Quantile Regression with Interval Data

Arie Beresteanu* Yuya Sasaki†

October 23, 2017‡

Abstract

This paper investigates the identification of quantiles and quantile regression parameters when observations are set valued. We define the identification set of quantiles of random sets in a way that extends the definition of quantiles for regular random variables. We then give sharp characterization of this set by extending concepts from random set theory. For quantile regression parameters, we show that the identification set is characterized by a system of conditional moment inequalities. This characterization extends that of parametric quantile regression for regular random variables. Estimation and inference theories are developed for continuous cases, discrete cases, nonparametric conditional quantiles, and parametric quantile regressions. A fast computational algorithm of set linear programming is proposed. Monte Carlo experiments support our theoretical properties.

Keywords: Partial Identification, Random Sets, Quantile Regression, Quantile Sets.

*Department of Economics, University of Pittsburgh, arie@pitt.edu.

†Department of Economics, Vanderbilt University, yuya.sasaki@vanderbilt.edu.

‡We have benefited from discussions with Francesca Molinari, Ilya Molchanov, and Roee Teper, and useful comments by participants in NASM 2016. All remaining errors are ours. This research was sponsored in part by NSF grant SES-0922373.

1 Introduction

Set valued observations are common in data based on surveys. One type of set valued data is generated by a response to a questioner that offers a distinct set of intervals to choose from. Another type of interval data is generated by willingness to pay surveys where respondents choose the minimum and/or maximum amount. Several papers investigate the impact of this kind of imprecise data on the identification of model parameters under various assumptions. Simple econometric methods, such as to apply the OLS using the midpoint of willingness-to-pay interval as a dependent variable, have long been known to suffer from substantial biases – see Cameron and Huppert (1989).

More recently, set identification and set inference approaches are proposed to solve this issue under a number of econometric contexts. Examples include the following list of papers. Binary choice models with interval regressors are discussed in Manski and Tamer (2002). Mean regressions when the outcome variables are interval valued are discussed in Beresteanu and Molinari (2008). These two models are generalized in Beresteanu, Molchanov, Molinari (2011). Median regression models under endogeneity when the outcome variables are censored are discussed in Hong and Tamer (2003). Median regression models when the outcome variables are endogenously censored are discussed in Khan and Tamer (2009) and Khan, Ponomareva, and Tamer (2011). Quantile panel data models when the outcome variables are censored are discussed in Li and Oka (2015).

To the best of our knowledge, no preceding paper in this literature has discussed quantile regression models with general quantile ranks and general set valued data. Some empirical papers (e.g., O'Garra and Mourato, 2007; Gamper-Rabindran and Timmins, 2013), however, use quantiles and quantile regressions where outcome data are interval-valued by taking the midpoint of the interval as the representative value. In this light, this paper investigates the identification of quantiles and of quantile regressions when the outcome variable is set valued. Estimators and their large sample properties are developed as well.

We first identify and estimate the unconditional and conditional quantiles of a random set. The concept of the quantile set of a random set is introduced. The quantile set is shown to be identified by the containment and capacity quantiles which we define in Section 2. We then discuss how this set can be estimated and derive asymptotic properties of the estimator in Section 3.

The identification argument for conditional quantiles is extended to set identification of quantile regression functions depending on a finite number of parameters in Section 4. We show that this identification set is defined by a system of conditional moment inequalities. These inequalities involve the cumulative containment and cumulative capacity functionals, which characterize the identification set for the quantile regression parameters. We also show that the sharp identification set is convex if the quantile regression is linear in parameters.

Generally, computing the identification set for quantile regression parameters can be computationally expensive without a prior knowledge of an approximate region in which these parameters should be. We next suggest a solution for linear quantile regression. Estimation of parameters for linear quantile regressions can be written as minimization of check loss function (see, Koenker and Bassett (1978) and Koenker (2005)). The solution to this minimization problem can be characterized as the solution to a linear programming (LP) problem. The set estimate is thus characterized as the set of solutions to a collection of LP problems. In Section 5, we propose a feasible computational algorithm to produce this set of infinite LP solutions by solving a finite number of LP problems.

We conduct a series of Monte Carlo experiments in Section 6, and demonstrate that the proposed methods work with the models we test.[1]

Literature: A number of notable papers in the literature discuss identification of quantiles and/or quantile regressions under set-valued observed outcomes. One of the most common causes of set-valued outcomes is censoring. Powell (1984) provides an estimator for the linear median regression model where the outcome variable is censored. Manski (1985) discusses identification

[1]The first version of this draft contained an empirical illustration in addition, which has been removed from the current version to economize the presentation. The original draft is available upon request from the authros.

of the linear median regression model where econometricians only observe the sign of an outcome variable. Hong and Tamer (2003) discuss inference on the linear median regression model where the outcome variable is censored and regressors are endogenous. Khan and Tamer (2009) discuss inference on the linear median regression model where the outcome variable is endogenously censored, for which Khan, Ponomareva, and Tamer (2011) provide the sharp identified region. The model considered in this paper includes the case of censored outcome variables as instances of set-valued outcome variables. A bottom-coded outcome $y = c$ can be treated as the set outcome $Y = (-\infty, c]$. Furthermore, compared with these preceding papers, we consider generalized quantile ranks $\tau \in (0, 1)$ in addition to the median $\tau = 0.5$. When we consider linear quantile regression models, we focus on exogenous regressors. Our framework, however, allows for a very large range of endogenous censoring for arbitrary set-valued outcomes. More recently, Li and Oka (2015) consider linear quantile regressions with a censored outcome variable in the framework of Rosen (2012). We do not deal with panel data, but the model considered in this paper includes the case of censored outcome variables as argued above. While the source of partial identification is not interval-valued outcomes, also related are partial identification of nonseparable models investigated by Chesher (2005, 2010) and generalized by Chesher and Rosen (2015). Also related is the literature on multivariate quantiles and quantile regression (Hallin, Paindaveine, and Šiman, 2010; Carlier, Chernozhukov, and Galichon, 2016; Chernozhukov, Galichon, Hallin, and Henry, 2017) in which quantiles are treated to be set-valued.

Notations and definitions: We introduce basic notations and definitions partly following those of Molchanov (2005). Let $(\Omega, \Im, \mathbf{P})$ be a complete probability space on which all random variables and sets are defined. Let $\mathcal{K}(\mathbb{R})$ denote the collection of all closed sets in $\mathbb{R}$. For an $\mathbb{R}$-valued random variable y, let $F_y(t) = \mathbf{P}(\{\omega : y(\omega) \in (-\infty, t]\})$ define the cumulative distribution function F_y of y. When it exists, the probability density function is denoted by f_y. For $\tau \in (0, 1)$, let $q_y(\tau) = \inf\{t : F_y(t) \geq \tau\}$ denote the τ-th quantile of y. A random variable $y : \Omega \to \mathbb{R}$ is a measurable selection of $Y : \Omega \to \mathcal{K}(\mathbb{R})$ if $y(\omega) \in Y(\omega)$ $\mathbf{P}$-a.s. The set of selections of Y is denoted by $Sel(Y)$. The containment functional C_Y and capacity functional T_Y of Y are defined by $C_Y(K) = \mathbf{P}(\{\omega : Y(\omega) \subset K\})$ and $T_Y(K) = \mathbf{P}(\{\omega : Y(\omega) \cap K \neq \emptyset\})$, respectively. For two sets, A and B, in a finite dimensional Euclidean space $\left(\mathbb{R}^k, \|\|\right)$, the directed Hausdorff distance from A to B is

$$d_H(A, B) = \sup_{a \in A} \inf_{b \in B} \|a - b\|$$

and the Hausdorff distance between A and B is

$$H(A, B) = \max\{d_H(A, B), d_H(B, A)\}.$$

2 Quantiles of Random Sets: Identification

We start by discussing identification and estimation of unconditional quantiles. Let $y^* : \Omega \to \mathbb{R}$ be a random variable, let $Y : \Omega \to \mathcal{K}(\mathbb{R})$ be a random set in $\mathbb{R}$, and let $\tau \in (0, 1)$. Throughout, we assume that $Y(\omega)$ is non-empty $\mathbf{P}$-a.s. y^* is unobserved, and we assume that $y^*(\omega) \in Y(\omega)$, $\mathbf{P}$-a.s. We would like to learn about $q_{y^*}(\tau)$. Define the τ-th quantile set of Y by

$$\Theta_0^Y(\tau) = \{q_y(\tau) : y \in Sel(Y)\}.$$

This is, by definition, the identification set for $q_{y^*}(\tau)$. In other words, with no further information the only thing we can say about $q_{y^*}(\tau)$ is that $q_{y^*}(\tau) \in \Theta_0^Y(\tau)$.

For any $t \in \mathbb{R}$, define $\tilde{C}_Y(t) = C_Y((-\infty, t])$ and $\tilde{T}_Y(t) = T_Y((-\infty, t])$ to be the cumulative containment and cumulative capacity functionals, respectively. Note that $\tilde{C}_Y$ and $\tilde{T}_Y$ are monotone increasing and right continuous. For $\tau \in (0, 1)$, define

$$\begin{aligned}\tilde{C}_Y^{-1}(\tau) &= \inf\left\{t : \tilde{C}_Y(t) \geq \tau\right\} \text{ and} \\ \tilde{T}_Y^{-1}(\tau) &= \inf\left\{t : \tilde{T}_Y(t) \geq \tau\right\}\end{aligned}$$

to be the containment and capacity quantiles of Y, respectively.

Theorem 2.1 *For every $\tau \in (0, 1)$, $\Theta_0^Y(\tau) \subset \left[\tilde{T}_Y^{-1}(\tau), \tilde{C}_Y^{-1}(\tau)\right]$.*

Proof. Let $t \in \Theta_0^Y(\tau) = \{q_y(\tau) : y \in Sel(Y)\}$. By the definitions of $\tilde{C}_Y$ and $\tilde{T}_Y$, $y \in Sel\,(Y)$ implies $\tilde{C}_Y\,(t) \leq F_y\,(t) \leq \tilde{T}_Y\,(t)$ for all $t \in \mathbb{R}$.[2] Therefore, for any $\tau \in (0,1)$, $\inf\left\{t : \tilde{T}_Y\,(t) \geq \tau\right\} \leq \inf\{t : F_y\,(t) \geq \tau\} \leq \inf\left\{t : \tilde{C}_Y\,(t) \geq \tau\right\}$. This proves the theorem. ■

This theorem generalizes Corollary 1.2.1 in Manski (2003). When $Y = \{y\}$ for a random variable y, $\Theta_0^Y\,(\tau) = \{q_y\,(\tau)\}$. Molchanov (1990) defines a quantile of a random set in a different way. His definition does not provide the same generalization of a quantile function in case of a $\mathbb{R}$-valued random set that we need in this paper.

The other direction of set inclusion for Theorem 2.1 need not hold. To see this, consider a simple random set Y such that $Y(\omega) = [-2,-1] \cup [1,2]$ $\mathbf{P}$-a.s. Then, we have $0 \in [-2,2] = \left[\tilde{T}_Y^{-1}(0.5), \tilde{C}_Y^{-1}(0.5)\right]$, but $0 \notin \Theta_0^Y(0.5) \subset [-2,-1] \cup [1,2]$. This example illustrates why a 'hole' in the set $Y(\omega)$ fails to establish the other direction of set inclusion. This observation in fact can be generalized. The following theorem shows that the identification set equality holds without such holes.

Theorem 2.2 *If Y is a convex valued random set in $\mathbb{R}$ with $Sel(Y) \neq \emptyset$, then $\left(\tilde{T}_Y^{-1}\,(\tau), \tilde{C}_Y^{-1}\,(\tau)\right) \subset \Theta_0^Y\,(\tau)$ for all $\tau \in (0,1)$. Furthermore, if in addition $\inf Y(\omega) \in Y(\omega) > -\infty$ $\mathbf{P}$-a.s., then $\tilde{T}_Y^{-1}\,(\tau) \in \Theta_0^Y\,(\tau)$ for all $\tau \in (0,1)$. Similarly, if in addition $\sup Y(\omega) < \infty$ $\mathbf{P}$-a.s., then $\tilde{C}_Y^{-1}\,(\tau) \in \Theta_0^Y\,(\tau)$ for all $\tau \in (0,1)$.*

Proof. If $\left(\tilde{T}_Y^{-1}\,(\tau), \tilde{C}_Y^{-1}\,(\tau)\right)$ is empty, then the first claim in the theorem is trivially satisfied. Now, suppose that it is non-empty. Fix $\tau \in (0,1)$ and take $t \in \left(\tilde{T}_Y^{-1}\,(\tau), \tilde{C}_Y^{-1}\,(\tau)\right)$. Let $\Omega_L = \{\omega : \sup Y\,(\omega) < t\}$, $\Omega_U = \{\omega : \inf Y\,(\omega) > t\}$, and $\Omega_M = \Omega \setminus (\Omega_L \cup \Omega_U)$. By definition, $\mathbf{P}\,(\Omega_L) < \tau$ and $\mathbf{P}\,(\Omega_U) < 1 - \tau$. By $Sel(Y) \neq \emptyset$, we choose $y \in Sel(Y)$. Let

$$\tilde{y}\,(\omega) = \begin{cases} y(\omega) & \omega \in \Omega_L \cup \Omega_U \\ t & \omega \in \Omega_M \end{cases}$$

[2]This implication corresponds to the necessity part of Artstein's Lemma, which does not restrict to compact sets. See discussion in Beresteanu, Molchanov, Molinari (2012).

Since $Y(\omega)$ is convex, $t \in Y(\omega)$ for all $\omega \in \Omega_M$, and thus the random variable $\tilde{y}$ defined above is a selection of Y. By construction, $q_{\tilde{y}}(\tau) = t$ and thus $\left(\tilde{T}_Y^{-1}(\tau), \tilde{C}_Y^{-1}(\tau)\right) \subset \Theta_0^Y(\tau)$.

Suppose that $\inf Y(\omega) > -\infty$ $\mathbf{P}$-a.s. Then, since $Y(\omega)$ is closed, the random variable $\tilde{y}$ defined by $\tilde{y}(\omega) = \inf Y(\omega)$ is a selection of Y. It also satisfies $T_Y^{-1}(\tau) = \inf\{t : \mathbf{P}(Y \cap (-\infty, t] \neq \emptyset) \geq \tau\} = \inf\{t : F_{\tilde{y}}(t) \geq \tau\}$. Therefore, $\tilde{T}_Y^{-1}(\tau) = q_{\tilde{y}}(\tau) \in \Theta_0^Y(\tau)$.

Finally, suppose that $\sup Y(\omega) < \infty$ $\mathbf{P}$-a.s. Then, since $Y(\omega)$ is closed, the random variable $\tilde{y}$ defined by $\tilde{y}(\omega) = \sup Y(\omega)$ is a selection of Y. It also satisfies $C_Y^{-1}(\tau) = \inf\{t : \mathbf{P}(Y \subset (-\infty, t]) \geq \tau\} = \inf\{t : F_{\tilde{y}}(t) \geq \tau\}$. Therefore, $\tilde{C}_Y^{-1}(\tau) = q_{\tilde{y}}(\tau) \in \Theta_0^Y(\tau)$. ■

Theorems 2.1 and 2.2 together show that $\Theta_0^Y(\tau) = \left[\tilde{T}_Y^{-1}(\tau), \tilde{C}_Y^{-1}(\tau)\right]$ for all $\tau \in (0,1)$, if Y has a selection and $Y(\omega)$ is a compact interval $\mathbf{P}$-a.s. In other words, $\left[\tilde{T}_Y^{-1}(\tau), \tilde{C}_Y^{-1}(\tau)\right]$ is a sharp characterization of the identification set $\Theta_0^Y(\tau)$. There are alternative sufficient conditions for the condition, $Sel(Y) \neq \emptyset$, of Theorem 2.2. One such condition is that Y is closed-valued and non-empty a.s., as stated in the Fundamental Selection Theorem (Molchanov, 2005, Theorem 2.13).

Outcome variables which are reported as convex valued sets include several important cases that an empirical researcher may encounter. First is the case where Y is generated by a response to a questioner that offers a distinct set of intervals to choose from. The case above allows these intervals to be distinct, intersect or even be included in each other. A second type of data which is covered by this condition is willingness to pay surveys. Contingent valuation surveys which employ the collapsing interval method are a prominent example for this case. Sometimes in these surveys, the interval is indeed $[c, \infty)$ and quantiles can be estimated while expectations cannot.

3 Estimation and Inference

Throughout this section, we assume that the conditions of Theorem 2.2 hold, and investigate methods of estimation of the identification set $\Theta_0^Y(\tau) = \left[\tilde{T}_Y^{-1}(\tau), \tilde{C}_Y^{-1}(\tau)\right]$. While we focus on $\Theta_0^Y(\tau)$ here, estimation and large sample properties of the cumulative capacity functional $T_Y(t)$

and the cumulative containment functional $C_Y(\tau)$ themselves are presented in Appendix A.1 for completeness.

For simplicity, we assume that $Y_1, Y_2, ...$ are independently and identically distributed throughout, although this assumption can be replaced by alternative dependence assumptions for relevant laws of large numbers and central limit theorems. Let $Y_i = [a_i, b_i]$ and let $a_{(1)} \le a_{(2)} \le ... \le a_{(n)}$ and $b_{(1)} \le b_{(2)} \le ... \le b_{(n)}$ be the order statistics of $\{a_i\}_{i=1}^n$ and $\{b_i\}_{i=1}^n$, respectively. For $t \in \mathbb{R}$, let $\lfloor t \rfloor$ denote the biggest integer smaller than t. Then, for $\tau \in (0,1)$, define

$$\begin{aligned} \tilde{T}_n^{-1}(\tau) &= a_{(\lfloor n\tau \rfloor)} \qquad \text{and} \\ \tilde{C}_n^{-1}(\tau) &= b_{(\lfloor n\tau \rfloor)}. \end{aligned}$$

to be our estimators of the lower bound and upper bound of $\Theta_0^Y(\tau)$, respectively. Specifically, we define $\hat{\Theta}_0^Y(\tau) = [\tilde{T}_n^{-1}(\tau), \tilde{C}_n^{-1}(\tau)] = [a_{(\lfloor n\tau \rfloor)}, b_{(\lfloor n\tau \rfloor)}]$ as our estimator for $\Theta_0^Y(\tau)$. The following theorem shows that this estimator is strongly consistent.

Theorem 3.1 *Suppose the random set takes the form $Y = [a, b]$ $\mathbf{P}$-a.s. If Y_1, Y_2, ... are independently and identically distributed, then for $\tau \in (0, 1)$*

$$H\left(\hat{\Theta}^Y(\tau), \Theta_0^Y(\tau)\right) \overset{a.s.}{\to} 0.$$

A proof of this theorem is provided in Appendix A.2. In addition to this consistency result, we can derive the asymptotic distribution and/or super-consistency results, depending on the type of distribution of intervals. As such, we branch into a number of different cases in the following subsections.

3.1 Continuous Case

Consider first the case where $Y = [a, b]$ and both a and b are continuous random variables. Specifically, we assume that their joint cumulative distribution function $F_{a,b}$ is twice continuously differentiable, and their marginal density functions are denoted by f_a and f_b. Under this condition, Theorem 2.1 of Babu and Rao (1988) yields

$$\sqrt{n}\left(a_{(\lfloor n\tau\rfloor)} - q_a(\tau), b_{(\lfloor n\tau\rfloor)} - q_b(\tau)\right)' \xrightarrow{D} (z_L(\tau), z_U(\tau))' \sim N\left(\mathbf{0}, \Sigma(\tau)\right), \tag{3.1}$$

where

$$\Sigma(\tau) = \begin{pmatrix} \frac{\tau(1-\tau)}{f_a(q_a(\tau))^2} & \frac{F_{a,b}(q_a(\tau),q_b(\tau))-\tau^2}{f_a(q_a(\tau))f_b(q_b(\tau))} \\ \frac{F_{a,b}(q_a(\tau),q_b(\tau))-\tau^2}{f_a(q_a(\tau))f_b(q_b(\tau))} & \frac{\tau(1-\tau)}{f_b(q_b(\tau))^2} \end{pmatrix}. \tag{3.2}$$

Once the asymptotic joint normal distribution in (3.1) has been established, the methods developed in Beresteanu and Molinari (2008) for the Hausodorff and directed Hausdorff distance can be used as well as for hypothesis testing using the asymptotic distribution and using bootstrap. These results are stated here for completeness. We introduce the short-hand notations, $(x)_+ = \max\{x, 0\}$ and $(x)_- = \max\{-x, 0\}$.

Theorem 3.2 *Suppose the random set takes the form $Y = [a, b]$ $\mathbf{P}$-a.s. where both a and b are continuously distributed with a twice continuously differentiable joint cumulative distribution function. If Y_1, Y_2, ... are independently and identically distributed, then for $\tau \in (0, 1)$ such that $f_a(q_a(\tau)) \neq 0$ and $f_b(q_b(\tau)) \neq 0$,*

$$\sqrt{n}H\left(\hat{\Theta}^Y(\tau), \Theta_0^Y(\tau)\right) \xrightarrow{D} \max\left\{|z_L(\tau)|, |z_U(\tau)|\right\}, \tag{3.3}$$

$$\sqrt{n}d_H\left(\hat{\Theta}^Y(\tau), \Theta_0^Y(\tau)\right) \xrightarrow{D} \max\left\{(z_L(\tau))_+, (z_U(\tau))_-\right\}, \qquad \textit{and} \tag{3.4}$$

$$n\left(d_H\left(\hat{\Theta}^Y(\tau), \Theta_0^Y(\tau)\right)\right)^2 \xrightarrow{D} \max\left\{(z_L(\tau))_+^2, (z_U(\tau))_-^2\right\}, \tag{3.5}$$

where the random vector $(z_L(\tau), z_U(\tau))$ is distributed according to (3.1)–(3.2).

This result establishes the equivalence between the QLR and the Wald test statistics for the empirical set quantile as well.

The quantile function and its empirical counterpart are (left) inverses of the distribution function and its empirical counterpart, respectively, which exhibits a functional convergence to a tight (Gaussian) process. Therefore, we can conduct uniform analyses over a continuum of quantiles.

Suppose that the marginal cumulative distribution functions, F_a and F_b, as well as the joint cumulative distribution function $F_{a,b}$ are twice continuously differentiable. Let the marginal density functions, f_a and f_b, be positive on their respective support – this condition can be relaxed at the expense of a restricted domain of the uniform inference. Under this condition, Theorems 4.1 and 4.2 of Babu and Rao (1988) imply that the process $\{Z_n\}_n$ defined by

$$Z_n(\tau, t) \equiv \sqrt{n}\left(a_{(\lfloor n\tau \rfloor)} - q_a(\tau), b_{(\lfloor nt \rfloor)} - q_b(t)\right)$$

converges weakly to the mean-zero Gaussian process $Z = (Z_L, Z_U)$ where its covariance function is given by

$$\begin{aligned} E[Z_L(\tau)Z_L(t)] &= \frac{\tau \wedge t - \tau t}{f_a(q_a(\tau))f_a(q_a(t))} \\ E[Z_L(\tau)Z_U(t)] &= \frac{F_{a,b}(q_a(\tau), q_b(t)) - \tau t}{f_a(q_a(\tau))f_a(q_b(t))} \\ E[Z_U(\tau)Z_U(t)] &= \frac{\tau \wedge t - \tau t}{f_a(q_b(\tau))f_a(q_b(t))} \end{aligned} \tag{3.6}$$

Here, we use the short-hand notation $\tau \wedge t := \min\{\tau, t\}$ following the convention. With this weak convergence of the joint process, together with the continuous mapping theorem, we can state the uniform counterpart of Theorem 3.2. We let $\rightsquigarrow$ denote the weak convergence.

Theorem 3.3 *Suppose the random set takes the form* $Y = [a, b]$ $\mathbf{P}$*-a.s. where both* a *and* b *are continuously distributed with a twice continuously differentiable joint cumulative distribution function and twice continuously differentiable marginal cumulative distribution functions. Suppose also that the marginal density functions are positive on their respective supports. If* Y_1, Y_2, $\ldots$ *are independently and identically distributed, then*

$$\sqrt{n}H\left(\hat{\Theta}^Y(\cdot), \Theta_0^Y(\cdot)\right) \rightsquigarrow \max\left\{|Z_L(\cdot)|, |Z_U(\cdot)|\right\}, \tag{3.7}$$

$$\sqrt{n}d_H\left(\hat{\Theta}^Y(\cdot), \Theta_0^Y(\cdot)\right) \rightsquigarrow \max\left\{(Z_L(\cdot))_+, (Z_U(\cdot))_-\right\}, \quad \textit{and} \tag{3.8}$$

$$n\left(d_H\left(\hat{\Theta}^Y(\cdot), \Theta_0^Y(\cdot)\right)\right)^2 \rightsquigarrow \max\left\{(Z_L(\cdot))_+^2, (Z_U(\cdot))_-^2\right\}, \tag{3.9}$$

where $Z = (Z_L, Z_U)$ *is the mean-zero Gaussian process with its covariance function given by (3.6).*

3.2 Discrete Case

In many surveys, the respondent is given a list of brackets to choose from. In this case both a and b are discrete random variables. The results described in the previous section do not hold here. Assume $Y = [a, b]$ as before, and hence $\Theta_0^Y(\tau) = [q_a(\tau), q_b(\tau)]$ by Theorems 2.1 and 2.2, where q_a and q_b vary discretely. In the current subsection, we define the set estimator by $\hat{\Theta}^Y(\tau) = [a_{(\lceil n\tau \rceil)}, b_{(\lceil n\tau \rceil)}]$, where $\lceil t \rceil$ denotes the smallest integer greater than t. By Theorem 2 of Ramachandramurty and Rao (1973), we have

$$\begin{aligned}
&\mathbf{P}\left(r_n(a_{(\lceil n\tau \rceil)} - q_a(\tau)) \le 1\right) \to 1, \\
&\mathbf{P}\left(r_n(a_{(\lceil n\tau \rceil)} - q_a(\tau)) \le -1\right) \to 0, \\
&\mathbf{P}\left(r_n(b_{(\lceil n\tau \rceil)} - q_b(\tau)) \le 1\right) \to 1, \qquad \text{and} \\
&\mathbf{P}\left(r_n(b_{(\lceil n\tau \rceil)} - q_b(\tau)) \le -1\right) \to 0,
\end{aligned}$$

as $n \to \infty$, if $r_n \to \infty$ as $n \to \infty$. Note that r_n can diverge at an arbitrary rate as a function of n – even faster than $\sqrt{n}$. Thus, we obtain the following result by the continuous mapping theorem.

Theorem 3.4 *Suppose the random set takes the form $Y = [a, b]$ $\mathbf{P}$-a.s. where a and b are discretely distributed. If Y_1, Y_2, ... are independently and identically distributed, then for $\tau \in (0, 1)$*

$$\begin{aligned}
&r_n H\left(\hat{\Theta}^Y(\tau), \Theta_0^Y(\tau)\right) \xrightarrow{P} 0 \qquad \textit{and} \\
&r_n d_H\left(\hat{\Theta}^Y(\tau), \Theta_0^Y(\tau)\right) \xrightarrow{P} 0
\end{aligned}$$

for $r_n \to \infty$ as $n \to \infty$.

Theorem 3.4 suggests that the estimator for the identification region when Y is a discrete random set is super-consistent. Super-consistency may be useful in cases where estimating a discrete quantile set is just a first step in a two-step estimation procedure. On the other hand, a drawback to this result is that we do not obtain a root-n non-degenerate asymptotic normal distribution.

The lack of the ability to conduct inference with the naive quantile estimators is unfortunate. However, in the special case where the discrete boundaries, a and b, of the random set Y are given as a count data, we can allow for inference even in the discrete case by using the idea of "jittering" by Machado and Silva (2005) – an alternative approach is via smoothing (Kordas, 2006). Suppose that a and b are supported in the set $\mathcal{J} = \{0, 1, ..., J-1\}$ of cardinality $J \in \mathbb{N}$. Construct the random variables $\tilde{a} = a + u$ and $\tilde{b} = b + v$ where $u, v \sim \text{Uniform}(0,1)$ and (u, v) is independent of (a, b). Let $F_{\tilde{a},\tilde{b}}$ denote the joint cumulative distribution function of $(\tilde{a}, \tilde{b})$, which is identified from the two-dimensional convolution of the distributions of (a, b) and (u, v). As a result of the convolution, the distribution of $F_{\tilde{a},\tilde{b}}$ is differentiable infinitely many times on $(\mathcal{J} \oplus (0,1))^2$. Furthermore, the above construction of the mixture distribution yields the marginal quantiles relations

$$q_{\tilde{a}}(\tau) = q_a(\tau) + \frac{\tau - \sum_{j=0}^{q_a(\tau)-1} \Pr(a = j)}{\Pr(a = q_a(\tau))}$$
$$q_{\tilde{b}}(\tau) = q_b(\tau) + \frac{\tau - \sum_{j=0}^{q_b(\tau)-1} \Pr(b = j)}{\Pr(b = q_b(\tau))}$$

See Machado and Silva (2005). As such, we can define a new set estimator by

$$\tilde{\Theta}^Y(\tau) = \left[\check{a}(\tau), \check{b}(\tau)\right] := \left[\tilde{a}_{(\lfloor n\tau \rfloor)} - \frac{\tau - \sum_{j=0}^{a_{(\lceil n\tau \rceil)}-1} \widehat{\Pr}(a = j)}{\widehat{\Pr}(a = a_{(\lceil n\tau \rceil)})}, \tilde{b}_{(\lfloor n\tau \rfloor)} - \frac{\tau - \sum_{j=0}^{b_{(\lceil n\tau \rceil)}-1} \widehat{\Pr}(b = j)}{\widehat{\Pr}(b = b_{(\lceil n\tau \rceil)})}\right]$$

where $\widehat{\Pr}(a = j)$ and $\widehat{\Pr}(b = j)$ denote the empirical mass for each $j \in \mathcal{J}$. Note that, in this estimator, we distinguish the $\sqrt{n}$-consistent estimator $(\tilde{a}_{(\lfloor n\tau \rfloor)}, \tilde{b}_{(\lfloor n\tau \rfloor)})'$ and the aforementioned super-consistent estimator $\sqrt{n}$-consistent estimator $(a_{(\lfloor n\tau \rfloor)}, b_{(\lfloor n\tau \rfloor)})'$ on purpose.

To analyze the asymptotic distribution of this estimator $\tilde{\Theta}^Y(\tau)$, we first need the joint asymptotic distribution of the $2(J+1)$-dimensional vector $\sqrt{n}(\tilde{a}_{(\lfloor n\tau \rfloor)} - q_{\tilde{a}}(\tau), \tilde{b}_{(\lfloor n\tau \rfloor)} - q_{\tilde{b}}(\tau), \widehat{\Pr}(a = 0) - \Pr(a = 0), ..., \widehat{\Pr}(a = J-1) - \Pr(a = J-1), \widehat{\Pr}(b = 0) - \Pr(b = 0), ..., \widehat{\Pr}(b = J-1) - \Pr(b = J-1))'$ which consists stochastic elements of the boundaries of the set estimator. For any $\tau \in (0,1)$ such that $(q_{\tilde{a}}(\tau), q_{\tilde{b}}(\tau)) \in (\mathcal{J} \oplus (0,1))^2$,

$$(3.10) \quad \sqrt{n} \begin{pmatrix} \left(\tilde{a}_{(\lfloor n\tau \rfloor)} - q_{\tilde{a}}(\tau), \tilde{b}_{(\lfloor n\tau \rfloor)} - q_{\tilde{b}}(\tau)\right)' \\ \left(\widehat{\Pr}(a=0) - \Pr(a=0), ..., \widehat{\Pr}(a=J-1) - \Pr(a=J-1)\right)' \\ \left(\widehat{\Pr}(b=0) - \Pr(b=0), ..., \widehat{\Pr}(b=J-1) - \Pr(b=J-1)\right)' \end{pmatrix} \xrightarrow{D} N\left(\mathbf{0}, \tilde{\Sigma}(\tau)\right),$$

where $\tilde{\Sigma}(\tau)$ is a $2(J+1) \times 2(J+1)$ matrix which is completely expressed in (A.1) in Appendix A.3. If $\Pr(a = q_a(\tau)) \neq 0$ and $\Pr(b = q_b(\tau)) \neq 0$, then we therefore obtain

$$(3.11) \quad \sqrt{n}\left(\check{a}(\tau) - q_a(\tau), \check{b}(\tau) - q_b(\tau)\right)' \xrightarrow{D} (\check{z}_L(\tau), \check{z}_U(\tau))' \sim N\left(\mathbf{0}, \Xi(\tau)\tilde{\Sigma}(\tau)\Xi(\tau)'\right),$$

where $\Xi(\tau) = (\Xi'_{1\cdot}, \Xi'_{2\cdot})'$ is a $2 \times 2(J+1)$ matrix, where the first row takes the form

$$\Xi_{1\cdot} = \left(1, \; 0, \; \left(\frac{1\{j \leq q_a(\tau)\}}{\Pr(a = q_a(\tau))} + 1\{j = q_a(\tau)\}\frac{\tau - \sum_{j'=0}^{q_a(\tau)-1}\Pr(a = j')}{\Pr(a = q_a(\tau))^2}\right)_{j=0}^{J-1}, \; 0, ..., 0\right)$$

and the second row takes the form

$$\Xi_{2\cdot} = \left(0, \; 1, \; 0, ..., 0, \; \left(\frac{1\{j \leq q_b(\tau)\}}{\Pr(a = q_b(\tau))} + 1\{j = q_b(\tau)\}\frac{\tau - \sum_{j'=0}^{q_b(\tau)-1}\Pr(b = j')}{\Pr(b = q_b(\tau))^2}\right)_{j=0}^{J-1}\right).$$

From this asymptotic joint normal distribution, we obtain the following theorem that can be used for inference on random sets where the boundaries are discrete.

Theorem 3.5 *Suppose the random set takes the form $Y = [a, b]$ $\mathbf{P}$-a.s. where both a and b are discretely distributed with support contained in $\mathcal{J} = \{0, 1, ..., J-1\}$. Construct the random variables $\tilde{a} = a + u$ and $\tilde{b} = b + v$ where $u, v \sim$ Uniform$(0,1)$ and (u, v) is independent of (a, b). For any $\tau \in (0,1)$ such that $(q_{\tilde{a}}(\tau), q_{\tilde{b}}(\tau)) \in (\mathcal{J} \oplus (0,1))^2$, $Pr(a = q_a(\tau)) \neq 0$, and $Pr(b = q_b(\tau)) \neq 0$,*

$$(3.12) \quad \sqrt{n}H\left(\check{\Theta}^Y(\tau), \Theta_0^Y(\tau)\right) \xrightarrow{D} \max\{|\check{z}_L(\tau)|, |\check{z}_U(\tau)|\},$$

$$(3.13) \quad \sqrt{n}d_H\left(\check{\Theta}^Y(\tau), \Theta_0^Y(\tau)\right) \xrightarrow{D} \max\{(\check{z}_L(\tau))_+, (\check{z}_U(\tau))_-\}, \qquad \textit{and}$$

$$(3.14) \quad n\left(d_H\left(\check{\Theta}^Y(\tau), \Theta_0^Y(\tau)\right)\right)^2 \xrightarrow{D} \max\left\{(\check{z}_L(\tau))_+^2, (\check{z}_U(\tau))_-^2\right\},$$

where the random vector $(\check{z}_L(\tau), \check{z}_U(\tau))$ is distributed according to (3.11).

3.3 Conditional Quantiles

Suppose that in addition to Y we observe a vector of p covariates, x, not including a constant. We are back to the case where Y is continuously distributed as in Section 3.1. We would like to estimate the sharp identification set $\Theta_{Y|x}(\tau \mid x^*)$ for the conditional quantile $q_{y|x}(\tau \mid x^*)$ at $x^* \in supp(x)$. The identification argument for the unconditional quantiles directly carry over to the conditional quantiles. For estimation, if X is a discrete random variable such that $\mathbf{P}(X = x^*) > 0$, then we can replace the capacity and containment functionals in previous subsections with their conditional counterparts and the analysis stays the same. The more interesting case is when x or some of its components are continuous and $\mathbf{P}(x = x^*) = 0$. The remainder of this subsection discusses this case.

For $\tau \in (0, 1)$, we introduce the check function ρ_τ defined by $\rho_\tau(u) = u \cdot (\tau - 1[u < 0])$. We propose to locally minimize the sample sum of this loss in a neighborhood of x^*. To this end, we use the p-dimensional vector of bandwidth parameters $h_n(\tau \mid x^*) = (h_{n1}(\tau \mid x^*), \ldots, h_{np}(\tau \mid x^*))$. Let $K(u) = \frac{1}{2} \cdot 1[|u| < 1]$ denote the indicator kernel, although we can substitute other kernel functions. For a short-hand notation, we write $K\left(\frac{x_i - x^*}{h_n(\tau|x^*)}\right) := \prod_{k=1}^{p} K\left(\frac{x_{ik} - x_k^*}{h_{nk}(\tau|x^*)}\right)$ for the product kernel. With these notations, the τ-th quantile of the local lower bound $a(\tau \mid x^*)$ is estimated by

$$\hat{a}(\tau \mid x^*) = \arg\min_a \sum_{i=1}^{n} \rho_\tau (a_i - a) \cdot K\left(\frac{x_i - x^*}{h_n(\tau \mid x^*)}\right).$$

Likewise, the τ-th quantile of the local upper bound $b(\tau \mid x^*)$ is estimated by

$$\hat{b}(\tau \mid x^*) = \arg\min_b \sum_{i=1}^{n} \rho_\tau (b_i - b) \cdot K\left(\frac{x_i - x^*}{h_n(\tau \mid x^*)}\right).$$

The local set estimator is thus given by $\hat{\Theta}_{Y|x}(\tau \mid x^*) = [\hat{a}(\tau \mid x^*), \hat{b}(\tau \mid x^*)]$. Finally, we introduce the short-hand notation $N_n(\tau \mid x^*) = \sum_{i=1}^{n} K\left(\frac{x_i - x^*}{h_n(\tau|x^*)}\right)$ for the local sample size. The following lemma summarizes conditions under which $(\hat{a}(\tau \mid x^*), \hat{b}(\tau \mid x^*))$ is asymptotically joint normal.

Lemma 3.1 *Suppose that (i)* (x_i, a_i, b_i) *is i.i.d., (ii) the distribution of* x *is absolutely continuous,* f_x *is bounded away from zero in a neighborhood of* x^**, and* f_x *is continuously differentiable with* *bounded derivatives in a neighborhood of* x^*, *(iii)* $f_{a|x}(\cdot \mid x^*)$ *is positive and Lipschitz continuous in a neighborhood of* $q_{a|x}(\tau \mid x^*)$, *(iv)* $f_{b|x}(\cdot \mid x^*)$ *is positive and Lipschitz continuous in a neighborhood of* $q_{b|x}(\tau \mid x^*)$, *(v)* $q_{a|x}(\tau \mid \cdot)$ *and* $q_{b|x}(\tau \mid \cdot)$ *are Hölder continuous with exponent* $\gamma > 0$, *and (vi)* $h_{nk}(\tau \mid x^*) = \kappa_k(\tau \mid x^*) \cdot n^{-1/(2\gamma+p)}$ *for each* $k \in \{1, \dots, p\}$. *Then for* $\tau \in (0,1)$ *such that* $f_{a|x}(q_{a|x}(\tau \mid x^*) \mid x^*) \neq 0$ *and* $f_{b|x}(q_{b|x}(\tau \mid x^*) \mid x^*) \neq 0$,

$$\sqrt{N_n(\tau \mid x^*)} \left(\hat{a}(\tau \mid x^*) - a(\tau \mid x^*), \hat{b}(\tau \mid x^*) - b(\tau \mid x^*) \right) \xrightarrow{D} (z_L(\tau \mid x^*), z_U(\tau \mid x^*)) \sim N(\mathbf{0}, \Sigma(\tau \mid x^*)),$$

where

$$\Sigma(\tau \mid x^*) = \begin{pmatrix} \frac{\tau(1-\tau)}{f_{a|x}(q_{a|x}(\tau|x^*)|x^*)^2} & \frac{F_{a,b|x^*}(q_{a|x}(\tau|x^*), q_{b|x}(\tau|x^*)) - \tau^2}{f_{a|x}(q_{a|x}(\tau|x^*)|x^*) \cdot f_{b|x}(q_{b|x}(\tau|x^*)|x^*)} \\ \frac{F_{a,b|x^*}(q_{a|x}(\tau|x^*), q_{b|x}(\tau|x^*)) - \tau^2}{f_{a|x}(q_{a|x}(\tau|x^*)|x^*) \cdot f_{b|x}(q_{b|x}(\tau|x^*)|x^*)} & \frac{\tau(1-\tau)}{f_{b|x}(q_{b|x}(\tau|x^*)|x^*)^2} \end{pmatrix}. \tag{3.15}$$

A proof is provided in Appendix A.4. The estimator considered above takes a specific and simple form due to the assumptions we make. Various extensions are possible. For example, we can allow for higher order local polynomial estimation with additional smoothness assumptions. Once the asymptotic joint distribution is obtained, the methods developed in Beresteanu and Molinari (2008) can be applied to obtain the asymptotic results analogous to (3.7)–(3.9).

Theorem 3.6 *Suppose the random set takes the form* $Y = [a, b]$ $\mathbf{P}$*-a.s. and the conditions of Lemma 3.1 are satisfied. Then for* $\tau \in (0,1)$ *such that* $f_{a|x}(q_{a|x}(\tau \mid x^*) \mid x^*) \neq 0$ *and* $f_{b|x}(q_{b|x}(\tau \mid x^*) \mid x^*) \neq 0$,

$$\begin{aligned} &\sqrt{N_n(\tau \mid x^*)} H \left(\hat{\Theta}_{Y|x}(\tau \mid x^*), \Theta_{Y|x}(\tau \mid x^*) \right) \xrightarrow{D} \max \left\{ |z_L(\tau \mid x^*)|, |z_U(\tau \mid x^*)| \right\}, \\ &\sqrt{N_n(\tau \mid x^*)} d_H \left(\hat{\Theta}_{Y|x}(\tau \mid x^*), \Theta_{Y|x}(\tau \mid x^*) \right) \xrightarrow{D} \max \left\{ (z_L(\tau \mid x^*))_+, (z_U(\tau \mid x^*))_- \right\}, \quad \text{and} \\ &N_n(\tau \mid x^*) \left(d_H \left(\hat{\Theta}_{Y|x}(\tau \mid x^*), \Theta_{Y|x}(\tau \mid x^*) \right) \right)^2 \xrightarrow{D} \max \left\{ (z_L(\tau \mid x^*))_+^2, (z_U(\tau \mid x^*))_-^2 \right\}. \end{aligned} \tag{3.16}$$

4 Parametric Quantile Regression Models

In previous sections we considered unconditional quantiles and nonparametric conditional quantiles. We now turn to parametric quantile regression models.

4.1 Identification

Suppose that in addition to Y we observe a vector of p covariates, x, including a constant. More precisely, we observe the joint random set $(x, Y) : \Omega \to \mathbb{R}^p \times \mathcal{K}(\mathbb{R})$, a selection of which is a random vector $(x, y^*) \in Sel(x, Y)$. Throughout, we assume the regular conditional probability measures of Y and y given x exist. We define a parametrized τ-th quantile regression function $q(\,\cdot\,, \theta(\tau))$ by

$$F_{y|x}(q(x, \theta(\tau)) \mid x) = \tau \qquad \mathbf{P}\text{-a.s.}$$

for all $\tau \in (0,1)$. A special case is the linear quantile regression function given by $q(x, \theta(\tau)) = x'\theta(\tau)$. We would like to identify $\theta(\tau)$ for each $\tau \in (0,1)$. The identification set is defined by

$$\Theta_0(\tau) = \left\{\theta(\tau) \in \Theta : F_{y|x}(q(x, \theta(\tau)) \mid x) = \tau \ \mathbf{P}\text{-a.s.}, (x, y) \in Sel(x, Y)\right\} \tag{4.1}$$

for each $\tau \in (0,1)$.

Theorem 4.1 *For every $\tau \in (0,1)$,*

$$\Theta_0(\tau) \subset \left\{ \theta(\tau) \in \Theta : \begin{array}{l} E\left[1[Y \subset (-\infty, q(x, \theta(\tau))]] - \tau \mid x\right] \leq 0 \\ E\left[\tau - 1[Y \cap (-\infty, q(x, \theta(\tau))] \neq \emptyset] \mid x\right] \leq 0 \end{array} \quad \mathbf{P} - a.s. \right\}. \tag{4.2}$$

Proof. As in the proof of Theorem 2.1, $y \in Sel(Y \mid x = \xi)$ implies $\tilde{C}_{Y|x}(t \mid \xi) \leq F_{y|x}(t \mid \xi) \leq \tilde{T}_{Y|x}(t \mid \xi)$ for all $t \in \mathbb{R}$. Therefore, $(x, y) \in Sel(x, Y)$ implies $\tilde{C}_{Y|x}(q(x, \theta(\tau)) \mid x) \leq F_{y|x}(q(x, \theta(\tau)) \mid x) \leq \tilde{T}_{Y|x}(q(x, \theta(\tau)) \mid x)$ $\mathbf{P}$-a.s. Thus,

$$\begin{aligned} \Theta_0(\tau) &= \left\{\theta(\tau) \in \Theta : F_{y|x}(q(x, \theta(\tau)) \mid x) = \tau \ \mathbf{P}\text{-a.s.}, (x, y) \in Sel(x, Y)\right\} \\ &\subset \left\{\theta(\tau) \in \Theta : \tilde{C}_{Y|x}(q(x, \theta(\tau)) \mid x) \leq \tau \leq \tilde{T}_{Y|x}(q(x, \theta(\tau)) \mid x) \ \mathbf{P} - a.s.\right\}. \end{aligned}$$

Writing the last expression in terms of conditional moments yields the expression in the statement of the theorem. ■

This theorem provides conditional moment inequality restrictions to characterize a super-set of the identification set $\Theta_0(\tau)$. Like Theorem 2.1 for unconditional quantiles, the other direction of set inclusion is not generally guaranteed. However, like Theorem 2.2 for unconditional quantiles, the following theorem provides an important case where the reverse inclusion is true.

Theorem 4.2 *If $Sel(x,Y) \neq \emptyset$ and Y is a convex valued random set in $\mathbb{R}$ (i.e. Y is interval valued), then*

$$\left\{\theta(\tau) \in \Theta : \begin{array}{l} E\left[1[Y \subset (-\infty, q(x,\theta(\tau))]] - \tau \mid x\right] < 0 \\ E\left[\tau - 1[Y \cap (-\infty, q(x,\theta(\tau))] \neq \emptyset] \mid x\right] < 0 \end{array} \quad \mathbf{P}-a.s. \right\} \subset \Theta_0(\tau)$$

for all $\tau \in (0,1)$. Furthermore, if in addition $\inf Y(\omega) > -\infty$ $\mathbf{P}$-a.s., then

$$\left\{\theta(\tau) \in \Theta : \ E\left[\tau - 1[Y \cap (-\infty, q(x,\theta(\tau))] \neq \emptyset] \mid x\right] = 0 \ \ \mathbf{P}-a.s.\right\} \subset \Theta_0(\tau)$$

for all $\tau \in (0,1)$. Similarly, if in addition $\sup Y(\omega) < \infty$ $\mathbf{P}$-a.s., then

$$\left\{\theta(\tau) \in \Theta : \ E\left[1[Y \subset (-\infty, q(x,\theta(\tau))]] - \tau \mid x\right] = 0 \ \ \mathbf{P}-a.s.\right\} \subset \Theta_0(\tau)$$

for all $\tau \in (0,1)$.

Proof. If $\left\{\theta(\tau) \in \Theta : \tilde{C}_{Y|x}(q(x,\theta(\tau)) \mid x) < \tau < \tilde{T}_{Y|x}(q(x,\theta(\tau)) \mid x) \ \mathbf{P}-a.s.\right\}$ is empty, then the first claim in the theorem is satisfied. Now, suppose that this set is non-empty. Take $\theta \in \left\{\theta(\tau) \in \Theta : \tilde{C}_{Y|x}(q(x,\theta(\tau)) \mid x) < \tau < \tilde{T}_{Y|x}(q(x,\theta(\tau)) \mid x) \ \mathbf{P}-a.s.\right\}$. Let $\Omega_L = \{\omega : \sup Y(\omega) < q(x(\omega),\theta)\}$, $\Omega_U = \{\omega : \inf Y(\omega) > q(x(\omega),\theta)\}$, and $\Omega_M = \Omega \backslash (\Omega_L \cup \Omega_U)$. Then, $\mathbf{P}(\Omega_L \mid x) < \tau$ and $\mathbf{P}(\Omega_U \mid x) < 1-\tau$, $\mathbf{P}$-a.s. By $Sel(x,Y) \neq \emptyset$, we choose $(x,y) \in Sel(x,Y)$. Let

$$\tilde{y}(\omega) = \begin{cases} y(\omega) & \omega \in \Omega_L \cup \Omega_U \\ q(x(\omega),\theta) & \omega \in \Omega_M \end{cases}$$

Since $Y(\omega)$ is convex, $q(x(\omega),\theta) \in Y(\omega)$ for all $\omega \in \Omega_M$, and thus the random variable $\tilde{y}$ defined above is a selection of Y. Therefore, $(x,\tilde{y}) \in Sel(x,Y)$. By construction, $F_{\tilde{y}|x}(q(x,\theta) \mid x) = \tau$, $\mathbf{P}$-a.s., and thus $\theta \in \Theta_0(\tau)$. This shows

$$\left\{\theta(\tau) \in \Theta : \tilde{C}_{Y|x}(q(x,\theta(\tau)) \mid x) < \tau < \tilde{T}_{Y|x}(q(x,\theta(\tau)) \mid x) \; \mathbf{P}-a.s.\right\} \subset \Theta_0(\tau).$$

Writing the expression on the left-hand side in terms of conditional moments yields the expression in the statement of the theorem.

Suppose that $\inf Y(\omega) > -\infty$ $\mathbf{P}$-a.s. Then, since $Y(\omega)$ is closed, the random variable $\tilde{y}$ defined by $\tilde{y}(\omega) = \inf Y(\omega)$ is a selection of Y. Therefore, $(x,\tilde{y}) \in Sel(x,Y)$. If we take $\theta \in \left\{\theta(\tau) \in \Theta : \tau = \tilde{T}_{Y|x}(q(x,\theta(\tau)) \mid x) \; \mathbf{P}-a.s.\right\}$, then $\tau = \tilde{T}_{Y|x}(q(x,\theta) \mid x) = \mathbf{P}(Y \cap (-\infty, q(x,\theta)] \neq \emptyset \mid x) = F_{\tilde{y}|x}(q(x,\theta) \mid x)$, $\mathbf{P}$-a.s., and thus $\theta \in \Theta_0(\tau)$. This shows

$$\left\{\theta(\tau) \in \Theta : \tilde{T}_{Y|x}(q(x,\theta(\tau)) \mid x) = \tau \; \mathbf{P}-a.s.\right\} \subset \Theta_0(\tau).$$

Writing the expression on the left-hand side in terms of conditional moments yields the expression in the statement of the theorem.

Suppose that $\sup Y(\omega) < \infty$ $\mathbf{P}$-a.s. Then, since $Y(\omega)$ is closed, the random variable $\tilde{y}$ defined by $\tilde{y}(\omega) = \sup Y(\omega)$ is a selection of Y. Therefore, $(x,\tilde{y}) \in Sel(x,Y)$. If we take $\theta \in \left\{\theta(\tau) \in \Theta : \tilde{C}_{Y|x}(q(x,\theta(\tau)) = \tau \mid x) \; \mathbf{P}-a.s.\right\}$, then $\tau = \tilde{C}_{Y|x}(q(x,\theta) \mid x) = \mathbf{P}(Y \subset (-\infty, q(x,\theta)] \mid x) = F_{\tilde{y}|x}(q(x,\theta) \mid x)$, $\mathbf{P}$-a.s., and thus $\theta \in \Theta_0(\tau)$. This shows

$$\left\{\theta(\tau) \in \Theta : \tilde{C}_{Y|x}(q(x,\theta(\tau)) \mid x) = \tau \; \mathbf{P}-a.s.\right\} \subset \Theta_0(\tau).$$

Writing the expression on the left-hand side in terms of conditional moments yields the expression in the statement of the theorem. ■

Theorems 4.1 and 4.2 together show that the conditional moment inequality restrictions in (4.2) provide a sharp characterization of the identification set $\Theta_0(\tau)$ for all $\tau \in (0,1)$ if (x,Y) has a selection and $Y(\omega)$ is a compact interval $\mathbf{P}$-a.s. One sufficient condition for the condition, $Sel(x,Y) \neq \emptyset$, of Theorem 4.2 is that Y is closed-valued and non-empty, as stated in the Fundamental Selection

Theorem (Molchanov, 2005, Theorem 2.13). The conditional moment inequalities can be rewritten more simply as

$$\Theta_0(\tau) = \left\{ \theta(\tau) \in \Theta : \begin{array}{c} E\left[1[y^u \leq q(x, \theta(\tau))] - \tau \mid x\right] \leq 0 \\ E\left[\tau - 1[y^l \leq q(x, \theta(\tau))] \mid x\right] \leq 0 \end{array} \quad \mathbf{P} - a.s. \right\} \tag{4.3}$$

where (x, y^l, y^u) is an $\mathbb{R}^{p+2}$-dimensional random vector generated by (x, Y) through the transformation $(x(\omega), Y(\omega)) \mapsto (x(\omega), \min Y(\omega), \max Y(\omega)) =: (x(\omega), y^l(\omega), y^u(\omega))$ for each $\omega \in \Omega$.

4.2 Convexity of the Identification Set

For the quantile set, the sharp identification set is guaranteed to be an interval (see Section 2). For the current setting where the sharp identification set is only implicitly characterized by a system of conditional moment inequalities, it is not clear if the identification set has nice geometric properties such as convexity. Suppose that the quantile regression is specified in the linear-in-parameters form $q(x, \theta(\tau)) = x'\theta(\tau)$. In this case, the identification set $\Theta_0(\tau)$ can be shown to be convex. Consequently, projections of the identification set $\Theta_0(\tau)$ on each coordinate is an interval.

Theorem 4.3 *Suppose that $q(x, \theta(\tau)) = x'\theta(\tau)$. If Y is a closed convex valued random set in $\mathbb{R}$, then the identification set $\Theta_0(\tau)$ is convex.*

Proof. Fix τ. By Theorems 4.1 and 4.2, the identification set $\Theta_0(\tau)$ is given by (4.3) under the given conditions. Let $\theta^1, \theta^2 \in \Theta_0(\tau)$ and $\lambda \in (0,1)$. Then, $q(x, \theta(\tau)) = x'\theta(\tau)$ implies $\{\omega \in x^{-1}(\{\xi\}) \subset \Omega : y^u(\omega) \leq q(x(\omega), \lambda\theta^1 + (1-\lambda)\theta^2)\} \subset \{\omega \in x^{-1}(\{\xi\}) \subset \Omega : y^u(\omega) \leq \max\{q(x(\omega), \theta^1), q(x(\omega), \theta^2)\}\}$ for every $\xi \in \mathbb{R}^p$, and thus

$$E\left[1[y^u \leq q(x, \lambda\theta^1 + (1-\lambda)\theta^2)] - \tau \mid x\right] \leq 0 \quad \mathbf{P} - a.s.$$

Also, $q(x, \theta(\tau)) = x'\theta(\tau)$ implies $\{\omega \in x^{-1}(\{\xi\}) \subset \Omega : y^l(\omega) \leq q(x(\omega), \lambda\theta^1 + (1-\lambda)\theta^2)\} \supset \{\omega \in x^{-1}(\{\xi\}) \subset \Omega : y^l(\omega) \leq \min\{q(x(\omega), \theta^1), q(x(\omega), \theta^2)\}\}$ for every $\xi \in \mathbb{R}^p$, and thus

$$E\left[\tau - 1[y^l \leq q(x, \lambda\theta^1 + (1-\lambda)\theta^2)] \mid x\right] \leq 0 \quad \mathbf{P} - a.s.$$

Therefore, $\lambda\theta^1 + (1-\lambda)\theta^2 \in \Theta_0(\tau)$, showing that $\Theta_0(\tau)$ is convex. ■

This geometric information is useful in practice. For example, it provides a guidance about the direction of computational search for a grid representation of set estimates.

4.3 Inference Based on Conditional Moment Inequalities

The conditional moment inequality restrictions in (4.3) to characterize the identification set $\Theta_0(\tau)$ can be rewritten as

$$E\left[m_j(w,\theta) \mid x\right] \geq 0 \quad \mathbf{P}-a.s. \quad \text{for } j = 1, 2,$$

where the moment functions, m_1 and m_2, are defined by

$$\begin{aligned} m_1(w,\theta) &= 1[y^l \leq q(x,\theta)] - \tau \qquad \text{and} \\ m_2(w,\theta) &= \tau - 1[y^u \leq q(x,\theta)] \end{aligned}$$

for $w = (x, y^l, y^u)$. This model fits in the framework for which the existing literature provides methods of inference via moment selection, e.g., Andrews and Shi (2013). For convenience of the readers, we describe the procedure of inference based on this existing literature in Appendix A.5.

5 Set of Best Linear Predictors

In this section, we focus on the linear quantile regression model, $q(x,\theta) = x'\theta$. We define the identification region $\Theta_0^*(\tau)$ for the best linear predictors (BLP) θ by extending Koenker and Bassett (1978) for the case of interval valued Y. In addition, we show in Theorem 5.1 that the identification region $\Theta_0^*(\tau)$ for the BLP model is a super set of the identification region $\Theta_0(\tau)$ defined in (4.3) for linear models. This result is important since finding the identification $\Theta_0(\tau)$ defined in (4.3) is challenging relatively to finding the identification region $\Theta_0^*(\tau)$ for the BLP model. Therefore, one can start by finding the identification region for the BLP model and use this super set as a starting region where we should look for the identification region of Section 4.

For a given selection $(x, y) \in Sel(x, Y)$, we can estimate the associated parameters by minimizing the risk

$$R_\tau(\theta; x, y) = E\left[\rho_\tau\left(y - x'\theta\right)\right] \tag{5.1}$$

where $\rho_\tau(u) = u \cdot (\tau - 1[u < 0])$. (See Koenker and Bassett (1978) and Koenker (2005)). We propose that the identification set $\Theta_0(\tau)$ be approximated by

$$\Theta_0^*(\tau) = \left\{\arg\min_{\theta \in \Theta} R_\tau(\theta; x, y) : (x, y) \in Sel(x, Y)\right\}.$$

In appendix A.6, we provide some useful geometric properties of this set of best linear predictors. More importantly, we show in the theorem below that it is useful to locate the identification set $\Theta_0(\tau)$.

Theorem 5.1 *If $q(x, \theta) = x^*\theta$, then $\Theta_0(\tau) \subset \Theta_0^*(\tau)$.*

Proof. Suppose that $\theta \in \Theta_0(\tau)$. In other words, $F_{y|x}(x'\theta \mid x) = \tau$ $\mathbf{P}$-a.s. for a selection $(x, y) \in Sel(x, Y)$. Taking the gradient of $R_\tau(\theta; x, y)$ with respect to θ, we have

$$\begin{aligned} \nabla_\theta \int_{\mathbb{R}^p} \left[(\tau - 1)\int_{-\infty}^{\xi'\theta} (\zeta - \xi'\theta) dF_{y|x}(\zeta \mid \xi) + \tau \int_{\xi'\theta}^{\infty} (\zeta - \xi'\theta) dF_{y|x}(\zeta \mid \xi)\right] dF_x(\xi) \\ = \int_{\mathbb{R}^p} x\left[F_{y|x}(\xi'\theta \mid \xi) - \tau\right] dF_x(\xi) \ = \ 0 \end{aligned}$$

where the last equality follows from our choice of θ satisfying $F_{y|x}(x'\theta \mid x) = \tau$ $\mathbf{P}$-a.s. Therefore, we obtain $\nabla_\theta R(\theta; x, y) = 0$. Since $R_\tau(\,\cdot\,; x, y)$ is convex, this implies $\theta \in \Theta_0^*(\tau)$. ■

The other direction of set inclusion does not hold. While the identification set $\Theta_0(\tau)$ contains only those parameter vectors θ that correctly specify the quantile regression $q(x, \theta) = x'\theta$ for some $(x, y) \in Sel(x, Y)$, the approximate set $\Theta_0^*(\tau)$ contains many other parameter vectors θ which only allow $q(x, \theta) = x'\theta$ to be a best linear predictor for some $(x, y) \in Sel(x, Y)$. By Theorems 5.1, the approximation set $\Theta_0^*(\tau)$ does not miss any element of the identification set $\Theta_0(\tau)$. Therefore, we propose to first compute this set $\Theta_0^*(\tau)$ of best linear predictors, and conduct the test of conditional moment inequalities on and around this set. If the identification set $\Theta_0(\tau)$ is empty, then the parametric quantile regression model is misspecified, and $\Theta_0^*(\tau)$ trivially contains $\Theta_0(\tau)$. In this case of misspecification, the set $\Theta_0^*(\tau)$ of best linear predictors itself may be of use for best linear prediction and for causal inference – see Angrist, Chernozhukov, and Fernández-Val (2006) and Kato and Sasaki (2017). The remainder of this section is devoted to a computational algorithm to obtain the approximation set $\Theta_0^*(\tau)$.

5.1 Linear Programming

Since a random set can be viewed as a collection of regular random variables, we start by reviewing the regular random variable case. It seems logical that what we define for random sets should yield a regular quantile regression for singleton random sets.

The canonical LP problem is written as the following constrained minimization problem,

$$\begin{aligned} \min \quad & c'\theta \\ s.t \quad & A\theta = b \\ & \theta \geq 0, \end{aligned} \tag{5.2}$$

where A is a $m \times k$ matrix, c is a $k \times 1$ vector of coefficients, b is a $m \times 1$ vector of right-hand side constraints, and θ is a $k \times 1$ vector of unknowns. It is assumed that $m < k$.

Consider a finite random sample of n observations. Let y be the $n \times 1$ vector of outcomes, X be the $n \times p$ matrix of covariates (which includes a column of ones). We can solve the least absolute deviation problem corresponding to the τ-quantile regression by using the following linear programming problem,

$$\begin{aligned} & \min_{(\beta,u,v)\in\mathbb{R}^p\times\mathbb{R}^{2n}_+} \sum_{i=1}^{n} \tau u_i + (1-\tau)\, v_i \\ & s.t. \quad X\beta + u - v = y. \end{aligned} \tag{5.3}$$

The vector $(\beta_j)_{j=1}^p$ consists of the coefficients of the linear τ-quantile regression, while u and v are slack parameters (variables). The LP problem in (5.3) is labeled $LP\,(\tau, X, y)$. The simplex algorithm provides a solution to the above problem. The first stage is to transform the linear programming problem in (5.3) into the canonical form. Note that (5.2) requires that all variables over which we minimize be positive while the coefficients β in (5.3) are unrestricted. The first step requires the user to transform the problem into the form in (5.2). So at first, we can write (5.3) as

$$\begin{aligned} \min \quad & c'x \\ s.t. \quad & Ax = b(y) \\ & x \in S \end{aligned} \tag{5.4}$$

where $c = (0_p; \tau 1_n; (1-\tau)\, 1_n)$, $x = (\beta, u, v)$, $A = [X : I_{n\times n}, -I_{n\times n}]$, $b(y) = y$, and $S = \mathbb{R}^p \times \mathbb{R}_+^{2n}$. The first p coordinates of x are unrestricted while the last $2n$ coordinates are restricted to be non-negative. Some software packages handle this kind of almost canonical form (e.g. Matlab) but some more traditional code may not. Assume w.l.o.g. that the first p rows of the matrix X form a $p \times p$ full rank (and thus invertible) matrix. Denote this matrix by X_p and similarly denote by u_p, v_p and y_p the first p lines of the the corresponding column vectors u, v and y. Denote by X_{-p}, u_{-p}, v_{-p} and y_{-p} the remaining $n-p$ rows of these matrices and vectors. The first p equations in $Ax = b$ as well as the unconstrained variable β can be eliminated by writing $\beta = X_p^{-1}(y_p - u_p - v_p)$ and substituting β into these p first equations in $Ax = b$. The remaining $n - p$ equations then can be written as

$$\left(X_{-p}X_p^{-1}\right) u_p - \left(X_{-p}X_p^{-1}\right) v_p + u_{-p} - v_{-p} = y_{-p} - \left(X_{-p}X_p^{-1}\right) y_p.$$

Therefore, the problem in (5.4) can be written as

$$\begin{aligned} \min \quad & c'\tilde{x} \\ s.t. \quad & \tilde{A}\tilde{x} = \tilde{b}(y) \\ & \tilde{x} \geq 0 \end{aligned} \tag{5.5}$$

where $c = (\tau 1_n, (1-\tau)\, 1_n)$, $\tilde{x} = (u, v)$, $\tilde{A} = \left[X_{-p}X_p^{-1} : I_{n-p\times n-p} : X_{-p}X_p^{-1} : I_{n-p\times n-p}\right]$, and $\tilde{b}(y) = y_{-p} - \left(X_{-p}X_p^{-1}\right) y_p$. Notice that the LP problem in (5.3) has $p + 2n$ variables and n

equality constraints and the LP problem in (5.5) has $2n$ variables and $n-p$ equality constraints. Out of the solution for u and v in (5.5) we can of course recover the vector of interest β by using $\beta = X_p^{-1}(y_p - u_p - v_p)$.

For any ordered set B, let $B(i)$ denote the i-th element of B. For any ordered set $B \subset \{1, \cdots, 2n\}$ of cardinality $|B|$ and for any $|B|$-dimensional vector ξ, we define the $2n$-dimensional vector $\Pi_B \xi$ whose j-th coordinate is given by

$$(\Pi_B \xi)_j = \begin{cases} \xi_i & \text{if there exists } i \in \{1, \cdots, |B|\} \text{ such that } B(i) = j \\ 0 & \text{otherwise} \end{cases}$$

for each $i \in \{1, \cdots, 2n\}$. The simplex algorithm yields a solution $\tilde{x}(y)$ with an ordered set $B_y \subset \{1, \cdots, 2n\}$ of basic indices. Also let $-B_y = \{1, \cdots, 2n\} \backslash B_y$ denote an ordered set of non-basic indices. It is required that $|B_y| = n-p$, $|-B_y| = n+p$, $\tilde{x}(y)_{B_y} \geqslant 0_{n-p}$, and $\tilde{x}(y)_{-B_y} = 0_{n+p}$. Using these notations, the solution is explicitly written as the $2n$-dimensional vector $\tilde{x}(y) = \Pi_{B_y} \tilde{A}_{B_y}^{-1} \tilde{b}(y)$. Optimality requires $c_{-B_y} - c_{B_y} \tilde{A}_{B_y}^{-1} \tilde{A}_{-B_y} \geqslant 0_{n+p}$ and feasibility requires that $\tilde{x}(y) = \Pi_{B_y} \tilde{A}_{B_y}^{-1} \tilde{b}(y) \geqslant 0_{2n}$. The simplex algorithm prescribes an efficient computational procedure to find such an index set B_y satisfying these requirements.

5.2 Set Linear Programming

A simple brute force approach to set estimation of β is to obtain the solution $\tilde{x}(y) = (u(y), v(y))$ to the optimization problem (5.5) for each $y \in \times_{i=1}^n [y_{Li}, y_{Ui}]$, and to take the union $\cup_{y \in \times_{i=1}^n [y_{Li}, y_{Ui}]} \beta(u)$ where $\beta(y) = X_p^{-1}(y_p - u_p(y) - v_p(y))$. However, this exhaustive approach (even with a lattice approximation) is computationally intensive. In this light, we use some convenient properties of linear programming to propose a faster algorithm to compute the set estimate for β.

Pick $y^1 \in \times_{i=1}^n [y_{Li}, y_{Ui}]$. The simplex algorithm yields a solution $\tilde{x}(y^1)$ with an ordered set $B_{y^1} \subset \{1, \cdots, 2n\}$ of basic indices. Also let $-B_{y^1} = \{1, \cdots, 2n\} \backslash B_{y^1}$ denote an ordered set of non-basic indices. The solution is explicitly written as the $2n$-dimensional vector $\tilde{x}(y^1) =$

$\Pi_{B_{y^1}} \tilde{A}_{B_{y^1}}^{-1} \tilde{b}(y^1)$. The next proposition shows that there is a set $\mathcal{Y}^1 \in \times_{i=1}^{n} [y_{Li}, y_{Ui}]$ containing y^1 such that the set B_{y^1} of basic indices for (5.5) remains unchanged for all $y \in \mathcal{Y}^1$. Therefore, once we solve (5.5) for $y = y^1$, we do not need to solve (5.5) again for any other $y \in \mathcal{Y}^1$, and we thus tremendously save our computational resources.

Proposition 5.2 *A solution to (5.5) is given by* $\tilde{x}(y) = \Pi_{B_{y^1}} \tilde{A}_{B_{y^1}}^{-1} \tilde{b}(y)$ *for all* $y \in \mathcal{Y}^1$ *where* $\mathcal{Y}^1 = \left\{ y \in \times_{i=1}^{n} [y_{Li}, y_{Ui}] \,\middle|\, \tilde{A}_{B_{y^1}}^{-1} \tilde{b}(y) \geqslant 0_{n-p} \right\}$. *In particular,* $y^1 \in \mathcal{Y}^1$.

Proof. We have $c_{-B_{y^1}} - c_{B_{y^1}} \tilde{A}_{B_{y^1}}^{-1} \tilde{A}_{-B_{y^1}} \geqslant 0_{n+p}$ by the definition of B_{y^1} and $-B_{y^1}$ as the sets of basic and non-basic indices, respectively, at the solution to (5.5) with $y = y^1$. Notice that c does not depend on y in (5.5). Hence, any feasible vertex $\tilde{x} \in \mathbb{R}_+^{2n}$ of the constraint set having non-zero elements only for those indices in $B(y^1)$ is optimal for (5.5). Consider (5.5) with $y \in \times_{i=1}^{n} [y_{Li}, y_{Ui}]$. A vertex $\tilde{x}$ having non-zero elements only for those indices in $B(y^1)$ is written as $\tilde{x} = \Pi_{B_{y^1}} \tilde{A}_{B_{y^1}}^{-1} b(y)$. It is feasible if $\tilde{x} = \Pi_{B_{y^1}} \tilde{A}_{B_{y^1}}^{-1} b(y) \geqslant 0_{2n}$, which is true if and only if $\tilde{A}_{B_{y^1}}^{-1} b(y) \geqslant 0_{n-p}$. Therefore, a solution to (5.5) is given by $\tilde{x}(y) = \Pi_{B_{y^1}} \tilde{A}_{B_{y^1}}^{-1} \tilde{b}(y)$ for all $y \in \times_{i=1}^{n} [y_{Li}, y_{Ui}]$ such that $\tilde{A}_{B_{y^1}}^{-1} \tilde{b}(y) \geqslant 0_{n-p}$. ■

In light of this proposition, we propose the following procedure. For any $y \in \mathcal{Y}^1$, let

$$
\begin{aligned}
u_p(y) &= \left(\left(\Pi_{B_{y^1}} \tilde{A}_{B_{y^1}}^{-1} \tilde{b}(y) \right)_1, \cdots, \left(\Pi_{B_{y^1}} \tilde{A}_{B_{y^1}}^{-1} \tilde{b}(y) \right)_p \right)' \\
v_p(y) &= \left(\left(\Pi_{B_{y^1}} \tilde{A}_{B_{y^1}}^{-1} \tilde{b}(y) \right)_{n+1}, \cdots, \left(\Pi_{B_{y^1}} \tilde{A}_{B_{y^1}}^{-1} \tilde{b}(y) \right)_{n+p} \right)'
\end{aligned}
$$

be two p-dimensional subvectors of the solution $\Pi_{B_{y^1}} \tilde{A}_{B_{y^1}}^{-1} \tilde{b}(y)$. We can then directly compute the estimate of β corresponding to this $y \in \mathcal{Y}^1$ by $\beta(y) = X_p^{-1} (y_p - u_p(y) - v_p(y))$. Thus, we construct the subset estimate

$$
\hat{\Theta}_0^*(\tau; y^1) = \left\{ X_p^{-1} \left(y_p - \begin{pmatrix} \left(\Pi_{B_{y^1}} \tilde{A}_{B_{y^1}}^{-1} \tilde{b}(y) \right)_1 \\ \vdots \\ \left(\Pi_{B_{y^1}} \tilde{A}_{B_{y^1}}^{-1} \tilde{b}(y) \right)_p \end{pmatrix} - \begin{pmatrix} \left(\Pi_{B_{y^1}} \tilde{A}_{B_{y^1}}^{-1} \tilde{b}(y) \right)_{n+1} \\ \vdots \\ \left(\Pi_{B_{y^1}} \tilde{A}_{B_{y^1}}^{-1} \tilde{b}(y) \right)_{n+p} \end{pmatrix} \right) : y \in \mathcal{Y}^1 \right\}.
$$

Since this subset estimate is an image of $\mathcal{Y}^1$ through a simple linear transformation, it conveniently circumvents the need to solve the optimization problem for each $y \in \mathcal{Y}^1$. Once this subset estimate has been computed, pick $y^2 \in \times_{i=1}^n [y_{Li}, y_{Ui}] \backslash \mathcal{Y}^1$, use the simplex algorithm to get the set $B(y^2)$ of basic indices under y^2, and obtain the resultant subset estimate $\hat{\Theta}_0^*(\tau; y^2)$. This is followed by the third step where $y^3 \in \times_{i=1}^n [y_{Li}, y_{Ui}] \backslash \left(\cup_{k=1}^2 \mathcal{Y}^k \right)$ produces the subset estimate $\hat{\Theta}_0^*(\tau; y^3)$, and so on. Repeat this process to obtain the set estimate $\hat{\Theta}_0^*(\tau) = \cup_{k=1}^K \hat{\Theta}_0^*(\tau; y^k)$ for K steps until we exhaust $\times_{i=1}^n [y_{Li}, y_{Ui}] = \prod_{k=1}^K \mathcal{Y}^k$.

6 Monte Carlo Experiments

This section evaluates the properties of the estimators and tests proposed in the previous sections via a series of Monte Carlo experiments. Section 3 introduced estimation and testing methods for the continuous case (Section 3.1), the discrete case (Section 3.2), and the conditional quantiles (Section 3.3). In addition, Section 4 discussed parametric quantile regression models. Reflecting this structure of our theoretical results, we consider the four corresponding cases for Monte Carlo studies organized into the four subsections below.

6.1 Continuous Case

To generate interval-valued sets with continuously distributed upper and lower bounds, we use the following model.

$$
\begin{aligned}
Y_i &= [a_i, b_i], \qquad \text{where} \\
a_i &= 0.5 \cdot v_i + 1.5 \cdot w_i \qquad b_i = 2.5 \cdot v_i + 1.5 \cdot w_i \\
v_i &\sim U(0,1) \qquad\qquad w_i \sim U(0,1).
\end{aligned}
\tag{6.1}
$$

This data generating model assures that $a_i < b_i$, and so $Y_i \neq \emptyset$. Furthermore, the upper and lower bounds are correlated by construction. Table 1 summarizes the identification sets for various quantiles under this data generating model. The second and third columns list the lower and upper

Table 1: Identification sets under the continuous model (6.1).

τ	$q_a(\tau)$	$q_b(\tau)$	$m(\Theta_0^Y(\tau))$
0.20	0.550	1.225	0.675
0.30	0.700	1.500	0.800
0.40	0.850	1.750	0.900
0.50	1.000	2.000	1.000
0.60	1.150	2.250	1.100
0.70	1.300	2.500	1.200
0.80	1.450	2.775	1.325

bounds of the sharp identification sets $\Theta_0^Y(\tau) = [q_a(\tau), q_b(\tau)]$. The fourth column lists the widths $m(\Theta_0^Y(\tau))$ of the identification sets.

For Monte Carlo experiments, 25,000 small samples are drawn of sizes 250, 500, 1,000, and 2,000 observations. We use the test statistic based on the Hausdorff distance to test $H_0 : \Theta_0^Y(\tau) = \Theta$ versus $H_1 : \Theta_0^Y(\tau) \neq \Theta$ for various Θ. For each quantile and for each sample, the critical value is estimated using the asymptotic distribution (3.7) for the test with size $\alpha = 0.05$. For the variance matrix $\Sigma(\tau)$ in (3.2), the joint CDF $F_{a,b}(q_a(\tau), q_b(\tau))$ is estimated by the joint empirical CDF, and marginal densities $f_a(q_a(\tau))$ and $f_b(q_b(\tau))$ are estimated with the Gaussian kernel using Silverman's rule of bandwidth choice. Critical values are estimated using 25,000 simulated statistics generated by this asymptotic normal distribution. The local alternatives, against which the null is tested, use $\Theta = \Theta_0^Y(\tau) + \delta \cdot \frac{m(\Theta_0^Y(\tau))}{\sqrt{n}}$ for $\delta \in \{0.0, 0.5, 1.0, 2.0, 4.0, 8.0\}$.

Table 2 reports the rejection frequencies of the test at each of the quantiles $\tau \in \{0.25, 0.50, 0.75\}$. Observe that the rejection rates are approximately 0.05 for the null $\delta = 0$, which evidences that the size is correct. Also observe that the rejection rates increase as δ increases at approximately a uniform rate across sample sizes, which evidences the root-n rate of convergence for $H(\hat{\Theta}_0^Y(\tau), \Theta_0^Y(\tau))$. These results support the theoretical properties developed in Section 3.1. Although we report the

Table 2: Rejection frequencies of the test for $\tau \in \{0.25, 0.50, 0.75\}$ based on the Hausdorff distace against local alternatives under the continuous model (6.1).

		δ					
τ	n	0.0	0.5	1.0	2.0	4.0	8.0
0.25	250	0.043	0.047	0.074	0.228	0.870	1.000
0.25	500	0.043	0.050	0.082	0.236	0.881	1.000
0.25	1,000	0.043	0.052	0.085	0.250	0.881	1.000
0.25	2,000	0.046	0.056	0.087	0.260	0.895	1.000
0.50	250	0.052	0.070	0.141	0.471	0.990	1.000
0.50	500	0.052	0.074	0.136	0.468	0.990	1.000
0.50	1,000	0.051	0.073	0.142	0.466	0.991	1.000
0.50	2,000	0.051	0.072	0.142	0.471	0.990	1.000
0.75	250	0.043	0.083	0.205	0.709	1.000	1.000
0.75	500	0.044	0.083	0.203	0.719	1.000	1.000
0.75	1,000	0.044	0.081	0.202	0.728	1.000	1.000
0.75	2,000	0.046	0.080	0.203	0.730	1.000	1.000

case for $\tau \in \{0.25, 0.50, 0.75\}$ here, similar results are obtained for the other quantiles.[3]

6.2 Discrete Case

To generate interval-valued sets of discrete categories, we use the following model.

$$(6.2) \qquad \begin{aligned} Y_i &= \sum_{t=-\infty}^{\infty} 1[y_i \in (t-0.5, t+0.5]] \cdot [t-0.5, t+0.5], \qquad \text{where} \\ y_i &\sim N(0.0, 10.0). \end{aligned}$$

[3]The code is available from the authors upon request.

Table 3: Identification sets under the discrete model (6.2).

τ	$q_a(\tau)$	$q_b(\tau)$	$m(\Theta_0^Y(\tau))$
0.20	-3.500	-2.500	1.000
0.30	-2.500	-1.500	1.000
0.40	-1.500	-0.500	1.000
0.50	-0.500	0.500	1.000
0.60	0.500	1.500	1.000
0.70	1.500	2.500	1.000
0.80	2.500	3.500	1.000

For example, if $y_i = 2.485$ is drawn from $N(0.0, 10.0)$, then the obtained set is $Y_i = [1.5, 2.5]$. Table 3 summarizes the identification sets for various quantiles under this data generating model. The second and third columns list the lower and upper bounds of the sharp identification sets $\Theta_0^Y(\tau) = [q_a(\tau), q_b(\tau)]$. The fourth column lists the widths $m(\Theta_0^Y(\tau))$ of the identification sets.

Recall from Section 3.2 that the discrete case allows for super-consistent set estimation in the sense that $r_n H(\hat{\Theta}_0^Y(\tau), \Theta_0^Y(\tau)) \xrightarrow{P} 0$ at any rate $r_n \to \infty$ as $n \to \infty$. We demonstrate this property by simulating the frequencies of the event $H(\hat{\Theta}_0^Y(\tau), \Theta_0^Y(\tau)) > 0$ across various sample sizes n. 25,000 small samples are drawn of sizes 250, 500, 1,000, and 2,000 observations. Table 4 reports the frequencies of the event $H(\hat{\Theta}_0^Y(\tau), \Theta_0^Y(\tau)) > 0$ at each of the quantiles $\tau \in \{0.25, 0.50, 0.75\}$. Consistently with the theory, the last column indicates that $\text{Freq}\left(r_{2,000} H(\hat{\Theta}_0^Y(\tau), \Theta_0^Y(\tau)) > 1\right) = 0.000$ for any choice of $r_{2,000} > 0$.

Table 4: Frequencies of the event $H(\hat{\Theta}_0^Y(\tau), \Theta_0^Y(\tau)) > 0$ under the discrete model (6.2).

	n			
τ	250	500	1,000	2,000
0.25	0.101	0.033	0.004	0.000
0.50	0.055	0.006	0.000	0.000
0.75	0.101	0.033	0.004	0.000

6.3 Conditional quantiles

To generate interval-valued sets with continuously distributed upper and lower bounds which vary with an observed covariate x, we use the following model.

$$
\begin{aligned}
Y_i &= [a_i, b_i], \qquad \text{where} \\
a_i &= (0.5 + x_i) \cdot v_i + 1.5 \cdot w_i \qquad b_i = (2.5 + x_i) \cdot v_i + 1.5 \cdot w_i \\
v_i &\sim U(0,1) \qquad w_i \sim U(0,1) \qquad x_i \sim N(0,1).
\end{aligned} \tag{6.3}
$$

This data generating model assures that $a_i < b_i$, and so $Y_i \neq \emptyset$. Furthermore, the upper and lower bounds are mutually correlated, as well as depend on the observed covariate x. In the locality of $x^* = 0.0$, the conditional distributions of a and b are the same as those unconditional quantiles under the data generating model (6.1). Both the upper and lower bounds tend to increase in a linear fashion as x increases. However, the widths of the sets are not designed to vary linearly in x to ensure $a_i < b_i$.

25,000 small samples are drawn of sizes 1,000, 2,000, and 4,000 observations. We use the test statistic based on the Hausdorff distance to test $H_0 : \Theta_0^Y(\tau \mid x^*) = \Theta$ versus $H_1 : \Theta_0^Y(\tau \mid x^*) \neq \Theta$ for various Θ. For each locality of x, for each quantile, and for each sample, the critical value is estimated using the asymptotic distribution (3.16) for the test with size $\alpha = 0.05$. For the variance matrix $\Sigma(\tau \mid x^*)$ in (3.15), the conditional joint CDF $F_{a,b|x}(q_a(\tau), q_b(\tau) \mid x^*)$ is estimated by the empirical CDF in the $h_n(\tau \mid x^*)$-neighborhood of x^*, and conditional densities $f_{a|x}(q_a(\tau) \mid x^*)$

and $f_{b|x}(q_b(\tau) \mid x^*)$ are estimated with the Gaussian kernel in the $h_n(\tau \mid x^*)$-neighborhood of x^*. Following Lemma 3.1 with the Lipschitz continuity of the conditional quantile functions $q_{a|x}(\tau \mid \cdot\,)$ and $q_{b|x}(\tau \mid \cdot\,)$, we set the bandwidth $h_n(\tau \mid x^*)$ to $\kappa(\tau \mid x^*) \cdot n^{-1/(2+p)}$ for a constant $\kappa(\tau \mid x^*)$ that does not vary with the sample size n. For the asymptotic theory, $\kappa(\tau \mid x^*)$ may be any constant, and so we set $\kappa(\tau \mid x^*) = f_x(x^*)^{-1}$ for the Monte Carlo study. Critical values are estimated using 25,000 simulated statistics generated by this asymptotic normal distribution. The local alternatives against which the null is tested use $\Theta = \Theta_{Y|x}(\tau \mid x^*) + \delta \cdot \frac{m(\Theta_{Y|x}(\tau|x^*))}{\sqrt{n}}$ for $\delta \in \{0.0, 0.5, 1.0, 2.0, 4.0, 8.0, 16.0\}$.

Table 5 reports the rejection frequencies of the test at each of the localities of $x^* \in \{-1.0, 0.0, 1.0\}$ and for each of the quantiles $\tau \in \{0.25, 0.50, 0.75\}$. Observe that the rejection rates are approximately 0.05 for the null $\delta = 0$, which evidences that the size is correct. Also observe that the rejection rates increase as δ increases at approximately a uniform rate across sample sizes, which evidences the root-$N_n(\tau \mid x^*)$ rate of convergence for $H(\hat{\Theta}_{Y|x}(\tau \mid x^*), \Theta_{Y|x}(\tau \mid x^*))$. These results support the theoretical properties developed in Section 6.3. Compared to the results for the model (6.1) for unconditional outcomes, not surprisingly, the local power is weaker in the current results for the mode (6.3) for conditional outcomes. We also conducted Monte Carlo simulation studies for models with multi-dimensional covariates x, and obtained similar results.[4]

6.4 Parametric Quantile Regression Models

To generate theoretically tractable parametric quantile regressions, we use the following model.

$$\begin{aligned} Y_i &= \sum_{t=-\infty}^{\infty} 1[y_i \in (t-0.1, t]] \cdot [t-0.1, t], \qquad \text{where} \\ y_i &= 1.0 + (1.0 + x) \cdot \varepsilon_i, \\ x_i &\sim U(0,1), \qquad \varepsilon_i \sim U(0,1), \qquad \text{and} \qquad x_i \perp\!\!\!\perp \varepsilon_i. \end{aligned} \tag{6.4}$$

[4] The code is available upon request from the authors.

Table 5: Rejection frequencies of the test for $x^* \in \{-1.0, 0.0, 1.0\}$ and $\tau \in \{0.25, 0.50, 0.75\}$ based on the Hausdorff distace against local alternatives under model (6.3) with a covariate.

			δ						
x^*	τ	n	0.0	0.5	1.0	2.0	4.0	8.0	16.0
-1.0	0.25	1,000	0.059	0.051	0.057	0.117	0.463	0.992	1.000
-1.0	0.25	2,000	0.054	0.051	0.060	0.113	0.410	0.977	1.000
-1.0	0.25	4,000	0.057	0.053	0.059	0.099	0.331	0.935	1.000
-1.0	0.50	1,000	0.059	0.042	0.044	0.102	0.474	0.993	1.000
-1.0	0.50	2,000	0.048	0.041	0.047	0.109	0.455	0.985	1.000
-1.0	0.50	4,000	0.051	0.042	0.043	0.084	0.342	0.954	1.000
-1.0	0.75	1,000	0.090	0.060	0.049	0.084	0.418	0.994	1.000
-1.0	0.75	2,000	0.066	0.053	0.053	0.109	0.440	0.991	1.000
-1.0	0.75	4,000	0.059	0.049	0.052	0.094	0.360	0.970	1.000
0.0	0.25	1,000	0.041	0.044	0.049	0.077	0.210	0.791	1.000
0.0	0.25	2,000	0.042	0.043	0.047	0.071	0.168	0.660	1.000
0.0	0.25	4,000	0.043	0.044	0.050	0.071	0.150	0.551	0.999
0.0	0.50	1,000	0.053	0.067	0.087	0.158	0.453	0.976	1.000
0.0	0.50	2,000	0.053	0.061	0.074	0.124	0.344	0.923	1.000
0.0	0.50	4,000	0.052	0.057	0.068	0.109	0.279	0.846	1.000
0.0	0.75	1,000	0.046	0.058	0.084	0.190	0.625	1.000	1.000
0.0	0.75	2,000	0.044	0.054	0.072	0.146	0.491	0.996	1.000
0.0	0.75	4,000	0.046	0.052	0.067	0.127	0.410	0.981	1.000
1.0	0.25	1,000	0.039	0.042	0.046	0.059	0.108	0.346	0.938
1.0	0.25	2,000	0.042	0.046	0.051	0.066	0.114	0.314	0.880
1.0	0.25	4,000	0.041	0.042	0.044	0.053	0.085	0.217	0.740
1.0	0.50	1,000	0.074	0.085	0.101	0.154	0.319	0.865	1.000
1.0	0.50	2,000	0.085	0.099	0.119	0.167	0.309	0.779	1.000
1.0	0.50	4,000	0.051	0.054	0.058	0.073	0.145	0.501	1.000
1.0	0.75	1,000	0.078	0.107	0.148	0.268	0.641	0.997	1.000
1.0	0.75	2,000	0.071	0.094	0.129	0.227	0.544	0.987	1.000
1.0	0.75	4,000	0.046	0.058	0.075	0.131	0.343	0.919	1.000

For example, if $x_i = 0.385$ and $\varepsilon_i = 0.573$, then $y_i \approx 1.794$ and $Y_i = [1.7, 1.8]$. This individual i is at the ($\tau = \varepsilon =$) 0.573-th quantile, and the corresponding quantile regression function is given by $q(x, \theta(0.573)) = 1.573 + 0.573x$. In fact, this model (6.4) yields the linear model for each quantile $\tau \in (0,1)$ with intercept $\theta_1(\tau) = 1 + \tau$ and slope $\theta_2(\tau) = \tau$. Hence, the true quantile regression parameter vector is $\theta(\tau) = (1 + \tau, \tau)$.

For Monte Carlo experiments, 1,000 small samples are drawn of sizes 100 and 200 observations. Based on the methods presented in Section 4, we conduct tests of $H_0 : \theta(\tau) = \theta$ versus $H_1 : \theta(\tau) \neq \theta$ for various parameter vectors θ for each quantile $\tau \in \{0.25, 0.50, 0.75\}$. The critical value is estimated using the procedure outlined in Appendix A.5 with 1,000 bootstrap replications for the size $\alpha = 0.05$.[5]

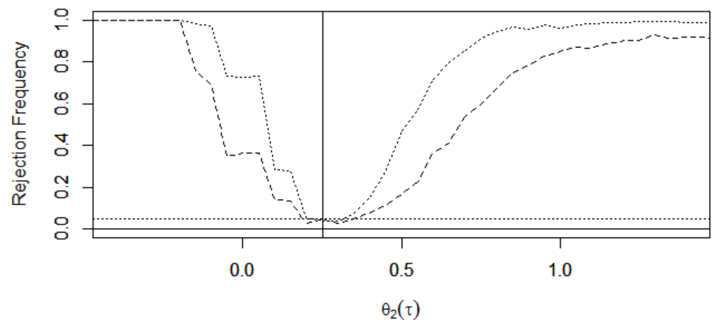
Figure 1 shows graphs of rejection frequencies. Parts (I), (II), and (III) of the figure focus on quantiles $\tau = 0.25$, 0.50, and 0.75, respectively. For ease of presentation, we focus on the one-dimensional slice of the two-dimensional parameter space running through the true parameter point at each quantile τ. Specifically, for the τ-th quantile, the results are displayed along the line segment $\{1+\tau\} \times [-0.5, 1.5]$ that passes through the true parameter point $(1+\tau, \tau)$. The true point is indicated in the figure by a solid vertical line. The rejection frequencies are a indicated by dashed and dotted curves for sample sizes $n = 100$ and 200, respectively. As expected, the frequencies at the true points fall below the designed size $\alpha = 0.05$. In fact, the frequencies fall below 0.05 not only at the true parameter point, but also in an interval containing the true parameter point. This is consistent with the fact that the quantile regression parameters are not point identified.

7 Summary

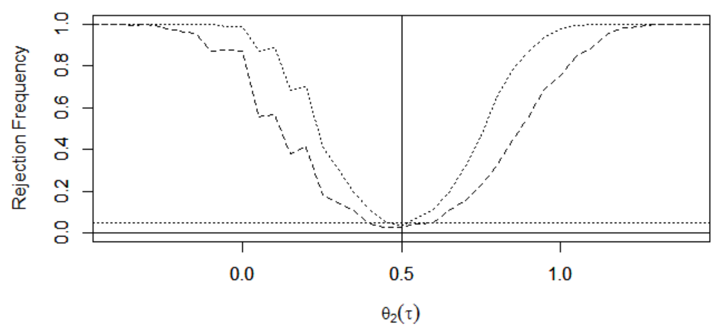
This paper investigates the identification of quantiles and of quantile regressions when the outcome variable is partially observed. We first identify and estimate the quantiles of a random set. The

[5]We used $R = 2$, $\epsilon_n = 0.05$, $\kappa_n = (0.3 \ln(n))^{\frac{1}{2}}$ and $B_n = (0.4 \ln(n) / \ln\ln(n))^{\frac{1}{2}}$.

(I) $\tau = 0.25$



(II) $\tau = 0.50$



(III) $\tau = 0.75$

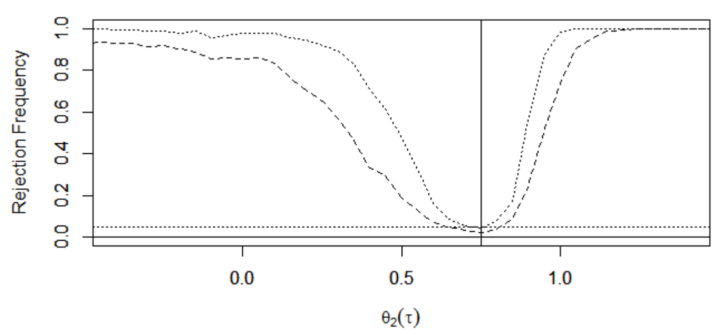


Figure 1: Rejection frequencies for inference of the parametric regression model (6.4). The horizontal axis measures $\theta_2(\tau)$ given $\theta_1(\tau)$ fixed at (I) 0.25, (II) 0.50, and (III) 0.75. The dashed and dotted curves indicate the sample sizes of $n = 100$ and 200, respectively.

quantile set is shown to be identified by the containment and capacity quantiles.

The identification argument for quantiles is extended to set identification of quantile regression parameters. We show that a system of conditional moment inequalities, involving the cumulative containment and cumulative capacity functionals, characterize the identification set for quantile regression parameters. A feasible computational algorithm is proposed to produce the set of infinite LP solutions by just solving a finite number of LP problems.

A series of Monte Carlo experiments are conducted for continuous cases, discrete cases, conditional quantiles, and quantile regression parameters. The results support our theoretical properties.

A Appendix

A.1 Estimation of Cumulative Containment and Capacity Functionals

Assume that we observe $Y_1, Y_2, ...$ i.i.d. $\mathbb{R}$-valued random closed sets. For a natural number n, define

$$\begin{aligned} \tilde{C}_n(t) &= \frac{1}{n}\sum_{i=1}^{n} 1\left[Y_i \subset (-\infty, t]\right] \qquad \text{and} \\ \tilde{T}_n(t) &= \frac{1}{n}\sum_{i=1}^{n} 1\left[Y_i \cap (-\infty, t] \neq \emptyset\right] \end{aligned}$$

to be the sample counterparts of the cumulative containment and capacity functionals. The following proposition shows the Glivenko-Cantelli property for these estimators.

Proposition A.1 *Under the above conditions*

$$\begin{aligned} &\sup_{t\in\mathbb{R}} \left|\tilde{T}_n(t) - \tilde{T}_Y(t)\right| \overset{a.s}{\to} 0 \qquad \textit{and} \\ &\sup_{t\in\mathbb{R}} \left|\tilde{C}_n(t) - \tilde{C}_Y(t)\right| \overset{a.s}{\to} 0 \end{aligned}$$

as $n \to \infty$.

Proof. We prove the claim for $\tilde{T}$. The proof for $\tilde{C}$ is analogous. Fix $t \in \mathbb{R}$ and let $z_i = 1\left[Y_i \cap (-\infty, t] \neq \emptyset\right]$ for $i = 1, 2, ...$ and $z = 1\left[Y \cap (-\infty, t] \neq \emptyset\right]$ with $E[z_i] = \tilde{T}_{Y_i}(t) = E[z] = \tilde{T}_Y(t)$.

By the Strong Law of Large Numbers we know that $\frac{1}{n}\sum_{i=1}^{n} z_i \overset{a.s.}{\to} z$. Define also the following random variables $w_i = 1\left[Y_i \cap (-\infty, t) \neq \emptyset\right]$ and $w = 1\left[Y \cap (-\infty, t) \neq \emptyset\right]$. $Ew_i = \mathbf{P}\left(Y_i \cap (-\infty, t)\right) = \lim_{a \nearrow t} \tilde{T}_{Y_i}(a) =: \tilde{T}_{Y_i}(t-)$. Since w and $w_i, i = 1, 2, ...$ are $i.i.d.$, $\tilde{T}_n(t-) \overset{a.s.}{\to} \tilde{T}_Y(t-)$ for each t. For an integer $k > 0$ and for j such that $1 \leq j \leq k-1$, let $a_{j,k} = \inf\left\{a : \tilde{T}_Y(a) \geq j/k\right\}$. We can pick an integer N_k such that for all $n \geq N_k$

$$\left|\tilde{T}_n(a_{j,k}) - \tilde{T}_Y(a_{j,k})\right| < k^{-1} \quad and \quad \left|\tilde{T}_n(a_{j,k}-) - \tilde{T}_Y(a_{j,k}-)\right| < k^{-1}$$

for $1 \leq j \leq k-1$. Take $a_{0,k} = -\infty$ and $a_{k,k} = \infty$ then the last two inequalities work for $j = 0$ and $j = k$ as well. For any $a \in \mathbb{R}$, we have $a_{j-1,k} < a < a_{j,k}$ for some $1 \leq j \leq k$ (if a is one of the end points we already have the above inequalities). For $n \geq N_k$, we have

$$\begin{aligned}
\tilde{T}_n(a) &\leq \tilde{T}_n(a_{j,k}-) \leq \tilde{T}_Y(a_{j,k}-) + k^{-1} \leq \tilde{T}_Y(a_{j-1,k}-) + 2k^{-1} \leq \tilde{T}_Y(a) + 2k^{-1} \\
\tilde{T}_n(a) &\geq \tilde{T}_n(a_{j-1,k}) \geq \tilde{T}_Y(a_{j-1,k}) - k^{-1} \geq \tilde{T}_Y(a_{j,k}-) - 2k^{-1} \geq \tilde{T}_Y(a) - 2k^{-1}.
\end{aligned}$$

Therefore, $\sup_a \left|\tilde{T}_n(a) - \tilde{T}_Y(a)\right| \leq 2k^{-1}$. This completes the proof. ■

This proposition shows that the sets

$$\begin{aligned}
\mathcal{T} &= \left\{t \mapsto 1\left[Y \cap (-\infty, t] \neq \emptyset\right] : t \in \mathbb{R}\right\} \\
\mathcal{C} &= \left\{t \mapsto 1\left[Y \subset (-\infty, t]\right] : t \in \mathbb{R}\right\}
\end{aligned}$$

are $\mathbf{P}$-Glivenko-Cantelli. In addition, by the CLT, for every $t \in \mathbb{R}$,

$$\begin{aligned}
\sqrt{n}\left(\tilde{T}_n(t) - \tilde{T}_Y(t)\right) &\overset{D}{\to} N\left(0, T_Y(t)\left(1 - T_Y(t)\right)\right) \\
\sqrt{n}\left(\tilde{C}_n(t) - \tilde{C}_Y(t)\right) &\overset{D}{\to} N\left(0, C_Y(t)\left(1 - C_Y(t)\right)\right)
\end{aligned}$$

We can further show the tightness of the cumulative capacity process and the cumulative containment process.

Proposition A.2 *$\mathcal{T}$ and $\mathcal{C}$ are $\mathbf{P}$-Donsker.*

Proof. We show the result for $\mathcal{T}$. The case of $\mathcal{C}$ is similar. For every $\varepsilon > 0$, as we saw in the proof for A.1, we can find a grid $-\infty < t_1 < t_2 < ... < t_k < \infty$ such that the bracket $[1[Y \cap (-\infty, t_{i-1}], 1[Y \cap (-\infty, t_i]]$ have L_1 bracket size which is at most ε. The number of grid points $k = k(\varepsilon)$ can be bounded by $2/\varepsilon$ to assure L_2 size of less than ε. This is true because the function $\tilde{T}_Y(t)$ cannot have more than $1/\varepsilon$ number of jumps which are larger than ε. Therefore, $N_{[]}(\varepsilon, \mathcal{T}, L_2(\mathbf{P})) \leq 2/\varepsilon$ and thus,

$$J_{[]}(\delta, \mathcal{T}, L_2(\mathbf{P})) \leq \int_0^\delta \sqrt{\log(2/\varepsilon)} d\varepsilon.$$

The last integral can be bounded by $C + \int_0^\delta \sqrt{-\log \varepsilon} d\varepsilon$ and for $\delta = 1$ the last integral equals $\frac{1}{2}\sqrt{\pi}$. Thus $J_{[]}(\delta, \mathcal{T}, L_2(\mathbf{P}))$ is finite, and the result follows by Theorem 19.5 of **?**. ■

A.2 Proof of Theorem 3.1

Proof. Since Y is closed and convex $\mathbf{P}$-a.s., we have $\Theta_0^Y(\tau) = [q_a(\tau), q_b(\tau)]$ by Theorems 2.1 and 2.2. Thus, $H\left(\hat{\Theta}^Y(\tau), \Theta_0^Y(\tau)\right) = \max\left\{\left|a_{(\lfloor n\tau \rfloor)} - q_a(\tau)\right|, \left|b_{(\lfloor n\tau \rfloor)} - q_b(\tau)\right|\right\}$. Since $a_{(\lfloor n\tau \rfloor)} - q_a(\tau) \overset{a.s.}{\to} 0$ and $b_{(\lfloor n\tau \rfloor)} - q_b(\tau) \overset{a.s.}{\to} 0$ under the i.i.d. sampling assumption, the claim is proved by the continuous mapping theorem. ■

A.3 The variance matrix $\tilde{\Sigma}(\tau)$

In this section, we give a complete expression for the $2(J+1) \times 2(J+1)$ matrix, which is a component of the asymptotic normal distribution for the $2(J+1)$-dimensional vector $\sqrt{n}(\tilde{a}_{(\lfloor n\tau \rfloor)} - q_{\tilde{a}}(\tau), \tilde{b}_{(\lfloor n\tau \rfloor)} - q_{\tilde{b}}(\tau), \widehat{\Pr}(a = 0) - \Pr(a = 0), ..., \widehat{\Pr}(a = J-1) - \Pr(a = J-1), \widehat{\Pr}(b = 0) - \Pr(b = 0), ..., \widehat{\Pr}(b = J-1) - \Pr(b = J-1))'$. See (3.10) in the main text.

$\tilde{\Sigma}(\tau)$ is a $2(J+1) \times 2(J+1)$ matrix

$$\text{(A.1)} \qquad \tilde{\Sigma}(\tau) = \begin{pmatrix} \multicolumn{2}{c}{\multirow{2}{*}{$\Sigma_{\tilde{a},\tilde{b}}(\tau)$}} & \Sigma_{\tilde{a},p_a}(\tau)' & \Sigma_{\tilde{a},p_b}(\tau)' \\ & & \Sigma_{\tilde{b},p_a}(\tau)' & \Sigma_{\tilde{b},p_b}(\tau)' \\ \Sigma_{\tilde{a},p_a}(\tau) & \Sigma_{\tilde{b},p_a}(\tau) & \Sigma_{p_a}(\tau) & \Sigma_{p_a,p_b}(\tau) \\ \Sigma_{\tilde{a},p_b}(\tau) & \Sigma_{\tilde{b},p_b}(\tau) & \Sigma_{p_a,p_b}(\tau)' & \Sigma_{p_b}(\tau) \end{pmatrix},$$

$\Sigma_{\tilde{a},\tilde{b}}(\tau)$ is a 2×2 matrix of the form

$$\Sigma_{\tilde{a},\tilde{b}}(\tau) = \begin{pmatrix} \frac{\tau(1-\tau)}{f_{\tilde{a}}(q_{\tilde{a}}(\tau))^2} & \frac{F_{\tilde{a},\tilde{b}}\left(q_{\tilde{a}}(\tau),q_{\tilde{b}}(\tau)\right)-\tau^2}{f_{\tilde{a}}(q_{\tilde{a}}(\tau))f_{\tilde{b}}\left(q_{\tilde{b}}(\tau)\right)} \\ \frac{F_{\tilde{a},\tilde{b}}\left(q_{\tilde{a}}(\tau),q_{\tilde{b}}(\tau)\right)-\tau^2}{f_{\tilde{a}}(q_{\tilde{a}}(\tau))f_{\tilde{b}}\left(q_{\tilde{b}}(\tau)\right)} & \frac{\tau(1-\tau)}{f_{\tilde{b}}\left(q_{\tilde{b}}(\tau)\right)^2}, \end{pmatrix}.$$

$\Sigma_{\tilde{a},p_a}(\tau)$ and $\Sigma_{\tilde{b},p_b}(\tau)$ is a $J \times 1$ are $J \times 1$ matrices of the forms

$$\Sigma_{\tilde{a},p_a}(\tau) = \left[\frac{\left(1\left\{a \leq q_a(\tau)\right\} - \tau\right)\Pr(a=j)}{f_{\tilde{a}}(q_{\tilde{a}}(\tau))}\right]_{j=0}^{J-1} \text{ and } \Sigma_{\tilde{b},p_b}(\tau) = \left[\frac{\left(1\left\{b \leq q_b(\tau)\right\} - \tau\right)\Pr(b=j)}{f_{\tilde{b}}(q_{\tilde{b}}(\tau))}\right]_{j=0}^{J-1},$$

$\Sigma_{\tilde{a},p_b}(\tau)$ and $\Sigma_{\tilde{b},p_a}(\tau)$ are a $K \times 1$ matrices of the forms

$$\Sigma_{\tilde{a},p_b}(\tau) = \left[\frac{\sum_{j'=0}^{q_a(\tau)} \Pr(a=j', b=j) - \tau\Pr(b=j)}{f_{\tilde{a}}(q_{\tilde{a}}(\tau))}\right]_{j=0}^{J-1} \text{ and}$$
$$\Sigma_{\tilde{b},p_a}(\tau) = \left[\frac{\sum_{j'=0}^{q_b(\tau)} \Pr(a=j, b=j') - \tau\Pr(a=j)}{f_{\tilde{b}}(q_{\tilde{b}}(\tau))}\right]_{j=0}^{J-1},$$

Σ_{p_a} and Σ_{p_b} are $J \times J$ matrices of the forms

$$\Sigma_{p_a} = \left[1\{j=j'\}\Pr(a=j) - \Pr(a=j)\Pr(a=j')\right]_{j=0,j'=0}^{J-1,J-1} \text{ and}$$
$$\Sigma_{p_a} = \left[1\{j=j'\}\Pr(b=j) - \Pr(b=j)\Pr(b=j')\right]_{j=0,j'=0}^{J-1,J-1},$$

and Σ_{p_a,p_b} is a $J \times J$ matrix of the form

$$\Sigma_{p_a,p_b} = \left[\Pr(a=j, b=j') - \Pr(a=j)\Pr(b=j')\right]_{j=0,j'=0}^{J-1,J-1}.$$

A.4 Proof of Lemma 3.1

Proof. Under conditions (ii), (iii), (v), and (vi), Theorem 3.3 of Chaudhuri (1991) provides the local Bahadur representation

$$\begin{aligned} & \sqrt{N_n(\tau \mid x^*)} f_{a|x}(q_{a|x}(\tau \mid x^*) \mid x^*) \left(\hat{a}(\tau \mid x^*) - a(\tau \mid x^*)\right) \\ = & \sqrt{\frac{f_x(x^*)}{\frac{1}{nh_{n1}(\tau|x^*)\cdots h_{np}(\tau|x^*)}\sum_{i=1}^n K\left(\frac{x_i - x^*}{h_n(\tau|x^*)}\right)}} \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^n V_{n,i}(\tau \mid x^*) + R_n^a(\tau \mid x^*), \end{aligned}$$

where

$$\begin{aligned} V_{n,i}(\tau \mid x^*) &= \frac{\tau - 1[a_i \leq q_{a|x}(\tau \mid x^*)]}{\sqrt{h_{n1}(\tau \mid x^*) \cdots h_{np}(\tau \mid x^*) \cdot f_x(x^*)}} \cdot K\left(\frac{x_i - x^*}{h_n(\tau \mid x^*)}\right) \\ R_n^a(\tau \mid x^*) &= O\left([\ln n]^{3/4} \cdot n^{(-\gamma+2p)/(4\gamma+2p)} \cdot h_{n1}(\tau \mid x^*)^{1/2} \cdots h_{np}(\tau \mid x^*)^{1/2}\right) = o(1) \end{aligned}$$

Likewise, under conditions (ii), (iv), (v), and (vi),

$$\begin{aligned} & \sqrt{N_n(\tau \mid x^*)} f_{b|x}(q_{b|x}(\tau \mid x^*) \mid x^*) \left(\hat{b}(\tau \mid x^*) - b(\tau \mid x^*)\right) \\ = & \sqrt{\frac{f_x(x^*)}{\frac{1}{nh_{n1}(\tau|x^*)\cdots h_{np}(\tau|x^*)}\sum_{i=1}^n K\left(\frac{x_i - x^*}{h_n(\tau|x^*)}\right)}} \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^n W_{n,i}(\tau \mid x^*) + R_n^b(\tau \mid x^*), \end{aligned}$$

where

$$\begin{aligned} W_{n,i}(\tau \mid x^*) &= \frac{\tau - 1[b_i \leq q_{b|x}(\tau \mid x^*)]}{\sqrt{h_{n1}(\tau \mid x^*) \cdots h_{np}(\tau \mid x^*) \cdot f_x(x^*)}} \cdot K\left(\frac{x_i - x^*}{h_n(\tau \mid x^*)}\right) \\ R_n^b(\tau \mid x^*) &= O\left([\ln n]^{3/4} \cdot n^{(-\gamma+2p)/(4\gamma+2p)} \cdot h_{n1}(\tau \mid x^*)^{1/2} \cdots h_{np}(\tau \mid x^*)^{1/2}\right) = o(1) \end{aligned}$$

Note that condition (vi) implies $h_{nk}(\tau \mid x^*) \to 0$ for each $k \in \{1, \ldots, p\}$ and $nh_{n1}(\tau \mid x^*) \cdots h_{np}(\tau \mid x^*) \to \infty$ as $n \to \infty$. Therefore, conditions (i), (ii), and (vi) yield

$$\frac{1}{nh_{n1}(\tau \mid x^*) \cdots h_{np}(\tau \mid x^*)} \sum_{i=1}^n K\left(\frac{x_i - x^*}{h_n(\tau \mid x^*)}\right) \xrightarrow{P} f_x(x^*).$$

Also, standard calculations show that

$$\begin{aligned} & EV_{n,i}(\tau \mid x^*) = 0, \qquad EW_{n,i}(\tau \mid x^*) = 0, \\ & Var(V_{n,i}(\tau \mid x^*)) = Var(W_{n,i}(\tau \mid x^*)) = \tau \cdot (1 - \tau) + o(1), \qquad \text{and} \\ & Cov(V_{n,i}(\tau \mid x^*), W_{n,i}(\tau \mid x^*)) = F_{a,b|x^*}(q_{a|x}(\tau \mid x^*), q_{b|x}(\tau \mid x^*)) - \tau^2 + o(1) \end{aligned}$$

under conditions (ii) and (vi). Furthermore, $\{V_{n,i}(\tau \mid x^*), W_{n,i}(\tau \mid x^*)\}$ is trivially shown to satisfy the Lindeberg-Feller property. Therefore, applying the Lindeberg-Feller CLT under condition (i), we obtain

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (V_{n,i}(\tau \mid x^*), W_{n,i}(\tau \mid x^*)) \xrightarrow{D} (z_V(\tau \mid x^*), z_W(\tau \mid x^*)) \sim N(\mathbf{0}, \Sigma_{VW}(\tau \mid x^*))$$

as $n \to \infty$, where

$$\Sigma_{VW}(\tau \mid x^*) = \begin{pmatrix} \tau(1-\tau) & F_{a,b|x^*}(q_{a|x}(\tau \mid x^*), q_{b|x}(\tau \mid x^*)) - \tau^2 \\ F_{a,b|x^*}(q_{a|x}(\tau \mid x^*), q_{b|x}(\tau \mid x^*)) - \tau^2 & \tau(1-\tau) \end{pmatrix}.$$

By Slutsky's Lemma, we obtain

$$\sqrt{N_n(\tau \mid x^*)} \left(\hat{a}(\tau \mid x^*) - a(\tau \mid x^*), \hat{b}(\tau \mid x^*) - b(\tau \mid x^*) \right) \xrightarrow{D} (z_L(\tau \mid x^*), z_U(\tau \mid x^*)) \sim N(\mathbf{0}, \Sigma(\tau \mid x^*))$$

as $n \to \infty$. ■

A.5 Procedure of Inference for Quantile Regression Parameters

This appendix section provides the procedure of inference for quantile regression parameters based on the conditional moment inequality restrictions derived in Section 4 via the method of Andrews and Shi (2013). Recall that, if Y is a closed convex value random set, then identification set $\Theta_0(\tau)$ of the quantile regression parameters θ is characterized by the conditional moment inequalities

$$E\left[m_j(w, \theta) \mid x\right] \geq 0 \quad \mathbf{P} - a.s. \quad \text{for } j = 1, 2,$$

where the moment functions, m_1 and m_2, are defined by

$$\begin{aligned} m_1(w, \theta) &= 1[y^l \leq q(x, \theta)] - \tau \qquad \text{and} \\ m_2(w, \theta) &= \tau - 1[y^u \leq q(x, \theta)] \end{aligned}$$

for $w = (x, y^l, y^u)$. Recall also that this set is convex if the quantile regression function $q(\,\cdot\,, \theta)$ is linear in parameters θ.

Normalize the vector x of p covariates into $[0,1]^p$, and define the sample moment functions and the sample variances by

$$\begin{aligned}
\bar{m}_{nj}(\theta, g) &= n^{-1} \sum_{i=1}^{n} m_j(w_i, \theta) \cdot g(x_i) \quad \text{for } j = 1, 2 \text{ and} \\
\hat{\sigma}^2_{nj}(\theta, g) &= n^{-1} \sum_{i=1}^{n} \left(m_j(w_i, \theta) \cdot g(x_i) - \bar{m}_{nj}(\theta, g) \right)^2 \quad \text{for } j = 1, 2,
\end{aligned}$$

for a function g to be defined below. To bound the sample variance away from zero, we use

$$\bar{\sigma}^2_{nj}(\theta, g) = \hat{\sigma}^2_{nj}(\theta, g) + \epsilon_n \cdot \hat{\sigma}^2_{nj}(\theta, 1) \quad \text{for } j = 1, 2.$$

With $g_{a,r}(x) = 1\left[x \in \prod_{u=1}^{p} \left(\frac{a_u - 1}{2r}, \frac{a_u}{2r} \right]\right]$, an approximated test statistic at θ is computed by

$$\bar{T}_{nR}(\theta) = \sum_{r=1}^{R} (r^2 + 100)^{-1} \sum_{a \in [1, \cdots, 2r]^p} (2r)^{-p} \left(\left[\frac{n^{\frac{1}{2}} \bar{m}_{n1}(\theta, g_{a,r})}{\bar{\sigma}_{n1}(\theta, g_{a,r})} \right]_-^2 + \left[\frac{n^{\frac{1}{2}} \bar{m}_{n2}(\theta, g_{a,r})}{\bar{\sigma}_{n2}(\theta, g_{a,r})} \right]_-^2 \right)$$

for some truncation number $R \in \mathbb{N}$, where $[x]_- = -x$ if $x < 0$ and $[x]_- = 0$ if $x \geq 0$.

Lemma 1 of Andrews and Shi (2013) guarantees that our definition of $\bar{T}_{nR}(\theta)$ satisfies Assumptions S1–S4 in their paper. Likewise, Lemma 3 of Andrews and Shi (2013) guarantees that the choice of $g_{a,r}$ defined above satisfies Assumptions CI and M in their paper. In order to assure that we can use the method of Andrews and Shi (2013), it remains to check their condition (2.3). The following conditions suffice: w is i.i.d.; $0 < \text{Var}(m_1(w, \theta)) < \infty$; $0 < \text{Var}(m_2(w, \theta)) < \infty$; $E|m_1(w, \theta)/\sigma_1(\theta)|^{2+\delta} < \infty$; and $E|m_2(w, \theta)/\sigma_2(\theta)|^{2+\delta} < \infty$; where $\delta > 0$, $\sigma_1(\theta) = \text{Var}(m_1(w, \theta))$, and $\sigma_2(\theta) = \text{Var}(m_2(w, \theta))$.

To compute the critical value for $\bar{T}_n(\theta)$, generate B bootstrap samples $\{w^*_{ib} : i = 1, \cdots, n\}$ for $b = 1, \cdots, B$. For each bootstrap sample $\{w^*_{ib} : i = 1, \cdots, n\}$, compute $\bar{m}^*_{nbj}(\theta, g)$ and $\bar{\sigma}^*_{nbj}(\theta, g)$ for $j = 1, 2$. For each bootstrap sample, compute the bootstrap test statistic

$$\begin{aligned}
\bar{T}^*_{nbR}(\theta) = \sum_{r=1}^{R} (r^2 + 100)^{-1} \sum_{a \in [1, \cdots, 2r]^p} (2r)^{-p} &\left(\left[\frac{n^{\frac{1}{2}} (\bar{m}^*_{nb1}(\theta, g_{a,r}) - \bar{m}_{n1}(\theta, g_{a,r}))/\hat{\sigma}_{n1}(\theta, 1) + \varphi_n(\theta, g_{a,r})}{\bar{\sigma}^*_{nb1}(\theta, g_{a,r})/\hat{\sigma}_{n1}(\theta, 1)} \right]_-^2 \right. \\
&\left. + \left[\frac{n^{\frac{1}{2}} (\bar{m}_{n2}(\theta, g_{a,r}) - \bar{m}_{n2}(\theta, g_{a,r}))/\hat{\sigma}_{n2}(\theta, 1) + \varphi_n(\theta, g_{a,r})}{\bar{\sigma}^*_{nb2}(\theta, g_{a,r})/\hat{\sigma}_{n2}(\theta, 1)} \right]_-^2 \right)
\end{aligned}$$

where $\varphi_{nj}(\theta, g)$ is given by

$$\varphi_{nj}(\theta, g) = B_n 1[\kappa_n^{-1} n^{\frac{1}{2}} \bar{m}_{nj}(\theta, g)/\bar{\sigma}_{nj}(\theta, g) > 1] \quad \text{for } j = 1, 2.$$

Andrews and Shi (2013) recommend $\epsilon_n = 0.05$, $\kappa_n = (0.3 \ln(n))^{\frac{1}{2}}$, and $B_n = (0.4 \ln(n)/ \ln \ln(n))^{\frac{1}{2}}$. The critical value $\bar{c}^*_{nRB,1-\alpha}(\theta)$ is set to be the $1 - \alpha + 10^6$ sample quantile of the bootstrap test statistics. Thus, a nominal level $1 - \alpha$ confidence set is approximated by

$$\left\{\theta \in \Theta : \bar{T}_{nR}(\theta) \leq \bar{c}^*_{nRB,1-\alpha}(\theta)\right\}.$$

Because finding the approximate region for this set is computationally burdensome when the dimension of the parameter set Θ is large, we provide in Section 5 an efficient algorithm to compute the estimate of the identification set.

A.6 Geometric Properties of the Set of Best Linear Predictors

In this section, we provide some geometric properties of the set of best linear predictors proposed in Section 5. It is shown that the set of connected, and therefore a projection of the set to each coordinate is interval-valued. To show this property, we go through several auxiliary lemmas. For the sake of rigorous proofs, we now formally define categorical sets and interval-valued sets below.

Definition A.1 *A random set $Y : \Omega \to \mathcal{K}(\mathbb{R})$ is categorical if we have either $Y(\omega) = Y(\omega')$ or $Y(\omega) \cap Y(\omega') = \emptyset$ for all pairs $\omega, \omega' \in \Omega$*

Definition A.2 *A random set $Y : \Omega \to \mathcal{K}(\mathbb{R})$ is interval-valued if $Y(\omega)$ is an interval for all $\omega \in \Omega$.*

For a random set Y, define the restricted set of selections

$$Sel_C(Y) = \left\{y \in Sel(Y) : F_y \text{ is continuous and strictly increasing on } F_y^{-1}((0,1))\right\}.$$

We first state the following auxiliary lemma of CDF equivalence for interval-valued categorical random sets.

Lemma A.1 *Let $y_0, y_1 \in Sel_C(Y)$ be two selections from an interval-valued categorical random set (cf. Definitions A.1 and A.2). For any $\omega \in \Omega$, we have $F_{y_0}(Y(\omega)) = F_{y_1}(Y(\omega))$.*

Proof. Define the set $\Omega_{L(\omega)} = \{\omega' \in \Omega : \sup Y(\omega') \leq \inf Y(\omega)\}$. Let $\tau \in F_{y_0}(Y(\Omega))$. By the strict increase of F_{y_0} for $y_0 \in Sel_C(Y)$, we can write $\tau = Pr(\{\omega' \in \Omega : y_0(\omega') \leq F_{y_0}^{-1}(\tau)\})$ where $F_{y_0}^{-1}(\tau) \in Y(\omega)$ by the definition of τ. Since $y_0 \in Sel_C(Y)$ and Y is an interval-valued categorical random set, we the obtain $\tau = Pr(\{\omega' \in \Omega : y_0(\omega') \leq F_{y_0}^{-1}(\tau)\}) \geq Pr(\Omega_{L(\omega)})$ by the monotone property of probability measures.

Assume by way of contradition that $\tau \notin F_{y_1}(Y(\omega))$. Since F_{y_1} is increasing and Y is interval-valued, this implies either $\tau < F_{y_1}(\zeta)$ for all $\zeta \in Y(\omega)$ or $\tau > F_{y_1}(\zeta)$ for all $\zeta \in Y(\omega)$. Without loss of generality, we consider the former case, which can be rewritten as $\tau < Pr(\Omega_{y_1}^{\zeta})$ for all $\zeta \in Y(\omega)$, where $\Omega_{y_1}^{\zeta} = \{\omega' \in \Omega : y_1(\omega') \leq \zeta\}$ for a short-hand notation. Consider a decreasing sequence $\{\zeta_n\}_{n=1}^{\infty} \in Y(\omega)$ such that $\zeta_n \to \inf Y(\omega)$. Since $y_1 \in Sel_C(Y)$, we have $\cap_{n=1}^{\infty} \Omega_{y_1}^{\zeta_n} = \Omega_{y_1}^{\inf Y(\omega)} \subset \Omega_{L(\omega)} \cup \{\omega' \in \Omega : y_1(\omega') = \inf Y(\omega)\}$ where the last set has a zero probability measure by the continuity of $y_1 \in Sel_C(Y)$. Apply the continuity theorem of probability measures to the decreasing sequence $\{\Omega_{y_1}^{\zeta_n}\}_{n=1}^{\infty}$, we obtain $\tau < \lim_{n\to\infty} Pr(\Omega_{y_1}^{\zeta_n}) \leq Pr(\Omega_{L(\omega)})$. Combining this result with the conclusion from the last paragraph, we obtain $\tau \geq Pr(\Omega_{L(\omega)}) > \tau$, a contradiction. Similarly, the case of $\tau > F_{y_1}(\zeta)$ for all $\zeta \in Y(\omega)$ leads to a contradiction. Therefore, $\tau \in F_{y_1}(Y(\omega))$ holds.

A symmetric argument by interchanging the roles of y_0 and y_1 shows that $\tau \in F_{y_1}(Y(\omega))$ implies $\tau \in F_{y_0}(Y(\omega))$. Therefore, $F_{y_0}(Y(\omega)) = F_{y_1}(Y(\omega))$ follows. ∎

Lemma A.2 *Let $y_0, y_1 \in Sel_C(Y)$ be two selections from an interval-valued categorical random set Y (cf. Definitions A.1 and A.2). Then, for any $\lambda \in [0,1]$, there exists $y_\lambda \in Sel_C(Y)$ such that $F_{y_\lambda} = (1-\lambda) \cdot F_{y_0} + \lambda \cdot F_{y_1}$.*

Proof. Define $y_\lambda : \Omega \to \mathbb{R}$ by $y_\lambda(\omega) = \inf\{y \in \mathbb{R} : F_{y_0}(y_0(\omega)) \leqslant (1-\lambda) \cdot F_{y_0}(y) + \lambda \cdot F_{y_1}(y)\}$. First, we show that y_λ is measurable. $F = (1-\lambda) \cdot F_{y_0} + \lambda \cdot F_{y_1}$ is continuous and strictly increasing on its support because $y_0, y_1 \in Sel_C(Y)$. Therefore, F has a strictly increasing inverse F^{-1} by Pfeiffer (1990, pp. 266), and it follows that $y_\lambda(\omega) = F^{-1} \circ F_{y_0} \circ y_0(\omega)$. Since F^{-1} and F_{y_0} are continuous and y_0 is measurable, it follows that y_λ is measurable.

Second, we show that y_λ is a selection of Y. Let $\omega \in \Omega$. Because $y_0 \in Sel_C(Y) \subset Sel(Y)$, we have $y_0(\omega) \in Y(\omega)$. Thus, we obtain $F \circ y_\lambda(\omega) = F_{y_0} \circ y_0(\omega) \in F_{y_0}(Y(\omega)) = F_{y_1}(Y(\omega))$, where the first equality is due to the definition of y_λ and the last equality is due to Lemma A.1. Taking a convex combination yields $F \circ y_\lambda(\omega) \in (1-\lambda)F_{y_0}(Y(\omega)) + \lambda F_{Y_1}(Y(\omega)) = F(Y(\omega))$. Therefore, it follows that $y_\lambda(\omega) \in Y(\omega)$, showing that y_λ is a selection of Y.

Finally, we show that $F_{y_\lambda} = (1-\lambda) \cdot F_{y_0} + \lambda \cdot F_{y_1}$. This claim follows from the following chain of equalities: $F_{y_\lambda}(y) = P\left(\{\omega \in \Omega : y_\lambda(\omega) \le y\}\right) = P\left(\left\{\omega \in \Omega : F^{-1} \circ F_{y_0}(y_0(\omega)) \le y\right\}\right) = P\left(\{\omega \in \Omega : F_{y_0}(y_0(\omega)) \le (1-\lambda) \cdot F_{y_0}(y) + \lambda \cdot F_{y_1}(y)\}\right) = (1-\lambda) \cdot F_{y_0}(y) + \lambda \cdot F_{y_1}(y)$, where the first equality is by the definition of the cdf F_{y_λ}, the second equality is due to the definition of y_λ, the third equality is by the short-hand notation for $F = (1-\lambda) \cdot F_{y_0} + \lambda \cdot F_{y_1}$, and the last equality uses $F_{y_0}(y_0) \sim U(0,1)$ by the probability integral transform. Therefore, we have $F_{y_\lambda} = (1-\lambda) \cdot F_{y_0} + \lambda \cdot F_{y_1}$. ■

We now consider the joint random set $(x, Y) : \Omega \to \mathbb{R}^p \times \mathcal{K}(\mathbb{R})$ and extend the restricted set $Sel_C(Y)$ of selections to

$$Sel_C(x,Y) = \Big\{(x,y) \in Sel(x,Y) : F_{y|x}(\,\cdot\, \mid \xi) \text{ is continuous and strictly increasing} \\ \text{on } F_{y|x}^{-1}((0,1) \mid \xi) \text{ for each } \xi \in \operatorname{Supp}(x)\Big\}.$$

Similar lines of a proof to those of Lemma A.2 yield the following extension to Lemma A.2.

Lemma A.3 *Suppose that x has a countable support. Let $(x, y_0), (x, y_1) \in Sel_C(x, Y)$ be two selections from an interval-valued categorical random set Y (cf. Definitions A.1 and A.2). Then, for any $\lambda \in [0,1]$, there exists $(x, y_\lambda) \in Sel_C(x, Y)$ such that $F_{y_\lambda|x} = (1-\lambda) \cdot F_{y_0|x} + \lambda \cdot F_{y_1|x}$.*

Proof. For each $\xi \in \operatorname{Image}(x)$, let $\Omega_\xi = x^{-1}(\{\xi\}) \subset \Omega$. Consider the restrictions $y_0 \mid_{\Omega_\xi}$, $y_1 \mid_{\Omega_\xi}$ and $Y \mid_{\Omega_\xi}$ of y_0, y_1 and Y, respectively, to the domain Ω_ξ. Because $\Omega_\xi = x^{-1}(\{\xi\})$ is a measurable subset of Ω due to the measurability of x, the restrictions $y_0 \mid_{\Omega_\xi}$ and $y_1 \mid_{\Omega_\xi}$ are also measurable functions. Furthermore, $y_0 \mid_{\Omega_\xi}$ and $y_1 \mid_{\Omega_\xi}$ are selection of $Y \mid_{\Omega_\xi}$ because they are restricted to the identical domain Ω_ξ. Therefore, by Lemma A.2, there exists a selection $y_{\xi,\lambda} : \Omega_\xi \to \mathbb{R}$ of $Y \mid_{\Omega_\xi}$ such that $F_{y_{\xi,\lambda}} = (1-\lambda) \cdot F_{y_0|x=\xi} + \lambda \cdot F_{y_1|x=\xi}$.

Define the function $y_\lambda : \Omega \to \mathbb{R}$ by the rule of assignment $y_\lambda(\omega) = y_{x(\omega),\lambda}(\omega)$. Further, define the function $(x, y_\lambda) : \Omega \to \mathbb{R}^{p+1}$ by the rule of assignment $(x, y_\lambda)(\omega) = (x(\omega), y_\lambda(\omega))$. We have $(x, y_\lambda)(\omega) = (x(\omega), y_\lambda(\omega)) = (x(\omega), y_{x(\omega),\lambda}(\omega)) \in \{x(\omega)\} \times Y(\omega)$ because the previous paragraph concluded that $y_{\xi,\lambda}$ is a selection of $Y \mid_{\Omega_\xi}$ for each $\xi \in \text{Image}(x)$. Because $y_{\xi,\lambda}$ is a measurable function for each $\xi \in \text{Image}(x)$ from the previous paragraph and $\text{Image}(x)$ is countable, this y_λ is a measurable function. But then, (x, y_λ) is also a measurable function. It also follows from the conclusion of the previous paragraph that $F_{y_\lambda|x=\xi} = F_{y_{\xi,\lambda}} = (1-\lambda) \cdot F_{y_0|x=\xi} + \lambda \cdot F_{y_1|x=\xi}$. These arguments together show that the desired conclusion holds. ■

The condition that the support of x is countable is restrictive, but is needed in our proof. This condition guarantees that y_λ is a measurable function. Without this condition, it is not clear if the same conclusion remains due to the fact that a sigma field is closed only under countable unions.

Now, for a random vector (x, y), we consider the best linear predictor β_τ defined by

$$\beta_\tau = \arg\min_{\beta \in B} E[\rho_\tau(y - x'\beta)]$$

where $\rho_\tau(u) = (\tau - 1[u \leq 0]) \cdot u$ and $B \subset \mathbb{R}^p$ is a convex and compact set.

Lemma A.4 *Suppose that x has a countable support. Let $(x, y_0), (x, y_1) \in Sel_C(x, Y)$ be two selections from an interval-valued categorical random set Y (cf. Definitions A.1 and A.2). If $E[\rho_\tau(y_0 - x'\beta)]$ and $E[\rho_\tau(y_1 - x'\beta)]$ are strictly convex in β, then, for any $\lambda \in [0, 1]$, there exists $(x, y_\lambda) \in Sel_C(x, Y)$ such that $E[\rho_\tau(y_\lambda - x'\beta)] = (1-\lambda)E[\rho_\tau(y_0 - x'\beta)] + \lambda E[\rho_\tau(y_1 - x'\beta)]$ is strictly convex.*

Proof. Lemma A.3 shows that there exists $(x, y_\lambda) \in Sel_C(x, Y)$ such that $F_{y_\lambda|x} = (1-\lambda) \cdot F_{y_0|x} + \lambda \cdot F_{y_1|x}$. Furthermore, $F_{y_\lambda|x} = (1-\lambda) \cdot F_{y_0|x} + \lambda \cdot F_{y_1|x}$ and the strict convexity of $E[\rho_\tau(y_0 - x'\beta)]$

and $E[\rho_\tau(y_1 - x'\beta)]$ in β imply that $E[\rho_\tau(y_\lambda - x'\beta)] = (1-\lambda)E[\rho_\tau(y_0 - x'\beta)] + \lambda E[\rho_\tau(y_1 - x'\beta)]$ is strictly convex. ■

Lemma A.5 *Suppose that x has a countable support. If $(x, y) \in Sel(x, Y)$ is compactly supported and admits a conditional density $f_{y|x}(\ \cdot\ \mid \xi) \in C^2$ for each $\xi \in Supp(x)$, then $E[\rho_\tau(y - x'\beta)]$ is twice continuously differentiable in β with*

$$\begin{aligned} \frac{\partial}{\partial \beta} E[\rho_\tau(y - x'\beta)] &= E\left[x\rho_\tau(y - x'\beta)\frac{\partial \log f_{x,y}(x,y)}{\partial y}\right] \quad \textit{and} \\ \frac{\partial^2}{\partial \beta \partial \beta'} E[\rho_\tau(y - x'\beta)] &= E\left[x\frac{\rho_\tau(y - x'\beta)}{f_{x,y}(x,y)}\frac{\partial^2 f_{x,y}(x,y)}{\partial y^2}x'\right] \end{aligned}$$

Proof. We first modify the check function ρ_τ by

$$\bar{\rho}_\tau(u) = \begin{cases} \rho_\tau(u) & \text{if } u = \zeta - \xi'\beta \text{ for some } (\xi, \zeta) \in \text{Supp}(x, y) \text{ and } \beta \in B \\ 0 & \text{otherwise} \end{cases}$$

With this modification, we have $\bar{\rho}_\tau \in L^1$ due to the compactness of $\text{Supp}(x, y)$ and B. We can write the BLP objective as

$$\begin{aligned} E[\rho_\tau(y - x'\beta)] &= \sum_{\xi \in \text{Supp}(x)} \int \bar{\rho}_\tau(\zeta - q(\xi, \beta)) f_{x,y}(\xi, \zeta) d\zeta \\ &= \sum_{\xi \in \text{Supp}(x)} \left(\bar{\rho}_\tau * f_{x,y}(\xi,\ \cdot\ )\right)(q(\xi, \beta)) \end{aligned}$$

where $q(x, \beta) = x'\beta$ and '$*$' denotes the convolution operator. Since q is clearly twice continuously differentiable with respect to β with its first and second derivatives given by x and 0, respectively, it suffices to show that $(\bar{\rho}_\tau * f_{x,y}(\xi,\ \cdot\ ))$ is twice continuously differentiable with its first and second derivatives given by $\bar{\rho}_\tau * \frac{\partial}{\partial y} f_{x,y}(\xi,\ \cdot\ )$ and $\bar{\rho}_\tau * \frac{\partial^2}{\partial y^2} f_{x,y}(\xi,\ \cdot\ )$, respectively. But this desired property follows from the fact that $f \in L^1$ and $g \in C^k$ implies $f * g \in C^k$ with $\partial^\alpha (f * g) = f * \partial^\alpha g$ for each $\alpha \in \{0, \cdots, k\}$, $\bar{\rho}_\tau \in L^1$, and our condition that (x, y) admits $f_{x,y}(\xi,\ \cdot\ ) = f_{y|x}(\ \cdot\ \mid \xi) f_x(\xi) \in C^2$ for each $\xi \in \text{Supp}(x)$. ■

We now define the set of best linear predictors by

$$B_{I,\tau} = \left\{ \arg\min_{\beta \in B} E[\rho_\tau(y - x'\beta)] : (x, y) \in Sel^*(x, Y) \right\}$$

for

$$Sel^*(x,Y) = \{(x,y) \in Sel(x,Y) : (x,y) \text{ satisfies Condition 1}\},$$

where the condition is given below.

Condition 1

(i) $F_{y|x}(\ \cdot \mid \xi)$ *is continuous and strictly increasing on* $F_{y|x}^{-1}((0,1) \mid \xi)$ *for each* $\xi \in Supp(x)$.

(ii) (x,y) *is compactly supported.*

(iii) (x,y) *admits a conditional density function* $f_{y|x}(\ \cdot \mid \xi) \in C^2$ *for each* $\xi \in Supp(x)$.

(iv) $E[\rho_\tau(y - x'\beta)]$ *are strictly convex in* β.

Proposition A.3 *Suppose that* x *has a countable support, and* Y *is an interval-valued categorical random set* Y *(cf. Definitions A.1 and A.2). If* $E[\rho_\tau(y - x'\beta)] = 0$ *holds for some* $\beta \in B$ *for each selection* $(x,y) \in Sel^*(x,Y)$*, then* $B_{I,\tau}$ *is connected. In particular, the projection of* $B_{I,\tau}$ *to each coordinate is interval-valued.*

Proof. Let $(x,y_0),(x,y_1) \in Sel^*(x,Y)$. By Lemma A.3, for any $\lambda \in [0,1]$, there exists $(x,y_\lambda) \in Sel_C(x,Y)$ such that $F_{y_\lambda|x} = (1-\lambda)\cdot F_{y_0|x} + \lambda \cdot F_{y_1|x}$. Condition 1 (ii) and (iii) are satisfied by such a selection (x,y_λ) due to $F_{y_\lambda|x} = (1-\lambda)\cdot F_{y_0|x} + \lambda \cdot F_{y_1|x}$. Furthermore, Lemma A.4 shows that such a selection (x,y_λ) also satisfies Condition 1 (iv). Therefore, $(x,y_\lambda) \in Sel^*(x,Y)$.

Let $U \supset B$ be an open subset of $\mathbb{R}^p$, $V = (0,1)$, and W be an open subset of $\mathbb{R}^p$. Define the function $\Psi : U \times (0,1) \to W$ by $\Psi(\beta,\lambda) = (1-\lambda)\frac{\partial}{\partial\beta}E[\rho_\tau(y_0 - x'\beta)] + \lambda\frac{\partial}{\partial\beta}E[\rho_\tau(y_1 - x'\beta)]$, which is guaranteed to exist by Lemma A.5. Note also that $F_{y_\lambda|x} = (1-\lambda)\cdot F_{y_0|x} + \lambda \cdot F_{y_1|x}$, $(x,y_\lambda) \in Sel^*(x,Y)$ and Lemma A.5 show that $\Psi(\beta,\lambda) = \frac{\partial}{\partial\beta}E[\rho_\tau(y_\lambda - x'\beta)]$ for each $\lambda \in [0,1]$.

First, observe that for each $\lambda \in V$ there is exactly one $\beta_\tau(\lambda) \in U$ satisfying $\Psi(\beta_\tau(\lambda),\lambda) = \vec{0}$ due to Lemma A.4. Second, the local solvability (i.e., the existence of a continuous explicit function $\beta_\tau(\lambda)$ at each $\lambda \in V$) follows from the implicit function theorem with Lemmas A.4 and A.5. Third, for each compact subset of $K \subset V = (0,1)$, $\lambda \in K$ implies that $\Psi(\beta_\tau(\lambda),\lambda) = \vec{0}$ holds for some $\beta_\tau(\lambda) \in B$ by the condition of the proposition. Therefore, by Theorem 1 of Sandberg (1981), the map $\lambda \mapsto \beta_\tau(\lambda)$ from $V = (0, 1)$ into U is continuous. Also, this continuity extends to the domain $[0, 1]$ by the definition of Ψ and Lemma A.5.

Therefore, $B_{I,\tau}$ is path-connected, and is therefore connected. ■

References

Andrews, D. W., and X. Shi (2013): "Inference Based on Conditional Moment Inequalities," Econometrica, 81(2), 609–666.

Angrist, J., V. Chernozhukov, and I. Fernández-Val (2006): "Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure," Econometrica, 74(2), 539–563.

Babu, J. G., and R. C. Rao (1988): "Joint Asymptotic Distribution of Marginal Quantiles and Quintile Functions in Samples from a Multivariate Population," Journal of Multivariate Analysis, 27(1), 15–23.

Beresteanu, A., I. Molchanov, and F. Molinari (2011): "Sharp Identification Regions in Models with Convex Moment Predictions," mimeo.

Beresteanu, A., I. Molchanov, and F. Molinari (2012): "Partial Identification Using Random Set Theory," Journal of Econometrics, 166(1), 17–32.

Beresteanu, A., and F. Molinari (2008): "Asymptotic Properties for a Class of Partially Identified Models," Econometrica, 76(4), 763–814.

Cameron, T. A., and D. D. Huppert (1989): "OLS Versus ML Estimation of Non-Market Resource Values with Payment Card Interval Data," Journal of Environmental Economics and Management, 17(3), 230–246.

Carlier, G., V. Chernozhukov, and A. Galichon (2016): "Vector Quantile Regression: an Optimal Transport Approach," Annals of Statistics, 44(3), 1165–1192.

Chaudhuri, P. (1991): "Nonparametric Estimates of Regression Quantiles and Their Local Bahadur Representation," Annals of Statistics, 19(2), 760–777.

Chernozhukov, V., A. Galichon, M. Hallin, and M. Henry (2017): "Monge-Kantorovich Depth, Quantiles, Ranks and Signs," Annals of Statistics, 45(1), 223–256.

Chesher, A. (2005): "Nonparametric Identification under Discrete Variation," Econometrica, 73(5), 1525–1550.

Chesher, A. (2010): "Instrumental Variable Models for Discrete Outcomes," Econometrica, 78(2), 575–601.

Chesher, A., and A. M. Rosen (2015): "Characterizations of Identified Sets delivered by Structural Econometric Models," CeMMAP Working Paper CWP63/15.

Gamper-Rabindran, S., and C. Timmins (2013): "Does Cleanup of Hazardous Waste Sites Raise Housing Values? Evidence of Spatially Localized Benefits," Journal of Environmental Economics and Management, 65(3), 345–360.

Hallin, M., D. Paindaveine, and M. Šiman (2010): "Multivariate Quantiles and Multiple-Output Regression Quantiles: From : L_1 Optimization to Halfspace Depth," Annals of Statistics, 38(2), 635–669.

Hong, H., and E. Tamer (2003): "Inference in Censored Models with Endogenous Regressors," Econometrica, 71(3), 905–932.

Kato, R., and Y. Sasaki (2017): "On Using Linear Quantile Regressions for Causal Inference," Econometric Theory, 33(3), 664–690.

Khan, S., M. Ponomareva, and E. Tamer (2011): "Sharpness in Randomly Censored Linear Models," Economics Letters, 113(1), 23–25.

Khan, S., and E. Tamer (2009): “Inference on Endogenously Censored Regression Models Using Conditional Moment Inequalities,” Journal of Econometrics, 152(2), 104–119.

Koenker, R. (2005): Quantile Regression, Vol. 38 of Econometric Society Monographs. Cambeidge University Press.

Koenker, R., and G. Bassett (1978): “Regression Quantiles,” Econometrica, 46(1), 33–50.

Kordas, G. (2006): “Smoothed Binary Regression Quantiles,” Journal of Applied Econometrics, 21(3), 387–407.

Li, T., and T. Oka (2015): “Set Identification of the Censored Quantile Regression Model for Short Panels with Fixed Effects,” Journal of Econometrics, 188(2), 363–377.

Machado, J. A., and J. S. Silva (2005): “Quantiles for Counts,” Journal of the American Statistical Association, 100(472), 1226–1237.

Manski, C. F. (1985): “Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator,” Journal of Econometrics, 27(3), 313–333.

Manski, C. F. (2003): Partial Identification of Probability Distributions. Springer Verlag, New York.

Manski, C. F., and E. Tamer (2002): “Inference on Regressions with Interval Data on a Regressor or Outcome,” Econometrica, 70(2), 519–546.

Molchanov, I. (2005): Theory of Random Sets. Springer Verlag, London.

Molchanov, I. S. (1990): “Empirical Estimation of Distribution Quantiles of Random Sets,” Theory of Probability and its Applications, 35(3), 594–600.

O’Garra, T., and S. Mourato (2007): “Public Preferences for Hydrogen Buses: Comparing Interval Data, OLS and Quantile Regression Approaches,” Environmental and Resource Economics, 36(4), 389–411.

Pfeiffer, P. E. (1990): Probability for Applications. Springer.

Powell, J. (1984): "Least Absolute Deviations Estimation for the Censored Regression Model," Journal of Econometrics, 53(3), 303–325.

Ramachandramurty, P., and M. S. Rao (1973): "Some Comments on Quantiles and Order Statistics," Canadian Mathematical Bulletin, 16(2), 289–293.

Rosen, A. M. (2012): "Set identification via Quantile Restrictions in Short Panels," Journal of Econometrics, 166(1), 127–137.

Sandberg, I. W. (1981): "Global Implicit Function Theorems," IEEE Transactions on Circuits and Systems, 28(2), 145–149.

van der Vaart, A. W. (1998): Asymptotic Statistics. Cambridge University Press.