Nonparametric Identification and Estimation of Panel Quantile Models with Sample Selection

Sungwon Lee*
Department of Economics
Sogang University
sungwon.lee83@gmail.com

First draft: September 23, 2019
This draft: February 20, 2023

Abstract

This paper develops nonparametric panel quantile regression models with sample selection. The class of models allows the unobserved heterogeneity to be correlated with time-varying regressors in a time-invariant manner. I adopt the correlated random effects approach proposed by Mundlak (1978) and Chamberlain (1980), and the control function approach to correct the sample selection bias. The class of models is general and flexible enough to incorporate many empirical issues, such as endogeneity of regressors and censoring. Identification of the static model requires that $T \geq 3$, where T is the number of time periods, and that there is an excluded variable that affects the selection probability. We consider a dynamic extension of the models and provide identification conditions. The condition on T in dynamic models is stronger than that in static models as it is needed that $T \geq 4$ for identification of dynamic models. We also propose semiparametric models for practical implementation of estimation. Based on the identification result, this paper proposes to use penalized sieve minimum distance estimation to estimate the parameters and establishes the asymptotic theory. A small Monte-Carlo simulation study confirms that the estimators perform well in finite samples.

Keywords: Sample selection, panel data, quantile regression, nonseparable models, correlated random effects, nonparametric identification, control function approach, penalized sieve minimum distance.
JEL Classification Numbers: C14, C21, C23.

*I am deeply grateful to Jason Abrevaya, Timothy Christensen, Gregory Cox, Christian Dustmann, Sukjin Han, Doosoo Kim, Frank Kleibergen, Patrick Kline, Andres Santos, Yongcheol Shin, and participants in the 2020 World Congress of the Econometric Society, the 2020 Korean Economic Review International Conference, the 2021 IAAE Annual Conference, the 2021 Asian Meeting of Econometric Society, the 2021 Econometric Society European Summer Meeting, and the seminars at Korean Econometric Society, Ryerson University, and Sogang University for their valuable comments and suggestions. Any errors are mine.

1 Introduction

Sample selection is prevalent in economics. Since the seminal work of Gronau (1974) and Heckman (1979), sample selection has considerably received a lot of attention from both theoretical and applied econometrics due to its relevance and importance in many empirical contexts (e.g. Ahn and Powell (1993), Donald (1995), Das et al. (2003), and Newey (2009)). At the same time, quantile regression models have become a popular alternative to conditional mean models since the seminal work of Koenker and Bassett (1978) as they allow to investigate the distribution of the outcome variable and recover heterogeneous effects. Although many papers have studied sample selection and quantile regression, the literature on the intersection of them is relatively scarce as most papers have considered sample selection issues for conditional mean regression models. In particular, sample selection issues in quantile regression models for panel data have not been well-addressed, whereas the availability of panel data has become larger.

In this paper, we develop a nonseparable panel quantile model with sample selection and study identification and estimation of the model. Specifically, we consider the following panel quantile model:

$$\begin{aligned} Y_t^* &= g(X_t, U_t), \\ Y_t &= D_t Y_t^*, \end{aligned} \tag{1}$$

where t indicates time, Y_t^* is an outcome variable of interest, X_t is a vector of time-varying covariates, U_t is an unobserved heterogeneity, and D_t is a dummy variable indicating if it is selected. The structural function g is assumed to be strictly increasing with respect to its second argument for almost all X_t.

One of distinct features of the model in (1) is nonseparability between X_t and U_t. Many papers in the literature on sample selection develop models and estimators under different sets of assumptions, but they share some common feature that they focus on additively separable models. For quantile regression in the presence of sample selection, Buchinsky (1998) considered an additively separable quantile regression model for cross-sectional data. The additive separability facilitates identification and estimation of model parameters, but it considerably restricts the type of heterogeneity that can be allowed in a model. Nonseparability is important in quantile regression as (i) it can allow for various types of heterogeneous effects and (ii) it is less vulnerable to model misspecification.[1] Nonseparable quantile regression with sample selection has been studied quite recently by Arellano and Bonhomme (2017) and Chernozhukov et al. (2018). While their models are semiparametric and mainly for cross-sectional data, this paper focuses on nonparametric quantile regression models for panel data. To my best knowledge, this paper is the first to consider nonseparable panel quantile regression models in the presence of sample selection.

[1] Huber and Melly (2015) point out that the additive separability may lead to inconsistency of the estimator in the linear quantile regression models and propose a test for the structure.

Panel data models can incorporate time-invariant heterogeneity that may be correlated with time-varying regressors. When time-invariant heterogeneity is correlated with time-varying regressors, it is called time-invariant endogeneity. One can resolve time-invariant endogeneity by taking some differencing-based approach when the model is linear or additively separable, but it is much harder to deal with time-invariant endogeneity for nonlinear or nonseparable models. To overcome this difficulty, we consider a correlated random effects (CRE) approach which was originally pioneered by Mundlak (1978) and Chamberlain (1980). The main idea of the CRE approach is to assume that the distribution of the unobserved heterogeneity depends on the whole history of the time-varying covariates. In doing so, one can allow for time-invariant endogeneity as well as improve tractability of the model..[2]

This paper provides conditions under which the model parameters are nonparametrically identified. The main idea of the identification strategy in this paper is to utilize variation in some excluded variables. The model in this paper contains two types of endogeneity - time-invarant endogeneity and endogenous selection. Therefore, it is expected to have at least two excluded variables for identification. We show that one can use the rich information in panel data to deal with the time-invariant endogeneity. This feature of the identification strategy requires that the number of time periods be greater than or equal to 3 and covariates have enough variation. On the other hand, we use a control function approach to correct for the selection bias, and this requires for an instrumental variable that varies the selection probability but does not directly affect the outcome. An exclusion restriction associated with the instrument, together with a conditional independence assumption, allows to resolve the endogenous selection, and this is a generalization of the approach of Heckman (1979). Under these standard identification conditions, the structural function of the outcome variable, which is denoted by $g(\cdot,\cdot)$, and the conditional distribution of the unobserved error term for the selected are nonparametrically identified. The identification strategy considered in this paper shares some common with those in Vytlacil and Yildiz (2007), Torgovitsky (2015), and Chen et al. (2020) in the sense that it relies on the notion of matching. We also consider several extensions of the model to address some important empirical issues such as time-varying endogeneity and censoring. It is shown that the model in this paper can easily be extended to incorporate those issues, and therefore the class of models in this paper is very general and flexible.

It is often of interest to investigate dynamic relationships when panel data are available. We develop dynamic panel quantile regression models in which lagged dependent variables enter as regressors with sample selection and provide conditions for nonparametric iden-

[2]In the standard linear panel data models, the unobserved heterogeneity U_t is decomposed into two parts: one is a time-invariant error term, and the other is an time-varying idiosyncratic error. This dependence is the main motivation of the fixed effects model where X_t and time-invariant components are correlated in an arbitrary manner. In this paper, we do not explicitly distinguish time-invariant components in U_t. However, the dependence between time-varying regressors and time-invariant components in the error term is allowed in this paper, and we introduce a time-invariant component by using the CRE approach.

tification. The presence of lagged dependent variables as regressors requires a stronger condition on the number of time periods than that in static models. Specifically, it is needed that $T \geq 4$ for nonparametric identification as the initial time period is treated as given. Panel data models with sample selection have been studied by, for example, Kyriazidou (2001), Gayle and Viauroux (2007), and Semykina and Wooldridge (2013), but none of them considered quantile regression models.

While the fully nonparametric models are robust to model misspecification, they may not be tractable in estimation. In this regard, we propose two classes of semiparametric models: (i) semiparametric index models and (ii) additively separable models. These classes of models are very useful in a sense that one can reduce the dimension of some nonparametric object. Then, we provide conditions under which the parameters of the models are identified.

The identification result suggests a nonlinear optimization problem for estimation that the selection probability enters as a control function. Based on the identification result, we propose to use the penalized sieve minimum distance (PSMD) estimation developed by Chen and Pouzo (2015). The method of sieves provides a very flexible and general way to estimate semi-nonparametric or nonparametric models. The sieve method is also easy to implement in practice, and therefore it has been widely used. This paper provides the asymptotic theory for the PSMD sieve estimators, including consistency, convergence rates, and asymptotic normality of smooth functionals. We also develop the asymptotic theory for the sieve quasi likelihood ratio (QLR) inference.

Unlike the cross-sectional or time-series data, there are multiple types of data in terms of the number of individuals and the number of time periods, which are denoted by n and T, respectively, for panel data models. The relative magnitude between these two quantities defines the data structure, and this feature of the data structure is very important for panel data models as they are related to estimation of models. This paper considers a fixed T-panel data model, and the fixed-T framework renders the model fit into data where T is much smaller than n. The large-T framework is frequently used in the literature on nonlinear fixed effects panel models to handle the incidental parameter problem (Neyman and Scott (1948)).[3] For panel quantile models with fixed effects, Koenker (2004), Canay (2011), Kato et al. (2012), and Besstremyannaya and Golovan (2019) make use of the large-T framework.[4] To adopt the large-T framework, however, the number of time periods in data should be larger than the number of individuals, and this requirement may not be appropriate to or suitable for some datasets, especially microdatasets or short panel

[3] Fernández-Val and Weidner (2018) provide a comprehensive review on the literature on large-T panel data models.

[4] Canay (2011) originally imposed a condition that $n/T^s \to 0$ for some $s > 1$ to establish consistency and asymptotic normality. Under this condition, one may use short panel data where n grows faster than T. Besstremyannaya and Golovan (2019), however, point out that the rate condition is not sufficient for existence of a limiting distribution or for zero mean of a limiting distribution. This result in Besstremyannaya and Golovan (2019) suggests that the estimator of Canay (2011) does not fit into short panel data.

datasets. In addition, not only the asymptotic properties of estimators, but finite-sample performances also depend on the magnitudes of n and T.[5] In this regard, estimators based on the large-T framework may be sensitive to the model specification and nature of data. On the other hand, I consider the fixed-T framework while incorporating time-invariant endogeneity, and this allows for a much wider applicability of the model in this paper.

We conduct a Monte-Carlo simulation study with a semiparametric model to examine the performance of estimators in finite samples. The results show that the semiparametric estimators have negligible biases and small standard deviations, which suggest that they perform well in finite samples.

Literature This paper is related to the literature on the panel data models with sample selection (e.g. Wooldridge (1995); Kyriazidou (1997); Semykina and Wooldridge (2010, 2013)).[6] For panel data in the presence of sample selection, Wooldridge (1995) and Kyriazidou (1997) propose estimators for panel data models where the outcome variable equation is linear in parameters. Wooldridge (1995) adopts the Mundlak-Chamberlain device (Mundlak (1978) and Chamberlain (1980)) to handle the time-invariant unobserved heterogeneity and uses the control function approach in the same spirit of Heckman (1979). The idea of Wooldridge (1995) is extended by Semykina and Wooldridge (2010) and Semykina and Wooldridge (2013) to incorporate time-varying endogeneity and dynamic panel data models, respectively. These papers, however, hugely rely on (semi-) parametric assumptions as well as the additive separability of the error term. Kyriazidou (1997) considers the conditional exchangeability assumption and the additively separable structure of the model. However, not only the additively separable error structure, but the conditional exchangeability condition may also fail to hold in some cases.[7] This paper differs from the aforementioned papers in that it considers nonseparable quantile regression models for panel data, whereas Wooldridge (1995) and Semykina and Wooldridge (2010) consider linear conditional mean models.

This paper is also related to the studies in the literature on (nonparametric) identification and estimation with endogeneity. This literature is too large to list all related papers, and one may refer to Matzkin (2007) for a comprehensive review. Focusing on sample selection, this paper is closely related to, for example, Buchinsky (1998), Das et al. (2003), and Newey (2009). The model varies across them, but they make use of the control function approach to correct for the sample selection bias. This paper shares some common with them as the identification strategy in this paper also utilizes a control function, but differs from them as the model in this paper is nonseparable and for panel data. As mentioned earlier, Arellano and Bonhomme (2017) and Chernozhukov et al. (2018) consider nonsep-

[5] The simulation results in Kato et al. (2012) show that the root mean squared error is quite large when T is small in the location-scale shift model.

[6] One can refer to Dustmann and Rochina-Barrachina (2007) for comparison of some estimators including Wooldridge (1995) and Kyriazidou (1997).

[7] A related discussion can be found in Altonji and Matzkin (2005).

arable quantile regression models with sample selection, but their models are more fitting into cross-sectional data. Furthermore, they consider semiparametric model specifications for estimation. In contrast, this paper considers a class of nonparametric models with sample selection for panel data, and the identification or estimation strategy developed in this paper does not impose such distributional assumptions. As a result, this paper extends Arellano and Bonhomme (2017) and Chernozhukov et al. (2018) to nonparametric quantile regression models for panel data.

The literature on (nonlinear) panel data models is another area that this paper is closely related to. One of the features of the model in this paper is that we adopt the CRE approach, and this approach is also widely used in the literature to address time-invariant endogeneity. Abrevaya and Dahl (2008) make use of the CRE approach for linear panel quantile regression, but this paper differs from them as it considers nonparametric nonseparable models. Bester and Hansen (2009) study identification of the marginal effects in general panel data models with the CRE approach with a focus on identification of marginal effects, and therefore this paper is different from theirs in terms of the parameter of interest and identification/estimation strategy. Arellano and Bonhomme (2016) recently consider CRE specifications and develop a class of tractable nonlinear panel models, but their identification strategy relies on a high-level condition called injectivity, which is related to the completeness condition. On the other hand, this paper adopts a control function approach for identification. For general nonlinear panel data models, Altonji and Matzkin (2005) study identification and estimation of local average responses (LARs) and structural functions under an assumption called exchangeability. While the exchangeability condition generally implies some shape restrictions on the distribution of the unobserved error term, we circumvent to restrict the shape of the distribution of the error term by taking the CRE approach. Hoderlein and White (2012), Chernozhukov et al. (2013), and Chernozhukov et al. (2015) study identification of average structural functions and quantile structural functions, but this paper considers identification and estimation of the structural functions. Evdokimov (2010) studies identification and estimation of a class of panel data models, but his identification is based on deconvolution. Therefore, the identification strategy in this paper is completely different from his. More importantly, none of them address sample selection issues which are the main focus of this paper. In the absence of sample selection, Galvao Jr (2011) develops a dynamic quantile regression model with fixed effects. Galvao Jr (2011) uses lagged regressors as instruments, which is a s well-established approach in the literature on dynamic panel data models (e.g., Arellano and Bond (1991) and Blundell and Bond (1998)), to deal with the correlation between lagged dependent variables and fixed effects. On the other hand, we rely on the CRE approach to resolve time-invariant endogeneity in this paper and focus on sample selection.

Outline The rest of this paper is organized as follows. In section 2, we introduce the model and parameters of interest. Section 3 considers nonparametric identification of the model. Sections 4 and 5 present an dynamic extension of the model and semiparametric models, respectively. Section 6 proposes two-step sieve estimation and establishes the asymptotic theory for the nonparametric sieve two-step estimators. Section 7 presents the results of a Monte-Carlo simulation study. Section 8 concludes and discusses future work. All mathematical proofs for the asymptotic theory are presented in the appendix.

Notation We introduce some notation. For a vector A, A' denotes the transpose of A. For a generic random variable A_t, the support of A_t is denoted by $Supp(A_t)$. Let $\mathbf{A} \equiv (A_1, A_2, ..., A_T)'$ be the random vector consisting of A_t's from time period 1 to T. and let $\mathbf{A}_{-t} \equiv (A_1, ..., A_{t-1}, A_{t+1}, ..., A_T)'$ be the random vector consisting of A_t's from time period 1 to T but not t. We use notation $\mathbf{A}_{-t,s}$ to denote the random vector consisting of A_t's from time period 1 to T but not t and s. Realizations of A and $\mathbf{A}$ are denoted by a and $\mathbf{a}$, respectively.[8] For two random variables A and B and for any $u \in (0,1)$, $Q_{A|B}(u|b)$ indicates the u-th conditional quantile of A on $B = b$, and $F_{A|B}(a|b)$ is the conditional distribution function of A given $B = b$. $\mathbb{E}[\cdot]$ is the expectation operator.

2 The Model

We consider the following general non-separable panel data model:

$$Y_t^* = g(X_t, U_t), \tag{2}$$

where $Y_t^* \in \mathbb{R}$ is an outcome variable of interest, $X_t \in \mathbb{R}^{d_x}$ is a vector of time-varying covariates, and $U_t \in \mathbb{R}$ is an unobserved error term. We assume that $g(x, \cdot)$ is strictly increasing for almost all $x \in Supp(X_t)$ for all $t = 1, 2, ..., T$ and that $\{U_t : t = 1, 2, ..., T\}$ is stationary. Since the quantile operator is preserved under a monotone transformation, it is straightforward to see that for any $u \in \mathcal{U} \subseteq (0,1)$,

$$Q_{Y_t^*|\mathbf{X}}(u|\mathbf{X}) = g(X_t, Q_{U_t|\mathbf{X}}(u|\mathbf{X}); u). \tag{3}$$

Note that the structural function g is allowed to vary across quantile levels.

It is common to assume that the unobserved error term U_t can be decomposed into time-invariant individual heterogeneity and time-varying idiosyncratic terms and that the time-invariant individual heterogeneity may be correlated with X_t. For the standard linear panel data model, such time-invariant heterogeneity can be eliminated by taking difference. For nonlinear models, however, the approach based on differencing does not work in general.

[8] Note that, however, I use u for the quantile level index throughout the paper, and thus u is not a realization of the random variable U_t in (1).

To overcome the difficulty in identification and estimation of the model with short panels, we adopt the CRE approach. Specifically, we assume that the conditional quantile function of U_t given $\mathbf{X}$ is an unknown function of $\mathbf{X}$. This is motivated by the CRE approach which was pioneered by Mundlak and Chamberlain (Mundlak (1978); Chamberlain (1980, 1982)). The CRE approach provides an effective way to deal with the unobserved heterogeneity in nonlinear panel models and it has been widely considered in the literature. Abrevaya and Dahl (2008) propose a linear panel quantile model, and Bester and Hansen (2009) investigate identification of marginal effects in a class of nonseparable panel models.[9] Both of them utilize some CRE approach to handle the unobserved individual effects with short panels. Arellano and Bonhomme (2016) recently develop a tractable estimation strategy for nonseparable panel data models based on the CRE approach.

The class of models in this paper is also related to the correlated random coefficient models in the literature (e.g. Arellano and Bonhomme (2012); Graham and Powell (2012); Laage (2019)). For quantile regression, Graham et al. (2018) consider linear panel quantile models with random coefficients, building upon Graham and Powell (2012). However, the model of this paper differs from those in that we consider a nonparametric structural function g with a scalar error term, whereas they consider a parametric structural function for g with a multi-dimensional error structure. Below we present some illustrative examples that fit into the class of CRE models in (3).

Example 2.1 (Random Coefficient Model)**.** *Suppose that $Supp(X_t) = \mathbb{R}$ and that the data generating process is as follows:*

$$Y_t^* = \exp(X_t)U_t.$$

It is obvious that $Q_{Y_t^*|\mathbf{X}}(u|\mathbf{X}) = \exp(X_t) \cdot Q_{U_t|\mathbf{X}}(u|\mathbf{X})$, and hence the structural function $g(x, \gamma; u) = x \cdot \gamma$ for all $u \in \mathcal{U} \subseteq (0, 1)$.

Example 2.2 (Linear Panel Quantile Model)**.** *Abrevaya and Dahl (2008) propose a class of linear panel quantile models as follows:*

$$Y_t^* = X_t'\beta(u) + \alpha(u) + \epsilon_t(u),$$
$$\alpha(u) = \mathbf{X}'\delta(u) + c(u),$$

where $\alpha(u)$ is an unobserved time-invariant heterogeneity, $c(u)$ is an unobserved error term, and $Q_{c(u)+\epsilon_t(u)}(u|\mathbf{X}) = 0$. It is straightforward to see that $Q_{Y_t^*|\mathbf{X}}(u|\mathbf{X}) = X_t'\beta(u) + \mathbf{X}'\delta(u)$ under the restriction on the model. This class of models is a special case of (3). Specifically, one can set $U_t = \alpha + \epsilon_t$ and $g(X_t, \gamma; u) = X_t'\beta(u) + \gamma$. The conditional quantile of U_t given $\mathbf{X}$ is equal to $\mathbf{X}'\delta(u)$.

[9] The class of models considered in this paper encompasses the linear panel quantile regression models in Abrevaya and Dahl (2008) as a special case, and thus it can be viewed as a nonparametric generalization of the linear panel quantile models with correlated random effects.

Example 2.3 (Panel Quantile Model)**.** *Arellano and Bonhomme (2016) consider the following model as an example:*

$$Y_t^* = X_t^{'}\beta(\epsilon_t) + \alpha\delta(\epsilon_t),$$
$$\alpha = \mathbf{X}^{'}\mu(V),$$

where for all $t = 1, 2, ..., T$, ϵ_t and V are uniformly distributed conditional on $\mathbf{X}$ and α is an unobserved time-invariant heterogeneity. As pointed out in Arellano and Bonhomme (2016), this model is a generalization of the standard linear quantile models of Koenker and Bassett (1978) to panel data. Assuming that the map $u \mapsto X_t^{'}\beta(u) + \mathbf{X}_i^{'}\mu(u) \cdot \delta(u)$ is strictly increasing and that ϵ_t and V are comonotonic, it can be shown that $Q_{Y_t^*|\mathbf{X}}(u|\mathbf{X}) = X_t^{'}\beta(u) + \mathbf{X}^{'}\theta(u)\delta(u)$.[10] Letting $U_t \equiv X_t^{'}\{\beta(\epsilon_t) - \beta(u)\} + \mathbf{X}^{'}\mu(V) \cdot \delta(\epsilon_t)$, $g(x, \gamma; u) = x^{'}\beta(u) + \gamma$ and $Q_{U_t|\mathbf{X}}(u) = \mathbf{X}^{'}\mu(u)\delta(u) \equiv \mathbf{X}^{'}\theta(u)$.

In examples 2.2 and 2.3, although there are two unobserved error terms, they can be collapsed into a scalar error term. While additivity plays the role of putting them together in example 2.2, comonotonicity of ϵ_t and V enables to collapse the error terms into a scalar error in example 2.3 where the unobserved error terms are nonlinearly enter. Therefore, the class of generalized CRE models in (3) is quite flexible and general.

Based on (2) and (3), we develop a panel quantile model with sample selection. Let $\Pr(D_t = 1|X_t = x, Z_t = z) \equiv p_t(x, z)$ be the propensity score (or selection probability), where $Z_t \in \mathbb{R}^{d_z}$ is a vector of excluded variables and $\mathbf{Z} = (Z_1, Z_2, ..., Z_T)^{'}$. The selection probability conditioning on X_t and Z_t is denoted by P_t (i.e. $P_t \equiv p_t(X_t, Z_t)$). The random vector $\mathbf{W} \equiv (\mathbf{Y}, \mathbf{X}, \mathbf{Z}, \mathbf{D})$ is observed from the data.

In the presence of sample selection, it is well-known that using only selected observations usually yields a sample selection bias, and thus it is necessary to correct such a bias. In this paper, we adopt the control function approach to correct the sample selection bias. The control function approach to sample selection was originally proposed by Heckman (1979), and it has been adapted to various models. Specifically, I impose the following assumption:

Assumption 1. *Let $u \in \mathcal{U}$ be given. For all $t \in \{1, 2, ., T\}$,*

$$Q_{U_t|\mathbf{X}, Z_t, D_t=1}(u|\mathbf{X}, Z_t, D_t = 1) = r(\mathbf{X}, p_t(X_t, Z_t); u), \tag{4}$$

where r is an unknown measurable function.

Note that r is allowed to take a different form across the quantile level u, and thus one can infer the conditional distribution function of U_t from the conditional quantile function of U_t, $r(\mathbf{X}, P_t; u)$. As a consequence, the way to correct the sample selection bias in this paper is to implicitly modify the conditional distribution function of the unobserved error

[10] For the definition of comonotonicity, one may refer to Koenker (2005, p.60).

term, and this is similar to that of Buchinsky (1998). However, it is different from the way that is considered in Arellano and Bonhomme (2017) or Chernozhukov et al. (2018) in that we do not impose any parametric or semiparametric structure on the conditional distribution of U_t. In sum, we consider the following model in this paper:

$$
\begin{aligned}
&Y_t^* = g(X_t, U_t), \\
&Y_t = D_t Y_t^*, \\
&Q_{U_t|\mathbf{X},Z_t,D_t=1}(u|\mathbf{X}, Z_t, D_t = 1) = r(\mathbf{X}, p_t(X_t, Z_t); u).
\end{aligned}
\tag{5}
$$

From (5), it is straightforward to see that

$$
\begin{aligned}
Q_{Y_t|\mathbf{X},Z_t,D_t=1}(u|\mathbf{X}, Z_t, D_t = 1) &= g(X_t, Q_{U_t|\mathbf{X},Z_t,D_t=1}(u|\mathbf{X}, Z_t, D_t = 1)) \\
&= g(X_t, r(\mathbf{X}, p_t(X_t, Z_t); u); u).
\end{aligned}
\tag{6}
$$

Assumption 1 shares a common feature with the models of Buchinsky (1998), Das et al. (2003) and Newey (2009) in that the selection bias is adjusted by including a control function, and this can be considered as a generalization of the control function approach in Heckman (1979). This paper, however, differs from Buchinsky (1998) in that neither any parametric restriction nor additivity of the error term is imposed on model (5).[11] Therefore, this paper extends the additive semiparametric quantile models for cross-sectional data in Buchinsky (1998) to nonseparable quantile regression models for panel data. Das et al. (2003) and Newey (2009) study identification and estimation of nonparametric sample selection models with an additive error term. This paper differs from them in that model (5) does not impose such an additive separability and thus it allows much various types of heterogeneity. In addition, this paper considers quantile regression, whereas they focus on conditional mean functions.

The structural functions g and r are related to many objects of interest. We first define the local structural function (LQSF) in time t as follows:

Definition 2.1 (Local Quantile Structural Function (LQSF)). *The local u-th quantile structural function (LQSF) at $(X_t, r(\mathbf{X}, P_t)) = (x, \gamma)$ in time t is*

$$
q_t^{local}(u, x, \gamma) \equiv Q_{Y_t^*|X_t, r(\mathbf{X},P_t)}(u|X_t = x, r(\mathbf{X}, P_t) = \gamma).
$$

Note that the definition of the LQSF is similar to, but slightly different from that of Fernández-Val et al. (2019). It is straightforward to see that the structural function $g(x, \gamma; u)$ in this paper corresponds to the LQSF. Related to the LQSF, one can consider the quantile structural function (QSF) which was introduced by Imbens and Newey (2009). The following definition of the QSF is a generalization of the QSF in Imbens and Newey

[11] Buchinsky (1998) considers a class of models where g is characterized by some finite-dimensional parameter. Specifically, his model is written as $g(X, \epsilon) = X^{'}\beta + \epsilon$.

(2009) to that for panel data:

Definition 2.2 (Quantile Structural Function (QSF))**.** *The u-th quantile structural function (QSF) in time t evaluated at $X_t = x$ is*

$$q_t(u, x) \equiv \mathbb{E}[q_t^{local}(u, x, r(\mathbf{X}, P_t))].$$

The LQSF and QSF are parameters of interest in many empirical analyses and closely related to the (local) quantile treatment effect of changing X_t. To make it concrete, we provide the definitions of the local and the average quantile treatment effects of X_t below:

Definition 2.3 (Local Quantile Treatment Effect (LQTE))**.** *The u-th local quantile treatment effect in time t of changing X_t from x_0 to x_1 at $r(\mathbf{X}, P_t) = \gamma$ is*

$$LQTE_t(u, x_0, x_1, \gamma) \equiv q_t^{local}(u, x_1, \gamma) - q_t^{local}(u, x_0, \gamma)$$

Definition 2.4 (Quantile Treatment Effect (QTE))**.** *The u-th average conditional quantile effect in time t of changing X_t from x_0 to x_1 is*

$$\begin{aligned} QTE_t(u, x_0, x_1) &\equiv \int q_t^{local}(u, x_1, r(\mathbf{x}, h(x, z))) - q_t^{local}(u, x_0, r(\mathbf{x}, h(x, z))) dF_{\mathbf{X}, Z_t}(\mathbf{x}, z) \\ &= q_t(u, x_1) - q_t(u, x_0). \end{aligned}$$

If X_t is continuous, then the LQTE and QTE can be interpreted as the local and average marginal effects, respectively. Many objects that are similar to the LQTE or the QTE are considered in the literature on nonseparable panel data models (e.g. Altonji and Matzkin (2005); Bester and Hansen (2009); Imbens and Newey (2009); Hoderlein and White (2012); Chernozhukov et al. (2013, 2015)). It is clear to see that the LQTE and QTE are functionals of the structural functions g and r from the definitions.

3 Nonparametric Identification

3.1 Main Results

In this section, we consider identification of the model parameters. The identification strategy is based on the model implication in (6), and the main objects of interest in (5) are $g(\cdot, \cdot; u)$ and $r(\cdot, \cdot; u)$. Note that the conditional selection probability at time t, $p_t(x, z)$, is identified from the data. As shown earlier, one can answer many questions that are empirically relevant, such as the marginal effect of X_t on the conditional quantile of Y_t^*, through identification of these objects. To achieve identification of $g(\cdot, \cdot; u)$ and $r(\cdot, \cdot; u)$, I impose the following assumption:

Assumption 2. *Let $T \geq 3$. For any $u \in \mathcal{U} \subseteq (0, 1)$, the following conditions hold:*

(i) *For each* $t = 1, 2, ..., T$, *there exists a known value* $\bar{x}(u) \in Supp(X_t) \subseteq \mathbb{R}^{d_x}$ *such that* $g(\bar{x}(u), \gamma; u) = \gamma$ *;*

(ii) *Let* $x \in Supp(X_t)$ *and* $\gamma \in Supp(r(\mathbf{X}, p_t(X_t, Z_t)))$ *be given. For any* $t, s \in \{1, 2, ..., T\}$ *with* $t \neq s$, *there exists a non-empty subset* $\tilde{\mathcal{X}}_{-t,s}(x, \bar{x}(u))$ *of* $Supp(\mathbf{X}_{-t,s}|(X_t, X_s) = (x, \bar{x}(u)))$ *and such that, for any* $\mathbf{x}_{-t,s} \in \tilde{\mathcal{X}}_{-t,s}(x, \bar{x}(u))$, $r(\mathbf{x}_0, p) = \gamma$ *for some* $p \in Supp(p_t(x, Z_t)|\mathbf{X} = \mathbf{x}_0)$ *and* $\Pr(\mathbf{X}_{-t,s} \in \tilde{\mathcal{X}}_{-t,s}(x, \bar{x}(u))) > 0$, *where* $\mathbf{x}_0 = (X_t = x, X_s = \bar{x}(u), \mathbf{X}_{-t,s} = \mathbf{x}_{-t,s})$*;*

(iii) *For any* $t = 1, 2, ..., T$ *and for any* $x \in Supp(X_t)$ *and* $z \in Supp(Z_t) \subseteq \mathbb{R}^{d_z}$, *there exists a non-empty set* $\mathcal{Z}_s(z) \subseteq Supp(Z_s)$ *for some* $s \in \{1, 2, ..., T\}$ *such that, for any* $\tilde{z} \in \mathcal{Z}_s(z)$, $p_t(x, z) = p_s(\bar{x}(u), \tilde{z})$ *and* $\Pr(Z_s \in \mathcal{Z}_s(z)) > 0$.

Condition (i) is a normalization. Theorem 3.1 in Matzkin (2007) implies that it is necessary to impose a normalization to identify function $g(\cdot, \cdot; u)$. Note that the value $\bar{x}(u)$ may differ across the quantile indices. We assume that $\bar{x}(u)$ remains the same across $u \in \mathcal{U}$ for simplicity.

Condition (ii) is implied by sufficient variation in $\mathbf{X}_{-t,s}$. This variable can be viewed as an excluded variable that provides a source of exogenous variation to r while fixing X_t and X_s for some t and s. To illustrate how this condition is used for identification, consider the linear panel quantile model in Example 2.2 with assuming that $T = 3$ and $d_x = 1$. In addition, we ignore the sample selection issues, and therefore Y_t^* is observed for everyone, to elucidate the role of condition (ii) in identification analysis. Note that $\bar{x}(u) = 0$, $g(x, \gamma) = x\beta(u) + \gamma$, and $r(\mathbf{x}) = \mathbf{x}'\delta(u)$. Let $x \in Supp(X_1)$ be given. Then, one can show that $Q_{Y_1^*|\mathbf{X}}(u|\mathbf{X}_0) = x\beta(u) + r(\mathbf{X}_0)$ and that $Q_{Y_2^*|\mathbf{X}}(u|\mathbf{X}_0) = r(\mathbf{X}_0) = x\delta_1(u) + X_3\delta_3(u)$, where $\delta(u) = (\delta_1(u), \delta_2(u), \delta_3(u))'$ and $\mathbf{X}_0 = (x, 0, X_3)'$. Condition (ii) ensures that one can find a set of values of X_3 such that for a given $\gamma \in Supp(r(\mathbf{X}))$, $x\delta_1(u) + X_3\delta_3(u) = \gamma$. Therefore, a necessary condition in this illustration that guarantees condition (ii) in Assumption 2 is that $\delta_3(u) \neq 0$. If X_3 has enough variation conditioning on $X_1 = x$ and $X_2 = 0$ and $\delta_3(u) \neq 0$, then condition (ii) for this example is satisfied. This is similar to Assumption 2 in Imbens and Newey (2009) that a large support condition for the excluded variable is satisfied. Since the model in this paper allows for time-invariant endogeneity and it is captured by the CRE specification, X_t can be considered endogenous (in the time-invariant manner), and the covariates in other time periods are used as an excluded variable that helps resolve the time-invariant endogeneity. It is worth pointing out that more than three time periods gives additional exogenous variation that can be used to identify the structural functions and therefore having more than 3 time periods provide additional identification power.

A related assumption to condition (ii) is exchangeability considered by Altonji and Matzkin (2005). Exchangeability typically places some restriction on the admissible class of functions for r, it may help weaken assumptions on variation in the excluded variable.

Compared to the exchangeability assumption, condition (ii) is likely to require stronger conditions on the support of $\mathbf{X}_{-t,s}$, but it does not impose any shape restrictions on the class of functions that r belongs to. More importantly, exchangeability may not be plausible to be assumed with panel data where t indicates time. Exchangeability is related to symmetry of the effects of covariates on the distribution of the unobserved error term, and thus the effect of a change in X_s is the same (or similar) to that of a change in X_t for some $t \neq s$. In this regard, exchangeability may be consistent with some variants of the form in Mundlak (1978) in a sense that the correlated random effects specification of Mundlak (1978) is the average of X_t's over time and therefore the effects of X_t and X_s with $t \neq s$ are symmetric. On the other hand, I do not impose such restrictions on the model so that one can consider more flexible specifications for r. One can refer to Altonji and Matzkin (2005, pp.1062-1066) for further discussion on the exchangeability condition and CRE approach.

Condition (iii) requires variation in the excluded variable Z_t, which is an instrumental variable. This condition also requires that the excluded variable Z_t affect the selection probability, so one can use the variation in Z_t and Z_s to match the selection probabilities in time periods t and s. This condition is needed to deal with the endogenous selection. For illustration, suppose that $D_t = \mathbf{1}(X_t\zeta + Z_t\pi \geq \nu_t)$, where $\nu_t \sim N(0,1)$, $(X_t, Z_t) \perp \nu_t$ and $d_x = d_z = 1$. Then, for given $x \in Supp(X_t)$, $\Pr(D_t = 1|X_t = x, Z_t) = \Phi(x\zeta + Z_t\pi)$ and $\Pr(D_s = 1|X_s = \bar{x}(u), Z_s) = \Phi(\bar{x}(u)\zeta + Z_s\pi)$, where Φ is the standard normal distribution function. In this case, condition (iii) in Assumption 2 is satisfied if $\pi \neq 0$ and variation in either Z_t or Z_s is large enough. The former condition $\pi \neq 0$ corresponds to the standard relevance condition for instrumental variables, and such relevance conditions are usually required for nonparametric identification with endogeneity. The latter condition which is about variation in Z_t (or Z_s) is similar to the large support condition in Imbens and Newey (2009). Similar assumptions to condition (iii) can be found in, for example, Altonji and Matzkin (2005) and Vytlacil and Yildiz (2007).

An informal description of the identification strategy in this paper is as follows: Fixing $X_t = x$, the information in time period s, together with the normalization, is used to derive an expression for r at $(X_t', X_s', \mathbf{X}_{-t,s}')' = (x', \bar{x}(u)', \mathbf{X}_{-t,s}')'$ and $p_s(\bar{x}(u), Z_s)$. Then, one can utilize the variation in $\mathbf{X}_{-t,s}$, Z_t, and Z_s to find values $\mathbf{x}_0$ and p such that $r(\mathbf{x}_0, p) = \gamma \in Supp(r(\mathbf{X}, h(X_t, Z_t)))$. Taking average over such values yields identification of $g(x, \gamma; u)$ over $Supp(X_t, r(\mathbf{X}, h_t(X_t, Z_t)))$ for each $t = 1, 2, ..., T$. Then, one can identify $r(\mathbf{x}, p)$ for all $(\mathbf{x}, p) \in Supp(\mathbf{X}, h_t(X_t, Z_t))$ by taking average of the inverse map of g conditional on $\mathbf{X} = \mathbf{x}$ and $p_t(X_t, Z_t) = p$. The following theorem demonstrates that Assumption 2 are sufficient for identification of $g(\cdot, \cdot; u)$ and $r(\cdot, \cdot; u)$ in (6).

Theorem 3.1. *Let $u \in \mathcal{U}$ be given. Suppose that Assumptions 1 and 2 hold. Then, for each $t = 1, 2, ..., T$, $g(\cdot, \cdot; u)$ is over $Supp(X_t, r(\mathbf{X}, p_t(X_t, Z_t)))$. Furthermore, $r(\cdot, \cdot; u)$ is identified over the set $\cup_{t=1}^{T} Supp(\mathbf{X}, p_t(X_t, Z_t))$.*

Proof. I drop u in functions g and r for simplicity of notation. Let $\mathbf{X}_0 = (X_t = x, X_s =$

$\bar{x}, \mathbf{X}_{-t,s})$. Note that $p_t(X_t, Z_t)$ is directly identified from the data. Then, one can show that under Assumption 2,

$$Q_{Y_t|\mathbf{X},Z_t,D_t=1}(u|\mathbf{X}_0, Z_t) = g(x, r(\mathbf{X}_0, P_{t0})),$$
$$Q_{Y_s|\mathbf{X},Z_s,D_s=1}(u|\mathbf{X}_0, Z_s) = r(\mathbf{X}_0, P_{s0}),$$

where $P_{t0} \equiv \Pr(D_t = 1|X_t = x, Z_t)$ and $P_{s0} \equiv \Pr(D_s = 1|X_s = \bar{x}, Z_s)$. Let $\gamma \in \mathbb{R}$ be given, then it is straightforward to see that

$$g(x, \gamma) = \mathbb{E}[Q_{Y_t|\mathbf{X},Z_t,D_t=1}(u|\mathbf{X}_0, Z_t)|Q_{Y_s|\mathbf{X},Z_s,D_s=1}(u|\mathbf{X}_0, Z_s) = \gamma, P_{t0} = P_{s0}]. \quad (7)$$

The conditioning event in (7) has a positive measure by conditions (ii) and (iii) in Assumption 2, and therefore $g(x, \gamma)$ is identified.

Since $g(\cdot, \cdot; u)$ is assumed to be strictly monotone in its second argument, there exists the inverse mapping with respect to the second argument. From (6), one obtains that

$$r(\mathbf{X}, P_t) = g^{-1}(X_t, Q_{Y_t|\mathbf{X},Z_t,D_t=1}(u|\mathbf{X}, Z_t)),$$

where $g^{-1}(x, y)$ is the inverse mapping of $g(x, e)$ with respect to e that is identified.[12] For any $(\mathbf{x}, p) \in Supp(\mathbf{X}, P_t)$, one obtains that

$$r(\mathbf{x}, p) = \mathbb{E}[g^{-1}(x, Q_{Y_t|X,Z,D_t=1}(u|\mathbf{X}, \mathbf{Z}))|\mathbf{X} = \mathbf{x}, p_t(x, Z_t) = p],$$

and this establishes identification of $r(\mathbf{x}, p)$ over $Supp(\mathbf{X}, p_t(X_t, Z_t))$ for each $t = 1, 2, ..., T$.

∎

The LQTE and QTE are objects that may be important and relevant to policy evaluation. These objects are functionals of the structural functions g, r, and h, and thus they are identified once those structural functions are identified. The following corollary demonstrates that the LQTE and QTE are identified under the same set of conditions for identification of the structural functions.

Corollary 3.2. *Suppose that the conditions in Theorem 3.1 hold. Let $t \in \{1, 2, ..., T\}$ and $x_0, x_1 \in Supp(X_t)$ be given. Then, for any $\gamma \in Supp(r(\mathbf{X}, h_t(X_t, Z_t); u))$, $LQTE_t(u, x_0, x_1, \gamma)$ is identified. In addition, $QTE(u, x_0, x_1)$ is also identified.*

Proof. Recall that $m(x, \gamma; u) = q_t^{local}(u, x, \gamma)$. Since m and r are identified by Theorem 3.1, one obtains that

$$\begin{aligned} LQTE_t(u, x_0, x_1, \gamma) &\equiv q_t^{local}(u, x_1, \gamma) - q_t^{local}(u, x_0, \gamma) \\ &= \mathbb{E}[g(x_1, r(\mathbf{X}, p_t(X_t, Z_t); u); u) - g(x_0, r(\mathbf{X}, p_t(X_t, Z_t); u); u)|r(\mathbf{X}, p_t(X_t, Z_t); u) = \gamma]. \end{aligned}$$

[12] $y = g(x, e)$ if any only if $e = g^{-1}(x, y)$.

Note that the conditioning event is of a positive probability because $\gamma \in Supp(r(\mathbf{X}, p_t(X_t, Z_t); u))$, and thus $LQTE_t(u, x_0, x_1, \gamma)$ is identified. Similarly, it follows from the definition of $QTE_t(u, x_0, x_1)$ that

$$QTE_t(u, x_0, x_1) = \mathbb{E}[g(x_1, r(\mathbf{X}, p_t(X_t, Z_t); u); u) - g(x_0, r(\mathbf{X}, p_t(X_t, Z_t); u); u)],$$

where the expectation is taken over $Supp(r(\mathbf{X}, p_t(X_t, Z_t); u))$. Therefore, $QTE_t(u, x_0, x_1)$ is also identified. ■

The CRE approach was also adopted by Bester and Hansen (2009) and Arellano and Bonhomme (2016), and they require that T to be greater than or equal to 3 for identification. The identification strategy in this paper, however, is different from theirs. Specifically, Bester and Hansen (2009) focus on the marginal effects of continuous covariates without completely specifying the data generating process. They use a derivative argument for identification of the marginal effects. In contrast, we focus on identification of the structural functions with specifying the data generating process (equation (1)), and the identification strategy in this paper is to use variation in excluded variables. The marginal effects in this paper are also identified as a by-product (Corollary 3.2). Arellano and Bonhomme (2016) consider nonparametric identification of structural functions, but the identification strategy in Theorem 3.1 is different from that of Arellano and Bonhomme (2016). Specifically, Arellano and Bonhomme (2016) use a high-level assumption, called an injectivity condition, and this condition resembles completeness conditions that are commonly used in the literature on nonparametric identification (e.g. Newey and Powell (2003) and Blundell et al. (2007)). The injectivity condition, however, is relatively difficult to interpret and verify in practice. More importantly, estimation and inference may suffer from an ill-posed inverse problem which leads to a slower convergence rate. On the other hand, the identification strategy in this paper does not rely on completeness conditions, and hence it is not subject to an ill-posed inverse problem.

The identification strategy in Theorem 3.1 does not require to specify the distribution of the unobserved error term. In contrast, Arellano and Bonhomme (2017) and Chernozhukov et al. (2018) consider some semiparametric specification of the joint distribution of Y^* and D. Furthermore, both papers focus on quantile regression models for cross-sectional data, whereas this paper considers models for panel data.

The implication of model (equation (6)) is similar to that of Lewbel and Linton (2007) or Escanciano et al. (2016), and hence the identification strategy of this paper shares some common with their strategies. However, the models of the papers are different from that of this paper. Specifically, the model of Lewbel and Linton (2007) differs from (6) in that they assume that X_t is excluded from $\mathbf{X}$ and that the selection probability does not depend on X_t. Therefore, their identification strategy cannot be directly applied to identify g and r in (6). Escanciano et al. (2016) study a class of models where there are two index functions and g relates these two index functions.[13] The focus of Escanciano et al. (2016) is on identification and estimation of the finite-dimensional parameter in one of the index functions, but this paper studies nonparametric identification and estimation of (6) without specifying an index function for X_t, which allows for more flexibility of the model.

While the linear correlated random coefficients models allow for multi-dimensional error terms, the identification comes at cost of a larger (but fixed) number of time periods (e.g. Arellano and Bonhomme (2012); Graham and Powell (2012); Graham et al. (2018); Laage (2019)). In contrast, this paper imposes a scalar error term, but identification requires T be greater than equal to 3 with some support condition. The requirement for T is much weaker than that in the correlated random coefficients models where T should be greater than or equal to the number of covariates. In addition, the model in this paper is completely nonparametric, whereas most of correlated random coefficients models are parametric or semiparametric.

3.2 Extensions

In this section, we discuss some extensions of the panel quantile models with sample selection in Section 2. We consider (i) endogeneity of X_t and (ii) censoring, which are useful and relevant to many empirical situations. We show that model (5) can be easily extended to incorporate these issues.

3.2.1 Endogenous Regressors

Endogeneity issues are prevalent in many empirical questions. The CRE specification effectively captures "time-invariant" endogeneity, but some regressors may exhibit "time-varying" endogeneity.[14] The model implication in (6) is closely related to the control function approach to deal with sample selection bias, and it can be extended to allow for endogeneity of X_t. To make it concrete, suppose that $X_t = (X_t^{e'}, \tilde{X}_t')'$, where X_t^e is a vector of endogenous regressors and $\tilde{X}_t$ is a vector of exogenous regressors. For brevity of the model, we assume that $X_t^e \in \mathbb{R}$, but it can be easily extended to the case where X_t^e is a vector. Assume that and $Z_t = (Z_{1t}', Z_{2t}')'$, and consider the following class of models:

$$\begin{aligned}
&Y_t^* = g(X_t, U_t),: \\
&Y_t = D_t Y_t^*, \\
&X_t^e = q(Z_{2t}, V_t), \\
&Q_{U_t|\mathbf{X},\mathbf{Z},D_t=1}(u|\mathbf{X}, Z_t, D_t = 1) = r^e(\tilde{\mathbf{X}}, V_t, p_t(X_t, Z_{1t}); u),
\end{aligned} \tag{8}$$

[13] One of the index functions is a linear-index function, and the other one is a known function from data.

[14] As mentioned earlier, the CRE specification is closely related to the dependence between the regressors and time-invariant unobserved heterogeneity, which is commonly assumed in fixed effects models. On the other hand, I use the term time-varying endogeneity to allow for dependence between the regressors and time-varying components of the error term U_t. See also Laage (2019).

where $V_t \in \mathbb{R}$ is unobserved and independent of Z_{2t}, and $q(z_2, v)$ is a non-trivial function of z_2 and strictly increasing in v for all z_2. Without loss of generality, V_t is assumed to be uniformly distributed on the unit interval, conditional on Z_{2t}. Model (8) is closely related to the sample selection model with endogeneity that is studied by Das et al. (2003). The conditional quantile restriction on U_t in model (8) implies that, conditional on V_t, X_t^e is no longer endogenous, and hence the following model implication is obtained:

$$Q_{Y_t|\mathbf{X},\mathbf{Z},D_t=1}(u|\mathbf{X},\mathbf{Z},D_t=1) = g(X_t, r^e(\tilde{\mathbf{X}}, V_t, p_t(X_t, Z_{1t}); u); u). \tag{9}$$

This extends the control function approach to handle the sample selection bias that is presented in (6) to a more general case where some regressors are endogenous. The variable V_t plays the role of a control function to handle endogeneity of X_t^e, and needs to be estimated in the first-stage. Since the model implication (6) suggests that the selection probability $p_t(X_t, Z_t)$ plays the role of control function to correct selection bias, the roles of $p_t(X_t, Z_t)$ and V_t are almost the same. Similar approaches for cross-sectional data models are considered by, for example, Newey et al. (1999), Lee (2007), Imbens and Newey (2009), and Chernozhukov et al. (2015). For panel data models, Semykina and Wooldridge (2010) develop a class of models that is similar to (9), but their focus is on the conditional mean function with additively separable error terms.

3.2.2 Censoring

Censoring is an issue that empirical researchers frequently face. I consider the following censoring rule with sample selection:

$$\begin{aligned} Y_t^* &= g(X_t, U_t), \\ Y_t &= D_t \cdot \max(Y_t^*, C_t), \end{aligned}$$

where C_t's are fixed constants. Since the quantile operator is preserved under monotone transformations, quantile regression models can easily incorporate censored data in a similar fashion of Powell (1986). Specifically, one can show that

$$Q_{Y_t|\mathbf{X},Z_t,D_t=1}(u|\mathbf{X},Z_t,D_t=1) = \max(Q_{Y_t^*|\mathbf{X},Z_t,D_t=1}(u|\mathbf{X},Z_t,D_t=1), C_t),$$

and the conditional quantile function of Y_t^* is the same as (6). It is worth noting that one can simultaneously incorporate endogeneity and censoring, as in Chernozhukov et al. (2015). This result can be further extended to the Tobit type-3 model considered by Fernández-Val et al. (2019). The model of Fernández-Val et al. (2019) is different from that of this paper in that the selection rule in their model is not binary and that the error term in the outcome equation can be multi-dimensional, and their identification strategy relies on the control function approach of Imbens and Newey (2009). Nevertheless, the model in this paper can incorporate such a class of selection rules with the control function approach as shown earlier.

4 Dynamic Models

Dynamic panel data models are of interest in many empirical situations. Sample selection issues in panel data models have been addressed by several papers including Kyriazidou (2001), Gayle and Viauroux (2007), and Semykina and Wooldridge (2013). Including lagged dependent variables as regressors usually causes many problems. For example, the within-estimator or first-differencing estimator in the standard linear fixed effects model may not be consistent when lagged dependent variables are included as regressors. It is much harder to resolve such issues when the model is nonlinear. In this section, we demonstrate that the CRE approach combined with a control function approach can effectively be incorporated into dynamic quantile regression with sample selection.

We consider the following class of dynamic autoregressive models:

$$\begin{aligned} Y_t^* &= g(Y_{t-1}^*, X_t, U_t), \\ Y_t &= D_t Y_t^*, \\ \Pr(D_t = 1|D_{t-1}, X_t, Z_t) &= p_t(D_{t-1}, X_t, Z_t), \end{aligned} \tag{10}$$

for $t = 2, 3, ..., T$, where $g(y, x, \cdot)$ is strictly increasing for almost all y and x. For $t = 1$, it is assumed that $Y_1 = g(X_1, U_1)$ and that $\Pr(D_1 = 1|X_1, Z_1) = p_1(X_1, Z_1)$. I adopt the idea of Wooldridge (2005) to formulate flexible and tractable dynamic quantile regression models using the CRE approach. Specifically, it is assumed that the conditional quantile of U_t given the lagged dependent variables, $\mathbf{X}$, $\mathbf{Z}$, and the selected for two adjacent time periods is a function of the conditioning variables and selection probabilities at time periods 1 and t. This assumption is similar to that on the dynamic model in Bester and Hansen (2009) or Arellano and Bonhomme (2016). To introduce a formal assumption, time period 1 is treated as the initial time period, and let $\mathbf{W}^t$ be the vector that collects all W_s's up to time period t, where W_s is a generic random variable at time s (i.e., $\mathbf{W}^t \equiv (W_1, W_2, ..., W_t)'$). Note that $\mathbf{W}^T = \mathbf{W}$.

Assumption DM.1. *Let $u \in \mathcal{U} \subseteq (0, 1)$ be given. For all $t \in \{2, ..., T\}$,*

$$Q_{U_t|\mathbf{Y}^{*,t-1},\mathbf{X},\mathbf{Z},D_t\cdot D_{t-1}\cdot D_1=1}(u|\mathbf{Y}^{*,t-1}, \mathbf{X}, \mathbf{Z}) = r(Y_1, \mathbf{X}, P_t, P_1; u), \tag{11}$$

where $P_t \equiv p_t(D_{t-1}, X_t, Z_t)$ for $t \geq 2$ and $P_1 \equiv \Pr(D_1 = 1|X_1, Z_1) = h_1(X_1, Z_1)$.

Assumption DM.1 is a dynamic generalization of Assumption 1. In the absence of sample selection, this assumption can be reduced to that the unknown function r does not depend on the selection probabilities at time 1 and t, which is similar to the CRE specification of the dynamic model in Bester and Hansen (2009) or Arellano and Bonhomme (2016). It is also worth mentioning that r is a function of Y_0 and P_1 as well as $\mathbf{X}$ and P_t. Conditioning on $D_1 = 1$ is necessary to observe Y_1 that is included in the CRE specification with lagged dependent variables, and this is why P_1 appears as a correction term in equation (11).

Below we provide a set of conditions under which the structural functions are nonparametrically identified. For simplicity, we assume that the support of Y_t for each $t = 1, 2, ..., T$ is the same.

Assumption DM.2. *Let $T \geq 4$. For any $u \in \mathcal{U} \subseteq (0, 1)$, the following conditions hold:*

(i) For each $t = 1, 2, ..., T$, there exists a known value $(\bar{y}(u), \bar{x}(u)')' \in Supp\left((Y_{t-1}, X_t')'\right) \subseteq \mathbb{R}^{1+d_x}$ such that $g(\bar{y}(u), \bar{x}(u), \gamma; u) = \gamma$;

(ii) Let $y_1 \in Supp(Y_1)$, $x_1 \in Supp(X_1)$, and $p_1 \in Supp(h_1(X_1, Z_1))$ be given. For $t \neq 1$, let $x \in Supp(X_t|X_1 = x_1)$ and $\gamma \in Supp(r(Y_1, \mathbf{X}, h_t(1, X_t, Z_t), h_1(X_1, Z_1); u))$ be given. For $s \in \{2, 3, ..., T\}$ with $s \neq t$, there exists a non-empty subset $\tilde{\mathcal{X}}_{-t,s,1}(x_1, x, \bar{x}(u))$ of $Supp(\mathbf{X}_{-t,s,1}|(X_1, X_t, X_s) = (x_1, x, \bar{x}(u)), D_t D_{t-1} D_1 = 1)$ such that, for any $p \in Supp(h_t(1, x, Z_t)|\mathbf{X} = \tilde{\mathbf{x}}_{-t,s,1})$, there exists $\mathbf{x}_{-t,s,1} \in \tilde{\mathcal{X}}_{-t,s,1}(x_1 x, \bar{x}(u))$ such that $r(y_1, \mathbf{x}_0, p, p_1) = \gamma$, and $\Pr(\mathbf{X}_{-t,s,1} \in \tilde{\mathcal{X}}_{-t,s,1}(x_1, x, \bar{x}(u))) > 0$, where $\mathbf{x}_0 = (X_1 = x_1, X_t = x, X_s = \bar{x}(u), \mathbf{X}_{-t,s,1} = \mathbf{x}_{-t,s,1})$;

(iii) Let $y_1 \in Supp(Y_1)$, $x_1 \in Supp(X_1)$, $p_1 \in Supp(P_1)$ be given. For $t \neq 1$, let $x \in Supp(X_t)$ be given. and for any $x \in Supp(X_t)$ and $z \in Supp(Z_t) \subseteq \mathbb{R}^{d_z}$, there exists a non-empty set $\mathcal{Z}_s(z) \subseteq Supp(Z_s|Z_t = z)$ for some $s \in \{1, 2, ..., T\}$ such that, for any $\tilde{z} \in \mathcal{Z}_s(z)$, $h_t(1, x, z) = h_s(1, \bar{x}(u), \tilde{z})$, and $\Pr(Z_s \in \mathcal{Z}_s(z)) > 0$

Assumption DM.2 is corresponding to Assumption 2 in Section 2 that imposes identifying conditions for static models, While it is needed that $T \geq 4$ to identify the parameters in static models, Assumption DM.2 requires that $T \geq 4$ identification. This is due to the presence of a lagged dependent variable, which results in that, under Assumption DM.1, it is not possible to utilize variation in X_1 with fixing Y_1. One can interpret it as that the information in time period 1 is treated as given or an initial condition. The identification strategy for dynamic models is similar to that for static models, but it is required that individuals be selected for two consecutive time periods. This is due to the fact that the structural function in the outcome equation depends on a lagged dependent variable, and therefore it is needed to observe the values of outcome variables for two adjacent time periods in order for the structural functions to be well-defined.

Theorem 4.1. *Let $u \in \mathcal{U} \subseteq (0, 1)$ be given and suppose that Assumptions DM.1 and DM.2 hold. Then, the structural functions g and r are nonparametrically identified over $\cup_{t=2}^{T} Supp(X_t, R_t)$ and $\cup_{t=2}^{T} Supp(Y_1, \mathbf{X}, P_t, P_1)$, respectively, where $R_t \equiv r(Y_1, \mathbf{X}, h_t(1, X_t, Z_t), h_1(X_1, Z_1); u)$, $P_t \equiv h_t(D_{t-1}, X_t, Z_t)$ for $t \geq 2$ and $P_1 = \Pr(D_1 = 1|X_1, Z_1)$.*

Proof. The proof is almost identical to that of Theorem 3.1. Let $y_1 \in Supp(Y_1)$, $x_1 \in Supp(X_1)$, and $p_1 \in Supp(h_1(X_1, Z_1))$ be given. For any $t \neq 1$, pick any $x \in Supp(X_t|X_1 = x_1)$, $\gamma \in Supp(R_t)$, and $s \in \{2, 3, ..., T\} - \{t\}$. Define $\mathbf{X}_0 \equiv (X_1 = x_1, X_t = x, X_s = \bar{x}(u), \mathbf{X}_{-t,s,1})$, and consider

$$\begin{aligned} &Q_{Y_t|\mathbf{Y}^{t-1},\mathbf{X},\mathbf{Z},D_tD_{t-1}D_1=1}(u|(y_1, Y_2, ..., y), \mathbf{X_0}, \mathbf{Z}) = g(y, x, r(y_1, \mathbf{X}_0, h_t(1, x, Z_t), P_1)), \\ &Q_{Y_s|\mathbf{Y}^{s-1},\mathbf{X},\mathbf{Z},D_sD_1=1}(u|(y_1, Y_2, ..., \bar{y}(u)), \mathbf{X_0}, \mathbf{Z}) = r(y_1, \mathbf{X}_0, h_s(1, \bar{x}(u), Z_s), P_1). \end{aligned}$$

By Assumption DM.2-(iii), one obtains $\Pr\left(h_t(1, X_t, Z_t) = h_s(1, X_s, Z_s)|X_t = x, X_s = \bar{x}(u)\right) > 0$. By Assumption DM.2-(ii), $\Pr\left(r(y_1, \mathbf{X}_0, P_t, P_1) = \gamma|Z_t = z, Z_s \in \mathcal{Z}_s(z)\right) > 0$, where $\mathcal{Z}_s(z)$ is defined in Assumption DM.2-(iii). Therefore, $g(y, x, \gamma)$ is identified through the following conditional expectation:

$$\begin{aligned} &g(y, x, \gamma) \\ =&\mathbb{E}\Big[Q_{Y_t|\mathbf{Y}^{t-1},\mathbf{X},\mathbf{Z},D_tD_{t-1}D_1=1}(u|(y_1, Y_2, ..., y), \mathbf{X_0}, \mathbf{Z}) \\ &\quad |Q_{Y_s|\mathbf{Y}^{s-1},\mathbf{X},\mathbf{Z},D_sD_1=1}(u|(y_1, Y_2, ..., \bar{y}(u)), \mathbf{X_0}, \mathbf{Z}) = \gamma, h_t(1, x, Z_t) = h_s(1, \bar{x}(u), Z_s)\Big]. \end{aligned}$$

Let $p \in Supp(h_t(X_t, Z_t))$ and pick any $z \in Supp(Z_t|X_t = x)$ such that $p = h_t(x, z)$. Then, one obtains that

$$g(y, x, r(y_1, \mathbf{x}, p, p_1)) = \mathbb{E}\left[Q_{Y_t|\mathbf{Y}^{t-1},\mathbf{X},\mathbf{Z},D_tD_{t-1}D_1=1}(u|(y_1, Y_2, ..., y), \mathbf{x}, \mathbf{Z}_0)\right],$$

where $\mathbf{x} = (x_1, x_2, ..., x_T)' \in Supp(\mathbf{X})$, $\mathbf{Z}_0 = (Z_1, Z_2, ..., Z_{t-1}, z, Z_{t+1}, ..., Z_T)' \in Supp(\mathbf{Z})$. Since $g(\cdot, \cdot, \cdot)$ is strictly increasing in its last argument, identification of r can be achieved by inverting the structural function m in the above equation. ■

Remark 4.2. *One can include a lag of covariates into the outcome equation in the same way to Bester and Hansen (2009). In such cases, the requirement on the number of time periods should be strengthened. For example, when the structural function m depends on Y_{t-1}, X_t, X_{t-1}, and U_t, it is required that $T \geq 5$ for nonparametric identification, with some proper normalization condition.*

5 Semiparametric Models

While fully nonparametric models are attractive as they are robust to model misspecification, one important and practical issue is that it is difficult to estimate parameters in them when the dimension of covariate is large. Although the CRE approach allows us to consider flexible and general models, the number of covariates involved in estimating parameters can be very large and thus the fully nonparametric model presented in the previous section may not be practically useful. To address such issues, we propose some semiparametric models and study identification of the parameters in them. The semiparametric models considered in this section are static, but their dynamic extensions are straightforward.

5.1 Index Models

One can impose an index structure on the structural function r , and this is originally motivated by the original CRE approach of Mundlak (1978) and Chamberlain (1980). Specifically, we impose the following assumption.

Assumption 1′. *Let $u \in \mathcal{U}$ be given. For all $t \in \{1, 2, ..., T\}$,*

$$Q_{U_t|\mathbf{X},Z_t,D_t=1}(u|\mathbf{X}, Z_t, D_t = 1) = r(\mathbf{X}'\psi(u), P_t), \tag{12}$$

where $\psi(u) = \left(\psi_1(u)', \psi_2(u)', ..., \psi_T(u)'\right)'$ and $\psi_t(u) \in \mathbb{R}^{d_x}$ for all t.

The index structure is consistent with the CRE specification of Chamberlain (1980) and reduces the dimension of the structural function r, while allowing for nonseparability between the index and P_t. This semiparametric specification requires additional assumptions for identification of the index coefficient vector ψ, and these assumptions depend on X_t being continuous or discrete. To make the role of each type of regressor, we assume that $X_t = (X_t^{c'}, X_t^{d'})'$, where $X_t^c \in \mathbb{R}^{d_{xc}}$ is a vector of continuous regressors and $X_t^d \in \mathbb{R}^{d_{xd}}$ is a vector of discrete regressors. Consequently, we partition the coefficient ψ into two parts, $\psi_t = \left((\psi_t^c)', (\psi_t^d)'\right)'$, where $\psi_t^c = (\psi_{1t}^c, \psi_{2t}^c, ..., \psi_{d_{xc}t}^c)'$ and $\psi_t^d = (\psi_{1t}^d, \psi_{2t}^d, ..., \psi_{d_{xd}t}^d)'$.

Assumption 2′. *Let $T \geq 3$. For any $u \in \mathcal{U} \subseteq (0, 1)$, the following conditions hold:*

(i) For each $t = 1, 2, ..., T$, there exists a known value $\bar{x}(u) \in Supp(X_t) \subseteq \mathbb{R}^{d_x}$ such that $g(\bar{x}(u), \gamma; u) = \gamma$;

(ii) Let $x \in Supp(X_t)$ and $\gamma \in Supp(r(\mathbf{X}'\psi, h(X_t, Z_t)))$ be given. For any $t, s \in \{1, 2, ..., T\}$ with $t \neq s$, there exists a non-empty subset $\tilde{\mathcal{X}}_{-t,s}^S(x, \bar{x}(u))$ of $Supp(\mathbf{X}_{-t,s}'\psi_{-t,s}|(X_t, X_s) = (x, \bar{x}(u)))$ and such that, for any $\mathbf{x}_{-t,s} \in \tilde{\mathcal{X}}_{-t,s}^S(x, \bar{x}(u))$, $r(\mathbf{x}_0'\psi, p) = \gamma$ for some $p \in Supp(p_t(x, Z_t)|\mathbf{X} = \mathbf{x}_0)$ and $\Pr(\mathbf{X}_{-t,s} \in \tilde{\mathcal{X}}_{-t,s}(x, \bar{x}(u))) > 0$, where $\mathbf{x}_0 = (X_t = x, X_s = \bar{x}(u)), \mathbf{X}_{-t,s} = \mathbf{x}_{-t,s})$;

(iii) For any $t = 1, 2, ..., T$ and for any $x \in Supp(X_t)$ and $z \in Supp(Z_t) \subseteq \mathbb{R}^{d_z}$, there exists a non-empty set $\mathcal{Z}_s(z) \subseteq Supp(Z_s)$ for some $s \in \{1, 2, ..., T\}$ such that, for any $\tilde{z} \in \mathcal{Z}_s(z)$, $p_t(x, z) = p_s(\bar{x}(u), \tilde{z})$ and $\Pr(Z_s \in \mathcal{Z}_s(z)) > 0$.

(iv) $\psi_{1t}^c = 1$ for all t.

(v) $g(\cdot, \cdot)$ is differentiable with respect to the second argument, and $r(\cdot, \cdot)$ is differentiable with respect to the first argument.

(vi) $r(\cdot,\cdot)$ is invertible with respect to the first argument.

Conditions (i), (ii), and (iii) in Assumption $2'$ are almost the same as conditions (i), (ii), and (iii) in Assumption 2, respectively. Condition (iv) – (vi) in Assumption 2 are additionally imposed to identify the finite-dimensional parameter ψ. Condition (iv) is a normalization, which is very standard in the literature (e.g. Escanciano et al. (2016)). It requires that there exist at least on continuous regressor whose the coefficient is nonzero. Condition (v) imposes some smoothness on g and r, and this condition allows one to identify ψ_t^c's. It is worth pointing out that if X_t consists only of continuous regressors, the coefficients ψ_t^c's are be identified without condition (vi) in Assumption $2'$. Condition (vi) can be implied by strict monotonicity of $r(\cdot,\cdot)$ with respect to its first argument, and Escanciano et al. (2016) also impose a similar condition to identify the coefficients on discrete regressors. To motivate this assumption, consider the linear panel quantile model in Example 2.2. It can easily be shown that $Q_{Y_t|\mathbf{X},Z_t,D_t=1}(u|\mathbf{X},Z_t,D_t=1) = X_t\beta + \mathbf{X}\psi + h(P_t)$ for some unknown function $h(\cdot)$ and that $r(a,p) = a + h(p)$. In this case, the structural function r is strictly increasing in its first argument, and therefore condition (vi) is satisfied. In the wage equation example, the $\mathbf{X}'\psi$ can be considered as the ability of individual, and it is natural to assume that the structural function r is monotonically increasing in $\mathbf{X}'\psi$. A similar assumption is made by Evdokimov (2010), without considering the CRE approach.

The following theorem demonstrates that the parameters of the semiparametric model in (12) are identified under Assumptions $1'$ and $2'$.

Theorem 5.1. *Let $u \in \mathcal{U}$ be given and Assumption $1'$ hold. Suppose that conditions (i) – (v) in Assumption $2'$ are satisfied. Then, for each $t = 1, 2, ..., T$, $g(\cdot,\cdot;u)$ and $r(\cdot,\cdot;u)$ are identified over $Supp(X_t, r(\mathbf{X}'\psi, p_t(X_t,Z_t)))$ and the set $\cup_t^T Supp(\mathbf{X}'\psi, p_t(X_t,Z_t))$, respectively, and ψ_t^c's are also identified. If condition (vi) in Assumption $2'$ additionally holds, then ψ_t^d's are also identified.*

Proof. Let $t \in \{1, 2, ..., T\}$ be given. Under conditions (i) through (iii) in Assumption $2'$, the structural function m is identified over it support and one can show that

$$r(\mathbf{X}'\psi, P_t) = g^{-1}(X_t, Q_{Y_t|\mathbf{X},Z_t,D_t=1}(u|\mathbf{X}, Z_t)) \tag{13}$$

by using the same argument of the proof of Theorem 3.1. In addition, the structural function r is also identified from (13) over its support in a similar way to the proof of Theorem 3.1.

I first identify the coefficients on the continuous regressors ψ_t^c's. Choose $s \in \{1, ..., t-1, t+1, ..., T\}$. Taking derivative with respect to X_{1s}^c, one obtains that

$$r_1(\mathbf{X}'\psi, P_t) = g_2^{-1}(X_t, Q_{Y_t|\mathbf{X},Z_t,D_t=1}(u|\mathbf{X}, Z_t))\frac{\partial Q_{Y_t|\mathbf{X},Z_t,D_t=1}(u|\mathbf{X}, Z_t)}{\partial X_{1s}^c}. \tag{14}$$

Pick any $k \in \{2, 3, ..., d_x\}$. Taking derivative with respect to X_{ks}^c yields that

$$r_1(\mathbf{X}'\psi, P_t)\psi_{ks}^c = g_2^{-1}(X_t, Q_{Y_t|\mathbf{X},Z_t,D_t=1}(u|\mathbf{X}, Z_t))\frac{\partial Q_{Y_t|\mathbf{X},Z_t,D_t=1}(u|\mathbf{X}, Z_t)}{\partial X_{ks}^c}. \tag{15}$$

As a result, one can see that ψ_{ks}^c is identified by the ratio between (14) and (15), and therefore one can identify $\psi_t'^c$ for all $t \in \{1, 2, ..., T\}$. Note that to identify ψ_t^c, one can consider the model restriction in (13) for some different time period s and use the same argument.

By the invertibility condition (condition (vi) in Assumption 2′), one obtains that

$$\mathbf{X}'\psi = r^{-1}(g^{-1}(X_t, Q_{Y_t|\mathbf{X},Z_t,D_t=1}(u|\mathbf{X}, Z_t)), P_t).$$

Using the variation in the discrete regressor X_t^d, one can identify ψ_t^d's. ■

One can also consider an index structure for structural function g, and this is in particular useful when the dimension of X_t is large. Specifically, if it is assumed that $g(X_t, \gamma) = g(X_t'\beta, \gamma)$ for some $\beta \in \mathbb{R}^{d_x}$, then one can use a similar argument in the proof of Theorem 5.1 under similar conditions for g to those for r. These conditions include (i) the differentiability and invertibility of g with respect to its first argument and (ii) a normalization condition for β.

5.2 Additively Separable Models

Additively separable models are very popular in empirical studies as they are very tractable. In particular, one can use a location-scale model for quantile regression to allow for general and flexible specifications even with additive separability. The following assumption imposes additive separability between X_t and r as well as a parametric specification for structural function h.

Assumption 1″. *Let $u \in \mathcal{U}$ be given. For all $t \in \{1, 2, ..., T\}$,*

$$Q_{Y_t|\mathbf{X},Z_t,D_t=1}(u|\mathbf{X}, Z_t, D_t = 1) = h(X_t; u) + \mathbf{X}'\psi(u) + k(P_t; u). \tag{16}$$

Assumption 2″. *Let $T \geq 2$.*

(i) For each $t = 1, 2, ..., T$, there exists a known value $\bar{x}(u) \in Supp(X_t) \subseteq \mathbb{R}^{d_x}$ such that $h(\bar{x}(u); u) = 0$;

(ii) For any $u \in \mathcal{U} \subseteq (0, 1)$ and for any $t \neq s$ and $(x_t', x_s')' \in Supp(X_t, X_s)$, there exists $\mathcal{Z}((x_t', x_s')') \subseteq Supp(Z_t, Z_s)$ such that $\Pr(\mathcal{Z}((x_t', x_s')')) > 0$ and $p_t(x_t, z_t) = p_s(x_s, z_s)$ for all $(z_t', z_s')' \in \mathcal{Z}((x_t', x_s')')$.

While it is evident that Assumption 1″ restricts the type of heterogeneity that can be allowed in the model, it provides substantial identifying power. Specifically, one can weaken the condition on the number of time periods that is needed for identification. Condition (i) in Assumption 2″ is a normalization condition. Condition (ii) in Assumption 2″ is similar to condition (iii) in Assumption 2, which requires the excluded variable Z_t to have sufficient variation. If Z_t has a large support and the selection is determined by a threshold crossing equation model, then this condition is likely to be met.

Theorem 5.2. *Suppose that Assumptions 1″ and 2″ hold. Then, for any $u \in \mathcal{U} \subseteq (0,1)$, $\delta(u)$ is identified and $h(\cdot;u)$ and $k(\cdot;u)$ are also identified over $\cup_{t=1}^{T} Supp(X_t)$ and $\cup_{t=1}^{T} Supp(P_t)$, respectively.*

Proof. For simplicity of notation, we assume that X_t is a continuous random variable.

Identification of ψ is from the following derivative:

$$\psi_s = \frac{\partial Q_{Y_t|\mathbf{X},Z_t,D_t=1}(u|\mathbf{X},Z_t,D_t=1)}{\partial X_s},$$

where $t \neq s$. Since $\psi \in \mathbb{R}^T$ and there are $T \times (T-1)$ equations, one can identify ψ.

For identification of β, pick any $t, s \in \{1,2,...,T\}$ such that $t \neq s$. For a given value $x \in Supp(X_t)$, define $\mathbf{x}_0 \equiv (x^{'}, \bar{x}^{'})^{'}$. Then,

$$\begin{aligned} Q_{Y_t|\mathbf{X},Z_t,D_t=1}(u|\mathbf{x}_0,Z_t,D_t=1) - Q_{Y_s|\mathbf{X},Z_s,D_s=1}(u|\mathbf{x}_0,Z_s,D_s=1) &= h(x) - h(\bar{x}) + k(P_t) - k(P_s) \\ &= h(x) + k(P_t) - k(P_s) \end{aligned}$$

Taking conditional expectation on $(Z_t^{'}, Z_s^{'})^{'} \in \mathcal{Z}((x^{'},\bar{x}^{'})^{'})$, one obtains that

$$\mathbb{E}[Q_{Y_t|\mathbf{X},Z_t,D_t=1}(u|\mathbf{x}_0,Z_t,D_t=1) - Q_{Y_s|\mathbf{X},Z_s,D_s=1}(u|\mathbf{x}_0,Z_s,D_s=1)|\mathbf{x}_0,(Z_t^{'},Z_s^{'})^{'} \in \mathcal{Z}((x^{'},\bar{x}^{'})^{'})] = h(x),$$

and this conditional expectation is well-defined by Assumption 2″. Therefore, $h(\cdot)$ is identified over $\cup_t^T Supp(X_t)$. Since $h(\cdot)$ and δ are identified, k is identified over $\cup_t^T Supp(P_t)$ by using (16). ■

The identification strategy used in this paper does not rely on the derivative of the reduced form parameter with respect to the excluded variables even for additively separable models. This is different from the identification strategy of Das et al. (2003) who use a derivative argument. While the identification strategy of Das et al. (2003) requires that the excluded variables be continuous, the identification strategy in this paper potentially allows for discrete excluded variables as long as their support is large enough.

One can choose a specific form of structural function $g(\cdot;u)$. A popular choice is a linear function, and one can write the model restriction with a linear specification as follows:

$$Q_{Y_t|\mathbf{X},Z_t,D_t=1}(u|\mathbf{X},Z_t,D_t=1) = X_t^{'}\beta(u) + \mathbf{X}^{'}\psi(u) + k(P_t;u). \tag{17}$$

Equation (17) is corresponding to the model considered in Example 2.2, and it is a special case of the nonparametric model in (5). The quantile restriction in equation (17) is similar to that of Buchinsky (1998) and the conditional mean restriction of Das et al. (2003). When structural function m is specified as a linear function, the following assumption, together with sufficient variation in excluded variables, guarantees identification of the model parameters:

Assumption 3″. *Let $T \geq 2$ and $u \in \mathcal{U} \subseteq (0,1)$ be given. For any $t \neq s$,*

$$\Pr\left(rank\left((X_t - X_s)\cdot(X_t - X_s)^{'}\right) = d_x\right) = 1.$$

Assumption 3″ is a rank condition that requires that the time-varying covariate X_t have sufficient variation across time periods, and this condition is standard in the literature on panel data models. It is worth noting that this condition rules out the case where X_t contains some time-invariant regressors, such as a constant regressor.

The result below is a direct consequence of Theorem 5.2 and establishes the identification of β, δ, and $k(\cdot)$ under Assumptions 1″, 2″, and 3″.

Corollary 5.3. *Let $u \in \mathcal{U}$ be given and suppose that $0 \in Supp(X_t)$ for all $t = 1, 2, ..., T$. If Assumptions 1″, 2″-(ii), and 3″ hold with $h(x;u) = x^{'}\beta(u)$, then, $\beta(u)$ and $\psi(u)$ are identified. Moreover, $k(\cdot;u)$ is identified over $\cup_{t=1}^{T} Supp(P_t)$.*

Proof. Identification of ψ is the same as in Theorem 5.2. Note that the normalization condition in Assumption 2″ is satisfied with $\bar{x} = 0$. For identification of β, pick any $t, s \in \{1, 2, ..., T\}$ such that $t \neq s$. Then,

$$Q_{Y_t|\mathbf{X},Z_t,D_t=1}(u|\mathbf{X}, Z_t, D_t = 1) - Q_{Y_s|\mathbf{X},Z_s,D_s=1}(u|\mathbf{X}, Z_s, D_s = 1) = (X_t - X_s)^{'}\beta + k(P_t) - k(P_s).$$

Taking conditional expectation on $(Z_t^{'}, Z_s^{'})^{'} \in \mathcal{Z}((X_t^{'}, X_s^{'})^{'})$, one obtains that

$$\begin{aligned}&\mathbb{E}[Q_{Y_t|\mathbf{X},Z_t,D_t=1}(u|\mathbf{X}, Z_t, D_t = 1) - Q_{Y_s|\mathbf{X},Z_s,D_s=1}(u|\mathbf{X}, Z_s, D_s = 1)|X_t, X_s, (Z_t^{'}, Z_s^{'})^{'} \in \mathcal{Z}((X_t^{'}, X_s^{'})^{'})]\\ &= (X_t - X_s)^{'}\beta\end{aligned}$$

and this condition expectation is well-defined by condition (ii) in Assumption 3″. Therefore, by condition (i) in Assumption 3″, one can show that

$$\begin{aligned}\beta &= \left((X_t - X_s)\cdot(X_t - X_s)^{'}\right)^{-1}\cdot(X_t - X_s)\\ &\times \mathbb{E}[Q_{Y_t|\mathbf{X},Z_t,D_t=1}(u|\mathbf{X}, Z_t, D_t = 1) - Q_{Y_s|\mathbf{X},Z_s,D_s=1}(u|\mathbf{X}, Z_s, D_s = 1)|X_t, X_s, (Z_t^{'}, Z_s^{'})^{'} \in \mathcal{Z}((X_t^{'}, X_s^{'})^{'})].\end{aligned}$$

Since β and ψ are identified, $k(\cdot)$ is also identified over $\cup_{t=1}^{T} Supp(P_t)$ from equation (17). ■

6 Estimation and Inference

We use a PSMD procedure to estimate the parameters developed by Chen and Pouzo (2015). Note that since the infinite-dimensional parameters do not depend on any endogenous regressors, the estimators do not suffer from an ill-posed inverse problem; and therefore, one can expect that the role of penalization is not significant. For this reason, the PSMD procedure employed in this paper is is the one with slowly growing finite-dimensional linear sieves and small flexible penalty (Chen and Pouzo (2012, 2015)). We also establish the asymptotic theory for the PSMD estimators for inference on general functionals of the parameters. Since the moment conditions used to estimate the parameters are non-smooth, the asymptotic theory in this paper considers sieve quasi likelihood ratio (SQLR) inference that allows one to avoid estiatming asymptotic variances. [15]

We introduce additional notation. For a generic vector A, $vec(A)$ denotes the vectorization of A. Let $\{\mathbf{A}_i \equiv (\mathbf{Y}_i, \mathbf{X}_i', \mathbf{Z}_i', \mathbf{D}_i')' : i = 1, 2, ...N\}$ be the data, where $\mathbf{X}_i \equiv (vec(X_{i1})', ..., vec(X_{iT})')'$ and $\mathbf{Z}_i \equiv (vec(Z_{i1})', ..., vec(Z_{iT})')'$. For a positive definite square matrix W and a comformable vector A, define $||A||_W \equiv \sqrt{A'WA}$. I also define several norms that are used in this paper. The Euclidean norm is denoted by $||\cdot||_E$. Let $||\cdot||_\infty$ and $||\cdot||_2$ denote the supremum-norm and L_2-norm on a function space, respectively. For any $\alpha_1, \alpha_2 \in \mathcal{A}$, define $||\alpha_1 - \alpha_2||_2^2 \equiv ||g_1 - g_2||_2^2 + ||r_1 - r_2||_2^2 + ||p_1 - p_2||_2^2$.

6.1 Penalized Sieve Minimum Distance Estimation

For a given value of $u \in \mathcal{U}$, let g_0, r_0, and p_0 be the true parameter values for g, r, and p, respectively.[16] Let $\mathcal{G}$, $\mathcal{R}$, and $\mathcal{P}$ be classes of admissible functions for g, r, and p, respectively. The parameter is denoted by $\alpha \equiv (g, r, p)'$, and the parameter space $\mathcal{A}$ is the Cartesian product of $\mathcal{G}$, $\mathcal{R}$, and $\mathcal{P}$ (i.e., $\mathcal{A} \equiv \mathcal{G} \times \mathcal{R} \times \mathcal{P}$).

Define

$$\begin{aligned}
&\rho_{1,t}(\mathbf{Y}, \mathbf{D}, \mathbf{X}, \mathbf{Z}; \alpha) \equiv D_t - p(X_t, Z_t), \\
&\rho_{2,t}(\mathbf{Y}, \mathbf{D}, \mathbf{X}, \mathbf{Z}; \alpha) \equiv D_t \left\{ \mathbf{1}\left(Y_t - g(X_t, r(\mathbf{X}, p(X_t, Z_t))) \leq 0 \right) - u \right\},
\end{aligned}$$

for each $t = 1, 2, ..., T$. For each $j = 1, 2$, let

$$\rho_j(\mathbf{Y}, \mathbf{D}, \mathbf{X}, \mathbf{Z}; \alpha) \equiv [\rho_{j,1}(\mathbf{Y}, \mathbf{D}, \mathbf{X}, \mathbf{Z}; \alpha), \rho_{j,2}(\mathbf{Y}, \mathbf{D}, \mathbf{X}, \mathbf{Z}; \alpha), ..., \rho_{j,T}(\mathbf{Y}, \mathbf{D}, \mathbf{X}, \mathbf{Z}; \alpha)]'$$

[15] It is worth noting that the identification the model implication (6) suggests a two-step estimation procedure in which the selection probabilitiy $\Pr(D_{it} = 1 | X_{it}, Z_{it})$ is estimated in the first step and the structural functions are estimated by using the standard quantile regression with the estimated selection probability. Hahn et al. (2018) develop the asymptotic thoery for nonparametric two-step sieve estimators, including consistency, convergence rates, and $\sqrt{n}$-asymptotic normality for regular functionals. However, the objective function for the standard quantile regression is non-differentiable; and therefore, it may be challenging to estimate the asymptotic variance of the two-step sieve estimator of a functional. On the other hand, I overcome such practical issues by using the SQLR inference and the bootstrap procedure developed by Chen and Pouzo (2015).

[16] We drop u from the parameters for notational simplicity.

and

$$m_{j,t}(\mathbf{X},\mathbf{Z};\alpha)\equiv\mathbb{E}\left[\rho_{j,t}(\mathbf{Y},\mathbf{D},\mathbf{X},\mathbf{Z};\alpha)|\mathbf{X},\mathbf{Z}\right].$$

Finally, define

$$m_t(\mathbf{X},\mathbf{Z};\alpha)\equiv\begin{bmatrix}m_{1,t}(\mathbf{X},\mathbf{Z};\alpha)\\ m_{2,t}(\mathbf{X},\mathbf{Z};\alpha)\end{bmatrix}=\begin{bmatrix}p_0(X_t,Z_t)-p(X_t,Z_t)\\ \left\{F_{Y_t^*|D_t=1,\mathbf{X},\mathbf{Z}}\left(g(X_t,r(\mathbf{X},p(X_t,Z_t)))|\mathbf{X},\mathbf{Z}\right)-u\right\}p_0(X_t,Z_t)\end{bmatrix}$$

for each $t=1,2,...,T$, and $m(\mathbf{X},\mathbf{Z};\alpha)\equiv\left[m_1(\mathbf{X},\mathbf{Z};\alpha)',.....,m_T(\mathbf{X},\mathbf{Z};\alpha)'\right]'$. Then, for any positive definite finite matrix $\Sigma(\mathbf{X},\mathbf{Z})$,

$$\mathbb{E}\left[||m(\mathbf{X},\mathbf{Z};\alpha)||^2_{\Sigma^{-1}}\right]\begin{cases}\geq 0 & \text{for any } \alpha\in\mathcal{A}\\ =0 & \text{if and only if } \alpha=\alpha_0\end{cases}$$

under the identification conditions. Let $\hat{m}_n(\mathbf{X},\mathbf{Z};\alpha)$ be a nonparametric consistent estimator of $m(\mathbf{X},\mathbf{Z};\alpha)$. Let $\hat{\Sigma}_n(\cdot,\cdot)$ be a consistent estimator of $\Sigma(\cdot,\cdot)$. Then, the PSMD estimator of α_0, $\hat{\alpha}_n$, is defined as the solution to the following minimization problem:

$$\min_{\alpha\in\mathcal{A}_n}\left\{\hat{Q}_n(\alpha)+\lambda_n\text{Pen}(\alpha)\right\},$$

where $\hat{Q}_n(\alpha)\equiv\frac{1}{n}\sum_i^n\hat{m}_n(\mathbf{X}_i,\mathbf{Z}_i;\alpha)'\hat{\Sigma}_n(\mathbf{X}_i,\mathbf{Z}_i)^{-1}\hat{m}_n(\mathbf{X}_i,\mathbf{Z}_i;\alpha)$, $\lambda_n\text{Pen}(\alpha)$ is a penalty term with $\lambda_n=o(1)$ and $\text{Pen}(\cdot):\mathcal{A}\to\mathbb{R}_+$, and $\mathcal{A}_n$ is a sieve space of $\mathcal{A}$. In this paper, I use a series estimator of $m(\mathbf{X},\mathbf{Z};\alpha)$; that is, for each $j=1,2$ and $t=1,2,...,T$,

$$\hat{m}_{j,t,n}(\mathbf{X},\mathbf{Z};\alpha)=\phi^{J_n}(\mathbf{X},\mathbf{Z})'\left(\sum_{i=1}^n\phi^{J_n}(\mathbf{X}_i,\mathbf{Z}_i)\cdot\phi^{J_n}(\mathbf{X}_i,\mathbf{Z}_i)'\right)^{-1}\sum_{i=1}^n\phi^{J_n}(\mathbf{X}_i,\mathbf{Z}_i)'\rho_{j,t}(\mathbf{Y}_i,\mathbf{D}_i,\mathbf{X}_i,\mathbf{Z}_i;\alpha),$$

where $\{\phi_j(\cdot,\cdot)\}_{j=1}^\infty$ is a sequence of some basis functions and $\phi^{J_n}(x,z)\equiv(\phi_1(x,z),\phi_2(x,z),...,\phi_{J_n}(x,z))'$ with $J_n\to\infty$ as $n\to\infty$. Then, $\hat{m}_{t,n}=(\hat{m}_{1,t,n},\hat{m}_{2,t,n})'$ is a series estimator of $m_t(\mathbf{X},\mathbf{Z};\alpha)$, and define $\hat{m}_n(\mathbf{X},\mathbf{Z};\alpha)\equiv\left[\hat{m}_{1,n}',...,\hat{m}_{T,n}'\right]'$.

We introduce one of the most popular classes of functions, which is called the Hölder class. Let $f:\mathbb{D}\to\mathbb{R}$ where $\mathbb{D}\subseteq\mathbb{R}^{d_x}$ for some integer $d_x\geq 1$. Let $\omega=(\omega_1,...,\omega_{d_x})$ be a d_x-tuple of nonnegative integers, and define the differential operator as $\nabla^\omega f\equiv\frac{\partial^{|\omega|}}{\partial x_1^{\omega_1}\partial x_2^{\omega_2}\cdots\partial x_{d_x}^{\omega_{d_x}}}f(x)$, where $x=(x_1,x_2,...,x_{d_x})\in\mathbb{D}$ and $|\omega|\equiv\sum_{i=1}^{d_x}\omega_i$. Let $[s]$ be the integer part of $s\in\mathbb{R}_+$, then a function $f:\mathcal{X}\to\mathbb{R}$ is called s-smooth if it is $[s]$ times continuously differentiable on $\mathcal{X}$ and for all ω such that $|\omega|=[s]$ and for some $\nu\in(0,1]$ and constant $c>0$, $|\nabla^\omega f(x)-\nabla^\omega f(y)|\leq c\cdot||x-y||_E^\nu$ for all $x,y\in\mathcal{X}$. Let $\mathcal{C}^{[p]}(\mathcal{X})$ denote the space of all $[p]$ times continuously differentiable real-valued functions on $\mathcal{X}$. A Hölder ball with smoothness s is defined as follows:

$$\Lambda_C^s(\mathcal{X}) \equiv \left\{ f \in \mathcal{C}^{[s]}(\mathcal{X}) : \sup_{|\omega|\leq[s]} \sup_{x\in\mathcal{X}} |\nabla^\omega f(x)| \leq C, \sup_{|\omega|=[s]} \sup_{x,y\in\mathcal{X},x\neq y} \frac{|\nabla^\omega f(x) - \nabla^\omega f(y)|}{||x-y||_E^\nu} \leq C \right\},$$

where C is a positive finite constant.

The choice of sieve spaces depends on the class of functions and support of unknown function. When an unknown function is in a Hölder ball and its support is the unit interval, one can use polynomial, trigonometric polynomial, or spline sieve spaces. If the support is unbounded, then Hermite polynomial sive spaces can be used. For the detailed discussion on the choice of sieve spaces, one can refer to Chen (2007).

6.2 Asymptotic Theory for PSMD Estimators

6.2.1 Consistency and convergence rates

Assumption 3. *(i) The data $\{\mathbf{A}_i = (A_{i1}, ..., A_{iT})' : i = 1, 2, ...n\}$ are i.i.d across i; (ii) for any $t = 1, 2, ..., T$, the conditional distribution of Y_t^* on $\mathbf{X}$ and Z_t is absolutely continuous with respect to the Lebesgue measure ; (iii) for any $t = 1, 2, ..., T$, $\mathbb{E}[|Y_t^*|]$ and $\mathbb{E}[|q_u(\mathbf{X}_i, Z_{it})|]$ are uniformly bounded; (iv) Supp$(\mathbf{X}, \mathbf{Z})$ is a compact connected subset of $\mathbb{R}^{(d_x+d_z)T}$ with Lipscthiz boundary; (v) the density function of $(\mathbf{X}, \mathbf{Z})$ is bounded and bounded away from zero over Supp$(\mathbf{X}, \mathbf{Z})$; (vi) $p_0(x, z)$ is bounded away from zero and one uniformly over $\cup_t^T$Supp(X_t, Z_t).*

Assumption 4. *(i) $g_0 \in \mathcal{G} \equiv \Lambda_{c_g}^{s_g}(\cup_t^T Supp(\mathbf{X}, p_0(X_t, Z_t))$, $r_0 \in \mathcal{R} \equiv \Lambda_{c_r}^{s_r}(Supp(\mathbf{X}, p_0(X_t, Z_t))$ with $s_m > 1$ and $s_r > 1$, $g_0(x, \gamma)$ and $r_0(\mathbf{x}, p)$ are continuously differentiable with respect to γ and p, respectively, and the derivatives are uniformly bounded; (ii) $\log\left(\frac{p_0}{1-p_0}\right) \in \mathcal{P} \equiv \Lambda_{c_p}^{s_p}(\cup_t^T Supp(X_t, Z_t))$ with $s_p > \frac{1}{2}$.*

Assumption 5. *(i) Let*

$$\mathcal{G}_n \equiv \left\{ g_n(x, \gamma) = \phi^{k_{g,n}}(x, \gamma)' \beta_{g,n} : \sup_{x,\gamma} |g_n(x, \gamma)| \leq c_g \right\},$$

$$\mathcal{R}_n \equiv \left\{ r_n(\mathbf{x}, p) = \phi^{k_{r,n}}(\mathbf{x}, p)' \beta_{r,n} : \sup_{\mathbf{x},p} |r_n(\mathbf{x}, p)| \leq c_r \right\},$$

$$\mathcal{P}_n \equiv \left\{ L\left(\phi^{k_{p,n}}(X_t, Z_t)' \beta_{p,n}\right) : \beta_{p,n} \in \mathbb{R}^{k_{p,n}} \right\},$$

where $\phi^{k_{g,n}}$, $\phi^{k_{r,n}}$, and $\phi^{k_{p,n}}$ are some basis functions, $k_{g,n}$, $k_{r,n}$, and $k_{p,n}$ are some positive non-decreasing integer sequences such that $k_{g,n}, k_{r,n}, k_{p.n} \to \infty$ as $n \to \infty$ and $k_n \equiv \max(k_{m,n}, k_{r,n}, k_{p,n}) = o(n)$, $L(\cdot)$ is the standard logistic distribution function; (ii) let $\mathbb{Q}_{g,t,n} \equiv \mathbb{E}\left[\phi^{k_{g,n}}(x_t, r_n(\mathbf{X}, P_t)) \cdot \phi^{k_{g,n}}(x_t, r_n(\mathbf{X}, P_t))'\right]$, $\mathbb{Q}_{r,t,n} \equiv \mathbb{E}\left[\phi^{k_{r,n}}(\mathbf{X}, P_t) \cdot \phi^{k_{r,n}}(\mathbf{X}, P_t)'\right]$ and $\mathbb{Q}_{p,t,n} \equiv \mathbb{E}\left[\phi^{k_{p,n}}(X_t, Z_t) \cdot \phi^{k_{p,n}}(X_t, Z_t)'\right]$, then for any t, the eigenvalues of $\mathbb{Q}_{g,t,n}$, $\mathbb{Q}_{r,t,n}$, and $\mathbb{Q}_{p,t,n}$ are bounded above and away from zero uniformly over all n; (iii) there exist

$\left\{\beta^0_{g,n}\right\}_{n=1}^{\infty}$, $\left\{\beta^0_{r,n}\right\}_{n=1}^{\infty}$, *and* $\left\{\beta^0_{p,n}\right\}_{n=1}^{\infty}$ *such that* $\sup|g_0(x,\gamma)-\phi^{k_{g,n}}(x,\gamma)'\beta^0_{g,n}|=O\left(k_{g,n}^{-\sigma_g}\right)$, $\sup|r_0(\mathbf{x},p)-\phi^{k_{r,n}}(\mathbf{x},p)'\beta^0_{r,n}|=O\left(k_{r,n}^{-\sigma_r}\right)$, *and* $\sup\left|\log\left(\frac{p_0(x,z)}{1-p_0(x,z)}\right)-\phi^{k_{p,n}}(x,z)'\beta^0_{p,n}\right|=O\left(k_{p,n}^{-\sigma_p}\right)$ *for some* $\sigma_g,\sigma_r,\sigma_p>0$.

Assumption 6. *(i) For any* $\alpha\in\mathcal{A}$, $m(\mathbf{X},\mathbf{Z};\alpha)\in\Lambda^{p_m}_{c_m}(Supp(\mathbf{X},\mathbf{Z}))$ *with* $p_m>\frac{T(d_x+d_z)}{2}$ *and some* $c_m>0$*; (ii)* $\mathcal{M}_n=\left\{m_n(\mathbf{x},\mathbf{z};\alpha)=\phi^{J_n}(\mathbf{x},\mathbf{z})'\beta_{m,n}:\sup_{\mathbf{x},\mathbf{z}}|m_n(\mathbf{x},\mathbf{z})|\le c_m\right\}$, *where* $\left\{\phi^{J_n}\right\}_{n=1}^{\infty}$ *is a sequence of some basis functions; (iii) there exists* $\left\{\beta^0_{m,n}\right\}_{n=1}^{\infty}$ *such that* $\sup_{(\mathbf{x},\mathbf{z})\in Supp(\mathbf{X},\mathbf{Z})}|m(\mathbf{x},\mathbf{z};\alpha)-\phi^{J_n}(\mathbf{x},\mathbf{z})'\beta^0_{m,n}|=O\left(J_n^{-\sigma_m}\right)$*; (iv) let* $\mathbb{Q}_{m,n}\equiv\mathbb{E}\left[\phi^{J_n}(\mathbf{X},\mathbf{Z})\phi^{J_n}(\mathbf{X},\mathbf{Z})'\right]$, *then the eigenvalues of* $\mathbb{Q}_{m,n}$ *are bounded above and away from zero uniformly over all* n*; (v)* $J_n\ge c\cdot(k_{g,n}+k_{r,n}+k_{p,n})$ *for some* $c\ge1$.

Assumption 7. *(i)* $\Sigma(\mathbf{X},\mathbf{Z})=I_{2T}$ *almost surely; (ii)* $\lambda_n=o\left(n^{-1}\right)$*; (iii)* $Pen(\alpha)=0$.

Assumption 3 imposes conditions on the data generating process. Note that the first condition of Assumption 3 allows for serial correlation as it only requires the data be i.i.d. across the individuals. Condition (ii) in Assumption 3 implies that Y_t^* is a continuous random variable, which is standard for quantile regression models. Condition (iii) is a mild condition on moments of the dependent variable and conditional quantile function. Conditions (iv) and (v) in Assumption 3 are standard in the literature on nonparametric estimation. Condition (vi) rules out the case where all individuals are selected or not selected.

Assumption 4 specifies the parameter spaces for the structural functions g, r, and the log odds ratio, and their sieve spaces. Assumption 5 defines sieve spaces for $\mathcal{G}$, $\mathcal{R}$, and $\mathcal{P}$. The choice of sieve spaces depends on the parameter spaces and support conditions. We use linear sieve approximations of g_0, r_0, and the log odds ratio of the selection probability $p_0(x,z)$. This is to ensure that $p_0(x,z)\in(0,1)$ for all possible values of x and z (cf. Hirano et al. (2003)). Condition (ii) of Assumption 5 is standard in the literature on sieve or series estimation (cf. Newey (1997) and Chen and Christensen (2018)). Condition (iii) of Assumption 5 implies that the unknown parameters are well-approximated over their sieve spaces, and this condition is easily met with an appropriate choice of sieve spaces.

Assumptions 4 and 5 together imply that when the polynomial or spline sieve spaces are used for g_0, r_0 and $\log\left(\frac{p_0}{1-p_0}\right)$, Assumption 4-(i) implies that $\sigma_g=p_g/(d_x+1)$ and $\sigma_r=p_r/(Td_x+1)$ (Newey (1997)). Likewise, under Assumption 4-(ii), one can show that $\sigma_p=\frac{p_p}{(d_x+d_z)}$. We denote the sequences of functions $\left\{\phi^{k_{g,n}}(x,\gamma)'\beta^0_{g,n}\right\}_{n=1}^{\infty}$, $\left\{\phi^{k_{r,n}}(\mathbf{x},p)'\beta^0_{r,n}\right\}_{n=1}^{\infty}$, and $\left\{L\left(\phi^{k_{p,n}}(X_t,Z_t)'\beta^0_{p,n}\right)\right\}_{n=1}^{\infty}$ by $\{\pi_n g_0\}_n$, $\{\pi_n r_0\}_n$, and $\{\pi_n p_0\}_n$, respectively. It is also worth mentioning that Assumption 4-(i) and Assumption 5 together imply that $\mathcal{G}_n\subseteq\mathcal{G}_{n+1}$, $\mathcal{R}_n\subseteq\mathcal{R}_{n+1}$, and $\mathcal{P}_n\subseteq\mathcal{P}_{n+1}$ for all $n\ge1$ and that $\overline{\cup_{n=1}^{\infty}\mathcal{G}_n}=\mathcal{G}$, $\overline{\cup_{n=1}^{\infty}\mathcal{R}_n}=\mathcal{R}$, and $\overline{\cup_{n=1}^{\infty}\mathcal{P}_n}=\mathcal{P}$ where, for a set A, $\overline{A}$ is the closure of A.

Assumption 6 imposes conditions required for a series estimator of m to perform well. This assumption is similar to Assumptions 4 and 5. Condition (v) of Assumption 6 , together with Assumption 5 -(i), implies that $J_n \to \infty$ and $J_n/n \to 0$. Assumption 7 specifies the penalty function that will be used in this paper and the rate of λ_n.[17]

The following theorem demonstrates that the sieve estimator of α_0, $\hat{\alpha}_n$, is consistent with respect to both $||\cdot||_\infty$ and $||\cdot||_2$ under a set of assumptions.

Theorem 6.1. *Suppose that Assumptions 1, 2, and 3–7 hold. Then,*

$$||\hat{\alpha}_n - \alpha_0||_\infty = o_p(1).$$

Let $\mathcal{A}_{os}$ be a convex $||\cdot||_2$-neighborhood of α_0 such that

$$\mathcal{A}_{os} \subseteq \{\alpha \in \mathcal{A} : ||\alpha - \alpha_0||_\infty < C_0, \lambda_n \text{Pen}(h) < \lambda_n C_0\}$$

for some positive constant C_0. For any $\alpha \in \mathcal{A}_{os}$ and $t \in \{1, 2, ..., T\}$, define a pathwise derivative of m_t as

$$\begin{aligned}
&\frac{dm_t(\mathbf{X}, \mathbf{Z}; \alpha_0)}{d\alpha}[\alpha - \alpha_0] \\
&\equiv \frac{d\mathbb{E}\left[\rho_t(\mathbf{A}; (1-\tau)\alpha_0 + \tau\alpha)|\mathbf{X}, \mathbf{Z}\right]}{d\tau}\Big|_{\tau=0} \\
&= \begin{bmatrix} -(p(X_t, Z_t) - p_0(X_t, Z_t)) \\ f_{Y_t^*|D_t=1,\mathbf{X},\mathbf{Z}}(Q_{t,0}|\mathbf{X}, \mathbf{Z}) p_0(X_t, Z_t) \left\{(g - g_0) + B_{1,t}(\mathbf{X}, \mathbf{Z}; \alpha_0) \cdot (r - r_0) + B_{2,t}(\mathbf{X}, \mathbf{Z}; \alpha_0) \cdot (p - p_0)\right\} \end{bmatrix},
\end{aligned}$$

where $Q_{t,0} \equiv g_0(X_t, r_0(\mathbf{X}, p_0(X_t, Z_t)))$, $B_{1,t}(\mathbf{X}, \mathbf{Z}; \alpha_0) \equiv g_0'(X_t, r_0(\mathbf{X}, p_0(X_t, Z_t)))$ and $B_{2,t}(\mathbf{X}, \mathbf{Z}; \alpha_0) \equiv g_0'(X_t, r_0(\mathbf{X}, p_0(X_t, Z_t)))r_0'(\mathbf{X}, p_0(X_t, Z_t))$. The corresponding pathwise derivative of m is

$$\frac{dm(\mathbf{X}, \mathbf{Z}; \alpha_0)}{d\alpha}[\alpha - \alpha_0] \equiv \left[\left(\frac{dm_1(\mathbf{X}, \mathbf{Z}; \alpha_0)}{d\alpha}[\alpha - \alpha_0]\right)', \cdots, \left(\frac{dm_T(\mathbf{X}, \mathbf{Z}; \alpha_0)}{d\alpha}[\alpha - \alpha_0]\right)'\right]'.$$

Assumption 8. *For all $t \in \{1, 2, ..., T\}$, $f_{Y_t^*|\mathbf{X},\mathbf{Z},D_t=1}(Q_{Y_t|D_t=1,\mathbf{X},\mathbf{Z}}(u|\mathbf{X}, \mathbf{Z})|\mathbf{X}, \mathbf{Z}) > 0$ and $f_{Y_t^*|\mathbf{X},\mathbf{Z},D_t=1}(y|\mathbf{x}, \mathbf{z})$ is uniformly bounded over $(y, \mathbf{x}, \mathbf{z}) \in \mathbb{R} \times Supp(\mathbf{X}, \mathbf{Z})$.*

For any $\alpha_1, \alpha_2 \in \mathcal{A}_{os}$, define

$$||\alpha_1 - \alpha_2||^2 \equiv \mathbb{E}\left[\left(\frac{dm(\mathbf{X}, \mathbf{Z}; \alpha_0)}{d\alpha}[\alpha_1 - \alpha_2]\right)' \cdot \Sigma(\mathbf{X}, \mathbf{Z})^{-1} \cdot \left(\frac{dm(\mathbf{X}, \mathbf{Z}; \alpha_0)}{d\alpha}[\alpha_1 - \alpha_2]\right)\right].$$

Then, under Assumption 8, $||\cdot||$ is a pseudo-metric on $\mathcal{A}_{os}$. I consider additional assumptions to establish the convergence rate of the sieve estimator of α_0, $\hat{\alpha}_n$.

Assumption 9. *For any $\alpha, \tilde{\alpha} \in \mathcal{A}_{os}$,*

$$\mathbb{E}[\{g(X_{it}, r(\mathbf{X}_i, p(X_{it}, Z_{it}))) - \tilde{g}(X_{it}, \tilde{r}(\mathbf{X}_i, \tilde{p}(X_{it}, Z_{it})))\}^2] \asymp ||\alpha - \tilde{\alpha}||^2_{\mathcal{A},2}.$$

[17] It is only required that $\lambda_n = o(1)$ for consistency of $\hat{\alpha}_n$. Assumption 10-(i) is strongether than the condition, but this rate condition on λ_n facilitates deriving the convergence rate of $\hat{\alpha}_n$.

Assumption 10. *Let* $\delta_{m,n}^2 = \max\left\{\frac{J_n}{n}, J_n^{-2p_m/T\cdot(d_x+d_z)}\right\}$. *Then,*

$$\max\left\{\delta_{m,n}^2, \left(k_{g,n}^{-2\sigma_m} + k_{r,n}^{-2\sigma_r} + k_{p,n}^{-2\sigma_p}\right), \lambda_n\right\} = \delta_{m,n}^2.$$

Assumption 9 is not very restrictive, in particular when focusing on a neighborhood of α_0, and is standard (see, for example, Van de Geer (2000, Section12.3)). Assumption 10 restricts the rates of J_n and λ_n. $\frac{J_n}{n}$ and $J_n^{-p_m/T\cdot(d_x+d_z)}$ are the convergence rates of the variance term and square of bias of $\hat{m}_n$, respectively. Under Assumption 10, the penalty term does not affect the convergence rate of the sieve estimator $\hat{\alpha}_n$.

Theorem 6.2. *Suppose that Assumptions 1, 2, and 3–7 hold. If, additionally, Assumptions 8, 9 and 10 hold, then*

$$\begin{aligned}
||\hat{\alpha}_n - \alpha_0||_2 &= O_p\left(||\alpha_0 - \pi_n\alpha_0||_2 + \delta_{m,n}\right) \\
&= O_p\left(\max\left(k_{g,n}^{-\sigma_g}, k_{r,n}^{-\sigma_r}, k_{p,n}^{-\sigma_p}\right) + \delta_{m,n}\right) \\
&= O_p\left(\delta_{m,n}\right)
\end{aligned}$$

Furthermore, if $J_n \asymp k_n = \max\{k_{g,n}, k_{r,n}, k_{p,n}\}$, *then*

$$||\hat{\alpha}_n - \alpha_0||_2 = O_p\left(\sqrt{\frac{k_n}{n}} + \max\left(k_{g,n}^{-\sigma_g}, k_{r,n}^{-\sigma_r}, k_{p,n}^{-\sigma_p}, k_n^{-\frac{p_m}{T\cdot(d_x+d_z)}}\right)\right).$$

The L_2 convergence rate of $\hat{\alpha}_n$ is decomposed into two parts. The first component $\frac{k_n}{n}$ is the convergence rate of the variance term of $\hat{\alpha}_n$, and the other component reflects the convergence rates of the bias terms of $\hat{\alpha}_n$ and $\hat{m}_n$. Once the sieve spaces for $\mathcal{A}$ and $\mathcal{M}$ are appropriately chosen, the convergence rates of the bias terms can be compared. In addition, since the model does not suffer from an ill-posed inverse problem, this convergence rate is consistent with (or similar to) the standard one in the literature on series/sieve estimation (e.g., Newey (1997); Chen (2007)).

6.2.2 Asymptotic distributions

We develop the distributional theory for the PSMD estimator. We establish the asymptotic normality of the plug-in estimator of a functional of interest and consider the sieve quasi likelihood ratio (SQLR) inference. One important advantange of the SQLR inference is that it is not needed to estimate the asymptotic variance of the sieve estimator of the functional, which is very useful for non-smooth moment conditions. It is also worth noting that the distributional theory in this section allows the functional of interest to be either regular or irregular. We then provide the validity of the generalized residual bootstrap procedure for the SQLR statistic.

Let $\delta_{2,n}$ be the L_2 convergence rate of the sieve estimator of α_0, $\hat{\alpha}_n$, given in Theorem

6.2. Define

$$\mathcal{N}_{os} \equiv \{\alpha \in \mathcal{A} : ||\alpha - \alpha_0||_2 \leq M_n \delta_{2,n}, \lambda_n \text{Pen}(\alpha) \leq \lambda_n C_0\},$$
$$\mathcal{N}_{osn} \equiv \mathcal{N}_{os} \cap \mathcal{A}_n,$$

where $M_n = \log(\log(n+1))$. Then, Theorem 6.2 implies that $\hat{\alpha}_n \in \mathcal{N}_{osn} \subseteq \mathcal{N}_{os}$ with probability approaching to 1.

Let $\alpha_{0,n} \in \mathcal{A}_n$ be such that $||\alpha_{0,n} - \alpha_0|| \leq ||\alpha - \alpha_0||$ for any $\alpha \in \mathcal{A}_n$. Denote a linear span of $\mathcal{A}_{os} - \{\alpha_0\}$ by $\mathbb{V}$. Similarly, let $\mathbb{V}_n$ be a linear span of $\mathcal{A}_{osn} - \{\alpha_{0.n}\}$. Let $\overline{\mathbb{V}}_n$ be $\text{clsp}(\mathbb{V}_n)$, where $\text{clsp}(\cdot)$ is the closed linear span under $||\cdot||$. Then, $\overline{\mathbb{V}}_n$ is a finite dimensional Hilbert space under $||\cdot||$ and dense in $\overline{\mathbb{V}} \equiv \text{clsp}(\mathbb{V})$. For simplicity, I assume that $\dim(\overline{\mathbb{V}}_n) = \dim(\mathcal{A}_n)$. Define an inner product on $\overline{\mathbb{V}} \times \overline{\mathbb{V}}$ as

$$< v_1, v_2 > \equiv \mathbb{E}\left[\left(\frac{dm(\mathbf{X}, \mathbf{Z}; \alpha_0)}{d\alpha}[v_1]\right)' \Sigma(\mathbf{X}, \mathbf{Z})^{-1}\left(\frac{dm(\mathbf{X}, \mathbf{Z}; \alpha_0)}{d\alpha}[v_2]\right)\right].$$

Define a pathwise derivative of the functional $f(\cdot)$ at α_0 in the direction of $v = \alpha - \alpha_0 \in \mathcal{A}$ as

$$\frac{df(\alpha_0)}{d\alpha}[v] = \frac{\partial f(\alpha_0 + \tau v)}{\partial \tau}\Big|_{\tau=0}.$$

Assume that t $\frac{df(\alpha_0)}{d\alpha}[\cdot]$ is a linear functional. Since $\overline{\mathbb{V}}_n$ is a finite-dimensional Hilbert space under $||\cdot||$, there exists $v_n^* \in \overline{\mathbb{V}}_n$ such that

$$\frac{df(\alpha_0)}{d\alpha}[v] =< v_n^*, v >$$

for all $v \in \overline{\mathbb{V}}_n$ and

$$||v_n^*|| \equiv \sup_{v \in \overline{\mathbb{V}}_n, ||v|| \neq 0} \frac{\left|\frac{df(\alpha_0)}{d\alpha}[v]\right|}{||v||} < \infty$$

by the Riesz representation theorem, where v_n^* is called the sieve Riesz representer of the functional $\frac{df(\alpha_0)}{d\alpha}[\cdot]$ on $\overline{\mathbb{V}}_n$.

Let

$$S_{n,i}^* \equiv \left(\frac{dm(\mathbf{X}_i, \mathbf{Z}_i; \alpha_0)}{d\alpha}[v_n^*]\right)' \Sigma(\mathbf{X}_i, \mathbf{Z}_i)^{-1} \rho(\mathbf{A}_i; \alpha_0)$$

be the sieve score associated with the i-th obsevcation, and $||v_n^*||_{sd}^2 \equiv Var(S_{n,i}^*)$ denote the sieve variance. We also define

$$u_n^* \equiv \frac{v_n^*}{||v_n^*||_{sd}}$$

as the scaled sieve Riesz representer. Denote

$$\mathbb{Z}_n \equiv \frac{1}{n}\sum_{i}^{n}\frac{S^*_{n,i}}{||v^*_n||_{sd}}.$$

Let $f:\mathcal{A}\to\mathbb{R}$ be a functional continuous in $||\cdot||_2$. To construct a confidence set of $f(\alpha_0)$ or perform hypothesis testing of $H_0: f(\alpha_0)=f_0$ against $H_1: f(\alpha_0)\neq f_0$, I consider the SQLR statistic. To this end, define $\mathcal{A}^R_n \equiv \{\alpha\in\mathcal{A}_n : f(\alpha)=f_0\}$ as the restricted sieve space and let $\hat{\alpha}^R_n\in\mathcal{A}^R_n$ be a restricted PSMD estimator; that is,

$$\hat{Q}_n\left(\hat{\alpha}^R_n\right)+\lambda_n \text{Pen}\left(\hat{\alpha}^R_n\right) \leq \inf_{\alpha\in\mathcal{A}^R_n}\left\{\hat{Q}_n\left(\alpha\right)+\lambda_n\text{Pen}(\alpha)\right\}+o_p\left(n^{-1}\right). \tag{18}$$

The SQLR statistic is defined as

$$\widehat{QLR}_n(f_0)\equiv n\left(\hat{Q}_n\left(\hat{\alpha}^R_n\right)-\hat{Q}_n\left(\hat{\alpha}_n\right)\right). \tag{19}$$

We consider the following assumptions to derive the asymptotic normality for the plug-in estimator of functional $f(\alpha_0)$ and the asymptotic distribution of the SQLR statistic.

Assumption 11. *Let $\mathcal{T}_n \equiv \{t\in\mathbb{R}: |t|\leq 4M_n^2\delta_{2,n}\}$. Then, the following conditions hold:*

(i) $v\mapsto \frac{df(\alpha_0)}{d\alpha}[v]$ is a nonzero linear functional mapping from $\mathbb{V}$ to $\mathbb{R}$, $\{\overline{\mathbb{V}}_n\}$ is an increasing sequence of finite-dimensional Hilbert spaces that is dense in $(\overline{\mathbb{V}},||\cdot||)$, and $\frac{||v^*_n||}{\sqrt{n}}=o(1)$;

(ii)

$$\sup_{(\alpha,t)\in\mathcal{N}_{osn}\times\mathcal{T}_n}\frac{\sqrt{n}\left|f(\alpha+tu^*_n)-f(\alpha_0)-\frac{df(\alpha_0)}{d\alpha}[\alpha+tu^*_n-\alpha_0]\right|}{||v^*_n||}=o(1);$$

(iii)

$$\frac{\sqrt{n}\left|\frac{df(\alpha_0)}{d\alpha}[\alpha_{0,n}-\alpha_0]\right|}{||v^*_n||}=o(1).$$

Assumption 12. *(i) $\frac{dm(\mathbf{X},\mathbf{Z};\alpha_0)}{d\alpha}[u^*_n]\in\Lambda^{s_d}_{c_d}(Supp(\mathbf{X},\mathbf{Z}))$ for some $c_d>0$ and $s_d>0$; (ii) $\sup_{y,\mathbf{x},\mathbf{z}}|f'_{Y^*_t|D_{t=1},\mathbf{X},\mathbf{Z}}(y|\mathbf{x},\mathbf{z})|$ is bounded; (iii) $g_0(x,\gamma)$ and $r_0(\mathbf{x},p)$ are twice continuously differentiable with respect to γ and p, respectively, and the derivatives are uniformly bounded;*

Assumption 13. *$M_n^2\delta_{2,n}=o(1)$; $M_n\delta^2_{2,n}=o\left(n^{-1/2}\right)$; $n\delta^2_{2,n}\left(M_n\delta_{2,n}(k_{g,n}+k_{r,n}+k_{p,n})\right)^{1/2}=o(1)$;*

Assumption 14. $\sqrt{n}\mathbb{Z}_n\xrightarrow{d}N(0,1)$.

Assumption 15. *The following condition holds:*

$$\left|\frac{1}{n}\sum_{i}^{n}\left(\frac{dm(\mathbf{X}_i,\mathbf{Z}_i;\alpha_0)}{d\alpha}[u_n^*]\right)'\left(\frac{dm(\mathbf{X}_i,\mathbf{Z}_i;\alpha_0)}{d\alpha}[u_n^*]\right)-\mathbb{E}\left[\left(\frac{dm(\mathbf{X}_i,\mathbf{Z}_i;\alpha_0)}{d\alpha}[u_n^*]\right)'\left(\frac{dm(\mathbf{X}_i,\mathbf{Z}_i;\alpha_0)}{d\alpha}[u_n^*]\right)\right]\right|$$
$$=o_p(1)$$

Assumption 11 is the same to Assumption 3.5 in Chen and Pouzo (2015), which restricts the local behavior of the functional $f(\cdot)$. Condition (i) restricts how fast k_n grows with n. Condition (ii) imposes a restriction on the nonlinearity bias of $f(\cdot)$. Note that this condition is satisfied when f is a linear functional. Condition (iii) is a undersmoothing condition which requires that the order of the sieve bias term $\frac{df(\alpha_0)}{d\alpha}[\alpha_{0,n}-\alpha_0]$ be smaller than that of the sieve standard deviation term. Similarly, Assumption 12 is imposed to control the higher order terms in the asymptotic expansion. Assumption 13 further restricts the rates of $k_{g,n}$, $k_{r,n}$, and $k_{p,n}$. Assumption 14 can be implied by, for example, Lindberg's condition. Assumption 15 is needed to establish the asymptotic behavior of the SQLR statistic.

Theorem 6.3. *Suppose that Assumptions 1, 2, and 3–10 hold. If Assumptions 11–14 are satisfied, then*

$$\sqrt{n}\frac{f(\hat{\alpha}_n)-f(\alpha_0)}{||v_n^*||_{sd}}\xrightarrow{d}N(0,1).$$

If Assumption 15 additionally holds, then under $H_0: f(\alpha_0)=f_0$,

$$||u_n^*||^2\times\widehat{QLR}_n(f_0)\xrightarrow{d}\chi^2(1).$$

The result of Theorem 6.3 holds for the functional f that may or may not be $\sqrt{n}$-estimable. However, it is needed to estimate the sieve Riesz representer v_n^* and $||v_n^*||_{sd}$, which may be challenging. We propose to use a bootstrap procedure to circumvent this difficulty.

Let $(B_i)_{i=1}^n$ be a sequence of i.i.d. nonnegative random variables independent of the data such that $\mathbb{E}[B]=1$, $Var(B)=\sigma_b^2$, and $\mathbb{E}\left[|B|^{2+\epsilon}\right]<\infty$ for some $\epsilon>0$. Define

$$\hat{m}^B_{j,t,n}(\mathbf{X},\mathbf{Z};\alpha)=\phi^{J_n}(\mathbf{X},\mathbf{Z})'\left(\sum_{i=1}^n\phi^{J_n}(\mathbf{X}_i,\mathbf{Z}_i)\cdot\phi^{J_n}(\mathbf{X}_i,\mathbf{Z}_i)'\right)^{-1}\sum_{i=1}^n\phi^{J_n}(\mathbf{X}_i,\mathbf{Z}_i)'\rho_{j,t}(\mathbf{Y}_i,\mathbf{D}_i,\mathbf{X}_i,\mathbf{Z}_i;\alpha)\cdot B_i$$

for each $j=1,2$ and $t=1,2,...,T$ as a bootstrap version of $\hat{m}_{j,t,n}(\mathbf{X},\mathbf{Z};\alpha)$, and $\hat{m}_n^B$ is defined in a similar way to $\hat{m}_n$ but with $\hat{m}^B_{j,t,n}(\mathbf{X},\mathbf{Z};\alpha)$. Let

$$\hat{Q}_n^B(\alpha)\equiv\frac{1}{n}\sum_{i}^{n}\hat{m}_n^B(\mathbf{X}_i,\mathbf{Z}_i;\alpha)'\hat{\Sigma}_n(\mathbf{X}_i,\mathbf{Z}_i)^{-1}\hat{m}_n^B(\mathbf{X}_i,\mathbf{Z}_i;\alpha)$$

and

$$\hat{\alpha}_n^B \equiv \arg\min_{\alpha \in \mathcal{A}_n} \left\{ \hat{Q}_n^B(\alpha) + \lambda_n \text{Pen}(\alpha) \right\}.$$

Denote $\hat{f}_n \equiv f(\hat{\alpha}_n)$, then we define the bootstrap SQLR test statistic as follows:

$$\widehat{QLR}_n^B(\hat{f}_n) \equiv n \left(\inf_{\alpha \in \mathcal{A}_n : f(\alpha) = \hat{f}_n} \hat{Q}_n^B(\alpha) - \hat{Q}_n^B\left(\hat{\alpha}_n^B\right) \right).$$

One can use a standard bootstrap procedure to mimic the asymptotic distribution of $||u_n^*||^2 \times \widehat{QLR}_n(f_0)$ and it is easy to test $H_0 : f(\alpha_0) = f_0$ or construct a confidence interval for $f(\alpha_0)$ without estimating $||v_n^*||_{sd}$ or $||u_n^*||$. Since, as pointed out by Chen and Pouzo (2015), the bootstrap validity holds under virtually the same conditions in Theorem 6.3, we do not provide its proof.

7 Simulation

In this section, we present results of Monte-Carlo simulations to examine the finite-sample performance of the estimators. We first generate (X_t^*, Z_t^*) from a bivariate normal distribution and let $X_t = \Phi(X_t^*)$ and $Z_t = \Phi(Z_t^*)$, where $\Phi(\cdot)$ is the standard normal distribution function. We consider several quantile levels ($u \in \{0.25, 0.5, 0.75\}$) and additively separable models given below:

$$\begin{aligned} Q_{Y_t|\mathbf{X},Z_t,D_t=1}(u|\mathbf{X}, Z_t, D_t = 1) =& g(X_t; u) + \mathbf{X}'\psi(u) + Q_{U_t|\mathbf{X},Z_t,D_t=1}(u|\mathbf{X}, Z_t, D_t = 1), \\ =& g(X_t; u) + \mathbf{X}'\psi(u) + h(P_t; u), \\ D_t =& \mathbf{1}(\xi_0 + X_t\xi_1 + Z_t\xi_2 \geq V_t), \end{aligned}$$

where $(U_t, V_t) \sim BVN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix}\right)$, $\psi(u) = (\psi_1(u), \psi_2(u), ..., \psi_T(u))'$ with $\psi_t(u) = 1 + 0.5 \times \Phi^{-1}(u)$ for each $u \in \{0.25, 0.5, 0.75\}$, $\xi_0 = -0.5$, $\xi_1 = -1$, and $\xi_2 = 1$. This data generating process yields that approximately 67.8% of individuals are selected for each time period. The number of time periods is set to be 3 (i.e., $T = 3$) for all simulations. The number of observations is set to be 500, and all simulation results are obtained from 500 iterations.

The structural function $g(\cdot; u)$ is specified as follows:

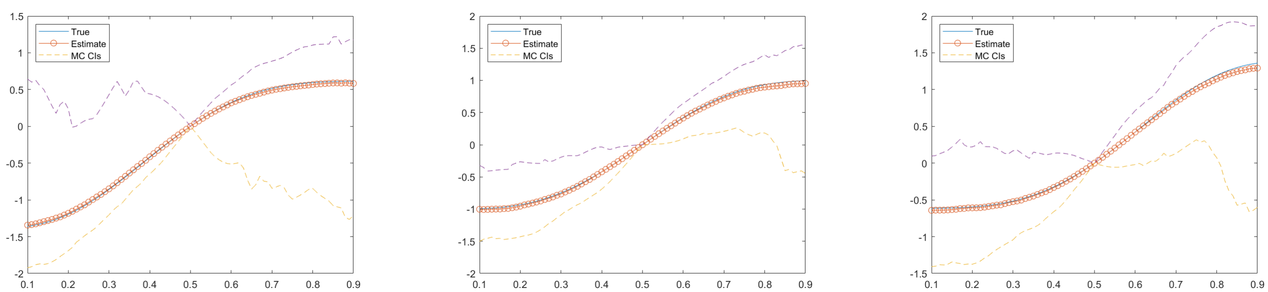
$$g(x; u) = 2 \times \{F_B(x; \alpha_m(u), \beta_m(u)) - F_B(0.5; \alpha_m(u), \beta_m(u))\},$$

where $F_B(\cdot; \alpha, \beta)$ the beta distribution function with parameters α and β. These parameters are set to be $\alpha_m(u) = 4 + \Phi^{-1}(u)$ and $\beta_m(u) = 4 - \Phi^{-1}(u)$ for each $u \in \{0.25, 0.5, 0.75\}$. This specification of g implies that the location normalization is satisfied with $\bar{x}(u) = 0.5$.

We use B-spline sieve spaces to approximate g and h, and the order of the sieve spaces is set to be proportional to $n^{1/7}$ with various combinations between the number of interior knots and the order of the B-spline function.[18] The selection probabilities are estimated by the series logit estimator of Hirano et al. (2003) with the B-Spline sieve spaces. The finite-sample performance of the sieve estimator of g is measured in terms of the integrated square bias ($IBIAS^2$), the integrated variance ($IVAR$), and the integrated mean squared error ($IMSE$) over $[0.1, 0.9]$.[19]

Figure 1 reports simulation results for each $u \in \{0.25, 0.5, 0.75\}$ with $n = 500$. Each panel show the true function $g(x; u)$ (the solid line), the median of sieve estimators $\hat{g}_n(x; u)$ for each x over 500 Monte-Carlo simulations (the circle-marked line), and 95% pointwise Monte-Carlo simulation confidence bands (the dashed lines). The order of spline functions and the number of interior knots are set to be 4.

Table 1 provides the Monte-Carlo simulation results with various combinations of the number of interior knots and the order of basis. All results are obtained from 500 simulations. One can find that the finite-sample performance of the sieve estimator $\hat{g}_n(\cdot; u)$ is not sensitive to the choice of k_{1n} and k_{2n}, where $k_{1n}$is the order of spline functions and k_{2n} is the number of interior knots, for $u = 0.5$. The results in Figure 1 and Table 1 confirm that the sieve estimator performs well even in finite samples.



$u = 0.25$ $\qquad$ $u = 0.5$ $\qquad$ $u = 0.75$

Figure 1: Simulation results for $u \in \{0.25, 0.5, 0.75\}$ with $n = 500$ and $k_n = 7$. The solid line is $g(\cdot; u)$, and the dashed line is the (medianl value of) sieve estimator $\hat{g}_n(\cdot; u)$. The dotted lines are 95% pointwise Monte-Carlo confidence bands. All simulation results are obtained from 500 iterations

[18]The complexity of B-spline sieve spaces (k_n) is determined by both the number of interior knots and the order of B-spline basis. Specifically, $k_n = k_{1n} + k_{2n}$, where k_{1n} and k_{2n} are the number of interior knots and the order of basis, respectively.

[19]The grid size is 0.01.

Table 1: Simulation results for various combinations of (k_{1n}, k_{2n}) and $n = 500$

u	k_n	(k_{1n}, k_{2n})	$IBIAS^2$	$IVAR$	$IMSE$
0.25	7	(3,4)	0.0046	0.1620	0.1666
0.5	7	(3,4)	0.0046	0.2077	0.2123
0.75	7	(3,4)	0.0023	0.1787	0.1810
0.25	7	(4,3)	0.0027	0.2891	0.2918
0.5	7	(4,3)	0.0014	0.1151	0.1165
0.75	7	(4,3)	0.0111	0.3162	0.3273
0.25	8	(4,4)	0.0053	0.2499	0.2552
0.5	8	(4,4)	0.0012	0.1016	0.1028
0.75	8	(4,4)	0.0032	0.1593	0.1625

8 Conclusion

In this paper, we develop a nonparametric panel quantile regression model with sample selection. The model is nonseparable and allows for time-invariant endogeneity in a similar spirit of the fixed effects models. To resolve the time-invariant endogeneity of the regressors and the sample selection bias, we adopt the CRE and control function approaches. In doing so, we avoid imposing any parametric or semiparametric restrictions on the distribution of the unobserved error terms, except for a conditional independence condition. The class of models is general and flexible enough to be extended to address many empirical issues about data, such as time-varying endogeneity and censoring. We study identification of the structural functions of the model. Identification requires that the number of time periods be greater than or equal to 3 ($T \geq 3$) and that there exist excluded variables that affect the selection probability. For practically tractable estimation, some semiparametric models are suggested, and a set of identification conditions for the semiparametric models is provided. Based on the identification result, we propose to use the PSMD estimation procedure to estimate the model parameters. We establish the consistency and convergence rates of the PSMD estimators under low-level conditions and provide a set of conditions under which the plug-in estimate of a functional of the parameter is asymptotically normal, regardless of whether the functional is $\sqrt{n}$-estimable or not. A small Monte-Carlo study confirms that the proposed estimators perform well in finite samples.

References

Abrevaya, J. and C. M. Dahl (2008). The effects of birth inputs on birthweight: evidence from quantile estimation on panel data. *Journal of Business & Economic Statistics 26*(4), 379–397. 1, 2, 2.2, 9

Ahn, H. and J. L. Powell (1993). Semiparametric estimation of censored selection models with a nonparametric selection mechanism. *Journal of Econometrics 58*(1-2), 3–29. 1

Altonji, J. G. and R. L. Matzkin (2005). Cross section and panel data estimators for nonseparable models with endogenous regressors. *Econometrica 73*(4), 1053–1102. 7, 1, 2, 3.1

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies 58*(2), 277–297. 1

Arellano, M. and S. Bonhomme (2012). Identifying distributional characteristics in random coefficients panel data models. *The Review of Economic Studies 79*(3), 987–1020. 2, 3.1

Arellano, M. and S. Bonhomme (2016). Nonlinear panel data estimation via quantile regressions. *The Econometrics Journal 19*(3), C61–C94. 1, 2, 2.3, 3.1, 4, 4

Arellano, M. and S. Bonhomme (2017). Quantile selection models with an application to understanding changes in wage inequality. *Econometrica 85*(1), 1–28. 1, 1, 2, 3.1

Besstremyannaya, G. and S. Golovan (2019). Reconsideration of a simple approach to quantile regression for panel data. *The Econometrics Journal 22*(3), 292–308. 1, 4

Bester, C. A. and C. Hansen (2009). Identification of marginal effects in a nonparametric correlated random effects model. *Journal of Business & Economic Statistics 27*(2), 235–250. 1, 2, 2, 3.1, 4, 4, 4.2

Blundell, R. and S. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics 87*(1), 115–143. 1

Blundell, R., X. Chen, and D. Kristensen (2007). Semi-nonparametric IV estimation of shape-invariant Engel curves. *Econometrica 75*(6), 1613–1669. 3.1

Buchinsky, M. (1998). The dynamics of changes in the female wage distribution in the USA: a quantile regression approach. *Journal of Applied Econometrics 13*(1), 1–30. 1, 1, 2, 2, 11, 5.2

Canay, I. A. (2011). A simple approach to quantile regression for panel data. *The Econometrics Journal 14*(3), 368–386. 1, 4

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *The Review of Economic Studies 47*(1), 225–238. (document), 1, 1, 2, 5.1, 5.1

Chamberlain, G. (1982). Multivariate regression models for panel data. *Journal of Econometrics 18*(1), 5–46. 2

Chen, S., S. Khan, and X. Tang (2020). Identification and estimation of weakly separable models without monotonicity. *arXiv preprint arXiv:2003.04337*. 1

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. *Handbook of Econometrics 6B*, 5549–5632. 6.1, 6.2.1

Chen, X. and T. M. Christensen (2018). Optimal sup-norm rates and uniform inference on nonlinear functionals of nonparametric IV regression. *Quantitative Economics 9*(1), 39–84. 6.2.1

Chen, X., O. Linton, and I. Van Keilegom (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica 71*(5), 1591–1608. A.1, A.3

Chen, X. and D. Pouzo (2012). Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals. *Econometrica 80*(1), 277–321. 6, A.1

Chen, X. and D. Pouzo (2015). Sieve Wald and QLR inferences on semi/nonparametric conditional moment models. *Econometrica 83*(3), 1013–1079. 1, 6, 15, 6.2.2, 6.2.2, A.1, A.2, A.1, A.1, A.2, A.3, A.4, A.3, A.3

Chernozhukov, V., I. Fernández-Val, J. Hahn, and W. Newey (2013). Average and quantile effects in nonseparable panel models. *Econometrica 81*(2), 535–580. 1, 2

Chernozhukov, V., I. Fernández-Val, S. Hoderlein, H. Holzmann, and W. Newey (2015). Nonparametric identification in panels using quantiles. *Journal of Econometrics 188*(2), 378–392. 1, 2

Chernozhukov, V., I. Fernández-Val, and A. E. Kowalski (2015). Quantile regression with censoring and endogeneity. *Journal of Econometrics 186*(1), 201–221. 3.2.1, 3.2.2

Chernozhukov, V., I. Fernández-Val, and S. Luo (2018). Distribution regression with sample selection, with an application to wage decompositions in the UK. *arXiv preprint arXiv:1811.11603*. 1, 1, 2, 3.1

Das, M., W. K. Newey, and F. Vella (2003). Nonparametric estimation of sample selection models. *The Review of Economic Studies 70*(1), 33–58. 1, 1, 2, 3.2.1, 5.2, 5.2

Donald, S. G. (1995). Two-step estimation of heteroskedastic sample selection models. *Journal of Econometrics 65*(2), 347–380. 1

Dustmann, C. and M. E. Rochina-Barrachina (2007). Selection correction in panel data models: an application to the estimation of females' wage equations. *The Econometrics Journal 10*(2), 263–293. 6

Escanciano, J. C., D. Jacho-Chávez, and A. Lewbel (2016). Identification and estimation of semiparametric two-step models. *Quantitative Economics 7*(2), 561–589. 3.1, 5.1

Evdokimov, K. (2010). Identification and estimation of a nonparametric panel data model with unobserved heterogeneity. *Department of Economics, Princeton University*. 1, 5.1

Fernández-Val, I., F. Peracchi, A. van Vuuren, and F. Vella (2019). Decomposing real wage changes in the United States. *arXiv preprint arXiv:1901.00419v2*. 2, 3.2.2

Fernández-Val, I. and M. Weidner (2018). Fixed effects estimation of large-T panel data models. *Annual Review of Economics 10*, 109–138. 3

Freyberger, J. and M. A. Masten (2019). A practical guide to compact infinite dimensional parameter spaces. *Econometric Reviews 38*(9), 979–1006. A.3

Galvao Jr, A. F. (2011). Quantile regression for dynamic panel data with fixed effects. *Journal of Econometrics 164*(1), 142–157. 1

Gayle, G.-L. and C. Viauroux (2007). Root-N consistent semiparametric estimators of a dynamic panel-sample-selection model. *Journal of Econometrics 141*(1), 179–212. 1, 4

Graham, B. S., J. Hahn, A. Poirier, and J. L. Powell (2018). A quantile correlated random coefficients panel data model. *Journal of Econometrics 206*(2), 305–335. 2, 3.1

Graham, B. S. and J. L. Powell (2012). Identification and estimation of average partial effects in "irregular" correlated random coefficient panel data models. *Econometrica 80*(5), 2105–2152. 2, 3.1

Gronau, R. (1974). Wage comparisons–A selectivity bias. *Journal of Political Economy 82*(6), 1119–1143. 1

Hahn, J., Z. Liao, and G. Ridder (2018). Nonparametric two-step sieve M estimation and inference. *Econometric Theory 34*(6), 1281–1324. 15

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica 47*(1), 153–161. 1, 1, 1, 2, 2

Hirano, K., G. W. Imbens, and G. Ridder (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica 71*(4), 1161–1189. 6.2.1, 7

Hoderlein, S. and H. White (2012). Nonparametric identification in nonseparable panel data models with generalized fixed effects. *Journal of Econometrics 168*(2), 300–314. 1, 2

Huber, M. and B. Melly (2015). A test of the conditional independence assumption in sample selection models. *Journal of Applied Econometrics 30*(7), 1144–1168. 1

Imbens, G. W. and W. K. Newey (2009). Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica 77*(5), 1481–1512. 2, 2, 3.1, 3.2.1, 3.2.2

Kato, K., A. F. Galvao Jr, and G. V. Montes-Rojas (2012). Asymptotics for panel quantile regression models with individual effects. *Journal of Econometrics 170*(1), 76–91. 1, 5

Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis 91*(1), 74–89. 1

Koenker, R. (2005). *Quantile Regression*. Econometric Society Monographs. Cambridge University Press. 10

Koenker, R. and G. Bassett (1978). Regression quantiles. *Econometrica 46*(1), 33–50. 1, 2.3

Kyriazidou, E. (1997). Estimation of a panel data sample selection model. *Econometrica 65*(6), 1335–1364. 1, 6

Kyriazidou, E. (2001). Estimation of dynamic panel data sample selection models. *The Review of Economic Studies 68*(3), 543–572. 1, 4

Laage, L. (2019). A correlated random coefficient panel model with time-varying endogeneity. *Working Paper*. 2, 3.1, 14

Lee, S. (2007). Endogeneity in quantile regression models: A control function approach. *Journal of Econometrics 141*(2), 1131–1158. 3.2.1

Lewbel, A. and O. Linton (2007). Nonparametric matching and efficient estimators of homothetically separable functions. *Econometrica 75*(4), 1209–1227. 3.1

Matzkin, R. L. (2007). Nonparametric identification. *Handbook of Econometrics 6B*, 5307–5368. 1, 3.1

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica 46*(1), 69–85. (document), 1, 1, 2, 3.1, 5.1

Newey, W. K. (1997). Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics 79*(1), 147–168. 6.2.1, 6.2.1, A.1

Newey, W. K. (2009). Two-step series estimation of sample selection models. *The Econometrics Journal 12*, S217–S229. 1, 1, 2

Newey, W. K. and J. L. Powell (2003). Instrumental variable estimation of nonparametric models. *Econometrica 71*(5), 1565–1578. 3.1

Newey, W. K., J. L. Powell, and F. Vella (1999). Nonparametric estimation of triangular simultaneous equations models. *Econometrica 67*(3), 565–603. 3.2.1

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

Powell, J. L. (1986). Censored regression quantiles. *Journal of Econometrics 32*(1), 143–155. 3.2.2

Semykina, A. and J. M. Wooldridge (2010). Estimating panel data models in the presence of endogeneity and selection. *Journal of Econometrics 157*(2), 375–380. 1, 3.2.1

Semykina, A. and J. M. Wooldridge (2013). Estimation of dynamic panel data models with sample selection. *Journal of Applied Econometrics 28*(1), 47–61. 1, 1, 4

Torgovitsky, A. (2015). Identification of nonseparable models using instruments with small support. *Econometrica 83*(3), 1185–1197. 1

Van de Geer, S. A. (2000). *Empirical processes in M-estimation*. Cambridge University Press. 6.2.1, A.3

Van der Vaart, A. W. and J. A. Wellner (1996). *Weak convergence and empirical processes*. Springer. A.3

Vytlacil, E. and N. Yildiz (2007). Dummy endogenous variables in weakly separable models. *Econometrica 75*(3), 757–779. 1, 3.1

Wooldridge, J. M. (1995). Selection corrections for panel data models under conditional mean independence assumptions. *Journal of Econometrics 68*(1), 115–132. 1, 6

Wooldridge, J. M. (2005). Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics 20*(1), 39–54. 4

A Proofs of the Results in Section 6

In this section, we provide mathematical proofs of the main results in Section 6. We introduce notation that will be used in the proofs. For any positive real sequences $\{a_n\}$ and $\{b_n\}$, $a_n \lesssim b_n$ means that there exist a finite constant $C > 0$ and $N \in \mathbb{N}$ such that $a_n \leq Cb_n$ for all $n \geq N$. If $a_n \lesssim b_n$ and $b_n \lesssim a_n$, it is denoted by $a_n \asymp b_n$. Let $(\mathcal{F}, ||\cdot||_{\mathcal{F}})$ be a metric space of real valued function $f : \mathcal{X} \to \mathbb{R}$. The covering number $N(\epsilon, \mathcal{F}, ||\cdot||_{\mathcal{F}})$ is the minimum number of $||\cdot||_{\mathcal{F}}$ ϵ-balls that cover $\mathcal{F}$. The entropy is the logarithm of the covering number. An ϵ-bracket in $(\mathcal{F}, ||\cdot||_{\mathcal{F}})$ is a pair of functions $l, u \in \mathcal{F}$ such that $||l||_{\mathcal{F}}, ||u||_{\mathcal{F}} < \infty$ and $||u - l||_{\mathcal{F}} \leq \epsilon$. The covering number with bracketing $N_{[]}(\epsilon, \mathcal{F}, ||\cdot||_{\mathcal{F}})$ is the minimum number of $||\cdot||_{\mathcal{F}}$ ϵ-brackets that cover $\mathcal{F}$. The entropy with bracketing is the logarithm of the covering number with bracketing. The bracketing integral is defined as $\int_0^\delta \sqrt{\log N_{[]}(\epsilon, \mathcal{F}, ||\cdot||_{\mathcal{F}})} d\epsilon$, and it is denoted by $J_{[]}(\delta, \mathcal{F}, ||\cdot||_{\mathcal{F}})$. $\otimes$ denotes the Kronecker product. Let C denote a generic positive and finite constant. It can be different across where it appears.

A.1 Proof of Theorem 6.1

Some empirical processes may not be measurable, and thus the expectation operator cannot be applied to those processes. In such a case, one can replace the expectation operator with the outer expectation operator. We use the notation $\mathbb{E}[\cdot]$ mainly to indicate the expectation operator, but it may also stand for the outer expectation if its argument is not measurable.

Lemma A.1. *Suppose that Assumptions 1, 2, and 3– 5 hold. Then, Assumption 3.1 in Chen and Pouzo (2015) is satisfied.*

Proof. We begin with verifying Condition (i) of Assumption 3.1 in Chen and Pouzo (2015). Suppose that there exists $\alpha \in \mathcal{A}$ such that $\mathbb{E}[\rho_u(\mathbf{Y}, \mathbf{D}, \mathbf{X}, \mathbf{Z}; \alpha)|\mathbf{X}, \mathbf{Z}] = 0$ a.s. Then, for each $t = 1, 2, ..., T$,

$$\begin{aligned}\mathbb{E}[\rho_{1,t}(\mathbf{Y}, \mathbf{D}, \mathbf{X}, \mathbf{Z}; \alpha)|\mathbf{X}, \mathbf{Z}] &= \mathbb{E}\left[D_t - p(X_t, Z_t)\Big|\mathbf{X}, \mathbf{Z}\right] \\ &= p_0(X_t, Z_t) - p(X_t, Z_t) = 0\end{aligned}$$

almost surely; and therefore, $||p_0(X_t, Z_t) - p(X_t, Z_t)||_2 = 0$. Consider

$$\begin{aligned}&\mathbb{E}[\rho_{2,t}(\mathbf{Y}, \mathbf{D}, \mathbf{X}, \mathbf{Z}; \alpha)|\mathbf{X}, \mathbf{Z}] \\ =&\mathbb{E}\left[D_t \left\{\mathbf{1}\left(Y_t - m(X_t, r(\mathbf{X}, p(X_t, Z_t); u); u) \leq 0\right) - u\right\} \Big|\mathbf{X}, \mathbf{Z}\right] \\ =&p_0(X_t, Z_t) \cdot \left\{F_{Y_t^*|\mathbf{X},\mathbf{Z},D_t=1}(g(X_t, r(\mathbf{X}, p(X_t, Z_t); u); u) - F_{Y_t^*|\mathbf{X},\mathbf{Z},D_t=1}(g_0(X_t, r_0(\mathbf{X}, p_0(X_t, Z_t); u); u)\right\}.\end{aligned}$$

Since $\inf_{(x,z)\in Supp(\mathbf{X},\mathbf{Z})} p_0(x,z) > 0$, one obtains that

$$\left\{F_{Y_t^*|\mathbf{X},\mathbf{Z},D_t=1}(g(X_t, r(\mathbf{X}, p(X_t, Z_t); u); u)|\mathbf{X}, \mathbf{Z}) - F_{Y_t^*|\mathbf{X},\mathbf{Z},D_t=1}(g_0(X_t, r_0(\mathbf{X}, p_0(X_t, Z_t); u); u)|\mathbf{X}, \mathbf{Z})\right\} = 0$$

almost surely. Since $F_{Y_t^*|\mathbf{X},\mathbf{Z},D_t=1}$ is strictly increasing, it follows that

$$g(X_t, r(\mathbf{X}, p(X_t, Z_t); u); u) = g_0(X_t, r_0(\mathbf{X}, p_0(X_t, Z_t); u); u)$$

almost surely, which implies that $g(X_t, r(\mathbf{X}, p(X_t, Z_t); u); u)$ is the u-th conditional quantile of Y_t given $D_t = 1$, $\mathbf{X}$, and $\mathbf{Z}$. Under Assumptions 1 and 2, the structural parameters are identified, and thus $||g - g_0||_\infty = ||r - r_0||_\infty = ||p - p_0||_\infty = 0$.Therefore, condition (i) of Assumption 3.1 in Chen and Pouzo (2015) is satisfied.

Condition (ii) of Assumption 3.1 in Chen and Pouzo (2015) is met by Assumption 5. Condition (iii) of Assumption 3.1 in Chen and Pouzo (2015) is implied by continuity of $F_{Y_t^*|D_t=1,\mathbf{X},\mathbf{Z}}(\cdot|\mathbf{x}, \mathbf{z})$ (Assumption 3). Condition (iv) of Assumption 3.1 in Chen and Pouzo (2015) holds as we choose the weighting matrix to be the identity matrix. As a result, Assumption 3.1 in Chen and Pouzo (2015) is satisfied. ■

Lemma A.2. *Suppose that Assumptions 4, 5, and 6 hold. Then, Assumption 3.3 in Chen and Pouzo (2015) holds.*

Proof. It is obvious that Condition (i) of Assumption 3.3 in Chen and Pouzo (2015) is met under Assumptions 4 and 5. I verify the conditions of Lemma C.2-(iii) in Chen and Pouzo (2012). Assumption C.1 in Chen and Pouzo (2012) is met under Assumptions 3 and 6. Note that $\sup_{\alpha\in\mathcal{A}_n} |\rho(\mathbf{W}, \alpha)| \leq 2$, and thus Condition (i) of Assumption C.2 in Chen and Pouzo (2012) is satisfied. Assumption 6 implies that condition (ii) of Assumption C.2 in Chen and Pouzo (2012) is satisfied with $b_{m,J} = J_n^{-p_m/T\cdot(d_x+d_z)}$ (e.g., Newey (1997)).

Finally, note that

$$\begin{aligned}
&\sup_{\tilde{\alpha}\in\mathcal{A}_n^{M_0}:||\alpha-\tilde{\alpha}||_\infty\leq\delta} \left|\rho(\mathbf{A}, \alpha) - \rho(\mathbf{A}, \tilde{\alpha})\right|^2 \\
\leq &\sup_{\tilde{\alpha}\in\mathcal{A}_n^{M_0}:||\alpha-\tilde{\alpha}||_\infty\leq\delta} D_t \left\{\mathbf{1}\left(Y_t \leq g(X_t, r(\mathbf{X}, p(X_t, Z_t))) - \delta\right) - \mathbf{1}\left(Y_t \leq g(X_t, r(\mathbf{X}, p(X_t, Z_t))) + \delta\right)\right\}
\end{aligned}$$

by the same logic to Chen et al. (2003, p.1600). Therefore, under Assumptions 4 and 6, it follows that

$$\max_{j\leq J_n}\mathbb{E}\left[\left[\phi_j^{J_n}(\mathbf{X},\mathbf{Z})\right]^2\cdot\sup_{\tilde{\alpha}\in\mathcal{A}_n^{M_0}:||\alpha-\tilde{\alpha}||_\infty\leq\delta}\left|\rho(\mathbf{A},\alpha)-\rho(\mathbf{A},\tilde{\alpha})\right|^2\right]$$
$$\lesssim\max_{j\leq J_n}\mathbb{E}\left[\left[\phi_j^{J_n}(\mathbf{X},\mathbf{Z})\right]^2\cdot\left\{F_{Y_t^*|\mathbf{X},\mathbf{Z},D_t=1}(g(X_t,r(\mathbf{X},p(X_t,Z_t)))-\delta)-F_{Y_t^*|\mathbf{X},\mathbf{Z},D_t=1}(g(X_t,r(\mathbf{X},p(X_t,Z_t)))+\delta)\right.\right.$$
$$\lesssim\max_{j\leq J_n}\mathbb{E}\left[\left[\phi_j^{J_n}(\mathbf{X},\mathbf{Z})\right]^2\cdot\delta\right]\leq K^2\delta$$

for some positive K. By Remark C.1 in Chen and Pouzo (2012), Assumption C.2-(iv) is also satisfied. In all, there exist finite constant $K>0$ and $\delta_{m,n}^2=\max\left\{\frac{J_n}{n},J_n^{-2p_m/T\cdot(d_x+d_z)}\right\}$ such that

$$K\mathbb{E}\left[||m(\mathbf{X},\mathbf{Z};\alpha)||_E^2\right]-O_p(\delta_{m,n}^2)\leq\frac{1}{n}\sum_i^n||\hat{m}_n(\mathbf{X}_i,\mathbf{Z}_i;\alpha)||_E^2\leq K\mathbb{E}\left[||m(\mathbf{X},\mathbf{Z};\alpha)||_E^2\right]+O_p(\delta_{m,n}^2)$$

uniformly over $\alpha\in\mathcal{A}_{osn}$. Under Assumption 6, $\delta_{g,n}^2=o(1)$; and therefore, Assumption 3.3-(ii) in Chen and Pouzo (2015) is satisfied. ■

Proof of Theorem 6.1

Proof. Under the set of conditions imposed in Theorem 6.1, Lemma A.1 shows that Assumption 3.1 in Chen and Pouzo (2015) is satisfied. Assumption 3.2 in Chen and Pouzo (2015) is directly imposed by Assumption 7. Assumption 3.3 in Chen and Pouzo (2015) is implied by Lemma 6. Therefore, applying Lemma 3.1 in Chen and Pouzo (2015) establishes the result of Theorem 6.1. ■

A.2 Proof of Theorem 6.2

Proof. Since the conditional density of Y_t^* on $D_t=1$, $\mathbf{X}$, and $\mathbf{Z}$ is uniformly bounded for any $t=1,2,...,T$ under Assumption 8, it is straightforward to see that

$$||\alpha_1-\alpha_2||^2\asymp||\alpha_1-\alpha_2||_2^2. \tag{20}$$

This implies that the sieve measure of local ill-posedness is constant (cf. Chen and Pouzo (2015, p.1030)) and that Assumption 3.4-(i) in Chen and Pouzo (2015) is satisfied. Assumption 9 implies that

$$\mathbb{E}\left[||m(\mathbf{X},\mathbf{Z};\alpha)||_E^2\right]\asymp||\alpha-\alpha_0||^2$$

for all $\alpha\in\mathcal{A}_{os}$; and therefore, condition (ii) of Assumption 3.4 in Chen and Pouzo (2015) is also met.

Define $Q_0(\alpha)\equiv\mathbb{E}\left[||m(\mathbf{X},\mathbf{Z};\alpha)||_E^2\right]$. Then, one obtains that, under Assumptions 5 and

8,

$$
\begin{aligned}
&Q_0(\pi_n\alpha_0)\\
\lesssim&\sum_{t=1}^{T}\left\{|\pi_n p_0(X_t,Z_t)-p_0(X_t,Z_t)|^2+|\{F_{Y_t^*|D_t=1,\mathbf{X},\mathbf{Z}}\left(\pi_n g_0(X_t,\pi_n r_0(\mathbf{X},\pi_n p_0(X_t,Z_t)))|\mathbf{X},\mathbf{Z}\right)-u\}p_0(X_t,Z_t)|^2\right\}\\
=&O\left(k_{p,n}^{-2\sigma_p}\right)+O\left(||\pi_n\alpha_0-\alpha_0||_2^2\right)\lesssim O\left(k_{g,n}^{-2\sigma_g}+k_{r,n}^{-2\sigma_r}+k_{p,n}^{-2\sigma_p}\right)
\end{aligned}
$$

since $|p_0(x,z)|\leq 1$ and $u=F_{Y_t^*|D_t=1,\mathbf{X},\mathbf{Z}}\left(g_0(X_t,r_0(\mathbf{X},p_0(X_t,Z_t)))|\mathbf{X},\mathbf{Z}\right)$. In addition, Lemma A.2 shows that $\hat{Q}_n(\alpha)\geq cQ(\alpha)-O_p\left(\delta_{m,n}^2\right)$ for some constant $c>0$ uniformly over $\mathcal{A}_{osn}$. This, together with Assumption 10-(i), results in that condition (iii) of Assumption 3.4 in Chen and Pouzo (2015) holds. Lastly, Assumptions 7-(i) directly implies condition (iv) of Assumption 3.4 in Chen and Pouzo (2015).

In all, the conditions of Lemma 3.2 Chen and Pouzo (2015) are satisfied, and thus the result of Theorem 6.2 follows from Lemma 3.2 in Chen and Pouzo (2015). ∎

A.3 Proof of Theorem 6.3

Let $\mathcal{A}_n^{C_0}\equiv\{\alpha\in\mathcal{A}_n:\lambda_n\text{Pen}(\alpha)\leq\lambda_n C_0\}$ and $\mathcal{O}_{on}\equiv\{\rho(\cdot,\alpha)-\rho(\cdot,\alpha_0):\alpha\in\mathcal{N}_{osn}\}$. Before proving Theorem 6.3, we provide a lemma that will be used in the proof.

Lemma A.3. *Let $\mathcal{F}$ be a class of function of a random vector X and $\mathcal{G}\equiv\{\mathbf{1}(Y\leq f):f\in\mathcal{F}\}$ with a random variable Y. Suppose that $F_{Y|X}$ is absolutely continuous with respect to the Lebesgue measure with its density function $f_{Y|X}$ uniformly bounded by a finite constant C_f. Then,*

$$N_{[]}(\sqrt{C_f\epsilon},\mathcal{G},||\cdot||_2)\leq N_{[]}(\epsilon,\mathcal{F},||\cdot||_2).$$

Proof. Let $\epsilon>0$ be given and $\{(l_i,u_i):i=1,2,...,N_{[]}(\epsilon,\mathcal{F},||\cdot||_2)\}$ be a set of ϵ-brackets for $\mathcal{F}$. Define $\{(\mathbf{1}(Y\leq l_i),\mathbf{1}(Y\leq u_i)):i=1,2,...,N_{[]}(\epsilon,\mathcal{F},||\cdot||_2)\}$, and it will be shown that it is a set of $C\epsilon$-brackets for $\mathcal{G}$. It is straightforward to see that, for any $f\in\mathcal{F}$, there exists (l_i,u_i) such that $l_i\leq f\leq u_i$ and $||u_i-l_i||_2\leq\epsilon$, and thus that $\mathbf{1}(Y\leq l_i)\leq\mathbf{1}(Y\leq f)\leq\mathbf{1}(Y\leq u_i)$. In addition, it follows that

$$
\begin{aligned}
||\mathbf{1}(Y\leq u_i)-\mathbf{1}(Y\leq l_i)||_2^2&=\mathbb{E}[\mathbf{1}(Y\leq u_i)-\mathbf{1}(Y\leq l_i)]\\
&=\mathbb{E}[F_{Y|X}(u_i|X)-F_{Y|X}(l_i)]\\
&\leq C_f\cdot||u_i-l_i||_2\leq C_f\epsilon
\end{aligned}
$$

by the definition of (l_i,u_i) and a property of indicator functions. Therefore, $\{(\mathbf{1}(Y\leq l_i),\mathbf{1}(Y\leq u_i)):i=1,2,...,N_{[]}(\epsilon,\mathcal{F},||\cdot||_2)\}$ is a set of $\sqrt{C_f\cdot\epsilon}$-brackets for $\mathcal{G}$. By the minimality of the bracketing number, one obtains the inequality. ∎

The following lemma shows that Assumption 3.6-(i) in Chen and Pouzo (2015), which is a sufficient condition for the result in Theorem 6.3, holds under the conditions imposed in Theorem 6.3.

Lemma A.4. *Suppose the conditions of Theorem 6.3 hold. Then, Assumption 3.6-(i) in Chen and Pouzo (2015) holds.*

Proof. We verify the conditions in Lemma 5.1 in Chen and Pouzo (2015). Note that Assumptions 1, 2, and 3–10 imply Assumptions 3.1 – 3.4 in Chen and Pouzo (2015). Therefore, it is enough to show that Assumption 3.6 in Chen and Pouzo (2015) is satisfied. Since Assumption 3.6-(i) in Chen and Pouzo (2015) is implied by Assumptions A.4 – A.7 in Chen and Pouzo (2015), we verify these sufficient conditions.

Assumptions 3 and 6 imply that Assumption A.4 in Chen and Pouzo (2015) is satisfied with $\sup_{(x,z)\in Supp(\mathbf{X},\mathbf{Z}),\alpha\in\mathcal{A}_n^{M_0}} |m(x,z;\alpha) - \phi^{J_n}(x,z)'\beta^0_{m,n}| = O\left(J_n^{-\sigma_m}\right) = o(1)$.

Let $\overline{\rho}_n(\mathbf{X},\mathbf{Z}) \equiv \mathbf{1}_T \otimes \begin{bmatrix} 2 \\ 2 \end{bmatrix}$, then $\sup_{\alpha\in\mathcal{A}_n^{C_0}} |\rho(\mathbf{X},\mathbf{Z};\alpha)| \leq \overline{\rho}_n(\mathbf{X},\mathbf{Z})$ almost surely and $\mathbb{E}\left[\overline{\rho}_n(\mathbf{X},\mathbf{Z})\right] < \infty$. Therefore, condition (i) of Assumption A.5 in Chen and Pouzo (2015) is met. Since, for any $\alpha, \tilde{\alpha} \in \mathcal{N}_{osn}$, $||\alpha - \tilde{\alpha}||_2 \leq \delta$ implies that $||\alpha - \tilde{\alpha}||_\infty \leq C\delta$ for some $C > 0$ by the compact embedding theorem in Freyberger and Masten (2019) (i.e., Theorem 1 in Freyberger and Masten (2019)), it follows that

$$\mathbb{E}\left[\sup_{\alpha\in\mathcal{N}_{osn}:||\alpha-\tilde{\alpha}||_2\leq\delta} ||\rho(\mathbf{A},\alpha) - \rho(\mathbf{A},\alpha_0)||_E^2 \Big| \mathbf{X},\mathbf{Z}\right] \leq K(\mathbf{X},\mathbf{Z}) \cdot \delta,$$

and this implies that Assumption A.5-(ii) in Chen and Pouzo (2015) is satisfied with $\kappa = 1/2$.

Let $\mathcal{Q}_n \equiv \{g(\cdot, r(\cdot, p(\cdot,\cdot))) : \alpha \in \mathcal{A}_{osn}\}$. Then, by Lemma A.3, it follows that

$$N_{[]}\left(\omega\left(M_n\delta_{2,n}\right)^\kappa, \mathcal{O}_{on}, ||\cdot||_2\right) \leq N_{[]}\left(\frac{\omega^2\left(M_n\delta_{2,n}\right)^{2\kappa}}{C}, \mathcal{Q}_n, ||\cdot||_2\right)$$

under Assumption 6 -(iii). Since the class of functions, $\mathcal{Q}_n$, is Lipschitz with respect to $\alpha \in \mathcal{A}_n$ with a finite Lipschitz constant, applying Theorem 2.7.11 in Van der Vaart and Wellner (1996) results in that $N_{[]}\left(\frac{\omega^2(M_n\delta_{2,n})^{2\kappa}}{C}, \mathcal{Q}_n, ||\cdot||_2\right) \leq N\left(\frac{\omega^2(M_n\delta_{2,n})^{2\kappa}}{C'}, \mathcal{A}_n, ||\cdot||_2\right)$. Since $\mathcal{A}_n = \mathcal{G}_n \times \mathcal{R}_n \times \mathcal{P}_n$, the covering number $N\left(\frac{\omega^2(M_n\delta_{2,n})^{2\kappa}}{C'}, \mathcal{A}_n, ||\cdot||_2\right)$ is bounded by the product of $N\left(\frac{\omega^2(M_n\delta_{2,n})^{2\kappa}}{C_G}, \mathcal{M}_n, ||\cdot||_2\right)$, $N\left(\frac{\omega^2(M_n\delta_{2,n})^{2\kappa}}{C_R}, \mathcal{R}_n, ||\cdot||_2\right)$, and $N\left(\frac{\omega^2(M_n\delta_{2,n})^{2\kappa}}{C_P}, \mathcal{P}_n, ||\cdot||_2\right)$ for some finite constants C_G,C_R, and C_P. In all, it follows from Lemma 2.5 in Van de Geer (2000) that

$$N_{[]}\left(\omega\left(M_n\delta_{2,n}\right)^\kappa, \mathcal{O}_{on}, ||\cdot||_2\right) \lesssim \left(1 + \frac{C}{\omega^2\left(M_n\delta_{2,n}\right)^{2\kappa}}\right)^{k_{g,n}+k_{r,n}+k_{p,n}}$$

for some constant C and $\kappa = 1/2$. Therefore,

$$\begin{aligned}
&\int_0^1 \sqrt{1+\log\left(N_{[]}\left(\omega\left(M_n\delta_{2,n}\right)^{\kappa}, \mathcal{O}_{on}, ||\cdot||_2\right)\right)}d\omega \\
\lesssim&\int_0^1 \sqrt{1+(k_{g,n}+k_{r,n}+k_{p,n})\log\left(1+\frac{C}{\omega^2\left(M_n\delta_{2,n}\right)^{2\kappa}}\right)}d\omega \\
\lesssim&\sqrt{(k_{g,n}+k_{r,n}+k_{p,n})}\cdot\frac{1}{\left(M_n\delta_{2,n}\right)^{\kappa}}\int_0^1\sqrt{\frac{1}{\omega}}d\omega \\
=&O\left(\sqrt{(k_{g,n}+k_{r,n}+k_{p,n})}\cdot\frac{1}{\left(M_n\delta_{2,n}\right)^{\kappa}}\right),
\end{aligned}$$

where the second inequality holds as $\log(1+t) \leq \sqrt{t}$ for all $t \geq 0$. Let $\sqrt{D_n} \equiv \sqrt{(k_{g,n}+k_{r,n}+k_{p,n})}\cdot \frac{1}{(M_n\delta_{2,n})^{\kappa}}$. Then,

$$\begin{aligned}
&n\delta_{2,n}^2\left(M_n\delta_{2,n}\right)^{1/2}\sqrt{D_n}\max\left\{\left(M_n\delta_{2,n}\right)^{1/2}\sqrt{D_n}, M_n\right\} \\
=&n\delta_{2,n}^2\left(M_n\delta_{2,n}\right)^{1/2}\cdot\sqrt{(k_{g,n}+k_{r,n}+k_{p,n})} \\
=&o(1)
\end{aligned}$$

under Assumption 13; and thus, condition (iii) of Assumption A.5 in Chen and Pouzo (2015) is satisfied. Condition (iv) of Assumption A.5 in Chen and Pouzo (2015) is trivial under Assumption 13.

Let

$$\begin{aligned}
&B_{1,t}(\mathbf{X},\mathbf{Z};\alpha_0) \equiv g_0^{'}(X_t, r_0(\mathbf{X}, p_0(X_t, Z_t))), \\
&B_{2,t}(\mathbf{X},\mathbf{Z};\alpha_0) \equiv g_0^{'}(X_t, r_0(\mathbf{X}, p_0(X_t, Z_t)))r_0^{'}(\mathbf{X}, p_0(X_t, Z_t)),
\end{aligned}$$

where $g_0^{'}(x,r) \equiv \frac{\partial g_0(x,r)}{\partial r}$ and $r_0^{'}(\mathbf{x},p) \equiv \frac{\partial r_0(\mathbf{x},p)}{\partial p}$. Then,

$$\frac{dm_t(\mathbf{X},\mathbf{Z};\alpha_0)}{d\alpha}[u_n^*] = \begin{bmatrix} -u_{p.n}^* \\ f_{Y_t^*|D_t=1,\mathbf{X},\mathbf{Z}}(q_0|\mathbf{X},\mathbf{Z})\cdot p_0(X_t,Z_t)\left\{u_{g,n}^* + B_{1,t}(\mathbf{X},\mathbf{Z};\alpha_0)\cdot u_{r,n}^* + B_{2,t}(\mathbf{X},\mathbf{Z};\alpha_0)\cdot u_{p,n}^*\right\}\end{bmatrix}$$

for each $t = 1, 2, ..., T$.

Define

$$\tilde{m}(\mathbf{X}_i,\mathbf{Z}_i;\alpha_0) \equiv \phi^{J_n}(\mathbf{X}_i,\mathbf{Z}_i)^{'}\left(\sum_i^n \phi^{J_n}(\mathbf{X},\mathbf{Z})\phi^{J_n}(\mathbf{X},\mathbf{Z})^{'}\right)^{-}\sum_i^n \phi^{J_n}(\mathbf{X}_i,\mathbf{Z}_i)m(\mathbf{X}_i,\mathbf{Z}_i;\alpha_0)$$

and

$$\frac{d\tilde{m}_t(\mathbf{X}_i,\mathbf{Z}_i;\alpha_0)}{d\alpha}[u_n^*] \equiv \phi^{J_n}(\mathbf{X}_i,\mathbf{Z}_i)^{'}\left(\sum_i^n \phi^{J_n}(\mathbf{X},\mathbf{Z})\phi^{J_n}(\mathbf{X},\mathbf{Z})^{'}\right)^{-}\sum_i^n \phi^{J_n}(\mathbf{X}_i,\mathbf{Z}_i)\frac{dm_t(\mathbf{X}_i,\mathbf{Z}_i;\alpha_0)}{d\alpha}[u_n^*]$$

for each $t = 1, 2, ..., T$.

Conditions (i) and (ii) of Assumption A.6 in Chen and Pouzo (2015) are satisfied by Assumptions 7-(i), 12-(i), and 13.

For condition (iii) of Assumption A.6 in Chen and Pouzo (2015), I verify the conditions of Lemma 1 in Chen et al. (2003). Let

$$\mathcal{F}_n \equiv \left\{ ||m(\mathbf{X}, \mathbf{Z}; \alpha)||_E^2 : \alpha \in \mathcal{A}_{osn} \right\}.$$

Note that

$$\begin{aligned} ||m(\mathbf{X}, \mathbf{Z}; \alpha)||_E^2 &= \sum_{t=1}^{T} ||m_t(\mathbf{X}, \mathbf{Z}; \alpha)||_E^2 \\ &= \sum_{t=1}^{T} \left\{ \left\{ p_0(X_t, Z_t) \left(F_{Y_t^*|D_t=1,\mathbf{X},\mathbf{Z}}(q|\mathbf{X}, \mathbf{Z}) - u \right) \right\}^2 + (p(X_t, Z_t) - p_0(X_t, Z_t))^2 \right\}, \end{aligned}$$

and this leads to that

$$\begin{aligned} &\mathbb{E}\left[\left(||m(\mathbf{X}, \mathbf{Z}; \alpha)||_E^2 - ||m(\mathbf{X}, \mathbf{Z}; \alpha_0)||_E^2 \right)^2 \right] \\ =&\mathbb{E}\left[||m(\mathbf{X}, \mathbf{Z}; \alpha)||_E^4 \right] \\ \lesssim&\mathbb{E}\left[\left\{ p_0(X_t, Z_t) \left(F_{Y_t^*|D_t=1,\mathbf{X},\mathbf{Z}}(q|\mathbf{X}, \mathbf{Z}) - u \right) \right\}^4 + (p(X_t, Z_t) - p_0(X_t, Z_t))^4 \right] \\ \lesssim&\mathbb{E}\left[\left(F_{Y_t^*|D_t=1,\mathbf{X},\mathbf{Z}}(q|\mathbf{X}, \mathbf{Z}) - u \right)^2 + (p(X_t, Z_t) - p_0(X_t, Z_t))^2 \right] \\ \lesssim&||\alpha - \alpha_0||_2^2 \end{aligned}$$

by the fact that $F_{Y_t^*|D_t=1,\mathbf{X},\mathbf{Z}}(q_0|\mathbf{X}, \mathbf{Z}) = u$ and Assumptions 4 and 6-(v). This implies that $||m(\mathbf{X}, \mathbf{Z}; \alpha)||_E^2$ is L_2-continuous at $\alpha = \alpha_0$. By using a similar argument and Theorem 2.7.11 in Van der Vaart and Wellner (1996), one obtains that

$$\int_0^\infty \log N_{[]}\left(\epsilon, \mathcal{F}_n, ||\cdot||_2\right) \leq \int_0^\infty \log N_{[]}\left(\frac{\epsilon}{C}, \mathcal{A}_n, ||\cdot||_2\right) = O\left(\sqrt{(k_{g,n} + k_{r,n} + k_{p,n})}\right) < \infty.$$

Applying Lemma 1 in Chen et al. (2003) results in that

$$\sup_{\mathcal{N}_{osn}} \frac{1}{n} \sum_{i}^{n} \left\{ ||m(\mathbf{X}, \mathbf{Z}; \alpha)||_E^2 - \mathbb{E}\left[||m(\mathbf{X}, \mathbf{Z}; \alpha)||_E^2 \right] \right\} = o_p\left(n^{-1/2}\right),$$

and thus, condition (iii) of Assumption A.6 in Chen and Pouzo (2015) is met. Define $\mathcal{F}_n \equiv \left\{ \frac{dm(\mathbf{X},\mathbf{Z};\alpha_0)}{d\alpha}[u_n^*]' \{ m(\mathbf{X}, \mathbf{Z}; \alpha) - m(\mathbf{X}, \mathbf{Z}; \alpha_0) \} : \alpha \in \mathcal{N}_{osn} \right\}$. By using the same argument to that of the proof of Proposition 6.1 in Chen and Pouzo (2015), it is enough to show that under Assumption 6,

$$||F_n(\mathbf{X}, \mathbf{Z})||_2 = o(1),$$

where $F_n(\mathbf{x},\mathbf{z}) \equiv \sup_{f\in\mathcal{F}_n} |f(\mathbf{x},\mathbf{z})|$. Observe that

$$
\begin{aligned}
||F_n||_2^2 &= \mathbb{E}\left[\left(\frac{dm(\mathbf{X},\mathbf{Z};\alpha_0)}{d\alpha}[u_n^*]' \sup_{\alpha\in\mathcal{N}_{osn}} |m(\mathbf{X},\mathbf{Z};\alpha) - m(\mathbf{X},\mathbf{Z};\alpha_0)|\right)^2\right] \\
&\leq \mathbb{E}\left[\frac{dm(\mathbf{X},\mathbf{Z};\alpha_0)}{d\alpha}[u_n^*]' \frac{dm(\mathbf{X},\mathbf{Z};\alpha_0)}{d\alpha}[u_n^*]\right] \cdot \mathbb{E}\left[\sup_{\alpha\in\mathcal{N}_{osn}} (m(\mathbf{X},\mathbf{Z};\alpha) - m(\mathbf{X},\mathbf{Z};\alpha_0))^2\right] \\
&\lesssim \mathbb{E}\left[\frac{dm(\mathbf{X},\mathbf{Z};\alpha_0)}{d\alpha}[u_n^*]' \frac{dm(\mathbf{X},\mathbf{Z};\alpha_0)}{d\alpha}[u_n^*]\right] \cdot \sup_{\alpha\in\mathcal{N}_{osn}} ||\alpha-\alpha_0||_2^2 \\
&\lesssim M_n^2 \delta_{2,n}^{2\kappa} = o(1).
\end{aligned}
$$

Then, one obatins that

$$
\sup_{\alpha\in\mathcal{N}_{osn}} \frac{1}{n}\sum_i^n \left\{\frac{dm(\mathbf{X},\mathbf{Z};\alpha_0)}{d\alpha}[u_n^*]' m(\mathbf{X},\mathbf{Z};\alpha) - \mathbb{E}\left[\frac{dm(\mathbf{X},\mathbf{Z};\alpha_0)}{d\alpha}[u_n^*]'\{m(\mathbf{X},\mathbf{Z};\alpha)\right]\right\} = o_p\left(n^{-1/2}\right).
$$

In all, all conditions of Assumptions A.6 in Chen and Pouzo (2015) are satisfied.

Condition (i) of Assumption A.7 in Chen and Pouzo (2015) is implied by Assumption 12-(ii).

Note that under Assumption 12-(ii), we have for each $t = 1, 2, ..., T$,

$$
\frac{d^2 m_{1,t}(\mathbf{X},\mathbf{Z};\alpha)}{d\alpha^2}[u_n^*, u_n^*] = 0
$$

and

$$
\begin{aligned}
&\frac{d^2 m_{2,t}(\mathbf{X},\mathbf{Z};\alpha)}{d\alpha^2}[u_n^*, u_n^*] \\
=& f'_{Y_t^*|D_t=1,\mathbf{X},\mathbf{Z}}(Q|\mathbf{X},\mathbf{Z}) p(X_t,Z_t)\left\{u_{g,n}^* + B_{1,t}(\mathbf{X},\mathbf{Z};\alpha)\cdot u_{r,n}^* + B_{2,t}(\mathbf{X},\mathbf{Z};\alpha)\cdot u_{p,n}^*\right\}\cdot u_{g,n}^* \\
&+ f'_{Y_t^*|D_t=1,\mathbf{X},\mathbf{Z}}(Q|\mathbf{X},\mathbf{Z}) p(X_t,Z_t)\left\{u_{g,n}^* + B_{1,t}(\mathbf{X},\mathbf{Z};\alpha)\cdot u_{r,n}^* + B_{2,t}(\mathbf{X},\mathbf{Z};\alpha)\cdot u_{p,n}^*\right\} g'(X_t, r(\mathbf{X}, p(X_t,Z_t))) u_{r,n}^* \\
&+ f_{Y_t^*|D_t=1,\mathbf{X},\mathbf{Z}}(Q|\mathbf{X},\mathbf{Z})\cdot p(X_t,Z_t)\left\{\partial_r B_{1,t}(\mathbf{X},\mathbf{Z};\alpha) u_{r,n}^* + \partial_r B_{2,t}(\mathbf{X},\mathbf{Z};\alpha) u_{p,n}^*\right\} u_{r,n}^* \\
&+ \left\{u_{g,n}^* + B_{1,t}(\mathbf{X},\mathbf{Z};\alpha)\cdot u_{r,n}^* + B_{2,t}(\mathbf{X},\mathbf{Z};\alpha)\cdot u_{p,n}^*\right\}\cdot C_t(\mathbf{X},\mathbf{Z};\alpha)\cdot u_{p,n}^* \\
&+ f_{Y_t^*|D_t=1,\mathbf{X},\mathbf{Z}}(Q_0|\mathbf{X},\mathbf{Z})\cdot p(X_t,Z_t)\cdot\left\{\partial_p B_{1,t}(\mathbf{X},\mathbf{Z};\alpha) u_{r,n}^* + \partial_p B_{2,t}(\mathbf{X},\mathbf{Z};\alpha) u_{p,n}^*\right\} u_{p,n}^*,
\end{aligned}
$$

where $C_t(\mathbf{X},\mathbf{Z};\alpha) \equiv f'_{Y_t^*|D_t=1,\mathbf{X},\mathbf{Z}}(Q_0|\mathbf{X},\mathbf{Z}) p_0(X_t,Z_t) g'(X_t, r(\mathbf{X}, p(X_t,Z_t)))\cdot r'(\mathbf{X}, p(X_t,Z_t)) + f_{Y_t^*|D_t=1,\mathbf{X},\mathbf{Z}}(Q_0|\mathbf{X},\mathbf{Z})$. Therefore, for any $\alpha\in\mathcal{N}_{osn}$, it follows from the mean value theorem that

$$
\begin{aligned}
\left|\left|\frac{dm(\mathbf{X},\mathbf{Z},\alpha)}{d\alpha}[u_n^*] - \frac{dm(\mathbf{X},\mathbf{Z},\alpha_0)}{d\alpha}[u_n^*]\right|\right|_E^2 &\leq \left|\left|\frac{d^2 m(\mathbf{X},\mathbf{Z};\tilde{\alpha})}{d\alpha^2}[\alpha-\alpha_0, u_n^*]\right|\right|_E^2 \\
&\lesssim ||u_n^*||_2^2 \cdot ||\alpha-\alpha_0||_E^2
\end{aligned}
$$

by conditions (ii) and (iii) of Assumption 12. Thus,

$$\mathbb{E}\left[\sup_{\alpha\in\mathcal{N}_{osn}}\left|\frac{dm(\mathbf{X},\mathbf{Z},\alpha)}{d\alpha}[u_n^*]-\frac{dm(\mathbf{X},\mathbf{Z},\alpha_0)}{d\alpha}[u_n^*]\right|\Big|_E^2\right](M_n\delta_{2,n})^2 \lesssim ||\alpha-\alpha_0||_2^2\,(M_n\delta_{2,n})^2 \lesssim (M_n\delta_{2,n})^4 = o\left(n^{-1}\right)$$

by Assumption 13, which implies that condition (ii) of Assumption A.7 in Chen and Pouzo (2015).

Condition (iii) of Assumption A.7 in Chen and Pouzo (2015) is satisfied under Assumptions 12 and 13 since

$$\mathbb{E}\left[\sup_{\alpha\in\mathcal{N}_{osn}}\left|\left|\frac{d^2m(\mathbf{X},\mathbf{Z};\alpha)}{d\alpha^2}[u_n^*,u_n^*]\right|\right|_E\right]\lesssim ||u_{g,n}^*||_2^2+||u_{r,n}^*||_2^2+||u_{p,n}^*||_2^2<\infty.$$

Lastly, note that one can show that for any $\alpha_1\in\mathcal{N}_{os}$ and $\alpha_2\in\mathcal{N}_{osn}$,

$$\begin{aligned}&\left|\left|\frac{dm(\mathbf{X},\mathbf{Z},\alpha_1)}{d\alpha}[\alpha_2-\alpha_0]-\frac{dm(\mathbf{X},\mathbf{Z},\alpha_0)}{d\alpha}[\alpha_2-\alpha_0]\right|\right|_E\\ \lesssim&||\alpha_2-\alpha_0||_E\cdot||\alpha_1-\alpha_0||_E\end{aligned}$$

by using a similar argument to before. In addition,

$$\begin{aligned}&\mathbb{E}\left[\left(\frac{dm(\mathbf{X},\mathbf{Z};\alpha_0)}{d\alpha}[u_n^*]\right)'\left(\frac{dm(\mathbf{X},\mathbf{Z},\alpha_1)}{d\alpha}[\alpha_2-\alpha_0]-\frac{dm(\mathbf{X},\mathbf{Z},\alpha_0)}{d\alpha}[\alpha_2-\alpha_0]\right)\right]\\ \leq&\sqrt{\mathbb{E}\left[\left|\left|\frac{dm(\mathbf{X},\mathbf{Z};\alpha_0)}{d\alpha}[u_n^*]\right|\right|_E^2\right]}\cdot\sqrt{\mathbb{E}\left[\left|\left|\left(\frac{dm(\mathbf{X},\mathbf{Z},\alpha_1)}{d\alpha}[\alpha_2-\alpha_0]-\frac{dm(\mathbf{X},\mathbf{Z},\alpha_0)}{d\alpha}[\alpha_2-\alpha_0]\right)\right|\right|_E^2\right]}\\ \lesssim&\sqrt{\mathbb{E}\left[\left|\left|\frac{dm(\mathbf{X},\mathbf{Z};\alpha_0)}{d\alpha}[u_n^*]\right|\right|_E^2\right]}\cdot||\alpha_2-\alpha_0||_2\cdot||\alpha_1-\alpha_0||_2\lesssim(M_n\delta_{2,n})^2=o\left(n^{-1/2}\right)\end{aligned}$$

by Assumptions 7-(i), 13, and 12-(iii). Thus, condition (iv) of Assumption A.7 in Chen and Pouzo (2015) is satisfied.

In all, it follows from Lemma 5.1 in Chen and Pouzo (2015) that Assumption 3.6(i) in Chen and Pouzo (2015) is satisfied. ■

Proof of the Theorem

Proof. Assumption 11 is identical to Assumption 3.5 in Chen and Pouzo (2015). Assumption 3.6(ii) in Chen and Pouzo (2015) is directly imposed by Assumption 14. By Lemma A.4, Assumptino 3.6(i) in Chen and Pouzo (2015) is satisfied. Therefore, applying Theorem 4.1 in Chen and Pouzo (2015) yields the result.

For the second result, Assumption B in Chen and Pouzo (2015) is imposed by Assumption 15. By Lemma 5.1 in Chen and Pouzo (2015) and the proof of Theorem 6.3, we conclude that conditions of Theorem 4.3 in Chen and Pouzo (2015) are all met. Since the PSMD estimator $\hat{\alpha}_n$ under Assumption 10 is not optimally weighted, the first part of Theorem 4.3 in Chen and Pouzo (2015) completes the proof. ■