博士論文

Essays on Identification and Estimation of Nonseparable Models
（非分離的モデルの識別と推定に関する研究）

石原　卓弥

Contents

*This chapter is based on Ishihara (2017).
†This chapter is based on Ishihara (2019).

Acknowledgement

I would like to express my sincere appreciation to my advisor Katsumi Shimotsu for his guidance, encouragement, and patience during my graduate studies. I could not have finished the work without his warm encouragement and insightful suggestions. I also would like to thank Fumiya Akashi, Hidehiko Ichimura, Kaoru Irie, Kengo Kato, Tatsuya Kubokawa, Yasuhiro Omori, and Shonosuke Sugasawa. Other faculty members at the University of Tokyo have provided an opportunity to learn cutting-edge knowledge, and seminar participants at various conferences have provided several comments.

Chapter 1

Overview

A purpose of this thesis is to propose a methodology for the identification and estimation of nonseparable models. A common practice when estimating many economic models is to use additive models such as $Y = g(X) + \epsilon$. However, in an economic model, it is rarely the case that the unobserved variables influence the dependent variables in an additive way. For example, if we consider the return to education, then additive models imply that the effect of education on earnings does not depend on the unobserved variables that include ability or family background. Hence, additive models cannot capture the unobserved heterogeneity. On the contrary, nonseparable models, $Y = g(X, \epsilon)$, capture the unobserved heterogeneity effect of explanatory variables on outcomes because these models allow the derivative of the structural function to depend on the unobserved variables.

This thesis provides novel methods for the identification and estimation of nonseparable models. First, we consider the partial identification of the nonseparable model using binary instruments. Second, we propose the identification and estimation approach of nonseparable panel data models. Finally, we generalize the nonseparable model proposed by Athey and Imbens (2006) and propose a tractable estimator of the quantile treatment effects. The organization of this thesis is as follows.

In Chapter 2, we explore the partial identification of nonseparable models with continuous endogenous and binary instrumental variables. We show that the structural function is partially identified when it is monotone or concave in the explanatory variable. D'Haultfœuille and Février (2015) and Torgovitsky (2015) prove the point identification of the structural function under two key assumptions: (1) the conditional distribution functions of the endogenous variable for different values of the instrumental variables have intersections and (2) the structural function is strictly increasing in the scalar unobservable variable. We demonstrate that, even if these two assumptions do not hold, monotonicity and concavity provide identifying power. Point identification is achieved when the structural function is flat or linear with respect to the explanatory variable over a given interval.

In Chapter 3, we explore the identification and estimation of nonseparable panel data models. We show that the structural function is nonparametrically identified when it is strictly increasing in a scalar unobservable variable, the conditional distributions of unobservable variables do not change over time, and the joint support of explanatory variables satisfies some weak assumptions. To identify the target parameters, existing studies assume that the structural function does not change over time, and that there are "stayers", namely individuals with the same regressor values in two time periods. Our approach, by contrast, allows the structural function to depend on the time period in an arbitrary manner and does not require the existence of stayers. In the estimation part, we propose parametric and nonparametric estimators that implement our identification results. Monte Carlo studies indicate that our parametric estimator performs well in finite samples. Finally, we extend our identification results to models with discrete outcomes, and show that the structural function is partially identified.

In Chapter 4, we explore the identification and estimation of the quantile treatment effects (QTE) by using panel data. We generalize the change-in-changes (CIC) model proposed by Athey and Imbens (2006) and propose a tractable estimator of the QTE. The CIC model allows for the estimation of the potential outcomes distribution and captures the heterogeneous effects of the treatment on the outcomes. However, there are two problems with the CIC model: (1) there is a lack of a tractable estimator in the presence of covariates and (2) the CIC estimator does not work when the treatment is continuous. Our model allows the presence of covariates and the continuous treatment. We propose a two-step estimation method based on the quantile regression and minimum distance method. We then show the consistency and asymptotic normality of our estimator. Monte Carlo studies indicate that our estimator performs well in finite samples. We use our method to estimate the impact of an insurance program on quantiles of household production.

Chapter 2

Partial Identification of Nonseparable Models using Binary Instruments*

2.1 Introduction

In this chapter, we examine the identification of a system of structural equations that takes the following form:

$$\begin{aligned} Y &= g(X, \epsilon) \\ X &= h(Z, \eta), \end{aligned} \tag{2.1}$$

where $Y \in \mathbb{R}$ is a scalar response variable, $X \in \mathbb{R}$ is a continuous endogenous variable, $Z \in \{0, 1\}$ is a binary instrument, and ϵ and η are unobservable scalar variables. For simplicity, we assume that X is a scalar variable. This specification is nonseparable in the unobservable variable ϵ and captures the unobserved heterogeneity in the effect of X on Y. Such models have also been considered by, for example, D'Haultfœuille and Février (2015) and Torgovitsky (2015).

D'Haultfœuille and Février (2015) and Torgovitsky (2015) show that g is point identified when $g(x, e)$ and $h(z, v)$ are strictly increasing in e and v and Z is independent of (ϵ, η). Their results are important for empirical analyses in which many instruments are binary or discrete, such as the intent-to-treat in a randomized controlled experiment or quarter of birth used by Angrist and Krueger (1991). For nonparametric models with a continuously distributed X, several point identification results require Z to be continuously distributed. See, for example, Newey, Powell, and Vella (1999) and Imbens and Newey (2009).

*This chapter is based on Ishihara (2017).

D'Haultfœuille and Février (2015) and Torgovitsky (2015) use two key assumptions when establishing point identification for g. First, $F_{X|Z}(x|0)$ and $F_{X|Z}(x|1)$ have intersections. Second, $g(x, e)$ is strictly increasing in e. However, many empirically important models do not satisfy these assumptions. For example, $F_{X|Z}(x|0)$ and $F_{X|Z}(x|1)$ do not have an intersection when Z has a strictly monotonic effect on X such as linear models $X = \beta_0 + \beta_1 Z + \eta$. Further, in many applications, instrumental variables have a strictly monotonic effect on endogenous variables (e.g. the LATE framework proposed by Imbens and Angrist (1994)). For example, as in Macours, Schady, and Vakis (2012), cash transfer programs have been implemented in several countries. As such, if we use treatment indicator Z as the instrumental variable for income X, Z has a strictly monotonic effect on X, which violates the intersection assumption. Hence, $F_{X|Z}(x|0)$ and $F_{X|Z}(x|1)$ never have an intersection in this example. Actually, in Section 2.5, we show that $F_{X|Z}(x|0)$ and $F_{X|Z}(x|1)$ do not have an intersection in the real data. When Y is discrete or censored, $g(x, e)$ is not strictly increasing in e. Moreover, many problems in economics involve dependent variables that are discrete or censored. For example, development economists may want to analyze the effects of income changes on child education. If school attendance is used as a dependent variable, then Y is discrete. As another example, assume that we want to analyze the effects of income changes on education expenditure. Then, education expenditure is censored at zero when children do not attend school.

This study shows that, when $g(x, e)$ is monotone or concave in x, we can partially identify g, even if $F_{X|Z}(x|0)$ and $F_{X|Z}(x|1)$ have no intersection and $g(x, e)$ is not strictly increasing in e. $g(x, e)$ is monotone or concave in x in many economic models. For example, the demand function is decreasing in price if the income effect is negligible, and economic analyses of production often suppose that the production function is monotone and concave in inputs. In general, the demand function is not decreasing in price. For instance, Hoderlein (2011) employs nonseparable models and analyzes consumer behavior without the monotonicity assumption. Many studies employ monotonicity or concavity to identify the target parameters (e.g., Manski (1997), Giustinelli (2011), D'Haultfoeuille, Hoderlein, and Sasaki (2013), and Okumura and Usui (2014)). Specifically, Manski (1997) imposes these assumptions and shows that the average treatment response is partially identified. The partial identification approach using the concavity assumption in this study is somewhat similar to that considered by D'Haultfoeuille et al. (2013).

In this model, monotonicity and concavity provide identifying power. D'Haultfœuille and Février (2015) and Torgovitsky (2015) show that when $F_{X|Z}(x|0)$ and $F_{X|Z}(x|1)$ have intersections, $g(x', g^{-1}(x, y))$ is identified for all x, x', and y, where $g^{-1}(x, y)$ is the inverse of g with respect to its last component. Then, g is point identified under appropriate normalization. By contrast, when $F_{X|Z}(x|0)$ and $F_{X|Z}(x|1)$ do not have intersections, we only identify $g(x', g^{-1}(x, y))$ for some x and x'. Although this information restricts the functional form of g, it does not provide the informative bounds of g. In this case, monotonicity and convexity allow us to interpolate or extrapolate $g(x', g^{-1}(x, y))$ and provide the informative bounds of $g(x', g^{-1}(x, y))$. For example, if $g(x', g^{-1}(x, y))$ is identified and $\tilde{x} \geq x'$, monotonicity implies $g(\tilde{x}, g^{-1}(x, y)) \geq g(x', g^{-1}(x, y))$, and hence, we obtain a lower bound of $g(\tilde{x}, g^{-1}(x, y))$. Using these bounds, we can achieve the partial identification of g.

There is a rich literature on the identification of nonseparable models using the control function approach. For example, Chesher (2007), Hoderlein and Mammen (2007), Florens, Heckman, Meghir, and Vytlacil (2008), Imbens and Newey (2009), Hoderlein and Mammen (2009), Hoderlein (2011), Kasy (2011), and Blundell, Kristensen, and Matzkin (2013) consider the identification of nonseparable models using the control function approach. Particularly, Imbens and Newey (2009) consider models similar to (2.1). Their study allows ϵ to be multivariate, showing that the quantile function of $g(x, \epsilon)$ is point identified, while in this analysis, ϵ is imposed as scalar. Their results need continuous instruments, whereas those of D'Haultfœuille and Février (2015), Torgovitsky (2015), and the present study do not.

We assume that the instrumental variable Z is binary. D'Haultfœuille and Février (2015) consider the case in which the instrumental variable takes more than two values, thus showing point identification can be achieved using group and dynamical systems theories even when $F_{X|Z}(x|z)$ and $F_{X|Z}(x|z')$ have no intersection.

Caetano and Escanciano (2017) provides alternative results for the identification of nonseparable models with continuous endogenous variables and binary instruments. To this end, they use the observed covariates to identify the structural function. Although their approach does not require $F_{X|Z}(x|z)$ and $F_{X|Z}(x|z')$ to intersect, they assume the structural function does not depend on the observed covariates. By contrast, our identification approach does not require the existence of covariates and allows the structural function to depend on the observed covariates.

The remainder of this study is organized as follows. Section 2.2 introduces the assumptions employed in the analysis. Sections 2.3 and 2.4 demonstrate the partial identification of g under the monotonicity and concavity assumptions when conditional distributions have no intersections. Section 2.5 computes the bounds using real data. Section 2.6 extends the result in Section 2.3 to a more general case, where we allow Y to be discrete or censored. Section 2.7 concludes the paper.

2.2 Model

For any random variable U and random vector W, let $F_{U|W}(u|w)$ denote the conditional distribution function of U conditional on W. In some places, we interchangeably use the notation $F_{U|W=w}(u)$ instead of $F_{U|W}(u|w)$. Let $\mathcal{X}$, $\mathcal{X}_z$, and $\mathcal{Y}_{x,z}$ denote the interiors of the support of X, $X|Z=z$, and $Y|X=x, Z=z$, respectively. The following two assumptions are the same as those in D'Haultfœuille and Février (2015) and Torgovitsky (2015):

Assumption 2.1. *The instrument is independent of the unobservable variables:* $Z \perp\!\!\!\perp (\epsilon, \eta)$.

Assumption 2.2. *(i) Function g is continuous and $g(x,e)$ is strictly increasing in e for all $x \in \mathcal{X}$. (ii) For all $z \in \{0,1\}$, $h(z,v)$ is continuous and strictly increasing in v.*

Assumptions 2.1 and 2.2 (ii) are typically employed when using the control function approach. See, for example, Imbens and Newey (2009), D'Haultfœuille and Février (2015), and Torgovitsky (2015). Although Assumption 2.2 (i) is strong, it is necessary for our identification approach. Hoderlein and Mammen (2007), Hoderlein and Mammen (2009), Hoderlein (2011), and Imbens and Newey (2009) do not employ this assumption. We relax part of Assumption 2.2 (i) in Section 2.6, where we assume $g(x,e)$ is nondecreasing in e.

The next assumption regarding the conditional distributions of X conditional on Z differs from that of D'Haultfœuille and Février (2015) and Torgovitsky (2015).

Assumption 2.3. *(i) $F_{X|Z}(x|z)$ is continuous in x for all $z \in \{0,1\}$ and $F_{X|Z}(x|0) < F_{X|Z}(x|1)$ for all $x \in \mathcal{X}$. (ii) $\mathcal{X}_0 = (\underline{x}_0, \overline{x}_0)$, $\mathcal{X}_1 = (\underline{x}_1, \overline{x}_1)$, and $-\infty < \underline{x}_1 < \underline{x}_0 < \overline{x}_1 < \overline{x}_0 < \infty$.*

Conditions (i) and (ii) above imply that $F_{X|Z}(x|z)$ is strictly increasing and continuous in x conditional on $\mathcal{X}_z$. Further, condition (i) implies that $F_{X|Z}(x|0)$ and $F_{X|Z}(x|1)$ do not have any intersection on the support of X and $X|Z=0$ stochastically dominates $X|Z=1$. Therefore, Z has a strictly monotonic effect on X. D'Haultfœuille and Février (2015) and Torgovitsky (2015) rule out this case because they assume $F_{X|Z}(x|0)$ and $F_{X|Z}(x|1)$ have intersections on $\mathcal{X}$.

When we have $\mathcal{X}_0 = \mathcal{X}_1 = (\underline{x}, \overline{x})$, then $F_{X|Z}(x|0)$ and $F_{X|Z}(x|1)$ must have intersections at the boundary points of the support of X. However, in this case, g is not identified unless $g(\underline{x}, e)$ (or $g(\overline{x}, e)$) exists and $g(\underline{x}, e)$ (or $g(\overline{x}, e)$) is strictly increasing in e. Torgovitsky (2015) shows that the point identification of g holds when $F_{X|Z}(x|0)$ and $F_{X|Z}(x|1)$ intersect at a boundary point $\underline{x}$, and $g(\underline{x}, e)$ exists and is strictly increasing in e.

Next, we impose restrictions on the conditional distributions of Y conditional on X and Z.

Assumption 2.4. *(i) For all* $(z, x, y) \in \{0, 1\} \times \mathcal{X}_z \times \mathcal{Y}_{x,z}$*,* $F_{Y|X,Z}(y|x, z)$ *is continuous in* x *and* y*. (ii) For all* $(z, x) \in \{0, 1\} \times \mathcal{X}_z$*, we have* $\mathcal{Y}_{x,z} = \mathcal{Y} = (\underline{y}, \overline{y})$*, where* $-\infty \leq \underline{y} < \overline{y} \leq \infty$*.*

D'Haultfœuille and Février (2015) and Torgovitsky (2015) also assume condition (i) but not condition (ii). Both conditions imply that $F_{Y|X,Z}(y|x, z)$ is strictly increasing and continuous in y on $\mathcal{Y}$. Hence, the conditional quantile function of Y conditional on X and Z is the inverse of $F_{Y|X,Z}(y|x, z)$. Condition (ii) is not necessary for this study's results but, without it, deriving the results can become cumbersome. Further, we relax condition (i) in Section 2.6 and allow Y to be discrete or censored.

Finally, we impose the normalization assumption on unobservable variables and support condition of $\epsilon | X = x, Z = z$.

Assumption 2.5. *(i)* $\epsilon \sim U(0, 1)$ *and* $\eta \sim U(0, 1)$*. (ii) For all* $(z, x) \in \{0, 1\} \times \mathcal{X}_z$*, the interior of the support of* $\epsilon | X = x, Z = z$ *is* $(0, 1)$*.*

Condition (i) is the usual normalization in a nonseparable model (see Matzkin (2003)). Torgovitsky (2015) does not use this normalization, while D'Haultfœuille and Février (2015) normalize ϵ to be uniformly distributed. Condition (ii) implies that $g(x, e) \in (\underline{y}, \overline{y}) = \mathcal{Y}$ for all $(x, e) \in \mathcal{X} \times (0, 1)$. Condition (ii) is necessary because, if the support of $\epsilon | X = x, Z = z$ is $[0, \bar{e}]$ for some $0 < \bar{e} < 1$, then the conditional support of Y given $X = x$ and $Z = z$ is equal to $\{g(x, e) : e \in [0, \bar{e}]\}$ and we have $g(x, e) \notin \mathcal{Y}$ for $e > \bar{e}$. This implies that we can not identify $g(x, e)$ for $e > \bar{e}$.

Example 2.1 (Cash Transfer Programs)**.** *Cash transfer programs have been conducted in many countries and many papers estimate their impacts on early childhood development by using randomized experiments. For example, Macours et al. (2012) analyze the impact of a cash transfer program on early childhood cognitive development. In this program, participants were randomly assigned to either the treatment or control groups. As such, we can consider the following model:*

$$
\begin{aligned}
Y &= g(X, \epsilon), \\
X &= \tilde{Z} h_1(\eta) + (1 - \tilde{Z}) h_0(\eta),
\end{aligned}
$$

where Y *is the child's outcome of cognitive development,* X *is the total expenditure, and* $\tilde{Z}$ *is the treatment indicator of the program. Because cash transfers usually increase total expenditure, we can assume* $h_1(\eta) - h_0(\eta) > 0$*. When participants are randomly assigned to either the treatment or control groups,* $Z \equiv 1 - \tilde{Z}$ *is independent of* (ϵ, η) *and hence Assumption 2.1 is satisfied. Because* Z *is independent of* η*, we have* $F_{X|Z}(x|1) = P(h_0(\eta) \leq x)$ *and* $F_{X|Z}(x|0) = P(h_1(\eta) \leq x)$*. Since* $h_1(\eta) > h_0(\eta)$*, we have* $F_{X|Z}(x|0) <$

$F_{X|Z}(x|1)$ for all x. In this case, Assumption 2.3 is satisfied, that is, $F_{X|Z}(x|0)$ and $F_{X|Z}(x|1)$ have no intersection. In Section 2.5, we show this assumption actually holds for the data used by Macours et al. (2012).

2.3 Partial Identification through Monotonicity

Let $\overline{\mathcal{Y}}$ be the closure of $\mathcal{Y}$. We establish the partial identification of g by showing we can identify functions $T^U_{x',x}(y) : \overline{\mathcal{Y}} \to \overline{\mathcal{Y}}$ and $T^L_{x',x}(y) : \overline{\mathcal{Y}} \to \overline{\mathcal{Y}}$ and they are (i) strictly increasing in y, (ii) surjective, that is, $T^U_{x',x}\left([\underline{y}, \overline{y}]\right) = T^L_{x',x}\left([\underline{y}, \overline{y}]\right) = [\underline{y}, \overline{y}]$, and (iii) satisfy the following inequalities:

$$g(x', e) \leq T^U_{x',x}(g(x, e)), \tag{2.2}$$
$$g(x', e) \geq T^L_{x',x}(g(x, e)). \tag{2.3}$$

From (2.2) and (2.3), $T^U_{x',x}(y)$ and $T^L_{x',x}(y)$ are the upper and lower bounds of $g(x', g^{-1}(x, y))$, respectively. If $T^U_{x',x}(y)$ is identified for all $x, x' \in \mathcal{X}$, we can obtain the lower bound of the structural function $g(x, e)$ in the following manner. Here, we define $G^L_x(u) \equiv \int F_{Y|X=x'}\left(T^U_{x',x}(u)\right) dF_X(x')$. If $T^U_{x',x}(y)$ satisfying (2.2) is obtained for all $x, x' \in \mathcal{X}$, then we have

$$\begin{aligned} G^L_x(g(x,e)) &= \int F_{Y|X=x'}\left(T^U_{x',x}(g(x,e))\right) dF_X(x') \\ &\geq \int F_{Y|X=x'}(g(x',e))\, dF_X(x') \\ &= \int P(g(x',\epsilon) \leq g(x',e)|X = x') dF_X(x') \\ &= \int P(\epsilon \leq e|X = x') dF_X(x') \;=\; e, \end{aligned} \tag{2.4}$$

where the first inequality follows from (2.2) and the third equality follows from the strict monotonicity of $g(x, e)$ in e. Furthermore, $G^L_x(u)$ is invertible because $T^U_{x',x}(y)$ is strictly increasing in y. Because $T^U_{x',x}(y)$ is surjective, we have $G^L_x\left([\underline{y}, \overline{y}]\right) = [0, 1]$. Hence, for all $e \in (0, 1)$, we have

$$g(x, e) \geq \left(G^L_x\right)^{-1}(e). \tag{2.5}$$

Similarly, we define $G^U_x(u) \equiv \int F_{Y|X=x'}\left(T^L_{x',x}(u)\right) dF_X(x')$, and thus, we have

$$g(x, e) \leq \left(G^U_x\right)^{-1}(e).$$

Next, we explain how to construct functions $T^U_{x',x}(y)$ and $T^L_{x',x}(y)$ that satisfy (2.2) and (2.3). For any random variable U and random vector W, let $Q_{U|W}(\tau|w)$ denote the conditional τ-th quantile of U conditional on $W = w$, that is, $Q_{U|W}(\tau|w) \equiv \inf\{u :$

$F_{U|W}(u|w) \geq \tau\}$. As in Torgovitsky (2015), we define $\pi(x) : \mathcal{X}_0 \to \mathcal{X}_1$ and $\pi^{-1}(x) : \mathcal{X}_1 \to \mathcal{X}_0$* as:

$$\begin{aligned}\pi(x) \equiv &\ Q_{X|Z}\left(F_{X|Z}(x|0)|1\right), \\ \pi^{-1}(x) \equiv &\ Q_{X|Z}\left(F_{X|Z}(x|1)|0\right).\end{aligned} \tag{2.6}$$

Figure 2.1 illustrates functions $\pi(x)$ and $\pi^{-1}(x)$. The following result is essentially proven by D'Haultfœuille and Février (2015) (Theorem 1). However, we state this result as a proposition because it plays a central role in the following and our assumptions differ somewhat from those of D'Haultfœuille and Février (2015).

Proposition 2.1. *Assume that $\pi(x)$ and $\pi^{-1}(x)$ exist. Define*

$$\begin{aligned}\tilde{T}_x^{(1)}(y) &\equiv Q_{Y|X,Z}\left(F_{Y|X,Z}\left(y|x,0\right)|\pi(x),1\right), \\ \tilde{T}_x^{(-1)}(y) &\equiv Q_{Y|X,Z}\left(F_{Y|X,Z}\left(y|x,1\right)|\pi^{-1}(x),0\right).\end{aligned}$$

Then, under Assumptions 2.1–2.5, we have

$$\begin{aligned}g\left(\pi(x),e\right) &= \tilde{T}_x^{(1)}\left(g(x,e)\right), \\ g\left(\pi^{-1}(x),e\right) &= \tilde{T}_x^{(-1)}\left(g(x,e)\right).\end{aligned}$$

In the first step of the proof, we show that

$$P(\epsilon \leq e|X = x, Z = 0) = P(\epsilon \leq e|X = \pi(x), Z = 1). \tag{2.7}$$

We then define

$$V \equiv F_{X|Z}(X|Z). \tag{2.8}$$

This is called "control variable" in Imbens and Newey (2009). From Assumptions 2.1 and 2.5 (i), we obtain $V = \eta$. Because $F_{X|Z}(x|z)$ is continuous and strictly increasing in x, we obtain

$$F_{\epsilon|X,Z}(e|x,z) = F_{\epsilon|V,Z}(e|F_{X|Z}(x|z), z).$$

By Assumption 1, this implies $F_{\epsilon|X,Z}(e|x,z) = F_{\epsilon|V}(e|F_{X|Z}(x|z))$. Hence, we obtain (2.7) by the definition of $\pi(x)$.

In the second step, we show that (2.7) implies $g\left(\pi(x),e\right) = \tilde{T}_x^{(1)}\left(g(x,e)\right)$. It follows from (2.7) and the strict monotonicity of g that

$$\begin{aligned}F_{Y|X,Z}(g(x,e)|x,0) &= P\left(g(x,\epsilon) \leq g(x,e)|X = x, Z = 0\right) \\ &= P\left(\epsilon \leq e|X = x, Z = 0\right) \\ &= P\left(\epsilon \leq e|X = \pi(x), Z = 1\right) \\ &= F_{Y|X,Z}(g(\pi(x),e)|\pi(x),1).\end{aligned}$$

*These functions correspond to s_{ij} in D'Haultfœuille and Février (2015).

Hence, we obtain $g(\pi(x), e) = \tilde{T}_x^{(1)}(g(x,e))$. Similarly, we also obtain $g(\pi^{-1}(x), e) = \tilde{T}_x^{(-1)}(g(x,e))$.

By definition, if $\pi(x)$ and $\pi^{-1}(x)$ exist, $\tilde{T}_x^{(1)}(y)$ and $\tilde{T}_x^{(-1)}(y)$ are strictly increasing, $\tilde{T}_x^{(1)}([\underline{y}, \overline{y}]) = [\underline{y}, \overline{y}]$, and $\tilde{T}_x^{(-1)}([\underline{y}, \overline{y}]) = [\underline{y}, \overline{y}]$. If $\pi^n(x)$ exists for $n \in \mathbb{N}$, we define $\pi^n(x) \equiv \pi \circ \cdots \circ \pi(x)$. Because the domain of π is $\mathcal{X}_0$, $\pi^n(x)$ does not exist when $\pi^{n-1}(x) \notin \mathcal{X}_0$. If $\pi^n(x)$ exists, we obtain $g(\pi^n(x), e) = \tilde{T}_{\pi^{n-1}(x)}^{(1)} \circ \cdots \circ \tilde{T}_x^{(1)}(g(x,e))$. We define $\tilde{T}_x^{(n)}(y) \equiv \tilde{T}_{\pi^{n-1}(x)}^{(1)} \circ \cdots \circ \tilde{T}_x^{(1)}(y)$ if $\pi^n(x)$ exists. Then, if $\pi^n(x)$ exists, we have

$$g(\pi^n(x), e) = \tilde{T}_x^{(n)}(g(x,e)),$$

$\tilde{T}_x^{(n)}(y)$ is strictly increasing in y, and $\tilde{T}_x^{(n)}([\underline{y}, \overline{y}]) = [\underline{y}, \overline{y}]$. Similarly, we define $\pi^{-n}(x) \equiv \pi^{-1} \circ \cdots \circ \pi^{-1}(x)$ and $\tilde{T}_x^{(-n)}(y) \equiv \tilde{T}_{\pi^{-(n-1)}(x)}^{(-1)} \circ \cdots \circ \tilde{T}_x^{(-1)}(y)$ if $\pi^{-n}(x)$ exists. If $\pi^{-n}(x)$ exists, we have

$$g(\pi^{-n}(x), e) = \tilde{T}_x^{(-n)}(g(x,e)),$$

$\tilde{T}_x^{(-n)}(y)$ is strictly increasing in y, and $\tilde{T}_x^{(-n)}([\underline{y}, \overline{y}]) = [\underline{y}, \overline{y}]$. For all $x \in \mathcal{X}$, we define $\tilde{T}_x^{(0)}(y) \equiv y$ and $\pi^0(x) \equiv x$.

These results imply that, if $\pi^n(x)$ exists, we have $\tilde{T}_x^{(n)}(y) = g(\pi^n(x), g^{-1}(x,y))$, where $g^{-1}(x,y)$ is the inverse function of $g(x,e)$ with respect to e. Hence, we can identify $g(\pi^n(x), g^{-1}(x,y))$ if $\pi^n(x)$ exists. This information restricts the functional form of g. However, as in Remark 1, it does not provide the informative bounds of g without other restrictions.

Here, we examine the properties of $\pi(x)$ and $\pi^{-1}(x)$. Because $F_{X|Z}(x|0) < F_{X|Z}(x|1)$ for $x \in \mathcal{X}$, we have

$$\begin{aligned} \pi(x) = \;& Q_{X|Z}\left(F_{X|Z}(x|0)|1\right) \;<\; Q_{X|Z}\left(F_{X|Z}(x|1)|1\right) \;=\; x, \\ \pi^{-1}(x) = \;& Q_{X|Z}\left(F_{X|Z}(x|1)|0\right) \;>\; Q_{X|Z}\left(F_{X|Z}(x|0)|0\right) \;=\; x. \end{aligned} \tag{2.9}$$

Figure 1 illustrates this intuitively. Because $X|Z=0$ stochastically dominates $X|Z=1$ and functions $\pi(x)$ and $\pi^{-1}(x)$ satisfy (2.28), the inequalities hold.

To facilitate the illustration of our identification results, we first review the identification approach of D'Haultfœuille and Février (2015) and Torgovitsky (2015) when $\underline{x}_0 = \underline{x}_1 = \xi$, although Assumption 2.3 rules out the case of $\underline{x}_0 = \underline{x}_1 = \xi$. Additionally, we assume that $g(\xi, e)$ exists and is strictly increasing in e.

D'Haultfœuille and Février (2015) and Torgovitsky (2015) use function $T_{x',x}(y)$ that satisfies $g(x', e) = T_{x',x}(g(x,e))$. This function corresponds to $Q_{x'x}$ in D'Haultfœuille and Février (2015). We define

$$G_x(u) \equiv \int F_{Y|X=x'}\left(T_{x',x}(u)\right) dF_X(x').$$

Then, similar to (2.4), we have $G_x(g(x,e)) = e$, and hence $g(x,e) = (G_x)^{-1}(e)$. If we can identify a function $T_{x',x}(y)$ for all x and x', we then can point identify the structural function g.

Pick an initial point $x_0 \in \mathcal{X}$ (i.e., $x_0 > \xi$) and form a recursive sequence $x_{n+1} = \pi(x_n)$ for $n > 0$. Because $\underline{x}_0 = \underline{x}_1 = \xi$ implies $\mathcal{X}_1 \subset \mathcal{X}_0$, we have $\pi(x) \in \mathcal{X}_0$ for all $x \in \mathcal{X}$ and there exists a sequence $\{\pi^n(x)\}_{n=1}^{\infty}$. The sequence $\{x_n\}$ is decreasing by (2.9) and $x_n > \xi$ for all $n \geq 0$ by the definition of $\pi(x)$. Hence, sequence $\{x_n\}$ converges to a limiting point. Because (2.28) implies

$$F_{X|Z}(x_{n+1}|1) = F_{X|Z}(x_n|0)$$

and $F_{X|Z}(x|z)$ is continuous in x, we have $F_{X|Z}(\lim_{n\to\infty} x_n|1) = F_{X|Z}(\lim_{n\to\infty} x_n|0)$. Because $F_{X|Z}(x|0) < F_{X|Z}(x|1)$ for all $x \in (\xi, \overline{x}_0)$ and $F_{X|Z}(\xi|0) = F_{X|Z}(\xi|1) = 0$, the sequence $\{x_n\}$ converges to ξ for any initial point $x_0 \in \mathcal{X}$. Figure 2.2 illustrates this intuitively. We define $\tilde{T}_x^{(\infty)}(y) \equiv \lim_{n\to\infty} \tilde{T}_x^{(n)}(y)$, which is strictly increasing and invertible in y. From the continuity of g, we obtain, for all $x \in \mathcal{X}$,

$$\tilde{T}_x^{(\infty)}(g(x,e)) = \lim_{n\to\infty} g(\pi^n(x), e) = g(\xi, e).$$

Because $\tilde{T}_x^{(\infty)}(g(x,e)) = \tilde{T}_{x'}^{(\infty)}(g(x',e))$ holds for any x, x', we have

$$g(x', e) = \left(\tilde{T}_{x'}^{(\infty)}\right)^{-1}\left(\tilde{T}_x^{(\infty)}(g(x,e))\right).$$

We define $T_{x',x}(y) \equiv \left(\tilde{T}_{x'}^{(\infty)}\right)^{-1}\left(\tilde{T}_x^{(\infty)}(y)\right)$. Then, $T_{x',x}(y)$ is strictly increasing and satisfies $g(x',e) = T_{x',x}(g(x,e))$. This implies that $g(x', g^{-1}(x,y))$ is identified for all x and x'. Hence, as previously discussed, g is point identified.

This approach is not available under Assumption 2.3 because a convergent sequence $\{\pi^n(x)\}_{n=1}^{\infty}$ does not exist. When $F_{X|Z}(x|0)$ and $F_{X|Z}(x|1)$ have no intersections, $\pi^n(x)$ lies in $\mathcal{X}_1 \cap \mathcal{X}_0^c = (\underline{x}_1, \underline{x}_0]$ when n is sufficiently large. If $\pi^n(x)$ is in $\mathcal{X}_1 \cap \mathcal{X}_0^c$, then $\pi^{n+1}(x)$ does not exist. From the proof of Lemma 2.1, for all $x \in \mathcal{X}$, $\{n : \pi^n(x) \text{ exists.}\}$ is a finite set under Assumption 2.3. For example, in Figure 2.1, $\pi(x)$, $\pi^{-1}(x)$, and $\pi^{-2}(x)$ exist but $\pi^2(x)$ and $\pi^{-3}(x)$ do not.

Remark 2.1. *If we do not impose additional restrictions, the identified set of $g(x,e)$ can become unbounded under Assumption 2.3. To show this, we derive the identified set of g. We define*

$$\mathcal{G} \equiv \{\tilde{g} : \mathcal{X} \times (0,1) \to \mathbb{R} : \tilde{g}(x,e) \text{ is continuous and strictly increasing in } e.\}.$$

Torgovitsky (2015) derives the identified set of g under another normalization assumption. Similarly, we obtain the following identified set:

$$\mathcal{G}_I \equiv \left\{\tilde{g} \in \mathcal{G} : (\tilde{g}^{-1}(X,Y), V) \perp\!\!\!\perp Z \text{ and } \tilde{g}^{-1}(X,Y) \sim U(0,1)\right\},$$

where $\tilde{g}^{-1}$ is the inverse of $\tilde{g}$ with respect to its last component and V is defined as in (2.8). The independence condition in the identified set is equivalent to the following condition:

$$P\left(Y \leq \tilde{g}(X,e)|V=v, Z=0\right) = P\left(Y \leq \tilde{g}(X,e)|V=v, Z=1\right) \quad \text{for all } v \in (0,1).$$

From the definition of V, for all $v \in (0,1)$, we have

$$F_{Y|X,Z}\left(\tilde{g}(x_{v,0},e)|x_{v,0},0\right) \;=\; F_{Y|X,Z}\left(\tilde{g}(x_{v,1},e)|x_{v,1},1\right),$$

where $x_{v,z} \equiv Q_{X|Z}(v|z)$. Hence, we can rewrite $\mathcal{G}_I$ as

$$\begin{aligned}\mathcal{G}_I \;=\; & \left\{\tilde{g} \in \mathcal{G} : \tilde{g}^{-1}(X,Y) \sim U(0,1) \quad \text{and} \right. \\ & \left. \tilde{g}\left(x_{v,1}, \tilde{g}^{-1}(x_{v,0},\cdot)\right) = Q_{Y|X,Z}\left(F_{Y|X,Z}(\cdot|x_{v,0},0)|x_{v,1},1\right) \text{ for all } v. \right\}. \qquad (2.10)\end{aligned}$$

This expression implies that $g\left(x_{v,1}, g^{-1}(x_{v,0},y)\right)$ is identified for all v. Proposition 2.1 provides the same result. The sharp lower and upper bounds of $g(x,e)$ are obtained by $\inf_{\tilde{g}\in\mathcal{G}_I}\tilde{g}(x,e)$ and $\sup_{\tilde{g}\in\mathcal{G}_I}\tilde{g}(x,e)$.

To show that the bounds of $g(x,e)$ can be unbounded, we consider the following simple model:

$$\begin{aligned} Y &= \Phi^{-1}(\epsilon), \\ X &= Z(\eta-1) + (1-Z)\eta, \end{aligned}$$

where $\Phi(\cdot)$ is the standard normal distribution function, $\epsilon \sim U(0,1)$, $\eta \sim U(0,1)$, Z is a random Bernoulli variable with $p = 0.5$, and (ϵ,η,Z) are mutually independent. Then, it follows from (2.10) that $\tilde{g} \in \mathcal{G}_I$ if and only if

$$\tilde{g}(v,e) = \tilde{g}(v-1,e) \quad \text{for all } e, v \in (0,1), \qquad (2.11)$$

$$P(Y \leq \tilde{g}(X,e)) = e \quad \text{for all } e \in (0,1). \qquad (2.12)$$

We construct $\tilde{g}_M$ as follows. First, we define

$$\tilde{g}_M(x,0.5) \equiv \begin{cases} \Phi^{-1}\left(4M(x+1)+0.5-M\right), & -1 < x \leq -0.5 \\ \Phi^{-1}\left(-4M(x+0.5)+0.5+M\right), & -0.5 < x \leq 0 \\ \Phi^{-1}\left(4Mx+0.5-M\right), & 0 < x \leq 0.5 \\ \Phi^{-1}\left(-4M(x-0.5)+0.5+M\right), & 0.5 < x \leq 1 \end{cases},$$

where $-0.5 < M < 0.5$. Second, for $e \neq 0.5$, we define $\tilde{g}_M(x,e)$ as

$$\tilde{g}_M(x,e) \equiv \begin{cases} \Phi^{-1}\left(2e\Phi(\tilde{g}(x,0.5))\right), & 0 < e < 0.5 \\ \Phi^{-1}\left(1-2(1-e)\{1-\tilde{g}(x,0.5)\}\right), & 0.5 < e < 1 \end{cases}.$$

Then, we confirm that $\tilde{g}_M$ satisfies (2.11) and (2.12) for all $-0.5 < M < 0.5$. Hence, $\tilde{g}_M$ is an element of $\mathcal{G}_I$ for all $-0.5 < M < 0.5$. Because $\tilde{g}_M(0, 0.5) = \Phi^{-1}(0.5 - M)$, the lower and upper bounds of $g(0, 0.5)$ are $-\infty$ and $+\infty$, respectively. Therefore, in this setting, the identified set of g can be unbounded.

If we do not impose additional restrictions, we cannot construct strictly increasing functions $T^U_{x',x}(y)$ and $T^L_{x',x}(y)$ that satisfy (2.2) and (2.3). First, we show that a set $\Pi^M_{x',x}$ defined below is nonempty and finite, when $F_{X|Z}(x|0)$ and $F_{X|Z}(x|1)$ have no intersections. Second, we show that we can partially identify $g(x,e)$ using $\Pi^M_{x',x}$ when $g(x,e)$ is nondecreasing in x.

For $(x, x') \in \mathcal{X} \times \mathcal{X}$, we define $\Pi^M_{x',x}$ as

$$\Pi^M_{x',x} \equiv \{(n,m) : n, m \in \mathbb{Z},\ \pi^n(x') \text{ and } \pi^m(x) \text{ exist, and } \pi^n(x') \leq \pi^m(x).\}. \quad (2.13)$$

In Figure 2.1, $\Pi^M_{x',x} = \{(-1,-2), (0,-2), (0,-1), (1,-2), (1,-1), (1,0)\}$. The following lemma shows that $\Pi^M_{x',x}$ is nonempty and finite when $F_{X|Z}(x|0)$ and $F_{X|Z}(x|1)$ have no intersections.

Lemma 2.1. *Under Assumptions 2.1–2.5, $\Pi^M_{x',x}$, as defined by (2.13), is nonempty and finite for all $(x, x') \in \mathcal{X} \times \mathcal{X}$.*

Under Assumptions 2.1–2.5, for any $x \in \mathcal{X}$ the set $\{n \in \mathbb{Z} : \pi^n(x) \text{ exists.}\}$ is finite from the proof of Lemma 2.1. Hence, g cannot be point identified using the method proposed by D'Haultfœuille and Février (2015) and Torgovitsky (2015)).

We impose the following assumption:

Assumption 2.6 (Monotonicity)**.** *For all $e \in (0,1)$, $g(x,e)$ is nondecreasing in x.*

The monotonicity assumption holds for many economic models. For example, the demand function is ordinarily decreasing in price if the income effect is negligible, and economic analyses of production often assume that the production function is monotonically increasing in input. Monotonicity assumptions of this type have been employed in many studies. For example, Manski (1997) imposes a monotonicity assumption on a response function and shows that the average treatment response is partially identified.

If $(n,m) \in \Pi^M_{x',x}$, Assumption 2.6 implies that

$$\tilde{T}^{(n)}_{x'}\left(g(x',e)\right) = g(\pi^n(x'), e) \leq g(\pi^m(x), e) = \tilde{T}^{(m)}_{x}\left(g(x,e)\right).$$

Because $\tilde{T}^{(n)}_{x'}(y)$ is strictly increasing in y and $\tilde{T}^{(n)}_{x'}\left([\underline{y}, \overline{y}]\right) = [\underline{y}, \overline{y}]$, we have $g(x',e) \leq \left(\tilde{T}^{(n)}_{x'}\right)^{-1}\left(\tilde{T}^{(m)}_{x}\left(g(x,e)\right)\right)$ for $(n,m) \in \Pi^M_{x',x}$. Hence, we have

$$g(x',e) \leq \min_{(n,m)\in \Pi^M_{x',x}} \left(\tilde{T}^{(n)}_{x'}\right)^{-1}\left(\tilde{T}^{(m)}_{x}\left(g(x,e)\right)\right).$$

Define

$$T^{MU}_{x',x}(y) \equiv \min_{(n,m)\in\Pi^M_{x',x}} \left(\tilde{T}^{(n)}_{x'}\right)^{-1}\left(\tilde{T}^{(m)}_{x}(y)\right),$$
$$T^{ML}_{x',x}(y) \equiv \max_{(n,m)\in\Pi^M_{x,x'}} \left(\tilde{T}^{(m)}_{x'}\right)^{-1}\left(\tilde{T}^{(n)}_{x}(y)\right). \tag{2.14}$$

Then, $T^{MU}_{x',x}(y)$ is strictly increasing and satisfies

$$g(x',e) \leq T^{MU}_{x',x}\left(g(x,e)\right). \tag{2.15}$$

Similarly, $T^{ML}_{x',x}(y)$ is strictly increasing and satisfies

$$g(x',e) \geq T^{ML}_{x',x}\left(g(x,e)\right). \tag{2.16}$$

As already mentioned, the functions that satisfy (2.2) and (2.3) are the upper and lower bounds of $g(x', g^{-1}(x,y))$, respectively. Hence, for any $(n,m) \in \Pi^M_{x',x}$, $\left(\tilde{T}^{(n)}_{x'}\right)^{-1}\left(\tilde{T}^{(m)}_{x}(y)\right)$ becomes the upper bound of $g(x', g^{-1}(x,y))$. This implies that $T^{MU}_{x',x}(y)$ is the lowest upper bound of $g(x', g^{-1}(x,y))$ in the sense that $T^{MU}_{x',x}(y)$ is lower than $\left(\tilde{T}^{(n)}_{x'}\right)^{-1}\left(\tilde{T}^{(m)}_{x}(y)\right)$ for any $(n,m) \in \Pi^M_{x',x}$. Similarly, $T^{ML}_{x',x}(y)$ is the largest lower bound of $g(x', g^{-1}(x,y))$.

We define

$$\begin{aligned}
G^{ML}_x(u) &\equiv \int F_{Y|X=x'}\left(T^{MU}_{x',x}(u)\right) dF_X(x'),\\
G^{MU}_x(u) &\equiv \int F_{Y|X=x'}\left(T^{ML}_{x',x}(u)\right) dF_X(x'),\\
B^{ML}(x,e) &\equiv \sup_{y: y\leq x}\left\{\left(G^{ML}_y\right)^{-1}(e)\right\},\\
B^{MU}(x,e) &\equiv \inf_{y: y\geq x}\left\{\left(G^{MU}_y\right)^{-1}(e)\right\}.
\end{aligned}$$

$G^{ML}_x(u)$ and $G^{MU}_x(u)$ provide the lower and upper bounds of $g(x,e)$ on the basis of arguments (2.4) and (2.5). $B^{ML}(x,e)$ and $B^{MU}(x,e)$ strengthen these bounds.

Theorem 2.1. *Under Assumptions 2.1–2.6, for all* $(x,e) \in \mathcal{X} \times (0,1)$*, we have*

$$B^{ML}(x,e) \;\leq\; g(x,e) \;\leq\; B^{MU}(x,e).$$

In the first step, we show that $\left(G^{ML}_x\right)^{-1}(e) \leq g(x,e) \leq \left(G^{MU}_x\right)^{-1}(e)$. In the second step, we strengthen these bounds to $B^{ML}(x,e) \leq g(x,e) \leq B^{MU}(x,e)$. Figure 2.3 intuitively illustrates this proof. The idea is similar to that of Manski (1997), who considers the case in which response function $y(t)$ is increasing, where $y(t)$ is a latent outcome with treatment t. He then uses the monotonicity of $y(t)$ to partially identify average response function $E[y(t)]$ when the support of the outcome is bounded. By contrast, our bounds are bounded even when the support of the outcome is unbounded.

Simulation 2.1. *To illustrate Theorem 2.1, we consider the following example:*

$$
\begin{aligned}
&Y = h(X)exp\left(\alpha + \beta\Phi^{-1}(\epsilon)\right)\\
&X = (0.2+\eta)Z + (1-Z)\{(2-\rho)(\eta-1)+2.2\},
\end{aligned}
\tag{2.17}
$$

where $h(x)$ is an increasing function specified below, $\Phi(\cdot)$ is the standard normal distribution function, Z is a random Bernoulli variable with $p = 0.5$, and $(\alpha, \beta) = (0.5, 0.5)$. Suppose that

$$
\begin{aligned}
\epsilon &= \Phi(U)\\
\eta &= \Phi(V)\\
(U, V) &\sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix}\right).
\end{aligned}
$$

Then, $\epsilon \sim U(0,1)$ and $\eta \sim U(0,1)$. In this example, $F_{X|Z}(x|1) = x - 0.2$ for $x \in [0.2, 1]$ and $F_{X|Z}(x|0) = \frac{1}{2-\rho}(x - 2.2) + 1$ for $x \in [\rho + 0.2, 2.2]$. These functions are depicted in Figure 4. Conditional distribution functions $F_{X|Z}(x|0)$ and $F_{X|Z}(x|1)$ do not intersect when $\rho > 0$. When $\rho = 0$, these functions intersect at $x = 0.2$. Torgovitsky (2015) shows that g is point identified when $\rho = 0$.

We calculate the bounds of $g(x, 0.5)$ using Theorem 2.1 when $h(x) = h_1(x) \equiv x$ or $h(x) = h_2(x) \equiv 2\exp(4(x-1.2))/\{1+\exp(4(x-1.2))\} + 0.2$. Figures 2.5 and 2.6 show these bounds for three different choices of ρ: 0.01, 0.1, and 0.3. For h_1 and h_2, the bounds become tighter as ρ become smaller. In particular, the bounds are very close to the true function when $\rho = 0.01$. This implies that $B^{ML}(x, e)$ and $B^{MU}(x, e)$ converge to $g(x, e)$ as $\rho \to 0$. When $\rho = 0.01$ and 0.1, the bounds of h_2 are tighter than that of h_1. This result is caused by $h_2(x)$ being flatter than $h_1(x)$ over a particular interval. As discussed later, Theorem 2.2 shows that g is point identified when $g(x, e)$ is flat with respect to x over a given interval.

The bounds become tighter as the difference between $g(x', e)$ and $T^U_{x',x}(g(x, e))$ (or $T^L_{x',x}(g(x, e))$) decreases. The following theorem shows that, if $g(x, e)$ is flat in x over a given interval, inequalities (2.2) and (2.3) become equalities and structural function g is point identified.

Theorem 2.2. *Under Assumptions 2.1–2.6, if there exists $\tilde{x} \in \mathcal{X}_0 \cap \mathcal{X}_1$ such that $x \mapsto g(x, e)$ is constant on $[\pi(\tilde{x}), \pi^{-1}(\tilde{x})]$ for each $e \in (0, 1)$, then $B^{ML}(x, e)$ and $B^{MU}(x, e)$ coincide with $g(x, e)$ for all $(x, e) \in \mathcal{X} \times (0, 1)$. Hence, g is point identified. This result holds even when the interval $[\pi(\tilde{x}), \pi^{-1}(\tilde{x})]$ is unknown.*

In the first step, we show that, for all $x \in \mathcal{X}$, $n \in \mathbb{Z}$ exists such that $\pi^n(x), \pi^{n+1}(x) \in [\pi(\tilde{x}), \pi^{-1}(\tilde{x})]$. In the second step, we show g is point identified. Because $g(x, e)$ is constant in x conditional on $[\pi(\tilde{x}), \pi^{-1}(\tilde{x})]$, we have $g(x', e) = T^{MU}_{x',x}(g(x,e))$ and $g(x', e) = T^{ML}_{x',x}(g(x,e))$ for all $x, x' \in \mathcal{X}$ and $e \in (0,1)$. Hence, $B^{ML}(x,e)$ and $B^{MU}(x,e)$ coincide with $g(x,e)$ because inequalities (2.15) and (2.16) become equalities.

Simulation 2.2. *To illustrate Theorem 2.2, we consider model (2.17). We set* $h(x) = \max\{0, x - \delta\} + 0.5$ *and* $\rho = 0.3$. *Figures 2.7–2.9 show* $B^{ML}(x, 0.5)$ *and* $B^{MU}(x, 0.5)$ *for three different choices of* δ*:* 0.4, 0.55, *and* 1.2. *In this model,* $g(x,e)$ *is constant on* $[0.2, \delta]$. *Because* $\pi(0.5) = 0.2$ *and* $\pi^{-1}(0.5) = 1.01$, *interval* $[0.2, \delta]$ *covers* $[\pi(0.5), 0.5]$ *when* $\delta = 0.55$ *and covers* $[\pi(0.5), \pi^{-1}(0.5)]$ *when* $\delta = 1.2$. *Hence, the condition of Theorem 2.2 is satisfied only when* $\delta = 1.2$. *In Figure 2.9,* $B^{ML}(x, 0.5)$ *and* $B^{MU}(x, 0.5)$ *coincide with* $g(x, 0.5)$ *when* $\delta = 1.2$. *By contrast, when* $\delta = 0.4$ *and* 0.55, $g(x, 0.5)$ *is not point identified.*

Remark 2.2. *The bounds of Theorem 2.1 are sharp in the sense that there can exist data generating processes that satisfy the conditions of the theorem such that the bounds are attained. As shown in Theorem 2.2, if* $g(x,e)$ *is flat with respect to* x *over a given interval, we have* $g(x,e) = B^{ML}(x,e) = B^{MU}(x,e)$. *This implies that* g *is point identified. Therefore, in this case,* $B^{ML}(x,e)$ *and* $B^{MU}(x,e)$ *are sharp bounds of* $g(x,e)$.

Remark 2.3. *Although our bounds may not be sharp in general, we can derive the identified set of* g *under Assumption 2.6. We define*

$$\mathcal{G}^M \equiv \{\tilde{g} \in \mathcal{G} : \tilde{g}(x,e) \text{ is nondecreasing in } x.\}.$$

Then, similar to (2.10), the identified set of g *under Assumption 2.6 is obtained by*

$$\begin{aligned} \mathcal{G}^M_I \; = \; & \big\{\tilde{g} \in \mathcal{G}^M : \tilde{g}^{-1}(X,Y) \sim U(0,1) \;\; \text{and} \\ & \tilde{g}\left(x_{v,1}, \tilde{g}^{-1}(x_{v,0}, \cdot)\right) = Q_{Y|X,Z}\left(F_{Y|X,Z}(\cdot|x_{v,0}, 0)|x_{v,1}, 1\right) \text{ for all } v.\big\}. \end{aligned}$$

Hence, the sharp lower and upper bounds of $g(x,e)$ *are* $\inf_{\tilde{g}\in\mathcal{G}^M_I} \tilde{g}(x,e)$ *and* $\sup_{\tilde{g}\in\mathcal{G}^M_I} \tilde{g}(x,e)$, *respectively. However, these bounds may not coincide with* $B^{ML}(x,e)$ *and* $B^{MU}(x,e)$. *Actually, in some settings,* $\left(G^{ML}_x\right)^{-1}(e)$ *and* $\left(G^{MU}_x\right)^{-1}(e)$ *are not nondecreasing in* x. *This implies that* $\left(G^{ML}_x\right)^{-1}(e)$ *and* $\left(G^{MU}_x\right)^{-1}(e)$ *are not sharp in general.*

It is difficult to compute $\mathcal{G}^M_I$ *because* $\mathcal{G}^M$ *is infinite dimensional. By contrast,* $B^{ML}(x,e)$ *and* $B^{MU}(x,e)$ *have closed-form expressions and are hence computable. In Simulations 2.1 and 2.2, we compute* $B^{ML}(x,e)$ *and* $B^{MU}(x,e)$ *in some settings, and in Section 2.5, we show that we can obtain informative bounds in real data.*

2.4 Partial Identification through Concavity

In this section, we propose a method to construct the lower and upper bounds of $g(x,e)$ when $g(x,e)$ is concave in x.

First, we show that a set $\Pi^C_{x',x}$ defined below is nonempty and finite. Second, we show that we can partially identify g using $\Pi^C_{x',x}$ when $g(x,e)$ is concave in x.

For $(x,x') \in \mathcal{X} \times \mathcal{X}$, we define $\Pi^C_{x',x}$ as

$$\begin{aligned}\Pi^C_{x',x} \equiv\ & \{(n,m) : n,m \in \mathbb{Z},\ \pi^n(x'),\ \pi^{n-1}(x) \text{ and } \pi^m(x) \text{ exist}, \\ & \text{and } \pi^n(x') \le \pi^m(x) \le \pi^{n-1}(x').\}\,. \end{aligned} \tag{2.18}$$

In Figure 2.1, $\Pi^C_{x',x} = \{(0,-1),(1,0)\}$. The following lemma shows that $\Pi^C_{x',x}$ is nonempty and finite, similar to Lemma 2.1.

Lemma 2.2. *Under Assumptions 2.1–2.5, $\Pi^C_{x',x}$ as defined by (2.18) is nonempty and finite for all $(x,x') \in \mathcal{X} \times \mathcal{X}$.*

Similar to Section 2.3, we impose the following assumption.

Assumption 2.7 (Concavity)**.** *For all $e \in (0,1)$, $g(x,e)$ is concave in x.*

The concavity assumption holds in many economic models. For example, economic analyses of production often assume that the production function is concave in inputs. For instance, Manski (1997) assumes concavity and shows that the average treatment response is partially identified. Further, D'Haultfoeuille et al. (2013) achieves the partial identification of the average treatment on the treated effect using a locally concavity assumption.

As in Section 2.3, if we identify functions $T^U_{x',x}(y)$ and $T^L_{x',x}(y)$ that are strictly increasing in y, surjective, and satisfy (2.2) and (2.3), we can obtain the lower and upper bounds of $g(x,e)$. Hence, we consider constructing functions $T^U_{x',x}(y)$ and $T^L_{x',x}(y)$ that are strictly increasing in y, surjective, and satisfy (2.2) and (2.3).

If $(n,m) \in \Pi^C_{x',x}$, from Assumption 2.7, we have

$$\left[t_{x',x}(n,m)\tilde{T}^{(n)}_{x'} + (1 - t_{x',x}(n,m))\tilde{T}^{(n-1)}_{x'}\right](g(x',e)) \le \tilde{T}^{(m)}_{x}\left(g(x,e)\right),$$

where $\left[t_{x',x}(n,m)\tilde{T}^{(n)}_{x'} + (1 - t_{x',x}(n,m))\tilde{T}^{(n-1)}_{x'}\right](y) = t_{x',x}(n,m)\tilde{T}^{(n)}_{x'}(y) + (1 - t_{x',x}(n,m))$ $\tilde{T}^{(n-1)}_{x'}(y)$ and $t_{x',x}(n,m) = \left(\pi^{n-1}(x') - \pi^m(x)\right) / \left(\pi^{n-1}(x') - \pi^n(x')\right)$. Because $\tilde{T}^{(n)}_{x'}(y)$ and $\tilde{T}^{(n-1)}_{x'}(y)$ are strictly increasing in y, $\tilde{T}^{(n)}_{x'}\left([\underline{y},\overline{y}]\right) = [\underline{y},\overline{y}]$, and $\tilde{T}^{(n-1)}_{x'}\left([\underline{y},\overline{y}]\right) = [\underline{y},\overline{y}]$, we have

$$g(x',e) \le \min_{(n,m)\in\Pi^C_{x',x}} \left[t_{x',x}(n,m)\tilde{T}^{(n)}_{x'} + (1 - t_{x',x}(n,m))\tilde{T}^{(n-1)}_{x'}\right]^{-1}\left(\tilde{T}^{(m)}_{x}\left(g(x,e)\right)\right).$$

We define

$$\begin{aligned} T^{CU}_{x',x}(y) &\equiv \min_{(n,m)\in\Pi^C_{x',x}} \left[t_{x',x}(n,m)\tilde{T}^{(n)}_{x'} + (1 - t_{x',x}(n,m))\tilde{T}^{(n-1)}_{x'}\right]^{-1}\left(\tilde{T}^{(m)}_{x}(y)\right), \\ T^{CL}_{x',x}(y) &\equiv \max_{(n,m)\in\Pi^C_{x,x'}} \left(\tilde{T}^{(m)}_{x'}\right)^{-1}\left(\left[t_{x,x'}(n,m)\tilde{T}^{(n)}_{x} + (1 - t_{x,x'}(n,m))\tilde{T}^{(n-1)}_{x}\right](y)\right). \end{aligned} \tag{2.19}$$

Then, $T^{CU}_{x',x}(y)$, as defined in (2.19), is strictly increasing and satisfies

$$g(x', e) \leq T^{CU}_{x',x}\left(g(x, e)\right). \tag{2.20}$$

Similarly, $T^{CL}_{x',x}(y)$, as defined in (2.19), is strictly increasing and satisfies

$$g(x', e) \geq T^{CL}_{x',x}\left(g(x, e)\right). \tag{2.21}$$

We define

$$\begin{aligned} G^{CL}_x(u) &\equiv \int F_{Y|X=x'}\left(T^{CU}_{x',x}(u)\right) dF_X(x'), \\ G^{CU}_x(u) &\equiv \int F_{Y|X=x'}\left(T^{CL}_{x',x}(u)\right) dF_X(x'), \end{aligned}$$

$$B^{CL}(x,e) \equiv \sup_{y,y':y<x<y'} \left\{ \left(\frac{x-y}{y'-y}\right) \left(G^{CL}_{y'}\right)^{-1}(e) + \left(\frac{y'-x}{y'-y}\right) \left(G^{CL}_{y}\right)^{-1}(e) \right\},$$

$$\begin{aligned} B^{CU}(x,e) \equiv \; & \min \Big[\inf_{y,y':x<y<y'} \left\{ \left(\tfrac{x-y}{y'-y}\right) B^{CL}(y',e) + \left(\tfrac{y'-x}{y'-y}\right) \left(G^{CU}_y\right)^{-1}(e) \right\}, \\ & \inf_{y,y':y'<y<x} \left\{ \left(\tfrac{y-x}{y-y'}\right) B^{CL}(y',e) + \left(\tfrac{x-y'}{y-y'}\right) \left(G^{CU}_y\right)^{-1}(e) \right\} \Big]. \end{aligned}$$

$G^{CL}_x(u)$ and $G^{CU}_x(u)$ provide the lower and upper bounds of $g(x, e)$ as per (2.4) and (2.5). $B^{CL}(x, e)$ and $B^{CU}(x, e)$ strengthen these bounds.

Theorem 2.3. *Under Assumptions 2.1–2.5 and 2.7, for all* $(x, e) \in \mathcal{X} \times (0, 1)$*, we have*

$$B^{CL}(x, e) \leq g(x, e) \leq B^{CU}(x, e).$$

Similar to Theorem 2.1, we can show that $\left(G^{CL}_x\right)^{-1}(e) \leq g(x, e) \leq \left(G^{CU}_x\right)^{-1}(e)$. We strengthen the bounds to $B^{CL}(x, e) \leq g(x, e) \leq B^{CU}(x, e)$ using the concavity of $g(x, e)$ in x. Figure 2.10 intuitively illustrates this proof. A similar approach is used by Manski (1997), namely utilizing the concavity of the response function to partially identify the average response function when the support of the outcome is bounded. However, our approach does not require information on the infimum and supremum of the support of the outcome.

This identification approach is somewhat similar to that of D'Haultfoeuille et al. (2013), who study the identification of nonseparable models with continuous, endogenous regressors, using repeated cross sections. Specifically, they consider the following model:

$$Y_t = g_t(X_t, A_t), \quad t = 1, \cdots, T,$$

where A_t is an unobserved heterogeneous factor. They show that, under the assumptions that $A_t|V_t \equiv F_{X_t}(X_t) = v \sim A_s|V_s \equiv F_{X_s}(X_s) = v$ and $g_t(x, a) = m_t(g(x, a))$, the average treatment on treated effect $\Delta^{ATT}(x, x') \equiv E[g_T(x, A_T) - g_T(x', A_T)|X_T = x]$ is identified when $F_{X_T}(x) = F_{X_t}(x')$. Under this assumption, $\Delta^{ATT}(x, x')$ is not identified if $F_{X_T}(x) \neq F_{X_t}(x')$ for all $t \in \{1, \cdots, T-1\}$. However, they show that $\Delta^{ATT}(x, x')$ is partially identified if $x \mapsto g(x, a)$ is locally concave.

In several cases, such as the production function, we can assume that both Assumptions 2.6 and 2.7 hold. Then, it follows from Theorems 2.1 and 2.3 that

$$\max\{B^{ML}(x, e), B^{CL}(x, e)\} \leq g(x, e) \leq \min\{B^{MU}(x, e), B^{CU}(x, e)\}. \tag{2.22}$$

In this case, we can obtain tighter bounds in the following manner. We define

$$\begin{aligned}
T^{MCU}_{x',x}(y) &\equiv \min\{T^{MU}_{x',x}(y), T^{CU}_{x',x}(y)\}, \\
T^{MCL}_{x',x}(y) &\equiv \max\{T^{ML}_{x',x}(y), T^{CL}_{x',x}(y)\}, \\
G^{MCL}_{x}(u) &\equiv \int F_{Y|X=x'}\left(T^{MCU}_{x',x}(u)\right) dF_X(x'), \\
G^{MCU}_{x}(u) &\equiv \int F_{Y|X=x'}\left(T^{MCL}_{x',x}(u)\right) dF_X(x').
\end{aligned}$$

Similarly to the above arguments, we have $g(x', e) \leq T^{MCU}_{x',x}(g(x, e))$ and $g(x', e) \geq T^{MCL}_{x',x}(g(x, e))$, and hence we can obtain

$$\left(G^{MCL}_{x}\right)^{-1}(e) \leq g(x, e) \leq \left(G^{MCU}_{x}\right)^{-1}(e).$$

Define

$$\begin{aligned}
\tilde{B}^{MCL}(x, e) &\equiv \sup_{y: y \leq x} \left\{\left(G^{MCL}_{y}\right)^{-1}(e)\right\}, \\
\tilde{B}^{MCU}(x, e) &\equiv \inf_{y: y \geq x} \left\{\left(G^{MCU}_{y}\right)^{-1}(e)\right\},
\end{aligned}$$

$$\hat{B}^{MCL}(x, e) \equiv \sup_{y, y': y < x < y'} \left\{\left(\frac{x - y}{y' - y}\right)\left(G^{MCL}_{y'}\right)^{-1}(e) + \left(\frac{y' - x}{y' - y}\right)\left(G^{MCL}_{y}\right)^{-1}(e)\right\},$$

$$\begin{aligned}
\hat{B}^{MCU}(x, e) \equiv \ \min\Big[&\inf_{y, y': x < y < y'} \left\{\left(\tfrac{x-y}{y'-y}\right)\hat{B}^{MCL}(y', e) + \left(\tfrac{y'-x}{y'-y}\right)\left(G^{MCU}_{y}\right)^{-1}(e)\right\}, \\
&\inf_{y, y': y' < y < x} \left\{\left(\tfrac{y-x}{y-y'}\right)\hat{B}^{MCL}(y', e) + \left(\tfrac{x-y'}{y-y'}\right)\left(G^{MCU}_{y}\right)^{-1}(e)\right\}\Big].
\end{aligned}$$

Then, from the above results, both $\tilde{B}^{MCU}(x, e)$ and $\hat{B}^{MCU}(x, e)$ are upper bounds of $g(x, e)$. Similarly, both $\tilde{B}^{MCL}(x, e)$ and $\hat{B}^{MCL}(x, e)$ are also lower bounds of $g(x, e)$. Therefore, we can obtain

$$\max\{\tilde{B}^{MCL}(x, e), \hat{B}^{MCL}(x, e)\} \leq g(x, e) \leq \min\{\tilde{B}^{MCL}(x, e), \hat{B}^{MCL}(x, e)\}. \tag{2.23}$$

Clearly, these bounds are tighter than (2.22).

Similar to Theorem 2.2, the following theorem shows that, if $g(x, e)$ is linear in x over a particular interval, inequalities (2.20) and (2.21) become equalities, and $B^{CL}(x, e)$ and $B^{CU}(x, e)$ coincide with $g(x, e)$.

Theorem 2.4. *Under Assumptions 2.1–2.5 and 2.7, if* $\tilde{x} \in \mathcal{X}$ *exists such that* $g(x, e)$ *is linear in* x *on* $[\pi(\tilde{x}), \pi^{-1}(\tilde{x})]$*, then* $B^{CL}(x, e)$ *and* $B^{CU}(x, e)$ *coincide with* $g(x, e)$*. Hence,* g *is point-identified. This result holds even if interval* $[\pi(\tilde{x}), \pi^{-1}(\tilde{x})]$ *is unknown.*

Example 2.2 (Quantile regression models)**.** *Assume* $g(X, \epsilon) = \theta_0(\epsilon) + \theta_1(\epsilon)X$*, where* $\theta_0(e) + \theta_1(e)x$ *is strictly increasing in* e *for all* $x \in \mathcal{X}$*. This model is a quantile regression model with endogeneity. The* τ*-th quantile function of* $g(x, \epsilon)$ *is* $\theta_0(\tau) + \theta_1(\tau)x$*. In this case, structural function* $g(x, e) = \theta_0(e) + \theta_1(e)x$ *is linear in* x*. Hence, Theorem 2.4 shows that* $\theta_0(e)$ *and* $\theta_1(e)$ *are identified if binary instruments are available.*

In this case, we can identify $\theta_0\left(Q_{\epsilon|\eta}(\tau|v)\right)]$ *and* $\theta_1\left(Q_{\epsilon|\eta}(\tau|v)\right)$ *by another approach. As in Section 2.3, we obtain* $\epsilon|X = Q_{X|Z}(v|z), Z = z \;\sim\; \epsilon|\eta = v$ *for all* $v \in (0, 1)$ *and* $z \in \{0, 1\}$*. This implies that*

$$\begin{aligned} Q_{Y|X,Z}\left(\tau|Q_{X|Z}(v|0), 0\right) &= \theta_0\left(Q_{\epsilon|\eta}(\tau|v)\right) + \theta_1\left(Q_{\epsilon|\eta}(\tau|v)\right) \times Q_{X|Z}(v|0), \\ Q_{Y|X,Z}\left(\tau|Q_{X|Z}(v|1), 1\right) &= \theta_0\left(Q_{\epsilon|\eta}(\tau|v)\right) + \theta_1\left(Q_{\epsilon|\eta}(\tau|v)\right) \times Q_{X|Z}(v|1). \end{aligned}$$

Because $Q_{X|Z}(v|0) \neq Q_{X|Z}(v|1)$ *under Assumption 2.3, for all* $\tau \in (0, 1)$ *and* $v \in (0, 1)$*, we can obtain* $\theta_0\left(Q_{\epsilon|\eta}(\tau|v)\right)$ *and* $\theta_1\left(Q_{\epsilon|\eta}(\tau|v)\right)$ *from the above equations. This result is similar to the identification results of Chesher (2003) and Jun (2009).*

The above model is a special case of the linear correlated random coefficients (CRC) model. Masten and Torgovitsky (2016) consider the linear CRC model and show that the expectations of coefficients are identified. In this model, we can also identify the expectations of coefficients as $E[\theta_j(\epsilon)]$*. Let* U *be a uniformly distributed random variable. Then, it follows from* $Q_{\epsilon|\eta}(U|v) \sim \epsilon|\eta = v$ *that* $\int_0^1 \theta_j\left(Q_{\epsilon|\eta}(\tau|v)\right) d\tau = E\left[\theta_j\left(Q_{\epsilon|\eta}(U|v)\right)\right] = E[\theta_j(\epsilon)|\eta = v]$*. Hence, since* η *is uniformly distributed, we have* $\int_0^1\int_0^1 \theta_j\left(Q_{\epsilon|\eta}(\tau|v)\right) d\tau dv = E[\theta_j(\epsilon)]$*.*

2.5 Calculating Bounds using Real Data

Here, we compute the bounds defined in Theorem 2.1 using the data in Macours et al. (2012) and show that our bounds are informative using real data. Specifically, Macours et al. (2012) analyze the income effects on early childhood cognitive development by using the *Atención a Crisis* program, which is a cash transfer program implemented in rural areas in Nicaragua. As in Example 2.1, we focus on the income effects on early childhood cognitive development.

In the analysis, we use only children between 5 and 7 years of age to control for age effects. The sample size for this analysis is 447, the size of the treatment group is 206, and that of the control group is 241. Following Macours et al. (2012), we use a standardized test score of receptive vocabulary (TVIP) as the outcome of a child's cognitive development. The average test score is 0.449 and the standard deviation 1.212. We use the logarithm of total consumption per capita as X, and let Z denotes the control indicator. We then estimate the conditional distribution and quantile functions and compute the bounds defined in Theorem 2.1 by treating these estimates as true functions.

Figure 2.11 shows the estimates of $F_{X|Z}(x|0)$ and $F_{X|Z}(x|1)$. This shows that these functions do not have any intersections. Hence, Assumption 2.3 is satisfied. Since the estimates of the tail of probability distributions are not reliable, we only use the estimates of $F_{X|Z}(x|0)$ and $F_{X|Z}(x|1)$ between 0.1 and 0.9. We thus compute $B^{ML}(x, 0.5)$ and $B^{MU}(x, 0.5)$ by using these estimates. Figure 2.12 shows the bounds. This implies that we can obtain informative bounds by using our identification approach. The average length of the difference is 0.045, which is rather small compared with the standard deviation of Y. These bounds show that the structural function is flat in x when x is low. As in Simulations 2.1 and 2.2, it is expected that this fact provides informative bounds of g.

2.6 Extension: General Models

In this section, we extend the results in Section 2.3 to more general models and allow Y to be discrete or censored. If outcomes are discrete or censored, then Assumptions 2.2 and 2.4 are not satisfied. Hence, we replace these assumptions with the following ones:

Assumption 2.2'. *(i) Function $g(x, e)$ is nondecreasing in e for all $x \in \mathcal{X}$. (ii) For all $z \in \{0, 1\}$, $h(z, v)$ is continuous and strictly increasing in v.*

Assumption 2.4'. *For all $(z, x) \in \{0, 1\} \times \mathcal{X}_z$, we have $\overline{\mathcal{Y}}_{x,z} = \overline{\mathcal{Y}}$. Here, we define $\overline{y} \equiv \sup\{y : y \in \overline{\mathcal{Y}}\}$ and $\underline{y} \equiv \inf\{y : y \in \overline{\mathcal{Y}}\}$.*

Assumption 2.2' (i) differs from Assumption 2.2 (i). Assumption 2.2 (i) imposes the strict monotonicity of $g(x, e)$ in e, while Assumption 2.2' (i) requires only the weak monotonicity of $g(x, e)$ in e. For example, if we consider

$$g(x, e) = \mathbf{1}\{e > (1 + \exp(\beta_0 + \beta_1 x))^{-1}\},$$

then $g(x, e)$ is not strictly increasing in e. Chesher (2010) and Shaikh and Vytlacil (2011) also employ a weak monotonicity condition. Assumption 2.4 implies that Y is continuously distributed, while Assumptions 2.2' and 2.4' allow outcomes that are discrete or censored.

D'Haultfœuille and Février (2015) and Torgovitsky (2015) do not consider the case in which the outcomes are discrete or censored because they assume that $g(x, e)$ is strictly increasing in e. Chesher (2010) and Shaikh and Vytlacil (2011) consider instrumental variable models for the discrete outcome. They also show that the structural or average structural functions are partially identified using instruments.

We show that $g(x,e)$ is partially identified under Assumptions 2.1, 2.2', 2.3, 2.4', 2.5, and 2.6.

We define

$$
\begin{aligned}
F^{+}_{Y|X,Z}(y|x,z) &\equiv P(Y \leq y|X=x, Z=z), \\
F^{-}_{Y|X,Z}(y|x,z) &\equiv P(Y < y|X=x, Z=z), \\
Q^{+}_{Y|X,Z}(\tau|x,z) &\equiv \sup\{y : F^{-}_{Y|X,Z}(y|x,z) \leq \tau\} \wedge \overline{y}, \\
Q^{-}_{Y|X,Z}(\tau|x,z) &\equiv \inf\{y : F^{+}_{Y|X,Z}(y|x,z) \geq \tau\} \vee \underline{y}.
\end{aligned}
$$

$F^{+}_{Y|X,Z}(y|x,z)$ and $Q^{+}_{Y|X,Z}(\tau|x,z)$ are right continuous in y and τ, and $F^{-}_{Y|X,Z}(y|x,z)$ and $Q^{-}_{Y|X,Z}(\tau|x,z)$ are left continuous in y and τ. Under Assumptions 2.2' and 2.4', Proposition 2.1 does not hold. Instead, we show the following proposition.

Proposition 2.2. *Define*

$$
\begin{aligned}
\hat{T}^{(1)}_{x}(y) &\equiv Q^{-}_{Y|X,Z}\left(F^{-}_{Y|X,Z}(y|x,0)\middle|\pi(x),1\right), \\
\check{T}^{(1)}_{x}(y) &\equiv Q^{+}_{Y|X,Z}\left(F^{+}_{Y|X,Z}(y|x,0)\middle|\pi(x),1\right), \\
\hat{T}^{(-1)}_{x}(y) &\equiv Q^{-}_{Y|X,Z}\left(F^{-}_{Y|X,Z}(y|x,1)\middle|\pi^{-1}(x),0\right), \\
\check{T}^{(-1)}_{x}(y) &\equiv Q^{+}_{Y|X,Z}\left(F^{+}_{Y|X,Z}(y|x,1)\middle|\pi^{-1}(x),0\right).
\end{aligned}
$$

Then, under Assumptions 2.1, 2.2', 2.3, 2.4', and 2.5, we have

$$
\begin{aligned}
g\left(\pi(x),e\right) &\geq \hat{T}^{(1)}_{x}\left(g(x,e)\right), \\
g\left(\pi(x),e\right) &\leq \check{T}^{(1)}_{x}\left(g(x,e)\right), \\
g\left(\pi^{-1}(x),e\right) &\geq \hat{T}^{(-1)}_{x}\left(g(x,e)\right), \\
g\left(\pi^{-1}(x),e\right) &\leq \check{T}^{(-1)}_{x}\left(g(x,e)\right).
\end{aligned}
$$

This approach is similar to the identification approaches of Athey and Imbens (2006) and Chesher (2010). Specifically, Athey and Imbens (2006) show that the counterfactual distribution is partially identified using right and left continuous quantile functions when outcomes are discrete. Chesher (2010) uses a result in which the weak monotonicity of $h(x,u)$ in u implies $\{u : h(x,u) \leq h(x,\tau)\} \supset \{u : u \leq \tau\}$ and $\{u : h(x,u) < h(x,\tau)\} \subset \{u : u < \tau\}$ and shows that structural function h is partially identified.

When the outcome is binary, this result is similar to Lemma 2.1 in Shaikh and Vytlacil (2011). They consider the following model:

$$
\begin{aligned}
Y &= \mathbf{1}\{v_1(D,X) \geq \epsilon_1\}, \\
D &= \mathbf{1}\{v_2(Z) \geq \epsilon_2\},
\end{aligned}
$$

where $(X,Z) \perp\!\!\!\perp (\epsilon_1, \epsilon_2)$. They then show that the sign of $v_1(1,x') - v_1(0,x)$ is identified under the appropriate support condition. Similarly, we can obtain the sign of $g(x,e) - g(\pi(x),e)$ from Proposition 2.2. When $P(Y=0|X=x,Z=0) > P(Y=0|X=\pi(x),Z=1)$, we have $\hat{T}_x^{(1)}(1)=1$ and $\hat{T}_x^{(1)}(0)=0$. It follows from Proposition 2.2 that $g(\pi(x),e) \geq g(x,e)$. Hence, we can identify the sign of $g(x,e)-g(\pi(x),e)$.

We define $\hat{T}_x^{(n)}(y) \equiv \hat{T}^{(1)}_{\pi^{n-1}(x)} \circ \cdots \circ \hat{T}_x^{(1)}(y)$ and $\check{T}_x^{(n)}(y) \equiv \check{T}^{(1)}_{\pi^{n-1}(x)} \circ \cdots \circ \check{T}_x^{(1)}(y)$ if $\pi^n(x)$ exists. Then, we have

$$\begin{aligned} g(\pi^n(x),e) &\geq \hat{T}_x^{(n)}\left(g(x,e)\right), \\ g(\pi^n(x),e) &\leq \check{T}_x^{(n)}\left(g(x,e)\right). \end{aligned}$$

Similarly, we define $\hat{T}_x^{(-n)}(y) \equiv \hat{T}^{(-1)}_{\pi^{-(n-1)}(x)} \circ \cdots \circ \hat{T}_x^{(-1)}(y)$ and $\check{T}_x^{(-n)}(y) \equiv \check{T}^{(-1)}_{\pi^{-(n-1)}(x)} \circ \cdots \circ \check{T}_x^{(-1)}(y)$ if $\pi^{-n}(x)$ exists. Then, we have

$$\begin{aligned} g(\pi^{-n}(x),e) &\geq \hat{T}_x^{(-n)}\left(g(x,e)\right), \\ g(\pi^{-n}(x),e) &\leq \check{T}_x^{(-n)}\left(g(x,e)\right). \end{aligned}$$

We define $\hat{T}_x^{(0)}(y) = y$ and $\check{T}_x^{(0)}(y) = y$ for any $x \in \mathcal{X}$.

If $(n,m) \in \Pi^M_{x',x}$, then Assumption 2.6 implies that

$$\hat{T}_{x'}^{(n)}\left(g(x',e)\right) \leq g(\pi^n(x'),e) \leq g(\pi^m(x),e) \leq \check{T}_x^{(m)}\left(g(x,e)\right).$$

If also define

$$\left(\hat{T}_{x'}^{(n)}\right)^{\rightarrow}(u) \equiv \sup\{y : \hat{T}_{x'}^{(n)}(y) \leq u\} \wedge \overline{y},$$

we have $\left(\hat{T}_{x'}^{(n)}\right)^{\rightarrow}\left(\hat{T}_{x'}^{(n)}(y)\right) = \sup\{y' : \hat{T}_{x'}^{(n)}(y') \leq \hat{T}_{x'}^{(n)}(y)\} \wedge \overline{y} \geq y$ for all $y \in \overline{\mathcal{Y}}$. Hence, we obtain

$$g(x',e) \leq \min_{(n,m)\in\Pi^M_{x',x}} \left(\hat{T}_{x'}^{(n)}\right)^{\rightarrow}\left(\check{T}_x^{(m)}(g(x,e))\right).$$

If defining $T^{GU}_{x',x}(y) \equiv \min_{(n,m)\in\Pi^M_{x',x}} \left(\hat{T}_{x'}^{(n)}\right)^{\rightarrow}\left(\check{T}_x^{(m)}(y)\right)$, then $T^{GU}_{x',x}(y)$ satisfies

$$g(x',e) \leq T^{GU}_{x',x}\left(g(x,e)\right). \tag{2.24}$$

Similarly, if we define $T^{GL}_{x',x}(y) \equiv \max_{(n,m)\in\Pi^M_{x,x'}} \left(\check{T}_{x'}^{(m)}\right)^{\leftarrow}\left(\hat{T}_x^{(n)}(y)\right)$ and $\left(\check{T}_{x'}^{(m)}\right)^{\leftarrow}(u) \equiv \inf\{y : \check{T}_{x'}^{(m)}(y) \geq u\} \vee \underline{y}$, then $T^{GL}_{x',x}(y)$ satisfies

$$g(x',e) \geq T^{GL}_{x',x}\left(g(x,e)\right). \tag{2.25}$$

We define

$$
\begin{aligned}
G_x^{GL}(u) &\equiv \int F^{+}_{Y|X}\left(T^{GU}_{x',x}(u)|x'\right) dF(x'), \\
G_x^{GU}(u) &\equiv \int F^{-}_{Y|X}\left(T^{GL}_{x',x}(u)|x'\right) dF(x'), \\
B^{GL}(x,e) &\equiv \sup_{y:y\leq x}\left\{\inf\{u : G_y^{GL}(u)\geq e\}\right\} \vee \underline{y}, \\
B^{GU}(x,e) &\equiv \inf_{y:y\geq x}\left\{\sup\{u : G_y^{GU}(u)\leq e\}\right\} \wedge \overline{y},
\end{aligned}
$$

where $F^{+}_{Y|X}(y|x) \equiv P(Y \leq y|X = x)$ and $F^{-}_{Y|X}(y|x) \equiv P(Y < y|X = x)$. $G_x^{GL}(u)$ and $G_x^{GU}(u)$ provide the lower and upper bounds of $g(x,e)$ by an argument similar to (2.4) and (2.5). $B^{GL}(x,e)$ and $B^{GU}(x,e)$ strengthen these bounds.

Theorem 2.5. *Under Assumptions 2.1, 2.2', 2.3, 2.4', 2.5, and 2.6, for all* $(x,e) \in \mathcal{X} \times (0,1)$*, we have*

$$B^{GL}(x,e) \leq g(x,e) \leq B^{GU}(x,e).$$

2.7 Conclusions

In this chapter, we consider the partial identification of nonseparable models using binary instruments. We show that partial identification can be achieved when $g(x,e)$ is monotone or concave in x, even if X is continuous and Z is binary. D'Haultfœuille and Février (2015) and Torgovitsky (2015) show that g is point identified without monotonicity and concavity. They use two key assumptions to establish the point identification of g. First, $F_{X|Z}(x|0)$ and $F_{X|Z}(x|1)$ have intersections and second, $g(x,e)$ is strictly increasing in a scalar unobservable. However, there are many empirically important models that do not satisfy these assumptions. As such, we provide bounds for structural functions without the use of these assumptions.

2.8 Appendix: Proofs

Proof of Proposition 2.1. Step.1 We show that, for all $e \in (0,1)$ and $x \in \mathcal{X}_0$,

$$P(\epsilon \leq e|X = x, Z = 0) = P(\epsilon \leq e|X = \pi(x), Z = 1). \tag{2.26}$$

First, we examine variable $V \equiv F_{X|Z}(X|Z)$. This is called "control variable" in Imbens and Newey (2009). Let $h^{-1}(z,x)$ be the inverse function of $h(z,v)$ with respect to v. We thus have, for all $(z,x) \in \{0,1\} \times \mathcal{X}_z$,

$$\begin{aligned} F_{X|Z}(x|z) &= P\left(h(z,\eta) \leq x|Z = z\right) \\ &= P\left(\eta \leq h^{-1}(z,x)|Z = z\right) \\ &= P\left(\eta \leq h^{-1}(z,x)\right) = h^{-1}(z,x), \end{aligned}$$

where the second equality follows from the strict monotonicity of $h(x,v)$ in v and the third equality follows from $Z \perp\!\!\!\perp (\epsilon, \eta)$. Therefore, we obtain

$$V = h^{-1}(Z,X) = \eta.$$

Next, we show that the conditional distribution of ϵ conditional on $(X,Z) = (x,z)$ is the same as that of ϵ conditional on $V = F_{X|Z}(x|z)$. Because $(x,z) \to (F_{X|Z}(x|z), z)$ is one-to-one and $F_{X|Z}(x|z)$ is continuous in x, the σ-field generated by X and Z is the same as that generated by V and Z. Hence, we have

$$P\left(\epsilon \leq e|X = x, Z = z\right) = P\left(\epsilon \leq e|V = F_{X|Z}(x|z), Z = z\right).$$

It follows from $Z \perp\!\!\!\perp (\epsilon, \eta)$ and $V = \eta$ that

$$P\left(\epsilon \leq e|X = x, Z = z\right) = P\left(\epsilon \leq e|V = F_{X|Z}(x|z)\right). \tag{2.27}$$

Hence, the conditional distribution of ϵ conditional on X and Z solely depends on $V = F_{X|Z}(X|Z)$.

By definition, functions $\pi(x)$ and $\pi^{-1}(x)$ satisfy

$$\begin{aligned} F_{X|Z}(\pi(x)|1) =& \; F_{X|Z}(x|0), \\ F_{X|Z}(\pi^{-1}(x)|0) =& \; F_{X|Z}(x|1). \end{aligned} \tag{2.28}$$

Hence, events $\{X = x, Z = 0\}$ and $\{X = \pi(x), Z = 1\}$ have the same $V = F_{X|Z}(X|Z)$, and (2.26) follows from (2.27).

Step.2 We show that (2.26) implies $g\left(\pi(x), e\right) = \tilde{T}_x^{(1)}\left(g(x,e)\right)$. For all $(x,e) \in \mathcal{X}_0 \times (0,1)$, we have

$$\begin{aligned}
\tilde{T}_x^{(1)}\left(g(x,e)\right) &= Q_{Y|X=\pi(x),Z=1}\left(F_{Y|X=x,Z=0}\left(g(x,e)\right)\right) \\
&= Q_{Y|X=\pi(x),Z=1}\left(P(\epsilon \le e|X=x, Z=0)\right) \\
&= Q_{Y|X=\pi(x),Z=1}\left(P(\epsilon \le e|X=\pi(x), Z=1)\right) \\
&= Q_{Y|X=\pi(x),Z=1}\left(F_{Y|X=\pi(x),Z=1}\left(g(\pi(x),e)\right)\right) \\
&= g(\pi(x),e),
\end{aligned}$$

where the third equality follows from (2.26).

Similarly, we can prove $g\left(\pi^{-1}(x),e\right) = \tilde{T}_x^{(-1)}\left(g(x,e)\right)$. □

Proof of Lemma 2.1. Observe that, if $\pi^n(x)$ exists and $\pi^n(x) \in \mathcal{X}_0$, then $\pi^{n+1}(x)$ also exists from (2.6). Suppose that there does not exist $n \in \mathbb{N} \cup \{0\}$ such that $\pi^n(x) \in \mathcal{X}_1 \cap \mathcal{X}_0^c = (\underline{x}_1, \underline{x}_0]$. Then, there exists sequence $\{x_n\}_{n=0}^{\infty}$ such that $x_n = \pi^n(x)$. By (2.9), $\{x_n\}_{n=0}^{\infty}$ is a decreasing sequence. Because $x_n > \underline{x}_0$, $\{x_n\}_{n=0}^{\infty}$ converges to $x_\infty \in [\underline{x}_0, \overline{x}^0)$. It follows from (2.28) that

$$F_{X|Z}(x_{n+1}|1) = F_{X|Z}(x_n|0),$$

meaning we have $F_{X|Z}(x_\infty|1) = F_{X|Z}(x_\infty|0)$ by the continuity of $F_{X|Z}$. However, this equation violates Assumption 2.3. Hence, for all $x \in \mathcal{X}$, there exists $n \in \mathbb{N} \cup \{0\}$ such that $\pi^n(x) \in \mathcal{X}_1 \cap \mathcal{X}_0^c$. Consequently, $\pi^{n'}(x)$ does not exist for $n' > n$. Similarly, for all $x \in \mathcal{X}$, we have $\pi^{-m}(x) \in \mathcal{X}_0 \cap \mathcal{X}_1^c$ for some $m \in \mathbb{N} \cup \{0\}$. Then, $\pi^{-m'}(x)$ does not exist for $m' > m$. Therefore, $\Pi_{x',x}^M$ is finite for all $(x,x') \in \mathcal{X} \times \mathcal{X}$ because the set $\{(n,m) \in \mathbb{Z} \times \mathbb{Z} : \pi^n(x') \text{ and } \pi^m(x) \text{ exist.}\}$ is finite.

We proceed to show the nonemptiness of $\Pi_{x',x}^M$. For all $x, x' \in \mathcal{X}$, $(n,m) \in \mathbb{Z} \times \mathbb{Z}$ exists such that $\pi^n(x') \in \mathcal{X}_1 \cap \mathcal{X}_0^c = (\underline{x}_1, \underline{x}_0]$ and $\pi^m(x) \in \mathcal{X}_0 \cap \mathcal{X}_1^c = [\overline{x}_1, \overline{x}_0)$. It follows from Assumption 2.3 (ii) that $\pi^n(x') < \pi^m(x)$. □

Proof of Theorem 2.1. As discussed in Section 2.3, it suffices to show that $T_{x',x}^{ML}(y)$ and $T_{x',x}^{MU}(y)$ are strictly increasing in y and surjective. If $\pi^n(x)$ exists, $\tilde{T}_x^{(n)}(y)$ is strictly increasing in y. Hence, $T_{x',x}^{ML}(y)$ and $T_{x',x}^{MU}(y)$ are strictly increasing in y because $\Pi_{x',x}^M$ is finite by Lemma 2.1. If $\pi^n(x)$ exists, we obtain $\tilde{T}_x^{(n)}([\underline{y}, \overline{y}]) = [\underline{y}, \overline{y}]$. Because $\Pi_{x',x}^M$ is finite, we have $T_{x',x}^{ML}(y)$ and $T_{x',x}^{MU}(y)$ are surjective. □

Proof of Theorem 2.2. Step.1 First, we show that, for all $x \in \mathcal{X}$, there exists $n^* \in \mathbb{Z}$ such that $\pi^{n^*}(x)$ and $\pi^{n^*+1}(x)$ are well defined and $\pi^{n^*}(x), \pi^{n^*+1}(x) \in [\pi(\tilde{x}), \pi^{-1}(\tilde{x})]$. If $\pi^n(x)$ and $\pi^n(y)$ are well defined, because $\pi^n(\cdot)$ is strictly increasing, we can obtain

$$x \leq y \quad \Rightarrow \quad \pi^n(x) \leq \pi^n(y). \tag{2.29}$$

We consider the following four cases: (i) $\pi(\tilde{x}) \leq x \leq \tilde{x}$, (ii) $\tilde{x} \leq x \leq \pi^{-1}(\tilde{x})$, (iii) $x < \pi(\tilde{x})$, and (iv) $x > \pi^{-1}(\tilde{x})$. In case (i), it follows from (2.29) that $\pi(\tilde{x}) \leq x \leq \tilde{x} \leq \pi^{-1}(x) \leq \pi^{-1}(\tilde{x})$. In case (ii), it follows from (2.29) that $\pi(\tilde{x}) \leq \pi(x) \leq \tilde{x} \leq x \leq \pi^{-1}(\tilde{x})$. In case (iii), it follows from the proof of Lemma 2.1 that $n \in \mathbb{N}$ exists such that $\pi^{-n}(x) \in \mathcal{X}_0 \cap \mathcal{X}_1^c$. This implies that $\pi^{-1}(x), ..., \pi^{-n}(x)$ exist. By the definition of π, we have $\pi(\tilde{x}) \in \mathcal{X}_1$, and hence $x < \pi(\tilde{x}) < \pi^{-n}(x)$. Therefore, there exists $n^* \in \mathbb{Z}$ such that $\pi^{n^*+2}(x) \leq \pi(\tilde{x}) \leq \pi^{n^*+1}(x)$ and we can obtain $\pi(\tilde{x}) \leq \pi^{n^*+1}(x) \leq \tilde{x} \leq \pi^{n^*}(x) \leq \pi^{-1}(\tilde{x})$ from (2.29). Similarly, in case (iv), there exists $n^* \in \mathbb{Z}$ such that $\pi^{n^*}(x), \pi^{n^*+1}(x) \in [\pi(\tilde{x}), \pi^{-1}(\tilde{x})]$.
Step.2 Next, we show that g is point identified. From step 1, for all $x, x' \in \mathcal{X}$, there exists $n, m \in \mathbb{Z}$ such that $\pi^n(x'), \pi^{n+1}(x'), \pi^m(x), \pi^{m+1}(x) \in [\pi(\tilde{x}), \pi^{-1}(\tilde{x})]$. Then, from (2.29), we have either $\pi^{n+1}(x') \leq \pi^{m+1}(x) \leq \pi^n(x') \leq \pi^m(x)$ or $\pi^{m+1}(x) \leq \pi^{n+1}(x') \leq \pi^m(x) \leq \pi^n(x')$. If $\pi^{n+1}(x') \leq \pi^{m+1}(x) \leq \pi^n(x') \leq \pi^m(x)$, then we have $(n+1, m+1), (n, m) \in \Pi^M_{x',x}$. If $\pi^{m+1}(x) \leq \pi^{n+1}(x') \leq \pi^m(x) \leq \pi^n(x')$, then we have $(n+1, m) \in \Pi^M_{x',x}$. Hence, there exists a pair $(n^*, m^*) \in \Pi^M_{x',x}$ such that $\pi^{n^*}(x'), \pi^{m^*}(x) \in [\pi(\tilde{x}), \pi^{-1}(\tilde{x})]$. As $g(x, e)$ is constant on $[\pi(\tilde{x}), \pi^{-1}(\tilde{x})]$, we obtain

$$\tilde{T}^{(n^*)}_{x'}\left(g(x', e)\right) = \tilde{T}^{(m^*)}_{x}\left(g(x, e)\right).$$

Therefore, $g(x', e) = T^{MU}_{x',x}\left(g(x, e)\right)$. Hence, $\left(G^{ML}_x\right)^{-1}(e)$ coincides with $g(x, e)$ because (2.15) becomes an equality. This implies that $B^{ML}(x, e)$ coincides with $g(x, e)$. Similarly, $B^{MU}(x, e)$ coincides with $g(x, e)$. □

Proof of Lemma 2.2. From the proof of Lemma 2.1, $\Pi^C_{x',x}$ is finite. Hence, we prove the nonemptiness of $\Pi^C_{x',x}$. From the proof of Lemma 2.1, for all $x, x' \in \mathcal{X}$, there exist $n, m \in \mathbb{Z}$ such that $\pi^m(x), \pi^n(x') \in \mathcal{X}_1 \cap \mathcal{X}_0^c$. Without loss of generality, we assume $\pi^n(x') \leq \pi^m(x)$. Then, $\pi^{m-1}(x)$ and $\pi^{n-1}(x')$ exist because $\pi^m(x), \pi^n(x') \in \mathcal{X}_1$. Because $\pi^m(x) \in \mathcal{X}_1 \cap \mathcal{X}_0^c$ and $\pi^{n-1}(x') \in \mathcal{X}_0$, we have $\pi^n(x') \leq \pi^m(x) \leq \pi^{n-1}(x') \leq \pi^{m-1}(x)$ from (2.29), and hence $(n, m) \in \Pi^C_{x',x}$. Therefore, $\Pi^C_{x',x}$ is nonempty. □

Proof of Theorem 2.3. Similar to the proof of Theorem 2.1, we can obtain

$$\left(G^{CL}_x\right)^{-1}(e) \leq g(x, e) \leq \left(G^{CU}_x\right)^{-1}(e).$$

Because $g(x,e)$ is concave in x, if $x = ty' + (1-t)y$ and $t \in (0,1)$, then we have $g(x,e) \geq tg(y',e) + (1-t)g(y,e) \geq t\left(G^{CL}_{y'}\right)^{-1}(e) + (1-t)\left(G^{CL}_{y}\right)^{-1}(e)$. Hence, we have

$$g(x,e) \geq \sup_{y,y':y<x<y'} \left\{ \left(\frac{x-y}{y'-y}\right)\left(G^{CL}_{y'}\right)^{-1}(e) + \left(\frac{y'-x}{y'-y}\right)\left(G^{CL}_{y}\right)^{-1}(e) \right\}.$$

Because $g(x,e)$ is concave in x, if $x = ty' + (1-t)y$ and $t < 0$, then we have $g(x,e) \leq tg(y',e) + (1-t)g(y,e)$. Because $B^{CL}(x,e) \leq g(x,e) \leq \left(G^{CU}_{x}\right)^{-1}(e)$, $t<0$, and $1-t>0$, we have $g(x,e) \leq tB^{CL}(y',e) + (1-t)\left(G^{CU}_{y}\right)^{-1}(e)$. Similarly, if $x = ty + (1-t)y'$ and $t > 1$, then we have $g(x,e) \leq tg(y,e) + (1-t)g(y',e) \leq t\left(G^{CU}_{y}\right)^{-1}(e) + (1-t)B^{CL}(y',e)$. Hence, we have

$$\begin{aligned} g(x,e) \leq \ \min\Big[& \inf_{y,y':x<y<y'} \left\{ \left(\tfrac{x-y}{y'-y}\right) B^{CL}(y',e) + \left(\tfrac{y'-x}{y'-y}\right)\left(G^{CU}_{y}\right)^{-1}(e) \right\}, \\ & \inf_{y,y':y'<y<x} \left\{ \left(\tfrac{y-x}{y-y'}\right) B^{CL}(y',e) + \left(\tfrac{x-y'}{y-y'}\right)\left(G^{CU}_{y}\right)^{-1}(e) \right\} \Big]. \end{aligned}$$

□

Proof of Theorem 2.4. Similar to Theorem 2.2, for all $x, x' \in \mathcal{X}$, there exist $(n,m) \in \Pi^C_{x',x}$ such that $\pi^n(x')$, $\pi^{n-1}(x')$, and $\pi^m(x)$ are in $[\pi(\tilde{x}), \pi^{-1}(\tilde{x})]$. Because $g(x,e)$ is linear in x, we have

$$\left[t_{x',x}(n,m)\tilde{T}^{(n)}_{x'} + (1 - t_{x',x}(n,m))\tilde{T}^{(n-1)}_{x'}\right](g(x',e)) = \tilde{T}^{(m)}_{x}(g(x,e)).$$

Similarly, for all $x, x' \in \mathcal{X}$, there exist $(n,m) \in \Pi^C_{x,x'}$ such that

$$\tilde{T}^{(m)}_{x'}(g(x',e)) = \left[t_{x,x'}(n,m)\tilde{T}^{(n)}_{x} + (1 - t_{x,x'}(n,m))\tilde{T}^{(n-1)}_{x}\right](g(x,e)).$$

Hence, as described above, $B^{CL}(x,e)$ and $B^{CU}(x,e)$ coincide with $g(x,e)$ because inequalities (2.20) and (2.21) become equalities. □

Proof of Proposition 2.2. Because $g(x,e)$ is nondecreasing in e, we have

$$\begin{aligned} F^-_{Y|X,Z}(g(x,e)|x,0) &= P(g(x,\epsilon) < g(x,e)|X=x, Z=0) \\ &\leq P(\epsilon < e|X=x, Z=0) \\ &= P(\epsilon \leq e|X=\pi(x), Z=1) \\ &\leq P(g(\pi(x),\epsilon) \leq g(\pi(x),e)\,|X=\pi(x), Z=1) \\ &= F^+_{Y|X,Z}(g(\pi(x),e)|\pi(x),1), \end{aligned} \tag{2.30}$$

where the first inequality follows from $\{\epsilon : g(x,\epsilon) < g(x,e)\} \subset \{\epsilon : \epsilon < e\}$ and the second from $\{\epsilon : \epsilon \leq e\} \subset \{\epsilon : g(x,\epsilon) \leq g(x,e)\}$. From the definition of $Q^-_{Y|X,Z}(\tau|x,z)$, it follows that $Q^-_{Y|X,Z}\left(F^+_{Y|X,Z}(y|x,z)\Big|x,z\right) = \inf\{y' : F^+_{Y|X,Z}(y'|x,z) \geq F^+_{Y|X,Z}(y|x,z)\} \vee \underline{y} \leq y$ for all $y \in \overline{\mathcal{Y}}$. Hence, inequality (2.30) implies that

$$\begin{aligned}
\hat{T}^{(1)}_x\left(g(x,e)\right) &= Q^-_{Y|X,Z}\left(F^-_{Y|X,Z}(g(x,e)|x,0)\Big|\pi(x),1\right) \\
&\leq Q^-_{Y|X,Z}\left(F^+_{Y|X,Z}\left(g(\pi(x),e)|\pi(x),1\right)\Big|\pi(x),1\right) \\
&\leq g(\pi(x),e).
\end{aligned}$$

Similarly, because $g(x,e)$ is nondecreasing in e, we have

$$F^+_{Y|X,Z}\left(g(x,e)|x,0\right) \geq F^-_{Y|X,Z}\left(g(\pi(x),e)|\pi(x),1\right).$$

Because $Q^+_{Y|X,Z}\left(F^-_{Y|X,Z}(y|x,z)\Big|x,z\right) = \sup\{y' : F^-_{Y|X,Z}(y'|x,z) \leq F^-_{Y|X,Z}(y|x,z)\} \wedge \overline{y} \geq y$ for all $y \in \overline{\mathcal{Y}}$, we have

$$g(\pi(x),e) \leq \check{T}^{(1)}_x\left(g(x,e)\right).$$

Similarly, we have two inequalities: $g(\pi^{-1}(x),e) \geq \hat{T}^{(-1)}_x\left(g(x,e)\right)$ and $g(\pi^{-1}(x),e) \leq \check{T}^{(-1)}_x\left(g(x,e)\right)$. □

Proof of Theorem 2.5. First, we show that

$$\inf\{u : G^{GL}_x(u) \geq e\} \vee \underline{y} \leq g(x,e) \leq \sup\{u : G^{GU}_x(u) \leq e\} \wedge \overline{y}. \tag{2.31}$$

Because $T^{GU}_{x',x}(y)$ satisfies (2.24), we have

$$\begin{aligned}
G^{GL}_x\left(g(x,e)\right) &= \int F^+_{Y|X=x'}\left(T^{GU}_{x',x}(g(x,e))\right) dF(x') \\
&\geq \int F^+_{Y|X=x'}\left(g(x',e)\right) dF(x') \\
&= \int P\left(g(x',\epsilon) \leq g(x',e)|X=x'\right) dF(x') \\
&\geq \int P\left(\epsilon \leq e|X=x'\right) dF(x') = e,
\end{aligned}$$

where the second inequality follows from $\{\epsilon : \epsilon \leq e\} \subset \{\epsilon : g(x',\epsilon) \leq g(x',e)\}$. Because $g(x,e) \geq \underline{y}$, we can obtain $g(x,e) \geq \inf\{u : G^{GL}_x(u) \geq e\} \vee \underline{y}$. Similarly, because $T^{GL}_{x',x}(y)$ satisfies (2.25), we have

$$\begin{aligned}
G^{GU}_x\left(g(x,e)\right) &\leq \int F^-_{Y|X=x'}\left(g(x',e)\right) dF(x') \\
&= \int P\left(g(x',\epsilon) < g(x',e)|X=x'\right) dF(x') \\
&\leq \int P\left(\epsilon < e|X=x'\right) dF(x') = e,
\end{aligned}$$

where the second inequality follows from $\{\epsilon : g(x', \epsilon) < g(x', e)\} \subset \{\epsilon : \epsilon < e\}$. Hence, we can obtain $g(x, e) \leq \sup\{u : G_x^{GU}(u) \leq e\} \wedge \overline{y}$.

Because $g(x, e)$ is nondecreasing in x and (2.31) holds, similar to Theorem 2.1, we have $B^{GL}(x, e) \leq g(x, e) \leq B^{GU}(x, e)$. □

2.9 Appendix: Figures

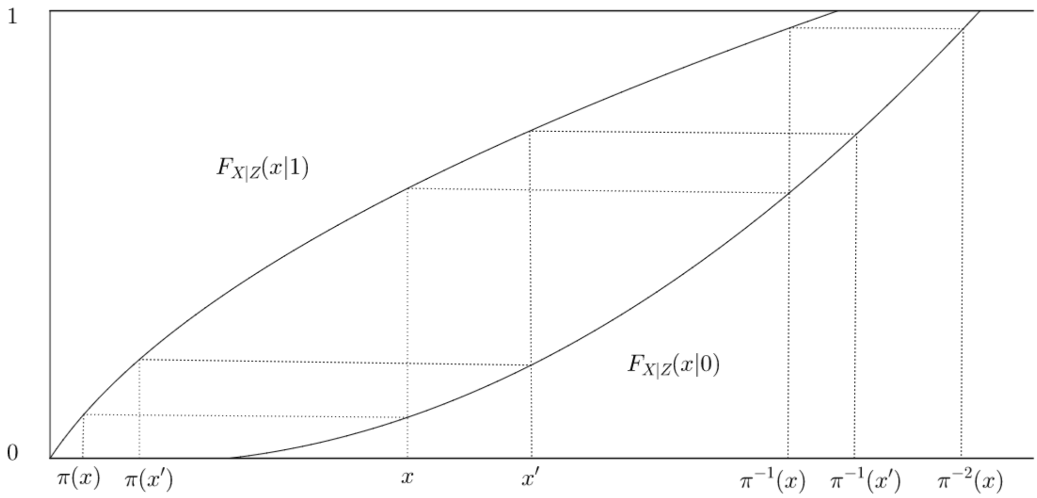

Figure 2.1: The case where Assumption 2.3 holds.

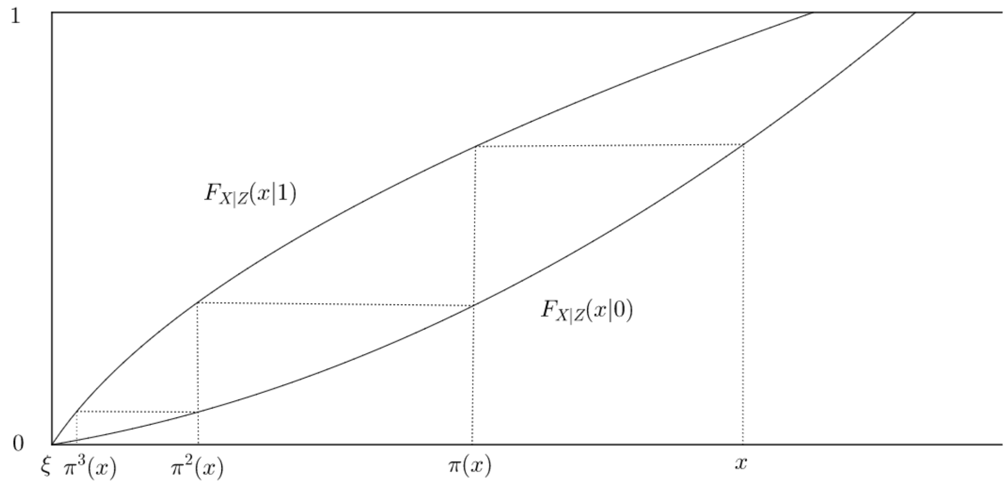

Figure 2.2: The case where Assumption 2.3 does not hold.

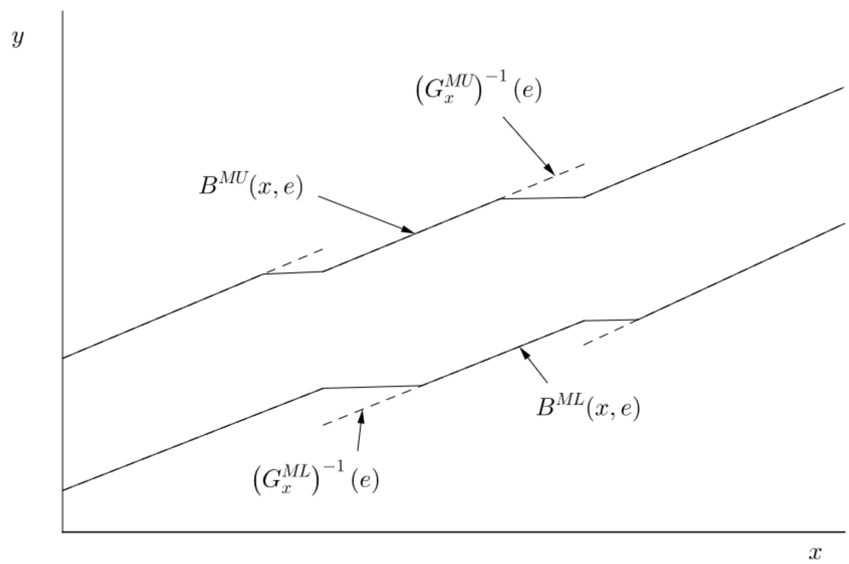

Figure 2.3: The dashed lines denote $\left(G_x^{ML}\right)^{-1}(e)$ and $\left(G_x^{MU}\right)^{-1}(e)$. The solid lines denote $B^{ML}(x,e)$ and $B^{MU}(x,e)$.

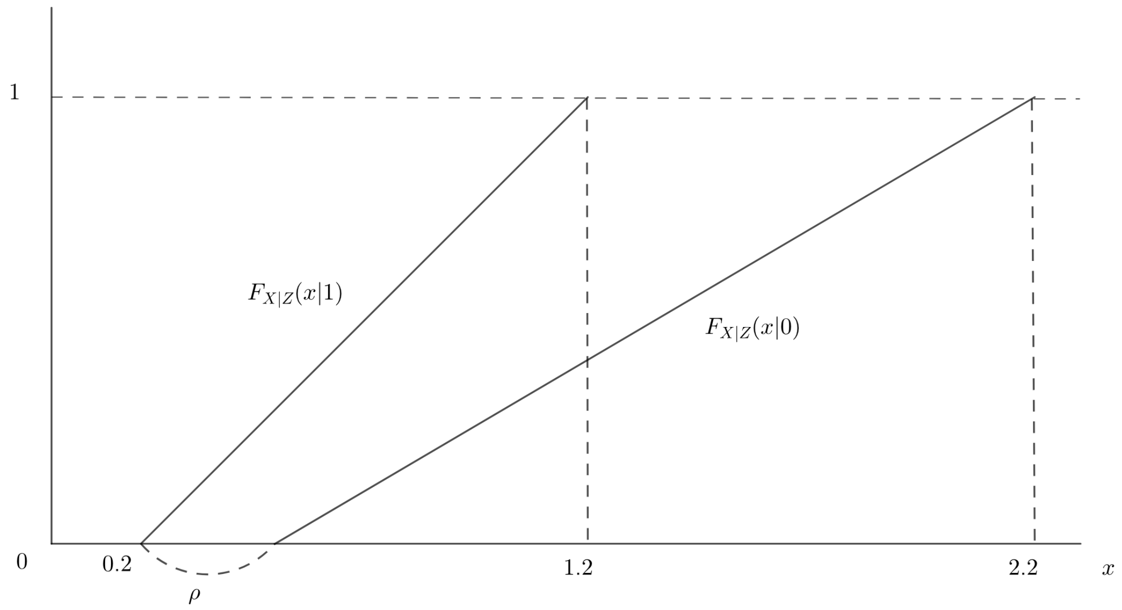

Figure 2.4: $F_{X|Z}(x|0)$ and $F_{X|Z}(x|1)$ for Simulation 2.1.

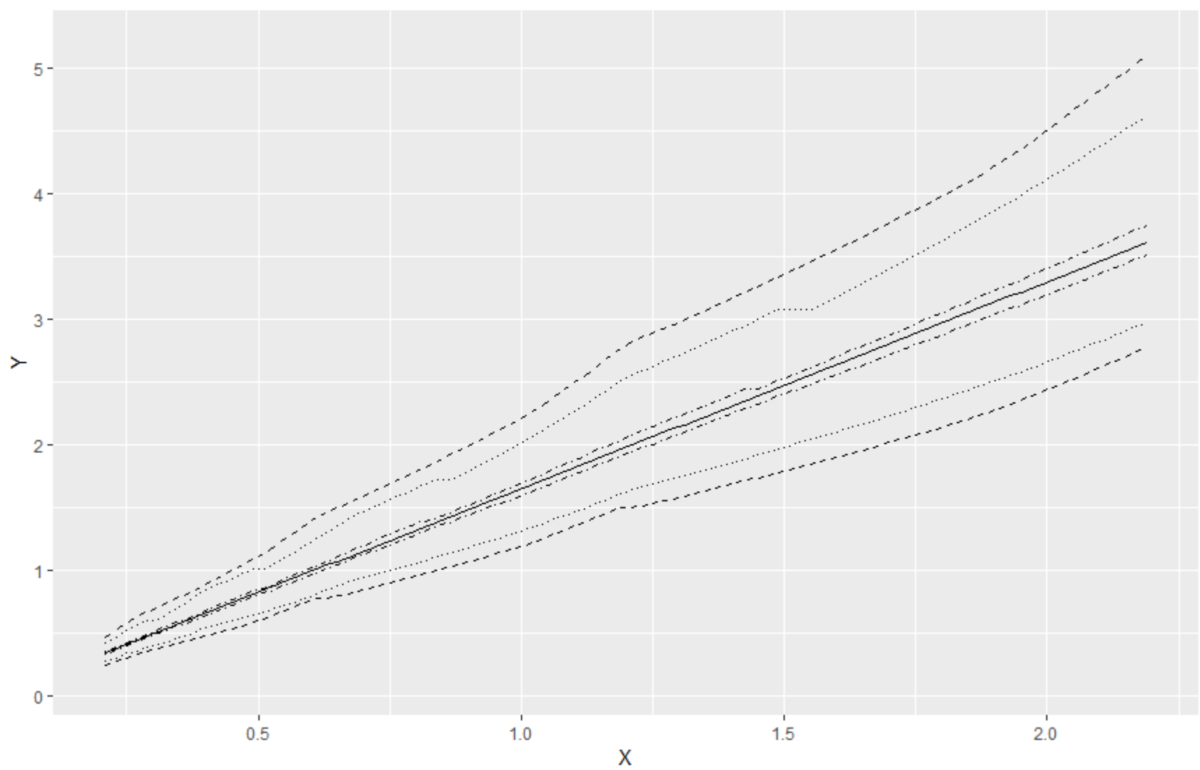

Figure 2.5: $h(x) = h_1(x)$. The solid line denotes $g(x, 0.5)$. The dashed, dotted, and dash-dotted lines denote $B^{ML}(x, 0.5)$ and $B^{MU}(x, 0.5)$ when $\rho = 0.3$, 0.1, and 0.01, respectively.

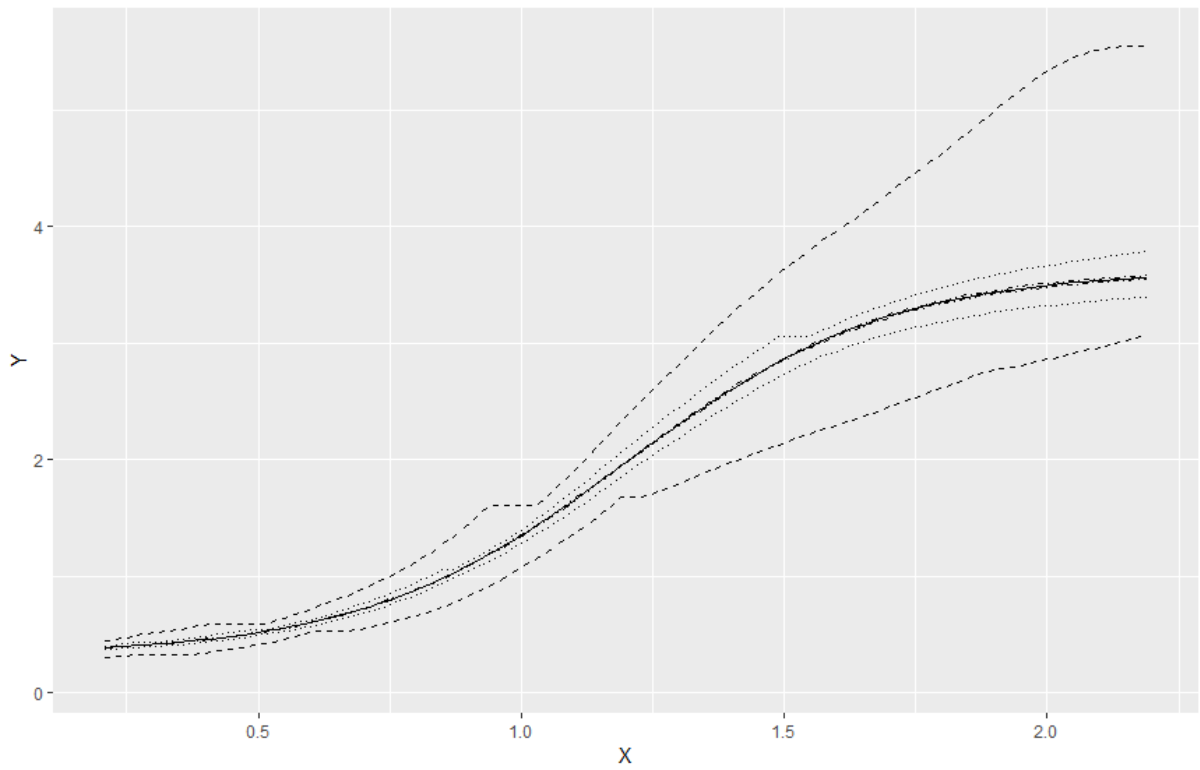

Figure 2.6: $h(x) = h_2(x)$. The solid line denotes $g(x, 0.5)$. The dashed, dotted, and dash-dotted lines denote $B^{ML}(x, 0.5)$ and $B^{MU}(x, 0.5)$ when $\rho = 0.3$, 0.1, and 0.01, respectively.

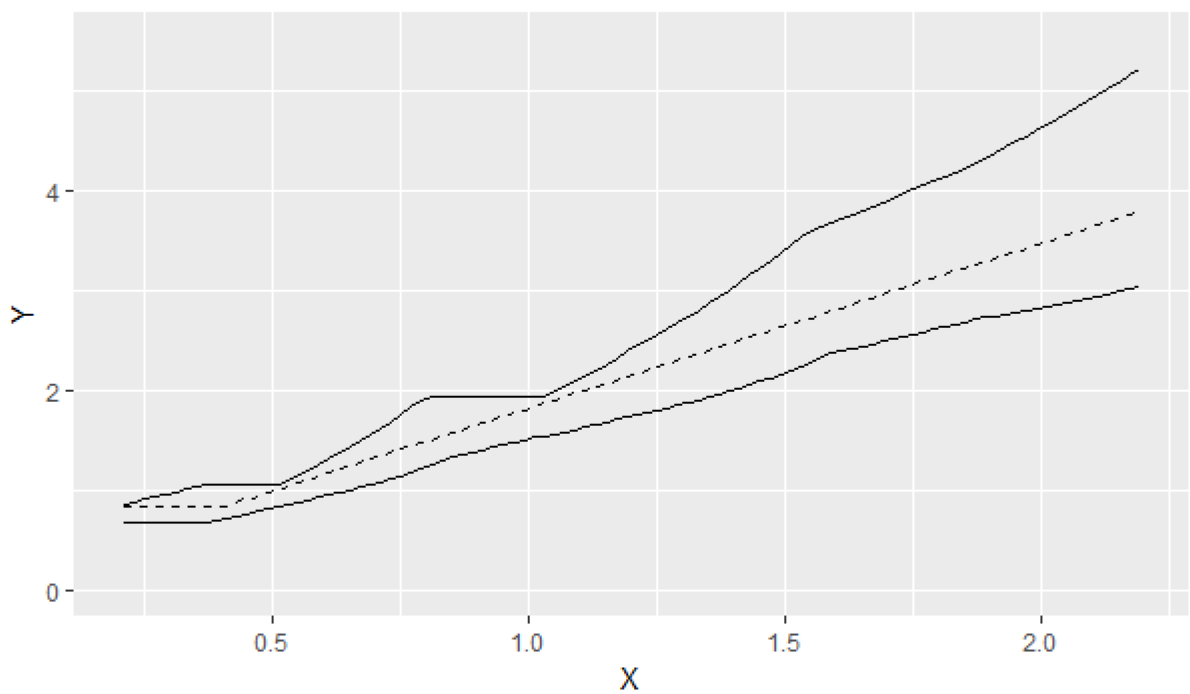

Figure 2.7: $\delta = 0.4$. The dashed line denotes $g(x, 0.5)$. The solid lines denote $B^{ML}(x, 0.5)$ and $B^{MU}(x, 0.5)$.

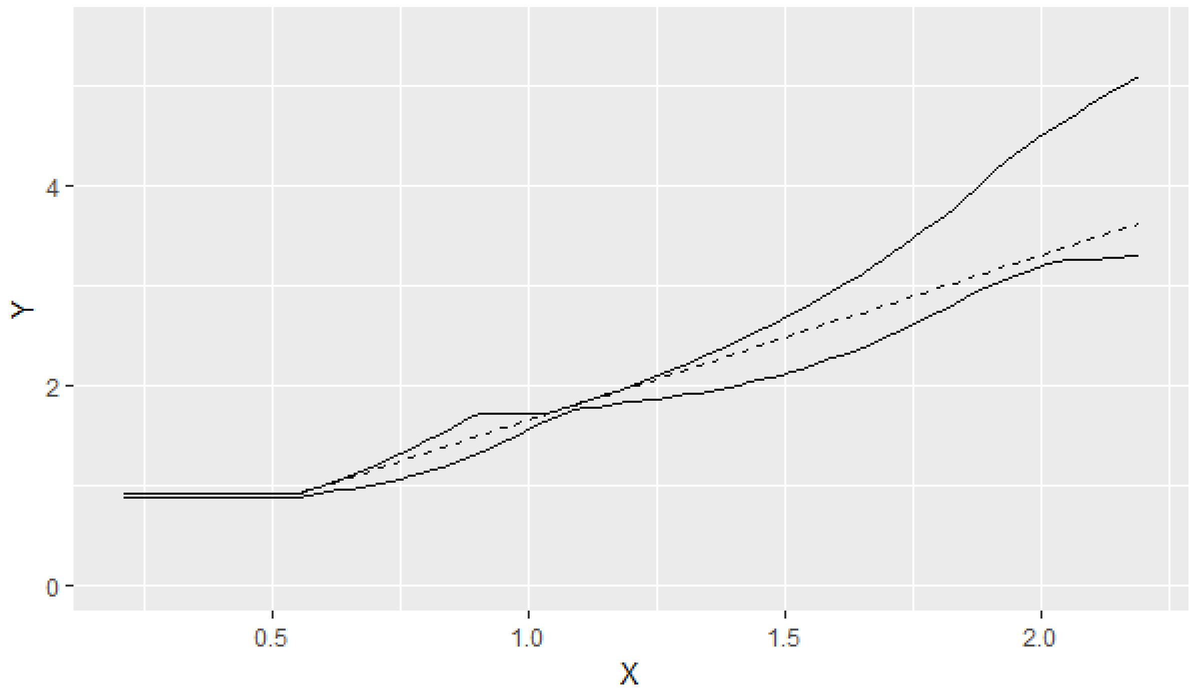

Figure 2.8: $\delta = 0.55$. The dashed line denotes $g(x, 0.5)$. The solid lines denote $B^{ML}(x, 0.5)$ and $B^{MU}(x, 0.5)$.

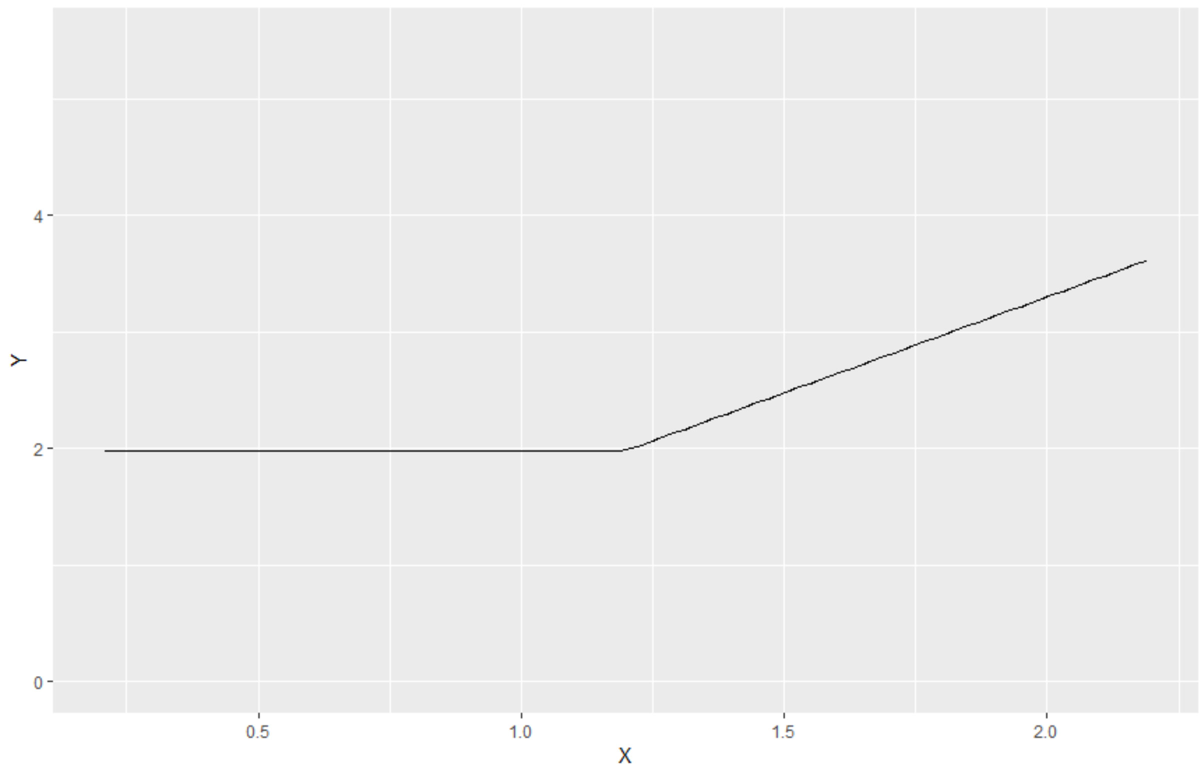

Figure 2.9: $\delta = 1.2$. The dashed line denotes $g(x, 0.5)$. The solid lines denote $B^{ML}(x, 0.5)$ and $B^{MU}(x, 0.5)$.

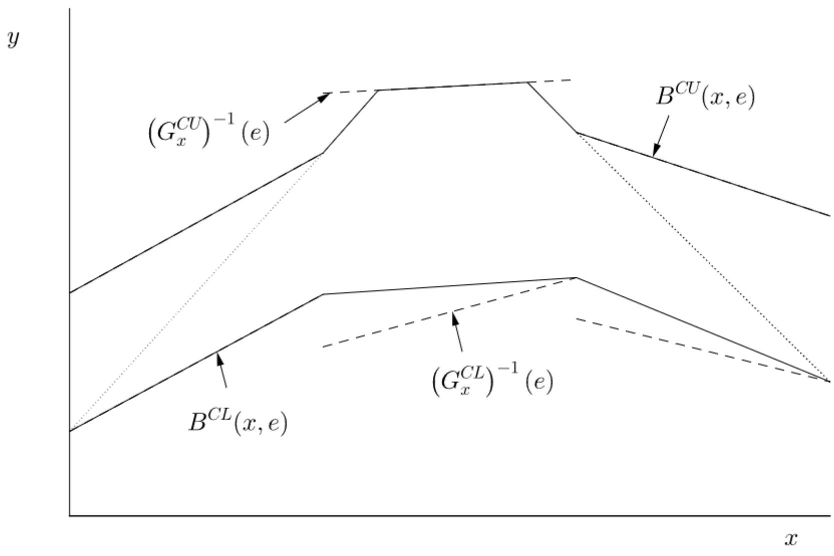

Figure 2.10: The dashed lines denote $\left(G_x^{CL}\right)^{-1}(e)$ and $\left(G_x^{CU}\right)^{-1}(e)$. The solid lines denote $B^{CL}(x,e)$ and $B^{CU}(x,e)$.

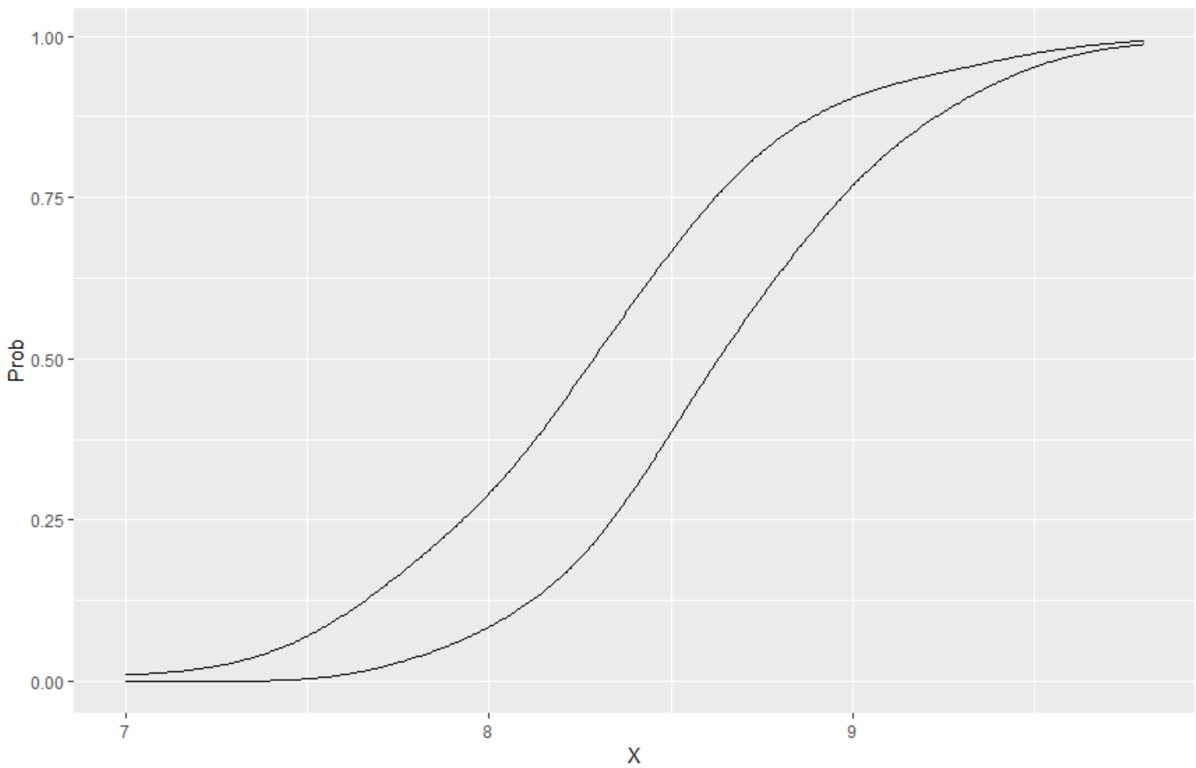

Figure 2.11: The right-hand line denotes $F_{X|Z}(x|0)$ and the left-hand one denotes $F_{X|Z}(x|1)$.

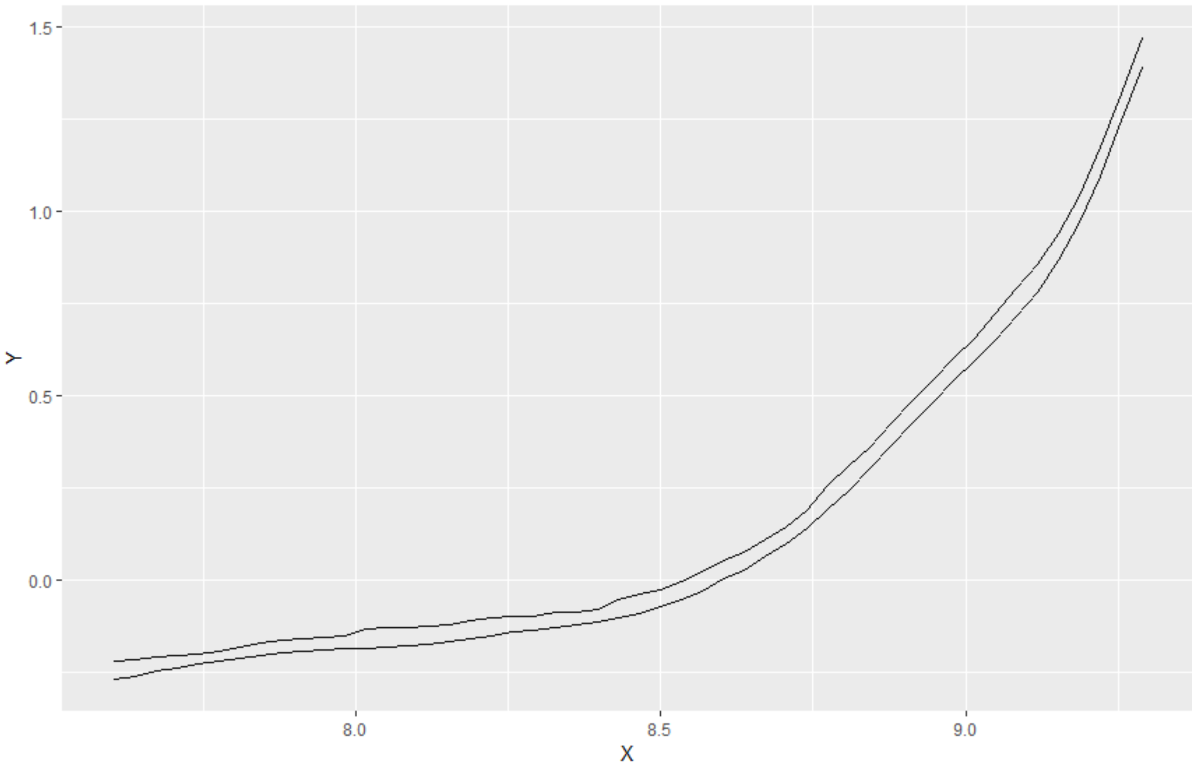

Figure 2.12: The lower line denotes $B^{ML}(x, 0.5)$ and the upper one denotes $B^{MU}(x, 0.5)$.

Chapter 3

Identification and Estimation of Time-Varying Nonseparable Panel Data Models without Stayers†

3.1 Introduction

In this chapter, we consider the identification and estimation of the following nonseparable panel data model:

$$Y_{it} = g_t(X_{it}, U_{it}), \quad i = 1, \cdots, n, \quad t = 1, \cdots, T, \tag{3.1}$$

where $Y_{it} \in \mathbb{R}$ is a scalar response variable, $X_{it} \in \mathbb{R}^k$ is a vector of explanatory variables, and $U_{it} \in \mathbb{R}$ is a scalar unobservable variable. Suppose that $\mathbf{Y}_i = (Y_{i1}, \cdots, Y_{iT})$ and $\mathbf{X}_i = (X'_{i1}, \cdots, X'_{iT})$ are observable. Many widely used panel data models fall into this category. For example, this specification contains the textbook linear panel data model

$$Y_{it} = X'_{it}\beta + \alpha_i + \epsilon_{it},$$

because we can regard $\alpha_i + \epsilon_{it}$ as U_{it}. Furthermore, it contains the following nonlinear panel data models:

$$Y_{it} = h^{-1}(X'_{it}\beta + \gamma_t + \alpha_i + \epsilon_{it}), \tag{3.2}$$

$$Y_{it} = c(U(\alpha_i, \epsilon_{it})) + X'_{it}\beta(U(\alpha_i, \epsilon_{it})), \tag{3.3}$$

where (3.2) is the transformation model proposed by Abrevaya (1999), and (3.3) is the random effects quantile regression model proposed by Galvao and Poirier (2017).

†This chapter is based on Ishihara (2019).

The importance of unobserved heterogeneity when modeling economic behavior is widely recognized. Nonseparable models capture the unobserved heterogeneity effect of explanatory variables on outcomes because these models allow the derivative of the structural function to depend on an unobserved variable. Indeed, there is extensive literature on nonseparable panel data models including Altonji and Matzkin (2005),Evdokimov (2010), Hoderlein and White (2012), Chernozhukov, Fernández-Val, Hahn, and Newey (2013), D'Haultfoeuille et al. (2013), and Chernozhukov, Fernandez-Val, Hoderlein, Holzmann, and Newey (2015).

This study shows that we can nonparametrically identify $g_t(x, u)$ when $g_t(x, u)$ is strictly increasing in u, the conditional distributions of U_{it} are the same over time, and the support of $\mathbf{X}_i$ satisfies some weak assumptions. To identify the target parameters, many nonseparable panel data models assume that the structural function does not change over time, and require the existence of "stayers", namely individuals with the same regressor values in two time periods. By contrast, our approach allows g_t to depend on the time period t in an arbitrary manner and does not require the existence of stayers.

Although modeling time trends is important in research on panel data, existing nonseparable panel data models assume that the structural function does not change over time or impose some restrictions on these time trends. For instance, Altonji and Matzkin (2005) do not allow g_t to depend on time period t; Evdokimov (2010) and Hoderlein and White (2012) use additive time effects; and Chernozhukov et al. (2013) and Chernozhukov et al. (2015) use additive location time effects and multiplicative-scale time effects. Moreover, Chernozhukov et al. (2015) assume that $g_t(x, u)$ can be written as $\mu_t(x) + \sigma_t(x)\phi(x, u)$. Thus, time effects are linearly conditional on explanatory variables in this model, and as such it does not allow for nonlinear time effects. Indeed, D'Haultfoeuille et al. (2013) allow for nonlinear time effects by assuming that $g_t(x, u)$ can be written as $m_t(h(x, u))$, where m_t is a monotonic transformation. While this transformation extends the typical additive location time effects and captures macro-shocks, it does not allow the effect of macro-shocks to depend directly on an unobserved variable, and stipulates that $\nabla_x g_t(x, u)/\nabla_u g_t(x, u)$ does not depend on time. For example, consider the relationship between consumption and income. We write the Engel function of the i-th household as

$$Y_{it} = \phi(X_{it}, M_t, U_{it}),$$

where Y_{it} is consumption, X_{it} is income, U_{it} is the scalar unobserved heterogeneity that represents preference, and M_t is a macroeconomic variable. However, such a model does not satisfy D'Haultfoeuille et al. (2013) since $\nabla_x \phi(x, M_t, u)/\nabla_u \phi(x, M_t, u)$ depends on M_t. By contrast, our assumptions can accommodate this model, because we can rewrite this as (3.1) by treating $\phi(x, M_t, u)$ as $g_t(x, u)$.

Many nonseparable panel data models require the existence of stayers: Evdokimov (2010), Hoderlein and White (2012), and Chernozhukov et al. (2015) require stayers in order to identify the structural functions or derivatives of the average and quantile structural functions. In particular, to identify the structural function, Evdokimov (2010) requires the support of (X_{i1}, X_{i2}) contains (x, x) for all x. Many empirically important models do not satisfy this assumption. For example, in standard difference-in-differences (DID) models, there are no individuals treated during both time periods. Our approach does not require the existence of stayers and allows the support condition employed in standard DID models.

Our identification approach is based on the conditional stationary condition, that is, the conditional distribution function of U_{it} given $\mathbf{X}_i$ does not change over time. Similar assumptions are employed in all the aforementioned papers except for Altonji and Matzkin (2005). Indeed, Manski (1987), Abrevaya (1999), Athey and Imbens (2006), Hoderlein and White (2012), Graham and Powell (2012), Chernozhukov et al. (2013), and Chernozhukov et al. (2015) essentially make the same assumption. Whereas Evdokimov (2010) does not impose this assumption explicitly, a similar assumption is made by considering the following model: $Y_{it} = m(X_{it}, \alpha_i) + U_{it}$. In this model, the unobservable variable α_i automatically satisfies the conditional stationarity because α_i does not depend on t. By contrast, D'Haultfoeuille et al. (2013) do not assume the conditional stationarity of U_{it} given $\mathbf{X}_i$ because they consider the identification of nonseparable models using repeated cross-sections. Rather, they assume that the conditional distribution of U_{it} given $V_{it} \equiv F_{X_t}(X_{it})$ does not depend on time.

In the literature on nonseparable panel data models, many papers allow the unobservable variable to be a vector or do not impose monotonicity on the structural function, for example, Altonji and Matzkin (2005), Evdokimov (2010), Hoderlein and White (2012), Chernozhukov et al. (2013), and Chernozhukov et al. (2015). On the other hand, our model assumes that the unobservable variable is scalar, and that the structural function is strictly increasing in the unobservable variable. These assumptions are restrictive but crucial for our identification results.

In estimation part of the paper, we propose parametric and nonparametric estimation methods. We develop estimators based on the conditional stationary condition. Our parametric method is similar to that of Torgovitsky (2017). This estimator is obtained by minimizing the distance between the conditional distributions of U_{i1} and U_{i2}. We then prove consistency and asymptotic normality of this estimator. Because the asymptotic variance is complicated, we also show the validity of the nonparametric bootstrap. Our nonparametric estimator is obtained by using a kernel function, and we show consistency of this nonparametric estimator. Monte Carlo studies indicate that our parametric estimator performs well in finite samples.

Finally, we extend our identification results to models in which outcomes are discrete. This class of models includes many empirically important models such as binary choice panel data models. Models in this class cannot point-identify g_t, but can partially identify it by using the suggestion developed in Chesher (2010). We also allows g_t to depend on the time period t in an arbitrary manner and does not require the existence of stayers. However, the support condition becomes stronger than it is in models with continuous outcomes.

The remainder of the paper is organized as follows. Section 3.2 demonstrates the nonparametric identification of g_t when outcome variables are continuous. In Section 3.3, we propose the estimator under the parametric assumption and discuss its consistency, asymptotic normality, and bootstrap. Section 3.4 reports the results of several Monte Carlo simulations. In Section 3.5, we consider the case where Y_{it} is discrete and show the partial identification of g_t. Section 3.6 offers concluding remarks. The proofs of the theorems and auxiliary lemmas are collected in the Appendix.

3.2 Identification

First, for notational convenience we drop the subscript i and let $T = 2$. It is straightforward to extend the results to the case with $T \geq 3$. For any random variables V and W, let $F_{V|W}$ and $Q_{V|W}$ denote the conditional distribution function and the conditional quantile function, respectively. $\mathcal{X}_t$, $\mathcal{X}_{12}$, and $\mathcal{U}_t$ denote, respectively, the supports of X_t, (X_1, X_2), and U_t.

First, we assume $g_t(x, u)$ is strictly increasing in u, and U_t is continuously distributed.

Assumption 3.I1. *(i) For all t, the function $g_t(x, u)$ is continuous and strictly increasing in u for all x. If X_t is continuously distributed, then $g_t(x, u)$ is also continuous in x. (ii) For all t, $U_t|\mathbf{X} = \mathbf{x}$ is continuously distributed for all $\mathbf{x}$.*

Assumption 3.I2. *For all t and $\mathbf{x} \in \mathcal{X}_{12}$, the conditional distribution of the Y_t conditional on $\mathbf{X} = \mathbf{x}$ is continuous and strictly increasing.*

Many nonseparable panel data models do not employ the strict monotonicity assumption, for example, Altonji and Matzkin (2005), Hoderlein and White (2012), Chernozhukov et al. (2013), and Chernozhukov et al. (2015). These models allow for unobserved variables to be multivariate. Hence, our model is more restrictive than theirs. However, as noted in the previous section, our model covers many widely used panel data models, such as typical linear fixed-effects models.

Assumptions 3.I1 and 3.I2 rule out the case where outcomes are discrete. In Section 3.5, we relax the strict monotonicity assumption by allowing $g_t(x, u)$ to be flat inside the support of U_t, and consider the case where outcomes are discrete.

Next, we impose the normalization assumption.

Assumption 3.I3. *For some $\bar{x} \in \mathcal{X}_1$, we have $g_1(\bar{x}, u) = u$ for all u.*

Assumption 3.I3 is a normalization assumption common in nonseparable models (see, e.g., Matzkin (2003)). Because we assume $U_1|\mathbf{X} = \mathbf{x} \overset{d}{=} U_2|\mathbf{X} = \mathbf{x}$ below, it is sufficient to normalize $g_1(x, u)$ exclusively. The functions $g_t(x, u)$ and distributions of U_t depend on the choice of $\bar{x}$. However, we can construct an alternative structural function that does not depend on the choice of $\bar{x}$ as the following:

$$h_t(x, \tau) \equiv g_t(x, Q_{U_t}(\tau)).$$

Nevertheless, we can normalize this model in an alternative way.

Assumption 3.I3'. *For all t, the marginal distribution of U_t is uniform on* $[0, 1]$.

Under this normalization and additional assumptions, we can regard this structural function as the quantile function of the potential outcome considered by Chernozhukov and Hansen (2005). They refer to U_t as the rank variable. As they show, under the rank invariance or rank similarity assumption, we can think of the function $g_t(x, u)$ as the quantile function of the potential outcome. It is easy to show that the function $g_t(x, \tau)$ under Assumption 3.I3' is the same as the function $h_t(x, \tau)$ under Assumption 3.I3.

Hereafter, we use Assumption 3.I3, but we can replace Assumption 3.I3 with Assumption 3.I3' and identify the structural function g_t, as we show below.

We assume the conditional stationarity of U_t by following Manski (1987), Abrevaya (1999), Athey and Imbens (2006), Hoderlein and White (2012), Graham and Powell (2012), Chernozhukov et al. (2013), and Chernozhukov et al. (2015).

Assumption 3.I4. *(i) The conditional distributions of the unobservable U_t conditional on $\mathbf{X}$ is the same across t. That is, for all $\mathbf{x} \in \mathcal{X}_{12}$, we have*

$$U_1|\mathbf{X} = \mathbf{x} \overset{d}{=} U_2|\mathbf{X} = \mathbf{x}, \tag{3.4}$$

which implies that $\mathcal{U}_1 = \mathcal{U}_2 \equiv \mathcal{U}$. (ii) For all t, the conditional support of $U_t|\mathbf{X} = \mathbf{x}$ is $\mathcal{U}$.

When we can decompose U_t into time-variant and time-invariant parts, this assumption does not impose any restrictions on the dependence between the time-invariant part and $\mathbf{X}$. To see this, let $U_t = U(\alpha, \epsilon_t)$, where α is time-invariant and ϵ_t is time-variant. Then, Assumption 3.I4 holds, if

$$\epsilon_1|\alpha = a, \mathbf{X} = \mathbf{x} \overset{d}{=} \epsilon_2|\alpha = a, \mathbf{X} = \mathbf{x}. \tag{3.5}$$

Because condition (3.5) allows α to be correlated with $\mathbf{X}$ arbitrarily, Assumption 3.I4 imposes no restrictions on the time-invariant unobservable variables.

Indeed, Evdokimov (2010) and D'Haultfoeuille et al. (2013) employ similar assumptions, although the former does not make this assumption explicitly. By considering the model $Y_{it} = m(X_{it}, \alpha_i) + U_{it}$, the unobservable variable α_i automatically satisfies the conditional stationarity. Moreover, since D'Haultfoeuille et al. (2013) consider the identification using repeated cross-sections, they do not impose this assumption. Instead, they impose the following:

$$U_1|\mathbf{V}_1 = \mathbf{v} \overset{d}{=} U_2|\mathbf{V}_2 = \mathbf{v},$$

where $\mathbf{V}_t \equiv (F_{X_{t,1}}(X_{t,1}), \cdots, F_{X_{t,k}}(X_{t,k}))$ and $\mathbf{v} \in (0,1)^k$.

To show the identification of g_t, we introduce the following sets. Define $\mathcal{S}_0^1 \equiv \{\bar{x}\}$, $\mathcal{S}_0^2 \equiv \{x \in \mathcal{X}_2 : (\bar{x}, x) \in \mathcal{X}_{12}\}$, namely the cross-section of $\mathcal{X}_{12}$ at $X_1 = \bar{x}$. For $m \geq 1$, define

$$\begin{aligned} \mathcal{S}_m^1 &\equiv \{x \in \mathcal{X}_1 : \text{there exists } x_2 \in \mathcal{S}_{m-1}^2 \text{ such that } (x, x_2) \in \mathcal{X}_{12}.\}, \\ \mathcal{S}_m^2 &\equiv \{x \in \mathcal{X}_2 : \text{there exists } x_1 \in \mathcal{S}_m^1 \text{ such that } (x_1, x) \in \mathcal{X}_{12}.\}. \end{aligned}$$

Figure 3.1 illustrates these sets. Because $U_1|\mathbf{X} = \mathbf{x} \overset{d}{=} U_2|\mathbf{X} = \mathbf{x}$, for all $(x_1, x_2) \in \mathcal{X}_{12}$, we have

$$\begin{aligned} F_{Y_1|\mathbf{X}}(g_1(x_1,u)|x_1,x_2) &= P(g_1(x_1,U_1) \leq g_1(x_1,u)|X_1 = x_1, X_2 = x_2) \\ &= P(U_1 \leq u|X_1 = x_1, X_2 = x_2) \\ &= P(U_2 \leq u|X_1 = x_1, X_2 = x_2) \\ &= P(g_2(x_2,U_2) \leq g_2(x_2,u)|X_1 = x_1, X_2 = x_2) \\ &= F_{Y_2|\mathbf{X}}(g_2(x_2,u)|x_1,x_2). \end{aligned}$$

Because $F_{Y_t|\mathbf{X}}(y|x_1, x_2)$ is invertible in y for all $(x_1, x_2) \in \mathcal{X}_{12}$, we obtain

$$\begin{aligned} g_1(x_1,u) &= Q_{Y_1|\mathbf{X}}\left(F_{Y_2|\mathbf{X}}(g_2(x_2,u)|x_1,x_2)\,|x_1,x_2\right) \\ g_2(x_2,u) &= Q_{Y_2|\mathbf{X}}\left(F_{Y_1|\mathbf{X}}(g_1(x_1,u)|x_1,x_2)\,|x_1,x_2\right). \end{aligned} \tag{3.6}$$

Equations (3.6) imply that if $g_1(x_1, u)$ (or $g_2(x_2, u)$) is identified and $(x_1, x_2) \in \mathcal{X}_{12}$, then $g_2(x_2, u)$ (or $g_1(x_1, u)$) is also identified. First, we can identify $g_1(\bar{x}, u)$ because $g_1(\bar{x}, u) = u$ holds by Assumption 3.I3. Hence, we can identify $g_2(x, u)$ for all $x \in \mathcal{S}_0^2$, because $g_2(x,u) = Q_{Y_2|\mathbf{X}}\left(F_{Y_1|\mathbf{X}}(g_1(\bar{x},u)|\bar{x},x)\,|\bar{x},x\right) = Q_{Y_2|\mathbf{X}}\left(F_{Y_1|\mathbf{X}}(u|\bar{x},x)\,|\bar{x},x\right)$. We now turn to identifying $g_1(x, u)$ for $x \in \mathcal{S}_1^1$.

First, we fix $x \in \mathcal{S}_1^1$. According to the definition of $\mathcal{S}_1^1$, there exists $x_2 \in \mathcal{S}_0^2$ such that $(x, x_2) \in \mathcal{X}_{12}$. Then, it follows from (3.6) that

$$g_1(x,u) = Q_{Y_1|\mathbf{X}}\left(F_{Y_2|\mathbf{X}}(g_2(x_2,u)|x,x_2)\,|x,x_2\right),$$

and hence, $g_1(x,u)$ is identified because $g_2(x_2,u)$ is already identified. Similarly, by using (3.6), we can identify $g_2(x,u)$ for all $x \in \mathcal{S}_1^2$. Repeating this argument provides the following theorem.

Theorem 3.1. *Suppose that Assumptions 3.I1, 3.I2, 3.I3, and 3.I4 are satisfied. For all t, if we have $\mathcal{X}_t = \overline{\cup_{m=0}^{\infty}\mathcal{S}_m^t}$, then the structural function $g_t(x,u)$ is identified for all $x \in \mathcal{X}_t$ and $u \in \mathcal{U}$.*

We also show the identification of g_t under Assumption 3.I3' instead of 3.I3.

Corollary 3.1. *Suppose that Assumptions 3.I1, 3.I2, 3.I3', and 3.I4 are satisfied. For all t, if $\mathcal{X}_t = \overline{\cup_{m=0}^{\infty}\mathcal{S}_m^t}$ holds for some $\overline{x} \in \mathcal{X}_1$, then the function $g_t(x,u)$ is identified for all $x \in \mathcal{X}_t$ and $u \in \mathcal{U}$.*

This identification approach is similar to that of D'Haultfœuille and Février (2015), Torgovitsky (2015), and Ishihara (2017), who all identify nonseparable models using the discrete instrumental variable. D'Haultfœuille and Février (2015) and Ishihara (2017) use the same normalization as Assumption 3.I3'. D'Haultfœuille and Février (2015) show that under appropriate assumptions, if for all x and x', we identify the function $T_{x',x}(y)$ that is strictly increasing in y and satisfies

$$g(x',u) = T_{x',x}(g(x,u)),$$

then we can identify the structural function $g(x,u)$. We can also construct similar functions and show that g_t is point identified.

We next introduce some examples that satisfy this support condition.

Example 3.1 (DID model)**.** *In standard DID models, if X_t is a treatment indicator, then we have $\mathcal{X}_{12} = \{(0,0),(0,1)\}$. Because $\mathcal{X}_1 = \{0\}$, we assume $\bar{x} = 0$. That is, $g_1(0,u) = u$ for all u. Hence, we identify $g_1(x,u)$ for all $x \in \mathcal{X}_1$ and $u \in \mathcal{U}$. Then, because $\mathcal{S}_0^2 = \{0,1\} = \mathcal{X}_{12}$, the support condition of Theorem 3.1 holds and we can identify $g_2(x,u)$ for all $x \in \mathcal{X}_2$ and $u \in \mathcal{U}$.*

Our identification approach does not require the joint distribution of (Y_1,Y_2). Hence, if we can observe $D \equiv \mathbf{1}\{X_2 = 1\}$, then we can identify the structural function g_t by using repeated cross-sections. If the potential outcome $Y_t(x)$ is equal to $g_t(x,U_t)$, then this setting is similar to Athey and Imbens (2006).

Similar to Athey and Imbens (2006), we can also identify the counterfactual distribution even when $\mathcal{X}_{12} \neq \{(0,0),(0,1)\}$. Let $Y_t(x) \equiv g_t(x,U_t)$ denote the potential outcomes. Then, we can identify $F_{Y_2(x)|X_2}(y|x')$, where $x \neq x'$. Suppose that there exists $x_1 \in \mathcal{X}_1$

such that $(x_1, x), (x_1, x') \in \mathcal{X}_{12}$. In this case, it follows from (3.6) that

$$\begin{aligned} & F_{Y_1|\mathbf{X}=(x_1,x')}\left(Q_{Y_1|\mathbf{X}=(x_1,x)}(F_{Y_2|\mathbf{X}=(x_1,x)}(y))\right) \\ = \; & F_{Y_1|\mathbf{X}}\left(g_1(x_1, g_2^{-1}(x,y))|x_1, x'\right) \\ = \; & P\left(g_1(x_1, U_1) \le g_1(x_1, g_2^{-1}(x,y))|X_1 = x_1, X_2 = x'\right) \\ = \; & P\left(g_2(x, U_1) \le y|X_1 = x_1, X_2 = x'\right) \\ = \; & P\left(Y_2(x) \le y|X_1 = x_1, X_2 = x'\right). \end{aligned}$$

Hence, we can obtain $F_{Y_2(x)|X_2}(y|x')$ by integrating out x_1. The left-hand side is similar to the counterfactual distribution of Athey and Imbens (2006). When $\mathcal{X}_{12} = \{(0,0),(0,1)\}$, this result is same as their result.

Example 3.2 (connected support)**.** *When the interior of $\mathcal{X}_{12}$ is connected, the support condition of Theorem 3.1 holds. Because the interior of $\mathcal{X}_{12}$ is connected, for any $x \in \mathcal{X}_1$, there exists a series $(x_1^0, x_2^0), (x_1^1, x_2^0), (x_1^1, x_2^1), (x_1^2, x_2^1), (x_1^2, x_2^2) \cdots$ such that $x_1^0 = \bar{x}$, $(x_1^m, x_2^m), (x_1^{m+1}, x_2^m) \in \mathcal{X}_{12}$ for all $m = 0, 1, \cdots$, and $\lim_{m \to \infty} x_1^m = x$. Figure 3.2 illustrates this result intuitively. From the definition of $\mathcal{S}_m^1$, $x_1^m \in \mathcal{S}_1^m$ for all m. Hence, we have $x \in \overline{\cup_{m=0}^{\infty} \mathcal{S}_m^1}$, and the support condition of Theorem 3.1 holds.*

The support condition of Theorem 3.1 rules out the case where $X_1 = X_2$. Hence, if the explanatory variables do not vary across time periods, such as sex or race, this support condition does not hold.

If we have panel data with more than two periods, we can relax this support condition. Similar to the case where $T = 2$, we define the following sets. Define $\tilde{\mathcal{S}}_0^1 \equiv \{\bar{x}\}$, $\tilde{\mathcal{S}}_0^t \equiv \{x \in \mathcal{X}_t : (\bar{x}, x) \in supp(X_1, X_t)\}$, $t = 2, \cdots, T$, and for $m \ge 1$ and $t = 1, \cdots, T$,

$$\tilde{\mathcal{S}}_m^t \equiv \bigcup_{s \ne t} \{x \in \mathcal{X}_t : \text{there exists } x_s \in \tilde{\mathcal{S}}_{m-1}^s \text{ such that } (x_s, x) \in supp(X_s, X_t).\}.$$

Then, we show that $g_t(x, u)$ is point-identified under a similar support condition to that of Theorem 3.1.

Corollary 3.2. *Suppose Assumptions 3.I1, 3.I2, 3.I3, and 3.I4 are satisfied for $T \ge 3$. For $t = 1, \cdots, T$, if $\mathcal{X}_t = \overline{\cup_{m=0}^{\infty} \tilde{\mathcal{S}}_m^t}$, then the function $g_t(x, u)$ is identified for all $x \in \mathcal{X}_t$ and $u \in \mathcal{U}$.*

3.3 Estimation and Inference

In this section, we propose parametric and nonparametric estimation methods. In Sections 3.3.1–3.3.4, we assume that the admissible collection of structural functions is indexed by a finite-dimensional parameter, and propose a parametric estimation method based on the conditional stationary condition. In Section 3.3.5, we propose a nonparametric estimation method and show the consistency of this estimator. Throughout Section 3.3, we assume that $\{(\mathbf{Y}_i, \mathbf{X}_i)\}_{i=1}^n$ are independent and identically distributed.

3.3.1 Parametric Estimation

Consider the following parametric model:

$$Y_{it} = g_t(X_{it}, U_{it}; \theta_0) \qquad i = 1, \cdots, n, \quad t = 1, \cdots, T. \tag{3.7}$$

The outcome functions are parameterized by $\theta \in \Theta \subset \mathbb{R}^{d_\theta}$, where $\theta_0 \in \Theta$ is the true parameter.

Indeed, Torgovitsky (2017) consider a similar setting, and develop an estimator based on the identification result of Torgovitsky (2015). Following Torgovitsky (2017), we develop a minimum distance estimator based on our identification results.

The following assumptions are the parametric versions of Assumptions 3.I1, 3.I2, 3.I3, and 3.I4.

Assumption 3.P1. *(i) For all t, the function $g_t(x, u; \theta)$ is continuous and strictly increasing in u for all $\theta \in \Theta$. (ii) For all t, $U_{it}|\mathbf{X}_i = \mathbf{x}$ is continuously distributed for all* $\mathbf{x}$.

Assumption 3.P2. *For all t and $\mathbf{x} \in supp(\mathbf{X})$, the conditional distribution of Y_{it} conditional on $\mathbf{X}_i = \mathbf{x}$ is continuous and strictly increasing.*

Assumption 3.P3. *(i) For some $\bar{x} \in \mathcal{X}_1$, $g_1(\bar{x}, u; \theta) = u$ holds for all $u \in \mathcal{U}$ and $\theta \in \Theta$. (ii) For all $\theta, \theta' \in \Theta$ with $\theta \neq \theta'$, we have $g_t(\cdot, \cdot; \theta) \neq g_t(\cdot, \cdot; \theta')$ for some t.*

Assumption 3.P4. *(i) For all $\mathbf{x} \in supp(\mathbf{X})$ and $s, t \in \{1, \cdots, T\}$, we have $U_{is}|\mathbf{X}_i = \mathbf{x} \overset{d}{=} U_{it}|\mathbf{X}_i = \mathbf{x}$. (ii) The support of $U_{it}|\mathbf{X}_i = \mathbf{x}$ is $\mathcal{U}$.*

These assumptions are similar to the assumptions in Section 3.2. Assumption 3.P3 (ii) allows that $g_t(x, u; \theta)$ does not depend on some part of θ. For example, consider $\theta = (\theta_1, \theta_2, \cdots, \theta_T)$. Then, this condition allows that g_t depends exclusively on θ_t.

Similar to Section 3.2, we suppose $T = 2$. Under Assumptions 3.P1–3.P4 and the support condition of Theorem 3.1, we have

$$U_{i1,\theta}|\mathbf{X}_i = \mathbf{x} \overset{d}{=} U_{i2,\theta}|\mathbf{X}_i = \mathbf{x} \text{ for all } \mathbf{x} \quad \Leftrightarrow \quad \theta = \theta_0, \tag{3.8}$$

where $U_{it,\theta} \equiv g_t^{-1}(X_{it}, Y_{it}; \theta)$, $t = 1, 2$. Therefore, (3.8) implies that the function

$$\begin{aligned} D_\theta(v) &\equiv P\left(U_{i1,\theta} \leq v_u,\ \mathbf{X}_i \leq v_\mathbf{x}\right) - P\left(U_{i2,\theta} \leq v_u,\ \mathbf{X}_i \leq v_\mathbf{x}\right) \\ &= E\left[\left(\mathbf{1}\{Y_{i1} \leq g_1(X_{i1}, v_u; \theta)\} - \mathbf{1}\{Y_{i2} \leq g_2(X_{i2}, v_u; \theta)\}\right) \mathbf{1}\{\mathbf{X}_i \leq v_\mathbf{x}\}\right] \end{aligned} \tag{3.9}$$

is zero for all $v = (v_{\mathbf{x}}, v_u) \in \mathcal{V} \equiv \mathcal{X}_{12} \times \mathcal{U}$ if and only if $\theta = \theta_0$. Let $\|\cdot\|_\mu$ denote the L_2-norm with respect to a probability measure with support $\mathcal{V}$. Then, we have $\|D_\theta\|_\mu \geq 0$ and

$$\|D_\theta\|_\mu = 0 \quad \Leftrightarrow \quad \theta = \theta_0. \tag{3.10}$$

Hence, θ_0 is the value that provides a global minimum for $\|D_\theta\|_\mu$.

We construct the estimator $\hat{D}_{n,\theta}(v)$ of $D_\theta(v)$ as a sample analogue of (3.9):

$$\hat{D}_{n,\theta}(v) \equiv \frac{1}{n}\sum_{i=1}^{n} \left(\mathbf{1}\{Y_{i1} \leq g_1(X_{i1}, v_u; \theta)\} - \mathbf{1}\{Y_{i2} \leq g_2(X_{i2}, v_u; \theta)\}\right) \mathbf{1}\{\mathbf{X}_i \leq v_{\mathbf{x}}\}. \tag{3.11}$$

This is a natural estimator of $D_\theta(v)$. We can obtain the estimator $\hat{\theta}_n$ by minimizing $\|\hat{D}_{n,\theta}\|_\mu$. That is,

$$\hat{\theta}_n = \arg\min_{\theta \in \Theta} \|\hat{D}_{n,\theta}\|_\mu. \tag{3.12}$$

This estimator is similar to the estimators proposed by Brown and Wegkamp (2002) and Torgovitsky (2017). They prove the consistency and the asymptotic normality of their estimators, and also show the consistency of the nonparametric bootstrap. In what follow, we likewise prove the consistency and the asymptotic normality of our estimator, and show that the nonparametric bootstrap is consistent.

First, we collect the observable data together into a single vector, $W_i = (\mathbf{Y}_i, \mathbf{X}_i) = (Y_{i1}, Y_{i2}, X_{i1}, X_{i2})$. Next, we define

$$A_\theta^v(w) \equiv \left[\mathbf{1}\{y_1 \leq g_1(x_1, v_u; \theta)\} - \mathbf{1}\{y_2 \leq g_2(x_2, v_u; \theta)\}\right] \mathbf{1}\{\mathbf{x} \leq v_{\mathbf{x}}\},$$

where $w = (y_1, y_2, x_1, x_2)$. Then, $\hat{D}_{n,\theta}(v) = \frac{1}{n}\sum_{i=1}^n A_\theta^v(W_i)$.

3.3.2 Consistency and Asymptotic Normality

First, we demonstrate the consistency of $\hat{\theta}_n$. Under condition (3.10), the following assumptions are sufficient for $\hat{\theta}_n$ to be consistent.

Assumption 3.C1. *$\hat{\theta}_n$ satisfies $\|\hat{D}_{n,\hat{\theta}_n}\|_\mu = \inf_{\theta \in \Theta} \|\hat{D}_{n,\theta}\|_\mu$.*

Assumption 3.C2. *Θ is compact.*

Assumption 3.C3. *For all θ', θ, $|g_t(x, u; \theta') - g_t(x, u; \theta)| \leq \bar{g}(x)\|\theta' - \theta\|$ holds for some strictly positive $\bar{g}(x)$ with $E[\bar{g}(X_t)] \leq K$, where $0 < K < \infty$.*

Assumption 3.C4. *For all t, Y_{it} is absolutely continuously distributed given $\mathbf{X}_i$, with a conditional pdf $f_{Y_t|\mathbf{X}}(y|\mathbf{x})$ that is uniformly bounded above and continuous in y.*

Assumption 3.C5. *For all t, there exists an integer J_t and functions $\{\beta_j\}_j^{J_t}$ such that for every $\theta \in \Theta$ and $u \in \mathcal{U}$ there is an $\alpha^t(\theta, u) \in \mathbb{R}^{J_t}$ with $g_t(x, u) = \sum_{j=1}^{J_t} \alpha_j^t(\theta, u)\beta_j(x)$.*

Assumption 3.C1 entails that $\hat{\theta}_n$ minimizes $\|\hat{D}_{n,\theta}\|_\mu$. Assumptions 3.C3 and 3.C4 imply that $\|D_\theta\|_\mu$ is continuous in θ. Assumption 3.C5 ensures that a class of functions, $\{A_\theta^v : \theta \in \Theta, v \in \mathcal{V}\}$, is P-Glivenko–Cantelli. Hence, we show that $\|\hat{D}_{n,\theta}\|_\mu$ uniformly convergences to $\|D_\theta\|_\mu$ almost surely. Brown and Wegkamp (2002) and Torgovitsky (2017) also make similar assumptions. From these results and the compactness of Θ, we show the consistency from the usual arguments of extremum estimators (e.g., Newey and McFadden (1994)).

Theorem 3.2. *Under Assumptions 3.P1–3.P4, 3.C1–3.C5, and (3.10), we have* $\hat{\theta}_n \to_{a.s.} \theta_0$.

Next, we show the asymptotic normality of $\hat{\theta}_n$. Because the objective function $\|\hat{D}_{n,\theta}\|_\mu$ is not differentiable in θ, our approach follows from Pakes and Pollard (1989). Similarly, although $\hat{D}_{n,\theta}(v)$ is not differentiable in θ, we also assume $D_\theta(v)$ is differentiable in θ. We let $\nabla_x f$ denote the column vector of partial derivatives of f with respect to x. We define $\Gamma_\theta(v) \equiv \nabla_\theta D_\theta(v)$ and $\Gamma_0(v) \equiv \Gamma_{\theta_0}(v)$.

Assumption 3.N1. θ_0 *is an interior point of* Θ.

Assumption 3.N2. *For all* t, $g_t(x, u; \theta)$ *is continuously differentiable in* θ *in the neighborhood of* θ_0. *In the neighborhood of* θ_0, $|\nabla_\theta g_t(x, u; \theta)|$ *is bounded by some positive function* $\nabla \bar{g}(x)$ *with* $E|\nabla \bar{g}(X_t)| < \infty$.

Assumption 3.N3. *(i) There exists* $c > 0$ *such that* $\|\Gamma_0(v)'a\|_\mu \geq c\|a\|$ *for all* $a \in \mathbb{R}^{d_\theta}$. *(ii)* $\{\Gamma_\theta(v) : v \in \mathcal{V}\}$ *is equicontinuous in* θ *at* θ_0. *(iii)* $\int \|\Gamma_0(v)\|^2 d\mu(v) < \infty$.

Assumption 3.N1 is a standard assumption. Combined with Assumption 3.C4, Assumption 3.N2 implies that $D_\theta(v)$ is continuously differentiable in θ in the neighborhood of θ_0. Assumption 3.N3 (i) is a rank condition that corresponds to Assumption D4 in Torgovitsky (2017). Assumption 3.N3 (ii) implies that $\Gamma_0(v)'(\theta - \theta_0)$ approximates $D_\theta(v)$ in the neighborhood of θ_0 uniformly over v.

Theorem 3.3. *Under Assumptions 3.P1–3.P4, 3.C1–3.C5, 3.N1–3.N3, and (3.10),*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \rightsquigarrow N(0, \Delta_0^{-1}\Sigma_0\Delta_0^{-1}),$$

where $\Delta_0 \equiv \int \Gamma_0(v)\Gamma_0(v)'d\mu(v)$ *and*

$$\Sigma_0 \equiv \int\int_{\mathcal{V}\times\mathcal{V}} \{\Psi(v, v')\Gamma_0(v)\Gamma_0(v')'\} \, d\mu(v)d\mu(v')$$

with $\Psi(v, v') \equiv E[A_{\theta_0}^v(W)A_{\theta_0}^{v'}(W)]$.

The proof of Theorem 3.3 is similar to the proof in Pakes and Pollard (1989) for their Theorem (3.3).

The asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ depends on the probability measure μ. Carrasco and Florens (2000) consider the generalized method of moments (GMM) procedure with a continuum of moment conditions, obtaining the optimal estimator. They consider the following type of GMM estimator to minimize

$$\int\int \hat{D}_{n,\theta}(v) a_n(v,v') \hat{D}_{n,\theta}(v') dv dv',$$

where $a_n(v,v')$ converges to a kernel $a(v,v')$. As in Torgovitsky (2017), we consider only the special case where $a_n(v,v') = a(v,v')$ and $a(v,v') = 0$ for $v \neq v'$. Although our setting appears to be similar to that of Carrasco and Florens (2000), their approach is not directly applicable because their objective function is smooth. Hence, we do not pursue this problem.

3.3.3 Bootstrap

Let $\{W_{in}^*\}_{i=1}^n$ denote a bootstrap sample drawn with replacement from $\{W_i\}_{i=1}^n$. That is, $\{W_{in}^*\}_{i=1}^n$ are independently and identically distributed from the empirical measure P_n, conditional on the realizations $\{W_i\}_{i=1}^n$. We define

$$\hat{D}_{n,\theta}^*(v) \equiv \frac{1}{n}\sum_{i=1}^n A_\theta^v(W_{in}^*)$$

as the bootstrap counterpart to $\hat{D}_{n,\theta}(v)$. Next, we suppose that $\hat{\theta}_n^*$ satisfies

$$\|\hat{D}_{n,\hat{\theta}_n^*}^*\|_\mu = \inf_{\theta\in\Theta} \|\hat{D}_{n,\theta}^*\|. \tag{3.13}$$

Then, we can obtain the following theorem.

Theorem 3.4. *Suppose that $\hat{\theta}_n^*$ satisfies (3.13). Under the assumptions of Theorem 3.3, $\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)$ converges weakly to the limit distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ in probability.*

The proof of this theorem is similar to the proof of Theorem 6 in Brown and Wegkamp (2002). From Theorem 3.3, we show that

$$\hat{\theta}_n - \theta_0 \;\; = \;\; \gamma_n + o_p(n^{-1/2}),$$

where $\gamma_n = \Delta_0^{-1}\frac{1}{n}\sum_{i=1}^n \int \Gamma_0(v)(A_{\theta_0}^v(W_i) - E[A_{\theta_0}^v(W_i)])d\mu(v)$. By using the bootstrap stochastic equicontinuity due to Giné and Zinn (1990), we show that

$$\sqrt{n}\|\hat{\theta}_n^* - \hat{\theta}_n - \gamma_n^*\|$$

converges to zero in probability, conditional on almost all samples, where γ_n^* is the bootstrap counterpart of γ_n. The term γ_n^* has the same limiting distribution as γ_n according to the bootstrap theorem for the mean in $\mathbb{R}^{d_\theta}$. Hence, we show that $\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)$ converges weakly to the limit distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ in probability.

3.3.4 Nonparametric Estimation

In this section, we propose a nonparametric estimation method that uses a kernel function. We assume that $\mathcal{X}_1$ and $\mathcal{X}_2$ are compact and $\mathcal{X}_{12} = \mathcal{X}_1 \times \mathcal{X}_2$. This implies that the support condition of Theorem 1 is satisfied. Then, under Assumptions I3 and I4, for all $x \in \mathcal{X}_2$, $g_2(x, u)$ satisfies

$$E[\{\mathbf{1}\{Y_{i1} \le u\} - \mathbf{1}\{Y_{i2} \le g_2(x, u)\}\}|X_{i1} = \bar{x}, X_{i2} = x] = 0.$$

Hence, we can obtain an estimator of $g_2(x, u)$ by

$$\hat{g}_{n,2}(x, u) = \arg\min_{\xi} \left| \frac{1}{n} \sum_{i=1}^{n} \{\mathbf{1}\{Y_{i1} \le u\} - \mathbf{1}\{Y_{i2} \le \xi\}\} K\left(\frac{X_{i1} - \bar{x}}{h_n}\right) K\left(\frac{X_{i2} - x}{h_n}\right) \right|,$$

where $K(\cdot)$ is a kernel function and h_n is a bandwidth. Then, this estimator can be written as

$$\hat{g}_{n,2}(x, u) = \arg\min_{\xi} \left| \hat{F}_{Y_1|\mathbf{X}}(u|\bar{x}, x) - \hat{F}_{Y_2|\mathbf{X}}(\xi|\bar{x}, x) \right|, \tag{3.14}$$

where

$$\hat{F}_{Y_t|\mathbf{X}}(y|x_1, x_2) \equiv \frac{\sum_{i=1}^{n} \mathbf{1}\{Y_{it} \le y\} K\left(\frac{X_{i1} - x_1}{h_n}\right) K\left(\frac{X_{i2} - x_2}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{X_{i1} - x_1}{h_n}\right) K\left(\frac{X_{i2} - x_2}{h_n}\right)}.$$

From Assumption I4, for all $x_1 \in \mathcal{X}_1$ and $x_2 \in \mathcal{X}_2$, we have

$$E[\{\mathbf{1}\{Y_{i1} \le g_1(x_1, u)\} - \mathbf{1}\{Y_{i2} \le g_2(x_2, u)\}\}|X_{i1} = x_1, X_{i2} = x_2] = 0.$$

Hence, similarly, we can obtain an estimator of $g_1(x, u)$ by

$$\hat{g}_{n,1}(x, u) = \arg\min_{\xi} \int \left| \hat{F}_{Y_1|\mathbf{X}}(\xi|x, x_2) - \hat{F}_{Y_2|\mathbf{X}}(\hat{g}_{n,2}(x_2, u)|x, x_2) \right|^2 d\mu_{X_2}(x_2), \tag{3.15}$$

where μ_{X_2} is a probability measure with support $\mathcal{X}_2$.

Let $\overline{\mathcal{U}}$ be a subset of $\mathcal{U}$. We impose the following assumptions.

Assumption 3.NP1. *Let $u \in \overline{\mathcal{U}}$. For any $\epsilon > 0$, there exists $\delta > 0$ such that $|g_t(x_t, u) - y| > \epsilon$ implies $\inf_{x_1, x_2} \left| F_{Y_t|\mathbf{X}}(g_t(x_t, u)|x_1, x_2) - F_{Y_t|\mathbf{X}}(y|x_1, x_2) \right| > \delta$ for all t.*

Assumption 3.NP2. *For all $u \in \overline{\mathcal{U}}$, $\{F_{Y_t|\mathbf{X}}(y|\mathbf{x})\}$ is equicontinuous in y at $y = g_t(x_t, u)$.*

Assumption 3.NP3. *We have* $\sup_{y,x_1,x_2}\left|\hat{F}_{Y_t|\mathbf{X}}(y|x_1,x_2) - F_{Y_t|\mathbf{X}}(y|x_1,x_2)\right| = o_p(1)$ *for all* t.

Assumptions 3.NP1 and 3.NP2 imply that for $u \in \overline{\mathcal{U}}$, $F_{Y_t|\mathbf{X}}(y|\mathbf{x})$ is strictly increasing and continuous in y at $y = g_t(x,u)$ uniformly. Assumption 3.NP3 implies that $\hat{F}_{Y_t|\mathbf{X}}(y|x_1,x_2)$ uniformly converges to $F_{Y_t|\mathbf{X}}(y|x_1,x_2)$ in probability.

Theorem 3.5. *We assume that* $\mathcal{X}_1$ *and* $\mathcal{X}_2$ *are compact and* $\mathcal{X}_{12} = \mathcal{X}_1 \times \mathcal{X}_2$*. Then, under Assumptions 3.I1–3.I4, 3.NP1–3.NP3, for all* $(x_1,x_2) \in \mathcal{X}_{12}$ *and* $u \in \overline{\mathcal{U}}$*, we have*

$$\hat{g}_{n,1}(x_1,u) \to_p g_1(x_1,u) \quad \textit{and} \quad \hat{g}_{n,2}(x_2,u) \to_p g_2(x_2,u),$$

where $\hat{g}_{n,1}(x_1,u)$ *and* $\hat{g}_{n,2}(x_2,u)$ *are defined in (3.15) and (3.14).*

3.4 Simulations

To evaluate the finite sample performance of our estimator, we conducted two Monte Carlo experiments.

Simulation 3.1. *The outcome equation is given by*

$$\begin{aligned} g_1(x,u) &= u + (\theta_1 + \theta_2 u)(x - \bar{x}), \\ g_2(x,u) &= u + (\theta_1 + \theta_3 u)(x - \bar{x}), \end{aligned}$$

where $\bar{x} = 0$*. Because* $g_1(\bar{x},u) = u$ *for all* x*, Assumption 3.I3 is satisfied. We assume*

$$\begin{aligned} X_t &= 4\Phi(Z_t) \quad t = 1,2, \\ U_t &= \alpha + \epsilon_t \quad t = 1,2, \end{aligned}$$

where $\Phi(\cdot)$ *is the standard normal distribution function and*

$$\begin{aligned} (Z_1, Z_2, \alpha) &\sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1.0 & 0.3 & 0.6 \\ 0.3 & 1.0 & 0.5 \\ 0.6 & 0.5 & 1.0 \end{pmatrix}\right), \\ (\epsilon_1, \epsilon_2) &\sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} (0.3)^2 & 0 \\ 0 & (0.3)^2 \end{pmatrix}\right). \end{aligned}$$

Because the correlations between α *and* $(Z_1,\ Z_2)$ *are not zero,* X_1 *and* X_2 *are correlated with* U_t*. Because* $\epsilon_1|\mathbf{X} = \mathbf{x}, \alpha = a \overset{d}{=} \epsilon_2|\mathbf{X} = \mathbf{x}, \alpha = a$*, the conditional stationarity assumption holds. We used* $\mu = Unif(0,4) \times Unif(0,4) \times N(0,1)$ *as the integrating measure.*

We considered the following two settings: (i) $(\theta_1,\theta_2,\theta_3) = (0.5, 1.0, 0.7)$*, (ii)* $\theta_2 = \theta_3$ *and* $(\theta_1,\theta_2) = (0.5, 1.0)$*. Under Setting (i), we cannot use estimation methods of other* *papers, because their time effects depend on* U_t. *On the other hand, under Setting (ii), there are no time effects. Hence, we can estimate* $E\left[\nabla_x g_t(X_t, U_t)|X_1 = X_2 = x\right]$ *by using the method proposed in Hoderlein and White (2012) because there are stayers. Thus, we estimated* $E\left[\nabla_x g_t(X_t, U_t)|X_1 = X_2 = 2\right]$ *using our method and their method and compared the results of both. Under Setting (ii), we have* $E\left[\nabla_x g_t(X_t, U_t)|X_1 = X_2 = 2\right] = \theta_1 + \theta_2 E[U_t|X_1 = X_2 = 2] = 0.5$. *Hence, we estimated* $E\left[\nabla_x g_t(X_t, U_t)|X_1 = X_2 = 2\right]$ *by using* $\hat{\theta}_1 + \hat{\theta}_2 \hat{E}[U_t|X_1 = X_2 = 2]$.

Table 3.1 contains the results under Setting (i) for sample sizes of 400, 800, *and* 1600. *The number of replications was set to* 1000 *throughout. Table 3.1 shows the bias, standard deviation, and the mean squared error (MSE) of the estimates of* $(\theta_1, \theta_2, \theta_3)$, *highlighting that the standard deviation and MSE decrease as the sample size increases. In some cases, the biases of the estimates do not decrease. However they are relatively small under all settings.*

Table 3.2 contains the results under Setting (ii) for sample sizes of 500 *and* 1000. *Table 3.2 shows that the standard error of our estimator is smaller than that of Hoderlein and White (2012) for all settings. Although the bias of our estimator is larger than their estimator, the MSE of our estimator is smaller.*

Simulation 3.2 (DID model)**.** *We considered the case where* $\mathcal{X}_{12} = \{(0,0), (0,1)\}$. *The outcome equation is given by*

$$\begin{aligned} g_1(x, u) &= u \\ g_2(x, u) &= (\theta_1 + \theta_2 u)(1 - x) + (\theta_3 + \theta_4 u)x, \end{aligned}$$

where $(\theta_1, \theta_2, \theta_3, \theta_4) = (0.5, 0.7, 0.5, 1.2)$. *Because* $g_1(x, u)$ *does not depend on* x, *Assumption 3.I3 holds for any* $\bar{x} \in \mathcal{X}_1$. *We assumed*

$$\begin{aligned} X_2 &= \mathbf{1}\{Z > 0\}, \\ U_t &= \alpha + \epsilon_t \quad t = 1, 2, \end{aligned}$$

where $\Phi(\cdot)$ *is the standard normal distribution function and*

$$\begin{aligned} (Z, \alpha) &\sim N\left(\begin{pmatrix} 0 \\ 2.0 \end{pmatrix}, \begin{pmatrix} 1.0 & 0.6 \\ 0.6 & 1.0 \end{pmatrix}\right), \\ (\epsilon_1, \epsilon_2) &\sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} (0.3)^2 & 0 \\ 0 & (0.3)^2 \end{pmatrix}\right). \end{aligned}$$

Because $\epsilon_1|X_2 = x \overset{d}{=} \epsilon_2|X_2 = x$ *for all* $\mathbf{x}$, *Assumption 3.I4 holds. When* $X_2 = 0$, *we have* $Y_2 = g_2(0, U_2) = \theta_1 + \theta_2 U_2$, *and when* $X_2 = 1$, *we have* $Y_2 = g_2(1, U_2) = \theta_3 + \theta_4 U_2$. *This specification is similar to that of a typical DID model. However, letting* $Y_t(x) = g_t(x, U_2)$

be potential outcomes, this model does not satisfy the parallel trend assumption if $\theta_2 \neq 1$, *because* $E[Y_2(0) - Y_1(0)|X_2 = x] = \theta_1 + (\theta_2 - 1)E[U_1|X_2 = x]$ *holds by the conditional stationarity of* U_t*. Hence, we cannot estimate the average treatment effect on the treated (ATT) or the average treatment effect (ATE) by using a standard DID method. Under this setting, we have*

$$
\begin{aligned}
ATE &\equiv E[Y_2(1) - Y_2(0)] = 1.00, \\
QTE_{25} &\equiv Q_{Y_2(1)}(0.25) - Q_{Y_2(0)}(0.25) \fallingdotseq 0.65, \\
QTE_{50} &\equiv Q_{Y_2(1)}(0.50) - Q_{Y_2(0)}(0.50) = 1.00, \\
QTE_{75} &\equiv Q_{Y_2(1)}(0.75) - Q_{Y_2(0)}(0.75) \fallingdotseq 1.35.
\end{aligned}
$$

We also estimated ATE and QTE as follows:

$$
\begin{aligned}
A\hat{T}E &= \left(\hat{\theta}_3 + \hat{\theta}_4 \hat{E}[\hat{U}]\right) - \left(\hat{\theta}_1 + \hat{\theta}_2 \hat{E}[\hat{U}]\right), \\
Q\hat{T}E_{100\tau} &= \left(\hat{\theta}_3 + \hat{\theta}_4 \hat{Q}_{\hat{U}}(\tau)\right) - \left(\hat{\theta}_1 + \hat{\theta}_2 \hat{Q}_{\hat{U}}(\tau)\right),
\end{aligned}
$$

where $\hat{E}[\hat{U}]$ *is a sample average of* $\hat{U} \equiv (Y_{i1}, \cdots, Y_{in}, g_2^{-1}(X_{12}, Y_{12}; \hat{\theta}), \cdots, g_2^{-1}(X_{n2}, Y_{n2}; \hat{\theta}))$, *and* $\hat{Q}_{\hat{U}}(\tau)$ *is a sample* τ*-th quantile of* $\hat{U}$*. Because* $X_1 = 0$ *and* X_2 *is discrete, we used*

$$
A_\theta^v(w) = \left(\mathbf{1}\{y_1 \leq v_u\} - \mathbf{1}\{y_2 \leq g_2(x_2, v_u; \theta)\}\right) \mathbf{1}\{x_2 = v_x\},
$$

where $v = (v_x, v_u)$*. We used* $\mu = Ber(0.5) \times N(\bar{Y}_1, s_{Y_1})$ *as the integrating measure, where* $\bar{Y}_1$ *is the sample average of* Y_1 *and* s_{Y_1} *is the standard deviation of* Y_1*. Table 3.3 contains the results of this experiment for sample sizes of* 400*,* 800*, and* 1600*. The number of replications was set to* 1000 *throughout. Table 3.3 shows the bias, standard deviation, and MSE of the estimates of parameters, the ATE, and QTE, highlighting that the standard deviation and MSE of estimates decrease as the sample size increases. The biases of the estimates of parameters, the ATE, and* QTE_{50} *are relatively small, whereas the biases of the estimates of* QTE_{25} *and* QTE_{75} *are large. This may be caused by the fact that the sample quantiles are biased.*

3.5 Discrete Outcomes

In this section, we consider the case where outcomes are discrete. In the case of discrete outcomes, we cannot point-identify $g_t(x, u)$. This is likewise true in Athey and Imbens (2006), Chesher (2010), and Ishihara (2017). They consider the case where outcomes are discrete, and instead show partial identification of the structural function. Hence, in this section, we also consider partial identification of $g_t(x, u)$.

First, we drop the i subscript and let $T = 2$, as in Section 3.2. Let $\mathcal{Y}_t$ denote the support of Y_t. The assumptions employed in Section 3.2 do not allow the outcomes to be discrete. Hence, we impose the following assumptions.

Assumption 3.D1. *(i) For all $t \in \{1, 2\}$, the function $g_t(x, u)$ is weakly increasing in u for all x. (ii) For all $t \in \{1, 2\}$, $U_t|\mathbf{X} = \mathbf{x}$ is continuously distributed for all $\mathbf{x}$.*

Assumption 3.D2. *(i) For all $t \in \{1, 2\}$, Y_t is discretely distributed. (ii) $\mathcal{Y}_1 = \mathcal{Y}_2 \equiv \mathcal{Y}$ with $\underline{y} \equiv \inf \mathcal{Y}$ and $\overline{y} \equiv \sup \mathcal{Y}$.*

Assumption 3.D3. *For all $t \in \{1, 2\}$, the marginal distribution of U_t is uniform on $[0, 1]$.*

Assumption 3.D4. *(i) For all $\mathbf{x} \in \mathcal{X}_{12}$, $U_1|\mathbf{X} = \mathbf{x} \overset{d}{=} U_2|\mathbf{X} = \mathbf{x}$ holds. (ii) The support of $U_t|\mathbf{X} = \mathbf{x}$ is $[0, 1]$.*

Assumption 3.I1 (i) stipulates that $g_t(x, u)$ is strictly increasing in u. If U_t is continuously distributed, then Y_t must be continuously distributed under Assumption 3.I1 (i). Hence, in this section, we relax Assumption 3.I1 by allowing g_t to be flat inside the support of U_t. Athey and Imbens (2006) and Chesher (2010) also employ this weakly monotonic assumption in models with discrete outcomes. Furthermore, when outcomes are discrete, we cannot use Assumption 3.I3, because U_t is continuously distributed. Hence, we use another normalization assumption. Assumption 3.D4 is identical to Assumption 3.I4.

We can thus obtain the following theorem.

Theorem 3.6. *Suppose that Assumptions 3.D1, 3.D2, 3.D3, and 3.D4 are satisfied. For all $t \in \{1, 2\}$, if $\mathcal{X}_{12} = \mathcal{X}_1 \times \mathcal{X}_2$ holds, then we have*

$$\begin{aligned} g_t(x, u) &\geq g_t^L(x, u) \equiv \inf\{y \in [\underline{y}, \overline{y}] : G_{t,x}^L(y) \geq u\}, \\ g_t(x, u) &\leq g_t^U(x, u) \equiv \sup\{y \in [\underline{y}, \overline{y}] : G_{t,x}^U(y) \leq u\}, \end{aligned}$$

where $G_{t,x}^L$ and $G_{t,x}^U$ are defined by (3.24) and (3.27), respectively.

This identification approach is similar to that in Ishihara (2017), who considers the identification of nonseparable models with binary instruments and shows that the structural functions are partially identified when outcomes are discrete.

In Theorem 3.6, we assume that $\mathcal{X}_{12} = \mathcal{X}_1 \times \mathcal{X}_2$. Although this support condition does not require stayers, it is nevertheless stronger than that of Theorem 3.1. Indeed, we can relax this condition and partially identify g_t under a weaker support condition. However, if we do, then the bounds of g_t may be looser.

To illustrate Theorem 3.6, we introduce two examples.

Example 3.3 (DID model with binary outcomes)**.** *Suppose that the outcomes are binary and* $\mathcal{X}_{12} = \{(0,0),(0,1)\}$. *Then,* $\mathcal{X}_{12} = \mathcal{X}_1 \times \mathcal{X}_2$, *where* $\mathcal{X}_1 = \{0\}$ *and* $\mathcal{X}_2 = \{0,1\}$. *This is the usual DID setting. Define* $D \equiv \mathbf{1}\{X_2 = 1\}$. *We consider the partial identification of* $g_2(0,u)$ *and* $g_2(1,u)$.

In this case, we have

$$G^L_{2,0}(y) = P(D=1)F^+_{Y_2|D=1}(T^U_{2,1,0}(y)) + P(D=0)F^+_{Y_2|D=0}(T^U_{2,0,0}(y)),$$

where $T^U_{2,1,0}(y) = Q^+_{Y_2|D=1} \circ F^+_{Y_2|D=1} \circ Q^+_{Y_1|D=0} \circ F^+_{Y_1|D=0}(y)$ *and* $T^U_{2,0,0}(y) = y$. *We define* $p_t(d) \equiv P(Y_t = 1|D = d)$, *then*

$$G^L_{2,0}(y) = \begin{cases} P(D=1, Y_2 \leq \mathbf{1}\{p_2(1) \geq p_1(1)\}) & \text{if } y < 0 \\ P(D=1, Y_2 \leq \mathbf{1}\{p_2(1) \geq p_1(1) \text{ or } p_1(0) \geq p_2(0)\}) + P(D=0, Y_2=0) & \text{if } 0 \leq y < 1 \\ 1 & \text{if } y \geq 1 \end{cases}.$$

Therefore, we can obtain a lower bound

$$g^L_2(0,u) = \begin{cases} \mathbf{1}\{u > P(Y_2=0)\} & \text{if } p_1(1) > p_2(1) \text{ and } p_1(0) < p_2(0) \\ \mathbf{1}\{u > P(Y_2=0) + P(D=1, Y_2=1)\} & \text{if } p_1(1) \leq p_2(1) \text{ or } p_1(0) \geq p_2(0) \end{cases}.$$

Similarly, we can obtain the following functions:

$$\begin{aligned} g^L_2(1,u) &= \begin{cases} \mathbf{1}\{u > P(Y_2=0)\} & \text{if } p_1(1) < p_2(1) \text{ and } p_1(0) > p_2(0) \\ \mathbf{1}\{u > P(D=0) + P(D=1, Y_2=0)\} & \text{if } p_1(1) \geq p_2(1) \text{ or } p_1(0) \leq p_2(0) \end{cases} \\ g^U_2(0,u) &= \begin{cases} \mathbf{1}\{u \geq P(D=0, Y_2=0)\} & \text{if } p_1(1) \geq p_2(1) \text{ or } p_1(0) \leq p_2(0) \\ \mathbf{1}\{u \geq P(Y_2=0)\} & \text{if } p_1(1) < p_2(1) \text{ and } p_1(0) > p_2(0) \end{cases} \\ g^U_2(1,u) &= \begin{cases} \mathbf{1}\{u \geq P(D=1, Y_2=0)\} & \text{if } p_1(1) \leq p_2(1) \text{ or } p_1(0) \geq p_2(0) \\ \mathbf{1}\{u \geq P(Y_2=0)\} & \text{if } p_1(1) > p_2(1) \text{ and } p_1(0) < p_2(0) \end{cases} \end{aligned}$$

If we define the potential outcomes as $Y_t(x) = g_t(x, U_t)$, *we can partially identify the ATE. Because* $g^L_2(x,u)$ *and* $g^U_2(x,u)$ *respectively denote the lower and upper bounds of* $g_2(x,u)$, *we have*

$$E[g^L_2(x,U)] \leq E[Y_2(x)] \leq E[g^U_2(x,U)], \quad \text{for } x = 1, 2,$$

where $U \sim Unif(0,1)$. *Hence, we have*

$$E[g^L_2(1,U)] - E[g^U_2(0,U)] \leq \mu_{ATE} \leq E[g^U_2(1,U)] - E[g^L_2(0,U)],$$

where $\mu_{ATE} \equiv E[Y_2(1) - Y_2(0)]$.

Hence, above bounds of g_2 imply that the lower (upper) bound of ATE is not larger (smaller) than 0. Actually, when $p_1(1) < p_2(1)$ and $p_1(0) > p_2(0)$, that is $E[Y_1(0)|D = 1] < E[Y_2(1)|D = 1]$ and $E[Y_1(0)|D = 0] > E[Y_2(0)|D = 0]$, a lower bound of ATE becomes 0. This situation implies that the mean of the treated group increases, although the time trend effect is negative. Hence, in this case, it is intuitive that the ATE is larger than 0. Contrarily, when $p_1(1) > p_2(1)$ and $p_1(0) < p_2(0)$, that is $E[Y_1(0)|D = 1] > E[Y_2(1)|D = 1]$ and $E[Y_1(0)|D = 0] < E[Y_2(0)|D = 0]$, an upper bound of ATE becomes 0. This situation implies that the mean of the treated group decreases, although the time trend effect is positive. Hence, in this case, it is intuitive that the ATE is smaller than 0.

As an example, we consider the following case:

$$\begin{aligned} &E[Y_1|D = 1] = 0.4, \quad E[Y_1|D = 0] = 0.3, \\ &E[Y_2|D = 1] = 0.5, \quad E[Y_1|D = 0] = 0.2, \\ &P(D = 1) = 0.5. \end{aligned}$$

In this case, we can obtain

$$\begin{aligned} g_2^L(0, u) &= \mathbf{1}\{u > 0.9\}, \\ g_2^L(1, u) &= \mathbf{1}\{u > 0.65\}, \\ g_2^U(0, u) &= \mathbf{1}\{u > 0.65\}, \\ g_2^U(1, u) &= \mathbf{1}\{u > 0.25\}. \end{aligned}$$

Hence, in this case, ATE is smaller than 0.65 and larger than 0. As discussed above, because $E[Y_1|D = 1] < E[Y_2|D = 1]$ and $E[Y_1|D = 0] > E[Y_2|D = 0]$, a lower bound of ATE becomes 0.

Example 3.4. *We consider the following model:*

$$Y_t = g_t(X_t, U_t) = \mathbf{1}\{U_t > (1 + \exp(\alpha_t + \beta_t X_t))^{-1}\}, \quad t = 1, 2,$$

where $U_t = \Phi(\epsilon_t)$ and

$$(X_1, X_2, \epsilon_t) \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1.0 & 0.6 & 0.4 \\ 0.6 & 1.0 & 0.4 \\ 0.4 & 0.4 & 1.0 \end{pmatrix}\right).$$

Hence, $U_t \sim Unif(0, 1)$ for all t and $U_1|\mathbf{X} = \mathbf{x} \overset{d}{=} U_2|\mathbf{X} = \mathbf{x}$ for all $\mathbf{x}$. We set $(\alpha_1, \alpha_2, \beta_1, \beta_2) = (0, 0.3, 0.5, 0.6)$. Under this setting, we calculate $g_t^L(x, u)$ and $g_t^U(x, u)$ defined by Theorem 3.5 for $x = -2, -1, 0, 1, 2$. Table 3.4 shows $g_t^L(x, u)$, $g_t^U(x, u)$, and $g_t(x, u)$ at $x = -2, -1, 0, 1, 2$.

When x is small, the lower (upper) bounds are uninformative (informative). Contrarily, when x is large, lower (upper) bounds are informative (uninformative). In this model, there is a positive time trend because $g_1(x,u) \leq g_2(x,u)$. These bounds reflect this fact. That is, they also satisfy $g_1^L(x,u) \leq g_2^L(x,u)$ and $g_1^U(x,u) \leq g_2^U(x,u)$.

We can extend Theorem 3.6 to panel data with more than two periods.

Corollary 3.3. *Suppose Assumptions 3.D1, 3.D2, 3.D3, and 3.D4 are satisfied for $T \geq 3$. For $t = 1, \cdots, T$, if supp$(\mathbf{X}) = \mathcal{X}_1 \times \cdots \times \mathcal{X}_T$, then we have*

$$\begin{aligned} g_t(x,u) &\geq g_t^L(x,u) \equiv \inf\{y \in [\underline{y}, \overline{y}] : G_{t,x}^L(y) \geq u\}, \\ g_t(x,u) &\leq g_t^U(x,u) \equiv \sup\{y \in [\underline{y}, \overline{y}] : G_{t,x}^U(y) \leq u\}, \end{aligned}$$

where $G_{t,x}^L(y)$ and $G_{t,x}^U(y)$ are defined by (3.32).

3.6 Conclusion

In this chapter, we explored the identification and estimation of nonseparable panel data models. We showed that the structural function is nonparametrically identified when the structural function $g_t(x,u)$ is strictly increasing in u, the conditional distributions of U_{it} are the same over time, and the joint support of $\mathbf{X}_i$ satisfies weak assumptions. Many nonseparable panel data models assume that the structural function does not change over time and that stayers exist. By contrast, our approach allows the structural function to depend on the time period in an arbitrary manner, and it does not require the existence of stayers. In estimation part of the paper, we propose parametric and nonparametric estimators that implement our identification results. Monte Carlo studies indicated that our parametric estimator performs well with finite samples. Finally, we extended our identification results to models with discrete outcomes and showed that the structural function is partially identified.

3.7 Appendix: Proofs

Proof of Theorem 3.1. First, we show that $g_t(x,u)$ is identified for all $x \in \cup_{m=0}^{\infty}\mathcal{S}_m^t$. By the monotonicity of g_t and (3.4), equations (3.6) hold for all $(x_1,x_2) \in \mathcal{X}_{12}$. First, we can identify $g_1(\overline{x},u) = u$ by Assumption 3.I3. We can also identify $g_2(x_2,u)$ for all $x_2 \in \mathcal{S}_0^2$ because $(\bar{x},x_2) \in \mathcal{X}_{12}$ and we have

$$g_2(x_2,u) = Q_{Y_2|\mathbf{X}}\left(F_{Y_1|\mathbf{X}}\left(g_1(\bar{x},u)|\bar{x},x_2\right)|\bar{x},x_2\right) = Q_{Y_2|\mathbf{X}}\left(F_{Y_1|\mathbf{X}}\left(u|\bar{x},x_2\right)|\bar{x},x_2\right).$$

We now turn to identifying $g_1(x_1,u)$ for $x_1 \in \mathcal{S}_1^1$. Fix $x_1 \in \mathcal{S}_1^1$. According to the definition of $\mathcal{S}_1^1$, there exists $x_2 \in \mathcal{S}_0^2$ such that $(x_1,x_2) \in \mathcal{X}_{12}$. Then, it follows from (3.6) that

$$g_1(x_1,u) = Q_{Y_1|\mathbf{X}}\left(F_{Y_2|\mathbf{X}}\left(g_2(x_2,u)|x_1,x_2\right)|x_1,x_2\right),$$

and hence, $g_1(x_1,u)$ is identified because $g_2(x_2,u)$ is already identified. Similarly, by using (3.6), we can identify $g_2(x,u)$ for all $x \in \mathcal{S}_1^2$. Repeating this argument gives the identification of $g_t(x,u)$ for all $x \in \cup_{m=0}^{\infty}\mathcal{S}_m^t$.

Next, we show that $g_t(x,u)$ is identified for all $x \in \mathcal{X}_t$. We fix $x' \in \mathcal{X}_t \backslash \left(\cup_{m=0}^{\infty}\mathcal{S}_m^t\right)$. Since $\mathcal{X}_t = \overline{\cup_{m=0}^{\infty}\mathcal{S}_m^t}$, there exists a sequence $\{x^m\}_{m=1}^{\infty} \subset \cup_{m=0}^{\infty}\mathcal{S}_m^t$ such that $\lim_{m\to\infty} x^m = x'$. By the continuity of g_t, we have $\lim_{m\to\infty} g_t(x^m,u) = g_t(x',u)$ for all $u \in \mathcal{U}$. Hence, we can also identify $g_t(x',u)$ because $g_t(x^m,u)$ is identified for all m. □

Proof of Corollary 3.1. First, we show that if for all $x,x' \in \mathcal{X}_t$, we can identify the strictly increasing function $T_{t,x',x}(y)$ that satisfies

$$g_t(x',u) = T_{t,x',x}\left(g_t(x,u)\right), \tag{3.16}$$

then, $g_t(x,u)$ is point identified. We define $G_x^t(y) \equiv \int F_{Y_t|X_t}\left(T_{t,x',x}(y)|x'\right)dF_{X_t}(x')$, and then we have

$$\begin{aligned} G_x^t\left(g_t(x,u)\right) &= \int F_{Y_t|X_t}\left(g_t(x',u)|x'\right)dF_{X_t}(x') \\ &= \int P\left(U_t \le u|X_t = x'\right)dF_{X_t}(x') \\ &= P(U_t \le u) = u, \end{aligned}$$

where the last equality follows from Assumption 3.I3'. Because $T_{t,x',x}(y)$ is strictly increasing in y, $G_x^t(y)$ is invertible. Hence, we obtain $g_t(x,u) = \left(G_x^t\right)^{-1}(u)$. This implies that $g_t(x,u)$ is identified if we can construct $T_{t,x',x}(y)$ for all $x,x' \in \mathcal{X}_t$.

To construct $T_{t,x',x}(y)$, we show that for all $x \in \mathcal{X}_t$, we can identify the strictly increasing function $T_{t,x}^*(y)$ that satisfies

$$g_t(x,u) = T_{t,x}^*(g_1(\bar{x},u)). \tag{3.17}$$

For all $x \in \cup_{m=0}^{\infty}\mathcal{S}_m^t$, the proof of Theorem 3.1 implies that we can construct $T_{t,x}^*(y)$ that satisfies (3.17). Because $F_{Y_t|\mathbf{X}}$ and $Q_{Y_t|\mathbf{X}}$ are strictly increasing, $T_{t,x}^*(y)$ is strictly increasing in y for all $x \in \cup_{m=0}^{\infty}\mathcal{S}_m^t$. We fix $x' \in \mathcal{X}_t \backslash (\cup_{m=0}^{\infty}\mathcal{S}_m^t)$. Since $\mathcal{X}_t = \overline{\cup_{m=0}^{\infty}\mathcal{S}_m^t}$, there exists a sequence $\{x^m\}_{m=1}^{\infty} \subset \cup_{m=0}^{\infty}\mathcal{S}_m^t$ such that $\lim_{m\to\infty} x^m = x'$. By the continuity of g_t and (3.17), we have $\lim_{m\to\infty} T_{t,x_m}^*(g_1(\overline{x},u)) = g_t(x,u)$. Because $g_t(x,u)$ is strictly increasing in u, $\lim_{m\to\infty} T_{t,x_m}^*(y)$ is also strictly increasing in y. Hence, for all $x \in \mathcal{X}_t$, we can identify the strictly increasing function $T_{t,x}^*(y)$ that satisfies (3.17).

By using $T_{t,x}^*(y)$, we identify $T_{t,x',x}(y)$ that satisfies (3.16). Because, for $x, x' \in \mathcal{X}_t$, we have

$$g_t(x',u) = T_{t,x'}^*\left((T_{t,x}^*)^{-1}(g_t(x,u))\right),$$

we can construct the function $T_{t,x',x}(y)$ that satisfies $g_t(x',u) = T_{t,x',x}(g_t(x,u))$. Therefore, we can identify $g_t(x,u)$. □

Proof of Corollary 3.2. The proof is the same as that for Theorem 3.1. □

Proof of Theorem 3.2. We fix $\delta > 0$. By Lemma 3.1, 3.C2, and (3.10), there exists $\epsilon > 0$ such that $\|\theta - \theta_0\| \geq \delta$ implies $\|D_\theta\|_\mu \geq \epsilon$. Therefore, we have

$$\|D_{\hat{\theta}_n}\|_\mu < \epsilon \quad \Rightarrow \quad \|\hat{\theta}_n - \theta_0\| < \delta$$

and it will suffice to show that $\|D_{\hat{\theta}_n}\|_\mu \to_{a.s.} 0$. By (3.33), we have

$$\sup_\theta \|\hat{D}_{n,\theta} - D_\theta\|_\mu \leq \sup_{\theta,v} |\hat{D}_{n,\theta}(v) - D_\theta(v)| = o_{a.s.}(1). \tag{3.18}$$

By the triangle inequality and 3.C1,

$$\begin{aligned}\|D_{\hat{\theta}_n}\|_\mu &\leq \|\hat{D}_{n,\hat{\theta}_n} - D_{\hat{\theta}_n}\|_\mu + \|\hat{D}_{n,\hat{\theta}_n}\|_\mu \\ &\leq \|\hat{D}_{n,\hat{\theta}_n} - D_{\hat{\theta}_n}\|_\mu + \|\hat{D}_{n,\theta_0}\|_\mu.\end{aligned}$$

By the uniform convergence (3.18), $\|\hat{D}_{n,\hat{\theta}_n} - D_{\hat{\theta}_n}\|_\mu = o_{a.s.}(1)$ and $\|\hat{D}_{n,\theta_0}\|_\mu = \|\hat{D}_{n,\theta_0} - D_{\theta_0}\|_\mu = o_{a.s.}(1)$. Hence, we can show that $\hat{\theta}_n \to_{a.s.} \theta_0$. □

Proof of Theorem 3.3. First, we prove the $\sqrt{n}$-consistency of $\hat{\theta}_n$. As seen in the previous theorem, $\hat{\theta}_n$ is a consistent estimator of θ_0. Because $\hat{\theta}_n$ is consistent, we can select a sequence $\{\delta_n\}$ that converges to zero sufficiently slowly to ensure

$$P(\|\hat{\theta}_n - \theta_0\| \geq \delta_n) \to 0.$$

For this sequence, the supremum in (3.35) runs over a range that includes $\hat{\theta}_n$. Hence, by the triangle inequality and Lemma 3.4, we have

$$\|D_{\hat{\theta}_n}\|_\mu - \|\hat{D}_{n,\hat{\theta}_n}\|_\mu - \|\hat{D}_{n,\theta_0}\|_\mu \leq \|\hat{D}_{n,\hat{\theta}_n} - D_{\hat{\theta}_n} - \hat{D}_{n,\theta_0}\|_\mu = o_p(n^{-1/2}).$$

From Assumption 3.C1,

$$\|D_{\hat{\theta}_n}\|_\mu \leq o_p(n^{-1/2}) + 2\|\hat{D}_{n,\theta_0}\|_\mu.$$

Because $E[A^v_{\theta_0}(W)] = 0$ for all v and $E|A^v_{\theta_0}(W)A^{v'}_{\theta_0}(W)| \leq 1$ for all v and v', we have $\sqrt{n}\hat{D}_{n,\theta_0}(v) = \frac{1}{n}\sum_{i=1}^n A^v_\theta(W_i) \rightsquigarrow N(0, \Psi(v,v))$. Since the proof for Lemma 3.2 shows that $\{A^v_{\theta_0} : v \in \mathcal{V}\}$ is a Donsker class, $\{\sqrt{n}\hat{D}_{n,\theta_0}(v) : v \in \mathcal{V}\}$ converges weakly in $l^\infty(\mathcal{V})$ to a mean-zero Gaussian process with covariance function $\Psi(v,v')$ and we have $\|\hat{D}_{n,\theta_0}\|_\mu = O_p(n^{-1/2})$. Hence, we have

$$\|D_{\hat{\theta}_n}\|_\mu = O_p(n^{-1/2}).$$

Because $D_{\theta_0}(v) = 0$ for all v, Lemma 3.3 implies that for all θ in a neighborhood of θ_0,

$$\begin{aligned}
\|D_\theta\|_\mu &= \|\Gamma_0(v)'(\theta-\theta_0) - (D_\theta(v) - D_{\theta_0}(v) - \Gamma_0(v)'(\theta-\theta_0))\,\|_\mu \\
&\geq \|\Gamma_0(v)'(\theta-\theta_0)\|_\mu - \|D_\theta(v) - D_{\theta_0}(v) - \Gamma_0(v)'(\theta-\theta_0)\|_\mu \\
&\geq (c - o(1)) \times \|\theta-\theta_0\|.
\end{aligned}$$

Therefore, $\|\hat{\theta}_n - \theta_0\| \leq \frac{1}{c-o_p(1)}\|D_{\hat{\theta}_n}\|_\mu = O_p(n^{-1/2})$.

Next we establish the asymptotic normality of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ by approximating $\hat{D}_{n,\theta}(v)$ as the linear function

$$\hat{L}_{n,\theta}(v) \equiv \Gamma_0(v)'(\theta-\theta_0) + \hat{D}_{n,\theta_0}(v).$$

We have

$$\begin{aligned}
\|\hat{D}_{n,\hat{\theta}_n} - \hat{L}_{n,\hat{\theta}_n}\|_\mu &\leq \|\hat{D}_{n,\hat{\theta}_n} - D_{\hat{\theta}_n} - \hat{D}_{n,\theta_0}\|_\mu + \|D_{\hat{\theta}_n}(v) - \Gamma_0(v)'(\hat{\theta}_n - \theta_0)\|_\mu \\
&\leq o_p(n^{-1/2}) + o_p(\|\hat{\theta}_n - \theta_0\|) = o_p(n^{-1/2}),
\end{aligned}$$

where the second inequality follows from Lemma 3.3 and Lemma 3.4, and the last equality follows from the $\sqrt{n}$-consistency of $\hat{\theta}_n$.

Let $\tilde{\theta}_n$ be the value that provides a global minimum for $\|\hat{L}_{n,\theta}\|$. Then, $\Gamma_0(\cdot)'(\tilde{\theta}_n - \theta_0)$ is the $L_2(\mu)$-projection of $-\hat{D}_{n,\theta_0}(\cdot)$ onto the subspace of $L_2(\mu)$ spanned by the components of $\Gamma_0(\cdot)$. Because $\Delta_0 = \int \Gamma_0(v)\Gamma_0(v)'d\mu(v)$ is finite and invertible by 3.N3, we have

$$\sqrt{n}(\tilde{\theta}_n - \theta_0) = -\Delta_0^{-1}\int \Gamma_0(v)\sqrt{n}\hat{D}_{n,\theta_0}(v)d\mu(v). \tag{3.19}$$

Then, we have

$$\int \Gamma_0(v)\sqrt{n}\hat{D}_{n,\theta_0}(v)d\mu(v) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int A^v_{\theta_0}(W_i)\Gamma_0(v)d\mu(v)$$
$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i,$$

for $\xi_i \equiv \int A^v_{\theta_0}(W_i)\Gamma_0(v)d\mu(v)$. By Fubini's theorem, $E[\xi_i] = \int E[A^v_{\theta_0}(W_i)]\Gamma_0(v)d\mu(v) = 0$, and

$$E[\xi_i\xi_i'] = \int\int_{\mathcal{V}\times\mathcal{V}} \left\{E[A^v_{\theta_0}(W_i)A^{v'}_{\theta_0}(W_i)]\Gamma_0(v)\Gamma_0(v')'\right\} d\mu(v)d\mu(v')$$
$$= \int\int_{\mathcal{V}\times\mathcal{V}} \{\Psi(v,v')\Gamma_0(v)\Gamma_0(v')'\} d\mu(v)d\mu(v'),$$

where all elements of $E[\xi\xi']$ are finite. Hence, $\sqrt{n}(\tilde{\theta}_n - \theta_0) \rightsquigarrow N(0,\Sigma_0)$ by (3.19). Consequently, $\tilde{\theta}_n = \theta_0 + O_p(n^{-1/2})$, and $\{\delta_n\}$ can be assumed to satisfy $P(\|\tilde{\theta}_n - \theta_0\| \geq \delta_n) \to 0$. Because θ_0 is an interior point of Θ, $\tilde{\theta}_n$ lies in Θ with probability approaching one. To simplify the argument, we assume that $\|\tilde{\theta}_n - \theta_0\| < \delta_n$ and $\tilde{\theta}_n$ always belongs to Θ.

Because $|D_\theta(v)| \leq |\Gamma_0(v)'(\theta-\theta_0)| + o(\|\theta-\theta_0\|)$ uniformly over v by Lemma 3.3, we have

$$\|D(\tilde{\theta}_n)\|_\mu \leq \|\Gamma_0(v)'(\tilde{\theta}_n - \theta_0)\|_\mu + o_p(\|\tilde{\theta}_n - \theta_0\|) = O_p(n^{-1/2}).$$

By the triangle inequality and Lemma 3.4, we have $\|\hat{D}_{n,\tilde{\theta}_n}\|_\mu - \|D_{\tilde{\theta}_n}\|_\mu - \|\hat{D}_{n,\theta_0}\|_\mu = o_p(n^{-1/2})$, and hence $\|\hat{D}_{n,\tilde{\theta}_n}\|_\mu = O_p(n^{-1/2})$. Then, we can argue as for $\hat{\theta}_n$ to deduce that

$$\|\hat{D}_{n,\tilde{\theta}_n} - \hat{L}_{n,\tilde{\theta}_n}\|_\mu = o_p(n^{-1/2}).$$

Above, we showed that $\|\hat{D}_{n,\hat{\theta}_n} - \hat{L}_{n,\hat{\theta}_n}\|_\mu = o_p(n^{-1/2})$ and $\|\hat{D}_{n,\tilde{\theta}_n} - \hat{L}_{n,\tilde{\theta}_n}\|_\mu = o_p(n^{-1/2})$. Therefore, we have

$$\|\hat{L}_{n,\hat{\theta}_n}\|_\mu - o_p(n^{-1/2}) \leq \|\hat{D}_{n,\hat{\theta}_n}\|_\mu$$
$$\leq \|\hat{D}_{n,\tilde{\theta}_n}\|_\mu + o_p(n^{-1/2})$$
$$\leq \|\hat{L}_{n,\tilde{\theta}_n}\|_\mu + o_p(n^{-1/2}).$$

That is,

$$\|\hat{L}_{n,\hat{\theta}_n}\|_\mu = \|\hat{L}_{n,\tilde{\theta}_n}\|_\mu + o_p(n^{-1/2}),$$

and by squaring both sides, we have

$$\|\hat{L}_{n,\hat{\theta}_n}\|^2_\mu = \|\hat{L}_{n,\tilde{\theta}_n}\|^2_\mu + o_p(n^{-1}),$$

where the cross product term is absorbed into $o_p(n^{-1})$ because $\|\hat{L}_{n,\tilde{\theta}_n}\|_\mu = O_p(n^{-1/2})$. Because $\hat{L}_{n,\tilde{\theta}_n}(\cdot)$ and $\Gamma_0(\cdot)$ are orthogonal according to the definition of $\tilde{\theta}_n$, we can obtain

$$\begin{aligned} \|\hat{L}_{n,\theta}\|_\mu^2 &= \|\hat{L}_{n,\tilde{\theta}_n}(v) + \Gamma_0(v)'(\theta - \tilde{\theta}_n)\|_\mu^2 \\ &= \|\hat{L}_{n,\tilde{\theta}_n}\|_\mu^2 + \|\Gamma_0(v)'(\theta - \tilde{\theta}_n)\|_\mu^2. \end{aligned}$$

By making θ equal to $\hat{\theta}_n$, we have

$$o_p(n^{-1}) = \|\Gamma_0(v)(\hat{\theta}_n - \tilde{\theta}_n)\|_\mu^2 \geq c^2\|\hat{\theta}_n - \tilde{\theta}_n\|^2.$$

Hence, $\sqrt{n}(\hat{\theta}_n - \theta_0) = \sqrt{n}(\tilde{\theta}_n - \theta_0) + o_p(1) \rightsquigarrow N(0, \Delta_0^{-1}\Sigma_0\Delta_0^{-1})$. □

Suppose that each W_i is a coordinate function of $(\prod_{i=1}^\infty S, \prod_{i=1}^\infty \sigma(S), \prod_{i=1}^\infty P)$. Let ω denote one of the realizations of W_i, and let $(W_{n1}^*, \cdots, W_{nn}^*)$ denote the bootstrap sample. Following Hahn (1996), we introduce the following notations. Let $\{\zeta_n^*\}$ be a sequence of some bootstrap statistic: each ζ_n^* is some function $f_n(W_{n1}^*, \cdots, W_{nn}^*)$ of the bootstrap sample. We write $\zeta_n^* = O_p^\omega(a_n)$ if ζ_n^*, when conditioned on ω, is $O_p(a_n)$ for almost all ω. If ζ_n^*, when conditioned on ω, is $o_p(a_n)$ for almost all ω, we write $\zeta_n^* = o_p^\omega(a_n)$. We write $\zeta_n^* = O_B(1)$ if, for a given subsequence $\{n'\}$, there exists a further subsequence $\{n''\}$ such that $\zeta_{n''}^* = O_p^\omega(1)$. If for any subsequence $\{n'\}$ there is a further subsequence $\{n''\}$ such that $\zeta_{n''}^* = o_p^\omega(1)$, we write $\zeta_n^* = o_B(1)$. Note that $\zeta_n^* = o_B(1)$ if and only if ζ_n^* converges weakly to zero in probability.

Proof of Theorem 3.4. The proof is similar to that of Brown and Wegkamp (2002). First, we define

$$\begin{aligned} M(\theta) &\equiv \int D_\theta(v)^2 d\mu(v), \\ M_n(\theta) &\equiv \int \hat{D}_{n,\theta}(v)^2 d\mu(v), \\ M_n^*(\theta) &\equiv \int \hat{D}_{n,\theta}^*(v)^2 d\mu(v). \end{aligned}$$

Then, for any $\theta \to \theta_0$, we have

$$\begin{aligned} M_n^*(\theta) - M_n(\theta) &= \int (\hat{D}_{n,\theta}^*(v) - \hat{D}_{n,\theta}(v))^2 d\mu(v) + 2\int \hat{D}_{n,\theta}(v)(\hat{D}_{n,\theta}^*(v) - \hat{D}_{n,\theta}(v))d\mu(v) \\ &= \int (\hat{D}_{n,\theta_0}^*(v) - \hat{D}_{n,\theta_0}(v))^2 d\mu(v) + \left[\int (\hat{D}_{n,\theta}^*(v) - \hat{D}_{n,\theta}(v))^2 d\mu(v) \right. \\ &\quad \left. - \int (\hat{D}_{n,\theta_0}^*(v) - \hat{D}_{n,\theta_0}(v))^2 d\mu(v)\right] + 2\int D_\theta(v)(\hat{D}_{n,\theta}^*(v) - \hat{D}_{n,\theta}(v))d\mu(v) \\ &\quad +2\int (\hat{D}_{n,\theta}(v) - D_\theta(v))(\hat{D}_{n,\theta}^*(v) - \hat{D}_{n,\theta}(v))d\mu(v). \end{aligned}$$

Suppose that $\|\theta-\theta_0\| \le \delta_n$ for $\delta_n \downarrow 0$. By Lemma 3.6, we obtain $\|\sqrt{n}(\hat{D}^*_{n,\theta}-\hat{D}_{n,\theta})\|_\mu - \|\sqrt{n}(\hat{D}^*_{n,\theta_0}-\hat{D}_{n,\theta_0})\|_\mu = o_B(1)$. Hence,

$$\int(\hat{D}^*_{n,\theta}(v)-\hat{D}_{n,\theta}(v))^2 d\mu(v) - \int(\hat{D}^*_{n,\theta_0}(v)-\hat{D}_{n,\theta_0}(v))^2 d\mu(v) = o_B(n^{-1}).$$

Similarly, by the Donsker property of $\{A^v_\theta : \theta\in\Theta, v\in\mathcal{V}\}$,

$$\begin{aligned} & \int(\hat{D}_{n,\theta}(v)-D_\theta(v))(\hat{D}^*_{n,\theta}(v)-\hat{D}_{n,\theta}(v))d\mu(v) \\ = & \int \hat{D}_{n,\theta_0}(v)(\hat{D}^*_{n,\theta}(v)-\hat{D}_{n,\theta}(v))d\mu(v) + o_p(n^{-1}). \end{aligned}$$

Therefore, we have

$$\begin{aligned} M^*_n(\theta)-M_n(\theta) = & \int(\hat{D}^*_{n,\theta_0}(v)-\hat{D}_{n,\theta_0}(v))^2 d\mu(v) + 2\int \hat{D}_{n,\theta_0}(v)(\hat{D}^*_{n,\theta}(v)-\hat{D}_{n,\theta}(v))d\mu(v) \\ & +2\int D_\theta(v)(\hat{D}^*_{n,\theta}(v)-\hat{D}_{n,\theta}(v))d\mu(v) + o_B(n^{-1}) \\ = & \int(\hat{D}^*_{n,\theta_0}(v)-\hat{D}_{n,\theta_0}(v))^2 d\mu(v) + 2\int \hat{D}_{n,\theta_0}(v)(\hat{D}^*_{n,\theta}(v)-\hat{D}_{n,\theta}(v))d\mu(v) \\ & +2(\theta-\theta_0)'\int \Gamma_0(v)(\hat{D}^*_{n,\theta}(v)-\hat{D}_{n,\theta}(v))d\mu(v) \\ & +o_B(n^{-1/2}\|\theta-\theta_0\|+n^{-1}). \end{aligned}$$

Consequently, for $\theta\to\theta_0$ and $\eta\to\theta_0$,

$$\begin{aligned} & M^*_n(\theta)-M^*_n(\eta) \\ = & [(M^*_n-M_n)(\theta)-(M^*_n-M_n)(\eta)] + [(M_n-M)(\theta)-(M_n-M)(\eta)] + [M(\theta)-M(\eta)] \\ = & 2(\theta-\eta)'\int\Gamma_0(v)\left[(\hat{D}^*_{n,\theta_0}(v)-\hat{D}_{n,\theta_0}(v))-(\hat{D}_{n,\theta_0}(v)-D_{\theta_0}(v))\right]d\mu(v) \\ & +\int\left[(\theta-\theta_0)'\Gamma_0(v)+o(\|\theta-\theta_0\|)\right]^2 d\mu(v) - \int\left[(\eta-\theta_0)'\Gamma_0(v)+o(\|\eta-\theta_0\|)\right]^2 d\mu(v) \\ & +o_B(n^{-1/2}\|\theta-\theta_0\|+n^{-1/2}\|\eta-\theta_0\|+n^{-1}) \\ = & 2(\theta-\eta)'\int\Gamma_0(v)\left[(\hat{D}^*_{n,\theta_0}(v)-\hat{D}_{n,\theta_0}(v))-(\hat{D}_{n,\theta_0}(v)-D_{\theta_0}(v))\right]d\mu(v) \\ & +(\theta-\theta_0)'\Delta_0(\theta-\theta_0)-(\eta-\theta_0)'\Delta_0(\eta-\theta_0) \\ & +o_B(\|\theta-\theta_0\|^2+\|\eta-\theta_0\|^2+n^{-1/2}\|\theta-\theta_0\|+n^{-1/2}\|\eta-\theta_0\|+n^{-1}). \end{aligned} \tag{3.20}$$

We define

$$\begin{aligned} \gamma_n & \equiv \Delta_0^{-1}\int\Gamma_0(v)(\hat{D}_{n,\theta_0}(v)-D_{\theta_0}(v))d\mu(v), \\ \gamma^*_n & \equiv \Delta_0^{-1}\int\Gamma_0(v)(\hat{D}^*_{n,\theta_0}(v)-\hat{D}_{n,\theta_0}(v))d\mu(v). \end{aligned}$$

Then, we can rewrite (3.20) by

$$\begin{aligned} M_n^*(\theta) - M_n^*(\eta) &= 2(\theta-\eta)'\Delta_0(\gamma_n+\gamma_n^*) + (\theta-\theta_0)'\Delta_0(\theta-\theta_0) - (\eta-\theta_0)'\Delta_0(\eta-\theta_0) \\ &\quad + o_B(\|\theta-\theta_0\|^2 + \|\eta-\theta_0\|^2 + n^{-1/2}\|\theta-\theta_0\| + n^{-1/2}\|\eta-\theta_0\| + n^{-1}). \end{aligned}$$

We take $\theta = \hat{\theta}_n^*$ and $\eta = \theta_0 - (\gamma_n + \gamma_n^*)$. Observe that $\eta \in \Theta$ for n is sufficiently large, because θ_0 is an interior point of Θ. Hence, we have

$$\begin{aligned} 0 &\geq M_n^*(\hat{\theta}_n^*) - M_n^*(\theta_0 - (\gamma_n+\gamma_n^*)) \\ &= 2\left[(\hat{\theta}_n^*-\theta_0) + (\gamma_n+\gamma_n^*)\right]'\Delta_0(\gamma_n+\gamma_n^*) \\ &\quad + (\hat{\theta}_n^*-\theta_0)'\Delta_0(\hat{\theta}_n^*-\theta_0) - (\gamma_n+\gamma_n^*)'\Delta_0(\gamma_n+\gamma_n^*) \\ &\quad + o_B(\|\hat{\theta}_n^*-\theta_0\|^2 + \|\gamma_n+\gamma_n^*\|^2 + n^{-1/2}\|\hat{\theta}_n^*-\theta_0\| + n^{-1/2}\|\gamma_n+\gamma_n^*\| + n^{-1}) \\ &= \left[(\hat{\theta}_n^*-\theta_0) + (\gamma_n+\gamma_n^*)\right]'\Delta_0\left[(\hat{\theta}_n^*-\theta_0) + (\gamma_n+\gamma_n^*)\right] \\ &\quad + o_B(\|\hat{\theta}_n^*-\theta_0\|^2 + \|\gamma_n+\gamma_n^*\|^2 + n^{-1/2}\|\hat{\theta}_n^*-\theta_0\| + n^{-1/2}\|\gamma_n+\gamma_n^*\| + n^{-1}). \end{aligned}$$

By the same argument in Theorem 3.4, we have $\|\hat{\theta}_n^* - \hat{\theta}_n\| = O_B(n^{-1/2})$. Hence, $\|\hat{\theta}_n^* - \theta_0\| \leq \|\hat{\theta}_n^* - \hat{\theta}_n\| + \|\hat{\theta}_n - \theta_0\| = O_B(n^{-1/2}) + O_P(n^{-1/2})$. Since $\gamma_n = O_p(n^{-1/2})$ and $\gamma_n^* = O_B(n^{-1/2})$, we have

$$n\|\hat{\theta}_n^* - \theta_0 + (\gamma_n + \gamma_n^*)\|^2 = o_B(1).$$

Because it follows from Theorem 3.4 that

$$\hat{\theta}_n - \theta_0 = -\gamma_n + o_p(n^{-1/2}),$$

and we can obtain $\hat{\theta}_n^* - \hat{\theta}_n = -\gamma_n^* + o_B(n^{-1/2})$. The term γ_n^* has the same limiting distribution as γ_n by the bootstrap theorem for the mean in $\mathbb{R}^{d_\theta}$. This concludes the proof. □

Proof of Theorem 3.5. Fix $u \in \overline{\mathcal{U}}$. First, we show consistency of $\hat{g}_{n,2}(x,u)$. We define

$$\begin{aligned} \psi(x,u;\xi) &\equiv F_{Y_1|\mathbf{X}}(u|\bar{x},x) - F_{Y_2|\mathbf{X}}(\xi|\bar{x},x), \\ \hat{\psi}_n(x,u;\xi) &\equiv \hat{F}_{Y_1|\mathbf{X}}(u|\bar{x},x) - \hat{F}_{Y_2|\mathbf{X}}(\xi|\bar{x},x). \end{aligned}$$

Assumptions 3.I3 and 3.I4 imply that $\psi(x,u;\xi) = 0$ at $\xi = g_2(x,u)$. By Assumptions NP1, for any $\epsilon > 0$, there exists $\delta > 0$ such that for all $x \in \mathcal{X}_2$,

$$|g_2(x,u) - \xi| > \epsilon \;\Rightarrow\; |F_{Y_2|\mathbf{X}}(g_2(x,u)|\bar{x},x) - F_{Y_2|\mathbf{X}}(\xi|\bar{x},x)| > \delta.$$

This implies that for all $x \in \mathcal{X}_2$, we have

$$|g_2(x,u) - \xi| > \epsilon \;\Rightarrow\; |\psi(x,u;\xi)| > \delta.$$

Hence, if $\sup_{x\in\mathcal{X}_2}|\psi(x,u;\hat{g}_{n,2}(x,u))| = o_p(1)$, we can show $\sup_{x\in\mathcal{X}_2}|\hat{g}_{n,2}(x,u) - g_2(x,u)| = o_p(1)$. It suffices to show $\sup_{x\in\mathcal{X}_2}|\psi(x,u;\hat{g}_{n,2}(x,u))| = o_p(1)$. It follows from Assumption 3.NP3 that

$$\sup_{x,\xi}|\hat{\psi}_n(x,u;\xi) - \psi(x,u;\xi)| = o_p(1). \tag{3.21}$$

We have

$$\begin{aligned}
\sup_x|\psi(x,u;\hat{g}_{n,2}(x,u))| &\leq \sup_x\left|\hat{\psi}_n(x,u;\hat{g}_{n,2}(x,u)) - \psi(x,u;\hat{g}_{n,2}(x,u))\right| \\
&\quad + \sup_x\left|\hat{\psi}_n(x,u;\hat{g}_{n,2}(x,u))\right| \\
&\leq \sup_x\left|\hat{\psi}_n(x,u;\hat{g}_{n,2}(x,u)) - \psi(x,u;\hat{g}_{n,2}(x,u))\right| \\
&\quad + \sup_x\left|\hat{\psi}_n(x,u;g_2(x,u)) - \psi(x,u;g_2(x,u))\right| \\
&= 2\sup_{x,\xi}|\hat{\psi}_n(x,u;\xi) - \psi(x,u;\xi)| \;=\; o_p(1),
\end{aligned}$$

where the second equality follows from the definition of $\hat{g}_{2,n}(x,u)$ and $\psi(x,u;g_2(x,u)) = 0$, and the last equality follows from (3.21). Therefore, we have $\sup_{x\in\mathcal{X}_2}|\hat{g}_{n,2}(x,u) - g_2(x,u)| = o_p(1)$.

Next, we show consistency of $\hat{g}_{n,1}(x,u)$. We define

$$\begin{aligned}
\phi(x,u;\xi) &\equiv \left\{\int\left|F_{Y_1|\mathbf{X}}(\xi|x,x_2) - F_{Y_2|\mathbf{X}}(g_2(x_2,u)|x,x_2)\right|^2 d\mu_{X_2}(x_2)\right\}^{-1/2}, \\
\hat{\phi}_n(x,u;\xi) &\equiv \left\{\int\left|\hat{F}_{Y_1|\mathbf{X}}(\xi|x,x_2) - \hat{F}_{Y_2|\mathbf{X}}(\hat{g}_{n,2}(x_2,u)|x,x_2)\right|^2 d\mu_{X_2}(x_2)\right\}^{-1/2}.
\end{aligned}$$

By Assumption 3.I4, we have $\phi(x,u;g_1(x,u)) = 0$ for all $x \in \mathcal{X}_1$. Similar to the above argument, if $\phi(x,u;\hat{g}_{1,n}(x,u)) = o_p(1)$, then we can show $\hat{g}_{1,n}(x,u) \to_p g_1(x,u)$. It follows from uniform consistency of $\hat{g}_{2,n}(x,u)$ and Assumptions 3.NP2 and 3.NP3 that

$$\begin{aligned}
&\sup_{x,\xi}|\hat{\phi}_n(x,u;\xi) - \phi(x,u;\xi)| \\
\leq\; &\sup_{x,\xi}\left\{\int\left|\left(\hat{F}_{Y_1|\mathbf{X}}(\xi|x,x_2) - F_{Y_1|\mathbf{X}}(\xi|x,x_2)\right)\right.\right. \\
&+\left(\hat{F}_{Y_2|\mathbf{X}}(\hat{g}_{n,2}(x_2,u)|x,x_2) - F_{Y_2|\mathbf{X}}(\hat{g}_{n,2}(x_2,u)|x,x_2)\right) \\
&\left.\left.+\left(F_{Y_2|\mathbf{X}}(\hat{g}_{n,2}(x_2,u)|x,x_2) - F_{Y_2|\mathbf{X}}(g_2(x_2,u)|x,x_2)\right)\right|^2 d\mu_{X_2}(x_2)\right\}^{-1/2} \\
=\; &o_p(1).
\end{aligned} \tag{3.22}$$

Similar to the above argument, by (3.22), we have

$$\begin{aligned}
\sup_x\phi(x,u;\hat{g}_{n,1}(x,u)) &\leq \sup_x\left|\hat{\phi}_n(x,u;\hat{g}_{n,1}(x,u)) - \phi(x,u;\hat{g}_{n,1}(x,u))\right| \\
&\quad + \sup_x\left|\hat{\phi}_n(x,u;g_1(x,u)) - \phi(x,u;g_1(x,u))\right| \\
&= o_p(1).
\end{aligned}$$

Therefore, we obtain $\hat{g}_{1,n}(x,u) \to_p g_1(x,u)$ for all $x \in \mathcal{X}_1$. □

Proof of Theorem 3.6. We establish the partial identification of g_t by showing that we can identify functions $T^U_{t,x',x} : \mathbb{R} \to \mathbb{R}$ and $T^L_{t,x',x} : \mathbb{R} \to \mathbb{R}$ that satisfy

$$\begin{aligned} g_t(x',u) &\leq T^U_{t,x',x}\left(g_t(x,u)\right), \\ g_t(x',u) &\geq T^L_{t,x',x}\left(g_t(x,u)\right). \end{aligned} \tag{3.23}$$

If $T^U_{x',x}$ is identified for all x, x', then we can obtain a lower bound of the function g as follows. For any random variables V, W, we define

$$\begin{aligned} F^+_{V|W}(v|w) &\equiv P\left(V \leq v | W = w\right), \\ F^-_{V|W}(v|w) &\equiv P\left(V < v | W = w\right), \end{aligned}$$

where F^+ is an usual distribution function. In addition, we define

$$G^L_{t,x}(y) \equiv \int F^+_{Y_t|X_t=x'}\left(T^U_{t,x',x}(y)\right) dF_{X_t}(x'). \tag{3.24}$$

Then, we have

$$\begin{aligned} G^L_{t,x}\left(g_t(x,u)\right) &= \int F^+_{Y_t|X_t=x'}\left(T^U_{t,x',x}\left(g_t(x,u)\right)\right) dF_{X_t}(x') \\ &\geq \int F^+_{Y_t|X_t=x'}\left(g_t(x',u)\right) dF_{X_t}(x') \\ &= \int P\left(g_t(x',U_t) \leq g_t(x',u) | X_t = x'\right) dF_{X_t}(x') \\ &\geq \int P\left(U_t \leq u | X_t = x'\right) dF_{X_t}(x') = u, \end{aligned} \tag{3.25}$$

where the first inequality follows from (3.23). Because $g_t(x,u)$ is weakly increasing in u, we have $\{U_t \leq u\} \subset \{g_t(x,U_t) \leq g_t(x,u)\}$ and the second inequality of (3.25) holds. Hence, because $G^L_{t,x}\left(g_t(x,u)\right) \geq u$, we can obtain a lower bound

$$g_t(x,u) \geq \inf\{y \in [\underline{y}, \overline{y}] : G^L_{t,x}(y) \geq u\}. \tag{3.26}$$

Similarly, we define

$$G^U_{t,x}(y) \equiv \int F^-_{Y_t|X_t=x'}\left(T^L_{t,x',x}(y)\right) dF_{X_t}(x'). \tag{3.27}$$

Then, we have

$$\begin{aligned} G^U_{t,x}\left(g_t(x,u)\right) &= \int F^-_{Y_t|X_t=x'}\left(T^L_{t,x',x}(g_t(x,u))\right) dF_{X_t}(x') \\ &\leq \int F^-_{Y_t|X_t=x'}\left(g_t(x',u)\right) dF_{X_t}(x') \\ &= \int P\left(g_t(x',U_t) < g_t(x',u) | X_t = x'\right) dF_{X_t}(x') \\ &\leq \int P\left(U_t < u | X_t = x'\right) dF_{X_t}(x') = u. \end{aligned} \tag{3.28}$$

Owing to the weak monotonicity of g_t, we have $\{g_t(x, U_t) < g_t(x, u)\} \subset \{U_t < u\}$ and the second inequality of (3.28) holds. Hence, similarly, we can obtain an upper bound

$$g_t(x, u) \geq \sup\{y \in [\underline{y}, \overline{y}] : G^U_{t,x}(y) \leq u\}. \tag{3.29}$$

We here describe the construction of the functions $T^U_{t,x',x}(y)$ and $T^L_{t,x',x}(y)$ that satisfy (3.23). We define

$$\begin{aligned} Q^+_{Y_t|\mathbf{X}}(\tau|\mathbf{x}) &\equiv \sup\{y \in [\underline{y}, \overline{y}] : F^-_{Y_t|\mathbf{X}}(y|\mathbf{x}) \leq \tau\}, \\ Q^-_{Y_t|\mathbf{X}}(\tau|\mathbf{x}) &\equiv \inf\{y \in [\underline{y}, \overline{y}] : F^+_{Y_t|\mathbf{X}}(y|\mathbf{x}) \geq \tau\}. \end{aligned}$$

Because $\{U_t : U_t \leq u\} \subset \{U_t : g_t(x, U_t) \leq g_t(x, u)\}$ and $\{U_t : g_t(x, U_t) < g_t(x, u)\} \subset \{U_t : U_t < u\}$, for all $(x_1, x_2) \in \mathcal{X}_{12}$ and $t, s \in \{(1, 2), (2, 1)\}$, we have

$$\begin{aligned} F^-_{Y_t|\mathbf{X}}\left(g_t(x_t, u)|x_1, x_2\right) &= P\left(g_t(x_t, U_t) < g_t(x_t, u)|X_1 = x_1, X_2 = x_2\right) \\ &\leq P\left(U_t < u|X_1 = x_1, X_2 = x_2\right) \\ &= P\left(U_s < u|X_1 = x_1, X_2 = x_2\right) \\ &\leq P\left(g_s(x_s, U_s) < g_s(x_s, u)|X_1 = x_1, X_2 = x_2\right) \\ &= F^+_{Y_s|\mathbf{X}}\left(g_s(x_s, u)|x_1, x_2\right). \end{aligned}$$

For $t \neq s$, we define

$$\begin{aligned} \tilde{T}^{U,t,s}_{x_t,x_s}(y) &\equiv Q^+_{Y_t|(X_t,X_s)=(x_t,x_s)}\left(F^+_{Y_s|(X_t,X_s)=(x_t,x_s)}(y)\right) \\ \tilde{T}^{L,t,s}_{x_t,x_s}(y) &\equiv Q^-_{Y_t|(X_t,X_s)=(x_t,x_s)}\left(F^-_{Y_s|(X_t,X_s)=(x_t,x_s)}(y)\right). \end{aligned}$$

Then, we have

$$g_t(x_t, u) \leq \tilde{T}^{U,t,s}_{x_t,x_s}\left(g_s(x_s, u)\right),$$

because $Q^+_{Y_t|\mathbf{X}}(F^-_{Y_t|\mathbf{X}}(y|\mathbf{x})|\mathbf{x}) = \sup\{y' \in [\underline{y}, \overline{y}] : F^-_{Y_t|\mathbf{X}}(y'|\mathbf{x}) \leq F^+_{Y_t|\mathbf{X}}(y|\mathbf{x})\} \geq y$. Hence, if $(x', \tilde{x}) \in supp(X_t, X_s)$ and $(x, \tilde{x}) \in supp(X_t, X_s)$, then we have

$$g_t(x', u) \leq \tilde{T}^{U,t,s}_{x',\tilde{x}} \circ \tilde{T}^{U,s,t}_{\tilde{x},x}\left(g_t(x, u)\right). \tag{3.30}$$

Similarly, we have

$$g_t(x', u) \geq \tilde{T}^{L,t,s}_{x',\tilde{x}} \circ \tilde{T}^{L,s,t}_{\tilde{x},x}\left(g_t(x, u)\right). \tag{3.31}$$

We define

$$\begin{aligned} T^U_{t,x',x}(y) &\equiv \begin{cases} \inf_{\tilde{x}}\{\tilde{T}^{U,t,s}_{x',\tilde{x}} \circ \tilde{T}^{U,s,t}_{\tilde{x},x}(y)\} & \text{if } x \neq x' \\ y & \text{if } x = x' \end{cases} \\ T^L_{t,x',x}(y) &\equiv \begin{cases} \sup_{\tilde{x}}\{\tilde{T}^{L,t,s}_{x',\tilde{x}} \circ \tilde{T}^{L,s,t}_{\tilde{x},x}(y)\} & \text{if } x \neq x' \\ y & \text{if } x = x' \end{cases}. \end{aligned}$$

Then, these functions satisfy (3.23). □

Corollary 3.3. Similarly to (3.30) and (3.31), for $t \neq s$, we have

$$
\begin{aligned}
g_t(x', u) &\leq \tilde{T}^{U,t,s}_{x',\tilde{x}} \circ \tilde{T}^{U,s,t}_{\tilde{x},x} \left(g_t(x, u) \right), \\
g_t(x', u) &\geq \tilde{T}^{L,t,s}_{x',\tilde{x}} \circ \tilde{T}^{L,s,t}_{\tilde{x},x} \left(g_t(x, u) \right),
\end{aligned}
$$

where

$$
\begin{aligned}
\tilde{T}^{U,t,s}_{x_t,x_s}(y) &\equiv \inf_{\mathbf{x}_{-(t,s)}} Q^{+}_{Y_t|(X_t,X_s,\mathbf{X}_{-(t,s)})=(x_t,x_s,\mathbf{x}_{-(t,s)})} \left(F^{+}_{Y_s|(X_t,X_s,\mathbf{X}_{-(t,s)})=(x_t,x_s,\mathbf{x}_{-(t,s)})}(y) \right), \\
\tilde{T}^{L,t,s}_{x_t,x_s}(y) &\equiv \sup_{\mathbf{x}_{-(t,s)}} Q^{-}_{Y_t|(X_t,X_s,\mathbf{X}_{-(t,s)})=(x_t,x_s,\mathbf{x}_{-(t,s)})} \left(F^{-}_{Y_s|(X_t,X_s,\mathbf{X}_{-(t,s)})=(x_t,x_s,\mathbf{x}_{-(t,s)})}(y) \right),
\end{aligned}
$$

and $\mathbf{X}_{-(t,s)}$ denotes a vector of $\mathbf{X}$ except X_t and X_s. Hence,

$$
\hat{T}^{U}_{t,x',x}(y) \equiv \begin{cases} \inf_{s \neq t, \tilde{x} \in \mathcal{X}_s} \{ \tilde{T}^{U,t,s}_{x',\tilde{x}} \circ \tilde{T}^{U,s,t}_{\tilde{x},x}(y) \} & \text{if } x \neq x' \\ y & \text{if } x = x' \end{cases}
$$

and

$$
\hat{T}^{L}_{t,x',x}(y) \equiv \begin{cases} \sup_{s \neq t, \tilde{x} \in \mathcal{X}_s} \{ \tilde{T}^{L,t,s}_{x',\tilde{x}} \circ \tilde{T}^{L,s,t}_{\tilde{x},x}(y) \} & \text{if } x \neq x' \\ y & \text{if } x = x' \end{cases}
$$

satisfy inequality (3.23). Define

$$
\begin{aligned}
\hat{G}^{L}_{t,x}(y) &\equiv \int F^{+}_{Y_t|X_t=x'} \left(\hat{T}^{U}_{t,x',x}(y) \right) dF_{X_t}(x'), \\
\hat{G}^{U}_{t,x}(y) &\equiv \int F^{-}_{Y_t|X_t=x'} \left(\hat{T}^{L}_{t,x',x}(y) \right) dF_{X_t}(x'). \qquad (3.32)
\end{aligned}
$$

By a similar argument to the proof for Theorem 3.2, we have $g_t(x, u) \geq \inf\{y \in [\underline{y}, \overline{y}] : G^{L}_{t,x}(y) \geq u\}$ and $g_t(x, u) \leq \sup\{y \in [\underline{y}, \overline{y}] : G^{U}_{t,x}(y) \leq u\}$. □

3.8 Appendix: Auxiliary Lemmas

Lemma 3.1. *Under Assumptions 3.C3 and 3.C4, $\|D_\theta\|_\mu$ is continuous in θ.*

Proof. By Assumption 3.C4, the density $f_{Y_t|\mathbf{X}}(y|\mathbf{x})$ is bounded above by a constant C. For any θ', θ and v, we have

$$
\begin{aligned}
& |D_{\theta'}(v) - D_\theta(v)| \\
\leq\ & 2\max_t |E\left[(\mathbf{1}\{Y_t \leq g_t(X_t, v_u; \theta')\} - \mathbf{1}\{Y_t \leq g_t(X_t, v_u; \theta)\})\,\mathbf{1}\{\mathbf{X} \leq v_\mathbf{x}\}\right]| \\
\leq\ & 2\max_t |E\left[\left(F_{Y_t|\mathbf{X}}\left(g_t(X_t, v_u; \theta')|\mathbf{X}\right) - F_{Y_t|\mathbf{X}}\left(g_t(X_t, v_u; \theta)|\mathbf{X}\right)\right)\mathbf{1}\{\mathbf{X} \leq v_\mathbf{x}\}\right]| \\
\leq\ & 2\max_t \int \left| \int_{g_t(x_t, v_u; \theta)}^{g_t(x_t, v_u; \theta')} f_{Y_t|\mathbf{X}}(y|\mathbf{x}) dy \right| dF_\mathbf{X}(\mathbf{x}) \\
\leq\ & 2C\max_t \int |g_t(x_t, v_u; \theta') - g_t(x_t, v_u; \theta)|\, dF_{X_t}(x_t) \leq 2CK\|\theta' - \theta\|.
\end{aligned}
$$

Hence, $|\|D_{\theta'}\|_\mu - \|D_\theta\|_\mu| \leq \|D_{\theta'} - D_\theta\|_\mu \leq 2CK\|\theta' - \theta\|$, which implies the continuity of $\|D_\theta\|_\mu$. □

Lemma 3.2. *Under Assumptions 3.C3, 3.C4, and 3.C5,*

$$\sup_{\theta, v} \left| \hat{D}_{n,\theta}(v) - D_\theta(v) \right| = o_{a.s.}(1), \tag{3.33}$$

and for any $\delta_n \downarrow 0$

$$\sup_{v \in \mathcal{V}, \|\theta - \theta_0\| < \delta_n} \left| \sqrt{n}\left(\hat{D}_{n,\theta}(v) - D_\theta(v)\right) - \sqrt{n}\left(\hat{D}_{n,\theta_0}(v) - D_{\theta_0}(v)\right) \right| = o_p(1). \tag{3.34}$$

Proof. The collection of indicator functions $\{\mathbf{x} \mapsto \mathbf{1}\{\mathbf{x} \leq v_\mathbf{x}\} : v_\mathbf{x} \in \mathcal{X}_{12}\}$ is a VC-class. By Assumption 3.C5, the collection of indicator functions for subgraphs of $\{g_t(\cdot, v_u; \theta) : \theta \in \Theta, v_u \in \mathcal{U}\}$ is also a VC-class. By reference to Examples 2.10.7 and 2.10.8 in van der Vaart and Wellner (1996), $\{A_\theta^v : \theta \in \Theta, v \in \mathcal{V}\}$ is P-Donsker, and also P-Glivenko–Cantelli. Hence, we have (3.33) and for any $\delta_n \downarrow 0$

$$
\begin{aligned}
& \sup_{(\theta,v),(\theta',v'):\mathbb{P}(A_\theta^v - A_{\theta'}^{v'})^2 < \delta_n} |\mathbb{G}_n A_\theta^v - \mathbb{G}_n A_{\theta'}^{v'}| = o_p(1) \\
\Rightarrow\ & \sup_{v,\theta,\theta_0:\mathbb{P}(A_\theta^v - A_{\theta_0}^v)^2 < \delta_n} |\mathbb{G}_n A_\theta^v - \mathbb{G}_n A_{\theta_0}^v| = o_p(1).
\end{aligned}
$$

Then, we have

$$
\begin{aligned}
& \mathbb{P}(A_\theta^v - A_{\theta_0}^v)^2 \\
\leq\ & E\left[\{(\mathbf{1}\{Y_1 \leq g_1(X_1, v_u; \theta)\} - \mathbf{1}\{Y_1 \leq g_1(X_1, v_u; \theta_0)\}) \right. \\
& \left. -(\mathbf{1}\{Y_2 \leq g_2(X_2, v_u; \theta)\} - \mathbf{1}\{Y_2 \leq g_2(X_2, v_u; \theta_0)\})\}^2\right] \\
\leq\ & 4 \max_t E\left[|\mathbf{1}\{Y_t \leq g_t(X_t, v_u; \theta)\} - \mathbf{1}\{Y_t \leq g_t(X_t, v_u; \theta_0)\}|\right] \\
=\ & 4 \max_t E\left[|\mathbf{1}\{g_t(X_t, v_u; \theta_0) < Y_t \leq g_t(X_t, v_u; \theta)\} + \mathbf{1}\{g_t(X_t, v_u; \theta)\} < Y_t \leq g_t(X_t, v_u; \theta_0)\}|\right] \\
\leq\ & 8 \max_t \int \left|F_{Y_t|\mathbf{X}}\left(g_t(x_t, v_u; \theta)|\mathbf{x}\right) - F_{Y_t|\mathbf{X}}\left(g_t(x_t, v_u; \theta_0)|\mathbf{x}\right)\right| dF_{\mathbf{X}}(\mathbf{x}) \\
\leq\ & 8CK\|\theta - \theta_0\|.
\end{aligned}
$$

Because $\|\theta - \theta_0\| \to 0$ implies that $\mathbb{P}(A_\theta^v - A_{\theta_0}^v)^2 \to 0$, we have (3.34). □

Lemma 3.3. *Under Assumptions 3.C4, 3.N2, and 3.N3, $D_\theta(v)$ is continuously differentiable in θ in a neighborhood of θ_0 for all v, and $|D_\theta(v) - D_{\theta_0}(v) - \Gamma_0(v)'(\theta - \theta_0)| = o(\|\theta - \theta_0\|)$ uniformly over v.*

Proof. First, we show continuous differentiability of $D_\theta(v)$. For all v and θ in the neighborhood,

$$
\begin{aligned}
\nabla_\theta D_\theta(v) &= \nabla_\theta E\left[\left(F_{Y_1|\mathbf{X}}(g_1(X_1, v_u; \theta)|\mathbf{X}) - F_{Y_2|\mathbf{X}}(g_2(X_2, v_u; \theta)|\mathbf{X})\right) \mathbf{1}\{\mathbf{X} \leq v_\mathbf{x}\}\right] \\
&= \nabla_\theta \int_{\{\mathbf{x} \leq v_\mathbf{x}\}} \left(F_{Y_1|\mathbf{X}}(g_1(x_1, v_u; \theta)|\mathbf{x}) - F_{Y_2|\mathbf{X}}(g_2(x_2, v_u; \theta)|\mathbf{x})\right) dF_\mathbf{X}(\mathbf{x}).
\end{aligned}
$$

Let C be a constant such that $f_{Y_t|\mathbf{X}}(y|\mathbf{x}) \leq C$. Because $|f_{Y_t|\mathbf{X}}(g_t(x_t, v_u; \theta)|\mathbf{x})\nabla_\theta g_t(x_t, v_u; \theta)|$ is bounded by the integrable function $C\nabla\bar{g}(x_t)$, we can interchange a differential operator with an integral. Hence, we have

$$
\begin{aligned}
\nabla_\theta D_\theta(v) &= E[f_{Y_1|\mathbf{X}}(g_1(X_1, v_u; \theta)|\mathbf{X})\nabla_\theta g_1(X_1, v_u; \theta)\mathbf{1}\{\mathbf{X} \leq v_\mathbf{x}\}] \\
&\quad - E[f_{Y_2|\mathbf{X}}(g_2(X_2, v_u; \theta)|\mathbf{X})\nabla_\theta g_2(X_2, v_u; \theta)\mathbf{1}\{\mathbf{X} \leq v_\mathbf{x}\}].
\end{aligned}
$$

According to the dominated convergence theorem, $\nabla_\theta D_\theta(v)$ is continuous in θ in a neighborhood of θ_0 for all v.

Next, we show the second statement. Because $D_\theta(v)$ is continuously differentiable in θ, for θ in a neighborhood of θ_0, there exists $\bar{\theta}_v$ between θ and θ_0 such that

$$
\begin{aligned}
|D_\theta(v) - D_{\theta_0}(v) - \Gamma_0(v)'(\theta - \theta_0)| &= |\{\Gamma_{\bar{\theta}_v}(v) - \Gamma_0(v)\}(\theta - \theta_0)| \\
&\leq \|\theta - \theta_0\| \times \sup_{v \in \mathcal{V}} \|\Gamma_{\bar{\theta}_v}(v) - \Gamma_0(v)\|
\end{aligned}
$$

It follows from Assumption 3.N3 (ii) that $\sup_{v \in \mathcal{V}} \|\Gamma_{\bar{\theta}_v}(v) - \Gamma_0(v)\| \to 0$ as $\|\theta - \theta_0\| \to 0$. Hence, we have $|D_\theta(v) - D_{\theta_0}(v) - \Gamma_0(v)'(\theta - \theta_0)| = o(\|\theta - \theta_0\|)$ uniformly over v. □

Lemma 3.4. *Under Assumptions 3.C3 and 3.C5, for every sequence $\{\delta_n\}$ of positive numbers that converges to zero,*

$$\sup_{\|\theta-\theta_0\|<\delta_n} \|\hat{D}_{n,\theta} - D_\theta - \hat{D}_{n,\theta_0}\|_\mu = o_p(n^{-1/2}). \tag{3.35}$$

Proof. Note that

$$\sup_{\|\theta-\theta_0\|<\delta_n} \|\hat{D}_{n,\theta} - D_\theta - \hat{D}_{n,\theta_0}\|_\mu \leq \sup_{v\in\mathcal{V},\|\theta-\theta_0\|<\delta_n} \left|\hat{D}_{n,\theta}(v) - D_\theta(v) - \hat{D}_{n,\theta_0}(v)\right|.$$

By Lemma 3.2, the right-hand side is $o_p(n^{-1/2})$. Hence, (3.35) holds. □

Lemma 3.5. *Under the assumptions for Theorem 3.3, $\hat{\theta}^* \to_{a.s} \theta_0$ for almost all samples $W_1, \cdots, W_n$.*

Proof. By the triangle inequality, for almost all samples $W_1, \cdots, W_n$, we have

$$\begin{aligned}\sup_\theta \|\hat{D}^*_{n,\theta} - D_\theta\|_\mu &\leq \sup_\theta \|\hat{D}^*_{n,\theta} - \hat{D}_{n,\theta}\|_\mu + \sup_\theta \|\hat{D}_{n,\theta} - D_\theta\|_\mu \\ &\leq \sup_{\theta,v} |\hat{D}^*_{n,\theta}(v) - \hat{D}_{n,\theta}(v)| + \sup_{\theta,v} |\hat{D}_{n,\theta}(v) - D_\theta(v)| \to_{a.s.} 0,\end{aligned}$$

since $\{A^v_\theta : \theta \in \Theta, v \in \mathcal{V}\}$ is a Donsker class. The reminder of the proof is same as for Theorem 3.3. □

Lemma 3.6. *Suppose that the assumptions of Theorem 3.4 hold. For every sequence $\{\delta_n\}$ of positive numbers that converges to zero,*

$$\sup_{\|\theta-\theta_0\|<\delta_n} \|\sqrt{n}(\hat{D}^*_{n,\theta} - \hat{D}_{n,\theta}) - \sqrt{n}(\hat{D}^*_{n,\theta_0} - \hat{D}_{n,\theta_0})\|_\mu = o_B(1). \tag{3.36}$$

Proof. The left-hand side of (3.36) is dominated above by

$$\sup_{v,\|\theta-\theta_0\|<\delta_n} |\sqrt{n}(\hat{D}^*_{n,\theta}(v) - \hat{D}_{n,\theta}(v)) - \sqrt{n}(\hat{D}^*_{n,\theta_0}(v) - \hat{D}_{n,\theta_0}(v))|.$$

The bootstrap equicontinuity due to Giné and Zinn (1990) implies that this random variable is $o_B(1)$. Hence, we can obtain (3.36). □

3.9 Appendix: Figures and Tables

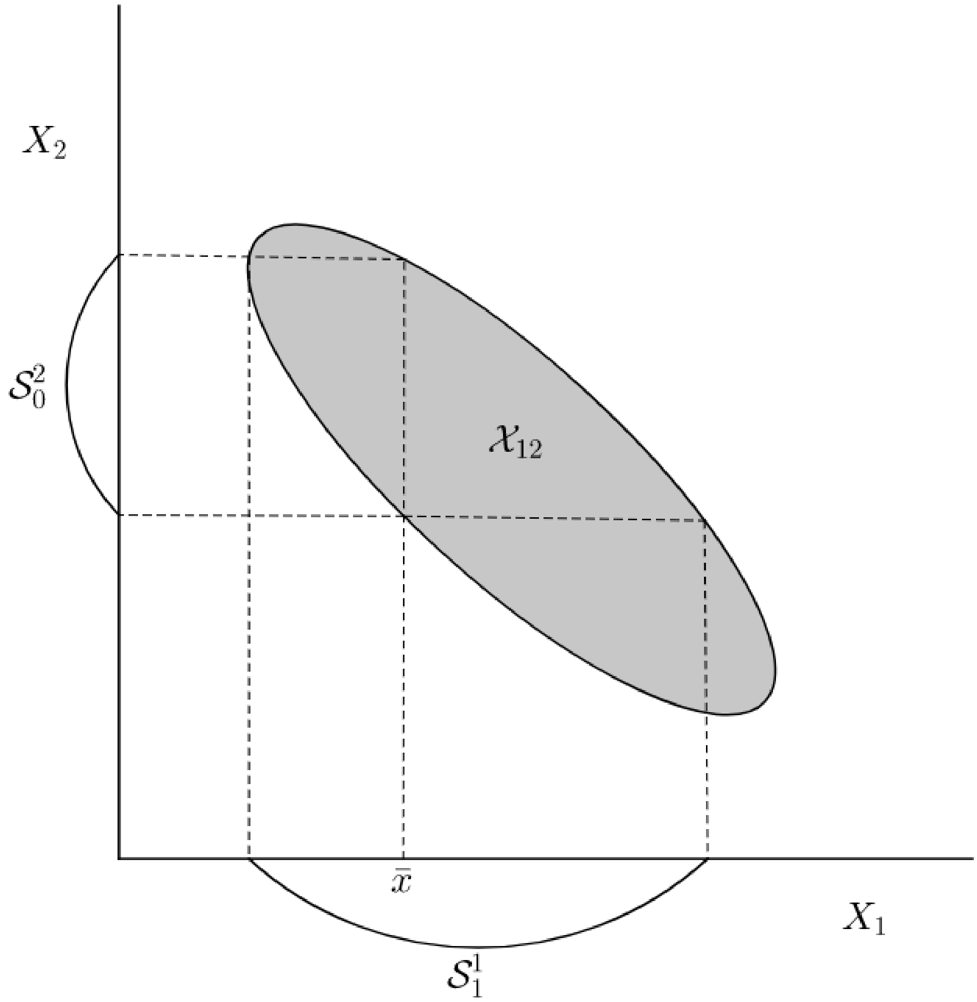

Figure 3.1: Description of $\mathcal{S}_m^t$.

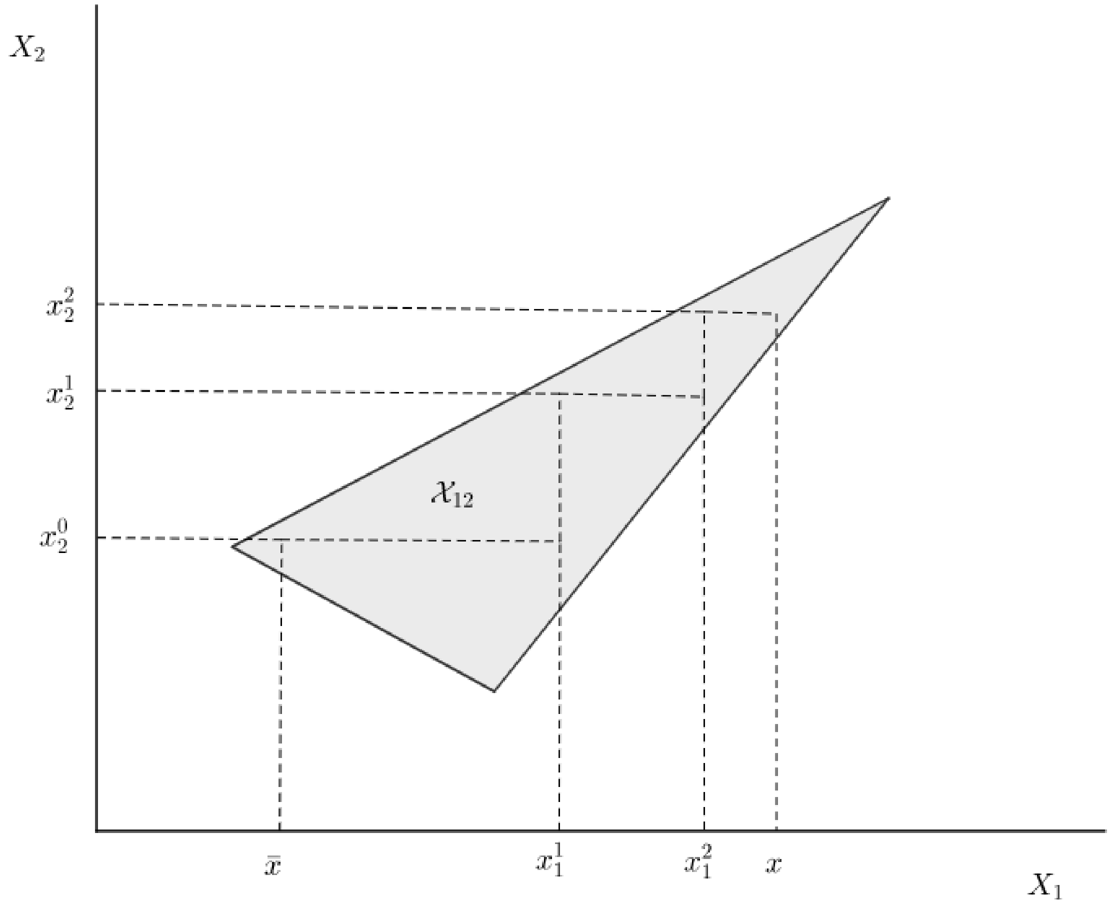

Figure 3.2: Connected support.

Table 3.1: Results of Simulation 3.1(i)

		$N = 400$	$N = 800$	$N = 1600$
	bias	0.0262	0.0195	0.0125
θ_1	std	0.1991	0.1709	0.1121
	mse	0.0403	0.0296	0.0127
	bias	0.0657	0.0733	0.0363
θ_2	std	0.1751	0.1375	0.1045
	mse	0.0350	0.0243	0.0122
	bias	0.0472	0.0605	0.0320
θ_3	std	0.1158	0.1069	0.0889
	mse	0.0231	0.0171	0.0089

Table 3.2: Results of Simulation 3.1(ii)

		$N = 500$	$N = 1000$
	bias	0.0101	0.0087
Our method	std	0.0978	0.0745
	mse	0.0097	0.0056
	bias	0.0042	0.0017
Hoderlein and White	std	0.1175	0.0855
	mse	0.0138	0.0073

Table 3.3: Results of Simulation 3.2

		$N = 400$	$N = 800$	$N = 1600$
θ_1	bias	-0.0025	-0.0031	0.0005
	std	0.0875	0.0604	0.0418
	mse	0.0077	0.0037	0.0018
θ_2	bias	0.0020	0.0022	0.0001
	std	0.0522	0.0351	0.0244
	mse	0.0027	0.0012	0.0006
θ_3	bias	-0.0082	0.0004	-0.0044
	std	0.1837	0.1256	0.0886
	mse	0.0338	0.0158	0.0079
θ_4	bias	0.0043	0.0001	0.0023
	std	0.0839	0.0562	0.0398
	mse	0.0071	0.0032	0.0016
ATE	bias	-0.0025	-0.0014	-0.0011
	std	0.0972	0.0673	0.0457
	mse	0.0095	0.0045	0.0021
QTE25	bias	-0.0333	-0.0303	-0.0343
	std	0.1148	0.0814	0.0565
	mse	0.0143	0.0075	0.0044
QTE50	bias	-0.0017	-0.0016	-0.0009
	std	0.0994	0.0685	0.0474
	mse	0.0099	0.0047	0.0022
QTE75	bias	0.0306	0.0292	0.0328
	std	0.1324	0.0896	0.0618
	mse	0.0185	0.0089	0.0049

Table 3.4: Lower and upper bounds for Example 3.4

x	$g_1^L(x,u)$	$g_1^U(x,u)$	$g_2^L(x,u)$	$g_2^U(x,u)$	$g_1(x,u)$	$g_2(x,u)$
-2	$\mathbf{1}\{u > 0.99\}$	$\mathbf{1}\{u > 0.47\}$	$\mathbf{1}\{u > 0.99\}$	$\mathbf{1}\{u > 0.41\}$	$\mathbf{1}\{u > 0.73\}$	$\mathbf{1}\{u > 0.71\}$
-1	$\mathbf{1}\{u > 0.98\}$	$\mathbf{1}\{u > 0.35\}$	$\mathbf{1}\{u > 0.98\}$	$\mathbf{1}\{u > 0.30\}$	$\mathbf{1}\{u > 0.62\}$	$\mathbf{1}\{u > 0.57\}$
0	$\mathbf{1}\{u > 0.85\}$	$\mathbf{1}\{u > 0.13\}$	$\mathbf{1}\{u > 0.82\}$	$\mathbf{1}\{u > 0.10\}$	$\mathbf{1}\{u > 0.50\}$	$\mathbf{1}\{u > 0.43\}$
1	$\mathbf{1}\{u > 0.64\}$	$\mathbf{1}\{u > 0.02\}$	$\mathbf{1}\{u > 0.59\}$	$\mathbf{1}\{u > 0.01\}$	$\mathbf{1}\{u > 0.38\}$	$\mathbf{1}\{u > 0.29\}$
2	$\mathbf{1}\{u > 0.53\}$	$\mathbf{1}\{u > 0.01\}$	$\mathbf{1}\{u > 0.47\}$	$\mathbf{1}\{u > 0.01\}$	$\mathbf{1}\{u > 0.27\}$	$\mathbf{1}\{u > 0.18\}$

Chapter 4

Panel Data Quantile Regression for Treatment Effect Models

4.1 Introduction

In the literature of program evaluation, it is important to learn about the distributional effects beyond the average effects of the treatment. Policy-makers are likely to prefer a policy that tends to increase outcomes in the lower tail of the outcome distribution to one that tends to increase outcomes in the middle or upper tail of the outcome distribution. One way to capture such effects is to compute the quantiles of the distribution of treated and control potential outcomes. Then, the parameter of interest is the quantile treatment effects (QTE) or the quantile treatment effects on the treated (QTT). For example, Abadie, Angrist, and Imbens (2002) estimates the distributional impacts of the Job Training Partnership Act (JTPA) program on earnings. They show that for women, the JTPA program had the largest proportional impact at low quantiles, but for men, the training impact was largest in the upper half of the distribution, with no significant effect on lower quantiles. Their result could not have been revealed using mean impact analysis. Empirical researchers estimated the distributional effects such as QTE or QTT in many areas of empirical economic research: e.g. Chernozhukov and Hansen (2004) estimate the QTE of participation in a 401(k) plan on several measures of wealth; James, Lahti, and Hoynes (2006) estimate the QTE of welfare reforms on earnings, transfers, and income; Martincus and Carballo (2010) estimate the QTE of trade promotion activities; and Havnes and Mogstad (2015) and Kottelenberg and Lehrer (2017) estimate the QTT of universal child care.

There is a rich literature on the identification and estimation of the QTE and QTT. Firpo (2007) shows the identification and estimation of the QTE parameters under unconfoundedness. Abadie et al. (2002), Chernozhukov and Hansen (2005), Chernozhukov and Hansen (2006), and Frölich and Melly (2013) show how instrumental variables can be used to identify the QTE. Athey and Imbens (2006), Melly and Santangelo (2015), and Callaway and Li (2017) provide the identification and estimation results for the QTT in the difference-in-differences (DID) setting by using repeated cross sections or panel data.

In this chapter, we consider the identification and estimation of the QTE by using panel data. We show that the QTE is identified under the rank invariance and rank stationarity assumptions. Our model corresponds to the change-in-changes (CIC) model proposed by Athey and Imbens (2006) in the DID setting. We generalize the CIC model and propose a tractable estimator of the QTE.

Athey and Imbens (2006) suggest the CIC model as an alternative to the DID model. The CIC model allows for the estimation of the potential outcomes distribution and captures the heterogeneous effects of the treatment on the outcomes. However, there are two problems with the CIC model. First, there is a lack of a tractable estimator in the presence of covariates. According to Lechner (2011) and Kottelenberg and Lehrer (2017), there have been a few applications of the CIC model for this reason. Second, the CIC estimator does not work when the treatment is continuous. Although Athey and Imbens (2006) provide extensions to settings with multiple groups and multiple time periods, they do not consider the case where the treatment is continuous.

Athey and Imbens (2006) provide nonparametric and semiparametric strategies in the presence of covariates. If the dimensionality of the observed covariates is high, the nonparametric strategy would be difficult to implement. Although the semiparametric strategy is more tractable, it assumes that the effects of the observed covariates do not depend on the unobserved factor, and the observed covariates are independent of the unobserved factor conditional on the treatment. On the contrary, our estimation method allows the effects of the observed covariates to depend on the unobserved factor and does not require the conditional independence between the observed covariates and the unobserved factor.

Melly and Santangelo (2015) and Kottelenberg and Lehrer (2017) also consider the estimation of the CIC model in the presence of covariates. Melly and Santangelo (2015) suggest a flexible semiparametric estimator based on quantile regression. They estimate the conditional distribution of outcomes for both treatment and control groups and both periods by using quantile regression, and then apply the changes-in-changes transformations. Kottelenberg and Lehrer (2017) rely on Firpo (2007)'s extension to quantiles of the inverse propensity scores method.

Athey and Imbens (2006) do not consider the case where the treatment variable is continuous. There are, however, many empirical applications where the treatment is continuous. For example, many researchers estimated the effects of class size on children's test score by using panel data. In this case, the policy maker may be interested in the effect of class size on the lower tail of the distribution of children's test score. Since Athey and Imbens (2006) only consider the DID setting, we cannot extend their estimation approach to the continuous treatment case directly.

We employ two key assumptions: the rank invariance and rank stationarity assumptions. The rank invariance assumption is introduced by Chernozhukov and Hansen (2005). This assumption implies that a scalar unobserved factor determines the potential outcomes across treatment states. As discussed in Chernozhukov and Hansen (2005), although the rank invariance is restrictive, we can relax this assumption. The rank stationarity assumption implies that the conditional distribution of the unobserved factor given explanatory variables and covariates does not change over time. In the literature of nonseparable panel data models, similar assumptions are employed by Athey and Imbens (2006), Hoderlein and White (2012), Graham and Powell (2012), D'Haultfoeuille et al. (2013), Chernozhukov et al. (2013), Chernozhukov et al. (2015), and Ishihara (2019). Our identification approach is essentially the same as that of Ishihara (2019).

We propose a two-step estimation method based on the quantile regression and minimum distance method. Ishihara (2019) considers a similar model and proposes a minimum distance estimator. However, the optimization of that estimator is computationally demanding when the dimensionality of covariates is high. To solve this problem, we use the quantile regression in the first step. Using the quantile regression, we can obtain the second stage estimator by optimizing the objective function over low dimensional parameters. This two-step estimation method is similar to the instrumental variable quantile regression proposed by Chernozhukov and Hansen (2006).

There is an alternative approach that estimates the distributional effects by using panel data. Callaway and Li (2017) provide the identification and estimation results for the QTT under a straightforward extension of the most common DID assumption. To identify the QTT, they employ two key assumptions, the Distributional Difference-in-Differences Assumption and the Copula Stability Assumption. The first assumption means the distribution of the change in potential untreated outcomes does not depend on whether or not the individual belongs to the treatment or the control group. The second assumption means the copula between the change in the untreated potential outcomes for the treated group and the initial untreated outcome for the treated group is stable over time.

The rest of the paper is organized as follows. Section 4.2 introduces the assumptions employed in this study and shows that our model is nonparametrically identified. In Section 4.3, we propose a two-step estimator and discuss its consistency and asymptotic normality. Section 4.4 contains the results of several Monte Carlo simulations. Section

4.5 illustrates the use of the derived estimator through a brief empirical example. Section 4.6 concludes. The proofs of the theorems and auxiliary lemmas are collected in the Appendix.

4.2 Model and Identification

First, in Section 4.2.1, we introduce the CIC model proposed by Athey and Imbens (2006) and discuss their estimation strategy. Next, in section 4.2.2, we propose the identification of the generalized CIC models.

Throughout this paper, for any random variables V and W, let $F_{V|W}$ denote the conditional distribution function of V conditional on W.

4.2.1 The Change-in-Changes Model

First, we introduce the CIC model proposed by Athey and Imbens (2006). We assume that an individual belongs to a group $G \in \{0, 1\}$ (where group 1 is the treatment group) and is observed in period $T \in \{0, 1\}$. Then, only individuals in group 1 in period 1 are treated. Hence, $I \equiv G \times T$ is an indicator for the treatment. Let Y^N denote the potential outcome if the individual does not receive the treatment, and let Y^I denote the potential outcome if the individual receives the treatment. Then, the realized outcome Y satisfies the following equation:

$$Y = Y^N \cdot (1 - I) + Y^I \cdot I.$$

Athey and Imbens (2006) impose the following four assumptions:

Assumption AI1. *The outcome of an individual in the absence of intervention satisfies the relationship* $Y^I = h(U, T)$.

Assumption AI2. $h(u, t)$ *is strictly increasing in* u *for* $t \in \{0, 1\}$.

Assumption AI3. *We have* $U \perp\!\!\!\perp T|G$.

Assumption AI4. *We have* $supp(U|G = 1) \subset supp(U|G = 0)$.

Under these assumptions, Athey and Imbens (2006) show that the distribution of $Y^N|G = 1, T = 1$ is identified and

$$F_{Y^N|G=1,T=1}(y) = F_{Y|G=1,T=0}\left(F^{-1}_{Y|G=0,T=0}\left(F_{Y|G=0,T=1}(y)\right)\right), \tag{4.1}$$

where $F^{-1}_{Y|G=g,T=t}$ is the conditional quantile function.

If we are interested in the effect of the intervention on the treated, we need to impose some additional assumptions. Similar to (4.1), Athey and Imbens (2006) show that in addition to Assumptions AI1–AI3, if we have $Y^I = h^I(U, T)$, $h^I(u, t)$ is strictly increasing in u, and $supp(U|G = 1) = supp(U|G = 0)$, then the distribution of $Y^I|G = 0, T = 1$ is identified and

$$F_{Y^I|G=0,T=1}(y) = F_{Y|G=0,T=0}\left(F^{-1}_{Y|G=1,T=0}\left(F_{Y|G=1,T=1}(y)\right)\right). \tag{4.2}$$

Therefore, we can identify the QTE and easily estimate the QTE by using a sample analogue.

4.2.2 Assumptions and Identification

In this section, we propose a model that generalizes the CIC model and introduce the assumptions employed in this paper. We consider the following potential outcome framework. Potential outcomes are indexed against the potential values x of the treatment variable $X_{it} \in \mathbb{R}^{d_X}$, and denoted by $Y_{it}(x)$. Then, we cannot observe $Y_{it}(x)$ directly, and the observed outcome is $Y_{it} \equiv Y_{it}(X_t)$. We consider the following model of potential outcomes:

$$Y_{it}(x) \quad = \quad q_t(x, Z_{it}, U_{it}), \quad i = 1, \cdots, n, \quad t = 1, \cdots, T,$$

where $q_t(x, z, \tau)$ is strictly increasing in τ, $Z_{it} \in \mathbb{R}^{d_Z+1}$ is a vector of covariates that are independent of U_{it}, and $U_{it} \in \mathbb{R}$ has the marginal uniform distribution. The CIC model considers the case of repeated cross sections and hence treats the time period as a random variable. However, in this model, we treat the time period as a fixed value because we are considering a case with panel data. Suppose that $\mathbf{Y}_i = (Y_{i1}, \cdots, Y_{iT})'$, $\mathbf{X}_i = (X'_{i1}, \cdots, X'_{iT})'$, and $\mathbf{Z}_i = (Z'_{i1}, \cdots, Z'_{iT})'$ are observable. Define $W_{it} \equiv (Y_{it}, X'_{it}, Z'_{it})'$ and $\mathbf{W}_i \equiv (W'_{i1}, \cdots, W'_{iT})'$. Let $\mathcal{X}_t$, $\mathcal{Z}_t$, $\mathcal{X}_{1,\cdots,T}$, and $\mathcal{Z}_{1,\cdots,T}$ denote the support of X_{it}, Z_{it}, $\mathbf{X}_i$, and $\mathbf{Z}_i$.

Remark 4.1 (Connection with the CIC model)**.** *In the standard DID setting, this model is essentially the same as the CIC model. To see this, we consider the same setting as that of Section 4.2.1. Then, the support of (X_{i0}, X_{i1}) becomes $\{(0, 0), (0, 1)\}$, and a variable $G_i \equiv X_{i1}$ denotes an indicator of the treatment group. We assume that there are no covariates. In this setting, $Y_{it}(0)$ and $Y_{it}(1)$ correspond to Y^N and Y^I, respectively. This implies that*

$$\begin{aligned} Y^N|T = t &\overset{d}{=} q_t(0, U_t), \\ Y^I|T = t &\overset{d}{=} q_t(1, U_t). \end{aligned}$$

Hence, by assuming that $U|T = t \overset{d}{=} U_t$, $h(u, t) = q_t(0, u)$, and $h^I(u, t) = q_t(1, u)$, we can think of this model as the CIC model introduced in the previous section.

First, we impose the following two assumptions.

Assumption 4.1 (Unobservable variables)**.** *(i) For all t, U_{it} is uniformly distributed on $(0,1)$ conditional on Z_{it}. (ii) For all t, $\mathbf{x}$, and z, the support of $U_{it}|\mathbf{X}_i = \mathbf{x}, Z_{it} = z$ is $[0,1]$.*

Assumption 4.2 (Continuous variable)**.** *For all t, the quantile function $q_t(x,z,\tau)$ is continuous and strictly increasing in τ for all $x \in \mathcal{X}_t$ and $z \in \mathcal{Z}_t$. If X_{it} is a continuous variable, then we assume that $q_t(x,z,\tau)$ is continuous in x.*

As seen above, Assumption 4.1 (i) implies that the conditional τ-th quantile of $Y_{it}(x)$ given $Z_{it} = z$ is equal to $q_t(x,z,\tau)$. Athey and Imbens (2006) do not assume that the unobserved variable is uniformly distributed. However, when the unobserved variable is continuous, this is a normalization, not a restriction (see, e.g. Matzkin (2003)). Assumption 4.2 rules out the case where outcomes are discrete or censored. Although Athey and Imbens (2006) consider the discrete outcomes, we do not consider the discrete outcome case in this study.

Assumption 4.3 (τ-th rank stationarity)**.** *For all $t \neq s$, $\mathbf{x}$, and z, we have $Pr(U_{it} \leq \tau|\mathbf{X}_i = \mathbf{x}, Z_{it} = z) = Pr(U_{is} \leq \tau|\mathbf{X}_i = \mathbf{x}, Z_{is} = z)$.*

This assumption implies that the probability that the ranking variable U_{it} is less than τ does not change across time conditional on $\mathbf{X}_i$ and Z_{it}.

Assumption 4.3 is a quantile version of the identification condition of the following conventional linear panel data model:

$$Y_{it} = X_{it}'\alpha + A_i + \epsilon_{it},\ E[X_{is}\epsilon_{it}] = 0 \text{ for all } t \text{ and } s,$$

where A_i is a fixed effect and ϵ_{it} is a time-variant unobserved variable. Let $\bar{E}[\cdot|\mathbf{X}_i]$ denote the linear projection on $\mathbf{X}_i$, as in Chamberlain (1982). Chernozhukov et al. (2013) show that above equation is satisfied if and only if there is $\tilde{\epsilon}_{it}$ with

$$Y_{it} = X_{it}'\alpha + \tilde{\epsilon}_{it},\ \bar{E}[\tilde{\epsilon}_{it}|\mathbf{X}_i] = \bar{E}[\tilde{\epsilon}_{is}|\mathbf{X}_i] \text{ for all } t \text{ and } s.$$

On the contrary, if the quantile function is linear and there are no covariates, then we can rewrite the model to

$$Y_{it} = X_{it}'\alpha(\tau) + \epsilon_{it}(\tau),$$

where $\epsilon_{it}(\tau) = X_{it}'(\alpha(U_{it}) - \alpha(\tau))$. Then, under Assumption 4.3, $\epsilon_{it}(\tau)$ satisfies $F_{\epsilon_t(\tau)|\mathbf{X}}(0|\mathbf{x}) = F_{\epsilon_s(\tau)|\mathbf{X}}(0|\mathbf{x})$ for all $t \neq s$ and $\mathbf{x}$. Hence, this assumption is a quantile version of the identification condition of the conventional linear panel data model.

If Assumption 4.3 holds for all $\tau \in (0,1)$, then the following assumption is satisfied:

Assumption 4.3' (Rank stationarity). *For all $t \neq s$, $\mathbf{x}$, and z, we have $U_{it}|\mathbf{X}_i = \mathbf{x}, Z_{it} = z \overset{d}{=} U_{is}|\mathbf{X}_i = \mathbf{x}, Z_{is} = z$.*

In the literature of nonseparable panel data models, similar assumptions are employed by Athey and Imbens (2006), Hoderlein and White (2012), Graham and Powell (2012), D'Haultfoeuille et al. (2013), Chernozhukov et al. (2013), Chernozhukov et al. (2015), and Ishihara (2019). Chernozhukov et al. (2013) refer to these assumptions as "time is randomly assigned" or "time is an instrument".

Remark 4.2 (Connection with the CIC model (continued)). *Consider the standard DID setting in Remark 1. When $\tau = 0.5$, Assumption 4.3 implies $P(Y_{i0}(0) \leq \mathrm{med}(Y_{i0}(0))|G_i = g) = P(Y_{i1}(0) \leq \mathrm{med}(Y_{i1}(0))|G_i = g)$ for all $g = 0, 1$, where $\mathrm{med}(Y_{it}(0))$ is the median of $Y_{it}(0)$. This does not imply that $P(Y_{it}(0) \leq \mathrm{med}(Y_{it}(0))|G_i = 0) = P(Y_{it}(0) \leq \mathrm{med}(Y_{it}(0))|G_i = 1)$, and hence this allows that the treatment group contains more high-ability people than the control group.*

Assumption 4.3' is equivalent to Assumption AI3. Assumption AI3 is satisfied if and only if we have $U|T = 0, G = g \overset{d}{=} U|T = 1, G = g$ for all g. On the contrary, Assumption 4.3' implies that $U_{it}|G_i = 0 \overset{d}{=} U_{it}|G_i = 1$ for all t. Because we have $U|T = t \overset{d}{=} U_{it}$, Assumption 4.3' is the same as Assumption AI3.

Define the sets $\mathcal{S}_t^m(\overline{x})$ in the following manner. First, define $\mathcal{S}_t^0(\overline{x}) \equiv \{\overline{x}\}$. For $m = 1, 2, \cdots$, we define

$$\begin{aligned} \mathcal{S}_t^m(\overline{x}) \equiv \{x \in \mathcal{X}_t \;\; : \;\; & \text{there exist } (x_t, x_s) \in \mathcal{X}_{t,s} \\ & \text{such that } x_t \in \mathcal{S}_t^{m-1}(\overline{x}) \text{ and } (x, x_s) \in \mathcal{X}_{t,s}.\}. \end{aligned}$$

When $T = 2$, we have

$$\mathcal{S}_1(\overline{x}) = \bigcup_{x_2 \in \mathcal{X}_2(\overline{x})} \mathcal{X}_1(x_2),$$

where $\mathcal{X}_1(x)$ and $\mathcal{X}_2(x)$ are the support of $X_{i1}|X_{i2} = x$ and $X_{i2}|X_{i1} = x$, respectively.

Assumption 4.4 (Support condition). *(i) For all t, $\overline{\cup_{n=0}^{\infty} \mathcal{S}_t^n(\overline{x})} = \mathcal{X}_t$ holds for some $\overline{x} \in \mathcal{X}_t$. (ii) The support of $\mathbf{X}_i|Z_{it} = z$ is equal to $\mathcal{X}_{1,\cdots,T}$ for all $z \in \mathcal{Z}_t$, and $\mathcal{Z}_t = \mathcal{Z}$ for all t.*

Assumption 4.4 (i) rules out the case where the endogenous variable does not change over time, but this assumption is satisfied in many cases. For example, Ishihara (2019) shows that this assumption holds when $\mathcal{X}_{1,2} = \{(0,0), (0,1)\}$ or the interior of $\mathcal{X}_{1,2}$ is connected. Assumption 4.4 (ii) does not require that exogenous variables change over time. Therefore, we can include the time-invariant variables into the covariates.

Under these assumptions, we can show that $q_t(x, z, \tau)$ is nonparametrically identified. The following proposition is essentially the same as Corollary 1 in Ishihara (2019).

Proposition 4.1. *Suppose that $Y_{it} = q_t(X_{it}, Z_{it}, U_{it})$ holds. Under Assumptions 4.1, 4.2, 4.3, and 4.4, $q_t(x, z, \tau)$ is point identified for all $x \in \mathcal{X}_t$ and $z \in \mathcal{Z}_t$. Furthermore, under Assumptions 4.1, 4.2, 4.3', and 4.4, q_t is point identified.*

Remark 4.3 (Connection with the CIC model (continued))**.** *Consider the standard DID setting with covariates. Then, under Assumptions 4.1, 4.2, 4.3', and 4.4, we have the following equations:*

$$\begin{aligned} F_{Y_1(0)|G=1,Z_1=z}(y) = & \quad F_{Y_0|G=1,Z_0=z}\left(F^{-1}_{Y_0|G=0,Z_0=z}\left(F_{Y_1|G=0,Z_1=z}(y)\right)\right), \\ F_{Y_1(1)|G=0,Z_1=z}(y) = & \quad F_{Y_0|G=0,Z_0=z}\left(F^{-1}_{Y_0|G=1,Z_0=z}\left(F_{Y_1|G=1,Z_1=z}(y)\right)\right). \end{aligned} \tag{4.3}$$

These results are essentially the same as (4.1) and (4.2).

4.3 Estimation and Inference

4.3.1 A Two-Step Estimator

We focus on the following linear-in-parameters model:

$$q_t(x, z, \tau) = x'\alpha(\tau) + z'\beta_t(\tau). \tag{4.4}$$

Hence, the model is written as

$$Y_{it} = X'_{it}\alpha(U_{it}) + Z'_{it}\beta_t(U_{it}), \quad U_{it}|Z_{it} \sim U(0,1).$$

If U_{it} is independent of X_{it} and Z_{it}, then this model becomes a standard linear quantile regression model. In this model, we allow U_{it} to be correlated with X_{it}. This model is similar to the IV quantile regression model proposed by Chernozhukov and Hansen (2006).

By Theorem 1, we can estimate $\alpha(\tau)$ and $\beta_t(\tau)$ by using the following conditions:

$$F_{Y_t|\mathbf{x},Z_t}(x'_t\alpha(\tau) + z'\beta_t(\tau)|\mathbf{x}, z) \;=\; F_{Y_s|\mathbf{x},Z_s}(x'_s\alpha(\tau) + z'\beta_s(\tau)|\mathbf{x}, z) \tag{4.5}$$

$$F_{Y_t - X'_t\alpha(\tau) - Z'_t\beta_t(\tau)|Z_t}(0|z) \;=\; \tau, \tag{4.6}$$

where $\mathbf{x} \equiv (x_1, \cdots, x_T)'$. Hence, we can construct an estimator by using the minimum distance approach. Ishihara (2019) uses a similar identification approach and provides a minimum distance estimator. However, if the dimensionality of the observed covariates is high, the optimization is quite difficult. Hence, we cannot directly apply the minimum distance approach.

We propose the following two-step estimator based on the quantile regression and minimum distance method. Fix $\tau \in (0,1)$. In the first step, we define $\tilde{\beta}_t(a,\tau)$ as

$$\begin{aligned} \tilde{\beta}_t(a,\tau) &\equiv \arg\min_{b_t \in \mathcal{B}_t} \frac{1}{n}\sum_{i=1}^{n} \rho_\tau\left(Y_{it} - X_{it}'a - Z_{it}'b_t\right) \\ &= \arg\min_{b_t \in \mathcal{B}_t} \frac{1}{n}\sum_{i=1}^{n} R_\tau(W_{it}; a, b_t), \end{aligned} \tag{4.7}$$

where $\rho_\tau(u) \equiv (\tau - \mathbf{1}\{u < 0\})u$, $\mathcal{B}_t$ is a parameter space of $\beta_t(\tau)$, and $R_\tau(W_{it}; a, b_t) \equiv \rho_\tau\left(Y_{it} - X_{it}'a - Z_{it}'b_t\right)$. This is an ordinary quantile regression of $Y_{it} - X_{it}'a$ on Z_{it}. Then, from (4.6), $\tilde{\beta}_t\left(\alpha(\tau),\tau\right)$ becomes a consistent estimator of $\beta_t(\tau)$.

In the second step, we construct estimators of $\alpha(\tau)$ and $\beta_t(\tau)$ using the minimum distance approach. Define

$$\begin{aligned} g_t(\mathbf{W}_i; a, b, v_z) \equiv\ & \mathbf{1}\{Y_{it} \le X_{it}'a + Z_{it}'b_t\}\mathbf{1}\{Z_{it} \le v_z\} \\ & -1/T\sum_{s=1}^{T} \mathbf{1}\{Y_{is} \le X_{is}'a + Z_{is}'b_s\}\mathbf{1}\{Z_{is} \le v_z\}, \end{aligned}$$

where $b = (b_1', \cdots, b_T')'$. Then, it follows from (4.5) that for all v_z and $v_{\mathbf{x}}$ we have

$$E\left[g_t(\mathbf{W}_i; \alpha(\tau), \beta(\tau), v_z)\mathbf{1}\{\mathbf{X}_i \le v_{\mathbf{x}}\}\right] = 0, \tag{4.8}$$

where $\beta(\tau) \equiv (\beta_1(\tau)', \cdots, \beta_T(\tau)')'$. Let $\|\cdot\|_\mu$ to be the L_2-norm with respect to a probability measure μ with support $\mathcal{V} \equiv \mathcal{X}_{1,\cdots,T} \times \mathcal{Z}$, that is, $\|f(v_{\mathbf{x}}, v_z)\|_\mu^2 = \int f(v_{\mathbf{x}}, v_z)^2 d\mu(v_{\mathbf{x}}, v_z)$. Using this norm, we obtain the following estimator of $\alpha(\tau)$:

$$\begin{aligned} \hat{\alpha}(\tau) &\equiv \arg\min_{a \in \mathcal{A}} \frac{1}{T}\sum_{t=1}^{T} \left\| \frac{1}{n}\sum_{i=1}^{n} g_t(\mathbf{W}_i; a, \tilde{\beta}(a,\tau), v_z)\mathbf{1}\{\mathbf{X}_i \le v_{\mathbf{x}}\} \right\|_\mu^2 \\ &= \arg\min_{a \in \mathcal{A}} \frac{1}{T}\sum_{t=1}^{T} \|\hat{D}_n^t(v; a, \tilde{\beta}(a,\tau))\|_\mu^2, \end{aligned} \tag{4.9}$$

where $\hat{D}_n^t(v; a, b) \equiv \frac{1}{n}\sum_{i=1}^{n} g_t(\mathbf{W}_i; a, b, v_z)\mathbf{1}\{\mathbf{X}_i \le v_{\mathbf{x}}\}$, $v = (v_{\mathbf{x}}', v_z')'$, and $\mathcal{A}$ is a parameter space of $\alpha(\tau)$. Because $\tilde{\beta}_t\left(\alpha(\tau),\tau\right)$ is a consistent estimator of $\beta_t(\tau)$, we can estimate $\beta_t(\tau)$ by $\hat{\beta}_t(\tau) \equiv \tilde{\beta}_t(\hat{\alpha}_t(\tau),\tau)$.

Our estimator is similar to the estimator proposed by Chernozhukov and Hansen (2006). They consider the IV quantile regression for heterogeneous treatment effect models and simultaneous equations models with nonadditive errors. Similar to Chernozhukov and Hansen (2006), our estimator is attractive from a computational point of view. Since the ordinary quantile regressions are obtained by convex optimization, our first step estimation (4.7) is computationally convenient. Our second step estimation (4.9) requires non-convex optimization, and hence it seems to be computationally demanding. However, we can obtain (4.9) by optimizing the objective function over α-parameter (typically one-dimensional). Therefore, this optimization problem is not so computationally expensive.

Remark 4.4 (Existing estimators of the CIC model with covariates)**.** *Athey and Imbens (2006) provide nonparametric and semiparametric strategies in the presence of covariates. Their nonparametric strategy is based on equations (4.3), and hence they estimate the QTT by estimating the conditional distribution and quantile functions nonparametrically. However, if the dimensionality of covariates is high, the nonparametric strategy would be difficult to implement. In their semiparametric strategy, they assume the following model:*

$$Y^I = h(T, U) + Z'\beta, \qquad Z \perp\!\!\!\perp (U, T)|G.$$

This model assumes that the effects of the observed covariates do not depend on the unobserved factor, and the observed covariates are independent of the unobserved factor conditional on the group. On the contrary, our model allows the effect of Z_{it} *on the outcome to depend on time and the unobserved variable, and does not require statistical independence between* U_{it} *and* Z_{it} *conditional on* G_i*.*

Melly and Santangelo (2015) also consider the estimation of the CIC model with covariates. They suggest a flexible semiparametric estimator based on quantile regression. They estimate the conditional distribution of outcomes for both treatment and control groups and both periods by using quantile regression, and then apply the changes-in-changes transformations (4.3). Hence, they assume that the conditional quantile function of observed outcomes is linear in covariates, that is,

$$F^{-1}_{Y_t|G=g,Z_t}(\tau|z) = z'\beta^g_t(\tau).$$

On the contrary, our model does not assume that the conditional quantile functions are linear. In model (4.4), we have

$$F^{-1}_{Y_t|G=g,Z_t}(\tau|z) = g\alpha(Q_t(\tau|g, z)) + z'\beta_t(Q_t(\tau|g, z)),$$

where $Q_t(\tau|g, z) \equiv F^{-1}_{U_t|G=g,Z_t=z}(\tau)$*. Hence, our model does not require the linearity of the conditional quantile functions because we allow the conditional distribution of* $U_{it}|D_i = d, Z_{it} = z$ *to depend on* z*.*

4.3.2 Consistency and Asymptotic Normality

First, we show that $\alpha(\tau)$ and $\beta(\tau) = (\beta_1(\tau)', \cdots, \beta_T(\tau)')'$ uniquely solves the limit problem. Define

$$\beta_t(a, \tau) \equiv \arg\min_{b_t \in \mathcal{B}_t} E[R_\tau(W_{it}; a, b_t)], \tag{4.10}$$

and

$$\alpha^*(\tau) \equiv \arg\min_{a \in \mathcal{A}} \frac{1}{T} \sum_{t=1}^{T} \|D^t(v; a, \beta(a,\tau))\|_\mu^2,$$

where $\beta(a,\tau) \equiv (\beta_1(a,\tau)', \cdots, \beta_T(a,\tau)')'$ and $D^t(v;a,b) \equiv E[g_t(\mathbf{W}_i; a, b, v_z)\mathbf{1}\{\mathbf{X}_i \leq v_{\mathbf{x}}\}]$. Hence, we show that $\alpha^*(\tau) = \alpha(\tau)$ and $\beta_t(\alpha^*(\tau),\tau) = \beta_t(\tau)$ for all t.

Define $e_t(a,\tau,z) \equiv P(Y_{it} \leq X_{it}'a + Z_{it}'\beta_t(a,\tau)|Z_{it} = z)$. We impose the following assumption:

Assumption 4.5. *(i) For all t, if $a, \tilde{a} \in \mathcal{A}$, $b_t, \tilde{b}_t \in \mathcal{B}_t$, and $(a, b_t) \neq (\tilde{a}, \tilde{b}_t)$, then $x'a + z'b_t$ is not equal to $x'\tilde{a} + z'\tilde{b}_t$ for some x and z. (ii) $E[|Y_{it}|]$ and $E[\|(X_{it}', Z_{it}')'\|]$ are finite for all t. (iii) For all a, τ, and t, $\beta_t(a,\tau)$ uniquely solves (4.10).*

Assumption 4.6. *For all t and $a \in \mathcal{A}$, $e_t(a,\tau,z) = \tau$ for some $z \in \mathcal{Z}$.*

Assumption 4.5 (i) is a usual identification condition. Assumption 4.6 is a technical condition. This assumption is satisfied for many situations. By the proof of Theorem 4.2 in Angrist, Chernozhukov, and Fernández-Val (2006), it follows from the first order condition of (4.7) that $E\left[\left(\mathbf{1}\{Y_{it} \leq X_{it}'a + Z_{it}'\beta_t(a,\tau)\} - \tau\right) Z_{it}\right] = 0$. Hence, $E[e_t(a,\tau,Z_{it})] = \tau$ holds because Z_{it} contains a constant. If $\{e_t(a,\tau,z) : z \in \mathcal{Z}\}$ is an interval, then $e_t(a,\tau,z) = \tau$ for some $z \in \mathcal{Z}$. When Z_{it} has continuous covariates and $e_t(a,\tau,z)$ is continuous in z, it is natural to assume that $\{e_t(a,\tau,z) : z \in \mathcal{Z}\}$ is an interval. Even when all covariates are discrete, if the model is saturated, that is the cardinality of $\mathcal{Z}$ is equal to the dimension of $\beta_t(a,\tau)$, then we have $e_t(a,\tau,z) = \tau$ for all $z \in \mathcal{Z}$.

Theorem 4.1. *Suppose that (4.4) and Assumptions 4.1, 4.2, 4.3', 4.4, 4.5, and 4.6 hold. Then, for all $\tau \in (0,1)$, $\alpha(\tau)$ and $\beta(\tau)$ uniquely solve the limit problems. That is, we have*

$$\frac{1}{T} \sum_{t=1}^{T} \|D^t(v; a, \beta(a,\tau))\|_\mu^2 = 0 \quad \text{and} \quad a \in \mathcal{A} \quad \Leftrightarrow \quad a = \alpha(\tau). \tag{4.11}$$

Next, we show the consistency of $\hat{\alpha}(\tau)$ and $\hat{\beta}(\tau)$. Let $\mathcal{T}$ be a finite subset of $(0,1)$, and we define $J_t^b(a,\tau) \equiv E\left[f_{Y_t - X_t'a|Z_t}(Z_{it}'\beta_t(a,\tau)|Z_{it}) Z_{it} Z_{it}'\right]$ and $J_t^b(\tau) \equiv J_t^b(\alpha(\tau),\tau)$. For example, $\mathcal{T} = \{0.1, \cdots, 0.9\}$.

Assumption 4.7. *(i) $\{\mathbf{W}_i\}_{i=1}^n$ are independent and identically distributed. (ii) $\mathcal{A}$ and $\mathcal{B}_t$ are compact for all t. (iii) For all t, $E[|Y_{it}|] < \infty$ and $\mathcal{X}_{1,\cdots,T}$, and $\mathcal{Z}$ are compact. (iv) For all a, τ, and t, $\beta_t(a,\tau)$ uniquely solves (4.10). (v) The conditional density $f_{Y_t - X_t'a|Z_t}(y|z)$ exists for all $a \in \mathcal{A}$ and t, and $f_{Y_t - X_t'a|Z_t}(y|z)$ is continuous in y and bounded above. (vi) For all t and τ, $J_t^b(a,\tau)$ is full rank uniformly over $a \in \mathcal{A}$, and $J_t^b(a,\tau)$ is continuous in a at $\alpha(\tau)$. (vii) For all t, $F_{Y_t|\mathbf{X},Z_t}(y|\mathbf{x},z)$ is uniformly continuous in y. (viii) $a \mapsto \beta(a,\tau)$ is continuous for all $\tau \in \mathcal{T}$.*

Let $\tilde{\mu}$ be a product measure $\mu \times \mu_T$, where $\mu_T(\{t\}) = 1/T$ for all $t \in \{1, \cdots, T\}$. Similar to $\|\cdot\|_\mu$, let $\|\cdot\|_{\tilde{\mu}}$ denote the L_2-norm with respect to $\tilde{\mu}$. Then, we have $\|D^t(v;a,b)\|_{\tilde{\mu}}^2 = \frac{1}{T}\sum_{t=1}^T \|D^t(v;a,b)\|_\mu^2$ and $\|\hat{D}_n^t(v;a,b)\|_{\tilde{\mu}}^2 = \frac{1}{T}\sum_{t=1}^T \|\hat{D}_n^t(v;a,b)\|_\mu^2$

Theorem 4.2. *Suppose that (4.11) holds for all $\tau \in \mathcal{T}$. Under Assumption 4.7, we have $\|\hat{\alpha}(\tau) - \alpha(\tau)\| \to_p 0$ and $\|\hat{\beta}(\tau) - \beta(\tau)\| \to_p 0$ for all $\tau \in \mathcal{T}$.*

For all t and τ, define

$$
\begin{aligned}
J_t^a(\tau) &\equiv E[f_{Y_t|X_t,Z_t}(X_{it}'\alpha(\tau) + Z_{it}'\beta_t(\tau)|X_{it}, Z_{it})Z_{it}X_{it}'], \\
\Gamma_1^t(v;a,\tau) &\equiv E[f_{Y_t|\mathbf{X},Z_t}(X_{it}'a + Z_{it}'b_t|\mathbf{X}_i, Z_{it})\mathbf{1}\{\mathbf{X}_i \le v_{\mathbf{x}}, Z_{it} \le v_z\}(X_{it} + B_t(a,\tau)'Z_{it})] \\
&\quad -\frac{1}{T}\sum_{s=1}^T E[f_{Y_s|\mathbf{X},Z_s}(X_{is}'a + Z_{is}'b_s|\mathbf{X}_i, Z_{is})\mathbf{1}\{\mathbf{X}_i \le v_{\mathbf{x}}, Z_{is} \le v_z\}(X_{is} + B_s(a,\tau)'Z_{is})], \\
\gamma_2^{t,s}(v;a,b) &\equiv \begin{cases} \frac{T-1}{T}E[f_{Y_t|\mathbf{X},Z_t}(X_{it}'a + Z_{it}'b_t|\mathbf{X}_i, Z_{it})\mathbf{1}\{\mathbf{X}_i \le v_{\mathbf{x}}, Z_{it} \le v_z\}Z_t], & \text{if } s = t \\ -\frac{1}{T}E[f_{Y_s|\mathbf{X},Z_s}(X_{is}'a + Z_{is}'b_s|\mathbf{X}_i, Z_{is})\mathbf{1}\{\mathbf{X}_i \le v_{\mathbf{x}}, Z_{is} \le v_z\}Z_{is}], & \text{if } s \ne t \end{cases}, \\
\Gamma_2^t(v;a,b) &\equiv (\gamma_2^{t,1}(v;a,b)', \cdots, \gamma_2^{t,T}(v;a,b)')',
\end{aligned}
$$

$\Gamma_1^t(v;\tau) \equiv \Gamma_1^t(v;\alpha(\tau),\tau)$, $\Gamma_2^t(v;\tau) \equiv \Gamma_2^t(v;\alpha(\tau),\beta(\tau))$, and $B_t(a,\tau) \equiv \frac{\partial}{\partial a'}\beta_t(a,\tau)$.

Assumption 4.8. *(i) For all $\tau \in \mathcal{T}$ and $a \in \mathcal{A}$, $\alpha(\tau)$ and $\beta_t(a,\tau)$ are inner points of $\mathcal{A}$ and $\mathcal{B}_t$, respectively. (ii) $\{y \mapsto f_{Y_t - X_t'a|Z_t}(y|z) : a \in \mathcal{A}\}$ is equicontinuous for all $z \in \mathcal{Z}$. (iii) The conditional density $f_{Y_t|\mathbf{X},Z_t}(y|\mathbf{x},z)$ exists, and $f_{Y_t|\mathbf{X},Z_t}(y|\mathbf{x},z)$ is uniformly continuous in y and bounded above. (vi) For all t and $\tau \in \mathcal{T}$, $\beta_t(a,\tau)$ is continuously differentiable in a. (v) For all $\tau \in \mathcal{T}$, there exists $c > 0$ such that $\|\Gamma_1^t(v;\tau)'a\|_{\tilde{\mu}} \ge c\|a\|$ for all $a \in \mathbb{R}^{d_X}$.*

Theorem 4.3. *Suppose that (4.11) holds for all $\tau \in \mathcal{T}$. Under Assumptions 4.7 and 4.8,*

$$\sqrt{n}(\hat{\alpha}(\tau) - \alpha(\tau)) = -\Delta_1(\tau)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \{\xi(\mathbf{W}_i;\tau) - \Delta_{12}(\tau)l(\mathbf{W}_i;\tau)\} + o_p(1), \tag{4.12}$$

and

$$
\begin{aligned}
\sqrt{n}(\hat{\beta}_t(\tau) - \beta_t(\tau)) &= o_p(1) + J_t^b(\tau)^{-1}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^n r_\tau(W_{it};\alpha(\tau),\beta_t(\tau))\right. \\
&\quad \left. - J_t^a(\tau)\Delta_1(\tau)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n(\xi(\mathbf{W}_i;\tau) - \Delta_{12}(\tau)l(\mathbf{W}_i;\tau))\right\}, \quad (4.13)
\end{aligned}
$$

where

$$
\begin{aligned}
\xi(\mathbf{W}_i;\tau) &\equiv \frac{1}{T}\sum_{t=1}^T\left[\int \Gamma_1^t(v;\tau)\mathbf{1}\{\mathbf{X} \le v_{\mathbf{x}}, Z_t \le v_z\}d\mu(v)\right]\mathbf{1}\{Y_{it} \le X_{it}'\alpha(\tau) + Z_{it}'\beta_t(\tau)\}, \\
l(\mathbf{W}_i;\tau) &\equiv (J_1^b(\tau)^{-1}r_\tau(W_{i1};\alpha(\tau),\beta(\tau))', \cdots, J_T^b(\tau)^{-1}r_\tau(W_{iT};\alpha(\tau),\beta(\tau))')',
\end{aligned}
$$

$r_\tau(W_{it};a,b_t) \equiv (\tau - \mathbf{1}\{Y_{it} \leq X_{it}'a + Z_{it}'b_t\})\, Z_{it},\ \Delta_1(\tau) \equiv \frac{1}{T}\sum_{t=1}^T \int \Gamma_1^t(v;\tau)\Gamma_1^t(v;\tau)'d\mu(v),$ *and* $\Delta_{12}(\tau) \equiv \frac{1}{T}\sum_{t=1}^T \int \Gamma_1^t(v;\tau)\Gamma_2^t(v;\tau)'d\mu(v)$.

The proof of this theorem is based on the argument of Brown and Wegkamp (2002), Chen, Linton, and Van Keilegom (2003), and Torgovitsky (2017).

When $\mathcal{T} = \{\tau_1, \cdots, \tau_J\}$, Theorem 4.3 implies that

$$\sqrt{n}\begin{pmatrix} \hat{\alpha}(\tau_1) - \alpha(\tau_1) \\ \vdots \\ \hat{\alpha}(\tau_J) - \alpha(\tau_J) \end{pmatrix} \rightsquigarrow N\left(\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma(\tau_1,\tau_1) & \cdots & \Sigma(\tau_1,\tau_J) \\ \vdots & \ddots & \vdots \\ \Sigma(\tau_J,\tau_1) & \cdots & \Sigma(\tau_J,\tau_J) \end{pmatrix}\right),$$

where

$$\Sigma(\tau,\tau') \equiv \Delta_1(\tau)^{-1}V(\tau,\tau')\Delta_1(\tau')^{-1}$$

and $V(\tau,\tau') \equiv E\left[(\xi(\mathbf{W}_i;\tau) - \Delta_{12}(\tau)l(\mathbf{W}_i;\tau))\,(\xi(\mathbf{W}_i;\tau') - \Delta_{12}(\tau')l(\mathbf{W}_i;\tau'))'\right]$. For example, we consider the case where $T=2$, X_{it} and Z_{it} are scalar, and the covariates are time invariant, that is, $Z_{it} = Z_i$. In this case, we have

$$\begin{aligned} \xi(\mathbf{W}_i;\tau) &= \frac{1}{4}\left[\int \gamma_1(v;\tau)\mathbf{1}\{\mathbf{X}_i \leq v_{\mathbf{x}}, Z_i \leq v_z\}d\mu(v)\right](\mathbf{1}\{U_{i1} \leq \tau\} - \mathbf{1}\{U_{i2} \leq \tau\}), \\ \Delta_{12}(\tau)l(\mathbf{W}_i;\tau) &= \frac{1}{4}\left[\int \gamma_1(v;\tau)\gamma_2^1(v;\tau)d\mu(v)\right] J_1^b(\tau)^{-1}(\tau - \mathbf{1}\{U_{i1} \leq \tau\})Z_i \\ &\quad -\frac{1}{4}\left[\int \gamma_1(v;\tau)\gamma_2^2(v;\tau)d\mu(v)\right] J_2^b(\tau)^{-1}(\tau - \mathbf{1}\{U_{i2} \leq \tau\})Z_i, \end{aligned}$$

where

$$\begin{aligned} \gamma_1(v;\tau) \equiv\ & E\big[\big\{f_{Y_1|\mathbf{X},Z}(Y_{i1}(\tau)|\mathbf{X}_i,Z_i)(X_{i1} + B_1(\alpha(\tau),\tau)Z_i) \\ & -f_{Y_2|\mathbf{X},Z}(Y_{i2}(\tau)|\mathbf{X}_i,Z_i)(X_{i2} + B_2(\alpha(\tau),\tau)Z_i)\big\}\,\mathbf{1}\{\mathbf{X}_i \leq v_{\mathbf{x}}, Z_i \leq v_z\}\big] \end{aligned}$$

and $\gamma_2^t(v;\tau) \equiv E\left[f_{Y_t|\mathbf{X},Z}(Y_{it}(\tau)|\mathbf{X}_i,Z_i)\mathbf{1}\{\mathbf{X}_i \leq v_{\mathbf{x}}, Z_i \leq v_z\}Z_i\right]$ for $t = 1,2$. By definition, we have $J_t^b(\tau) > 0$, and $\gamma_2^1(v;\tau)$ and $\gamma_2^2(v;\tau)$ have the same sign. Hence, when U_{i1} and U_{i2} are positively correlated, the variance of $\xi(\mathbf{W}_i;\tau)$ and $\Delta_{12}(\tau)l(\mathbf{W}_i;\tau)$ become small. In the extreme case, when $U_{i1} = U_{i2}$, $\xi(\mathbf{W}_i;\tau)$ is equal to zero.

4.4 Simulations

Simulation 4.1. *The outcome equation is given by*

$$\begin{aligned} Y_{i1} &= (\alpha_1 + \alpha_2 U_{i1})X_{i1} + \beta_{11}Z_{i,1} + \beta_{12}Z_{i,2} + U_{i1}, \\ Y_{i2} &= (\alpha_1 + \alpha_2 U_{i2})X_{i2} + \beta_{21}Z_{i,1} + \beta_{22}Z_{i,2} + U_{i2}, \end{aligned}$$

where $(\alpha_1, \alpha_2, \beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}) = (1, 0.5, 1, 1, 1.5, 1.5)$, $X_{it} = \tilde{X}_{it}^2$, $Z_{i,1}$ *and* $Z_{i,2}$ *are time-invariant covariates, and* $U_{it} = A_i + \tilde{U}_{it}$. *We assume that* $Z_i \equiv (Z_{i,1}, Z_{i,2}) \sim N(\mu_Z, \Sigma_Z)$, $(\tilde{X}_{i1}, \tilde{X}_{i2}, A_i) \sim N(0, \Sigma_{\mathbf{X}A})$, *and* $\tilde{U}_{it} \sim N(0, 1-\rho^2)$, *where* $\rho \in [0,1]$, $\mu_Z = (1,1)'$,

$$\Sigma_Z = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}, \text{ and } \Sigma_{\mathbf{X}A} = \begin{pmatrix} 1 & 0.5 & 0.6\rho \\ 0.5 & 1 & 0.4\rho \\ 0.6\rho & 0.4\rho & \rho^2 \end{pmatrix}.$$

Then, we have $Corr(U_1, U_2) = \rho^2$. *Hence,* U_{i1} *and* U_{i2} *are uncorrelated when* $\rho = 0$, *and* U_{i1} *and* U_{i2} *are perfectly correlated when* $\rho = 1$. *In this setting,* $\alpha(0.25) = 0.66$, $\alpha(0.5) = 1$, *and* $\alpha(0.75) = 1.34$.

Table 4.1 contains the results of this experiment for three different choices of ρ^2, 0.1, 0.5, *and* 0.9, *and two different choices for the sample size,* 1000 *and* 2000. *The number of replications is set at* 1000 *throughout. Table 4.1 shows the bias, the standard deviation, and MSE of the estimates of* $\alpha(\tau)$, $\tau = 0.25, 0.5,$ *and* 0.75. *Table 4.1 shows that the bias, the standard deviation, and MSE decrease in all experiments as the sample size increases. As expected, as* ρ *increases, the standard deviation decreases.*

Table 4.2 contains the coverage probabilities for nonparametric bootstrap confidence intervals of $\alpha(\tau)$ *when* $N = 1000$ *and* $\rho^2 = 0.5$. *These experiments are the result of* 1000 *replications with* 500 *bootstrap samples for each replication. Table 4.2 shows that the nominal and actual coverage probabilities are similar for all settings.*

Simulation 4.2. *We consider the following model. Following usual DID settings, we assume that* $\mathcal{X}_{1,2} = \{(0,0), (0,1)\}$, *that is* $X_{i1} = 0$ *for all* i. *The potential outcome is generated from*

$$\begin{aligned} Y_{i1}(0) &= U_{i1} + (1 + 0.5U_{i1})Z_{i1,1} + (1 + 0.5U_{i1})Z_{i,2}, \\ Y_{i2}(x) &= (0.5 + \alpha U_{i2})x + (1 + 0.5U_{i2}) + (1 + 0.5U_{i2})Z_{i2,1} + (1 + 0.5U_{i2})Z_{i,2}, \end{aligned}$$

where $X_{i2} = \mathbf{1}\{\tilde{X}_i \leq 0\}$, $Z_{it,1} = \tilde{Z}_{it,1}^2$, $Z_{i,2} = \tilde{Z}_{i,2}^2$, *and* $U_{it} = A_i + \tilde{U}_{it}$. *Define* $\tilde{Z}_i \equiv (\tilde{Z}_{i1,1}, \tilde{Z}_{i2,1}, \tilde{Z}_{i,2})'$ *and* $D_i \equiv X_{i2}$. *We assume that* $(\tilde{X}_i, A_i) \sim N(0, \Sigma_{XA})$, $\tilde{Z}_i \sim N(0, \Sigma_Z)$, *and* $\tilde{U}_{it} \sim N(0, 1-\rho^2)$, *where* $\mu_Z = (0.5, 0.5, 0.5)'$,

$$\Sigma_{XA} = \begin{pmatrix} 1 & 0.5\rho \\ 0.5\rho & \rho^2 \end{pmatrix}, \quad \Sigma_Z = \begin{pmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{pmatrix},$$

and $\rho \in [0,1]$. *In this setting,* D_i *and* U_{it} *are correlated unless* $\delta = 0$, *and* $\alpha(\tau) = 0.5 + (\theta_2 - \theta_1)\Phi^{-1}(\tau)$. *Define* $Z_{it} \equiv (Z_{it,1}, Z_{i,2})'$ *for all* t. *Because* $U_{i1}|D = d, Z_{i1} \overset{d}{=} U_{i1}|D = d, Z_{i2}$ *for all* $d = 0, 1$, *we obtain*

$$\begin{aligned} F^{-1}_{Y_{i2}(0)|D=d,Z_{i2}}(\tau|Z_{i2}) - F^{-1}_{Y_{i2}(0)|D=d,Z_{i1}}(\tau|Z_{i1}) &= h_2(0, Q(\tau|d, Z_{i2})) - h_1(0, Q(\tau|d, Z_{i1})) \\ &= 1 + (\theta_1 - 1)Q(\tau|d, Z_{i1}), \end{aligned}$$

where $Q(\tau|d,z) \equiv F^{-1}_{U_{i1}|D=d,Z_{i1}}(\tau|z) = F^{-1}_{U_{i2}|D=d,Z_{i2}}(\tau|z)$*. Hence, when* $\theta_1 = 1$*, this model satisfies the quantile parallel trend assumption. When* $\theta_1 \neq 1$*, we can not estimate the quantile treatment effects on the treated by using the quantile difference-in-differences methods.*

Table 4.3 contains the results of this experiment for three different choices of ρ^2*,* 0.1*,* 0.5*, and* 0.9*, and two different choices for the sample size,* 1000 *and* 2000*. The number of replications is set at* 1000 *throughout. Table 4.3 shows the bias, the standard deviation, and MSE of the estimates of* $\alpha(\tau)$*,* $\tau = 0.25, 0.5,$ *and* 0.75*. The results are similar to that of Simulation 4.1.*

Simulation 4.3. *To compare our estimation method with that of Athey and Imbens (2006), we consider the following model. We assume* $\mathcal{X}_{1,2} = \{(0,0),(0,1)\}$ *and the potential outcome is generated from*

$$
\begin{aligned}
Y_{i1}(0) &= U_{i1}, \\
Y_{i2}(x) &= (0.5 + \alpha U_{i2})x + (1 + 0.5 U_{i2}),
\end{aligned}
$$

where $\alpha = 1$*,* $X_{i2} = \mathbf{1}\{\tilde{X}_i \leq 0\}$*, and* $U_{it} = A_i + \tilde{U}_{it}$*. We assume that* $(\tilde{X}_i, A_i) \sim N(0, \Sigma_{XA})$ *and* $\tilde{U}_{it} \sim N(0, 1-\rho^2)$*, where* Σ_{XA} *is defined in Simulation 2 and* $\rho \in [0,1]$*. In this setting,* $\alpha(\tau) = 0.5 + \Phi^{-1}(\tau)$*.*

Athey and Imbens (2006) show that we can estimate $F_{Y_2(0)|D=1}(y)$ *and* $F_{Y_2(1)|D=0}(y)$ *by*

$$
\begin{aligned}
\hat{F}_{Y_2(0)|D=1}(y) &\equiv \hat{F}_{Y_1|D=1}\left(\hat{F}^{-1}_{Y_1|D=0}\left(\hat{F}_{Y_2|D=0}(y)\right)\right) \text{ and} \\
\hat{F}_{Y_2(1)|D=0}(y) &\equiv \hat{F}_{Y_1|D=0}\left(\hat{F}^{-1}_{Y_1|D=1}\left(\hat{F}_{Y_2|D=1}(y)\right)\right),
\end{aligned}
$$

where $\hat{F}_{Y_t|D=d}(\cdot)$ *and* $\hat{F}^{-1}_{Y_t|D=d}(\cdot)$ *are the empirical distribution and quantile. By using these estimators and empirical distributions of* $Y_2|D=0$ *and* $Y_2|D=1$*, we can estimate the distributions of* $Y_2(0)$ *and* $Y_2(1)$*. Hence, we can obtain an estimator of the QTE. We use this estimator as Athey and Imbens (2006)'s (AI) estimator.*

Table 4.4 shows the bias, the standard deviation, and MSE of our estimator and the AI estimator for three different choice of ρ^2*,* 0.1*,* 0.5*, and* 0.9*. For all settings, the MSE of our estimator is similar to that of the AI estimator. Hence, our estimation method is not worse than that of Athey and Imbens (2006).*

4.5 Application

In this section, we use our method to study the impact of an agricultural insurance program on household production. We use the data employed by Cai (2016) to estimate the QTE of insurance provision on tobacco production.

The empirical analysis is based on data obtained from 12 tobacco production counties in Jiangxi province of China. Across these 12 counties, only tobacco farmers in the county of Guangchang were eligible to buy the tobacco insurance policy. In 2003, the People's Insurance Company of China (PICC) designed and offered the first tobacco production insurance program to households in Guangchang. Hence, we use this county as a treatment group.

The sample includes information on around 3,400 tobacco households during the year 2002 and 2003. Table 1 provides summary statistics for the year 2002. Table 4.5 shows that treatment regions are quite different from control regions in terms of their observed characteristics. For example, control regions include more educated people than treatment regions. The proportion of high school or college educated people in treatment regions is 0.025, but that in control regions is 0.257. This implies that tobacco households in treatment regions are quite different from those in control regions. Hence, controlling the observed characteristics is important to adjust the difference between treatment and control regions.

We estimate the following linear-in-parameter model:

$$\begin{aligned} Y_{i,2002} &= Z_i'\beta_{2002}(U_{i,2002}), \\ Y_{i,2003} &= D_i\alpha(U_{i,2003}) + Z_i'\beta_{2003}(U_{i,2002}), \end{aligned}$$

where Y_{it} is the area of tobacco production (mu), D_i a treatment indicator equal to one for treatment regions and zero for control regions, and Z_i is a control variable with a constant term. We estimate $\alpha(\tau)$ at $\tau = 0.1, ..., 0.9$. Following Cai (2016), we employ the age of the head of the household, household size, and indicators of education level as control variables.

The main results from using our method are presented in Figure 1. The DID estimate is 0.239, and the 95 % confidence interval is $[0.084, 0.389]$. We use bootstrap to generate this confidence interval. Figure 4.1 shows that the estimates of $\alpha(\tau)$ differ across τ, and the QTE is increasing in τ. The impact of insurance provision is nearly zero at the lower and middle quantiles, and positive at the upper quantiles. The 95 % confidence intervals of the QTEs contain zero when $\tau \leq 0.7$.

Cai (2016) analyzes the welfare impact of the insurance program by using a calibration. The values of the parameters of the production function are chosen to match the DID (or triple difference) estimate. From this analysis, he concludes that providing a heavily subsidized compulsory insurance program has a positive welfare impact on rural households. However, our result shows that the insurance program does not change households' investment behavior so much at the lower and middle quantiles, and hence it may not affect household welfare at such quantiles.

4.6 Conclusion

In this chapter, we explore the identification and estimation of the QTE by using panel data. We generalize the CIC model and propose the tractable estimator of the QTE. Athey and Imbens (2006) suggest the CIC model as an alternative to the DID model. The CIC model allows for the estimation of the potential outcomes distribution and captures the heterogeneous effects of the treatment on the outcomes. However, there are two problems in the CIC model: (1) there is a lack of a tractable estimator in the presence of covariates and (2) the CIC estimator does not work when the treatment is continuous. Our model allows the presence of covariates and the continuous treatment. We propose the two-step estimation method based on the quantile regression and minimum distance method. We then show the consistency and asymptotic normality of our estimator. Monte Carlo studies indicate that our estimator performs well in finite samples. We use our method to estimate the impact of an insurance program on quantiles of household production.

4.7 Appendix: Proofs

Proof of Proposition 4.1. It is enough to show the first statement. First, we show that, if for all $x, x' \in \mathcal{X}_t$ and $z \in \mathcal{Z}_t$, we can identify the strictly function $Q^t_{x',x|z}(y)$ that satisfies

$$q_t(x', z, \tau) = Q^t_{x',x|z}(q_t(x, z, \tau)), \tag{4.14}$$

then $q_t(x, z, \tau)$ is identified for all $x \in \mathcal{X}_t$ and $z \in \mathcal{Z}_t$. Define

$$G^t_{x|z}(y) \equiv \int F_{Y_t|X_t,Z_t}(Q^t_{x',x|z}(y)|x', z)dF_{X_t|Z_t}(x'|z).$$

It follows from (4.14) that

$$\begin{aligned} G^t_{x|z}(q_t(x, z, \tau)) &= \int F_{Y_t|X_t,Z_t}(q_t(x', z, \tau)|x', z)dF_{X_t|Z_t}(x'|z) \\ &= \int P(U_{it} \le \tau | X_{it} = x', \tilde{Z}_{it} = z)dF_{X_t|Z_t}(x'|z) \\ &= P(U_{it} \le \tau | Z_{it} = z) \ = \ \tau. \end{aligned}$$

Hence we have $q_t(x, z, \tau) = (G^t_{x|z})^{-1}(\tau)$. This implies that $q_t(x, z, \tau)$ is point identified for all $x \in \mathcal{X}_t$.

Next, we show that for all $x, x' \in \mathcal{X}_t$ and $z \in \mathcal{Z}_t$, we can identify the strictly increasing function $Q^t_{x',x|z}(y)$ that satisfies (4.14). Observe that for $\mathbf{x} = (x'_1, \cdots, x'_T) \in \mathcal{X}_{1,\cdots,T}$ and $z \in \mathcal{Z}$,

$$\begin{aligned} F_{Y_t|\mathbf{X},Z_t}(q_t(x_t, z, \tau)|\mathbf{x}, z) &= P(U_{it} \le \tau | \mathbf{X}_i = \mathbf{x}, Z_{it} = z) \\ &= P(U_{is} \le \tau | \mathbf{X}_i = \mathbf{x}, Z_{is} = z) \\ &= F_{Y_s|\mathbf{X},Z_s}(q_s(x_s, z, \tau)|\mathbf{x}, z), \end{aligned} \tag{4.15}$$

where the second equality holds by Assumptions 4.1 (ii) and 4.2. By Assumption 4.3, we have $q_s(x_s, z, \tau) = F^{-1}_{Y_s|\mathbf{X},Z_s}\left(F_{Y_t|\mathbf{X},Z_t}(q_t(x_t, z, \tau)|\mathbf{x}, z)|\mathbf{x}, z\right)$. Hence, we can identify the strictly increasing function $\tilde{Q}^{s,t}_{x_s,x_t|z}(y)$ such that

$$q_s(x_s, z, \tau) = \tilde{Q}^{s,t}_{x_s,x_t|z}\left(q_t(x_t, z, \tau)\right). \tag{4.16}$$

Fix $z \in \mathcal{Z}$. We show that for all $x' \in \mathcal{S}^1_t(\overline{x})$ we can identify the strictly increasing function $\tilde{Q}^t_{x',\overline{x}|z}(y)$ such that

$$q_t(x', z, \tau) = \tilde{Q}^t_{x',\overline{x}|z}\left(q_t(\overline{x}, z, \tau)\right). \tag{4.17}$$

By the definition of $\mathcal{S}^1_t(\overline{x})$, there exist $x_s \in \mathcal{X}_s$ such that $(\overline{x}, x_s), (x', x_s) \in \mathcal{X}_{t,s}$. Thus, by (4.16), we have

$$q_t(x', z, \tau) = \tilde{Q}^{t,s}_{x',x_s|z}\left(\tilde{Q}^{s,t}_{x_s,\overline{x}|z}\left(q_t(\overline{x}, z, \tau)\right)\right).$$

Similarly, for all $x' \in \mathcal{S}_t^n(\overline{x})$, we can identify the strictly increasing function $\tilde{Q}^t_{x',\overline{x}|z}(y)$ that satisfy (4.17). By the continuity of q_t, for all $x' \in \overline{\cup_{n=0}^{\infty}\mathcal{S}_t^n(\overline{x})}$, we can identify the strictly increasing function $\tilde{Q}^t_{x',\overline{x}|z}(y)$ that satisfy (4.17). Because we have

$$q_t(x', z, \tau) = \tilde{Q}^t_{x',\overline{x}|z}\left(\left(\tilde{Q}^t_{x,\overline{x}|z}\right)^{-1}(q_t(x, z, \tau))\right),$$

we can identify the strictly increasing function $Q^t_{x',x|z}(y)$ that satisfies (4.14). Therefore, we can point identify $q_t(x, z, \tau)$. □

Proof of Theorem 4.1. It follows from the usual argument of quantile regression that $\beta_t(\alpha(\tau), \tau) = \beta_t(\tau)$. Because $\|D^t(v; a, \beta(a, \tau))\|^2_\mu \geq 0$, it is sufficient to prove that

$$\frac{1}{T}\sum_{t=1}^{T}\|D^t(v; a, \beta(a, \tau))\|^2_\mu = 0 \quad \Leftrightarrow \quad a = \alpha(\tau). \tag{4.18}$$

Suppose that $a = \alpha(\tau)$. Because $\beta_t(\alpha(\tau), \tau) = \beta_t(\tau)$, we have $D^t(v; a, \beta(a, \tau)) = D^t(v; \alpha(\tau), \beta(\tau)) = 0$ for all v and t.

Suppose that $a^* \in \mathcal{A}$ satisfies $\frac{1}{T}\sum_{t=1}^{T}\|D^t(v; a^*, \beta(a^*, \tau))\|^2_\mu = 0$. Then, it follows from the definition of $D^t(v; a, b)$ that for all t, s, $\mathbf{x} \in \mathcal{X}_{1,\cdots,T}$, and $z \in \mathcal{Z}$,

$$\begin{aligned} & P(Y_{it} \leq X'_{it}a^* + Z'_{it}\beta_t(a^*, \tau)|\mathbf{X}_i = \mathbf{x}, Z_{it} = z) \\ = & P(Y_{is} \leq X'_{is}a^* + Z'_{is}\beta_s(a^*, \tau)|\mathbf{X}_i = \mathbf{x}, Z_{is} = z). \end{aligned} \tag{4.19}$$

Define $\tilde{q}_t(x, z, \tau; a^*) \equiv x'a^* + z'\beta_t(a^*, \tau)$, then we have

$$F_{Y_t|\mathbf{X},Z_t}\left(\tilde{q}_t(x_t, z, \tau; a^*)|\mathbf{x}, z\right) = F_{Y_s|\mathbf{X},Z_s}\left(\tilde{q}_s(x_s, z, \tau; a^*)|\mathbf{x}, z\right),$$

where $\mathbf{x} = (x_1, \cdots, x_T)' \in \mathcal{X}_{1,\cdots,T}$. Similar to the proof of Proposition 4.1, by (4.19), for all $x, \tilde{x} \in \mathcal{X}_t$ we have

$$\tilde{q}_t(\tilde{x}, z, \tau; a^*) = Q^t_{\tilde{x},x|z}\left(\tilde{q}_t(x, z, \tau; a^*)\right),$$

where $Q^t_{\tilde{x},x|z}(y)$ are defined in the proof of Proposition 4.1. We define $G^t_{x|z}(y) \equiv \int F_{Y_t|X_t,Z_t}\left(Q^t_{\tilde{x},x|z}(y)\right) dF$ then we have

$$\begin{aligned} G^t_{x|z}\left(\tilde{q}_t(x, z, \tau; a^*)\right) &= \int F_{Y_t|X_t,\tilde{Z}_t}\left(\tilde{q}_t(\tilde{x}, z, \tau; a^*)|\tilde{x}, z\right) dF_{X_t|Z_t}(\tilde{x}|z) \\ &= P\left(Y_{it} \leq X'_{it}a^* + Z'_{it}\beta_t(a^*, \tau)|Z_{it} = z\right) = e_t(a^*, \tau, z). \end{aligned}$$

By the proof of Proposition 4.1, $\left(G^t_{x|z}\right)^{-1}(\tau) = q_t(x, z, \tau)$. Hence, we obtain

$$\tilde{q}_t(x, z, \tau; a^*) = q_t(x, z, e_t(a^*, \tau, z)) = x'\alpha(e_t(a^*, \tau, z)) + z'\beta_t(e_t(a^*, \tau, z)).$$

By Assumptions 4.4 (ii) and 4.5 (i), this implies that $a^* = \alpha(e_t(a^*, \tau, z))$ holds for all $z \in \mathcal{Z}$. Hence, it follows from Assumption 4.6 that $a^* = \alpha(\tau)$. □

Proof of Theorem 4.2. To prove the consistency of $\hat{\alpha}(\tau)$, we show the consistency of $\|D^t(v;a,b)\|_{\tilde{\mu}}$ in a and b. Then, we observe that

$$\begin{aligned} & |D^t(v;a,b) - D^t(v;\tilde{a},\tilde{b})| \\ = \; & \left|E\left[(E[g_t(\mathbf{W};a,b,v_Z)|\mathbf{X}] - E[g_t(\mathbf{W};\tilde{a},\tilde{b},v_Z)|\mathbf{X}])\mathbf{1}\{\mathbf{X} \leq v_{\mathbf{x}}\}\right]\right| \\ \leq \; & 2\max_s E\left[|F_{Y_s|\mathbf{X},Z_s}(X_s'a + Z_s'b_s|\mathbf{X},Z_s) - F_{Y_s|\mathbf{X},Z_s}(X_s'\tilde{a} + Z_s'\tilde{b}_s|\mathbf{X},Z_s)|\right]. \end{aligned}$$

Hence, $\{(a,b) \mapsto D^t(v;a,b) : v \in \mathcal{V}\}$ is equicontinuous by continuity of the conditional distribution. This implies that $\|D^t(v;a,b)\|_{\tilde{\mu}}$ is continuous in a and b.

Fix $\tau \in \mathcal{T}$. We show the consistency of $\hat{\alpha}(\tau)$ and $\hat{\beta}(\tau)$. By Lemma 4.2 and the definition of $\hat{\alpha}(\tau)$, we have

$$\begin{aligned} \left\|D^t\left(v;\hat{\alpha}(\tau),\tilde{\beta}(\hat{\alpha}(\tau),\tau)\right)\right\|_{\tilde{\mu}} &= \left\|\hat{D}_n^t\left(v;\hat{\alpha}(\tau),\tilde{\beta}(\hat{\alpha}(\tau),\tau)\right)\right\|_{\tilde{\mu}} + o_p(1) \\ &\leq \left\|\hat{D}_n^t\left(v;\alpha(\tau),\tilde{\beta}(\alpha(\tau),\tau)\right)\right\|_{\tilde{\mu}} + o_p(1) \\ &= \left\|D^t\left(v;\alpha(\tau),\tilde{\beta}(\alpha(\tau),\tau)\right)\right\|_{\tilde{\mu}} + o_p(1). \qquad (4.20) \end{aligned}$$

Because $F_{Y_t|\mathbf{X},Z_t}(y|\mathbf{x},z)$ is uniform continuous in y, it follows from Lemma 4.2 that

$$\|D^t(v;a,\tilde{\beta}(a,\tau))\|_{\tilde{\mu}} = \|D^t(v;a,\beta(a,\tau))\|_{\tilde{\mu}} + o_p(1).$$

Hence, (4.20) implies that

$$\left\|D^t\left(v;\hat{\alpha}(\tau),\beta(\hat{\alpha}(\tau),\tau)\right)\right\|_{\tilde{\mu}} \leq \left\|D^t\left(v;\alpha(\tau),\beta(\tau)\right)\right\|_{\tilde{\mu}} + o_p(1). \qquad (4.21)$$

Pick any $\delta > 0$. Then, by (4.11), Assumption 4.7 (ii), and continuity of $\|D^t(v;a,\beta(a,\tau))\|_{\tilde{\mu}}$, we obtain

$$\inf_{a\in\mathcal{A},\|a-\alpha(\tau)\|>\delta} \|D^t(v;a,\beta(a,\tau))\|_{\tilde{\mu}} > \|D^t(v;\alpha(\tau),\beta(\tau))\|_{\tilde{\mu}}.$$

By (4.21), wp $\to 1$ we have

$$\left\|D^t\left(v;\hat{\alpha}(\tau),\beta(\hat{\alpha}(\tau),\tau)\right)\right\|_{\tilde{\mu}} < \inf_{a\in\mathcal{A},\|a-\alpha(\tau)\|>\delta} \|D^t(v;a,\beta(a,\tau))\|_{\tilde{\mu}}.$$

Hence, we obtain $\|\hat{\alpha}(\tau) - \alpha(\tau)\| \to_p 0$. Because $\beta(a,\tau)$ is continuous in a, it follows from Lemma 4.2 that $\|\hat{\beta}(\tau) - \beta(\tau)\| \to_p 0$. □

Proof of Theorem 4.3. For $b : \mathcal{A} \times (0,1) \mapsto \mathbb{R}^{T(d_Z+1)}$, define

$$\begin{aligned} M^t(v;a,b,\tau) &\equiv D^t(v;a,b(a,\tau)), \\ M_n^t(v;a,b,\tau) &\equiv \hat{D}_n^t(v;a,b(a,\tau)). \end{aligned}$$

For $f : \mathcal{A} \times (0,1) \mapsto \mathbb{R}^d$, define $\|f\|_\infty \equiv \sup_{a\in\mathcal{A},\tau\in\mathcal{T}} \|f(a,\tau)\|$.

Fix $\tau \in \mathcal{T}$. First, we prove $\sqrt{n}$-consistency of $\hat{\alpha}(\tau)$. Because $\|\hat{\alpha}(\tau) - \alpha(\tau)\| \to_p 0$ and $\|\tilde{\beta} - \beta\|_\infty \to_p 0$ holds by Theorem 4.2 and Lemma 4.2, we choose a positive sequence $\delta_n = o(1)$ such that $P(\|\hat{\alpha}(\tau) - \alpha(\tau)\| \geq \delta_n,\ \|\tilde{\beta} - \beta\|_\infty \geq \delta_n) \to 0$. It follows from Lemma 4.6 that

$$\|M^t(v;\hat{\alpha}(\tau),\beta,\tau)\|_{\tilde{\mu}} + o_p(\|\hat{\alpha}(\tau)-\alpha(\tau)\|) \ \ \geq \ \ \|\Gamma_1^t(v;\tau)'(\hat{\alpha}(\tau)-\alpha(\tau))\|_{\tilde{\mu}}.$$

By Assumption 8 (v), we obtain

$$\|M^t(v;\hat{\alpha}(\tau),\beta,\tau)\|_{\tilde{\mu}} \geq (c - o_p(1)) \times \|\hat{\alpha}(\tau) - \alpha(\tau)\|. \tag{4.22}$$

Because $M_n^t(v;\alpha(\tau),\beta,\tau) = O_p(n^{-1/2})$ uniformly over $v \in \mathcal{V}$, $\|M^t(v;\hat{\alpha}(\tau),\beta,\tau)\|_{\tilde{\mu}}$ is bounded above by

$$\begin{aligned} & \|M^t(v;\hat{\alpha}(\tau),\beta,\tau) - M^t(v;\hat{\alpha}(\tau),\tilde{\beta},\tau)\|_{\tilde{\mu}} \\ + & \|M^t(v;\hat{\alpha}(\tau),\tilde{\beta},\tau) - M_n^t(v;\hat{\alpha}(\tau),\tilde{\beta},\tau) + M_n^t(v;\alpha(\tau),\beta,\tau)\|_{\tilde{\mu}} \\ + & \|M_n^t(v;\hat{\alpha}(\tau),\tilde{\beta},\tau)\|_{\tilde{\mu}} + O_p(n^{-1/2}), \end{aligned}$$

where $O_p(n^{-1/2})$ is uniform over $\tau \in \mathcal{T}$. Because $\{r_\tau(w;a,b) : a \in \mathcal{A}, b \in \mathcal{B}\}$ is Donsker, it follows from Lemma 4.4 and 4.6 that

$$\begin{aligned} & \|M^t(v;\hat{\alpha}(\tau),\beta,\tau) - M^t(v;\hat{\alpha}(\tau),\tilde{\beta},\tau)\|_{\tilde{\mu}} \\ \leq & \left\| M^t(v;\hat{\alpha}(\tau),\tilde{\beta},\tau) - M^t(v;\hat{\alpha}(\tau),\beta,\tau) - \Gamma_2^t(v,\hat{\alpha}(\tau),\tau)'[\tilde{\beta}(\hat{\alpha}(\tau),\tau) - \beta(\hat{\alpha}(\tau),\tau)] \right\|_{\tilde{\mu}} \\ & + \left\| \left(\Gamma_2^t(v,\hat{\alpha}(\tau),\tau) - \Gamma_2^t(v,\alpha(\tau),\tau)\right)' [\tilde{\beta}(\hat{\alpha}(\tau),\tau) - \beta(\hat{\alpha}(\tau),\tau)] \right\|_{\tilde{\mu}} \\ & + \left\| \Gamma_2^t(v,\alpha(\tau),\tau)' \left([\tilde{\beta}(\hat{\alpha}(\tau),\tau) - \beta(\hat{\alpha}(\tau),\tau)] - [\tilde{\beta}(\alpha(\tau),\tau) - \beta(\alpha(\tau),\tau)] \right) \right\|_{\tilde{\mu}} \\ & + \left\| \Gamma_2^t(v,\alpha(\tau),\tau)'[\tilde{\beta}(\alpha(\tau),\tau) - \beta(\alpha(\tau),\tau)] \right\|_{\tilde{\mu}} \\ \leq & \ o_p(\|\tilde{\beta} - \beta\|_\infty) + o_p(1) \times \|\tilde{\beta} - \beta\|_\infty + O_p(n^{-1/2}) = O_p(n^{-1/2}). \end{aligned} \tag{4.23}$$

By Lemma 4.3, we obtain

$$\|M^t(v;\hat{\alpha}(\tau),\tilde{\beta},\tau) - M_n^t(v;\hat{\alpha}(\tau),\tilde{\beta},\tau) + M_n^t(v;\alpha(\tau),\beta,\tau)\|_{\tilde{\mu}} = o_p(n^{-1/2}).$$

Hence, by (4.22) and (4.23), we have

$$(c - o_p(1)) \times \|\hat{\alpha}(\tau) - \alpha(\tau)\| \leq O_p(n^{-1/2}) + \|M_n^t(v;\hat{\alpha}(\tau),\tilde{\beta},\tau)\|_{\tilde{\mu}}. \tag{4.24}$$

By definition of $\hat{\alpha}(\tau)$,

$$
\begin{aligned}
& \|M_n^t(v;\hat{\alpha}(\tau),\tilde{\beta},\tau)\|_{\tilde{\mu}} \\
\leq\ & \|M_n^t(v;\alpha(\tau),\tilde{\beta},\tau)\|_{\tilde{\mu}} \\
\leq\ & \left\|M_n^t(v;\alpha(\tau),\tilde{\beta},\tau) - M^t(v;\alpha(\tau),\tilde{\beta},\tau) - M_n^t(v;\alpha(\tau),\beta,\tau)\right\|_{\tilde{\mu}} \\
& + \left\|M^t(v;\alpha(\tau),\tilde{\beta},\tau) - \Gamma_2^t(v;\tau)'[\tilde{\beta}(\alpha(\tau),\tau) - \beta(\alpha(\tau),\tau)]\right\|_{\tilde{\mu}} \\
& + \left\|\Gamma_2^t(v;\tau)'[\tilde{\beta}(\alpha(\tau),\tau) - \beta(\alpha(\tau),\tau)]\right\|_{\tilde{\mu}} + \left\|M_n^t(v;\alpha(\tau),\beta,\tau)\right\|_{\tilde{\mu}} \\
\leq\ & o_p(n^{-1/2}) + o_p(\|\tilde{\beta}-\beta\|_\infty) + O_p(n^{-1/2}) = O_p(n^{-1/2}).
\end{aligned}
$$

Therefore, by (4.24), we have $\|\hat{\alpha}(\tau) - \alpha(\tau)\| \leq O_p(n^{-1/2})$.

Next we show (4.12) by approximating $M_n^t(v;a,\tilde{\beta},\tau)$ as

$$
L_n^t(v;a,\tau) \equiv M_n^t(v;\alpha(\tau),\beta,\tau) + \Gamma_1^t(v;\tau)'(a-\alpha(\tau)) + \Gamma_2^t(v;\tau)'[\tilde{\beta}(\alpha(\tau),\tau) - \beta(\alpha(\tau),\tau)].
$$

Let $\bar{\alpha}(\tau)$ be the value that provides a global minimum for $\|L_n^t(v;a,\tau)\|_{\tilde{\mu}}$. Then, $\Gamma_1^t(v;\tau)'(\bar{\alpha}(\tau) - \alpha(\tau))$ is the $L_2(\tilde{\mu})$-projection of $-M_n^t(v;\alpha(\tau),\beta,\tau) - \Gamma_2^t(v;\tau)'[\tilde{\beta}(\alpha(\tau),\tau) - \beta(\alpha(\tau),\tau)]$ onto the subspace of $L_2(\mu)$ spanned by $\Gamma_1^t(v;\tau)$. Hence,

$$
\begin{aligned}
& \sqrt{n}(\bar{\alpha}(\tau) - \alpha(\tau)) \\
=\ & -\Delta_1(\tau)^{-1}\sqrt{n}\int \Gamma_1^t(v;\tau)\left\{M_n^t(v;\alpha(\tau),\beta,\tau) + \Gamma_2^t(v;\tau)'[\tilde{\beta}(\alpha(\tau),\tau) - \beta(\alpha(\tau),\tau)]\right\} d\tilde{\mu}(v,t),
\end{aligned}
$$

where $\Delta_1(\tau) \equiv \frac{1}{T}\sum_{t=1}^T \int \Gamma_1^t(v;\tau)\Gamma_1^t(v;\tau)' d\mu(v)$. Here, we have

$$
\begin{aligned}
& \sqrt{n}\int \Gamma_1^t(v;\tau) M_n^t(v;\alpha(\tau),\beta,\tau) d\tilde{\mu}(v,t) \\
=\ & \frac{1}{\sqrt{n}T}\sum_{i=1}^n\sum_{t=1}^T \int \Gamma_1^t(v;\tau) g_t(\mathbf{W}_i;\alpha(\tau),\beta(\tau),v_z)\mathbf{1}\{\mathbf{X}_i \leq v_{\mathbf{x}}\} d\mu(v).
\end{aligned}
$$

Define $Y_{it}(\tau) \equiv X_{it}'\alpha(\tau) + Z_{it}'\beta_t(\tau)$. Because $\sum_{t=1}^T \Gamma_1^t(v;\tau) = 0$, we have

$$
\begin{aligned}
& \frac{1}{T}\sum_{t=1}^T \int \Gamma_1^t(v;\tau) g_t(\mathbf{W}_i;\alpha(\tau),\beta(\tau),v_z)\mathbf{1}\{\mathbf{X}_i \leq v_{\mathbf{x}}\} d\mu(v) \\
=\ & \frac{1}{T}\sum_{t=1}^T \int \Gamma_1^t(v;\tau)\mathbf{1}\{\mathbf{X}_i \leq v_{\mathbf{x}}, Z_{it} \leq v_z\} d\mu(v)\mathbf{1}\{Y_{it} \leq Y_{it}(\tau)\} \\
& - \frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T \int \Gamma_1^t(v;\tau)\mathbf{1}\{\mathbf{X}_i \leq v_{\mathbf{x}}, Z_{is} \leq v_z\} d\mu(v)\mathbf{1}\{Y_{is} \leq Y_{it}(\tau)\} \\
=\ & \frac{1}{T}\sum_{t=1}^T \int \Gamma_1^t(v;\tau)\mathbf{1}\{\mathbf{X}_i \leq v_{\mathbf{x}}, Z_{it} \leq v_z\} d\mu(v)\mathbf{1}\{Y_{it} \leq Y_{it}(\tau)\}.
\end{aligned}
$$

It follows from Lemma 4.4 that $\sqrt{n}(\tilde{\beta}(\alpha(\tau),\tau) - \beta(\alpha(\tau),\tau)) = -1/\sqrt{n}\sum_{i=1}^{n} l(\mathbf{W}_i;\tau)$. Therefore, we obtain

$$\begin{aligned}
& \sqrt{n}\int \Gamma_1^t(v;\tau)\Gamma_2^t(v;\tau)'[\tilde{\beta}(\alpha(\tau),\tau) - \beta(\alpha(\tau),\tau)]d\tilde{\mu}(v,t) \\
= & -\frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\sum_{t=1}^{T}\left[\int \Gamma_1^t(v;\tau)\Gamma_2^t(v;\tau)'d\mu(v)\right] l(\mathbf{W}_i;\tau) + o_p(1).
\end{aligned}$$

This implies that $\sqrt{n}(\bar{\alpha}(\tau)-\alpha(\tau)) = -\Delta_1(\tau)^{-1}\left[1/\sqrt{n}\sum_{i=1}^{n}\{\xi(\mathbf{W}_i;\tau) - \Delta_{12}(\tau)l(\mathbf{W}_i;\tau)\}\right] + o_p(1)$, and hence it is sufficient to show that $\|\bar{\alpha}(\tau) - \hat{\alpha}(\tau)\| = o_p(n^{-1/2})$.

Because $\|\hat{\alpha}(\tau) - \alpha(\tau)\| = O_p(n^{-1/2})$, we have

$$\begin{aligned}
& \|M_n^t(v;\hat{\alpha}(\tau),\tilde{\beta},\tau) - L_n^t(v;\hat{\alpha}(\tau),\tau)\|_{\tilde{\mu}} \\
\le & \|M^t(v;\hat{\alpha}(\tau),\tilde{\beta},\tau) - M^t(v;\hat{\alpha}(\tau),\beta,\tau) - \Gamma_2^t(v;\hat{\alpha}(\tau),\tau)'[\tilde{\beta}(\hat{\alpha}(\tau),\tau) - \beta(\hat{\alpha}(\tau),\tau)]\|_{\tilde{\mu}} \\
& + \left\|\Gamma_2^t(v;\hat{\alpha}(\tau),\tau)'\left\{[\tilde{\beta}(\hat{\alpha}(\tau),\tau) - \beta(\hat{\alpha}(\tau),\tau)] - [\tilde{\beta}(\alpha(\tau),\tau) - \beta(\alpha(\tau),\tau)]\right\}\right\|_{\tilde{\mu}} \\
& + \left\|\left\{\Gamma_2^t(v;\hat{\alpha}(\tau),\tau) - \Gamma_2^t(v;\alpha(\tau),\tau)\right\}'[\tilde{\beta}(\alpha(\tau),\tau) - \beta(\alpha(\tau),\tau)]\right\|_{\tilde{\mu}} \\
& + \|M^t(v;\hat{\alpha}(\tau),\beta,\tau) - \Gamma_1^t(v;\tau)'(\hat{\alpha}(\tau) - \alpha(\tau))\|_{\tilde{\mu}} \\
& + \|M_n^t(v;\hat{\alpha}(\tau),\tilde{\beta},\tau) - M^t(v;\hat{\alpha}(\tau),\tilde{\beta},\tau) - M_n^t(v;\alpha(\tau),\beta,\tau)\|_{\tilde{\mu}} \\
\le & o_p(\|\tilde{\beta} - \beta\|_\infty) + o_p(\|\hat{\alpha}(\tau) - \alpha(\tau)\|) + o_p(n^{-1/2}) = o_p(n^{-1/2}).
\end{aligned} \tag{4.25}$$

Similarly, we have

$$\|M_n^t(v;\bar{\alpha}(\tau),\tilde{\beta},\tau) - L_n^t(v;\bar{\alpha}(\tau),\tau)\|_{\tilde{\mu}} = o_p(n^{-1/2}). \tag{4.26}$$

Hence, it follows from (4.25) and (4.26) that

$$\begin{aligned}
\|L_n^t(v;\hat{\alpha}(\tau),\tau)\|_{\tilde{\mu}} - o_p(n^{-1/2}) & \le \|M_n^t(v;\hat{\alpha}(\tau),\tilde{\beta},\tau)\|_{\tilde{\mu}} \\
& \le \|M_n^t(v;\bar{\alpha}(\tau),\tilde{\beta},\tau)\|_{\tilde{\mu}} \\
& \le \|L_n^t(v;\bar{\alpha}(\tau),\tau)\|_{\tilde{\mu}} + o_p(n^{-1/2}).
\end{aligned}$$

By definition of $\bar{\alpha}(\tau)$, we have $\|L_n^t(v;\hat{\alpha}(\tau),\tau)\|_{\tilde{\mu}} = \|L_n^t(v;\bar{\alpha}(\tau),\tau)\|_{\tilde{\mu}} + o_p(n^{-1/2})$. Because $\|L_n^t(v;\bar{\alpha}(\tau),\tau)\|_{\tilde{\mu}} = O_p(n^{-1/2})$, we have

$$\|L_n^t(v;\hat{\alpha}(\tau),\tau)\|_{\tilde{\mu}}^2 = \|L_n^t(v;\bar{\alpha}(\tau),\tau)\|_{\tilde{\mu}}^2 + o_p(n^{-1}).$$

Because $L_n^t(v;\bar{\alpha}(\tau),\tau)$ is orthogonal to $\Gamma_1^t(v;\tau)$, we obtain

$$\begin{aligned}
\|L_n^t(v;\bar{\alpha}(\tau),\tau)\|_{\tilde{\mu}}^2 & = \|L_n^t(v;a,\tau) - \Gamma_1^t(v;\tau)'(a - \bar{\alpha}(\tau))\|_{\tilde{\mu}}^2 \\
& = \|L_n^t(v;a,\tau)\|_{\tilde{\mu}}^2 + \|\Gamma_1^t(v;\tau)'(a - \bar{\alpha}(\tau))\|_{\tilde{\mu}}^2.
\end{aligned}$$

Hence, $\|\hat{\alpha}(\tau) - \bar{\alpha}(\tau)\| = o_p(n^{-1/2})$ holds, and we obtain (4.12).

Finally, we show asymptotic normality of $\hat{\beta}(\tau)$. By the proof of Lemma 4.4, we have

$$\begin{aligned} O(1/\sqrt{n}) &= \mathbb{G}_n r_\tau(W_t; \alpha(\tau), \beta_t(\tau)) + o_p(1) + \sqrt{n} E r_\tau(W_t; \hat{\alpha}(\tau), \hat{\beta}_t(\hat{\alpha}(\tau), \tau)) \\ &= \mathbb{G}_n r_\tau(W_t; \alpha(\tau), \beta_t(\tau)) + o_p(1) + (J_t^a(\tau) + o_p(1))\sqrt{n}(\hat{\alpha}(\tau) - \alpha(\tau)) \\ &\quad + (J_t^b(\tau) + o_p(1))\sqrt{n}(\hat{\beta}(\tau) - \beta(\tau)). \end{aligned}$$

Hence,

$$\begin{aligned} \sqrt{n}(\hat{\beta}(\tau) - \beta(\tau)) &= -J_t^b(\tau)^{-1} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n r_\tau(W_{it}; \alpha(\tau), \beta_t(\tau)) \right. \\ &\quad \left. - J_t^a(\tau)\Delta_1(\tau)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n (\xi(\mathbf{W}_i; \tau) + \Delta_{12}(\tau) l(\mathbf{W}_i; \tau)) \right\} + o_p(1), \end{aligned}$$

and we obtain (4.30). □

4.8 Appendix: Auxiliary Lemmas

Lemma 4.1. *Under the assumptions of Theorem 4.2, for all $\tau \in \mathcal{T}$, we have*

$$\sup_{a \in \mathcal{A}, b_t \in \mathcal{B}_t} \left| \frac{1}{n} \sum_{i=1}^n R_\tau(W_{it}; a, b_t) - E[R_\tau(W_{it}; a, b_t)] \right| = o_{a.s.}(1), \tag{4.27}$$

$$\sup_{\substack{a \in \mathcal{A}, b \in \mathcal{B}, \\ v \in \mathcal{V}}} \left| \hat{D}_n^t(v; a, b) - D^t(v; a, b) \right| = o_{a.s.}(1). \tag{4.28}$$

Proof. By definition of $R_\tau(w; a, b_t)$, the collection of functions $\{R_\tau(\cdot; a, b_t) : a \in \mathcal{A}, b_t \in \mathcal{B}_t\}$ is a VC-class. Hence, $\{R_\tau(\cdot; a, b_t) : a \in \mathcal{A}, b_t \in \mathcal{B}_t\}$ is Donsker, and also Glivenko-Cantelli. This implies (4.27).

Because $\{(x, z) \mapsto x'a + z'b_t : a \in \mathcal{A}, b_t \in \mathcal{B}_t\}$ is a VC-class, $\{\mathbf{w} \mapsto g_t(\mathbf{w}; a, b, v_z)\mathbf{1}\{\mathbf{x} \leq v_{\mathbf{x}}\} : a \in \mathcal{A}, b \in \mathcal{B}, v \in \mathcal{V}\}$ is also Donsker by Example 2.10.7 and 2.10.8 of van der Vaart and Wellner (1996). Therefore, we have (4.28). □

Lemma 4.2. *Under the assumptions of Theorem 4.2, for all $\tau \in \mathcal{T}$, we have*

$$\sup_{a \in \mathcal{A}} \left\| \tilde{\beta}_t(a, \tau) - \beta_t(a, \tau) \right\| = o_p(1). \tag{4.29}$$

Proof. Fix $\tau \in \mathcal{T}$. Lemma 4.1 implies that uniformly in a,

$$\begin{aligned} E[R_\tau(W_{it}; a, \tilde{\beta}_t(a, \tau))] &= \frac{1}{n} \sum_{i=1}^n R_\tau(W_{it}; a, \tilde{\beta}_t(a, \tau)) + o_p(1) \\ &< \frac{1}{n} \sum_{i=1}^n R_\tau(W_{it}; a, \beta_t(a, \tau)) + o_p(1) \\ &= E[R_\tau(W_{it}; a, \beta_t(a, \tau))] + o_p(1). \end{aligned}$$

Pick any $\delta > 0$. Let $\{B_\delta(a,\tau) : a \in \mathcal{A}\}$ be a collection of balls with diameter $\delta > 0$, each centered at $\beta_t(a,\tau)$. Because $\rho_\tau(u) - \rho_\tau(u') \le |u - u'|$, we have $E[R_\tau(W_{it}; a, b_t)] - E[R_\tau(W_{it}; \tilde{a}, \tilde{b}_t)] \le C\|(a', b_t')' - (\tilde{a}', \tilde{b}_t')'\|$. Hence, the function $b_t \mapsto E[R_\tau(W_{it}; a, b_t)]$ is continuous uniformly over $a \in \mathcal{A}$. Because $\frac{\partial^2}{\partial b_t \partial b_t'} E[R_\tau(W_t; a, b_t)]|_{b_t = \beta_t(a,\tau)} = J_t(a,\tau)$, it follows from Assumption 4.7 (iv) that

$$\inf_{a \in \mathcal{A}} \left[\inf_{b_t \in \mathcal{B}_t \setminus B_\delta(a,\tau)} E[R_\tau(W_{it}; a, b_t)] - E[R_\tau(W_{it}; a, \beta_t(a,\tau))] \right] > 0.$$

Uniformly in $a \in \mathcal{A}$, wp $\to 1$ we have

$$E[R_\tau(W_{it}; a, \tilde{\beta}_t(a,\tau))] \quad < \inf_{b_t \in \mathcal{B}_t \setminus B_\delta(a,\tau)} E[R_\tau(W_{it}; a_t, b_t)].$$

Therefore, wp $\to 1$ we have $\sup_{a \in \mathcal{A}} \|\tilde{\beta}_t(a,\tau) - \beta_t(a,\tau)\| \le \delta$. □

Lemma 4.3. *Define $f(W_t; a, b_t, \tau) \equiv (\tau - \mathbf{1}\{Y_t \le X_t'a + Z_t'b_t\}) Z_t$. Under the assumptions of Theorem 4.3, for any sequence of positive numbers $\{\delta_n\}$ that converges to zero, we have*

$$\sup_{\|\tilde{a}-a\| \le \delta_n, \|\tilde{b}_t - b_t\| \le \delta_n} \left| \mathbb{G}_n f(W_t; \tilde{a}, \tilde{b}_t, \tau) - \mathbb{G}_n f(W_t; a, b_t, \tau) \right| = o_p(1),$$

$$\sup_{\substack{\|\tilde{a}-a\| \le \delta_n, \|\tilde{b}-b\| \le \delta_n, \\ v \in \mathcal{V}}} \left| \sqrt{n}(\hat{D}_n^t(v; \tilde{a}, \tilde{b}) - D^t(v; \tilde{a}, \tilde{b})) - \sqrt{n}(\hat{D}_n^t(v; a, b) - D^t(v; a, b)) \right| = o_p(1).$$

Proof. Because $\{w \mapsto f(w; a, b_t, \tau) : a \in \mathcal{A}, b_t \in \mathcal{B}_t, \tau \in \mathcal{T}\}$ and $\{\mathbf{w} \mapsto g_t(\mathbf{w}; a, b, v_z)\mathbf{1}\{\mathbf{x} \le v_{\mathbf{x}}\} : a \in \mathcal{A}, b \in \mathcal{B}, v \in supp(\mu)\}$ are Donsker, we prove the statements of this lemma. □

Lemma 4.4. *Under the assumptions of Theorem 4.3, we have*

$$\sqrt{n}(\tilde{\beta}_t(a,\tau) - \beta_t(a,\tau)) = -J_t^b(a,\tau)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n r_\tau(W_{it}; a, \beta_t(a,\tau)) + o_p(1), \tag{4.30}$$

where $o_p(1)$ is uniform over $a \in \mathcal{A}$.

Proof. By the computational properties of the ordinary quantile regression estimator (see Theorem 3.3 in Koenker and Bassett (1978)), we obtain

$$O(1/\sqrt{n}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n r_\tau(W_{it}; a, \hat{\beta}_t(a,\tau))$$

uniformly over $a \in \mathcal{A}$. By Lemma 3, we have

$$\begin{aligned} O(1/\sqrt{n}) &= \sqrt{n}\mathbb{E}_n r_\tau(W_t; a, \hat{\beta}_t(a,\tau)) \\ &= \mathbb{G}_n r_\tau(W_t; a, \beta_t(a,\tau)) + o_p(1) + \sqrt{n} E r_\tau(W_t; a, \hat{\beta}_t(a,\tau)), \end{aligned} \tag{4.31}$$

where the term $o_p(1)$ is uniform over $a \in \mathcal{A}$. Because $Er_\tau(W_t; a, \beta_t(a,\tau)) = 0$ by first order condition, we obtain

$$\begin{aligned} Er_\tau(W_t; a, \hat\beta_t(a,\tau)) &= \left[\frac{\partial}{\partial b_t'} Er_\tau(W_t; a, b_t)\Big|_{b_t = \bar b^t_{a,\tau}} \right] (\hat\beta_t(a,\tau) - \beta_t(a,\tau)) \\ &= E\left[f_{Y_t - X_t'a|Z_t}(Z_t'\bar b^t_{a,\tau}|Z_t) Z_t Z_t' \right] (\hat\beta_t(a,\tau) - \beta_t(a,\tau)), \end{aligned}$$

where $\bar b^t_{a,\tau}$ is between $\hat\beta_t(a,\tau)$ and $\beta_t(a,\tau)$. Because $\{y \mapsto f_{Y_t - X_t'a|Z_t}(y|z) : a \in \mathcal{A}\}$ is equicontinuous for all z, we have

$$E\left[f_{Y_t - X_t'a|Z_t}(Z_t'\bar b^t_{a,\tau}|Z_t) Z_t Z_t' \right] = J^b_t(a,\tau) + o_p(1) \quad \text{uniformly over } a \in \mathcal{A}.$$

Therefore, it follows from (4.31) that

$$\sqrt{n}(\hat\beta_t(a,\tau) - \beta_t(a,\tau)) = -J^b_t(a,\tau)^{-1} \mathbb{G}_n r_\tau(W_t; a, \beta_t(a,\tau)) + o_p(1),$$

where the term $o_p(1)$ is uniform over $a \in \mathcal{A}$. □

Lemma 4.5. *Under the assumptions of Theorem 4.3, $D^t(v; a, \beta(a,\tau))$ is continuously differentiable in a, $D^t(v,a,b)$ is continuously differentiable in b, and*

$$\begin{aligned} \frac{\partial}{\partial a} D^t(v; a, \beta(a,\tau)) &= \Gamma^t_1(v; a, \tau), \\ \frac{\partial}{\partial b_s} D^t(v; a, b) &= \gamma^{t,s}_2(v; a, b). \end{aligned}$$

Proof. First, we show the continuous differentiability of $D^t(v; a, \beta(a,\tau))$ and $D^t(v,a,b)$. We observe that

$$\begin{aligned} D^t(v; a, b) &= E[\mathbf{1}\{Y_t \le X_t'a + Z_t'b_t\}\mathbf{1}\{\mathbf{X} \le v_{\mathbf{x}}, Z_t \le v_z\}] \\ &\quad - \frac{1}{T}\sum_{s=1}^T E[\mathbf{1}\{Y_s \le X_s'a + Z_s'b_s\}\mathbf{1}\{\mathbf{X} \le v_{\mathbf{x}}, Z_s \le v_z\}] \\ &= E[F_{Y_t|\mathbf{X},Z_t}(X_t'a + Z_t'b_t|\mathbf{X}, Z_t)\mathbf{1}\{\mathbf{X} \le v_{\mathbf{x}}, Z_t \le v_z\}] \\ &\quad - \frac{1}{T}\sum_{s=1}^T E[F_{Y_s|\mathbf{X},Z_s}(X_s'a + Z_s'b_s|\mathbf{X}, Z_s)\mathbf{1}\{\mathbf{X} \le v_{\mathbf{x}}, Z_s \le v_z\}]. \end{aligned}$$

Because we have

$$\begin{aligned} &\frac{\partial}{\partial b_t} E[F_{Y_t|\mathbf{X},Z_t}(X_t'a + Z_t'b_t|\mathbf{X}, Z_t)\mathbf{1}\{\mathbf{X} \le v_{\mathbf{x}}, Z_t \le v_z\}] \\ = \; & E[f_{Y_t|\mathbf{X},Z_t}(X_t'a + Z_t'b_t|\mathbf{X}, Z_t)\mathbf{1}\{\mathbf{X} \le v_{\mathbf{x}}, Z_t \le v_z\}Z_t], \end{aligned}$$

$D^t(v;a,b)$ is continuously differentiable in b and $(\partial/\partial b_s)D^t(v;a,b) = \gamma_2^{t,s}(v;a,b)$. Similarly, we have

$$\begin{aligned} & \frac{\partial}{\partial a} E[F_{Y_t|\mathbf{X},Z_t}(X_t'a + Z_t'\beta_t(a,\tau)|\mathbf{X},Z_t)\mathbf{1}\{\mathbf{X} \le v_{\mathbf{x}}, Z_t \le v_z\}] \\ = \;& E[f_{Y_t|\mathbf{X},Z_t}(X_t'a + Z_t'\beta_t(a,\tau)|\mathbf{X},Z_t)\mathbf{1}\{\mathbf{X} \le v_{\mathbf{x}}, Z_t \le v_z\}(X_t + B_t(a,\tau)'Z_t)]. \end{aligned}$$

Hence, $D^t(v;a,\beta(a,\tau))$ is also continuously differentiable in a and $(\partial/\partial a)D^t(v;a,\beta(a,\tau)) = \Gamma_1^t(v;a,\tau)$. □

Lemma 4.6. *Under the assumptions of Theorem 4.3, for any sequence of positive numbers $\{\delta_n\}$ that converges to zero, for all $\tau \in \mathcal{T}$, we have*

$$\sup_{\|a-\alpha(\tau)\|\le\delta_n} \|M^t(v;a,\beta,\tau) - \Gamma_1^t(v;\tau)'(a-\alpha(\tau))\|_{\tilde{\mu}} = o(\delta_n), \tag{4.32}$$

and

$$\begin{aligned} \sup_{a\in\mathcal{A},\|b-\beta\|_\infty\le\delta_n} & \big\| M^t(v;a,b,\tau) - M^t(v;a,\beta,\tau) \\ & -\Gamma_2^t(v;a,\tau)'[b(a,\tau)-\beta(a,\tau)]\big\|_{\tilde{\mu}} = o(\delta_n). \end{aligned} \tag{4.33}$$

Proof. First, we show (4.32). Because $M^t(v;a,\beta,\tau) = D^t(v;a,\beta(a,\tau))$ is continuously differentiable in a for all τ, there exists $\overline{a}_{v,\tau}^t$ between $\alpha(\tau)$ and a such that

$$M^t(v;a,\beta,\tau) - M^t(v;\alpha(\tau),\beta,\tau) = \Gamma_1(v;\overline{a}_{v,\tau}^t,\tau)'(a-\alpha(\tau)).$$

Because $M^t(v;\alpha(\tau),\beta,\tau) = 0$, we have

$$\begin{aligned} & \|M^t(v;a,\beta,\tau) - \Gamma_1^t(v;\alpha(\tau),\tau)'(a-\alpha(\tau))\|_{\tilde{\mu}} \\ = \;& \left\| \left(\Gamma_1(v;\overline{a}_{v,\tau}^t,\tau) - \Gamma_1^t(v;\alpha(\tau),\tau)\right)'(a-\alpha(\tau))\right\|_{\tilde{\mu}} \\ \le \;& \max_t \sup_{v\in\mathcal{V}} \|\Gamma_1(v;\overline{a}_{v,\tau}^t,\tau) - \Gamma_1^t(v;\alpha(\tau),\tau)\| \times \|a-\alpha(\tau)\|. \end{aligned}$$

Then, we have

$$\begin{aligned} & \|\Gamma_1(v;\overline{a}_{v,\tau}^t,\tau) - \Gamma_1^t(v;\alpha(\tau),\tau)\| \\ \le \;& 2\max_s E\left[\|f_{Y_s|\mathbf{X},Z_s}(X_s'\overline{a}_{v,\tau}^s + Z_s'\beta(\overline{a}_{v,\tau}^s,\tau)|\mathbf{X},Z_s)(X_s + B_s(\overline{a}_{v,\tau}^s,\tau)'Z_s)\right. \\ & \left. -f_{Y_s|\mathbf{X},Z_s}(X_s'\alpha(\tau) + Z_s'\beta_s(\alpha(\tau),\tau)|\mathbf{X},Z_s)(X_s + B_s(\alpha(\tau),\tau)'Z_s)\|\right]. \end{aligned}$$

Because $f_{Y_t|\mathbf{X},Z_t}(y|\mathbf{x},z)$ and $B_t(a,\tau)$ is continuous in y and a respectively, we obtain (4.32).

Next, we show (4.33). Because $D^t(v;a,b)$ is continuously differentiable in b, there exists $\overline{b}_{v,a,\tau}$ between $\tilde{b}(a,\tau)$ and $\beta(a,\tau)$ such that

$$D^t(v;a,\tilde{b}(a,\tau)) - D^t(v;a,\beta(a,\tau)) = \Gamma_2^t(v;a,\overline{b}_{v,a,\tau}^t)'[\tilde{b}(a,\tau)-\beta(a,\tau)].$$

Hence, we have

$$
\begin{aligned}
& \left| M^t(v;a,\tilde{b},\tau) - M^t(v;a,\beta,\tau) - \Gamma_2(v;a,\tau)'[\tilde{b}(a,\tau) - \beta(a,\tau)] \right| \\
= & \left| \left(\Gamma_2^t(v;a,\bar{b}_{v,a,\tau}) - \Gamma_2^t(v;a,\beta(a,\tau)) \right)' [\tilde{b}(a,\tau) - \beta(a,\tau)] \right| \\
\leq & \sup_{v\in\mathcal{V}} \|\Gamma_2^t(v;a,\bar{b}_{v,a,\tau}) - \Gamma_2^t(v;a,\beta(a,\tau))\| \times \|\tilde{b}(a,\tau) - \beta(a,\tau)\|.
\end{aligned}
$$

Similarly to (4.32), $\sup_{a\in\mathcal{A},v\in\mathcal{V}} \|\Gamma_2^t(v;a,\bar{b}_{v,a,\tau}) - \Gamma_2^t(v;a,\beta(a,\tau))\| = o(1)$ by the uniform continuity of $f_{Y_t|\mathbf{X},Z_t}(y|\mathbf{x},z)$ in y. Therefore, we obtain (4.33). □

4.9 Appendix: Figures and Tables

Table 4.1: Results of Simulation 1

		$N = 1000$			$N = 2000$		
		$\rho^2 = 0.1$	$\rho^2 = 0.5$	$\rho^2 = 0.9$	$\rho^2 = 0.1$	$\rho^2 = 0.5$	$\rho^2 = 0.9$
	bias	0.0211	0.0085	0.0055	0.0182	0.0104	0.0125
$\tau = 0.25$	std	0.2609	0.2087	0.1440	0.1797	0.1534	0.1042
	mse	0.0685	0.0436	0.0208	0.0327	0.0236	0.0110
	bias	0.0091	-0.0017	0.0037	0.0056	0.0076	-0.0033
$\tau = 0.50$	std	0.2274	0.1981	0.1226	0.1677	0.1404	0.0873
	mse	0.0518	0.0392	0.0150	0.0282	0.0198	0.0076
	bias	-0.0264	-0.0130	-0.0014	-0.0141	-0.0095	-0.0060
$\tau = 0.75$	std	0.2550	0.2114	0.1326	0.1771	0.1559	0.1052
	mse	0.0657	0.0449	0.0176	0.0316	0.0244	0.0111

Table 4.2: Results of Simulation 1

	τ	$a = 0.9$	$a = 0.95$	$a = 0.99$
	0.25	0.882	0.934	0.974
$N = 1000$	0.50	0.900	0.930	0.954
	0.75	0.874	0.938	0.982

Table 4.3: Results of Simulation 2

		$N = 1000$			$N = 2000$		
		$\rho^2 = 0.1$	$\rho^2 = 0.5$	$\rho^2 = 0.9$	$\rho^2 = 0.1$	$\rho^2 = 0.5$	$\rho^2 = 0.9$
	bias	0.0020	0.0016	-0.0033	-0.0030	-0.0022	0.0015
$\tau = 0.25$	std	0.2078	0.1858	0.1516	0.1508	0.1364	0.1088
	mse	0.0432	0.0345	0.0230	0.0227	0.0186	0.0118
	bias	0.0078	0.0082	0.0084	-0.0018	-0.0003	0.0024
$\tau = 0.50$	std	0.1832	0.1573	0.1218	0.1309	0.1168	0.0847
	mse	0.0336	0.0248	0.0149	0.0171	0.0136	0.0072
	bias	0.0107	0.0119	0.0078	0.0010	0.0018	0.0006
$\tau = 0.75$	std	0.2071	0.1793	0.1462	0.1477	0.1308	0.1039
	mse	0.0430	0.0322	0.0214	0.0218	0.0171	0.0108

Table 4.4: Results of Simulation 3, $N = 1000$

		Our method			Athey and Imbens		
		$\rho^2 = 0.1$	$\rho^2 = 0.5$	$\rho^2 = 0.9$	$\rho^2 = 0.1$	$\rho^2 = 0.5$	$\rho^2 = 0.9$
	bias	0.0056	0.0015	0.0009	-0.0031	-0.0107	-0.0150
$\tau = 0.25$	std	0.1304	0.1261	0.0932	0.1288	0.1238	0.0900
	mse	0.0170	0.0159	0.0087	0.0166	0.0154	0.0083
	bias	0.0000	0.0008	0.0002	-0.0097	-0.0094	-0.0119
$\tau = 0.50$	std	0.1160	0.0980	0.0733	0.1165	0.1002	0.0760
	mse	0.0134	0.0096	0.0054	0.0137	0.0101	0.0059
	bias	0.0006	-0.0042	0.0001	-0.0203	-0.0286	-0.0269
$\tau = 0.75$	std	0.1191	0.1013	0.0749	0.1226	0.1061	0.0812
	mse	0.0142	0.0103	0.0056	0.0154	0.0121	0.0073

Table 4.5: Summary Statistics

	Treatment	Control	Diff	P-val on Diff
Number of households	1260	2128		
Area of tobacco production (mu)	5.578	4.874	0.705	0.000
Age	41.119	41.522	-0.403	0.173
Household size	4.877	4.665	0.212	0.000
Education (Primary)	0.367	0.323	0.044	0.009
Education (Secondary)	0.602	0.338	0.263	0.000
Education (High school or College)	0.025	0.257	-0.232	0.000

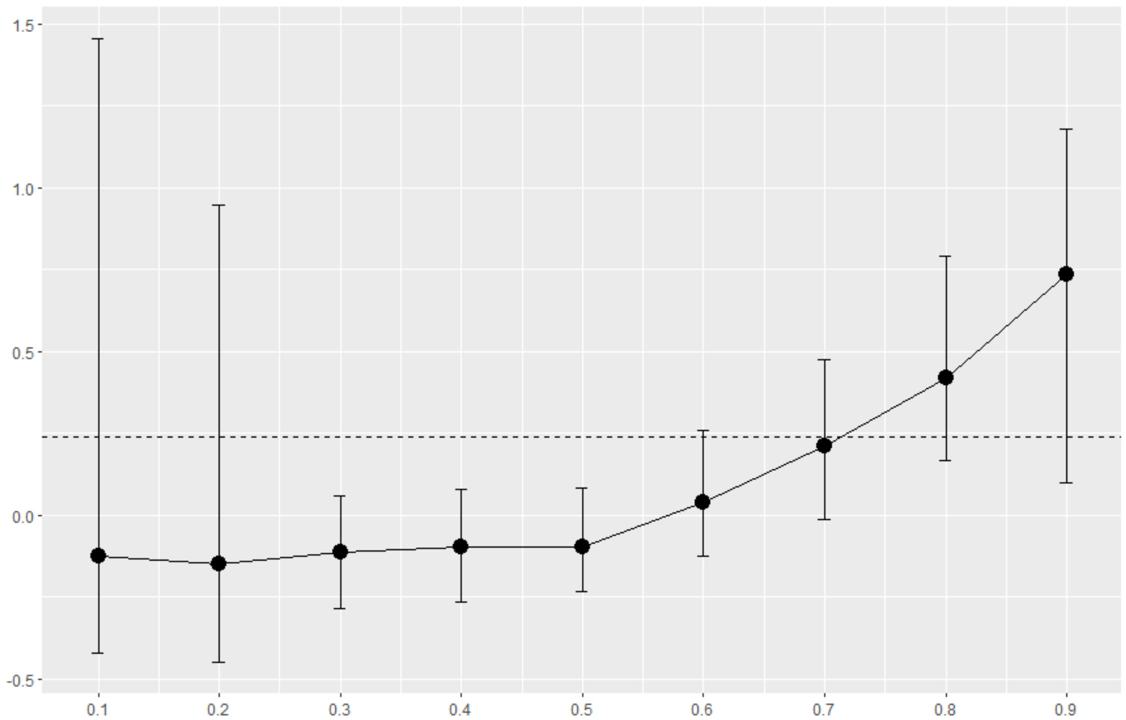

Figure 4.1: The estimates of the QTE and the 95 % confidence intervals. The dashed line denotes the DID estimate.

Bibliography

Abadie, A., J. Angrist, and G. Imbens (2002): "Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings," *Econometrica*, 70, 91–117.

Abrevaya, J. (1999): "Leapfrog estimation of a fixed-effects model with unknown transformation of the dependent variable," *Journal of Econometrics*, 93, 203–228.

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

Angrist, J., V. Chernozhukov, and I. Fernández-Val (2006): "Quantile regression under misspecification, with an application to the US wage structure," *Econometrica*, 74, 539–563.

Angrist, J. D. and A. B. Krueger (1991): "Does Compulsory School Attendance Affect Schooling and Earnings?" *The Quarterly Journal of Economics*, 106, 979–1014.

Athey, S. and G. W. Imbens (2006): "Identification and inference in nonlinear difference-in-differences models," *Econometrica*, 74, 431–497.

Blundell, R., D. Kristensen, and R. L. Matzkin (2013): "Control functions and simultaneous equations methods," *American Economic Review*, 103, 563–69.

Brown, D. J. and M. H. Wegkamp (2002): "Weighted Minimum Mean–Square Distance from Independence Estimation," *Econometrica*, 70, 2035–2051.

Caetano, C. and J. C. Escanciano (2017): "Identifying Multiple Marginal Effects with a Single Instrument," Tech. rep., working paper.

Cai, J. (2016): "The impact of insurance provision on household production and financial decisions," *American Economic Journal: Economic Policy*, 8, 44–88.

Callaway, B. and T. Li (2017): "Quantile Treatment Effects in Difference in Differences Models with Panel Data," .

Carrasco, M. and J.-P. Florens (2000): "Generalization of GMM to a continuum of moment conditions," *Econometric Theory*, 16, 797–834.

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

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

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

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

Chernozhukov, V. and C. Hansen (2004): "The effects of 401 (k) participation on the wealth distribution: an instrumental quantile regression analysis," *the Review of Economics and Statistics*, 86, 735–751.

——— (2005): "An IV model of quantile treatment effects," *Econometrica*, 73, 245–261.

——— (2006): "Instrumental quantile regression inference for structural and treatment effect models," *Journal of Econometrics*, 132, 491–525.

Chesher, A. (2003): "Identification in nonseparable models," *Econometrica*, 71, 1405–1441.

——— (2007): "Instrumental values," *Journal of Econometrics*, 139, 15–34.

——— (2010): "Instrumental variable models for discrete outcomes," *Econometrica*, 78, 575–601.

D'Haultfœuille, X. and P. Février (2015): "Identification of nonseparable triangular models with discrete instruments," *Econometrica*, 83, 1199–1210.

D'Haultfoeuille, X., S. Hoderlein, and Y. Sasaki (2013): "Nonlinear difference-in-differences in repeated cross sections with continuous treatments," Tech. rep., Boston College Department of Economics.

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

Firpo, S. (2007): "Efficient semiparametric estimation of quantile treatment effects," *Econometrica*, 75, 259–276.

Florens, J.-P., J. J. Heckman, C. Meghir, and E. Vytlacil (2008): "Identification of treatment effects using control functions in models with continuous, endogenous treatment and heterogeneous effects," *Econometrica*, 76, 1191–1206.

Frölich, M. and B. Melly (2013): "Unconditional quantile treatment effects under endogeneity," *Journal of Business & Economic Statistics*, 31, 346–357.

Galvao, A. and A. Poirier (2017): "Quantile regression random effects," .

Giné, E. and J. Zinn (1990): "Bootstrapping general empirical measures," *The Annals of Probability*, 851–869.

Giustinelli, P. (2011): "Non-parametric bounds on quantiles under monotonicity assumptions: with an application to the Italian education returns," *Journal of Applied Econometrics*, 26, 783–824.

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, 2105–2152.

Hahn, J. (1996): "A note on bootstrapping generalized method of moments estimators," *Econometric Theory*, 12, 187–197.

Havnes, T. and M. Mogstad (2015): "Is universal child care leveling the playing field?" *Journal of Public Economics*, 127, 100–114.

Hoderlein, S. (2011): "How many consumers are rational?" *Journal of Econometrics*, 164, 294–309.

Hoderlein, S. and E. Mammen (2007): "Identification of marginal effects in nonseparable models without monotonicity," *Econometrica*, 75, 1513–1518.

——— (2009): "Identification and estimation of local average derivatives in non-separable models without monotonicity," *The Econometrics Journal*, 12, 1–25.

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

Imbens, G. W. and J. D. Angrist (1994): "Identification and estimation of local average treatment effects," *Econometrica*, 62, 467–475.

Imbens, G. W. and W. K. Newey (2009): "Identification and estimation of triangular simultaneous equations models without additivity," *Econometrica*, 77, 1481–1512.

Ishihara, T. (2017): "Partial Identification of Nonseparable Models using Binary Instruments," *arXiv preprint arXiv:1707.04405*.

——— (2019): "Identification and estimation of time-varying nonseparable panel data models without stayers," *Journal of Econometrics*.

James, S., T. Lahti, and H. W. Hoynes (2006): "What mean impacts miss: Distributional effects of welfare reform experiments," *The American Economic Review*, 96, 988–1012.

Jun, S. J. (2009): "Local structural quantile effects in a model with a nonseparable control variable," *Journal of Econometrics*, 151, 82–97.

Kasy, M. (2011): "Identification in triangular systems using control functions," *Econometric Theory*, 27, 663–671.

Koenker, R. and G. Bassett (1978): "Regression quantiles," *Econometrica*, 33–50.

Kottelenberg, M. J. and S. F. Lehrer (2017): "Targeted or Universal Coverage? Assessing Heterogeneity in the Effects of Universal Child Care," *Journal of Labor Economics*, 35, 609–653.

Lechner, M. (2011): "The estimation of causal effects by difference-in-difference methods," *Foundations and Trends® in Econometrics*, 4, 165–224.

Macours, K., N. Schady, and R. Vakis (2012): "Cash Transfers, Behavioral Changes, and Cognitive Development in Early Childhood: Evidence from a Randomized Experiment," *American Economic Journal: Applied Economics*, 4, 247–273.

Manski, C. F. (1987): "Semiparametric analysis of random effects linear models from binary panel data," *Econometrica: Journal of the Econometric Society*, 357–362.

——— (1997): "Monotone treatment response," *Econometrica*, 1311–1334.

Martincus, C. V. and J. Carballo (2010): "Beyond the average effects: The distributional impacts of export promotion programs in developing countries," *Journal of Development Economics*, 92, 201–214.

Masten, M. A. and A. Torgovitsky (2016): "Identification of instrumental variable correlated random coefficients models," *Review of Economics and Statistics*, 98, 1001–1005.

Matzkin, R. L. (2003): "Nonparametric estimation of nonadditive random functions," *Econometrica*, 71, 1339–1375.

Melly, B. and G. Santangelo (2015): "The changes-in-changes model with covariates," *Universität Bern, Bern.*

Newey, W. K. and D. McFadden (1994): "Large sample estimation and hypothesis testing," *Handbook of econometrics*, 4, 2111–2245.

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

Okumura, T. and E. Usui (2014): "Concave-monotone treatment response and monotone treatment selection: With an application to the returns to schooling," *Quantitative Economics*, 5, 175–194.

Pakes, A. and D. Pollard (1989): "Simulation and the asymptotics of optimization estimators," *Econometrica: Journal of the Econometric Society*, 1027–1057.

Shaikh, A. M. and E. J. Vytlacil (2011): "Partial identification in triangular systems of equations with binary dependent variables," *Econometrica*, 79, 949–955.

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

——— (2017): "Minimum distance from independence estimation of nonseparable instrumental variables models," *Journal of Econometrics*, 199, 35–48.

van der Vaart, A. and J. Wellner (1996): *Weak Convergence and Empirical Processes: With Applications to Statistics*, Springer Science & Business Media.