamic tobit models and the semi-parametric two-part dynamic regression models in (3.1) and (3.2), satisfy Assumptions 2.3, 2.4(b) and 2.5.*

This discussion also applies to models with heteroscedasticity, which allow a more general functional form in corner solution models. If ξ_{it} has a heteroscedastic normal distribution such

© 2017 Royal Economic Society.

that $\xi_{it} \sim N(0, h(X_{it}, Y_{it-1}; \sigma_\xi))$, then the parametric censored density functions in (3.1) and (3.2) become, respectively,

$$f_{Y_{it}|X_{it},Y_{it-1},U_{it};\theta} = \Big(1 - \Phi\Big(\frac{X_{it}'\beta + \gamma Y_{it-1} + U_{it}}{h(X_{it}, Y_{it-1}; \sigma_\xi)}\Big)\Big)^{\mathbf{1}(Y_{it}=0)} \times\Big(\frac{1}{h(X_{it}, Y_{it-1}; \sigma_\xi)}\phi\Big(\frac{Y_{it} - X_{it}'\beta - \gamma Y_{it-1} - U_{it}}{h(X_{it}, Y_{it-1}; \sigma_\xi)}\Big)\Big)^{\mathbf{1}(Y_{it}>0)} \quad (3.5)$$

and

$$f_{Y_{it}|X_{it},Y_{it-1},U_{it};\theta} = (1 - \Phi(X_{it}'\beta_d + \gamma_d Y_{it-1} + U_{it}))^{\mathbf{1}(Y_{it}=0)} \times\Big(\Phi(X_{it}'\beta_d + \gamma_d Y_{it-1} + U_{it})\phi\Big(\frac{\log(Y_{it}) - X_{it}'\beta - \gamma Y_{it-1} - U_{it}}{h(X_{it}, Y_{it-1}; \sigma_\xi)}\Big) \times\frac{1}{h(X_{it}, Y_{it-1}; \sigma_\xi)Y_{it}}\Big)^{\mathbf{1}(Y_{it}>0)}. \quad (3.6)$$

Adding the heterogeneous structure does not affect the functional form, which dominates the relationship between the dependent variables $\widetilde{Y}_{it}$ and the unobserved covariate U_{it}. The derivations in homoscedastic cases can be extended to heteroscedastic cases in a similar fashion.

The assumptions that are not related to the parametric censored density functions include Assumptions 2.4(a) and 2.6. Recall that Assumption 2.6 implies Assumption 2.4(a). These assumptions require functional form restrictions on the conditional density function $f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},X_{it-1}}$ of observables. With the well-known completeness from the normal distributions in Lemma 3.1, we can construct $f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},X_{it-1}}$ satisfying Assumptions 2.4(a) and 2.6 for the dynamic tobit models and two-part dynamic regression models.

Consider the conditional version of (2.8),

$$f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},X_{it-1}} = \int f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta} f_{U_{it}|X_{it},Y_{it-1},X_{it-1};\theta} du_{it}. \quad (3.7)$$

In the dynamic tobit models, for $f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta}$ in (3.3), we use a parametric normal assumption for

$$f_{U_{it}|X_{it},Y_{it-1},X_{it-1};\theta} = \frac{1}{\sigma_\xi}\phi\Big(\frac{U_{it} - X_{it}'\beta - \gamma Y_{it-1} + \psi_1(X_{it}, Y_{it-1}; \theta_1) + \beta\psi_2(X_{it-1})}{\sigma_\xi}\Big).$$

Then, given a fixed (X_{it}, Y_{it-1}), and $\mathcal{X}_{it-1}$ containing an open set, we obtain[2]

$$f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},X_{it-1}} = \frac{1}{\sqrt{2}\sigma_\xi}\phi\Big(\frac{\widetilde{Y}_{it} - \psi_1(X_{it}, Y_{it-1}; \theta_1) - \beta\psi_2(X_{it-1})}{\sqrt{2}\sigma_\xi}\Big). \quad (3.8)$$

[2] Set $A = \widetilde{Y}_{it} - X_{it}'\beta - \gamma Y_{it-1}$, $B = \widetilde{Y}_{it} - \psi_1(X_{it}, Y_{it-1}; \theta_1) - \beta\psi_2(X_{it-1})$ and $M = B - A$. The result follows directly from $e^{-((A-U_{it})^2/2\sigma_\xi^2)}e^{-((U_{it}+M)^2/2\sigma_\xi^2)} = e^{-((U_{it}-(A-M)/2)^2/\sigma_\xi^2)}e^{-((A+M)^2/4\sigma_\xi^2)}$.

© 2017 Royal Economic Society.

1368423x, 2018, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ectj.12086 by Johns Hopkins University, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

This implies

$$\frac{\partial}{\partial X_{it-1}} f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},X_{it-1}} = \frac{1}{2\sigma_\xi^2}\beta\psi_2'(X_{it-1})(\widetilde{Y}_{it} - \psi_1(X_{it}, Y_{it-1};\theta_1) - \beta\psi_2(X_{it-1})) \times\phi\Big(\frac{\widetilde{Y}_{it} - \psi_1(X_{it}, Y_{it-1};\theta_1) - \beta\psi_2(X_{it-1})}{\sqrt{2}\sigma_\xi}\Big).$$

A sufficient condition to satisfy Assumption 2.6 for this specification of $f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},X_{it-1}}$ is $\beta\psi_2'(X_{it-1}) \neq 0$ and the range of ψ_2 contains an open set, according to Lemma 3.3. As for the two-part dynamic regression models, we can use the similar parametric normal assumption as

$$f_{U_{it}|X_{it},Y_{it-1},X_{it-1};\theta} = c(X_{it}, Y_{it-1}, X_{it-1})\Phi(X_{it}'\beta_d + \gamma_d Y_{it-1} + U_{it})\frac{1}{\sigma_\xi} \times\phi\Big(\frac{U_{it} - X_{it}'\beta - \gamma Y_{it-1} + \psi_1(X_{it}, Y_{it-1};\theta_1) + \beta\psi_2(X_{it-1})}{\sigma_\xi}\Big),$$

where $c(\cdot)$ is a density normalization coefficient. This conditional density function also fulfils Assumption 2.6.

For a model with a larger number of observed covariates, suppose $X_{it-1} = (X_{i1t-1}, X_{i2t-1})$, where X_{i1t-1} and X_{i2t-1} are $k_1 \times 1$ and $k_2 \times 1$ vectors of explanatory variables and X_{i2t-1} takes continuous values. Consider

$$f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},X_{it-1}} = \frac{1}{\sigma_\xi} \times\phi\Big(\frac{\widetilde{Y}_{it} - \psi_3(X_{it}, Y_{it-1}, X_{i1t-1};\theta_1) - \psi_4(X_{it}, Y_{it-1}, X_{i1t-1};\beta)X_{i2t-1}}{\sigma_\xi}\Big). \tag{3.9}$$

In this case, we only use the dependence between $\widetilde{Y}_{it}$ and X_{i2t-1} and the completeness in Assumption 2.6 becomes that the family

$$\Big\{\frac{\partial}{\partial x_{it-1}} f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},X_{it-1}} : x_{i2t-1} \in \mathcal{X}_{i2t-1}\Big\}$$

is complete over $L^2(\widetilde{\mathcal{Y}}_{it})$ for each $x_{it}, y_{it-1}, x_{i1t-1}$, where $\mathcal{X}_{i2t-1}$ is the support of x_{i2t-1}. A sufficient condition for Assumption 2.6 is $Pr(\text{rank}[\psi_4(X_{it}, Y_{it-1}, X_{i1t-1};\beta)] = 1) = 1$ and $\mathcal{X}_{i2t-1}$ contains an open set.

The examples in this section rely on the normality of the random shock ξ_{it} and it is possible to relax the normality assumption. However, in limited dependent variable models, the key issue is comparing estimated average partial effects across different models rather than parameter estimates. These models are likely to do an appropriate job of providing average partial effects under more general settings because these models require observing two periods of data without the initial condition for the dependent variable.

4. SEMI-PARAMETRIC ESTIMATION

The parameter in (2.6) is identified in Theorem 2.1 and it can be determined using (2.8). Optimizing certain empirical criteria in general parameter spaces produces a sieve MLE. The

© 2017 Royal Economic Society.

1368423x, 2018, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ectj.12086 by Johns Hopkins University, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

integral (2.8) suggests a corresponding sieve MLE:

$$(\hat{\theta}, \hat{f}_1)^T = \arg \max_{(\theta, f_1)^T \in \mathcal{A}_n} \frac{1}{N} \sum_{i=1}^{N} \ln \int f_{Y_{it}|X_{it},Y_{it-1},U_{it};\theta}(y_{it}|x_{it}, y_{it-1}, u_{it}) \times f_1(x_{it}, y_{it-1}, x_{it-1}, u_{it}; \theta) du_{it}, \tag{4.1}$$

using a two-period independent and identically distributed (i.i.d.) sample $\{y_{it}, x_{it}, y_{it-1}, x_{it-1}\}_{i=1}^{N}$. A general review of the semi-parametric sieve MLE appears in Shen (1997), Chen and Shen (1998) and Ai and Chen (2003). The space $\mathcal{A}_n$ is a sequence of approximating sieve spaces containing sieve approximations of the parameter because maximization over the whole parameter space $\mathcal{A}$ is undesirable. In addition, θ is a finite-dimensional parameter of interest and f_1 is a potentially infinite-dimensional nuisance parameter or non-parametric component that varies with θ. The following subsection provides a detailed implementation of sieve approximations of the non-parametric component f_1.

4.1. *Restrictions on sieve coefficients*

As for a non-parametric series estimator of $f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta}$, constructing a sieve approximating series that varies with the model parameter θ is an essential issue for the proposed sieve MLE. Denote δ_1 as a vector of sieve coefficients and $f_{X_{it},Y_{it-1},X_{it-1},U_{it};\delta_1}$ as the sieve approximation function. The function $f_{X_{it},Y_{it-1},X_{it-1},U_{it};\delta_1}$ in dynamic censored models with a lagged dependent variable consists of two different parts, $Y_{it-1} = 0$ and $Y_{it-1} > 0$, and these parts can be built according to their numerical structures. Set $f_{Y_{it-1}=0} = \text{Prob}(Y_{it-1} = 0)$. A way to split these two parts is

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

The corresponding density restrictions are

$$\int f_{X_{it},X_{it-1},U_{it}|Y_{it-1}=0} dx_{it} dx_{it-1} du_{it} = 1$$

and

$$f_{Y_{it-1}=0} + \int f_{X_{it},Y_{it-1}>0,X_{it-1},U_{it}} dy_{it-1} dx_{it} dx_{it-1} du_{it} = 1.$$

Set

$$z_{1,\sigma_\xi} \equiv \frac{x_{it}'\beta - x_{it-1}'\beta - u_{it}}{\sigma_\xi} \qquad \text{and} \qquad z_{2,\sigma_\xi} \equiv \frac{x_{it}'\beta - u_{it}}{\sigma_\xi}.$$

For the $Y_{it-1} = 0$ part, consider

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

© 2017 Royal Economic Society.

1368423x, 2018, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ectj.12086 by Johns Hopkins University, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

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

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

$$q_0(u_{it}) = \frac{1}{\sqrt{l_3}}, \, q_k(u_{it}) = \sqrt{\frac{2}{l_3}} \cos\left(\frac{k\pi}{l_3} u_{it}\right).$$

However, suppose that $y_{it-1} \in (0, l_4]$. Write

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

where

$$z'_{1,\sigma_\xi} \equiv \frac{x'_{it}\beta - \gamma y_{it-1} - x'_{it-1}\beta - u_{it}}{\sigma_\xi},$$

$$z'_{2,\sigma_\xi} \equiv \frac{x'_{it}\beta - \gamma y_{it-1} - u_{it}}{\sigma_\xi}$$

$$q_0(z_4) = \frac{1}{\sqrt{l_4}}$$

$$q_l(z_4) = \sqrt{\frac{2}{l_4}} \cos\left(\frac{l\pi}{l_4} z_4\right).$$

The density restrictions for these sieve coefficients are

$$\sum_{i=0}^{i_n} \sum_{j=0}^{j_n} \sum_{k=0}^{k_n} (\hat{a}_{ijk})^2 = 1 \qquad \text{and} \qquad f_{Y_{it-1}=0} + \sum_{i=0}^{i_n} \sum_{j=0}^{j_n} \sum_{k=0}^{k_n} \sum_{l=0}^{l_n} (\tilde{a}_{ijkl})^2 = 1. \tag{4.2}$$

4.2. Estimating average partial effects

Denote $f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}; \hat{\theta}, \delta_1}$ as the sieve MLE of the joint distribution $f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}}$ in the dynamic censored model, where $\hat{\theta}$ is the estimated finite dimensional parameter of the proposed sieve MLE. This parameter can be used to obtain the sieve approximations of the marginal distribution of the unobserved covariate U_{it}:

$$\hat{f}_{U_{it}} = \int_{\mathcal{X}_{it}} \int_{\mathcal{Y}_{it-1}} \int_{\mathcal{X}_{it-1}} f_{X_{it}, Y_{it-1}, X_{it-1}, U_{it}; \hat{\theta}, \delta_1} dx_{it} dy_{it-1} dx_{it-1}. \tag{4.3}$$

Therefore, under the assumptions made in Theorem 2.1, it is possible to consistently estimate average partial effects at interesting values of the explanatory variables. For example, the average

© 2017 Royal Economic Society.

1368423x, 2018, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ectj.12086 by Johns Hopkins University, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

structural functions in the dynamic tobit models are based on

$$\hat{\mu}(X_{it}, Y_{it-1}) \equiv \int_{U_{it}} \Big(\int_{Y_{it}} \max(0, Y_{it}) f_{Y_{it}|X_{it}, Y_{it-1}, U_{it};\hat{\theta}} dY_{it} \Big) \hat{f}_{1,U_{it}} dU_{it}$$
$$= \int_{U_{it}} \Big(\Phi\Big(\frac{X_{it}'\hat{\beta} + \hat{\gamma} Y_{it-1} + U_{it}}{\hat{\sigma}_{\xi}} \Big) (X_{it}'\hat{\beta} + \hat{\gamma} Y_{it-1} + U_{it})$$
$$+ \hat{\sigma}_{\xi} \phi\Big(\frac{X_{it}'\hat{\beta} + \hat{\gamma} Y_{it-1} + U_{it}}{\hat{\sigma}_{\xi}} \Big) \Big) \hat{f}_{1,U_{it}} dU_{it}. \tag{4.4}$$

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

$$\hat{\mu}(X_{it}, Y_1) - \hat{\mu}(X_{it}, Y_0 = 0). \tag{4.5}$$

However, the average partial effect of a continuous explanatory variable can be defined using derivatives of the average structural functions in (4.4).

5. EMPIRICAL APPLICATION

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

The empirical analysis in this study is based on the MEPS Panel 4. The MEPS data provide nationally representative information on health-care use, expenditures, sources of payment and insurance coverage for the US population from 1999 to 2000. The MEPS, which contains detailed data on annual total health-care expenditure, demographic characteristics, health conditions, health status, use of medical care services and income, is appropriate for our empirical application. Table 1 presents summary statistics of health insurance variables, socio-economic variables and health status regressors for the first and second years of the data. We have two periods of data with 7,669 cross-sectional observations. There are sizable fractions of the sample with zero medical expenditure: 18.646% (1,430/7,669) and 20.576% (1,578/7,669) in periods 1 and 2, respectively.

The estimated equation of dynamic health expenditures is

$$Y_{it} = \max\{0, X_{it}'\beta + \gamma Y_{it-1} + V_i + \underbrace{\eta_{it} + \xi_{it}}_{\varepsilon_{it}}\} \qquad \forall i = 1, \ldots, N;\ t = 1, 2,$$
$$= \max\{0, X_{it}'\beta + \gamma Y_{it-1} + U_{it} + \xi_{it}\}. \tag{5.1}$$

© 2017 Royal Economic Society.

1368423x, 2018, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ectj.12086 by Johns Hopkins University, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 1. Sample statistics.

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

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

The dependent variable $Y_{it} = Lnexp_{it}$ is the natural logarithm of medical expenditure plus one. The vector of covariates $X_{it} = (Lninc_{it}, Lnfam_{it}, Age_{it}, Male_{it}, Black_{it}, Education_{it}, Physical_{it}, Ndental_{it}, Good_{it}, Fair_{it}, Poor_{it}, \ldots, \textit{Time dummies})$. The unobserved heterogeneity V_i represents time-invariant individual heterogeneity factors, such as inherent ability or personal regimen to resist negative health shock. Assume that Assumptions 2.1 and 2.2 split the transitory error term ε_{it} into η_{it} and ξ_{it}, and that ξ_{it} is normally distributed. This normality assumption guarantees that Assumptions 2.3, 2.4(b), 2.5 are fulfilled, as Section 3 shows. Assumptions 2.4(a) and 2.5 demand the completeness conditions related to the family of conditional distribution of positive health expenditure $\widetilde{y}_{it}$ over x_{it-1}, $\{f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},X_{it-1}} : x_{it-1} \in \mathcal{X}_{it-1}\}$. These assumptions, along with the mild regularity condition

© 2017 Royal Economic Society.

1368423x, 2018, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ectj.12086 by Johns Hopkins University, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 2. Panel censored estimates for health expenditure.

	Linear fixed effects	RE tobit		Semi-parametric dynamic tobit	
	Coefficient (1)	Coefficient (2)	APE (3)	Coefficient (4)	APE (5)
$Lnexp_{it-1}$	–	–	–	1.052***	1.448***
	–	–	–	(0.001)	(0.006)
Lninc	0.031***	0.042***	0.039***	0.041***	0.056***
	(0.008)	(0.011)	(0.001)	(0.001)	(0.001)
Lnfam	−0.252***	−0.299***	−0.276***	−0.301***	−0.414***
	(0.045)	(0.056)	(0.003)	(0.001)	(0.002)
Age	0.040***	0.048***	0.044***	0.050***	0.068***
	(0.002)	(0.003)	(0.001)	(0.001)	(0.001)
Male	−1.130***	−1.399***	−1.294***	−1.399***	−1.927***
	(0.049)	(0.062)	(0.032)	(0.001)	(0.008)
Black	−0.581***	−0.717***	−0.653***	−0.717***	−0.987***
	(0.069)	(0.086)	(0.012)	(0.001)	(0.004)
Education	0.145***	0.184***	0.170***	0.181***	0.250***
	(0.008)	(0.011)	(0.002)	(0.001)	(0.001)
Physical	0.806***	0.854***	0.788***	0.855***	1.177***
	(0.098)	(0.119)	(0.007)	(0.001)	(0.005)
Ndental	0.442***	0.496***	0.458***	0.502***	0.691***
	(0.012)	(0.015)	(0.004)	(0.001)	(0.003)
Good	0.342***	0.391***	0.362***	0.392***	0.539***
	(0.047)	(0.059)	(0.008)	(0.001)	(0.002)
Fair	1.037***	1.180***	1.111***	1.178***	1.621***
	(0.080)	(0.098)	(0.031)	(0.001)	(0.007)
Poor	1.777***	1.956***	1.865***	1.957***	2.694***
	(0.142)	(0.173)	(0.054)	(0.001)	(0.011)
Deduction	0.384***	0.432***	0.402***	0.429***	0.590***
	(0.089)	(0.108)	(0.010)	(0.001)	(0.003)
Medicare	0.900***	0.995***	0.936***	0.995***	1.370***
	(0.143)	(0.175)	(0.035)	(0.001)	(0.006)
Medicaid	1.138***	1.346***	1.270***	1.343***	1.849***
	(0.092)	(0.114)	(0.001)	(0.001)	(0.008)

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

stated in Theorem 2.1, provide the identification of model (5.1) and the sieve MLE developed in Section 4 is applicable.

Table 2 shows the results of the estimation of panel data model (5.1) using three specifications, including a static linear fixed effect model (column 1), a static tobit model with random effect (RE; columns 2 and 3) and a semi-parametric dynamic tobit model (columns 4 and 5). The three sets of estimates present similar results in terms of directions of effects and estimated coefficients. As expected, there are differences in the magnitudes of the estimated average partial effects in the RE tobit and semi-parametric dynamic tobit specifications. The

© 2017 Royal Economic Society.

1368423x, 2018, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ectj.12086 by Johns Hopkins University, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

average partial effects of the semi-parametric dynamic tobit specification have greater effects after controlling for the dynamic effect of health expenditures. The coefficient estimate of state dependence effect of health expenditures is up to 1.052. As a result, the effect of previous health expenditures on the future health expenditures is estimated to be, in the average partial effect, 1.448. The estimated coefficient shows that the previous health expenditures have persistent effects or there is large first-order state dependence of health expenditures. One of the variables of interest is $Lninc_{it}$, the natural logarithm of the family income plus one. The coefficient of $Lninc_{it}$ in regression on $Lnexp_{it}$ represents the income elasticity of demand for health care. The result of the semi-parametric tobit specification indicates that individuals consume more health care when their incomes increase, after controlling for past health expenditures.

6. CONCLUSION

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

ACKNOWLEDGEMENTS

We thank the participants of the 2013 Annual Meeting of American Economic Association, and the 2013 China Meeting of the Econometric Society for various suggestions regarding this paper. The usual disclaimer applies.

REFERENCES

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

Ai, C. and X. Chen (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica 71*, 1795–843.

© 2017 Royal Economic Society.

1368423x, 2018, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ectj.12086 by Johns Hopkins University, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

Andrews, D. W. K. (2011). Examples of L2-complete and boundedly-complete distributions. Cowles Foundation Discussion Paper 1801, Yale University.

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies 58*, 277–97.

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

Blundell, R. and S. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics 87*, 115–43.

Bowden, R. (1973). The theory of parametric identification. *Econometrica 41*, 1069–74.

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

Chen, X., V. Chernozhukov, S. Lee and W. K. Newey (2014). Local identification of nonparametric and semiparametric models. *Econometrica 82*, 785–809.

Chernozhukov, V. and C. Hansen (2005). An IV model of quantile treatment effects. *Econometrica 73*, 245–61.

Chernozhukov, V., I. Fernández-Val, J. Hahn and W. Newey (2013). Average and quantile effects in nonseparable panel models. *Econometrica 81*, 535–80.

D'Haultfoeuille, X. (2011). On the completeness condition in nonparametric instrumental problems. *Econometric Theory 27*, 1–12.

Graham, B. and J. Powell (2012). Identification and estimation of average partial effects in "irregular" correlated random coefficient panel data models. *Econometrica 80*, 2105–52.

Hahn, J. (1999). How informative is the initial condition in the dynamic panel model with fixed effects? *Journal of Econometrics 93*, 309–26.

Hoderlein, S. and H. White (2012). Nonparametric identification in nonseparable panel data models with generalized fixed effects. *Journal of Econometrics 168*, 300–14.

Honoré, B. (1993). Orthogonality conditions for censored models with fixed effects and lagged dependent variables. *Journal of Econometrics 59*, 35–61.

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

Hsiao, C., P. Hashem and T. Kamil (2002). Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics 109*, 107–50.

Hu, L. (2002). Estimation of a censored dynamic panel data model. *Econometrica 70*, 2499–517.

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

Hu, Y. and J. Shiu (2017). Nonparametric identification using instrumental variables: sufficient conditions for completeness. Forthcoming in *Econometric Theory*.

Hu, Y., S. Schennach and J. Shiu (2017). Injectivity of a class of integral operators with compactly supported kernels. Forthcoming in *Journal of Econometrics* (https://doi.org/10.1016/j.jeconom.2017.05.013).

Kiviet, J. (1995). On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *Journal of Econometrics 68*, 53–78.

Kyriazidou, E. (2001). Estimation of dynamic panel data sample selection models. *Review of Economic Studies 68*, 543–72.

Mattner, L. (1993). Some incomplete but bounded complete location families. *Annals of Statistics 21*, 2158–62.

© 2017 Royal Economic Society.

1368423x, 2018, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ectj.12086 by Johns Hopkins University, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Newey, W. and J. Powell (2003). Instrumental variable estimation of nonparametric models. *Econometrica 71*, 1565–78.

Neyman, J. and E. Scott (1948). Consistent estimates based on partially consistent observations. *Econometrica 16*, 1–32.

Schennach, S. M. (2014). Entropic latent variable integration via simulation. *Econometrica 82*, 345–85.

Shen, X. (1997). On methods of sieves and penalization. *Annals of Statistics 25*, 2555–91.

Shiu, J. and Y. Hu (2013). Identification and estimation of nonlinear dynamic panel data models with unobserved covariates. *Journal of Econometrics 175*, 116–31.

Wooldridge, J. (2005). Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics 20*, 39–54.

APPENDIX A: PROOFS OF RESULTS

Proof of Lemma 2.1: First, we have shown $f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0} = f_{X_{it},Y_{it-1},X_{it-1},U_{it}}$. Next, given (x_{it}, y_{it-1}), define integral operators

$$L_{f_{\widetilde{Y}_{it},X_{it},Y_{it-1},X_{it-1}}} : L^2(\widetilde{\mathcal{Y}}_{it}) \to L^2(\mathcal{X}_{it-1}), \tag{A.1}$$

with

$$(L_{f_{\widetilde{Y}_{it},X_{it},Y_{it-1},X_{it-1}}} h)(x_{it-1}) = \int f_{\widetilde{Y}_{it},X_{it},Y_{it-1},X_{it-1}}(\widetilde{y}_{it}, x_{it}, y_{it-1}, x_{it-1})h(\widetilde{y}_{it})d\widetilde{y}_{it},$$

$$\widetilde{L}_{f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_0}} : L^2(\widetilde{\mathcal{Y}}_{it}) \to L^2(\mathcal{U}_{it}, \omega) \tag{A.2}$$

with

$$(\widetilde{L}_{f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_0}} h)(u_{it}) = \int f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_0}(\widetilde{y}_{it}|x_{it}, y_{it-1}, u_{it})h(\widetilde{y}_{it})d\widetilde{y}_{it},$$

and

$$L_{f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0}} : L^2(\mathcal{U}_{it}, \omega) \to L^2(\mathcal{X}_{it-1}) \tag{A.3}$$

with

$$(L_{f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0}} h)(x_{it-1}) = \int f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0} h(u_{it})du_{it}.$$

For each $h \in L^2(\widetilde{\mathcal{Y}}_{it})$.

$$\begin{aligned}
(L_{f_{\widetilde{Y}_{it},X_{it},Y_{it-1},X_{it-1}}})(h)(x_{it-1}) &= \int_{\widetilde{\mathcal{Y}}_{it}} f_{\widetilde{Y}_{it},X_{it},Y_{it-1},X_{it-1}} h(\widetilde{y}_{it})d\widetilde{y}_{it} \\
&= \int_{\widetilde{\mathcal{Y}}_{it}} \Big(\int_{\mathcal{U}_{it}} f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_0} f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0} du_{it} \Big) h(\widetilde{y}_{it})d\widetilde{y}_{it} \\
&= \int_{\mathcal{U}_{it}} f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0} \Big(\int_{\widetilde{\mathcal{Y}}_{it}} f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_0} h(\widetilde{y}_{it})d\widetilde{y}_{it} \Big) du_{it} \\
&= (L_{f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0}} \widetilde{L}_{f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_0}})(h)(x_{it-1}),
\end{aligned}$$

© 2017 Royal Economic Society.

1368423x, 2018, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ectj.12086 by Johns Hopkins University, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

based on (2.8). Because this derivation holds for arbitrary h, this amounts to the operator relationship

$$\underbrace{L_{f_{\widetilde{Y}_{it},X_{it},Y_{it-1},X_{it-1}}}}_{\text{Assumption 2.4(a)}} = L_{f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0}} \underbrace{\widetilde{L}_{f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_0}}}_{\text{Assumption 2.4(b)}}.$$

Combining the condition $f_{X_{it},Y_{it-1},X_{it-1}} > 0$ and Assumption 2.4(a) results in

$$\{f_{\widetilde{Y}_{it},X_{it},Y_{it-1},X_{it-1}} : x_{it-1} \in \mathcal{X}_{it-1}\}$$

being complete over $L^2(\widetilde{\mathcal{Y}}_{it})$ and then $L_{f_{\widetilde{Y}_{it},X_{it},Y_{it-1},X_{it-1}}}$ is invertible. In addition, because Assumption 2.4(b) ensures that the operator $\widetilde{L}_{f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_0}}$ is invertible, the operator relationship implies that the invertibility of the operator $L_{f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0}}$, i.e. $\{f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0} : x_{it-1} \in \mathcal{X}_{it-1}\}$ is complete over $L^2(\mathcal{U}_{it}, \omega)$. Suppose that the parameter θ_0 is not locally identifiable. Then, there exists $\theta_k \neq \theta_0$ and $\theta_k \mapsto \theta_0$ such that $f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0} = f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_k}$. Using the definition of $f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0}$ and $f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_k}$,

$$f_{\widetilde{Y}_{it},X_{it},Y_{it-1},X_{it-1}} = \int f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_0} f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0} du_{it}, \tag{A.4}$$

$$f_{\widetilde{Y}_{it},X_{it},Y_{it-1},X_{it-1}} = \int f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_k} f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_k} du_{it}. \tag{A.5}$$

By subtracting (A.5) from (A.4), it follows that

$$\begin{aligned} 0 = &\int f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_k} f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_k} \\ &- f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_0} f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0} du_{it}, \\ = &\int f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_k} f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_k} \\ &- f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_k} f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0} du_{it} \\ &+ \int f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_k} f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0} \\ &- f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_0} f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0} du_{it}, \\ = &\int \underbrace{f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_k}}_{\text{Assumption 2.3}} (f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_k} - f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0}) du_{it} \\ &+ \int (f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_k} - f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_0}) \underbrace{f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0}}_{\text{Assumptions 2.4(a) and (b)}} du_{it}. \end{aligned} \tag{A.6}$$

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

$$0 = \int (f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_k} - f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_0}) f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0} du_{it},$$

for all x_{it-1} in $\mathcal{X}_{it-1}$. Because Assumptions 2.4(a) and (b) imply that $\{f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0} : x_{it-1} \in \mathcal{X}_{it-1}\}$ is complete over $L^2(\mathcal{U}_{it}, \omega)$, we obtain $f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_k} = f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_0}$ for $\theta_k \neq \theta_0$ and θ_k approaches θ_0 as k approaches ∞. This contradicts the local identifiability of θ_0 in $f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta}$, proving the lemma. □

Proof of Lemma 2.2: Because $f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta}$ is correctly specified at θ_0 by Lemma 2.1, $\widetilde{f}_{X_{it},Y_{it-1},X_{it-1};\theta}$ is also correctly specified at θ_0 after integrating out. However, denote two integral kernels as

© 2017 Royal Economic Society.

1368423x, 2018, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ectj.12086 by Johns Hopkins University, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$K_{A;\theta_0}(x_{it}, y_{it-1}, x_{it-1}, u_{it}) \equiv (\partial/\partial\theta) f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0}$ and $K_{B;\theta_0}(\widetilde{y}_{it}, x_{it}, y_{it-1}, u_{it}) \equiv (\partial/\partial\theta) f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_0}$. Divide (A.6) by $\theta - \theta_0 \neq 0$ and rewrite it as follows:

$$0 = \int f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta} \frac{f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta} - f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0}}{\theta - \theta_0} du_{it} + \int \frac{f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta} - f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_0}}{\theta - \theta_0} f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0} du_{it}.$$

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

$$0 = \int f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_0} K_{A;\theta_0}(x_{it}, y_{it-1}, x_{it-1}, u_{it}) du_{it} + \int K_{B;\theta_0}(\widetilde{y}_{it}, x_{it}, y_{it-1}, u_{it}) f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0} du_{it}. \tag{A.7}$$

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

$$L_{K_{A;\theta_0}} : L^2(\mathcal{U}_{it}, \omega) \to L^2(\mathcal{X}_{it-1}) \tag{A.8}$$

with

$$(L_{K_{A;\theta_0}} h)(x_{it-1}) = \int \frac{\partial}{\partial\theta} f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0}(x_{it}, y_{it-1}, x_{it-1}, u_{it}) h(u_{it}) du_{it},$$

and

$$L_{K_{B;\theta_0}} : L^2(\widetilde{\mathcal{Y}}_{it}) \to L^2(\mathcal{U}_{it}, \omega) \tag{A.9}$$

with

$$(L_{K_{B;\theta_0}} h)(u_{it}) = \int \frac{\partial}{\partial\theta} f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_0}(\widetilde{y}_{it}|x_{it}, y_{it-1}, u_{it}) h(\widetilde{y}_{it}) d\widetilde{y}_{it}.$$

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

$$\begin{aligned}
&(L_{K_{A;\theta_0}} \widetilde{L}_{f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_0}})(h)(x_{it-1}) \\
&= \int_{\mathcal{U}_{it}} K_{A;\theta_0}(x_{it}, y_{it-1}, x_{it-1}, u_{it}) \Big(\int_{\mathcal{Y}_{it}} f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_0} h(\widetilde{y}_{it}) d\widetilde{y}_{it} \Big) du_{it} \\
&= \int_{\mathcal{Y}_{it}} \Big(\int_{\mathcal{U}_{it}} f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_0} K_{A;\theta_0}(x_{it}, y_{it-1}, x_{it-1}, u_{it}) du_{it} \Big) h(\widetilde{y}_{it}) d\widetilde{y}_{it} \\
&= -\int_{\mathcal{Y}_{it}} \Big(\int_{\mathcal{U}_{it}} K_{B;\theta_0}(\widetilde{y}_{it}, x_{it}, y_{it-1}, u_{it}) f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0} du_{it} \Big) h(\widetilde{y}_{it}) d\widetilde{y}_{it} \\
&= -\int_{\mathcal{U}_{it}} f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0} \Big(\int_{\mathcal{Y}_{it}} K_{B;\theta_0}(\widetilde{y}_{it}, x_{it}, y_{it-1}, u_{it}) h(\widetilde{y}_{it}) d\widetilde{y}_{it} \Big) du_{it} \\
&= -(L_{f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0}} L_{K_{B;\theta_0}})(h)(x_{it-1}),
\end{aligned}$$

where we have used the following: an interchange of the order of integration (justified by Fubini's theorem); (A.7); the definitions of these operators in (A.2), (A.3), (A.8) and (A.9). This derivation yields the following operator relationship

$$L_{K_{A;\theta_0}} \underbrace{\widetilde{L}_{f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_0}}}_{\text{Assumption 2.4(b)}} + \underbrace{L_{f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0}}}_{\text{Lemma 2.1}} \underbrace{L_{K_{B;\theta_0}}}_{\text{Assumption 2.5}} = 0. \tag{A.10}$$

© 2017 Royal Economic Society.

1368423x, 2018, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ectj.12086 by Johns Hopkins University, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Whereas Assumptions 2.4(a) and (b) imply that $\widetilde{L}_{f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_0}}$ and $L_{f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0}}$ are invertible, Assumption 2.5 guarantees that $L_{K_{B;\theta_0}}$ is invertible. Because the operators other than $L_{K_{A;\theta_0}}$ in (A.10) are all invertible, the integral operator $L_{K_{A;\theta_0}}$ is also invertible. This implies that the family of its corresponding kernel functions $\{(\partial/\partial\theta) f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0} : x_{it-1} \in \mathcal{X}_{it-1}\}$ is complete over $L^2(\mathcal{U}_{it}, \omega)$.

Suppose θ_0 is not locally identifiable in $\widetilde{f}_{X_{it},Y_{it-1},X_{it-1};\theta}$. This implies that there exists $\theta_k \neq \theta_0$ and $\theta_k \mapsto \theta_0$ such that

$$\widetilde{f}_{X_{it},Y_{it-1},X_{it-1};\theta_k}(x_{it}, y_{it-1}, x_{it-1}) = \widetilde{f}_{X_{it},Y_{it-1},X_{it-1};\theta_0}(x_{it}, y_{it-1}, x_{it-1}).$$

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

$$\int \Big(\frac{f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_k} - f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0}}{\theta_k - \theta_0} \Big) du_{it} = 0 \quad \text{for all } x_{it-1}.$$

If $\theta_k \mapsto \theta_0$, the equation becomes

$$\int \Big(\frac{\partial}{\partial\theta} f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0} \Big) du_{it} = 0 \quad \text{for all } x_{it-1}. \tag{A.11}$$

Because $L^2(\mathcal{U}_{it}, \omega)$ contains the constant function, (A.11) is in contradiction with the completeness of $\{(\partial/\partial\theta) f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0} : x_{it-1} \in \mathcal{X}_{it-1}\}$ over $L^2(\mathcal{U}_{it}, \omega)$. Therefore, under Assumptions 2.3–2.5, θ_0 is locally identifiable. □

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

LEMMA A.1. *Under Assumptions 2.3–2.6, the family of functions* $\{(\partial/\partial x_{it-1}) f_{U_{it}|X_{it},Y_{it-1},X_{it-1}} : x_{it-1} \in \mathcal{X}_{it-1}\}$ *is complete over* $L^2(\mathcal{U}_{it}, \omega)$.

Proof: In a similar manner to (2.8), write the conditional version of (2.8) for $\theta = \theta_0$,

$$f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},X_{it-1}} = \int f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it}} f_{U_{it}|X_{it},Y_{it-1},X_{it-1}} du_{it}. \tag{A.12}$$

Taking the derivative with respect to X_{it-1} results in

$$\frac{\partial}{\partial x_{it-1}} f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},X_{it-1}} = \int f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it}} \frac{\partial}{\partial x_{it-1}} f_{U_{it}|X_{it},Y_{it-1},X_{it-1}} du_{it}. \tag{A.13}$$

Set $\kappa_1 = (\partial/\partial x_{it-1}) f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},X_{it-1}}$ and $\phi = (\partial/\partial x_{it-1}) f_{U_{it}|X_{it},Y_{it-1},X_{it-1}}$. For each (x_{it}, y_{it-1}), define operators

$$L_{\kappa_1} : L^2(\widetilde{\mathcal{Y}}_{it}) \to L^2(\mathcal{X}_{it-1})$$

with

$$(L_{\kappa_1} h)(x_{it-1}) = \int \kappa_1(\widetilde{y}_{it}, x_{it}, y_{it-1}, x_{it-1}) h(\widetilde{y}_{it}) d\widetilde{y}_{it},$$

and

$$L_\phi : \mathcal{L}^p(\mathcal{U}_{it}, \omega) \to L^2(\mathcal{X}_{it-1})$$

with

$$(L_\phi h)(x_{it-1}) = \int \phi(u_{it}, x_{it}, y_{it-1}, x_{it-1}) h(u_{it}) du_{it}.$$

© 2017 Royal Economic Society.

1368423x, 2018, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ectj.12086 by Johns Hopkins University, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

For $h \in L^2(\widetilde{\mathcal{Y}}_{it})$.

$$
\begin{aligned}
(L_{\kappa_1})(h)(x_{it-1}) &= \int_{\widetilde{\mathcal{Y}}_{it}} \kappa_1(\widetilde{y}_{it}, x_{it}, y_{it-1}, x_{it-1}) h(\widetilde{y}_{it}) d\widetilde{y}_{it} \\
&= \int_{\widetilde{\mathcal{Y}}_{it}} \Big(\int_{\mathcal{U}_{it}} f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it}} \phi(u_{it}, x_{it}, y_{it-1}, x_{it-1}) du_{it} \Big) h(\widetilde{y}_{it}) \widetilde{y}_{it} \\
&= \int_{\mathcal{U}_{it}} \phi(u_{it}, x_{it}, y_{it-1}, x_{it-1}) \Big(\int_{\mathcal{Y}_{it}} f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it}} h(y_{it}) dy_{it} \Big) du_{it} \\
&= (L_\phi \widetilde{L}_{f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_0}})(h)(x_{it-1}),
\end{aligned}
$$

where (A.2) defines the operator $\widetilde{L}_{f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_0}}$. With the definitions of the operators, this equation can be rewritten as an operator relationship

$$
L_{\kappa_1} = L_\phi \widetilde{L}_{f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_0}}. \tag{A.14}
$$

Assumptions 2.4(b) and 2.6 guarantee the invertibility of the operators $\widetilde{L}_{f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_0}}$ and L_{κ_1}, respectively. Applying this invertibility to (A.14) results in the invertibility of L_ϕ. Thus, the family $\{\phi(u_{it}, x_{it}, y_{it-1}, x_{it-1}) : x_{it-1} \in \mathcal{X}_{it-1}\}$ is complete over $L^2(\mathcal{U}_{it}, \omega)$ for each x_{it}, y_{it-1}. □

Proof of Lemma 2.3: First, $f_{X_{it},Y_{it-1},X_{it-1};\theta}$ is correctly specified at θ_0 because by Lemma 2.2, $\widetilde{f}_{X_{it},Y_{it-1},X_{it-1};\theta}$ is correctly specified at θ_0. Suppose that θ_0 is not locally identifiable in the observable joint density function $f_{X_{it},Y_{it-1},X_{it-1};\theta}$. There exists $\theta_k \neq \theta_0$ and $\theta_k \mapsto \theta_0$ such that, $f_{X_{it},Y_{it-1},X_{it-1};\theta_k} = f_{X_{it},Y_{it-1},X_{it-1};\theta_0}$. This implies that

$$
\frac{\widetilde{f}_{X_{it},Y_{it-1},X_{it-1};\theta_k}}{\int\int\int \widetilde{f}_{X_{it},Y_{it-1},X_{it-1};\theta_k} dx_{it} dy_{it-1} dx_{it-1}} = \frac{\widetilde{f}_{X_{it},Y_{it-1},X_{it-1};\theta_0}}{1} = f_{X_{it},Y_{it-1},X_{it-1}}. \tag{A.15}
$$

This equation can be expressed as

$$
\frac{\int f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_k} du_{it}}{f_{X_{it},Y_{it-1},X_{it-1}}} = \int\int\int \widetilde{f}_{X_{it},Y_{it-1},X_{it-1};\theta_k} dx_{it} dy_{it-1} dx_{it-1}. \tag{A.16}
$$

The multiple integral on the right-hand side of (A.16) only depends on the parameter θ_k, and is independent of x_{it-1}. This suggests that given $x_{1t-1} \neq x_{2t-1}$,

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

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

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

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

$$
\begin{aligned}
0 &= \int \frac{\partial}{\partial x_{it-1}} f_{U_{it}|X_{it},Y_{it-1},X_{1t-1}} du_{it} \\
&= \int \frac{\partial}{\partial x_{it-1}} f_{U_{it}|X_{it},Y_{it-1},X_{1t-1}} du_{it},
\end{aligned}
$$

which contradicts the completeness in Lemma A.1. Therefore, the parameter θ_0 is locally identifiable in the observable joint density function $f_{X_{it},Y_{it-1},X_{it-1};\theta}$. □

© 2017 Royal Economic Society.

1368423x, 2018, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ectj.12086 by Johns Hopkins University, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Proof of Lemma 3.3: First, suppose $\widetilde{\mathcal{Y}}$ is a domain such that $\widetilde{\mathcal{Y}} \subset \mathcal{R}$. Let the family $\{f(y|u) : u \in \mathcal{U}\}$ be complete in $L^2(\mathcal{R})$. For each $h \in L^2(\widetilde{\mathcal{Y}})$ such that $\int_{\widetilde{\mathcal{Y}}} h(y)f(y|u)dy = 0$ for all u. Extend h to a function in $L^2(\mathcal{R})$ by

$$\tilde{h}(x) = \begin{cases} h(x) & \text{if } x \in \widetilde{\mathcal{Y}}, \\ 0 & \text{otherwise.} \end{cases}$$

It follows that $\int_{\mathcal{R}} \tilde{h}(y)f(y|u)dy = 0$ for all u. By the completeness of $f(y|u)$ over $L^2(\mathcal{R})$, $\tilde{h} = 0$ almost everywhere. Thus, $h = 0$ almost everywhere and $f(y|u)$ is complete over $L^2(\mathcal{R})$. Thus, the completeness of a function over a smaller domain is implied by the completeness of the function over a larger domain, and sufficient conditions for the completeness of these two families can be reduced to the completeness in $L^2(\mathcal{R})$.

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

$$\int h(y)\frac{\partial}{\partial y}\phi\Big(\frac{y-c-u}{\sigma_\xi}\Big)dy = h(y)\phi\Big(\frac{y-c-u}{\sigma_\xi}\Big)\Big|_{-\infty}^{\infty} - \int \frac{\partial}{\partial y}h(y)\phi\Big(\frac{y-c-u}{\sigma_\xi}\Big)dy$$
$$= -\int \frac{\partial}{\partial y}h(y)\phi\Big(\frac{y-c-u}{\sigma_\xi}\Big)dy.$$

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

$$0 = \int h(y)\frac{\partial^2}{\partial^2 y}\phi\Big(\frac{y-c-u}{\sigma_\xi}\Big)dy$$
$$= h(y)\frac{\partial}{\partial y}\phi\Big(\frac{y-c-u}{\sigma_\xi}\Big)\Big|_{-\infty}^{\infty} - \int \frac{\partial}{\partial y}h(y)\frac{\partial}{\partial y}\phi\Big(\frac{y-c-u}{\sigma_\xi}\Big)dy$$
$$= -\frac{\partial}{\partial y}h(y)\phi\Big(\frac{y-c-u}{\sigma_\xi}\Big)\Big|_{-\infty}^{\infty} + \int \frac{\partial^2}{\partial^2 y}h(y)\phi\Big(\frac{y-c-u}{\sigma_\xi}\Big)dy$$
$$= \int \frac{\partial^2}{\partial^2 y}h(y)\phi\Big(\frac{y-c-u}{\sigma_\xi}\Big)dy.$$

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

APPENDIX B: IDENTIFICATION IN THE DISCRETE CASE

In this appendix, we present a simple case in which the observed variables Y_{it}, X_{it}, Y_{it-1}, X_{it-1} and the unobserved covariate U_{it} are all discrete. We show how to use the identification techniques in Theorem 2.1 for this discrete case. For simplicity, assume that the variables $\widetilde{Y}_{it}$, X_{it-1} and U_{it} have the same size J (i.e. $\widetilde{Y}_{it}, X_{it-1}, U_{it} \in \{1, 2, \ldots, J\}$). For this setting, the integral operators used previously can be represented by

© 2017 Royal Economic Society.

1368423x, 2018, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ectj.12086 by Johns Hopkins University, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

J-by-J matrices. The idea of using the identification strategy in the discrete case for ease of exposition comes from the fact that a complete integral operator is associated with an invertible matrix.[3]

Equation (2.8) in the discrete case is

$$f_{\widetilde{Y}_{it},X_{it},Y_{it-1},X_{it-1}} = \sum_{U_{it}=1}^{J} f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta} f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta}. \tag{B.1}$$

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

$$\begin{aligned} M_{f_{\widetilde{Y}_{it},x_{it},y_{it-1},X_{it-1}}} &= (f_{\widetilde{Y}_{it},X_{it},Y_{it-1},X_{it-1}}(\widetilde{Y}_{it}, x_{it}, y_{it-1}, X_{it-1}))_{\widetilde{y}_{it},x_{it-1}} \\ L_{f_{\widetilde{Y}_{it}|x_{it},y_{it-1},U_{it};\theta}} &= (f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta}(\widetilde{Y}_{it}, x_{it}, y_{it-1}, U_{it}))_{\widetilde{y}_{it},u_{it}} \\ M_{f_{x_{it},y_{it-1},X_{it-1},U_{it};\theta}} &= (f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta}(x_{it}, y_{it-1}, X_{it-1}, U_{it}))_{u_{it},x_{it-1}}. \end{aligned}$$

Rewrite the equality (B.1) in terms of these matrices as follows:

$$\underbrace{M_{f_{\widetilde{Y}_{it},x_{it},y_{it-1},X_{it-1}}}}_{\text{observed from data}} = \underbrace{L_{f_{\widetilde{Y}_{it}|x_{it},y_{it-1},U_{it};\theta}}}_{\text{model specification}} M_{f_{x_{it},y_{it-1},X_{it-1},U_{it};\theta}}. \tag{B.2}$$

Assumption 2.3 implies that the square matrix $L_{f_{\widetilde{Y}_{it}|x_{it},y_{it-1},U_{it};\theta}}$ is invertible, leading to

$$M_{f_{x_{it},y_{it-1},X_{it-1},U_{it};\theta}} = (L_{f_{\widetilde{Y}_{it}|x_{it},y_{it-1},U_{it};\theta}})^{-1} M_{f_{\widetilde{Y}_{it},x_{it},y_{it-1},X_{it-1}}}. \tag{B.3}$$

As discussed earlier, it is necessary to ensure that $M_{f_{x_{it},y_{it-1},X_{it-1},U_{it};\theta}}$ is identifiable at θ_0. According to the proof of Lemma 2.1, there are two steps for identifiability. First, given (x_{it}, y_{it-1}), define

$$\begin{aligned} L_{f_{\widetilde{Y}_{it},x_{it},y_{it-1},X_{it-1}}} &= (f_{\widetilde{Y}_{it},X_{it},Y_{it-1},X_{it-1}}(\widetilde{Y}_{it}, x_{it}, y_{it-1}, X_{it-1}))_{x_{it-1},\widetilde{y}_{it}} \\ \widetilde{L}_{f_{\widetilde{Y}_{it}|x_{it},y_{it-1},U_{it};\theta}} &= (f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta}(\widetilde{Y}_{it}, x_{it}, y_{it-1}, U_{it}))_{u_{it},\widetilde{y}_{it}} \\ L_{f_{x_{it},y_{it-1},X_{it-1},U_{it};\theta}} &= (f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta}(x_{it}, y_{it-1}, x_{it-1}, U_{it}))_{x_{it-1},u_{it}}. \end{aligned}$$

Equality (B.1) can then be expressed by these matrices as follows:

$$\underbrace{L_{f_{\widetilde{Y}_{it},x_{it},y_{it-1},X_{it-1}}}}_{\text{Assumption 2.4(a)}} = L_{f_{x_{it},y_{it-1},X_{it-1},U_{it};\theta_0}} \underbrace{\widetilde{L}_{f_{\widetilde{Y}_{it}|x_{it},y_{it-1},U_{it};\theta_0}}}_{\text{Assumption 2.4(b)}}. \tag{B.4}$$

Notice that in this simple case,

$$\begin{aligned} L_{f_{\widetilde{Y}_{it},x_{it},y_{it-1},X_{it-1}}} &= M^T_{f_{\widetilde{Y}_{it},x_{it},y_{it-1},X_{it-1}}}, \\ \widetilde{L}_{f_{\widetilde{Y}_{it}|x_{it},y_{it-1},U_{it};\theta}} &= L^T_{f_{\widetilde{Y}_{it}|x_{it},y_{it-1},U_{it};\theta}}, \\ L_{f_{x_{it},y_{it-1},X_{it-1},U_{it};\theta}} &= M^T_{f_{x_{it},y_{it-1},X_{it-1},U_{it};\theta}}, \end{aligned}$$

which we might not have for a general continuous case. The matrix notations used here are based on integral operators in the proofs of the lemmata. Assumption 2.4 makes $L_{f_{x_{it},y_{it-1},X_{it-1},U_{it};\theta_0}}$ invertible. Hence,

[3] If $y, u \in \{1, 2\}$ and $\int_{\mathcal{U}} h(u) f(y|u) du = 0$, then the condition is equivalent to

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

The function h can be uniquely determined as $h = 0$ iff the first matrix representing $f_{y|u}$ is invertible.

© 2017 Royal Economic Society.

1368423x, 2018, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ectj.12086 by Johns Hopkins University, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

its transpose $M_{f_{x_{it},y_{it-1},X_{it-1},U_{it};\theta_0}}$ is also invertible. Then, suppose that there exists $\theta_k \neq \theta_0$ and $\theta_k \mapsto \theta_0$ such that

$$M_{f_{x_{it},y_{it-1},X_{it-1},U_{it};\theta_k}} = M_{f_{x_{it},y_{it-1},X_{it-1},U_{it};\theta_0}}. \tag{B.5}$$

Following the derivation in (A.6), we have a matrix expression

$$\begin{aligned}0 = {} & L_{f_{\widetilde{Y}_{it}|x_{it},y_{it-1},U_{it};\theta_k}}(M_{f_{x_{it},y_{it-1},X_{it-1},U_{it};\theta_k}} - M_{f_{x_{it},y_{it-1},X_{it-1},U_{it};\theta_0}}) \\ & + (L_{f_{\widetilde{Y}_{it}|x_{it},y_{it-1},U_{it};\theta_k}} - L_{f_{\widetilde{Y}_{it}|x_{it},y_{it-1},U_{it};\theta_0}})M_{f_{x_{it},y_{it-1},X_{it-1},U_{it};\theta_0}}.\end{aligned}$$

The invertibility of $M_{f_{x_{it},y_{it-1},X_{it-1},U_{it};\theta_0}}$ and (B.5) implies that $L_{f_{\widetilde{Y}_{it}|x_{it},y_{it-1},U_{it};\theta}}$ is not identifiable at θ_0, which is a contradiction.

Set $J \times 1$-vector $\mathbf{J}_1 = (1, 1, \ldots, 1)^T$. Integrating out the unobserved covariate in the discrete case leads to $\mathbf{J}_1^T M_{f_{x_{it},y_{it-1},X_{it-1},U_{it};\theta}}$. Suppose that there exists $\theta_k \neq \theta_0$ and $\theta_k \mapsto \theta_0$ such that $\mathbf{J}_1^T M_{f_{x_{it},y_{it-1},X_{it-1},U_{it};\theta_k}} = \mathbf{J}_1^T M_{f_{x_{it},y_{it-1},X_{it-1},U_{it};\theta_0}}$. It then follows that

$$0 = \mathbf{J}_1^T\left(\frac{M_{f_{x_{it},y_{it-1},X_{it-1},U_{it};\theta_k}} - M_{f_{x_{it},y_{it-1},X_{it-1},U_{it};\theta_0}}}{\theta_k - \theta_0}\right). \tag{B.6}$$

If $\theta \mapsto \theta_0$, the above equation implies $0 = \mathbf{J}_1^T M_{(\partial/\partial\theta)f_{x_{it},y_{it-1},X_{it-1},U_{it};\theta_0}}$, where

$$M_{(\partial/\partial\theta)f_{x_{it},y_{it-1},X_{it-1},U_{it};\theta}} = \left(\frac{\partial}{\partial\theta} f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta}(x_{it}, y_{it-1}, X_{it-1}, U_{it})\right)_{u_{it},x_{it-1}}.$$

Rewrite (A.7) in the discrete case as

$$\begin{aligned}0 = {} & \sum_{U_{it}=1}^{J} f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_0} \frac{\partial}{\partial\theta} f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0} \\ & + \sum_{U_{it}=1}^{J} \frac{\partial}{\partial\theta} f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},U_{it};\theta_0} f_{X_{it},Y_{it-1},X_{it-1},U_{it};\theta_0}.\end{aligned} \tag{B.7}$$

This leads to the following matrix expression

$$\begin{aligned} & M_{(\partial/\partial\theta)f_{x_{it},y_{it-1},X_{it-1},U_{it};\theta_0}} \underbrace{\widetilde{L}_{f_{\widetilde{Y}_{it}|x_{it},y_{it-1},U_{it};\theta_0}}}_{\text{Assumption 2.4(b)}} \\ & + \underbrace{L_{f_{x_{it},y_{it-1},X_{it-1},U_{it};\theta_0}}}_{\text{Lemma 2.1}} \underbrace{M_{(\partial/\partial\theta)f_{\widetilde{Y}_{it}|x_{it},y_{it-1},U_{it};\theta_0}}}_{\text{Assumption 2.5}} = 0,\end{aligned} \tag{B.8}$$

where

$$M_{(\partial/\partial\theta)f_{\widetilde{Y}_{it}|x_{it},y_{it-1},U_{it};\theta_0}} = \left(\frac{\partial}{\partial\theta} f_{\widetilde{Y}_{it}|x_{it},y_{it-1},U_{it};\theta_0}\right)_{u_{it},\widetilde{y}_{it}}.$$

Applying assumptions to (B.8) shows that $M_{(\partial/\partial\theta)f_{x_{it},y_{it-1},X_{it-1},U_{it};\theta_0}}$ is invertible, which contradicts $0 = \mathbf{J}_1^T M_{(\partial/\partial\theta)f_{x_{it},y_{it-1},X_{it-1},U_{it};\theta_0}}$.

Finally, the normalization in the discrete case is equivalent to

$$\mathbf{V}_{f_{x_{it},y_{it-1},X_{it-1};\theta}} \equiv \frac{\mathbf{J}_1^T M_{f_{x_{it},y_{it-1},X_{it-1},U_{it};\theta}}}{\sum_{x_{it}=1}^{J_{x_{it}}} \sum_{y_{it-1}=1}^{J_{y_{it-1}}} (\mathbf{J}_1^T M_{f_{x_{it},y_{it-1},X_{it-1},U_{it};\theta}} \mathbf{J})}, \tag{B.9}$$

where $J_{x_{it}}$ and $J_{y_{it-1}}$ represent the sizes of the discrete variables x_{it} and y_{it-1}, respectively. Suppose the normalization step does not lead to local identifiability at θ_0. This implies that there exists $\theta_k \neq \theta_0$ and

© 2017 Royal Economic Society.

1368423x, 2018, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ectj.12086 by Johns Hopkins University, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$\theta_k \mapsto \theta_0$ such that

$$\frac{\mathbf{J}_1^T M_{f_{x_{it},y_{it-1},X_{it-1},U_{it};\theta_k}}}{\sum_{x_{it}=1}^{J_{x_{it}}} \sum_{y_{it-1}=1}^{J_{y_{it-1}}} (\mathbf{J}_1^T M_{f_{x_{it},y_{it-1},X_{it-1},U_{it};\theta_k}} \mathbf{J})} = \frac{\mathbf{J}_1^T M_{f_{x_{it},y_{it-1},X_{it-1},U_{it};\theta_0}}}{1}$$
$$= \mathbf{V}_{f_{x_{it},y_{it-1},X_{it-1};\theta_0}}. \tag{B.10}$$

Rearrange the term

$$(\mathbf{J}_1^T M_{f_{x_{it},y_{it-1},X_{it-1},U_{it};\theta_k}})./\mathbf{V}_{f_{x_{it},y_{it-1},X_{it-1};\theta_0}}$$
$$= \Big(\sum_{x_{it}=1}^{J_{x_{it}}} \sum_{y_{it-1}=1}^{J_{y_{it-1}}} (\mathbf{J}_1^T M_{f_{x_{it},y_{it-1},X_{it-1},U_{it};\theta_k}} \mathbf{J}) \Big) \mathbf{J}_1,$$

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

$$(\mathbf{J}_1^T M_{f_{x_{it},y_{it-1},X_{1t-1},U_{it};\theta_k}})./\mathbf{V}_{f_{x_{it},y_{it-1},X_{1t-1};\theta_0}}$$
$$= (\mathbf{J}_1^T M_{f_{x_{it},y_{it-1},X_{2t-1},U_{it};\theta_k}})./\mathbf{V}_{f_{x_{it},y_{it-1},X_{2t-1};\theta_0}}.$$

Using (B.3), rewrite this equation as

$$0 = \mathbf{J}_1^T (L_{f_{\widetilde{Y}_{it}|x_{it},y_{it-1},U_{it};\theta_k}})^{-1} (M_{f_{\widetilde{Y}_{it},x_{it},y_{it-1},X_{1t-1}}} - M_{f_{\widetilde{Y}_{it},x_{it},y_{it-1},X_{2t-1}}})./\mathbf{V}_{f_{x_{it},y_{it-1},X_{2t-1};\theta_0}}.$$

Denote $M_{f_{\widetilde{Y}_{it}|x_{it},y_{it-1},\triangle X_{it-1}}}$ as a matrix of the difference of $f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},X_{it-1}}$ with respect to X_{it-1}. If $\theta_k \mapsto \theta_0$, then

$$0 = \mathbf{J}_1^T (L_{f_{\widetilde{Y}_{it}|x_{it},y_{it-1},U_{it};\theta_0}})^{-1} (M_{f_{\widetilde{Y}_{it}|x_{it},y_{it-1},X_{1t-1}}} - M_{f_{\widetilde{Y}_{it}|x_{it},y_{it-1},X_{2t-1}}})$$
$$\equiv \mathbf{J}_1^T \underbrace{(L_{f_{\widetilde{Y}_{it}|x_{it},y_{it-1},U_{it};\theta_0}})^{-1}}_{\text{Assumption 2.3}} \underbrace{M_{f_{\widetilde{Y}_{it}|x_{it},y_{it-1},\triangle X_{it-1}}}}_{\text{Assumption 2.6}},$$

where $M_{f_{\widetilde{Y}_{it}|x_{it},y_{it-1},X_{it-1}}} \equiv (f_{\widetilde{Y}_{it}|X_{it},Y_{it-1},X_{it-1}}(\widetilde{Y}_{it}|x_{it}, y_{it-1}, X_{it-1}))_{\widetilde{y}_{it},x_{it-1}}$. This contradicts the invertibility under Assumptions 2.3 and 2.6, showing that the density function is still locally identifiable at θ_0.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web site:

Replication files

© 2017 Royal Economic Society.

1368423x, 2018, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ectj.12086 by Johns Hopkins University, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License